



HAL
open science

Link prediction in dynamic and human-centered mobile wireless networks

Mohamed-Haykel Zayani

► **To cite this version:**

Mohamed-Haykel Zayani. Link prediction in dynamic and human-centered mobile wireless networks. Other [cs.OH]. Institut National des Télécommunications, 2012. English. NNT : 2012TELE0031 . tel-00787564

HAL Id: tel-00787564

<https://theses.hal.science/tel-00787564>

Submitted on 12 Feb 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Ecole Doctorale EDITE

**Thèse présentée pour l'obtention du diplôme de
Docteur de Télécom SudParis**

Doctorat conjoint Télécom SudParis et Université Pierre et Marie Curie

**Spécialité
Informatique**

**Par
Mohamed-Haykel Zayani**

**Titre
La Prédiction de Liens dans les Réseaux Sans-Fil Dynamiques Centrés sur
l'Être Humain**

Soutenue le 20 Septembre 2012 devant le jury composé de :

Président	Marcelo Dias de Amorim	Chargé de Recherche, LIP6/CNRS
Rapporteur	Hassnaa Moustafa	Ingénieur de Recherche Senior, Orange Labs
Rapporteur	André-Luc Beylot	Professeur, IRIT/ENSEEIH
Examineur	Pascale Minet	Chercheur HDR, INRIA Rocquencourt
Examineur	Farid Benbadis	Ingénieur Chercheur, Thales Communications
Directeur de thèse	Djamal Zeglache	Professeur, TELECOM SudParis
Co-directeur de thèse	Sami Tabbane	Professeur, SUP'COM Tunis
Encadrant	Vincent Gauthier	Maître de Conférences, TELECOM SudParis

Thèse n° 2012TELE0031



Doctoral School EDITE

**Thesis presented for obtaining the
Doctor of Philosophy Degree of Telecom SudParis**

Doctorate jointly delivered by Telecom SudParis and Pierre et Marie Curie University

**Speciality
Computer Science**

**Presented by
Mohamed-Haykel Zayani**

Title

Link Prediction in Dynamic and Human-Centered Mobile Wireless Networks

Defended in September, 20th 2012 in front of the committee:

President	Marcelo Dias de Amorim	Research Scientist, LIP6/CNRS
Reviewer	Hassnaa Moustafa	Senior Research Engineer, Orange Labs
Reviewer	André-Luc Beylot	Professor, IRIT/ENSEEIH
Examiner	Pascale Minet	Researcher, INRIA Rocquencourt
Examiner	Farid Benbadis	Research Engineer, Thales Communications
Director	Djamal Zeglache	Professor, TELECOM SudParis
Co-director	Sami Tabbane	Professor, SUP'COM Tunis
Advisor	Vincent Gauthier	Associate Professor, TELECOM SudParis

Thèse n° 2012TELE0031

Acknowledgments

I would like to thank my advisor Associate Professor Vincent Gauthier and my director Professor Djamel Zeglache for their unceasing support and dedication during all the time I spent working on this thesis. I am very grateful for the time they spent and the effort they have made to advise me and to guide my research works. I would like also to thank them for granting me the freedom of developing my ideas and for their suggestions and advices throughout this work. I really learned a lot by working with them and benefited from their experience as well as their excellent research and technical skills.

Special thanks also to my co-director Professor Sami Tabbane, my co-advisor Associate Professor Houda Khedher from SUP'COM Tunis and Doctor Ines Slama for their advices and implication in the progression of this thesis and for their continuous support to reach the objectives of the thesis.

More thanks go to my thesis evaluation committee consisting of Research scientist Marcelo Dias de Amorim, Researcher Hasnaa Moustafa, Professor André-Luc Beylot, Researcher Pascale Minet and Research Engineer Farid Benbadis.

Many thanks also to all my friends and colleagues, in no particular order, Bilel Romdhani, Jihen Ben Salem, Soumaya Ben Hassine Guetari, Mohamed Amine Guetari, Nouha Sghaier, Leila Ben Saad, Myriam Ben Saad, Ahmad Ahmad, Makhoul Hadji, Aroua Biri, Mohamed Ahmed Mohamed Sidi, Farouk Aissanou, Mazen Al Maarabani, Aymen Boudguiga, Anis Jdidi, Rachit Agarwal, Abhik Barnerjee, Khalifa Toumi, Emad Abderrahmane, Ahmed Soua, Imen Ben Lahmar, Dorsaf Zekri, Houssemed Medhioub, Marouene Mechtri, Wassim Drira, Tayyab Mehmood Mirza, Fabien Delestre and anyone who supported me to give the best of myself.

I wish to express my gratitude to Professor Tijani Chahed, Professor Walid Bennameur, Research Scientist Evrim Acar, Associate Professor Dimitrios Katsaros, Doctor Massimo Franceschet, Postdoctoral Researcher Pangun Park, Principal Scientist Sofie Pollin and Doctor Zhu Han for their help and support to concretize the contributions provided in this thesis.

Last and not least, I would like to thank wholeheartedly my mother Syhem, my father Hachemi and my brothers Mohamed Haythem, Mohamed Hachem and Mohamed Hichem for their continuous encouragement and support. I am deeply indebted to them. Also, I am thankful to God that by His grace and bounty, I am able to present this thesis.

Thanks a lot.

Abstract

During the last years, we have observed a progressive and continuous expansion of human-centered mobile wireless networks. The advent of these networks has encouraged the researchers to think about new solutions in order to ensure efficient evaluation and design of communication protocols. In fact, these networks are faced to several constraints as the lack of infrastructure, the dynamic topology, the limited resources and the deficient quality of service and security. We have been interested in the dynamicity of the network and in particular in human mobility. The human mobility has been widely studied in order to extract its intrinsic properties and to harness them to propose more accurate approaches. Among the prominent properties depicted in the literature, we have been specially attracted by the impact of the social interactions on the human mobility and consequently on the structure of the network. To grasp structural information of such networks, many metrics and techniques have been borrowed from the Social Network Analysis (SNA). The SNA can be seen as another network measurement task which extracts structural information of the network and provides useful feedback for communication protocols. In this context, the SNA has been extensively used to perform link prediction in social networks relying on their structural properties.

Motivated by the importance of social ties in human-centered mobile wireless networks and by the possibilities that are brought by SNA to perform link prediction, we are interested by designing the first link prediction framework adapted for mobile wireless networks as Mobile Ad-hoc Networks (MANETs) and Delay/Disruption Tolerant Networks (DTN). Our proposal tracks the evolution of the network through a third-order tensor over T periods and computes the sociometric Katz measure for each pair of nodes to quantify the strength of the social ties between the network entities. Such quantification gives insights about the links that are expected to occur in the period $T + 1$ and the new links that are created in the future without being observed during the tracking time. To attest the efficiency of our framework, we apply our link prediction technique on three real traces and we compare its performance to the ones of other well-known link prediction approaches. The results prove that our method reaches the highest level of accuracy and outperforms the other techniques. One of the major contributions behind our proposal highlights that the link prediction in such networks can be made in a distributed way. In other words, the nodes can predict their future links relying on the local information (one-hop and two-hop neighbors) instead of a full knowledge about the topology of the network.

Furthermore, we are keen to improve the link prediction performance of our

tensor-based framework. To quantify the social closeness between the users, we take into consideration two aspects of the relationships: the recentness of the interactions and their frequency. From this perspective, we wonder if we can consider a third criterion to improve the link prediction precision. Asserting the heuristic that stipulates that persistent links are highly predictable, we take into account the stability of the relationships (link and proximity stabilities). To measure it, we opt for the entropy estimation of a time series proposed in the Lempel-Ziv data compression algorithm. As we think that our framework measurements and the stability estimations complement each other, we combine them in order to provide new link prediction metrics. The simulation results emphasize the pertinence of our intuition.

Providing a tensor-based link prediction framework and proposing relative enhancements tied to stability considerations represent the main contributions of this thesis. Along the thesis, our concern was also focused on mechanisms and metrics that contribute towards improving communication protocols in these mobile networks. Nevertheless, our efforts were not limited to the major contributions previously mentioned and we had the opportunity to propose two other approaches that can be useful to improve the design and the evaluation of protocols in mobile multi-hop networks. Firstly, we propose a joint model for the IEEE 802.15.4 physical and medium access control layers. Secondly, we advance a self-learning repeated game framework, inspired by "The Weakest Link" TV game, to enforce cooperation in non-cooperative ad-hoc networks.

Keywords: Human-centered mobile wireless networks, link prediction, tensor, social ties, Katz measure, link and proximity stabilities, entropy.

Contents

1	Introduction	12
1.1	Constraints and Challenges in Mobile Multi-Hop Networks	12
1.1.1	Major Problems Encountered	13
1.1.2	The Human Mobility: Understanding it to Harness its Properties	14
1.2	Motivations and Contributions	15
1.3	Dissertation Organization	17
I	Proposition of a Link Prediction Framework and Metrics for Human-Centered Mobile Wireless Networks	20
2	Human-Centered Mobile Wireless Networks: Human Mobility in the Service of Link Prediction	21
2.1	Introduction	22
2.2	Mobile Wireless Networks	22
2.2.1	Mobile Ad-Hoc Networks	22
2.2.2	Delay Tolerant Networks	23
2.2.3	Opportunistic Networks	23
2.2.4	A Key Parameter for Mobile Wireless Networks: Human Mobility	23
2.3	Human Mobility	24
2.3.1	Towards Understanding Human Mobility	24
2.3.2	Encounter Traces	24
2.3.3	Human Mobility Properties	25
2.4	Social Network Analysis and Link Prediction	26
2.4.1	Overview on the Social Network Analysis	26
2.4.2	Link Prediction in Social Networks	27
2.5	Conclusion	32

3	Tensor-Based Link Prediction Framework for Mobile Wireless Networks	33
3.1	Introduction	33
3.2	Related Work	35
3.3	Description of the Tensor-Based Link Prediction Method	37
3.3.1	Notation	38
3.3.2	Matrix of Scores Computation	38
3.3.3	Matrix of Scores Interpretation	41
3.4	Performance Evaluation and Simulation Results	41
3.4.1	Simulation Traces	42
3.4.2	Performance Analysis	43
3.4.3	Discussion	55
3.5	Conclusion	57
4	Improving Link Prediction in Mobile Wireless Networks by Considering Link and Proximity Stabilities	59
4.1	Introduction	60
4.2	Related Work	61
4.3	Tensor-Based Link Prediction Framework: A Reminder	62
4.3.1	Notation	63
4.3.2	Overview on Tensor-Based Link Prediction Technique	63
4.3.3	Matrix of Scores Computation	64
4.3.4	Matrix of Scores Interpretation	65
4.4	How To Quantify Link and Proximity Stabilities and How To Use Them?	66
4.4.1	How Can Link and Proximity Stabilities Improve Link Prediction?	66
4.4.2	The Entropy Measure	67
4.4.3	Quantifying Link and Proximity Stabilities by Means of Time Series Entropy Estimation	70
4.4.4	Joining the Entropy Estimations to the Tensor-Based Link Prediction Framework	71
4.5	Simulations Scenarios and Performance Evaluation	75
4.5.1	Simulation Traces	75
4.5.2	Simulation Results and Performance Analysis	78
4.6	Conclusion	84

II	Other Contributions Proposed in the Thesis	85
5	A Joint Model for IEEE 802.15.4 Physical and Medium Access Control Layers	86
5.1	Introduction	86
5.2	Related Work	88
5.3	Developed IEEE 802.15.4 Model for Smart Grid	89
5.3.1	IEEE 802.15.4 PHY Model Description	89
5.3.2	Operation Details for the IEEE 802.15.4 MAC Model and the Interactions with the PHY Model	90
5.4	Simulation and Results	95
5.4.1	Cheking the Validity of the Joint Model	96
5.4.2	Comparison between Combined PHY and MAC Layers and Simple MAC Layer Models	96
5.4.3	Evolution of Node Performance with Growing Densities	98
5.5	Conclusion	99
6	Cooperation Enforcement for Packet Forwarding Optimization in Multi-hop Mobile Ad-hoc Networks	102
6.1	Introduction	102
6.2	Related Work	104
6.3	System Model And Problem Formulation	111
6.3.1	"The Weakest Link" TV Game Principle	111
6.3.2	The Proposed Model: Analogy with "The Weakest Link" TV Game	112
6.4	Self-Learning Repeated Game Framework and Punishment Mechanism	113
6.4.1	The Punishment Mechanism	115
6.4.2	The Self-Learning Repeated Game Framework Description	115
6.5	Performance Evaluation and Simulations Results	116
6.6	Conclusion	121
7	Conclusion	123
7.1	Contributions of the Thesis	123
7.1.1	Proposing a Tensor-Based Link Prediction Framework	123
7.1.2	Improving Link Prediction Efficiency by Considering Link and Proximity Stabilities	124
7.1.3	Other Works Towards the Improvement of the Evaluation and the Design of Communication Protocols in Human-Centered Wireless Networks	124

Contents

7.2	Future Research Directions	125
7.2.1	Implementing the Tensor-based Link Prediction Framework on Real Testbeds	125
7.2.2	Investigating the Contributions of Other Expressions for the Stability Measure	126
7.2.3	Exploiting the Joint Model for IEEE 802.15.4 PHY and MAC Layers for Proposing New Performance Metrics	127
7.2.4	Enhancing the Self-Learning Repeated Game	127
A	Thesis Publications	128
B	Résumé en Français	130
B.1	Introduction	130
B.2	Proposition de Méthode et de Métriques pour la Prédiction des liens dans les Réseaux Mobiles Sans-Fil Centrés sur l'Aspect Humain	131
B.2.1	Mobilité Humaine, Analyse des Réseaux Sociaux et Prédiction des Liens	131
B.2.2	Méthode de Prédiction des Liens Basée sur les Tenseurs pour les Réseaux Sans-Fil Mobiles	133
B.2.3	Amélioration de la Méthode de Prédiction des Liens en Con- sidérant les Stabilités de Lien et de Proximité	140
B.3	Autres Contributions Avancées dans la Thèse	146
B.4	Conclusion	147
	Bibliography	149

List of Figures

3.1	Example of the matrix \mathbf{S} computation	39
3.2	ROC Curves for different prediction cases applied on Dartmouth Campus trace	46
3.3	ROC Curves for different prediction cases applied on MIT Campus trace	47
3.4	ROC Curves for different prediction cases applied on Infocom 2006 trace	48
3.5	Prediction performance of the tensor-based technique distributed version for different values of β and θ	54
4.1	Reminder of the steps followed for matrix \mathbf{S} computation	65
4.2	Example of an application of the algorithm used to identify the shortest substrings in the Lempel-Ziv data compression mechanism	72
4.3	Evaluation metrics for the prediction applied on the MIT Campus trace for different tracking periods	76
4.4	Evaluation metrics for the prediction applied on the Dartmouth Campus trace for different tracking periods	77
4.5	Evaluation metrics for the prediction applied on the Infocom 2006 trace for different tracking periods	78
4.6	Evaluation metrics for the prediction applied on the MIT Campus trace for different tracking periods and with knowledge limited to direct neighbors	82
5.1	IEEE 802.15.4 PHY and MAC model flowchart	93
5.2	Comparison between IEEE 802.15.4 PHY & MAC Model and IEEE 802.15.4 MAC Model	98
5.3	Zoom in on the average wait time versus load offered for IEEE 802.15.4 PHY & MAC and IEEE 802.15.4 MAC models	99

List of Figures

5.4	Performances evolutions for different densities using IEEE 802.15.4 PHY & MAC Model	100
5.5	Zoom in on the average wait time versus load offered for different densities using IEEE 802.15.4 PHY & MAC Model	101
6.1	The self-learning repeated game flowchart proposed by Han et al. . .	107
6.2	The self-learning repeated game flowchart	114
6.3	The evolution of the average forwarding probability for different self-learning repeated game schemes in the case of the ring network . . .	117
6.4	The evolution of the average forwarding probability for different self-learning repeated game schemes in the case of a random network . .	118
B.1	Exemple de calcul de la matrice \mathbf{S}	135
B.2	Courbes ROC pour les différents cas de prédictions de liens appliqués sur la trace du campus de Dartmouth	138
B.3	Métriques d’Evaluation pour Différentes Techniques de Prédiction Appliquées sur la trace du Campus de MIT	144

List of Tables

3.1	Table of confusion of a binary prediction technique	44
3.2	Evaluation metrics for the prediction of all links applied on Dartmouth Campus trace	50
3.3	Evaluation metrics for the prediction of all links applied on MIT Campus trace	51
3.4	Evaluation metrics for the prediction of all links applied on Infocom 2006 trace	52
3.5	Evaluation metrics for the prediction of all links in $T+1$ applied on Infocom 2006 trace	56
4.1	Settings for each trace to assess the efficiency of the SE_{2hops} score in improving the prediction of new links	83
4.2	Top Scores Ratio in the future: comparison between the efficiency of the SE_{2hops} score and the one of the Katz measure in predicting new links	83
5.1	Main Parameters used in MAC Layer	95
5.2	Main Parameters used in PHY Layer	95
5.3	Comparison of the values for α and β obtained through the joint model and the one of Pollin et al. according to a growing density with saturated traffic	97
6.1	Evaluation metrics in the case of the ring network (all forwarding probabilities initialized to 0)	119
6.2	Evaluation metrics in the case of a random network (all forwarding probabilities initialized to 0)	119
6.3	Evaluation metrics in the case of the ring network (all forwarding probabilities initialized to 0.5)	120

List of Tables

6.4	Evaluation metrics in the case of the random network (all forwarding probabilities initialized to 0.5)	120
B.1	Métriques d’Evaluation pour la Prédiction de tous les Liens Appliquée sur la Trace du Campus de Dartmouth	139
B.2	Paramètres de Simulation pour la Vérification de l’Efficacité de la Métrique SE_{2hops} score pour l’Amélioration de la Prédiction des Nouveaux Liens	145
B.3	Ratio des Top Scores dans le Futur: Comparaison entre la performance de la métrique SE_{2hops} score et celle de la mesure de Katz dans la prédiction des nouveaux liens	146

Chapter 1

Introduction

The advent of mobile wireless networks, in last decades, has encouraged researchers to think about new solutions to ensure communication between the users of the network. The deployment of a centralized architectures can be costly which limits mobile wireless networks to densely populated area. As devices are carried by humans and the mobility is not restricted to the deployment areas, alternatives are needed to ensure the connectivity. To address this problem, Mobile Ad-hoc Networks (MANETs) [33] and Delay/Disruption Tolerant Networks (DTNs) [43, 42] have been proposed.

In ad-hoc networks and especially in MANETs, each pair of nodes can be connected by a wireless link or by multi-hop paths. Then, source and destination nodes can be out of range of each other and even though communicate. The network topology can evolve over time as nodes are expected to move. Nonetheless, the MANET routing protocols tend to maintain end-to-end paths between all the nodes. This property highlights the difference with opportunistic networks as the DTNs. Indeed, in DTNs, it is assumed that no end-to-end paths exist and forwarding a packet from a source to a destination relies only on the mobility of the nodes and the contacts occurred between them.

These networks are then characterized by a multi-hop communication to maintain connectivity between nodes and are faced to several constraints.

1.1 Constraints and Challenges in Mobile Multi-Hop Networks

Mobile ad-hoc and delay tolerant networks are faced to specific constraints on top of those related to wireless communication. In the following, we try to depict an overview on some constraints observed in such networks and we highlight the topo-

logical evolution due to the human mobility.

1.1.1 Major Problems Encountered

Some major problems that these networks can be faced with are summarized in the following points:

- **Lack of centralized infrastructure:** as previously stated, the multi-hop networks are not based on a centralized infrastructure. The network entities have to communicate in a distributed manner (end-to-end or opportunistic communication) to ensure the connectivity.
- **Dynamic topology:** the nodes in such networks are expected to move, to appear or to disappear. Hence, the topology is rather dynamic than static and is brought to evolve as time goes on. Therefore, it is important to understand this dynamicity and especially when it is tied to human mobility. Identifying structural information of the network through the human behavior can be very useful in order to accurately evaluate and design communication protocols for human-based networks.
- **Resource limitation:** generally, the nodes that form the network have limited resources in terms of energy, processing capacity and memory which may induce negative impacts on guaranteeing the delivery of messages. Then, these nodes may be non-cooperative in order to preserve their resources for their own benefit. Consequently, the communication protocols have to be efficient and adapted to these limitations in order to maximize the lifetime of the networks and its entities.
- **Quality of service and security:** the quality of service can be needed in order to ensure a fair access to the shared resources. The security also is an important aspect to take into consideration given that the collaboration to carry the information is distributed and the the wireless communication leads to unintentional accessibility (message broadcast).

The topology of such networks, as mentioned, is expected to be dynamic. This evolution can be explained by the appearance of new nodes in the networks, the disappearance of some nodes due to failures or resource limitations or the mobility pattern that characterize people who carry the nodes. This last detail has caught the attention of many researches and it has been shown that the human mobility has high impacts on the network and consequently on its performance [29, 59]. We give some details about this human mobility in the following section.

1.1.2 The Human Mobility: Understanding it to Harness its Properties

Human mobility is considered as an important property of human-centered mobile wireless networks [29]. Therefore, understanding its intrinsic properties is imperative to evaluate and/or to design communication protocols. For this purpose, many initiatives have targeted to collect information about the moves of several users in a network into traces [5, 16, 52, 106, 58, 39, 66, 95, 72, 29, 82]. The data collected gives indications of the locations visited by the users, the time they spend in these locations and sometimes details about their encounters. These traces have been used in many studies and some properties have been pointed out. We cite the skewed location visiting preferences [57], the time-dependent mobility behavior [57], the spatio-temporal regularity [16, 52, 106, 39, 77], the power-law property of the contact and inter-contact times between nodes [58, 29, 63] and the high predictability of human mobility [101]. Some other works have demonstrated that the human mobility is influenced by the social interactions that exist between the network entities and is directed by social intentions [39, 78, 105]. These findings attest that the users of the networks are strongly interdependent and their interactions govern the structure of the network. Hence, stressing on the social ties in such networks will considerably contribute in understanding how the network topology evolves. Consequently, several works have borrowed concepts from the Social Network Analysis (SNA) [111] to harness such a feedback for designing communication protocols [35, 59] or proposing new techniques to improve these protocols [110].

The use of the Social Network Analysis has been favored by the lack of infrastructure and the opportunistic aspect that can characterize the communication between nodes. The SNA can be seen as another network measurement task as stressed in [64] which focus on analyzing the relationships between the network entities et interpreting the findings to extract network properties as the network robustness, the most central nodes or the emerging communities. These properties describes the structural information of the network which can be used to design communication protocols as [35, 59] or to perform link prediction in GSM networks as made in [110].

Having investigated several works based on SNA and particularly those aiming at performing link prediction in social networks [8, 38, 110] and believing that it is possible to harness the social interactions in human-centered mobile wireless networks, one of our motivations in this thesis aims at proposing a framework which is able to perform link prediction in such networks. In the following section, we detail in deep the major motivations that brought us to provide the contributions of this thesis

1.2 Motivations and Contributions

In last decades, mobile multi-hop wireless networks and especially those which are human-centered have been widely studied and many approaches have been proposed in order to understand the properties of these networks and propose adapted communication protocols.

As discussed in the previous section, the topology in these network is dynamic and the human mobility has been highlighted as one of the key parameters which has an important impact on the structure of the network and consequently on the performance of the communication protocols [29, 59]. Also, it has been shown that the human mobility is dependent on the interactions between the nodes carried by the users which form a social network [39, 78]. Therefore, it is crucial to understand the properties for the human mobility to create more realistic mobility models to mimic real human-centered wireless networks and to design more efficient communication protocols. For this aim, many researchers have borrowed centrality metrics [111, 84, 60] and community detection techniques [111, 20, 85, 81] of the Social Network Analysis (SNA) [111] to propose, for example, new designs for communication protocols [35, 59].

In data-mining, the Social Network Analysis tool has been used to introduce the link prediction problem [74]. Moreover, many other works [109, 8, 38, 15, 98] have highlighted the effectiveness of link prediction in social networks as the co-authorship networks [4, 6] or social network websites as Facebook. The link prediction concept has been used in mobile phone networks [110] and has shown interesting results.

These findings have motivated us to design a **link prediction framework** which can be used in MANETs and in DTNs. To do so, we propose the following salient steps for our framework inspired by the data-mining works [8, 38]. First, it tracks the evolution of the topology or the neighborhood through a third-order tensor (a set of adjacency matrices tracking contacts over successive time periods). Then, it collapses the collected data by allocating a weight to each link according to its lifetime and its recentness. Finally, it applies the Katz measure [65] on the collapsed data to obtain prediction scores that express the strength of the social ties for each pair of nodes. For the best of our knowledge, we are the first to propose such a link prediction technique for mobile multi-hop wireless networks. Moreover, instead of data-mining approaches which assume that there is a central entity which has full knowledge of the topology evolution, we propose to perform link prediction in a distributed way. That is to say that each node constructs its own tensor and performs link prediction relying on local information (the detected neighbors) which is more adapted to the context of mobile wireless networks.

1.2. Motivations and Contributions

Furthermore, we aim to **improve the performance of our link prediction framework** by considering a new criterion that characterizes the interactions between nodes. In our framework, we have designed link prediction using the recency and the lifetime of contacts and we wonder if the **stability of the interactions** can be useful for our task. We derive this intuition from the following heuristic: if two nodes have been related through a stable link during a tracking period, this link is expected to be persistent in the future. To fulfil this requirement and to quantify stability, we use the entropy estimator used in Lempel-Ziv data compression [127] as it estimates well the entropy of a time series. Afterwards, we propose to combine the outputs of our link prediction framework with the stability quantification in order to **promote new entropy-based link prediction metrics**.

In this thesis, we have been concerned by providing an **efficient link prediction technique** for these mobile networks as we strongly believe that the feedback provided by this framework can be **helpful** for communication protocols. In MANETs, link prediction can be used to detect the robustness of the routes and estimate the time needed to refresh routing tables. For DTNs, where communication is rather opportunistic, link prediction can be beneficial in detecting possible future interactions which may enhance the message delivery.

Providing a tensor-based link prediction framework and proposing relative enhancements tied to stability considerations represent the **main contributions** of this thesis. As we have emphasized, our concern has focused on **mechanisms and metrics that contribute towards improving communication protocols** in these mobile networks. Nevertheless, our concern has not been limited to this major contribution and we have had the opportunity to propose **two other approaches** that can be useful to improve the design and the evaluation of protocols in mobile multi-hop networks.

Firstly, we have been interested in the IEEE 802.15.4 standard [2]. Many works have been proposed to model the physical layer (PHY layer) [128, 116] and the medium access layer (MAC layer) [89, 87]. These models have been designed to mimic the processing in the PHY and MAC layers and to estimate performance indicators as the probability of a packet loss. In the literature, the models have been distinct: they are reproducing either the PHY layer or the MAC layer. From this perspective, we have been motivated to provide a **comprehensive model** which is able to mimic more faithfully the functionalities of the standard by considering **both PHY and MAC layers of IEEE 802.15.4**. We propose a combination of two relevant models for the two layers. The PHY layer behavior is reproduced by a mathematical framework, which is based on radio and channel models, in order to quantify link reliability [128, 116]. Regarding the MAC layer, it is mimed by an

enhanced Markov chain [87]. This contribution allows us fully and more precisely to estimate the network performance with different network sizes, as well as different metrics such as node reliability and delay. Our contribution enables us to track possible failures at both layers.

Secondly, we have pointed out that the nodes in multi-hop networks and especially for MANETs have limited energy resources. Thus, a node tends to behave selfishly when it is asked to forward the packets of other nodes. Indeed, it would rather reject a forwarding request in order to save its energy. To overcome this problem, the nodes need to be motivated to cooperate. To this end, we propose a **self-learning repeated game framework**, as designed in [51, 86], to enforce cooperation between the nodes of a network. This framework is inspired by the concept of "The Weakest Link" TV game. Each node has a utility function whose value depends on its cooperation in forwarding packets on a route as well as the cooperation of all the nodes that form this same route. The more these nodes cooperate the higher is their utility value. This would establish a cooperative spirit within the nodes of the networks. All the nodes will then more or less equally participate to the forwarding tasks which would then eventually guarantee a more efficient packets forwarding from sources to respective destinations. In the same time, defecting nodes are automatically punished which causes a dramatic decrease of their benefits and encourages them to be more cooperative.

1.3 Dissertation Organization

The rest of the dissertation is organized in two major parts.

The first part is dedicated to the tensor-based link prediction framework and its relative improvements. It is built around three chapters.

In Chapter 2, we explain how human mobility is crucial for human-centered mobile wireless networks. To do so, we give an overview on the mobile wireless networks based on human interactions and we emphasize the importance of human mobility as a key parameter which may impact the structure of the network and consequently its performance. Then, we depict the major properties of the human mobility presented in the literature and we stress how social interactions between the users influence their mobility patterns. From this perspective, we present the Social Network Analysis (SNA) which concepts and mechanisms have been borrowed to extract structural information about the network. We point out the implication of SNA in data-mining and especially in performing link prediction in social networks and our motivation to address the problem of link prediction in the case of human-centered mobile wireless networks.

Then, we present in Chapter 3 a tensor-based link prediction framework devoted to our context, human-centered wireless networks. We detail in this chapter the salient steps followed by this framework in order to perform link prediction: collecting the contact occurrences during an historical, collapsing the data and applying a sociometric measure which is the Katz measure [65] for each link. The obtained link scores are used to interpret if a link is expected to occur in the future. After presenting our link prediction framework, we assess its efficiency by performing link prediction on three real traces and comparing its performance to other well-known link prediction techniques.

In Chapter 4, we highlight our motivation to improve the performance of our tensor-based link prediction framework. We emphasize our belief that measuring the stability of interactions between nodes is helpful to enhance the prediction performance of our framework. Indeed, we explain how quantifying such a stability can improve link prediction efficiency and how we design measuring such a parameter with applying an entropy estimator used in Lempel-Ziv data compression [127]. Afterwards, we advance some combinations between the outputs of our framework and the stability estimation to express new link prediction measures. In order to verify the contribution of our proposal and as done in the previous chapter, we evaluate our advanced metrics using real traces and we compare them to some other well-known link prediction metrics.

Regarding the second part, it consists of two chapters in which we present two other contributions proposed in the thesis. It expresses our concern to provide other techniques and metrics towards improving the evaluation and the conception of communication protocols in mobile multi-hop wireless networks.

In Chapter 5, we depict a joint model for IEEE 802.15.4 physical (PHY) and medium access control (MAC) layers for the Smart Grid project. We detail the approaches that have inspired us to design the PHY and MAC layers and how we join them to obtain a more comprehensive model. The joint model considers errors at both levels which enables us to reach better precision in estimating network performance. We present also some simulation scenarios to point out the precision provided by our joint model compared to a relevant model [87] that mimics the MAC layer only.

In Chapter 6, we present a self-learning repeated game dedicated to non-cooperative MANETs in order to maintain the network connectivity. We describe the system model based on the concept of the famous "Weakest Link" TV game and the problem formulation. Then, we explain why we have opted for a repeated game and how we have employed a self-learning mechanism joined to punishment threats. Moreover, we evaluate our proposal by comparing it to two other self-learning repeated

Chapter 1. Introduction

games proposed in the literature [51, 86].

Finally, we conclude this dissertation, in Chapter 7, by summarizing the contributions of this thesis and proposes future directions of research.

We provide all the thesis publications in Appendix A. This dissertation is written in English and for the sake of accessibility, we provide a summary in French language in Appendix B.

Part I

Proposition of a Link Prediction Framework and Metrics for Human-Centered Mobile Wireless Networks

Chapter 2

Human-Centered Mobile Wireless Networks: Human Mobility in the Service of Link Prediction

Contents

2.1	Introduction	22
2.2	Mobile Wireless Networks	22
2.2.1	Mobile Ad-Hoc Networks	22
2.2.2	Delay Tolerant Networks	23
2.2.3	Opportunistic Networks	23
2.2.4	A Key Parameter for Mobile Wireless Networks: Human Mobility	23
2.3	Human Mobility	24
2.3.1	Towards Understanding Human Mobility	24
2.3.2	Encounter Traces	24
2.3.3	Human Mobility Properties	25
2.4	Social Network Analysis and Link Prediction	26
2.4.1	Overview on the Social Network Analysis	26
2.4.2	Link Prediction in Social Networks	27
2.5	Conclusion	32

2.1 Introduction

In this chapter, we present the state of the art about the studies that have focused on the human mobility in human-centered mobile wireless networks (MANETS, DTNs, ...). The human mobility in this context is a key parameter as it has an impact on the structure of the network. Understanding the properties of the human mobility is crucial to evaluate and improve communication protocols in such networks. For this aim, many techniques used in Social Network Analysis (SNA) have been borrowed by the researchers. The information extracted through these techniques has notably contributed in providing solutions and insights for the link prediction in social networks problem. These findings widely used in data-mining have motivated us to propose our tensor-based link prediction in mobile wireless networks. In this chapter, we detail how human mobility impacts on an underlying social network and how to exploit the Social Network Analysis. The SNA enables us to extract intrinsic properties of the network and we use them to perform link prediction.

2.2 Mobile Wireless Networks

We propose in this section to study various mobile ad-hoc networking architectures. There are three main categories and each category is defined by the mobility of the nodes and by the network dynamicity. The three categories are:

- Mobile ad-hoc networks (MANETs).
- Delay or disruption tolerant networks (DTNs).
- Opportunistic networks.

2.2.1 Mobile Ad-Hoc Networks

Wireless networks can be classified into two categories: infrastructure-based networks and ad-hoc networks (Sensor networks, MANETs, DTNs). When the network is deployed with an infrastructure, the nodes communicate with one or many base stations. The set of all base stations (such as access points) are connected to a backbone. On the other hand, the ad-hoc networks have no infrastructure and are totally decentralized. Indeed, the networks consists of autonomous mobile nodes that are connected through wireless links and which may play the role of routers to ensure communication between any pair of source-destination [33].

Chapter 2. Human-Centered Mobile Wireless Networks: Human Mobility in the Service of Link Prediction

In ad-hoc networks and especially in MANETs, each pair of nodes can be connected by a wireless link or by multi-hop paths. Then, source and destination nodes can be out of range of each other and even though communicate. The network topology can evolve over time as nodes are expected to move. Nonetheless, the MANET routing protocols tend to maintain end-to-end paths between all the nodes. This property highlights the difference with opportunistic networks where it is assumed that no end-to-end paths exist and forwarding a packet from a source to a destination relies only on the mobility of the nodes and the contacts occurred between them.

2.2.2 Delay Tolerant Networks

The Delay or Disruption Tolerant Network (DTN) architecture is a paradigm that has been advanced in order to provide services for challenging networks [43] (ensuring connectivity in rural areas [37, 96], military networks [93], underwater networks [88], ...). The DTNs are deployed in environments where the connectivity is intermittent or scheduled and may undergo high error rates, long delays and low data rates. These networks cannot maintain end-to-end paths between nodes due to the mentioned constraints. Nevertheless, forwarding a packet from a source to a destination in these networks relies on the store-and-forward multi-hop routing based on the occurrence of contacts between intermediate nodes.

2.2.3 Opportunistic Networks

Opportunistic networks have the same characteristics as DTNs in terms of delay and disruption tolerance but designed with the consideration of more unpredictable mobility (the connectivity is not really scheduled) [42]. From this perspective, the opportunistic networks can be viewed as a generalization of DTNs. The Pocket Switched Networks (PSN) [58] represent an example of opportunistic networks.

2.2.4 A Key Parameter for Mobile Wireless Networks: Human Mobility

With the growth of these mobile networks and the sophistication of portable computation and communication devices, the human mobility becomes a paramount phenomenon that impacts the structure of the network as shown in [29, 59]. Hence, it is important to understand the human mobility to ensure accurate evaluations and analyses of the performance of communication protocols and applications. In the following section, we give a more deep information about human mobility and list its major properties detailed in the literature.

2.3 Human Mobility

Human mobility represents a key property of mobile wireless networks. Then, it is crucial to study it in order to understand its intrinsic characteristics and exploit these findings to design and/or evaluate communication mechanisms in such networks. In this section, we provide a related work about analyzing human mobility and designing models that mimic the human behavior.

2.3.1 Towards Understanding Human Mobility

To reproduce human mobility, two first individual models have been proposed: the Random Walk or Brownian motion [41] and the Random Waypoint model [62]. Nodes that use Random Walk move, at each step, with a speed chosen from a distribution towards a random position on a moving area. When a node reaches its destination, it fixes a new speed and picks new random position to which it have to move and the process is repeated. When the Random Waypoint model is considered, the same process as Random Walk is reproduced with adding pause times between each step. Many other mobility models that are similar to two cited approaches are detailed in [26, 79]. There are also group mobility models proposed in the literature, such as [53], where the movement of nodes are correlated to the one of the clusterhead.

Such models have been widely used in simulations to attest the efficiency of communication protocols in the context of intermittent connections. Despite the popularity of such synthetic models, it has been shown that they are unable to really reproduce the complexity of human behavior [29, 79]. From these perspective, many researchers have focused on collecting data and especially encounters between nodes from real life scenarios (conferences, student campuses, meetings, . . .). Relying on real traces, it is possible to identify salient properties of human mobility and perform more accurate simulations and more reliable evaluations.

2.3.2 Encounter Traces

In last years, many data collections about encounters in intermittently connected wireless networks have been led [3]. The corresponding traces are divided into two categories according to the encounters between mobile devices: the traces with derived encounters and those with direct encounters.

The derived encounters are deduced from data sets that describe the correspondences between mobile devices and access points in the network. In the literature, we find several traces of this type [5, 16, 52, 106, 39, 66]. The encounters between

access points and mobile devices are used to analyze mobility and evaluate communication protocols by assuming that two nodes are direct neighbors if they are connected to the same access point at the same time [29]. As done in [29, 103], we use parts of the traces [5] and [16] to derive encounters from the correspondences between access points and nodes connected to.

On the other hand, many initiatives [58, 95, 72, 29, 82] have targeted to collect direct encounters in various environments. The studies [58, 95, 29, 82] have asked people to carry iMotes (Intel Motes) [82] in limited areas (conferences, campuses, city locations, . . .) and the one proposed by Leguay et al. [72] in wide areas (city-scale measurements). In our work, we use also this kind of traces and we have been interested in a part of the trace of Infocom 2006 [95].

2.3.3 Human Mobility Properties

Using the encounter traces has enabled researchers to highlight interesting properties of the human properties. Indeed, the studies [16, 52, 106, 39, 77] have observed and analyzed the periodic aspect of human mobility. It has been shown that people move between well defined locations and that the time spent in different locations follows power-law [106]. These observation has been used to design new mobility models as the TVC model (Time-Variant Community mobility model) proposed in [57]. This mobility model is based on two human mobility properties: the skewed location visited preferences and the time-dependent mobility behavior. Many other works [58, 29, 63] have also emphasized the power-law property of the contact and inter-contact times between nodes. Also, Song et al. [101] have been interested in mobility patterns of cellular network users and have demonstrated that human mobility is highly predictable. They have found a 93% potential predictability in user mobility which is typically characterized by a lack of variability and which predictability is independent of the distance that can be covered by any user.

Recently, some studies have been interested in the impact of the social interactions on human mobility [39, 78, 105]. Eagle and Pentland [39] have outlined how a common context (work, hobby, . . .) in which two users evolve can be used to deduce their social relationship. Mtibaa et al. [78] have shown correlations between the contact graph and the social graph of mobile users participating in a conference. Thakur et al. [105] have highlighted that the characteristics of the traces obtained through the Time-Variant Community mobility model (TVC model) are not similar to those collected in real scenarios. The major cause behind this observation is that the TVC model reproduces the individual behavior (location preferences and time-dependent mobility behavior (circadian rhythm for example)) without correlating it

with other behaviors. In other words, using such a model generate users that have not preferred contacts (strong friends) and consequently no consistent social interactions. To emphasize these findings, Thakur et al. have proposed to compute a similarity metric. It measures the degree of similarity of the behaviors of two mobile nodes and the behavior of each node is expressed by an *association matrix*. The columns of the matrix represent the possible locations that a node can visit and the rows express time granularity (hours, days, weeks, etc.). The dominant behavioral patterns are tracked using the Singular Value Decomposition (SVD) [54]. For more details about the similarity metric computation, we refer the reader to [105].

From this perspective, the nodes in such networks are strongly interdependent and the interactions between them govern the structure of the network. This fact has motivated researchers to apply the Social Network Analysis (SNA) [111] to extract intrinsic properties of the network and to exploit them to design more efficient communication protocols.

2.4 Social Network Analysis and Link Prediction

2.4.1 Overview on the Social Network Analysis

Given that many of wireless opportunistic networks are human-centered and due to the lack of infrastructure and/or the opportunistic forwarding, the use of SNA has been favored. The SNA tends to analyze the social interactions in a network and to exploit the underlying findings in order to ensure a better packet delivery. Instead of network measurements tied to the network performance (latency, throughput, ...), the SNA can be viewed as another technique to grasp other network characteristics. As highlighted by Katsaros et al. [64], the SNA relies on the centrality metrics in order to identify which are the most important and influent nodes in the network. Moreover, it provides manifold community detection mechanisms. Such metrics and mechanisms contribute in the analysis of the social aspects in the network. In fact, they provide structural information of the network: existence and strength of communities, central nodes, network robustness to dynamicity, topology evolution as time goes on, ...

Several algorithm have been based on SNA to increase the reliability of mobile wireless networks. These algorithms closely rely on human mobility and then sensitive to human social interactions. Hui et al. [59] have proposed BubbleRap, a social based forwarding algorithm. It combines two major aspects of society. Indeed, Bubble takes into consideration the knowledge of community structure and the betweenness centrality characterization (fraction of shortest paths crossing a

node). Given that a node belongs at least to one community and has two ranking parameters (a global ranking, in the whole system, related to global centrality parameter and a local ranking, within the local community), the algorithm is as following: when a node has a message to deliver to another one, it bubbles it up the hierarchical ranking tree based on centrality (nodes called hubs) until it is received by a node belonging to the same community as the destination. After that, the local ranking (inside the community) is considered and the message is forwarded till it reaches the destination. In the same way, Daly and Haahr present SimBet Routing [35] based on small world dynamics [112]. This proposition addresses the problem of difficult communication in sparse mobile ad hoc networks. To overstep the cited difficulty, the authors lean on discovering a route that maximize the delivery rate and minimize the end-to-end delay. This challenge is raised even if the network is assimilated to a disconnected graph. To do this, they define a forwarding metric that relies on node's centrality (or ego-centrality) within the network and on node's social similarity (common neighbors) to the destination. Hossmann et al. [56] highlight the crucial impact of efficient mappings of mobility contacts on aggregated social graphs, used by DTN algorithms to determine forwarding decisions. They note that social DTN routing protocols BubbleRap [59] and SimBet [35], that we have detailed, contribute to achieve better performances compared to not explicitly social ones. They target to identify what are the conditions that improve the performances and why.

The SNA has been also exploited in link prediction in social networks and especially in data-mining, for co-authorship networks. We detail in the following subsection how the problem of link prediction has been stated and we detail which are the different categories of techniques used to perform link prediction.

2.4.2 Link Prediction in Social Networks

Social networks are dynamic as they can grow and change very rapidly as time goes on. These variations can be translated by the creation of new edges in the network which matches with the apparition of new interactions in the social relationships. Meanwhile, strong social ties are expected to "resist" to these changes and infer persistence and stability. Therefore, understanding the mechanisms that influence the evolution of the social networks is a fundamental question. From this perspective, in order to understand how social networks evolve, Liben-Nowell and Kleinberg [74] have focused on the link prediction problem. In fact, they have defined this basic computational problem: given a snapshot of a social network at time t , how can we accurately predict the edges that will be added in the network in the time interval

2.4. Social Network Analysis and Link Prediction

$t' - t$ where t' is a given future time?

Liben-Nowell and Kleinberg have advanced an interesting insight about the social networks. Indeed, they have stated that the link prediction problem is about identifying the extent of the social network using intrinsic properties of the network itself. Hence, they have considered features based on the common neighborhood and the link structure of the network and they have highlighted the importance of the link prediction in improving the applications in social networks. In the same way, Lü and Zhou [70] proposed a detailed presentation of the features that can be used to perform link prediction in social networks. They have classified them into three different categories:

- Local Similarity Indices which are built on the common direct neighbors that two nodes can have.
- Global Similarity Indices that relies on the paths that connect two nodes.
- Quasi-Local Similarity Indices which correspond to measures that take into consideration local paths (limited number of hops).

In our major contribution, we have been interested in some of these indices. We define them briefly:

Common Neighbors: for a node i , let $\Gamma(i)$ be the set of direct neighbors of i . When two nodes i and j have several common neighbors, it is expected that a link occurs between them. The number of common neighbors between the nodes i and j is defined as:

$$\text{Common_Neighbors}(i, j) = |\Gamma(i) \cap \Gamma(j)| \quad (2.1)$$

Salton Index [94]: it also called cosine similarity in the literature. Given that k_i is the degree of the node i , the Salton Index is expressed by:

$$\text{Salton_Index}(i, j) = \frac{|\Gamma(i) \cap \Gamma(j)|}{\sqrt{k_i \times k_j}} \quad (2.2)$$

Jaccard's Coefficient [61]: is defined as the ratio of the number of common neighbors of two nodes i and j and the cumulative number of all their neighbors. Is expressed as following:

$$\text{Jaccard's_Coefficient}(i, j) = \frac{|\Gamma(i) \cap \Gamma(j)|}{|\Gamma(i) \cup \Gamma(j)|} \quad (2.3)$$

Sørensen Index [104]: the use of this index is generally limited to ecological

Chapter 2. Human-Centered Mobile Wireless Networks: Human Mobility in the Service of Link Prediction

community data. It is defined as:

$$Sorensen_Index(i, j) = \frac{2|\Gamma(i) \cap \Gamma(j)|}{k_i + k_j} \quad (2.4)$$

Hub Promoted Index (HPI) [92]: this measure is used to quantify the *topological overlap of pairs of substrates in metabolic networks* [92]. It expressed as follows:

$$HPI(i, j) = \frac{|\Gamma(i) \cap \Gamma(j)|}{\min(k_i, k_j)} \quad (2.5)$$

Hub Depressed Index (HDI): this index, which is analogous to the previous measure, was advanced by Lü and Zhou [70]. It is defined by:

$$HDI(i, j) = \frac{|\Gamma(i) \cap \Gamma(j)|}{\max(k_i, k_j)} \quad (2.6)$$

Adamic-Adar Index [10]: Adamic and Adar have defined their similarity index by giving weights to all common neighbors of a pair of nodes (i, j) . The lower the degree of the common neighbor is, the higher its weight gets. The Adamic-Adar Index is given by:

$$Adamic - Adar_Index(i, j) = \sum_{z \in \Gamma(i) \cap \Gamma(j)} \frac{1}{\log k_z} \quad (2.7)$$

Resource Allocation Index [126]: it is exploited in the resource allocation in dynamic networks [83]. It is similar to the Adamic-Adar Index and defined by:

$$Resource_Allocation_Index(i, j) = \sum_{z \in \Gamma(i) \cap \Gamma(j)} \frac{1}{k_z} \quad (2.8)$$

Rooted PageRank Index [102]: this index is considered as an application of the PageRank algorithm [23]. The score matrix is given by:

$$PageRank_Score_Matrix = (1 - \beta)(I - \beta T)^{-1} \quad (2.9)$$

where β is the probability that a random walk moves from a node to one of its random neighbors and $T = DA$ with A is the adjacency matrix of the network and $D_{ii} = 1/\sum A_{ij}, D_{ij} = 0$ when $i \neq j$.

Katz Measure [65]: it is defined as the weighted sum of the number of paths that

2.4. Social Network Analysis and Link Prediction

connect two nodes i and j . It is expressed by:

$$Katz_Measure(i, j) = \sum_{l=1}^{\infty} \beta^l \cdot |paths_{ij}^{<l>}| \quad (2.10)$$

where $paths_{ij}^{<l>}$ denotes the set of all paths of length l that connect i to j and β represents the damping factor (the value of β is between 0 and 1 and then the longer a path is, the lower its weight gets).

Many studies such as [74, 8, 15, 38, 98] have demonstrated that using intrinsic properties of the network leads to predict the links in a social network. Liben-Nowell and Kleinberg [74] have demonstrated that measuring network proximity is able to give insights on link prediction in co-authorship networks [4, 6]. Acar et al. [8] and Dunlavy et al. [38] have explored different matrix- and tensor-based techniques that track the evolution of the network during a number of periods and predict the social network topology in the following period(s). These techniques and especially the tensor-based one have been coupled to the Katz measure as this measure quantify the network proximity of a pair of nodes based on the paths that connects them. Social network services and websites have interested researches about link prediction in social networks. Backstrom and Leskovic [15] have developed an algorithm based on Supervised Random Walks to grasp the structure of the social network and to assign weights to the edges. Applied notably on Facebook, the proposed algorithm predicts for each user a list of potential "friends". Shin et al. [98] have proposed a general framework for multi-scale link prediction (MSLP) which handles large-scale social networks as LiveJournal, Flickr and MySpace. This algorithm constructs low-rank approximations of a social network for different scales. At each scale, the approximated network proximity of each pair of users is determined and Shin et al. have opted for the Katz measure. Afterwards, the MSLP algorithm combines the approximative network proximity scores for all scales and make the prediction based on combined scores.

It is important to mention that most of the cited works have highlighted that the link prediction performance of the Katz measure outperforms the ones of the other indexes. We have used this metric in our work and we will give more details about it in the following chapter.

Concerning link prediction in mobile wireless networks, we have remarked the work of Wang et al. [110]. Through their study on records of 6 million mobile phone users, they have aimed to highlight the relationship between the individual mobility patterns and the social network formed by the users. They have proposed to measure the behavior similarity of two users through two categories of metrics:

Chapter 2. Human-Centered Mobile Wireless Networks: Human Mobility in the Service of Link Prediction

the network proximity metrics (Common Neighbors, Adamic-Adar, Jaccard's Coefficient, Katz Measure) and the mobility homophily metrics such as the Spatial Cosine Similarity and the Co-Location Rate [110]. On top of finding that the similarity of movements of two users is highly correlated to their proximity in the network, Wang et al. have shown that all these measures have an interesting predictive power and that combining them through a supervised learning improves significantly the link prediction performance.

We have been interested in the two mobility homophily metrics cited above to compare our work to other link prediction techniques in chapter 4. Wang et al. defined the most likely location of a mobile phone user i as:

$$ML(i) = \arg \max_{l \in Loc} PV(i, l) \quad (2.11)$$

where Loc designates the set of all locations which are cell phone towers. and

$$PV(i, l) = \sum_{a=1}^{n(i)} \delta(l, L_a(i)) / n(i) \quad (2.12)$$

represents the probability that user i is in the location l . The function δ has two inputs x and y . If $x = y$ then $\delta(x, y) = 1$ and 0 otherwise. The parameter a represents the step number and $L_a(i)$ returns the location visited by the user i at the step a .

Using this probability, Wang et al. defined many homophily metrics and we have retained the two following metrics:

Spatial Cosine Similarity: it is used to capture the cosine similarity of the trajectories of two users i and j . If these two users share similar visited locations, the cosine similarity tends to be high and a link is expected to occur as these two users seem to be close. The spatial cosine similarity is expressed as follows:

$$SCos(i, j) = \sum_{l \in Loc} \frac{PV(i, l) \times PV(j, l)}{\|PV(i, l)\| \times \|PV(j, l)\|} \quad (2.13)$$

Co-Location Rate: this metric computes the probability that two users i and j appear in the same location l during approximatively the same time:

$$CoL(i, j) = \frac{\sum_{a=1}^{n(i)} \sum_{b=1}^{n(j)} \Theta(\Delta T - |T_a(i) - T_b(j)|) \delta(L_a(i), L_b(j))}{\sum_{a=1}^{n(i)} \sum_{b=1}^{n(j)} \Theta(\Delta T - |T_a(i) - T_b(j)|)} \quad (2.14)$$

where $\Theta(x)$ is the Heaviside step function and ΔT is the maximum delay between of two nodes i and j visiting timestamps in the same location l which enables us

to say that these two nodes can be considered as co-located in l . The parameters a and b correspond to the step numbers that characterize the nodes i and j . The function $\delta(L_a(i), L_b(j))$ returns 1 if the node i in the step a of its move is in the same location as node j during its moving step b .

2.5 Conclusion

In this chapter we have shown how interactions between the entities in a social network can be important to perform link prediction. In this way, we want to exploit such interactions in human-centered mobile wireless networks as MANETs and DTNs to perform efficient link prediction. In the following chapter, we present a detailed description about our tensor-based link prediction framework which is inspired from data-mining and adapted to the context of human-centered mobile wireless networks.

Chapter 3

Tensor-Based Link Prediction Framework for Mobile Wireless Networks

Contents

3.1	Introduction	33
3.2	Related Work	35
3.3	Description of the Tensor-Based Link Prediction Method	37
3.3.1	Notation	38
3.3.2	Matrix of Scores Computation	38
3.3.3	Matrix of Scores Interpretation	41
3.4	Performance Evaluation and Simulation Results	41
3.4.1	Simulation Traces	42
3.4.2	Performance Analysis	43
3.4.3	Discussion	55
3.5	Conclusion	57

3.1 Introduction

In recent years extensive research has addressed challenges and problems raised in mobile, sparse and intermittently connected networks (i.e. DTN). In this case, forwarding packets greatly depends on the occurrence of contacts. Since the existence of links is crucial to deliver data from a source to a destination, the contacts and

their properties emerge as a key issue in designing efficient communication protocols [56]. Obviously, the occurrence of links is determined by the behavior of the nodes in the network [29]. It has been widely shown in [57, 105] that human mobility is directed by social intentions and reflects spatio-temporal regularity. A node can follow other nodes to a specific location (spatial level) and may bring out a behavior which may be regulated by a schedule (temporal level). The social intentions that govern the behavior of mobile users have also been observed through statistical analyses in [29, 63] by showing that the distribution of inter-contact times follow a truncated power law.

With the intention of improving the performance of intermittently connected wireless network protocols, it is paramount to track and understand the behavior of the nodes. We aim to propose an approach that analyzes the network statistics, quantifies the social relationship between each pair of nodes and exploits this measure as a score which indicates if a link would occur in the immediate future. We strongly believe that the social ties between nodes govern the status of a link and establish an indication for the link prediction: it would not occur if two nodes have no common interactions or willingness and would be effective and persistent with more correlated moving patterns.

In this chapter, we adapt a tensor-based link prediction algorithm successfully designed for data-mining [8, 38]. Our proposal records the network structure for T time periods and predicts all link occurrences for the $(T + 1)^{th}$ period and estimates the birth of new links (that have not occurred during the preceding T recording periods). This link prediction technique is designed through two steps. First, tracking time-dependent network snapshots in adjacency matrices which form a third-order tensor (two dimensions for nodes as adjacency matrices and one dimension for time to obtain the evolution of the network structure). Second, applying of the Katz measure [65] inspired by sociometry. The link prediction technique computes the degree of behavioral similarity of each pair of nodes relying on the tensor obtained in the first step. A high degree of behavior similarity means that the two nodes have the same "social" intentions. These common intentions are expressed by the willingness to meet each other and/or by similar moving patterns to visit the same location. They also promote the link occurrence between two socially similar nodes in the immediate future (prediction for the period $T+1$ after tracking the behavior of nodes during T time periods).

We discuss how we have designed the tensor-based prediction method and detail the two main steps in order to achieve link prediction. On the one hand, we describe how to track the network topology over time with a tensor. On the other hand, we explain how to compute and interpret the Katz measure. We then evaluate the

effectiveness of predictability through several simulation scenarios depending on the trace, the number of recording periods and the similarity metric computation which can be used in a centralized or a distributed way. To the best of our knowledge, this work is the first to perform the prediction technique in a distributed way. By "distributed way", we mean that each node in the network uses local information (the contacts with neighbors) to predict future links. The assessment of its efficiency can be beneficial for the improvement of the design of communication protocols in mobile, sparse and intermittently connected networks.

This chapter is organized as follows: Section 3.2 presents the related work that highlights the growing interest in social analysis and justifies the recourse to the tensors and to the Katz measure to perform predictions. In Section 3.3, we detail the two main steps that characterize our proposal. Section 3.4 details simulation scenarios used to evaluate the tensor-based prediction approach, analyzes the results obtained, assesses its efficiency and proposes a discussion about the described link prediction technique. Finally, we conclude the chapter in Section 3.5.

3.2 Related Work

Social Network Analysis (SNA) [111, 64] and ad-hoc networking have provided new perspectives for the design of network protocols [59, 35, 55]. These protocols aim to exploit the social aspects and relationship features between the nodes. Studies conducted in the field of SNA have mainly focused on two kinds of concepts: the most well-known centrality metrics suggested in [111, 84, 60, 31] and the community detection mechanisms proposed in [20, 81, 85, 111]. From this perspective, several works have tried to develop synthetic models that aim to reproduce realistic moving patterns [57, 71]. Nonetheless, the study done in [56] has underlined the fact that synthetic models cannot faithfully reproduce human behavior because these synthetic models are only location-driven and they do not track social intentions explicitly.

In their survey, Katsaros et al. [64] have underlined the limits of these protocols when the network topology is time-varying. The main drawback comes down to their inability to model topology changes as they are based on graph theory tools. To overcome this limit, tensor-based approaches have been used in some works to build statistics on the behavior of nodes in wireless networks over time as in [9]. Thakur et al. [105] have also developed a model using a collapsed tensor that tracks user's location preferences (characterized by probabilities) with a considered time granularity (week days for example) in order to follow the emergence of "behavior-aware" delay tolerant networks closely.

As previously mentioned, tracking the social ties between network entities enables us to understand how the network is structured. Such tracking has led to the design of techniques for link prediction. Link prediction in social networks has been addressed in data-mining applications as in [8, 38]. In the same context, Wang et al. [109] have advanced a method based on a local probabilistic graphical model that can be used for large graphs. This method is led by a supervised learning to predict if a link will occur between two nodes through the computation of the co-occurrence probability. Also, Backstrom and Leskovic [15] have developed an algorithm which uses Supervised Random Walks to learn the structure of the social network and to set the edge weights. Applied notably on Facebook, such an algorithm estimates for each user a list of people with whom ties are likely to be created. Concerning link prediction in community-based communication networks, [110] has highlighted salient measures that allow link occurrence between network users to be predicted. These metrics determine if a link occurrence is likely by quantifying the degree of proximity of two nodes (Katz measure [65], the number of common neighbors, Adamic-Adar measure [10], Jaccard's coefficient [61, 94], ...) or by computing the similarity of their mobility patterns (spatial cosine similarity, co-location rate, ...).

Through their exhaustive study, Liben-Nowell and Kleinberg [74] have presented several similarity measures based on node neighborhoods, the ensemble of all paths and higher-level approaches. They have compared the prediction performance of these measures applied as predictors on five co-authorship networks. In the same way, Lü and Zhou [70] have described in detail well-known metrics used in link prediction and they have classified them into three categories: the local metrics (computed from common neighbors and/or node degree), the global metrics (determined from the set of paths that connect two nodes) and the quasi-local metrics (a mix of the two previous categories). The efficiency of the Katz measure has been emphasized and compared to other prediction metrics in [74, 110].

In this chapter, we propose a link prediction technique that tracks the temporal network topology evolution in a tensor and computes a metric in order to characterize the social-based behavior similarity of each pair of nodes. Some approaches have addressed the same problem in data-mining in order to perform link prediction. Acar et al. [8] and Dunlavy et al. [38] have provided detailed methods based on matrix and tensor factorizations for link prediction in social networks such as the DBLP data set [6]. These methods have been successfully applied to predict a collaboration between two authors relying on the data set of the structure of relationships over time. Moreover, they have highlighted the use of the Katz measure [65], which can be seen as a behavior similarity metric, by assigning a link prediction score for each pair of nodes. The efficiency of the Katz measure in link prediction

has been also demonstrated in [109, 74].

3.3 Description of the Tensor-Based Link Prediction Method

It has been highlighted that a human mobility pattern shows a high degree of temporal and spatial regularity, and each individual is characterized by a time-dependent mobility pattern and a trend to return to preferred locations [29, 57, 105]. In order to improve the design of wireless network protocols, and especially the intermittently connected networks, it is important to exploit this knowledge since these interactions usually have an impact on the network structure and consequently on the network performance. Thus, we propose an approach that aims to exploit the similar behavior of nodes in order to predict link occurrence referring to the social closeness.

Predicting future links based on their social closeness is a challenge that is worth an investigation. Indeed, a good link prediction technique contributes to improving the opportunistic forwarding of packets and also enhances the delivery rate and/or decreases latency. Moreover, it helps avoid situations where packets encumber the queue of the nodes that are not able to forward these packets towards their final destinations.

To quantify the social closeness between each pair of nodes in the network, we use the Katz measure [65] inspired by sociometry. This measure aims at quantifying the social distance between people inside a social network. We also need to use a structure that records link occurrence between each pair of nodes over a certain period of time in order to perform the similarity measure computation. The records represent the network behavior statistics in time and space. To this end, a third-order tensor is considered. A tensor \mathbf{Z} consists of a set of slices and each slice corresponds to an adjacency matrix of the network tracked over a given period of time p . After the tracking phase, we reduce the tensor into a matrix (or collapsed tensor) which expresses the weight of each link according to its lifetime and its recentness. A high weight value in this matrix denotes a link whose corresponding nodes share a high degree of closeness. We apply the Katz measure to the collapsed tensor to compute a matrix of scores \mathbf{S} that not only considers direct links but also indirect links (multi-hop connections). The matrix of scores expresses the degree of similarity of each pair of nodes according to the spatial and the temporal levels. The higher the score is, the better the similarity pattern gets. Therefore, two nodes that have a high similarity score are more likely to have a common link in the future.

3.3. Description of the Tensor-Based Link Prediction Method

3.3.1 Notation

Scalars are denoted by lowercase letters, e.g., a . Vectors are denoted by boldface lowercase letters, e.g., \mathbf{a} . Matrices are denoted by boldface capital letters, e.g., \mathbf{A} . The r^{th} column of a matrix \mathbf{A} is denoted by \mathbf{a}_r . Higher-order tensors are denoted by bold Euler script letters, e.g., \mathcal{T} . The n^{th} frontal slice of a tensor \mathcal{T} is denoted by \mathbf{T}_n . The i^{th} entry of a vector \mathbf{a} is denoted by $\mathbf{a}(\mathbf{i})$, element (i, j) of a matrix \mathbf{A} is denoted by $\mathbf{A}(i, j)$, and element (i, j, k) of a third-order tensor \mathcal{T} is denoted by $\mathbf{T}_i(j, k)$.

3.3.2 Matrix of Scores Computation

The computation of the similarity scores is modeled in two distinct steps. First, we store the inter-contact between nodes in a tensor \mathcal{Z} and reduce it to a matrix \mathbf{X} called the collapsed tensor. In the second step, we compute the matrix of similarity scores \mathbf{S} relying on the matrix \mathbf{X} (cf. Fig. 4.1).

Collapsing the data from the tensor

We consider that the data is collected into the tensor \mathcal{Z} . The slice $\mathbf{Z}_p(i, j)$ describes the status of a link between a node i and a node j during a time period between $[(p-1) \cdot t, p \cdot t[$ ($p > 0$) where $\mathbf{Z}_p(i, j)$ is 1 if the link exists, even for few instants, during the time period p and 0 otherwise. The tensor is formed by a succession of adjacency matrices \mathbf{Z}_1 to \mathbf{Z}_T where the subscript letters designate the observed period. The simplest way to collapse these records into a single $N \times N$ matrix \mathbf{X} is to sum all the entries over the tracked time as in the following:

$$\mathbf{X}(i, j) = \sum_{p=1}^T \mathbf{Z}_p(i, j) \quad (3.1)$$

We follow another alternative, motivated by [8, 38], to collapse the data into one matrix. This alternative consists in damping backward the structure of the network. The links are considered over time and the more recent the adjacency matrix, the more weighted the structure. Collapsing the tensor records in this way is applied using the following expression:

$$\mathbf{X}(i, j) = \sum_{p=1}^T (1 - \theta)^{T-p} \mathbf{Z}_p(i, j) \quad (3.2)$$

where the matrix \mathbf{X} is called collapsed weighted tensor of \mathcal{Z} , and θ is a parameter used to adjust the weight of recentness and is between 0 and 1.

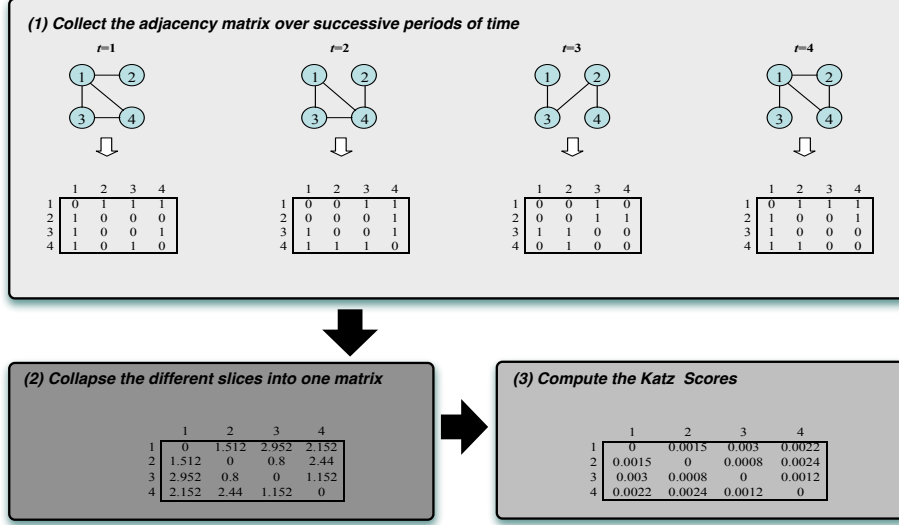


Figure 3.1: Example of the matrix \mathbf{S} computation

According to the results for the evaluation of the Katz measure prediction performance in [8, 38], collapsing the data relying on Eq. (4.1) gives better prediction results (compared to collapsing the records by using Eq. (3.1)). As we aim to perform a prediction using the Katz measure, we opt to collapse the tensor using Eq. (4.1). Below, we define the Katz measure and we explain how to compute it after determining the collapsed weighted tensor.

Katz Measure

The Katz measure, which is affiliated to sociometry, was first proposed by Leo Katz in [65]. He considered a social network as a directed graph $G = (V, E)$ where each vertex $v_i \in V$ represents a person i and each edge $(v_i, v_j) \in E$ connects the vertex i to the vertex j when a person i votes for (or endorses) a person j . We denote the subset $P_\ell(v_i, v_j)$ as the set of paths of length ℓ that connect vertex v_i to vertex v_j . The status that characterizes the couple (v_i, v_j) is defined by the weight and the number of paths of length ℓ , $|P_\ell(v_i, v_j)|$, connecting vertex v_i to vertex v_j .

Katz defined the score $\mathbf{S}(i, j)$ of a pair of nodes (i, j) as depicted in Eq.(3.3):

$$\mathbf{S}(i, j) = \sum_{\ell=1}^{+\infty} \beta^\ell |P_\ell(v_i, v_j)| \quad (3.3)$$

Where β is a user defined parameter, also called the damping factor, strictly superior to 0 and strictly inferior to 1. The expression β^ℓ denotes the weight of a path of

3.3. Description of the Tensor-Based Link Prediction Method

length ℓ .

This score is dependent on a function that proportionally decreases with the path length ℓ . The weight of such paths is β^ℓ and it is clear that the longer the path is, the lower the weight gets.

The Katz centrality of a person i , which is the sum of Katz measures involving the person i , can be seen as a generalization of the in-degree measure (like followers in Twitter). The status of the node according to the in-degree centrality measures the number of direct neighbors. The Katz centrality goes beyond that and measures the number of all nodes that are connected through one or multiple paths (these paths can be a direct link and/or multi-hop paths). Nevertheless, the contribution provided by a distant node is penalized by the damping factor β . Hence, the Katz centrality can be viewed as an extension of the in-degree measure to long paths. By reconsidering the social network defined by Katz, the Katz centrality takes into account the number of votes that a person can receive through direct or indirect endorsements (respectively paths of length 1 and those of length ℓ with $\ell > 1$). The damping factor, as mentioned, is used in order to emphasize that the strength of the endorsements fades over a chain of recommendations.

This metric is widely used in studies whose aim is to predict the occurrence of links [8, 38, 74], especially in social networks like co-authorship communities such as the DBLP [6] and arXiv [4] databases. Given that there are "social relationships" between nodes in networks with intermittent connections, it is challenging to exploit this measure and to apply it to collected data.

There is also another formulation to compute Katz scores by means of collapsed weighted tensor as detailed previously. Then, the score matrix \mathbf{S} can be rewritten as:

$$\mathbf{S} = \sum_{\ell=1}^{+\infty} \beta^\ell \cdot \mathbf{X}^\ell = (\mathbf{I} - \beta \cdot \mathbf{X})^{-1} - \mathbf{I} \quad (3.4)$$

where \mathbf{I} is the identity matrix and \mathbf{X} is the obtained collapsed weighted tensor.

The computation of matrix \mathbf{S} , as described before, is done in a **centralized way**. It means that the matrix \mathbf{S} is computed based on the full knowledge of the network topology over time. This may not be suitable with ad-hoc wireless networks where no central entity is considered and could in addition be very costly. A **distributed mechanism** should then be examined. In a distributed mechanism, each node would apply the prediction method relying only on information related to its nearest neighbors. It is paramount to remember that a Katz formulation gives more weight to short paths and assigns low scores to long paths. Therefore, the scores with neighbors located a few hops away should be sufficient and strong

enough compared to scores with further ones.

In the next section, in order to evaluate the efficiency of our proposal to perform prediction relying only on local information, we consider two ways to compute Katz measures:

- **The Centralized Computation:** The centralized way assumes that there is a central entity which has full knowledge of the network structure at each period and applies the Katz measure to the global adjacency matrices.
- **The Distributed Computation:** Each node has a limited knowledge of the network structure. We assume that a node is aware of its two-hop neighborhood. Hence, computation of Katz measures is performed on a local-information-basis.

The validity of the distributed mechanism will be discussed in Section 3.4.

3.3.3 Matrix of Scores Interpretation

The relationship between each pair of nodes is expressed by a score $\mathbf{S}(i, j)$, which reflects the degree of similarity between node i and node j . As mentioned in the Katz measure analysis, shorter paths lead to higher scores. Thus, two nodes that share a high score are nodes that are connected through short paths during some period of time and therefore have similar behavior (similar social intentions). The similarity here is related to common preferences in spatial and temporal space. Two nodes maintain their connectivity when they move in the same direction and at the same time. Therefore, these scores can be considered as indicators to a possible link existence in the future. Thus, the link prediction is done through measuring behavior similarity for each pair of nodes in the matrix \mathbf{S} .

3.4 Performance Evaluation and Simulation Results

To evaluate the efficiency of the tensor-based link prediction in intermittently connected wireless networks, we consider three different real traces. For each trace, we compute the corresponding scores matrix \mathbf{S} as described earlier and assess the performance of the link prediction method through evaluation techniques. In the following, we firstly present the traces used for the link prediction evaluation. Then, we give the corresponding results and analyze the effectiveness of the prediction method.

3.4.1 Simulation Traces

We consider three real traces to evaluate the link prediction approach. We exploit them to construct the tensor by generating adjacency matrices (with different tracking periods t : 5, 10 and 30 minutes). For each case, we track the required statistics about the nodes' behavior within T periods. We consider the adjacency matrix corresponding to the period $T+1$ as a benchmark to evaluate the Katz scores matrix in predicting the occurrence of links in the immediate future. We also take into account some periods beyond the period T to check the ability to predict links that have not occurred during the tracking phase and that have been tracked starting from the period $T+1$. We detail, in the following, the traces used.

- **First Trace: Dartmouth Campus trace:** We choose the trace of 01/05/2006 [5] from 8 a. m. to 6 p. m. We construct the tensor slices relying on SYSLOG traces between 8 a.m. and 3 p.m. (7 hours). The encounters that took place in the three remaining hours are considered as future events. The number of nodes is 1018.
- **Second Trace: MIT Campus trace:** We focus on the trace of 07/23/2002 [16] and also consider the same time slots as for the previous trace. The number of nodes is 646.
- **Third Trace: Infocom 2006 trace:** We consider the encounters which happened on 04/25/2006 [95] between 8 a.m. and 6 p.m among the 98 nodes involved in the trace (70 of them were carried by humans). We construct the tensor relying on the events that have happened from 8 a.m. to 2 p.m. The remaining time is used to identify new links that were not pointed out during the tracking phase and that occurred between 2 p.m. and 6 p.m.

For each scenario, we generate adjacency matrices corresponding to a different period t : 5, 10 and 30 minutes. Then, to record the network statistics over each trace historical, we use the tensor to track the encounters. The number of periods T depends on the trace (the retained historical) and on the length of the tracking period t . For example, for the Infocom 2006 trace, the tensor has respectively a number of slices T equal to 72, 36 and 12 periods (for the case where $t=5$ minutes, it is necessary to have 72 periods to cover 6 hours). Our aim behind considering this set of traces is to evaluate the efficiency of our proposal (predicting using local information) in predicting the links that would occur in the immediate future (at the period $T+1$) and new links that would occur for the first time after the period T .

Concerning the two first traces, we have to underline that the data represent records that describe the association or the disassociation of a device (identified with its MAC address) with an access point at a corresponding timestamp. In such a case, the tracked contacts are not involved in a mobile ad-hoc network. Nevertheless, we assumed that two users can communicate with each other when they are affiliated to the same access point. Chaintreau et al. [29] made the same assumption with the Dartmouth Campus trace in order to analyze, with a theoretical model, the impact of human mobility on the opportunistic forwarding algorithms. Song and Kotz [103] stated that this assumption may not be accurate but remains a good first approximation. The last trace (Infocom 2006 trace) was proposed to focus on the user mobility characterization and the content distribution evaluation. In particular, this trace was registered during an event in which encounters between people are frequent and repetitive. Hence, they can be viewed as a more faithful benchmark to evaluate our proposal.

As mentioned earlier, we take into account both centralized and distributed cases for the computation of scores. In both cases, we fix θ and β to 0.2 et 0.001 respectively. Later, in the Section 3.4.2, we explain the impact of these parameters and why we chose these values.

3.4.2 Performance Analysis

As described in the previous section, we apply the link prediction method to the three traces while considering the different tensor periods both for centralized and distributed cases. In order to assess the efficiency of the proposed method, we consider several link prediction scenarios (according to the trace, the tensor period and the way to compute scores) and we use different evaluation techniques (ROC curves, AUC metric, top scores ratio at $T+1$, accuracy and F-measure). Below we detail in the following the results obtained with each evaluation technique and analyze the link prediction efficiency.

Analysis of the ROC Curves

First of all, we aim to evaluate the performance of the prediction in a distributed way and compare this performance to that of the centralized approach. As a first evaluation step, we use the ROC curves (Receiver Operating Characteristic curves) [44].

A ROC curve is a graphical plot highlights the performance of a binary classifier (which is the prediction technique in this context) with varying the discrimination threshold. For us, the threshold takes values from the minimum to the maximum

3.4. Performance Evaluation and Simulation Results

Table 3.1: Table of confusion of a binary prediction technique

Prediction outcome \ Actual value	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

scores provided by the matrix of scores \mathbf{S} . To determine the points that form the ROC curve, we compare all the obtained scores with the current value of the threshold. If a score is higher to the threshold, it means that the prediction is positive. Otherwise, the prediction is negative. Hence, the ROC curve represents the evolution of the true positive rate (given by the ratio of true positive predictions among all positive predictions) expressed by $\frac{TP}{TP+FN}$ according to the false positive rate (given by the ratio of false positive predictions among all negative predictions) expressed by $\frac{FP}{FP+TN}$. The meaning of the variables TP , FP , TN and FN are detailed in the table of confusion given by Tab. 3.1. The ROC curve indicates a good prediction method if the bend of the curve is tending to be close to the upper left corner of the plot. The analysis of the obtained results will detail this appreciation.

The Fig. 3.2, 3.3 and 3.4 depict the ROC curves for both distributed and centralized computing approaches respectively obtained from the Dartmouth Campus trace, the MIT Campus trace and the Infocom 2006 trace. For each trace figure, (a), (b) and (c) the curves correspond to a tensor tracking period of 5, 10 and 30 minutes respectively.

We first notice that, for all scenarios, the prediction of all links is quite efficient, compared to the random guess (the curve's bends are in the upper left corner). Moreover, two other observations have to be mentioned. First, it is highlighted that the smaller the tensor tracking period, the more reliable the prediction. This observation is obvious for two reasons. On the one hand, with a short tensor period, tracking a short and occasional contact between two nodes is less likely compared to the case where the tensor periods are longer. On the other hand, if we take the example of the Infocom 2006 trace, recording 6 hours of encounters requires 72 adjacency matrices of 5-minute periods, instead of 12 matrices when the tracking is considered over periods of 30 minutes. Thus, tracking a short contact between two nodes has less influence when the tensor slices are more numerous. Moreover, short periods enable us to appreciate better short contacts as well as long ones, whereas predicting using long periods is not able to clearly estimate the length of a contact. As an example, in the case where the tensor slice time is 5 minutes, a fleeting contact can be tracked by one adjacency matrix among 72. However, for the case where the

slice time is 30 minutes, the fleeting contact is tracked by one tensor slice among 12, which gives it significantly more weight compared to the former case. Hence, short tensor slice periods enable us to minimize the probability of tracking a short contact and to restrict its impact. On the other hand, they allow the short and long encounters to be discerned more accurately.

Short tensor slice periods also allow us to better track the social interactions (meetings in a cafeteria, courses in an lecture theater, etc) between nodes which determine the occurrence of links. Successive adjacency matrices of 5 minutes give a more accurate description of the network structure over time as both analyzing and identifying these social events are easier through shorter periods.

The second observation concerns the similar results obtained at the centralized and distributed matrix of scores computation. In fact, the similarity is higher when the paths considered between a pair of nodes are short. Therefore, paths that have more than two hops have weaker scores and so are less weighted compared to shorter ones. The distributed case assumes that each node knows its neighbors at most at two hops. That is why distributed scores computation presents performances which are so similar to the centralized ones.

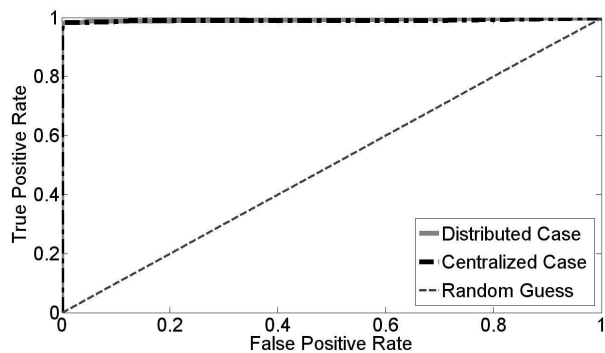
On top of that, we can highlight that the link prediction performance is better when the tensor-based technique is applied on the Dartmouth Campus and the MIT Campus traces than than the one obtained with the Infocom 2006 trace. This observation can be explained by the different natures of the traces. Indeed, the Dartmouth Campus and the MIT Campus traces are collected from access points and we adopted the same assumption as Chaintreau et al. [29] (detailed in the Section 3.4.1). Meanwhile, the Infocom 2006 trace is obtained after recording all the encounters between several mobile nodes which is impacted by lower range and density of nodes.

Evaluation of the Link Prediction Technique through Performance Metrics

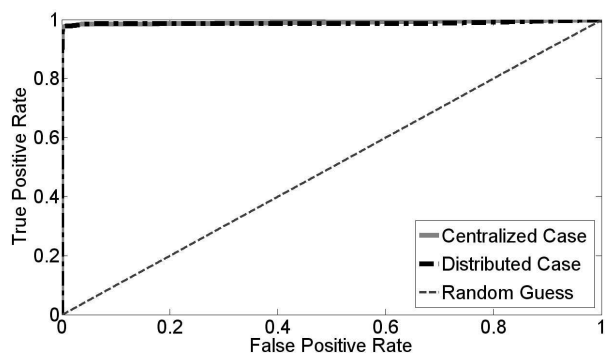
In this subsection, we consider another evaluation step. We use adapted evaluation metrics in order to further weigh the performance of the proposed link prediction technique. At this step, as well as evaluating the prediction of all links in the period $T+1$, we also try to assess the efficiency of our technique in predicting new links that occurred for the first time after the period T . For both link prediction cases, we compute a set of metrics:

- **Area Under the ROC Curve metric (AUC) [44]:** which is considered as a good performance indicator in our case. It consists in computing the area

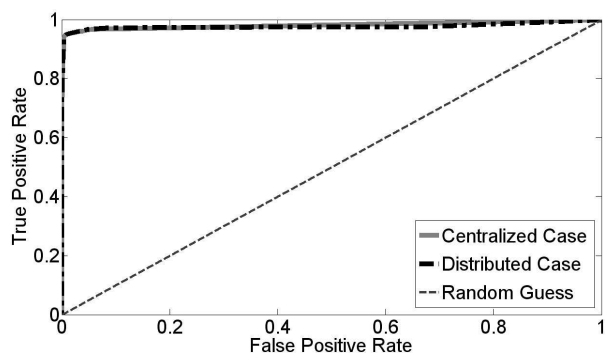
3.4. Performance Evaluation and Simulation Results



(a) 5 minutes tensor slice period



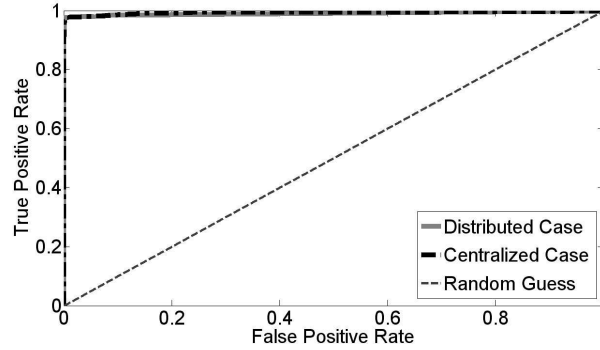
(b) 10 minutes tensor slice period



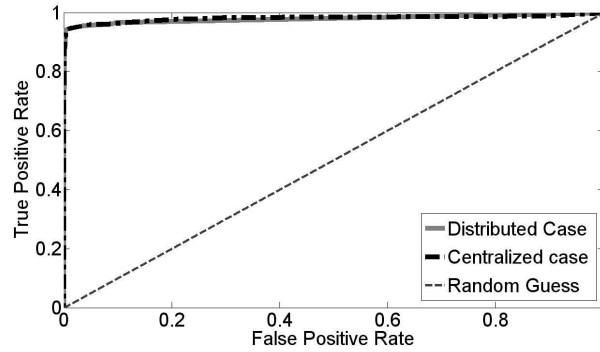
(c) 30 minutes tensor slice period

Figure 3.2: ROC Curves for different prediction cases applied on Dartmouth Campus trace

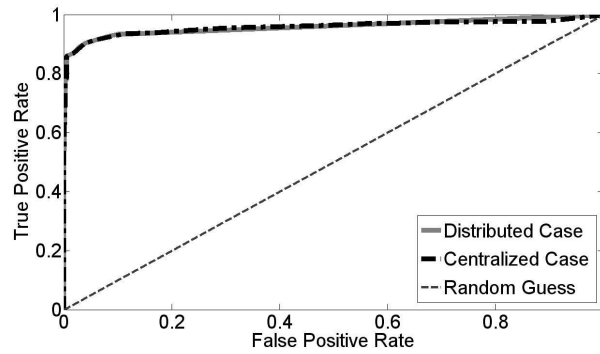
under the ROC curve. The closer the value of the area is to 1, the better the prediction gets.



(a) 5 minutes tensor slice period



(b) 10 minutes tensor slice period

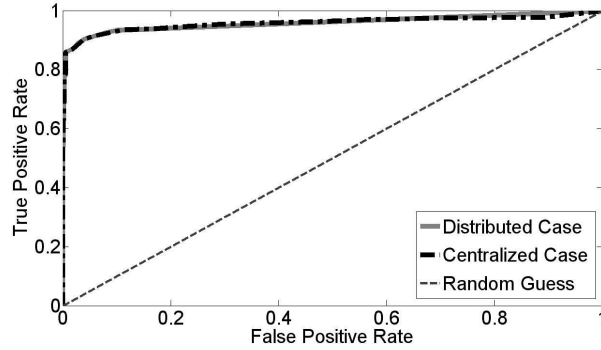


(c) 30 minutes tensor slice period

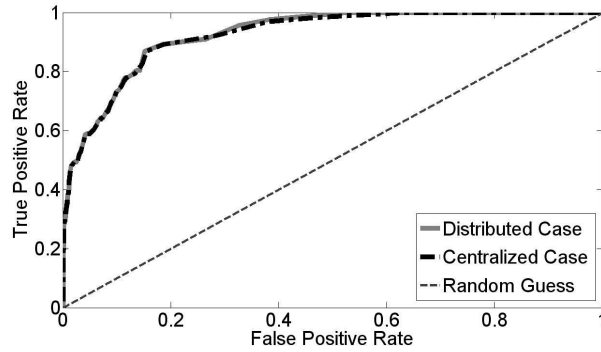
Figure 3.3: ROC Curves for different prediction cases applied on MIT Campus trace

- **Top Scores Ratio at the period $T+1$ (TSR):** we compute the number of occurring links in the period $T+1$ (the first period coming after the record of the network statistics). We call the number of existing links L . Then,

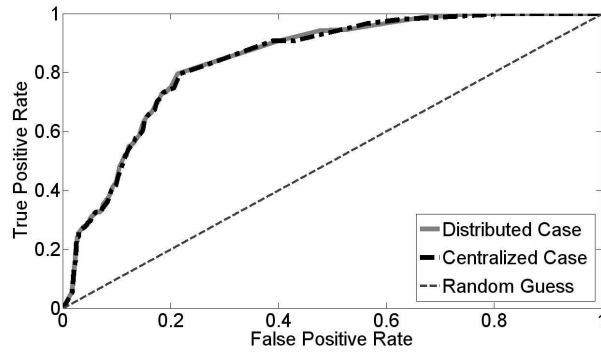
3.4. Performance Evaluation and Simulation Results



(a) 5 minutes tensor slice period



(b) 10 minutes tensor slice period



(c) 30 minutes tensor slice period

Figure 3.4: ROC Curves for different prediction cases applied on Infocom 2006 trace

we extract the links having the L highest scores found after applying the prediction technique and determine the percentage of existing links in both sets.

- **Accuracy (ACC):** this measure is defined in [44] as the ratio of correct prediction (true positive and true negative predictions) over all predictions (true positive, true negative, false positive and false negative predictions). In other words, it is computed by the ratio $\frac{TP+TN}{TP+FP+TN+FN}$ (see Table 3.1). We identify for each scenario the maximum value of the accuracy which indicates the degree of precision that can reach each prediction metric.
- **F-Measure or balanced F1 score:** the F-Measure [108] is the harmonic mean of precision¹ and recall². The F-Measure is expressed by $2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$. The higher the F-Measure is, the better the tradeoff of precision and recall gets and the more efficient the prediction metric is.

On top of doing the evaluation for the centralized and distributed ways to perform prediction, we compare their performances with those of some well-known prediction metrics: Common Neighbors, Salton Index [94], Jaccard Index [61], Sørensen Index [104], Hub Promoted Index (HPI) [92], Hub Depressed Index (HDI), Adamic-Adar Index [10], Resource Allocation Index [126] and Rooted PageRank Index [102]. We also include in the comparison the weights obtained through the computation of the Collapsed Weighted Tensor (CWT). The results corresponding to all links prediction at the period $T+1$ and those corresponding to new links prediction are listed in Table 6.2 (Dartmouth Campus trace), Table 6.3 (MIT Campus trace) and Table 6.4 (Infocom 2006 trace).

Also, we propose to use all the evaluation metrics, in each prediction scenario, to compute the **distance** (denoted Dist.) with the "perfect" prediction performance (AUC=1, TSR=100%, ACC=100% and F1=1). For a prediction approach p , this distance is expressed by:

$$Distance(p) = \sqrt{(1 - AUC(p))^2 + (1 - TSR(p))^2 + (1 - ACC(p))^2 + (1 - F1(p))^2} \quad (3.5)$$

By computing such a distance, we can have an idea about the performance closeness compared with the perfect case and assess how much the tradeoff of having simultaneous good evaluations is ensured.

Regarding all links prediction results at the period $T+1$ and based on the high values of AUC metric (more than 0.9 at most cases), top scores ratio obtained

¹represents to the proportion of links with positive prediction (occurring in the future) which are correctly identified [44]. Based on Table 3.1, the precision is equal to $\frac{TP}{TP+FP}$. This value is determined according to the deduced accuracy value.

²quantifies the ratio of correctly identified links over the occurring links in the future [44]. Referring to Table 3.1, the recall is defined by the expression $\frac{TP}{TP+FN}$. This value is also computed according to the retained accuracy value.

3.4. Performance Evaluation and Simulation Results

Table 3.2: Evaluation metrics for the prediction of all links applied on Dartmouth Campus trace

Prediction Cases		Metrics	All Links Prediction at $T+1$					New Links Prediction				
			AUC	TSR	ACC	F1	Dist.	AUC	TSR	ACC	F1	Dist.
$t=5$ mins	Distributed Case	0.9932	93.70%	99.90%	0.9407	0.0868	0.6834	7.14%	99.81%	0.0063	1.3964	
	Centralized Case	0.9905	93.61%	99.90%	0.9416	0.0871	0.5920	7.04%	99.81%	0.0063	1.4206	
	Collapsed Weighted Tensor	0.9932	93.48%	99.90%	0.9428	0.0870	–	–	–	–	–	
	Common Neighbors	0.9911	83.48%	99.57%	0.6946	0.3474	0.6836	7.46%	99.81%	0.0063	1.3942	
	Salton Index	0.9923	86.96%	99.77%	0.8692	0.1849	0.6834	5.46%	99.81%	0.0042	1.4091	
	Jaccard Index	0.9921	87.38%	99.78%	0.8741	0.1785	0.6830	5.67%	99.81%	0.0042	1.4078	
	Sorensen Index	0.9922	85.88%	99.76%	0.8591	0.1996	0.6834	5.67%	99.81%	0.0042	1.4077	
	Hub Prom. Index	0.9922	86.36%	99.77%	0.8612	0.1948	0.6834	5.57%	99.81%	0.0042	1.4084	
	Hub Depr. Index	0.9919	84.12%	99.73%	0.8404	0.2253	0.6832	5.25%	99.81%	0.0041	1.4107	
	Adamic-Adar Index	0.9923	87.88%	99.79%	0.8787	0.1717	0.6835	8.09%	99.81%	0.0021	1.3931	
	Resource Alloc. Index	0.9926	91.87%	99.85%	0.9176	0.1160	0.6837	7.14%	99.81%	0.0021	1.3993	
	Rooted PageRank Index	0.9889	90.53%	99.81%	0.8864	0.1483	0.5638	3.99%	99.80%	0.0041	1.4505	
$t=10$ mins	Distributed Case	0.9915	90.26%	99.84%	0.9093	0.1334	0.6831	6.09%	99.81%	0.0063	1.4035	
	Centralized Case	0.9883	90.19%	99.84%	0.9091	0.1343	0.6097	6.30%	99.81%	0.0164	1.4134	
	Collapsed Weighted Tensor	0.9900	93.48%	99.84%	0.9032	0.1171	–	–	–	–	–	
	Common Neighbors	0.9890	80.41%	99.52%	0.6575	0.3947	0.6837	6.41%	99.81%	0.0063	1.4012	
	Salton Index	0.9904	84.20%	99.73%	0.8427	0.2232	0.6835	5.36%	99.81%	0.0042	1.4098	
	Jaccard Index	0.9902	83.85%	99.71%	0.8349	0.2312	0.6830	5.46%	99.81%	0.0144	1.4020	
	Sorensen Index	0.9903	83.24%	99.72%	0.8351	0.2353	0.6836	5.46%	99.81%	0.0144	1.4019	
	Hub Prom. Index	0.9903	83.46%	99.73%	0.8317	0.2362	0.6835	5.04%	99.81%	0.0041	1.4120	
	Hub Depr. Index	0.9899	81.17%	99.69%	0.8144	0.2646	0.6834	4.73%	99.81%	0.0041	1.4141	
	Adamic-Adar Index	0.9904	83.87%	99.71%	0.8377	0.2290	0.6838	6.30%	99.81%	0.0063	1.4019	
	Resource Alloc. Index	0.9908	88.34%	99.80%	0.8857	0.1636	0.6838	5.78%	99.81%	0.0063	1.4054	
	Rooted PageRank Index	0.9871	87.67%	99.74%	0.8439	0.1994	0.5855	3.57%	99.80%	0.0020	1.4483	
$t=30$ mins	Distributed Case	0.9813	82.31%	99.69%	0.8261	0.2488	0.6860	5.57%	99.81%	0.0042	1.4078	
	Centralized Case	0.9764	82.56%	99.69%	0.8272	0.2467	0.6180	6.51%	99.81%	0.0042	1.4183	
	Collapsed Weighted Tensor	0.9742	81.49%	99.70%	0.8343	0.2498	–	–	–	–	–	
	Common Neighbors	0.9782	71.25%	99.52%	0.5450	0.5387	0.6871	6.83%	99.81%	0.0042	1.3991	
	Salton Index	0.9799	76.00%	99.57%	0.7535	0.3447	0.6867	5.78%	99.81%	0.0021	1.4077	
	Jaccard Index	0.9796	75.12%	99.55%	0.7454	0.3566	0.6866	5.88%	99.81%	0.0021	1.4071	
	Sorensen Index	0.9797	75.47%	99.56%	0.7471	0.3529	0.6866	5.67%	99.81%	0.0021	1.4085	
	Hub Prom. Index	0.9798	76.02%	99.58%	0.7573	0.3418	0.6865	5.15%	99.81%	0.0021	1.4120	
	Hub Depr. Index	0.9793	73.89%	99.52%	0.7321	0.3747	0.6867	5.78%	99.81%	0.0021	1.4077	
	Adamic-Adar Index	0.9800	74.16%	99.52%	0.7374	0.3690	0.6874	6.83%	99.81%	0.0042	1.3991	
	Resource Alloc. Index	0.9806	80.32%	99.64%	0.7976	0.2830	0.6878	7.14%	99.81%	0.0021	1.3984	
	Rooted PageRank Index	0.9750	78.24%	99.56%	0.7826	0.3086	0.6076	4.52%	99.80%	0.0021	1.4358	

at $T+1$, accuracy, F-Measure and consequently the distance from the best performance, we note that the prediction using the Katz measure (both in centralized and distributed ways) and the collapsed weighted tensor achieve the best performances. It is trivial that links that occur during the period T are the most expected to be effective at $T+1$. This property is well captured by the collapsed weighted tensor as it gives higher weights to recent links. This observation is well interpreted by the top score ratio in period $T+1$ metric. Computing the Katz measure (for the two treated cases) from the collapsed weighted tensor ensures that the links occurring at the period T are identified as the Katz measure gives highest scores to most recent and persistent links. Hence, we clearly observe that performing a prediction using the Katz measure achieves a similar prediction efficiency to estimate occurring links

Chapter 3. Tensor-Based Link Prediction Framework for Mobile Wireless Networks

Table 3.3: Evaluation metrics for the prediction of all links applied on MIT Campus trace

Prediction Cases		All Links Prediction at $T+1$					New Links Prediction				
		AUC	TSR	ACC	F1	Dist.	AUC	TSR	ACC	F1	Dist.
$t=5$ mins	Distributed Case	0.9907	91.48%	99.84%	0.9243	0.1144	0.6668	12.47%	99.45%	0.0071	1.3649
	Centralized Case	0.9929	91.48%	99.84%	0.9223	0.1155	0.6865	11.93%	99.45%	0.0071	1.3637
	Collapsed Weighted Tensor	0.9883	91.48%	99.84%	0.9234	0.1152	–	–	–	–	–
	Common Neighbors	0.9864	79.27%	99.43%	0.6489	0.4080	0.6668	12.47%	99.45%	0.0071	1.3649
	Salt on Index	0.9886	82.87%	99.65%	0.8274	0.2435	0.6666	10.49%	99.45%	0.0053	1.3791
	Jaccard Index	0.9886	82.31%	99.65%	0.8194	0.2531	0.6662	10.58%	99.45%	0.0053	1.3786
	S/orensen Index	0.9885	81.00%	99.64%	0.8169	0.2641	0.6665	10.67%	99.45%	0.0053	1.3779
	Hub Prom. Index	0.9884	81.19%	99.63%	0.8128	0.2657	0.6667	9.96%	99.42%	0.0051	1.3826
	Hub Depr. Index	0.9883	79.93%	99.60%	0.7867	0.2931	0.6665	10.76%	99.45%	0.0071	1.3760
	Adamic-Adar Index	0.9886	84.89%	99.63%	0.8080	0.2446	0.6669	11.57%	99.45%	0.0071	1.3707
	Resource Alloc. Index	0.9891	87.04%	99.75%	0.8803	0.1768	0.6665	11.57%	99.45%	0.0071	1.3708
	Rooted PageRank Index	0.9918	86.24%	99.67%	0.8551	0.2000	0.6885	8.79%	99.44%	0.0035	1.3864
$t=10$ mins	Distributed Case	0.9797	85.18%	99.70%	0.8609	0.2043	0.6685	12.11%	99.45%	0.0071	1.3668
	Centralized Case	0.9809	85.14%	99.70%	0.8609	0.2045	0.6859	11.93%	99.45%	0.0071	1.3639
	Collapsed Weighted Tensor	0.9725	85.14%	99.70%	0.8624	0.2044	–	–	–	–	–
	Common Neighbors	0.9752	74.86%	99.33%	0.5921	0.4798	0.6687	12.29%	99.45%	0.0071	1.3656
	Salt on Index	0.9774	77.07%	99.54%	0.7676	0.3273	0.6686	9.96%	99.45%	0.0089	1.3794
	Jaccard Index	0.9775	77.03%	99.54%	0.7654	0.3291	0.6681	10.31%	99.45%	0.0089	1.3773
	S/orensen Index	0.9774	76.04%	99.53%	0.7620	0.3385	0.6686	10.40%	99.45%	0.0089	1.3766
	Hub Prom. Index	0.9773	76.98%	99.53%	0.7559	0.3363	0.6684	9.60%	99.43%	0.0035	1.3857
	Hub Depr. Index	0.9771	75.09%	99.49%	0.7344	0.3649	0.6686	10.67%	99.45%	0.0106	1.3736
	Adamic-Adar Index	0.9776	79.95%	99.50%	0.7417	0.3278	0.6685	11.75%	99.45%	0.0071	1.3692
	Resource Alloc. Index	0.9780	81.71%	99.62%	0.8176	0.2593	0.6685	11.03%	99.45%	0.0036	1.3763
	Rooted PageRank Index	0.9730	80.59%	99.50%	0.7922	0.2857	0.6777	8.61%	99.44%	0.0035	1.3900
$t=30$ mins	Distributed Case	0.9589	73.31%	99.41%	0.7131	0.3940	0.6751	10.13%	99.45%	0.0054	1.3793
	Centralized Case	0.9578	73.76%	99.41%	0.7545	0.3619	0.6816	10.22%	99.45%	0.0054	1.3772
	Collapsed Weighted Tensor	0.9316	73.51%	99.41%	0.7149	0.3952	–	–	–	–	–
	Common Neighbors	0.9544	64.20%	99.11%	0.4664	0.6442	0.6757	10.76%	99.45%	0.0036	1.3764
	Salt on Index	0.9563	68.04%	99.30%	0.6411	0.4826	0.6755	8.52%	99.45%	0.0071	1.3885
	Jaccard Index	0.9565	68.94%	99.33%	0.6799	0.4482	0.6747	7.89%	99.45%	0.0107	1.3903
	S/orensen Index	0.9563	67.96%	99.30%	0.6634	0.4668	0.6754	8.43%	99.45%	0.0107	1.3866
	Hub Prom. Index	0.9563	66.45%	99.29%	0.6231	0.5065	0.6755	7.71%	99.44%	0.0053	1.3952
	Hub Depr. Index	0.9559	66.98%	99.29%	0.6702	0.4688	0.6750	8.52%	99.45%	0.0124	1.3849
	Adamic-Adar Index	0.9567	68.73%	99.26%	0.6107	0.5013	0.6752	10.67%	99.45%	0.0054	1.3758
	Resource Alloc. Index	0.9573	71.47%	99.36%	0.6868	0.4259	0.6753	9.96%	99.45%	0.0124	1.3753
	Rooted PageRank Index	0.9290	63.67%	98.98%	0.6274	0.5253	0.6608	4.84%	99.42%	0.0017	1.4203

at the period $T+1$ (similar distances with the perfect prediction). Meanwhile, the other prediction metrics show a lower performance in the context of predicting such links. Indeed, they quantify the relationship of a pair of nodes without seeing if a direct link connects them. For example, two isolated neighbors (or having few common neighbors) would have a weak score even if they were connected. On the other hand, two other nodes can be 2-hop neighbors and share several common neighbors which means that the score in this case is relatively high even if no link connects the corresponding pair of nodes.

Nevertheless, these prediction approaches depict an interesting capacity to predict future links (which have occurred for the first time after the period T) as they allocate the highest values to nodes which are relatively close in terms of hops and

3.4. Performance Evaluation and Simulation Results

Table 3.4: Evaluation metrics for the prediction of all links applied on Infocom 2006 trace

Prediction Cases		All Links Prediction at $T+1$					New Links Prediction				
		AUC	TSR	ACC	F1	Dist.	AUC	TSR	ACC	F1	Dist.
$t=5$ mins	Distributed Case	0.9675	63.97%	98.27%	0.6147	0.5288	0.6927	10.84%	97.14%	0.0682	1.3261
	Centralized Case	0.9686	63.24%	98.25%	0.6145	0.5339	0.6963	9.64%	97.10%	0.0233	1.3651
	Collapsed Weighted Tensor	0.9650	64.71%	98.33%	0.6400	0.5056	–	–	–	–	–
	Common Neighbors	0.8770	36.76%	97.20%	0.2766	0.9691	0.6835	10.84%	97.07%	0.0870	1.3151
	Salton Index	0.9082	43.38%	97.28%	0.0959	1.0710	0.6810	10.84%	97.03%	0.0860	1.3164
	Jaccard Index	0.9068	42.65%	97.32%	0.1216	1.0535	0.6857	10.84%	96.86%	0.0217	1.3608
	Sorensen Index	0.9085	43.38%	97.26%	0.0828	1.0821	0.6841	12.05%	96.93%	0.0222	1.3529
	Hub Prom. Index	0.8991	36.76%	97.22%	0.1401	1.0725	0.6792	10.84%	96.96%	0.0440	1.3464
	Hub Depr. Index	0.9054	41.91%	97.26%	0.1074	1.0695	0.6841	10.84%	97.00%	0.0444	1.3449
	Adamic-Adar Index	0.8956	37.50%	97.34%	0.1783	1.0380	0.6825	10.84%	97.14%	0.1277	1.2874
	Resource Alloc. Index	0.9092	42.65%	97.38%	0.3420	0.8780	0.6819	10.84%	97.14%	0.1087	1.3005
	Rooted PageRank Index	0.9282	22.06%	97.34%	0.1103	1.1853	0.6548	9.64%	96.19%	0.0180	1.3789
$t=10$ mins	Distributed Case	0.9260	48.81%	97.30%	0.4279	0.7717	0.7137	10.84%	96.86%	0.0625	1.3254
	Centralized Case	0.9234	48.81%	97.30%	0.4279	0.7720	0.6740	9.64%	96.89%	0.0220	1.3712
	Collapsed Weighted Tensor	0.9144	49.40%	97.38%	0.4502	0.7525	–	–	–	–	–
	Common Neighbors	0.8464	19.64%	96.58%	0.0349	1.2657	0.7017	9.64%	96.82%	0.0215	1.3653
	Salton Index	0.8612	27.38%	96.58%	0.1443	1.1314	0.6992	9.64%	96.82%	0.0215	1.3658
	Jaccard Index	0.8589	25.00%	96.60%	0.1451	1.1465	0.7048	8.43%	96.65%	0.0204	1.3735
	Sorensen Index	0.8611	27.98%	96.58%	0.1616	1.1145	0.7057	9.64%	96.86%	0.0217	1.3642
	Hub Prom. Index	0.8559	23.21%	96.37%	0.0330	1.2437	0.6945	7.23%	96.89%	0.0220	1.3825
	Hub Depr. Index	0.8589	26.79%	96.58%	0.0349	1.2200	0.7087	9.64%	96.75%	0.0211	1.3641
	Adamic-Adar Index	0.8506	22.62%	96.60%	0.0884	1.2055	0.6946	9.64%	96.86%	0.0217	1.3667
	Resource Alloc. Index	0.8695	31.55%	96.62%	0.1458	1.1029	0.6891	9.64%	97.03%	0.0230	1.3670
	Rooted PageRank Index	0.9259	43.45%	96.70%	0.1011	1.0651	0.6194	10.84%	95.36%	0.0148	1.3830
$t=30$ mins	Distributed Case	0.8435	26.70%	95.60%	0.0184	1.2358	0.6620	7.23%	96.86%	0.0217	1.3903
	Centralized Case	0.8420	26.21%	95.60%	0.0184	1.2389	0.6365	7.23%	96.58%	0.0200	1.3980
	Collapsed Weighted Tensor	0.8420	28.64%	95.60%	0.0184	1.2246	–	–	–	–	–
	Common Neighbors	0.7768	09.71%	95.49%	0.0352	1.3409	0.6471	7.23%	96.82%	0.0215	1.3941
	Salton Index	0.7959	11.17%	95.57%	0.0092	1.3470	0.6428	6.02%	97.00%	0.0227	1.4024
	Jaccard Index	0.7925	10.68%	95.67%	0.0094	1.3506	0.6459	4.82%	96.79%	0.0213	1.4107
	Sorensen Index	0.7956	11.65%	95.61%	0.0093	1.3438	0.6501	7.23%	97.00%	0.0227	1.3925
	Hub Prom. Index	0.7920	10.19%	94.29%	0.0546	1.3217	0.6439	4.82%	96.72%	0.0208	1.4116
	Hub Depr. Index	0.7915	12.62%	95.65%	0.0186	1.3312	0.6451	7.23%	96.86%	0.0217	1.3945
	Adamic-Adar Index	0.7891	10.68%	95.67%	0.0094	1.3511	0.6399	7.23%	96.82%	0.0215	1.3960
	Resource Alloc. Index	0.8050	12.62%	95.71%	0.0095	1.3358	0.6464	8.43%	96.79%	0.0213	1.3865
	Rooted PageRank Index	0.4589	15.53%	95.86%	0.0474	1.3840	0.4905	4.82%	95.84%	0.0165	1.4610

consequently social ties (if two nodes share a high number of common neighbors or an important proportion of common neighbors, they are expected to be close and then a link may occur between them in the future). We can notice that the prediction performance of such links is obviously less efficient than predicting all links. This observation is highlighted by a higher distance with the perfect prediction. The Katz measure follow a similar way to predict future links. The highest scores in this case are for pairs of nodes which are connected through 2-hop paths (as a direct link has never existed). The score gets higher with a more important set of common neighbors for two given nodes. This information totally agrees with the result advanced by Wang et al. [110] which stipulates that a new link has more chance to occur between two nodes if they are separated by a distance of 2 hops.

From this perspective, the use of the Katz measure enables us to reach one of the best performances in predicting new links. The performance of predicting such links is slightly impacted by the length of the tracking period (compared to the all links prediction in $T+1$ where the shorter the tracking period is, the better the prediction gets). This observation can be interpreted by the weak proportion of new links (that occur beyond the period T) compared to the overwhelming potential links (which never occurred during T periods and can occur starting from the period $T+1$). It is also important to mention that the collapsed weighted tensor has no ability to predict new links as it allocates weights only for links that have been tracked in the historical data. Then, any potential new link has no weight.

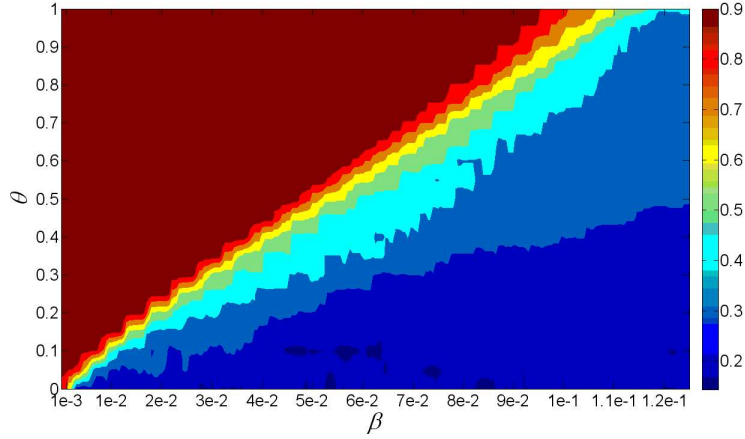
After evaluating both link prediction abilities, it is clear that performing the prediction using the Katz measure is the only way to simultaneously guarantee high efficiency in predicting all links in the period $T+1$ and estimating new links that have not occurred during the historical, compared to all the other approaches involved in the evaluation. Moreover, according to all the evaluation metrics and in particular the distance with the perfect performance, the results confirm that the proposal of computing the Katz measure relying only on the local information achieves similar or even better performance than the centralized case.

The Importance of the Parameters θ and β

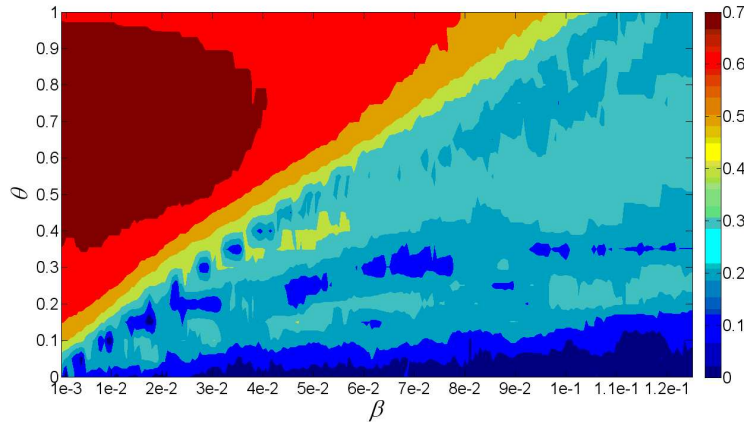
It is important to underline that the efficiency of our mechanism is dependent on the values of the parameters θ (a parameter used to adjust the weight of recentness and is between 0 and 1 to compute the collapsed weighted tensor in eq. (4.1)) and β (a damping factor used in the computation of the Katz scores matrix \mathbf{S} expressed by eq. (4.2)). We depict in Fig. 3.5(a) and Fig. 3.5(b) the top scores ratio at $T+1$ and the AUC, respectively, obtained for different values of θ and β . We can note that the values set to θ (i.e. 0.2) and to β (i.e. 0.001) enable us to reach quite an efficient level of prediction. These results are relative to a prediction set performed on the MIT Campus trace with the distributed version of the prediction framework.

After the investigation into the value to be chosen for the parameter β , we found that the convergence of the Katz measure is closely tied to the spectral radius of the adjacency matrix, as mentioned in [46]. In fact, β must be greatly inferior to the reverse of the latter value. In our case, we use the collapsed weighted tensor which can be considered as an adjacency matrix with weights of links. If we take the example of the collapsed weighted tensor of the scenario where we use the Infocom 2006 trace, with a period of tracking t equal to 5 minutes and a number of tracking periods equal to 72, we find that the lowest reverse of the spectral radius

3.4. Performance Evaluation and Simulation Results



(a) Area Under the ROC Curve



(b) Top Scores Ratio at $T+1$

Figure 3.5: Prediction performance of the tensor-based technique distributed version for different values of β and θ

is around 0.01. This lowest value is observed when θ tends to 0. With higher values of θ , the interval of values of β with good performance is larger (AUC higher than 0.8 and TSR higher than 0.6). Indeed, when the parameter θ is important, taking into consideration the past event is less significant. Then, the weights in the collapsed weighted tensor get lower which decreases the value of the spectral radius and increases the value of the highest β which ensures convergence. Hence, the interval of satisfying values of β is larger.

3.4.3 Discussion

In wireless networks and specifically in intermittently connected ones, it is important to exploit social relationships that influence the nodes' mobility. Taking advantage of the social aspect within these networks could ensure a better routing strategy and therefore improve the packet delivery rate and reduce latency. Through our proposal, we aim to track eventual similarities between mobility patterns of nodes and exploit them intelligently for a better link prediction.

As seen earlier, a link occurrence between two nodes is more likely when they have similar social behavior. Then, identifying nodes that have a similar mobility pattern could help to predict effective links between nodes in the future. The more accurate the link prediction is, the more optimized the routing scheme could become. In fact, an efficient link prediction would help to make better decisions in the forwarding process. For example, a node would rather decide to postpone sending a packet to a current next hop because the link prediction scheme estimates that a better forwarder (closer to the destination for example) is going to appear in the immediate future. Also, link prediction could prevent buffer overloading. Indeed, an overloaded node would rather drop a packet if the link prediction scheme indicates that there will be no possible route toward the destination in the future and before the packet's TTL expires. Through this approach, we can get quite efficient prediction results.

The results detailed previously show that the tensor-based link prediction framework can identify more than 90% of links (with a tracking period of 5 minutes). The link prediction relies on measuring the similarity of the mobility of nodes. Song et al. [101] investigated the limits of predictability in human mobility. Relying on data collected from mobile phone carriers, they found that 93% of user mobility is potentially predictable. The best predictability percentage reached by our approach analogously joins the conclusion of Song et al.

We have also shown through simulations that prediction efficiency is similar whether the Katz measures are computed in a centralized and distributed manner. As we have explained, the distributed scheme is only able to maintain high scores (link occurrence is likely) as nodes record neighbors at one and two hops. The seeming lack of information does not impact on predicting effectiveness. This observation also tallies with the conclusion of Acar and al. [8]. Indeed, in the data mining context, they tried to make the method scalable and proposed the Truncated Katz technique (expressed by eq. (11) in [8]). It consists in determining Katz scores replacing the collapsed weighted tensor by a low-rank approximation one. The results show that this latter technique retains a high prediction efficiency. Hence,

3.4. Performance Evaluation and Simulation Results

Table 3.5: Evaluation metrics for the prediction of all links in $T+1$ applied on Infocom 2006 trace

Prediction Cases		Metrics				
		All Links Prediction at $T+1$				
		AUC	TSR	ACC	F1	Dist.
$t=5\text{mins}$	Distributed Case and 1-hop knowledge	0.9722	64.71%	98.51%	0.6870	0.4728
	Distributed Case and 2-hop knowledge	0.9675	63.97%	98.27%	0.6147	0.5288
	Collapsed Weighted Tensor	0.9650	64.71%	98.33%	0.6400	0.5056
$t=10\text{ mins}$	Distributed Case and 1-hop knowledge	0.9142	49.40%	97.40%	0.4769	0.7333
	Distributed Case and 2-hop knowledge	0.9260	48.81%	97.30%	0.4279	0.7717
	Collapsed Weighted Tensor	0.9144	49.40%	97.38%	0.4502	0.7525
$t=30\text{ mins}$	Distributed Case and 1-hop knowledge	0.8265	28.16%	95.65%	0.0094	1.2367
	Distributed Case and 2-hop knowledge	0.8435	26.70%	95.60%	0.0184	1.2358
	Collapsed Weighted Tensor	0.8420	28.64%	95.60%	0.0184	1.2246

restricting the scores computation on most weighted links (in terms of recentness and duration) does not incur dramatic consequences on the prediction efficiency.

To perform prediction with local information, we assumed for the computation of scores that nodes know their 2-hop neighbors. It is obvious that exchanging information between nodes about neighbors causes additional overhead and consequently more solicited resources. From this perspective, a question can be highlighted: would the tensor-based link prediction method remain effective if the knowledge of nodes is limited to the direct neighbors? To answer this question, we take into consideration the scenario where the distributed computation of scores is based on knowledge of 1-hop neighbors and we compare it to the scenario which uses the 2-hop knowledge when the prediction is interested in the links in the period $T+1$. We also considered the prediction performance of the collapsed weighted tensor. We used the Infocom 2006 trace and kept the same parameters used in the previous simulations (see subsection 3.4.1). The comparison is made with the evaluation metrics described in the subsection 3.4.2. The results are reported in Table 3.5.

When the knowledge is limited to the neighbors at one hop, the closeness only means that there is a direct link between two nodes. Thereby, computing the scores through the tensor-based link prediction framework becomes too similar to determining weights of links with the collapsed weighted tensor. Then, limiting the knowledge to 1-hop neighbors deprive our proposal of its ability to estimate the occurrence of new links beyond the period T . Nonetheless, such knowledge allows better link prediction in the period $T+1$ compared to the other prediction methods involved in the comparison, especially with 5-minute and 10-minute tracking periods. As a reminder, the collapsed weighted tensor allocates a weight for a pair of nodes considering the recentness and the persistence of their contacts. Meanwhile, our proposal, in the case of knowledge of 1-hop neighbors, computes the scores in the same way and includes an additive value if there are paths connecting the same

pair of nodes. That is the reason behind obtaining a similar TSR metric with the collapsed weighted tensor. Moreover, the lack of the 2-hop knowledge enables the tensor-based link prediction framework to achieve better accuracy by setting the scores of all potential links (which did not exist between period 1 and period T and that can occur starting from period $T+1$) to 0. Indeed, the majority of potential links are not expected to occur as the contacts are influenced by the social ties (a node would meet a low percentage of the other nodes). That's why the accuracy and the F-Measure (F1) are better than the other case (and especially the case of distributed computation of scores with knowledge of 2-hop neighbors) which brings about shorter distance with the perfect performance. With longer tracking periods, the prediction with such knowledge loses precision and becomes very similar to those of the other approaches (similar distance).

From this perspective, it is not clear which is the best scenario to be retained if we compare the cost of exchanging local information between nodes to the cost of ignoring prediction of new links. Future simulations and real deployments will enable us to determine which setting is preferable.

3.5 Conclusion

Human mobility patterns are mostly driven by social intentions and correlations appear in the behavior of people forming the network. These similarities highly govern the mobility of people and then directly influence the structure of the network. The knowledge about the behavior of nodes greatly helps in improving the design of communication protocols. Intuitively, two nodes that follow the same social intentions over time promote the occurrence of a link in the immediate future.

In this chapter, we presented a link prediction technique inspired by data-mining and exploit it in the context of wireless networks. Our contribution in this chapter, a new link prediction technique for the intermittently connected wireless networks, is designed through two major steps. First, the network topology is tracked over several time periods in a tensor. Secondly, after collapsing the structural information, a Katz measure is computed for each pair of nodes as a score. A high score means similar moving patterns implying the closeness of the nodes and indicates that a link occurrence is likely in the future.

Through the link prediction evaluation, we obtained relevant results that attest the efficiency of our contribution. We can summarize them in the following points:

- The tensor-based link prediction technique is quite efficient. The result are supported by the ROC curves and the evaluation metrics (AUC metric, top

scores ratio (at $T+1$ or in the future), accuracy, F-Measure and consequently the distance with the perfect prediction).

- Applied on real traces, the proposed prediction technique provides more accurate results with shorter tensor tracking periods. In fact, short periods make the tracking more sensitive in terms of identifying fleeting and persistent contacts. Predicting through long tracking periods is mostly efficient with only long contacts.
- The link prediction framework guarantees good performance when prediction is applied to all links in the period $T+1$. Moreover, it has the capacity to identify some future links which never occur during the tracking of contacts in order to build the tensor.
- Applying the prediction technique in a distributed way (nodes only know their neighbors at most at two hops) achieves similar predicting performance compared to the use of the same framework in a centralized way (an entity has full-knowledge about network structure over time).
- The temporal tensor-based link prediction described in this chapter is based on an encounter metric which takes into account the occurring contacts at the same location and at the same time. We provided a performance comparison with other similar approaches and we found that our proposal is the only one that is able to ensure the tradeoff of reaching the highest performances for both predicting all links at the period $T+1$ and estimating some new links that would appear in the future.

Good link prediction offers the possibility to further improve opportunistic packet forwarding strategies by making better decisions in order to enhance the delivery rate or limit latency. Therefore, it will be relevant to supply some routing protocols with prediction information and to assess the contribution of our approach in enhancing the performance of the network especially as we propose an efficient distributed version of the prediction method. The proposed technique also motivates us to inquire into future enhancements as a more precise tracking of the behavior of nodes and a more efficient similarity computation. This willingness is rigorously detailed in the following chapter.

Chapter 4

Improving Link Prediction in Mobile Wireless Networks by Considering Link and Proximity Stabilities

Contents

4.1	Introduction	60
4.2	Related Work	61
4.3	Tensor-Based Link Prediction Framework: A Reminder	62
4.3.1	Notation	63
4.3.2	Overview on Tensor-Based Link Prediction Technique	63
4.3.3	Matrix of Scores Computation	64
4.3.4	Matrix of Scores Interpretation	65
4.4	How To Quantify Link and Proximity Stabilities and How To Use Them?	66
4.4.1	How Can Link and Proximity Stabilities Improve Link Prediction?	66
4.4.2	The Entropy Measure	67
4.4.3	Quantifying Link and Proximity Stabilities by Means of Time Series Entropy Estimation	70
4.4.4	Joining the Entropy Estimations to the Tensor-Based Link Prediction Framework	71
4.5	Simulations Scenarios and Performance Evaluation	75
4.5.1	Simulation Traces	75

4.5.2 Simulation Results and Performance Analysis	78
4.6 Conclusion	84

4.1 Introduction

Disruption Tolerant Networks (DTN) paradigm is an emerging wireless networking application where we have to deal with sparse and intermittent connectivity. In order to achieve a reasonable packet delivery rate, we have to rely on opportunistic or mobility-assisted routing, where messages are forwarded only when two nodes are in contact. Consequently, the packet delivery rate at the destination is strongly tied to the network structure during the forwarding process. Basically, as these networks are human-centered, the mobility patterns are governed by human behavior. Such a behavior highly impacts on the structure of the network as shown in [29, 56]. Moreover, it has been demonstrated in [57, 105, 115] that the human mobility is directed by social intentions that the network users share at the spatial and temporal levels. When the intentions of some people are correlated (to be present in the same locations at the same time), this favors their meeting and thereby the occurrence of links between them. In this way, mathematical models have been proposed to characterize the inter-contact time between two people through statistical analysis [29, 63]. Thus, it is crucial to better understand how links are created and to track their properties in order to design efficient communication protocols.

To analyze the network topology evolution, it is important to rely on records that describe the status of each link over time during a tracking period [22]. Extracting information about correlations between the willingness of people carrying the nodes is an important support which provides insights for predicting links. From this perspective, we have proposed in the previous chapter and in [117] a tensor-based link prediction technique. Our approach is based on a spatio-temporal framework that tracks the contacts between nodes. Hence, tracking the occurrence of links over successive time periods has enabled us to detect the degree of spatial closeness between the network users and then quantify their behavior similarity. Afterwards, this parameter has been used as an indicator to predict the occurrence of links in the immediate future.

In this chapter, we aim to improve the performance of the tensor-based link prediction technique by refining the measure of behavior similarity. As we are convinced that link prediction enhances the performance of communication protocols, we want the feedback provided by our framework to be the most reliable possible. We find

that the link and the proximity (at two hops) stabilities can also be considered as parameters to predict future links. Indeed, when some network users are related by strong social ties, the link between them tends to be persistent. In other words, when two network users have high correlated behaviors, their closeness is expected to lengthen and to remain stable. When nodes have information about their neighbors at two hops, the stability of such proximity, joined to the tensor-based link prediction method, also leads to the performance of more efficient link prediction. We detail how proximity stability can be beneficial later. Besides, measuring such stability measures is a key parameter which is interesting to exploit. To quantify them, we propose a metric inspired by the Lempel-Ziv estimator [127] which converges to the entropy of a time series. To enhance the prediction performance of the tensor-based link prediction framework, we join the stability measures with its outputs by designing new prediction measures that identify future links as the ones which are relative to users that have strong (behavior similarity) and stable (proximity stability) closeness. Afterwards, we assess if our proposal really contributes to improve the prediction performance of the tensor-based link prediction method.

The chapter is organized as follows: Section 4.2 briefly presents the related work. In Section 4.3, we give an overview of the tensor-based link prediction technique. Section 4.4 introduces the measurement of the link and the proximity stabilities, presents the estimator used to approximate them and details some combinations of the stability measures with our link prediction technique in order to provide new prediction metrics. In section 4.5, we present the simulation scenarios used to evaluate the contribution of these stability measures, define how to evaluate them and we analyze the obtained results. Finally, we conclude the chapter in Section 4.6.

4.2 Related Work

Due to the opportunistic aspect of forwarding in intermittently connected wireless networks and as these networks are basically human-centered networks, Social Network Analysis (SNA) [111] has been used in order to provide more efficient communication protocols. It has been applied to track and understand the relationships between the network entities and to extract structural information about the network (network robustness, topology variance, emerging communities, ...). Several communication protocols such as [59, 35, 55] have been based on SNA: they have used the centrality metrics proposed in [111, 84, 60, 31] and/or have exploited community detection mechanisms such as [111, 20, 81, 85]. These two categories of “social tools” have been defined as the two major concepts of SNA in the design of

4.3. Tensor-Based Link Prediction Framework: A Reminder

wireless ad-hoc network protocols by Katsaros et al. [64].

Tracking the social ties between the network entities has led to the design of techniques for link prediction. The link prediction in social networks has been addressed in the data-mining context [8, 38, 109, 74] and recently for the community-based communication networks [117, 110]. These works have highlighted salient measures that make possible to predict wireless links between network users. These metrics determine if an occurrence of a link is likely by quantifying the degree of proximity of two nodes (Katz measure [65], the number of common neighbors, the Adamic-Adar measure [10], the Jaccard's coefficient [61, 94], ...) or by computing the similarity of their mobility patterns (through such metrics as the spatial cosine similarity, the co-location rate, ...). The efficiency of the Katz measure has been especially emphasized compared to other prediction metrics in [74, 110]. Thanks to the SNA, it has been highlighted that the relationships between individuals has a major impact on the structure of the network [29, 56].

For the human-centered wireless networks, the social closeness between some people influences their mobility patterns. In [101], Song et al. have demonstrated that human mobility is potentially predictable in 93% of cases by using the traces of mobile users of a cellular phone network. This limit has been investigated through an estimation of the entropy proposed by Ziv and Lempel [127]. The entropy estimation has been applied on the sequence of visited locations for each mobile phone user during a tracking period. Song et al. have been motivated by the findings of [80] which has highlighted that the entropy is a very appropriate metric to measure the degree of predictability of such sequences.

In this chapter, we aim to stress that quantifying the degree of proximity of two nodes joined to the stability of their relationship enables us to enhance the link prediction performance of our method [117] presented in the previous chapter. For this objective, we redesign the tensor-based link technique by taking into consideration the feedback provided by measuring the link and the proximity at two hops stabilities. Hence, we describe how to quantify the stability and we propose some designs for new prediction metrics.

4.3 Tensor-Based Link Prediction Framework: A Reminder

The human mobility patterns highlight correlations in the behavior of network users. The researches done in [29, 57] have emphasized that the human mobility depicts a spatio-temporal regularity and the claims advanced in [56, 105] have demonstrated

that social ties characterizing the relationships between users are translated by correlations between human mobility patterns. From the perspective that these relationships govern how the network is structured, the tensor-based link prediction framework aims at identifying and exploiting these correlations to perform prediction. This is made by tracking and quantifying the degree of proximity of each pair of nodes. If two nodes have common intentions at the spatio-temporal level, they tend to be closer to each other and then nearby. Thus, a link occurrence between them is likely.

Predicting future links based on their social closeness is a challenge that is worth an investigation. Indeed, a good link prediction technique contributes to improving the opportunistic forwarding of packets and also enhances the delivery rate and/or decreases latency. Moreover, it helps to avoid situations where packets overload the queue of the nodes that are unable to forward these packets towards their final destinations. Motivated by the enhancement that can provide the prediction to communication protocols, we propose the tensor-based link prediction framework which we describe and explain in this section.

4.3.1 Notation

Scalars are denoted by lowercase letters, e.g., a . Vectors are denoted by boldface lowercase letters, e.g., \mathbf{a} . Matrices are denoted by boldface capital letters, e.g., \mathbf{A} . The r^{th} column of a matrix \mathbf{A} is denoted by \mathbf{a}_r . Higher-order tensors are denoted by bold Euler script letters, e.g., \mathcal{T} . The n^{th} frontal slice of a tensor \mathcal{T} is denoted \mathbf{T}_n . The i^{th} entry of a vector \mathbf{a} is denoted by $\mathbf{a}(\mathbf{i})$, element (i, j) of a matrix \mathbf{A} is denoted by $\mathbf{A}(i, j)$, and element (i, j, k) of a third-order tensor \mathcal{T} is denoted by $\mathbf{T}_i(j, k)$.

4.3.2 Overview on Tensor-Based Link Prediction Technique

In order to quantify the degree of spatial closeness of two nodes, we compute the Katz measure [65]. It is used in sociometry and, in the case of wireless networks, expresses the similarity of the behavior (i.e. mobility patterns) or the degree of proximity of two nodes. The Katz measure is dependent on the lengths of paths (one-hop or multi-hop paths) that separate these two nodes. It is computed from a third-order tensor which records the network statistics (i.e. occurrence of links between each pair of nodes during different tracking periods). Hence, a tensor \mathcal{Z} consists in a set of adjacency matrices which form successive slices. Each slice corresponds to the contacts which occurred during a period of time t ($\forall t, 1 \leq t \leq T$ where T is the total number of periods). Then, we determine a collapsed weighted

4.3. Tensor-Based Link Prediction Framework: A Reminder

tensor \mathbf{X} which allocates weights to the links according to the frequency of their occurrence and their recentness. The use of such a collapsing way is motivated by the results highlighted in [8]. Indeed, predicting links in the period $T+1$ using such a method to aggregate the data achieved best link prediction performance. Applying the Katz measure on the collapsed weighted tensor allows us to obtain the matrix of scores \mathbf{S} which quantifies the behavior similarity or the degree of proximity for each pair of nodes. An example provided by Fig. 4.1 details how the network structure is tracked and how the prediction scores are determined for the period $T+1$.

4.3.3 Matrix of Scores Computation

For each node, computing the matrix of scores is performed through two salient steps. Firstly, the link prediction framework records the adjacency matrices in a tensor \mathcal{Z} and determines the collapsed weighted tensor (or matrix) \mathbf{X} . Secondly, it applies the Katz measure on the matrix \mathbf{X} to obtain the matrix of Katz scores \mathbf{S} .

We consider that the data is collected into the tensor \mathcal{Z} . The slice $\mathbf{Z}_p(i, j)$ describes the status of a link between a node i and a node j during a time period between $[(p-1) \cdot t, p \cdot t[$ ($p > 0$) where $\mathbf{Z}_p(i, j)$ is 1 if the link exists during the time period p and 0 otherwise. The tensor is formed by a succession of adjacency matrices \mathbf{Z}_1 to \mathbf{Z}_T where the subscript letters designate the observed period. To determine the collapsed weighted tensor, we apply the following expression:

$$\mathbf{X}(i, j) = \sum_{p=1}^T (1 - \theta)^{T-p} \mathbf{Z}_p(i, j) \quad (4.1)$$

where the matrix \mathbf{X} is the collapsed weighted tensor of \mathcal{Z} , and θ is a parameter used to adjust the weight of recentness and its value is between 0 and 1.

The matrix \mathbf{X} can be seen as a generalized adjacency matrix with weighted links. A node can determine the behavior similarity with the other nodes by computing the Katz measure. This similarity between two nodes i and j is determined by the length of paths that connect i to j and is impacted by the weight of each path length.

Then, Eq.(4.2) is used to compute matrix of Katz scores \mathbf{S} as following:

$$\mathbf{S} = \sum_{\ell=1}^{+\infty} \beta^\ell \cdot \mathbf{X}^\ell = (\mathbf{I} - \beta \cdot \mathbf{X})^{-1} - \mathbf{I} \quad (4.2)$$

where β is a user defined parameter which is strictly positive and β^ℓ is the weight of a ℓ hops path length. It is clear that the longer the path is, the lower its weight gets. The matrix \mathbf{I} is the identity matrix and \mathbf{X} is the obtained collapsed weighted

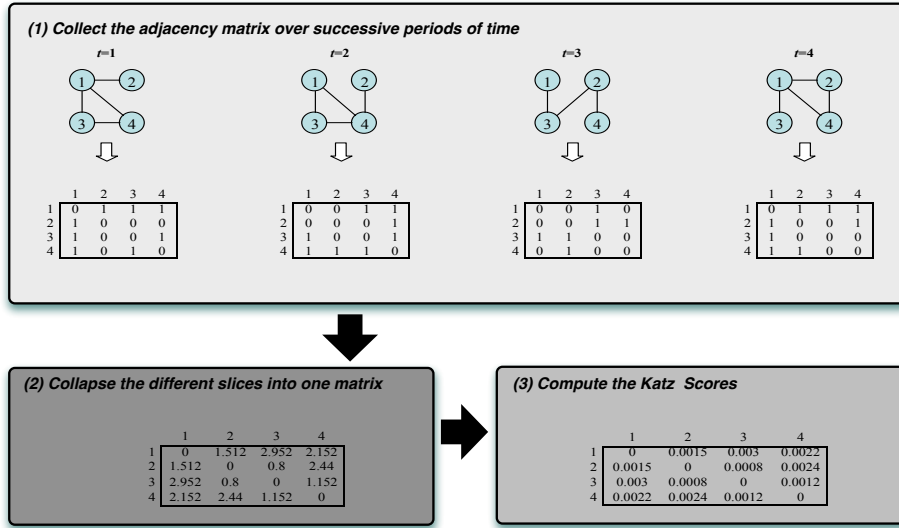


Figure 4.1: Reminder of the steps followed for matrix \mathbf{S} computation

tensor.

We depict as previously mentioned in Fig. 4.1 an example which details the two major steps described before. We take into consideration a network consisting of 4 nodes and having a dynamic topology over 4 time periods and we highlight how similarity scores are obtained (θ and β are respectively set to 0.2 and 0.001, as motivated in the previous chapter, for the example and later for the simulations). In this example, we assume that all nodes have the full knowledge of the network structure.

4.3.4 Matrix of Scores Interpretation

The social ties between each pair of nodes are quantified by the similarity measure given by $\mathbf{S}(i, j)$. When two nodes share an important score, this means that the paths that connect them tend to be short. Two nodes separated by few hops rimes with a geographical closeness which underlines a correlation between their behavior and emphasizes a spatial proximity. Therefore, the link occurrence between these two nodes is strongly plausible. Otherwise, if the similarity score is low or null, the two corresponding nodes share occasional or even no correlation in their behavior.

4.4 How To Quantify Link and Proximity Stabilities and How To Use Them?

As detailed previously, to predict links in intermittently connected wireless networks, we have used a tensor-based framework to record network structure for a number of periods and have measured Katz scores for each pair of nodes to quantify their behavior similarity. In [117], we have shown that the tensor-based link prediction framework achieves similar performance either when done in a centralized way (i.e. assuming that a central unit records the network statistics and performs prediction) or in a distributed way (i.e. the nodes predict links relying on local information). Obviously, we are interested in the distributed application of the prediction framework as we are concerned by the intermittently connected wireless networks and the effectiveness of our proposal in this case is strongly prominent. Nonetheless, we strongly conceive that we can further improve the prediction efficiency. We think that the prediction becomes more precise if we take into consideration other salient characteristics of ties between nodes. Indeed, we guess that focusing on the stabilities of links and of proximity at two hops are worth investigating and the predictability of human mobility made in [101] has further motivated us to continue in this direction.

4.4.1 How Can Link and Proximity Stabilities Improve Link Prediction?

Regarding link stability, when two nodes have strong social ties, a link occurrence between them is likely. If a link between them occurs, it is expected to be persevering and then stable. Indeed, they are expected to be close for a fairly substantial time as they share similar behavior. On the other hand, when the proximity at two hops between two nodes is stable, it can be interpreted in two separate manners. When the stability is expressed by the absence of proximity at two hops, this means that the two nodes are either tied by a link or separated by more than two hops. In this case, the outputs of the tensor-based link prediction technique can be used to identify if the corresponding nodes are directly tied. When this stability is expressed by two nodes constantly separated by two hops, the information provided by the tensor-based link prediction technique is able to attest that a link occurrence is unlikely. Moreover, the corresponding Katz score is expected to be lower than those of pair of nodes having a link between them. If the proximity is unstable, it is obvious to conclude that the link occurrence is unlikely. Indeed, even if there is a link between the two nodes, its occurrence is intermittent and then the link is

unstable. The information provided by the Katz measure gives indications about the real state of the link.

Hence, measuring such stability parameters is paramount to enhance the performance of the tensor-based link prediction framework. Therefore, we should refine the tensor-based link prediction framework to make it sensitive to these parameters. To measure stability and predictability in complex networks, the entropy measure has been widely used in the literature.

4.4.2 The Entropy Measure

The measure of entropy has been used in several works. Its definition has varied from one work to another. We will expose a little overview on the origin of entropy and a brief description of works having used this measure especially in the context of social and community-based networks.

Original Definition of Entropy

The concept of entropy was introduced for the first time by Clausius [32] as a unique measure of reversible change in thermal energy concerning the absolute temperature. He focused on the macroscopic behavior of chemical microscopic reactions and proposed thermodynamic entropy. Based on Clausius works, Boltzmann defined the combination microstates statistic entropy [21] as:

$$S = -k_B \sum_i p_i \ln p_i \quad (4.3)$$

where p_i is the probability that the microstate i is verified during system fluctuations and k_B is the constant of Boltzmann. This definition is applied to characterize the order in the system and how the system self-organizes among different entities.

Later, Shannon introduced the concept of information entropy H [97]. This measure has been used to quantify the capacity of a transmission channel and has been extended to other domains.

Using Entropy to Characterize the Information in a Network

Many works propose the use of entropy, for different aims. In [75], Lu et al. have presented, firstly, the principles of self-organization. Indeed, wireless networks use the self-organization to minimize configuration needs, to facilitate the deployment of the network and to support applications and services. The recourse to self-organization schemes permits to improve the order in the network. This organization take place on two levels: microscopic (logic links between nodes) and macroscopic

4.4. How To Quantify Link and Proximity Stabilities and How To Use Them?

(formation of flexible structure). Secondly, Lu et al. justify the use of entropy with three major reasons:

- The organization in a wireless network is similar to a thermal dynamic system.
- Many metrics have been proposed to evaluate self-organization strategies (protocol overhead, algorithmic complexity, . . .) but they do not give any idea on order degree.
- The statistic entropy used in thermodynamics is a key measure because it describes the behavior of self-organization protocols compared to changes of inherent parameters in the network as the reliability of links and nodes.

Lu et al. consider the entropy as a system macroscopic description taking into consideration microscopic interactions. Similarly, the equilibrium between two perfect gazes (macroscopic level) is the result of molecular interactions (microscopic level). Since self-organization limits interactions, so it limits also the entropy. Then, high entropy values indicate an important disorder, whereas low ones signify a better organization. This analysis is inspired from thermodynamics: when the entropy is low, the equilibrium is more stable and the disorder is less significant at molecular level.

In [100], Sneppen et al. use measures applied on network topologies. These measures characterize the ability of a node to lead and send a signal to its destination(s). The entropy is considered and is described as the capacity to predict from which neighbor the message arrives. It quantifies the predictability (or the order/disorder) of a traffic around a node. The probabilities represent the fraction of messages received from each neighbor. In [99], they present measures the investigation on constraints posed by the network structure on communication. They define two measures of entropy: the predictability relative to messages targeting a specified node (Target Entropy) and the predictability relative to messages crossing a specified node and one of its neighbors. The analysis of these measures lead to some conclusions:

- When the entropy values are high, the predictability is low. In the opposite case, a little number of links is used.
- The traffic to nodes with high degrees is unpredictable.
- Low values of entropy show that the traffic is concentrated. It is more distributed and logically more robust when the entropy is higher.

Chapter 4. Improving Link Prediction in Mobile Wireless Networks by Considering Link and Proximity Stabilities

In [107], Van Dyke Parunak and Brueckner try to show that the relationship between self-organization in multi-agent systems and the second law of thermodynamics is not a metaphor and that this relation can provide analytic and quantitative directives in the aim of the conception and the deployment of these systems. The self-organizing model is inspired from [68], which suggests that the key idea permitting to reduce the disorder in a multi-agents system is to copulate it with an another where the disorder increase. For a system, the self-organization is done at macroscopic level. Considering this, such behavior contradicts the second law of thermodynamics. Nevertheless, the system includes a microscopic level which dynamic increase the disorder. To reproduce a system with the two levels, the example of the pheromones used by ants is considered. The movements of ants constitute the macroscopic level while the molecules of pheromones represent the microscopic level. The movements of ants permit to define a little number of ways between the nest and the source of food. The disorder at macroscopic level is inconspicuous. The latter observation is the result of coupling macroscopic level agents with microscopic level, where the evaporation of molecules of pheromones takes place according to random mobility which increases the disorder.

The entropy, inspired from Shannon entropy and thermodynamic entropy, defines a disorder measure describing the trend of a system to be chaotic. This measure is applied at two levels: localization (microscopic level) and direction (macroscopic level). The results show that the two measures are antagonist which confirms that the second law of thermodynamics can be applied to multi-agent systems.

Another approach [113] tries to provide a schema able to classify a connection by three categories: Ethernet, WLAN or connection with low bandwidth. An algorithm is proposed to identify the connection type using the sending of packets pairs. This choice is led by the motivation to follow the random aspect at the reception of the pairs and use this aspect to identify the type of connection. The measure of random aspect is done by Shannon entropy. The major reason of this recourse instead the use of variance is that entropy is a better metric catching the random aspect of a random variable. As final note, it is interesting to mention that some works have used entropy to determine path stability in MANET and wireless sensor networks for their respective routing protocols. EQMGA (Entropy-based model to support QoS Multicast routing Genetic Algorithm) [30], ERPM (Entropy-based Routing Protocol using Mobility) [12], the An and Papavassiliou model [13], ELMR (Entropy-based Long-life Multipath Routing algorithm) [50] and QARPE (QoS-Aware Routing Protocol based on Entropy) [73] share the same idea of constructing a new entropy and select the most stable path relying on the entropy to reduce the number of route reconstruction when the topology is continuously changing.

4.4. How To Quantify Link and Proximity Stabilities and How To Use Them?

In the context of dynamical social networks, Zhao et al. [124] have been interested in human-based networks. These social networks encode information and are highly adaptive. From this perspective, Zhao et al. advance the entropy of dynamical social networks to describe the information about the dynamics of social interactions. Through a large dataset of interactions of cellular network users, it has been shown that the entropy that characterizes the dynamical social network is closely dependent on the time of a typical week-day (circadian rhythms) and that the social networks are extremely adaptive and whose structures are sensitive to the used technologies. In the same context, Bianconi et al. [19] have highlighted, by using entropy measures, how much important the annotated features of the nodes (age, gender, ...) or of the links (social ties, common attributes, ...) impact the structure of the network.

4.4.3 Quantifying Link and Proximity Stabilities by Means of Time Series Entropy Estimation

In order to measure these stabilities, we opt for the entropy metric. For our approach, we have been interested in the entropy estimator used for the Lempel-Ziv data compression [127, 67]. This tool enables us to estimate the entropy of a time series [80]. It has been applied by Song et al. [101] to demonstrate that the human mobility is highly predictable. For this purpose, they have analyzed traces of several cellular phone network users and they have constructed for each user a sequence of letters and each letter corresponds to the label that designates a specific location (a cell). After that, they applied the entropy estimator on the sequences of locations visited by several cellular network clients and they have obtained low values for the entropies which indicates a high predictability. Then, for a record of n steps, the entropy is estimated by:

$$S^{est} = \left(\frac{1}{n} \sum_i \Lambda_i \right)^{-1} \ln n \quad (4.4)$$

where Λ_i is the length of the shortest substring starting at position i which does not previously appear between position 1 and $i - 1$.

Before exploiting this estimator in link prediction, we have to define the time sequences in order to measure the stability of each link. Instead of Song's et al. approach which consists in filling in the user's location labels for each step, we rely on statistics of the link state between each pair of nodes. To construct this sequence, each node has to record the state of a link with every detected neighbor and at each period. Tracking the status of each link through a third order tensor accurately

achieves this requirement and for each link we obtain a sequence of zeros and/or ones. Therefore, we propose the entropy estimator $E_T^l(i, j)$ which quantifies the stability of the link between the nodes i and j over T periods. This is given by:

$$E_T^l(i, j) = \left(\frac{1}{n} \sum_{t=1}^T \Lambda_t(\mathbf{Z}_t(i, j)) \right)^{-1} \ln n \quad (4.5)$$

where $\Lambda_t(\mathbf{Z}_t(i, j))$ is the length of the shortest substring (consisting of a sequence of zeros and/or ones) starting at position t which does not previously appear between position 1 and $t - 1$. The parameter n corresponds to the number of substrings which are identified.

In the same way, we define $E_T^p(i, j)$ as estimator of the entropy which quantifies the stability of the proximity at two hops of the pair of nodes (i, j) . It is computed as in Eq. (4.5) but by substituting, for the value $\mathbf{Z}_t(i, j)$, the state of proximity at two hops at period t (1 if i and j are separated by 2 hops at the period t and 0 otherwise).

The Lempel-Ziv entropy estimator identifies at each step the shortest sequence which is not detected before. Therefore, we are tracking the length of the substrings step by step. If the shortest substrings become quickly too long, this means that there is redundancy. Indeed, the first chains added in the set of shortest substrings are repetitive in the whole sequence and it is necessary to concatenate new strings to make new shortest chains. Then, redundancy matches with stability as long chains decrease the value of the estimator. Otherwise, if the shortest substrings too often take the smallest possible length, this means that there vastly different combination of zeros and ones. This remark suggests that the ties highlight randomness and variation rather than regularity. So, the status of a tie is more unstable and the entropy estimation gets higher (due to short length of the substrings). In Fig. 4.2, we depict an example of the application of the algorithm used to identify the shortest substrings in the Lempel-Ziv data compression mechanism.

In the following subsection, we present new prediction scores. They are determined by joining the entropy estimations to the tensor-based link prediction framework.

4.4.4 Joining the Entropy Estimations to the Tensor-Based Link Prediction Framework

We have shown that the measures that we advance can be used to quantify the stability of a link or the proximity at two hops. Nevertheless, these measures, as mentioned previously, are unable to determine if the stability is quantified for an

4.4. How To Quantify Link and Proximity Stabilities and How To Use Them?

Example of an evolution of the state of a link between two nodes for successive time periods:

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
0	1	0	0	1	1	1	0

Short Substrings Identification Algorithm
Initially: Set of shortest substrings: {}
Step 1: Pointing t_1 Current substring is 0, not met before. Set of shortest substrings: {0}
Step 2: Pointing t_2 Current substring is 1, not met before. Set of shortest substrings: {0,1}
Step 3: Pointing t_3 Current substring is 0, met before. Continue the concatenation Set of shortest substrings: {0,1}
Step 4: Pointing t_4 Current substring is 00, not met before. Set of shortest substrings: {0,1,00}
.
.
.
Final set of shortest substrings: {0,1,00,11,10}

Figure 4.2: Example of an application of the algorithm used to identify the shortest substrings in the Lempel-Ziv data compression mechanism

occurring tie (series of ones in the tensor) or for a tie that is occasionally or never created (series of zeros in the tensor). Then it is important to combine them to a metric that expresses the lifetime or the perseverance of a link. From this perspective, we propose the combination with the Katz measure and/or the weight provided by the collapsed weighted tensor. Therefore, after evaluating several ways to combine the entropy estimations with the tensor-based link prediction framework, we propose four different techniques to join them. Our aim is to demonstrate that mea-

asuring the stability and exploiting it are really beneficial to improve the prediction performance. In this work, we want to prove that considering the stability of ties between nodes improves the link prediction of our framework. We are not seeking the design that ensures the best prediction efficiency. This investigation will be the aim of a future work. In the following, we detail how we design the combinations in order to predict better a link occurrence between a node i and a node j through new metrics.

- **Combining the collapsed weighted tensor value $\mathbf{X}(i, j)$ with the link entropy estimation $E_T^l(i, j)$ (XE scores):** We suggest to join the entropy estimation $E_T^l(i, j)$ (as expressed in Eq. (4.5) and where (i, j) is a pair of nodes) to the weight collected by the matrix $\mathbf{X}(i, j)$ (Eq. (4.1)) as we are seeking the most stable occurring links. In other words, we want to identify the links that have high weight and in the same time low estimation of the entropy. We use normalized values for $E_T^l(i, j)$ and $X(i, j)$.
- **Combining the Katz measure $\mathbf{S}(i, j)$ with the link entropy estimation $E_T^l(i, j)$ (SE scores):** We proceed as for the previous suggestion but we join the entropy estimation with the behavior similarity metric $\mathbf{S}(i, j)$ (Eq. (4.2)). We aim to check and assess that joining sociometric and stability measures can be helpful to make more precise predictions. As in the previous combination, we use normalized values for both parameters.
- **Computing the Katz measure $\mathbf{S}(i, j)$ from the combination of the collapsed weighted tensor value $\mathbf{X}(i, j)$ with the link entropy estimation $E_T^l(i, j)$ (XES scores):** We propose to apply the tensor-based link prediction technique but combining the collapsed weighted tensor \mathbf{X} with the link entropy estimation. Indeed, we combine each weight $\mathbf{X}(i, j)$ (Eq. (4.1)) with the measure $E_T^l(i, j)$ (Eq. (4.5)) and we apply the Katz formulation on the described combination in order to obtain a new matrix of scores. Also, we use normalized values for the entropy estimation and the collapsed weighted tensor values.
- **Computing the Katz measure $\mathbf{S}(i, j)$ with the definition of the new collapsed weighted tensor value $\mathbf{X}_{\text{new}}(i, j)$ (XNS scores):** We also propose to apply the tensor-based link prediction technique but with a new collapsed weighted tensor denoted \mathbf{X}_{new} . To determine it, we compute a coefficient, at each period p and for each pair of nodes (i, j) , that combines the occurrence weight obtained from \mathbf{X} (i. e. lifetime and recentness) with the link stability and/or the proximity stability. Afterwards, we apply the Katz

4.4. How To Quantify Link and Proximity Stabilities and How To Use Them?

formulation on the new collapsed weighted tensor \mathbf{X}_{new} in order to obtain a new matrix of scores. The matrix \mathbf{X}_{new} can be expressed in different ways which are detailed in the following.

Among the scores involving the entropy estimation that we have tested, we hold from each category the metric/metrics which highlights/highlight regularity and efficiency for all simulation scenarios. Then, for T tracking periods, we define the following matrices of scores:

$$XE_Score = ([1]_{N \times N} - \mathbf{X}) \bullet \times E_T^l \quad (4.6)$$

$$SE_Score = ([1]_{N \times N} - \mathbf{S}) \bullet \times E_T^l \quad (4.7)$$

$$XES_Score = (\mathbf{I} - \beta \cdot [E_T^l \bullet \times ([1]_{N \times N} - \mathbf{X})])^{-1} - \mathbf{I} \quad (4.8)$$

$$XNS_Score = (\mathbf{I} - \beta \cdot \mathbf{X}_{\text{new}})^{-1} - \mathbf{I} \quad (4.9)$$

where N is the number of nodes involved in the statistics and, for two nodes i and j , $\mathbf{X}_{\text{new}}(i, j)$ is given by:

$$\mathbf{X}_{\text{new}}(i, j) = \sum_{t=1}^T (1 - \theta)^{2(T-t)} \cdot (\mathbf{Z}_t(i, j) \cdot [\max_t(E_t) - E_t(i, j)]) \quad (4.10)$$

The parameter $E_t(i, j)$ and $\max_t(E_t)$ respectively correspond to the current entropy estimation (whether for link stability or proximity stability) and the maximum entropy value that we can obtain for t periods.

We derive three variants of the XNS_Score . When $E_t^l(i, j)$ is used to compute $\mathbf{X}_{\text{new}}(i, j)$, we define the measure $XNS1_Score$. If $E_t^p(i, j)$ is chosen, we express the metric $XNS2_Score$. In addition, we consider the case in which we compute $\mathbf{X}_{\text{new}}(i, j)$ using the link and the proximity stabilities. In this case, we propose the measure $XNS3_Score$ where $\mathbf{X}_{\text{new}}(i, j)$ is given by:

$$\mathbf{X}_{\text{new}}(i, j) = \sum_{t=1}^T (1 - \theta)^{3(T-t)} \cdot (\mathbf{Z}_t(i, j) \cdot [\max_t(E_t) - E_t^l(i, j)] \cdot [\max_t(E_t) - E_t^p(i, j)]) \quad (4.11)$$

In the following section, we detail the different scenarios to evaluate all these metrics and we assess their ability to achieve better performance in link prediction.

4.5 Simulations Scenarios and Performance Evaluation

To evaluate the efficiency of the tensor-based link prediction method joined to the entropy estimations, we consider two real traces. We present them in the following subsection. Then, we analyze the results obtained and we assess the efficiency of our contribution.

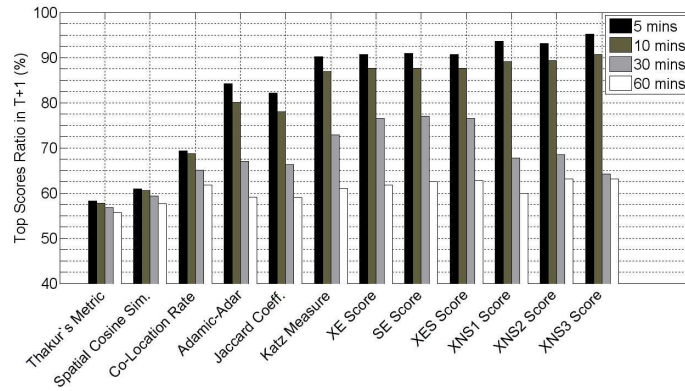
4.5.1 Simulation Traces

We consider three real traces to evaluate the contribution of the entropy estimations. We exploit them to construct the tensor by generating adjacency matrices with different time periods lengths. We detail, in the following, the traces used for the evaluation:

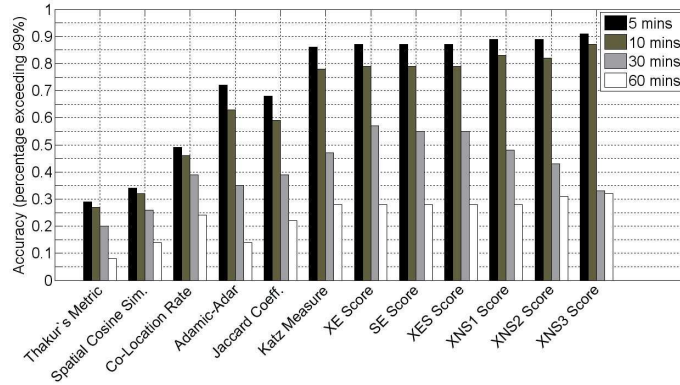
- **First Trace: MIT Campus trace:** We take the trace of 07/23/02 [16] and consider the events between 8 a.m. and 4 p.m. (8 tracking hours) to build up the tensors. The number of nodes is 646 and the number of locations (i.e. access points) is 174.
- **Second Trace: Dartmouth Campus trace:** We choose the trace of 01/05/06 [5] and construct the tensor slices relying on SYSLOG traces between 8 a.m. and 4 p.m. also. The number of nodes is 1018 and the number of locations (i.e. access points) is 128.
- **Third Trace: Infocom 2006 trace:** We consider the encounters which happened on 04/25/2006 [95] between 8 a.m. and 4 p.m among the 98 nodes involved in the trace (70 of them were carried by humans).

For each scenario (a fixed tracking period length), we track the occurrences of contacts during T periods. We also consider the adjacency matrix corresponding to the period $T+1$ as a benchmark to evaluate the effectiveness of our proposal. We construct tensors for the following period lengths: 5, 10, 30 and 60 minutes. That is to say that we record the network statistics for respectively a number of periods T equal to 96, 48, 16 and 8 slices (for the case where $t=5$ minutes, it is necessary to have 96 periods to cover 8 hours, 48 periods are needed to do the same when $t=10$ minutes, ...). We consider the distributed case for the computation of scores. We assume that each node has the knowledge about its 1-hop and 2-hop neighbors to

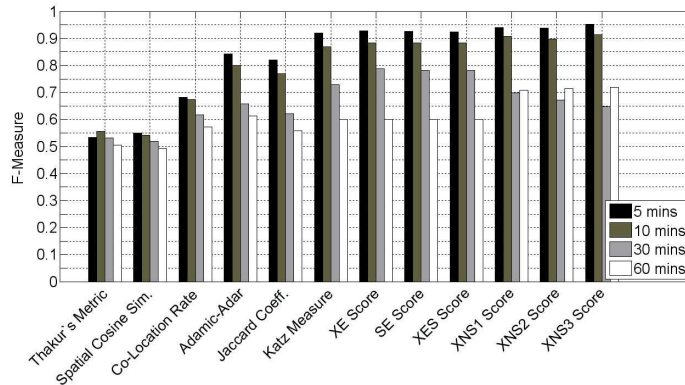
4.5. Simulations Scenarios and Performance Evaluation



(a) Top Scores Ratio in $T+1$



(b) Accuracy (percentage exceeding 99%)

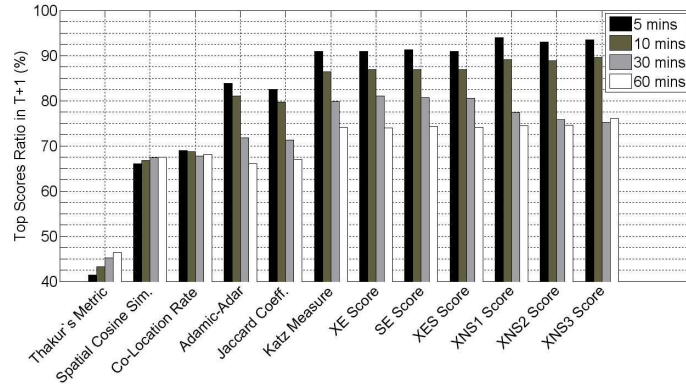


(c) F-Measure

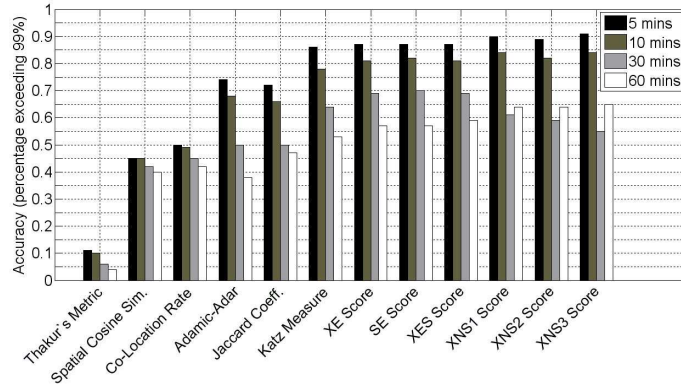
Figure 4.3: Evaluation metrics for the prediction applied on the MIT Campus trace for different tracking periods

compute the Katz measure for the tensor-based link prediction technique and the proposed measures of stability.

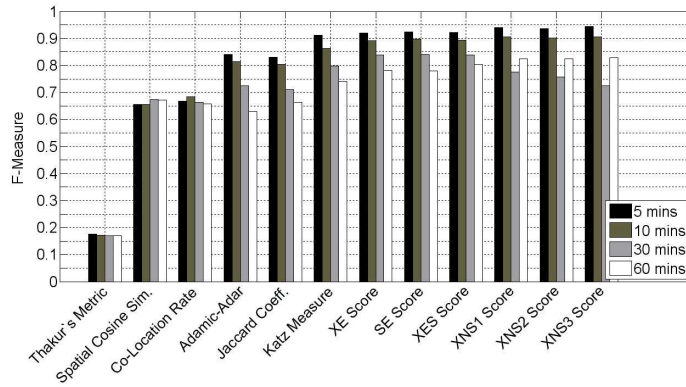
Chapter 4. Improving Link Prediction in Mobile Wireless Networks by Considering Link and Proximity Stabilities



(a) Top Scores Ratio in $T+1$



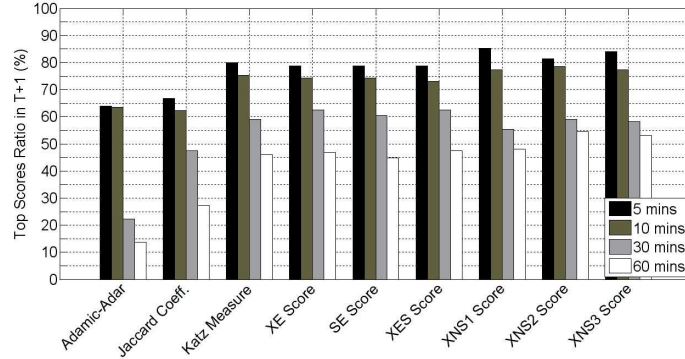
(b) Accuracy (percentage exceeding 99%)



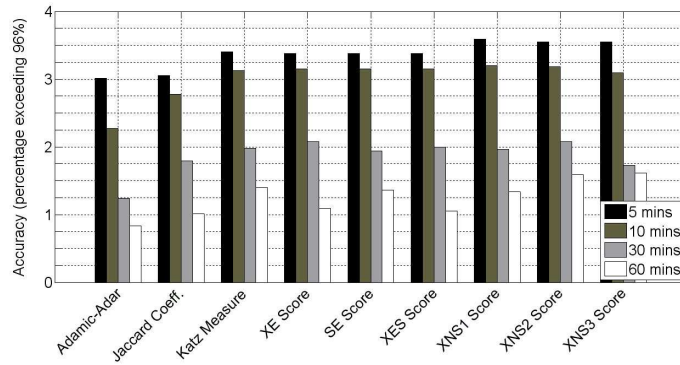
(c) F-Measure

Figure 4.4: Evaluation metrics for the prediction applied on the Dartmouth Campus trace for different tracking periods

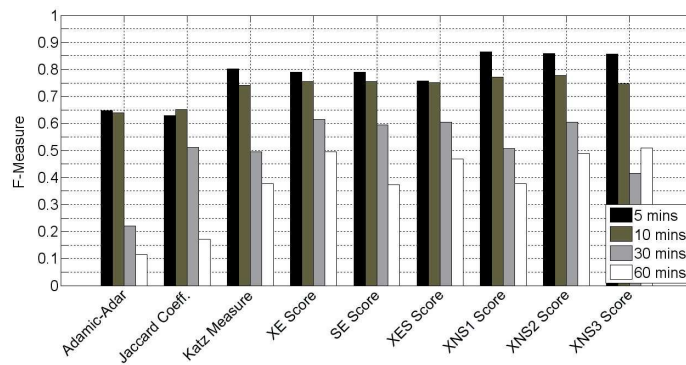
4.5. Simulations Scenarios and Performance Evaluation



(a) Top Scores Ratio in $T+1$



(b) Accuracy (percentage exceeding 96%)



(c) F-Measure

Figure 4.5: Evaluation metrics for the prediction applied on the Infocom 2006 trace for different tracking periods

4.5.2 Simulation Results and Performance Analysis

To assess if the proposed metrics enhance the prediction performance of the tensor-based link prediction framework, we consider three evaluation measures detailed in

the previous chapter:

- Top Scores Ratio at the period $T+1$ (TSR)
- Accuracy (ACC)
- F-Measure or balanced F1 score

Link Prediction of Links Occurring in $T+1$

We compare the prediction efficiency of the proposed prediction metrics with the one of the tensor-based link prediction technique [117] based on the Katz measure. To propose a comprehensive comparison, we also consider well-known prediction metrics presented in the literature. On the one hand, we consider behavioral-based link prediction metrics such as the similarity metric of Thakur et al. [105] and two metrics expressing mobile homophily proposed by Wang et al. in [110]: the spatial cosine similarity and the co-location rate. On the other hand, we take two link prediction metrics based on measuring the degree of proximity such as the Katz measure. They are the Adamic-Adar measure [10] and the Jaccard's coefficient [61, 94]. All the metrics that we propose are involved in the latter category.

The evaluation metrics are computed for all traces with different tracking periods lengths. Regarding the MIT Campus trace, the results are reported in Fig. 4.3. For the Dartmouth Campus trace, the prediction results are listed in Fig. 4.4. Also, a performance comparison is depicted in Fig. 4.5 between proximity-based link prediction metrics applied on the Infocom 2006 trace. For each figure, the plots (a), (b) and (c) respectively represent the top scores ratio in $T+1$, the accuracy (the percentage exceeding 99% for the two first traces and the percentage exceeding 96% for the last trace) and the F-Measure obtained for each prediction technique with different tracking periods.

The results obtained enable us to attest that the use of the Katz measure has been one of the best choices to perform prediction through the tensor-based technique. Using this metric achieves better performance than those of the other link prediction metrics proposed in the literature. Hence, the Katz measure is the best metric that we can use to interpret the meaning of the entropy estimations. The other well-known prediction metrics show a lower performance in the context of predicting such links. Indeed, instead of the Katz measure, they quantify the relationship of a pair of nodes without seeing if a direct link connects them. For example, two isolated neighbors (or having few common neighbors) would have a weak score even if they were connected. On the other hand, two other nodes can be

4.5. Simulations Scenarios and Performance Evaluation

2-hop neighbors and share several common neighbors which means that the score in this case is relatively high even if no link connects the corresponding pair of nodes.

Comparing the performance of the tensor-based link prediction technique with those of the proposed metrics based on link and/or proximity entropy estimations leads us to assess that our proposal of combining our prediction technique with stability measures is coherent and effective. We remark that for each scenario, there is at least three among the six proposed metrics that achieve better prediction performance (better top scores ratio at $T + 1$, accuracy and F-Measure) than the Katz measure. The Katz measure quantifies the social ties between two nodes by a score. Nevertheless, such a score sometimes cannot indicate if the contacts between these two nodes are stable or interspersed over time. The contribution of the entropy estimation is to identify the stable and persistent links but not only them. Also, the entropy allocates low values when a link never or rarely occurs between two nodes (long sequences of 0 for the status of a link). Hence, combining the Katz measure and the entropy estimation enables us to distinguish more clearly a stable link (low entropy estimation and high Katz measure). Such a combination highlights more clearly how nodes are tied which makes the prediction of the topology more efficient in the period $T+1$.

Considering all traces, we find that it is better to opt for the SE , the XE and the XES scores when the tracking periods are long (30 and 60 minutes). When these periods are shorter (10 and 5 minutes), the XNS scores are the most suitable metrics. In fact, shortening the length of the tensor period leads to obtaining a more precise tracking of the properties of the links (i.e. better distinction between persistent and fleeting links), more faithful estimation of the stability and then more efficient link prediction (more information about the impact of the tracking period length is provided in [117]). Meanwhile, when we use long tensor periods, tracking become less precise as the method considers long and short contact as the same. For example, if we consider a tracking period of 30 minutes, a contact that occurs during all this period is considered as the same as a contact that only lasts a few seconds during the same period (the status of the link is set to 1). It is clear that we lose precision when the contacts tend to be short. Moreover, we divide the historical data into T periods and when the tracking period are longer, the number of total periods T becomes less. Therefore, the sequence of 0 and 1 that characterizes a link over time becomes shorter and tends to be unstable (due to the lack of precision). Afterwards, the Katz measure, as well as the entropy estimation, is less efficient to characterize the relationship between nodes, especially for short contacts.

In these simulations, we have assumed that each node has knowledge about its one-hop and two-hop neighbors. Limiting the local information to one hop is not

compromising the performance either of the tensor-based link prediction technique or of the proposed metrics based on entropy estimation. It is true that we are not able, in this case, to compute the $XNS2$ and the $XNS3$ scores. Nevertheless, the Katz measure is no longer affected with this limitation. In fact, we have proved in [117] that the prediction efficiency of the tensor-based link prediction technique achieves similar performance in predicting links in the period $T+1$ whether the knowledge is limited to one-hop neighbors or extended to the two-hop neighbors. Fig. 4.6 depicts the Top Scores Ratio in $T+1$, the accuracy and the F-Measure obtained from the MIT Campus trace when prediction is performed by the Katz measure and by the proposed link stability based measures using only the knowledge of direct neighbors (XE , SE , XES and $XNS1$ scores). The results confirm the findings cited above: the performance of each prediction technique is similar in both scenarios. Obviously, the contribution of the entropy estimation is always effective. This effectiveness is maintained for link prediction in the period $T+1$.

Link Prediction of New Links

We have shown in the previous chapter that the Katz measure is able to predict some future new links (that have never occurred during the historical). We have tried to see if the prediction metrics that we propose are able to predict new links and we found that only the SE score is able to exactly achieve the same performance as the one of the Katz measure. This observation is trivial to explain : in fact, the entropy measures of all new links are the same as they are never occurring in the historical and then characterized by a sequence of zeros in the link status over time. The other proposed metrics depict limited or no ability to detect such links.

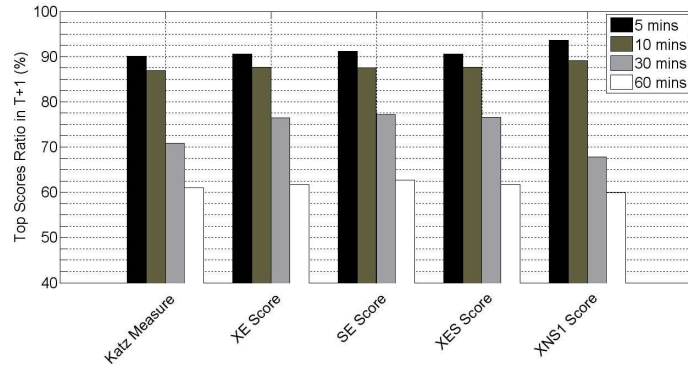
From this perspective, we wonder what would be the performance of the SE score if we substitute the link stability between two nodes by the one of proximity at 2 hops. Hence, we express the matrix of scores SE_{2hops} Score as follows:

$$SE_{2hops_Score} = ([1]_{N \times N} - \mathbf{S}) \bullet \times E_T^p \quad (4.12)$$

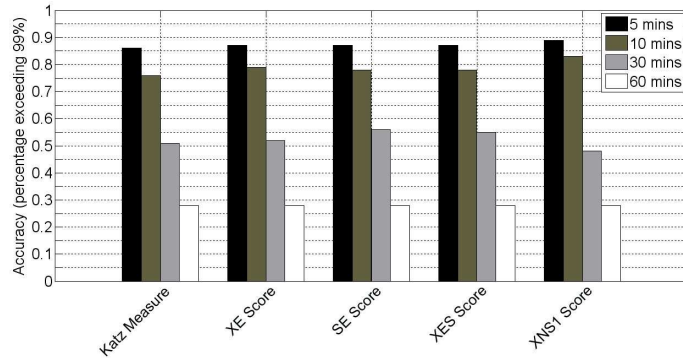
where N defines the number of nodes involved in the statistics, \mathbf{S} is the Katz scores matrix (normalized scores) and E_T^p represents the entropy estimations matrix of the proximity at two hops stability of each pair of nodes (i, j) .

To verify the effectiveness of this metric in improving the prediction of new links, we extract all the contacts happened during a tracking time and we consider the contacts that have occurred after this tracking time in order to assess the ability of detecting the birth of new links. The contacts are extracted as explained in the Table 4.1. In Table 4.2, we compare the performance of predicting new links of the

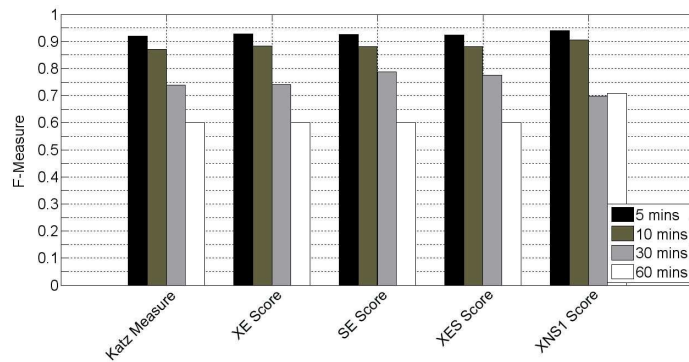
4.5. Simulations Scenarios and Performance Evaluation



(a) Top Scores Ratio in $T+1$



(b) Accuracy (percentage exceeding 99%)



(c) F-Measure

Figure 4.6: Evaluation metrics for the prediction applied on the MIT Campus trace for different tracking periods and with knowledge limited to direct neighbors

proposed metric with the one of the Katz measure (it is assumed that the nodes have the knowledge of their neighbors at one and two hops). For that aim, we

Chapter 4. Improving Link Prediction in Mobile Wireless Networks by Considering Link and Proximity Stabilities

Table 4.1: Settings for each trace to assess the efficiency of the SE_{2hops} score in improving the prediction of new links

Trace	Tracking time	Time in which contacts are considered as future
MIT Campus (07/23/02)	from 8 a.m. to 2 p.m.	2 p.m to 6 p.m.
Dartmouth Campus (01/05/06)	from 8 a.m. to 2 p.m.	2 p.m to 6 p.m.
Infocom 2006 (04/25/06)	from 8 a.m. to 1:30 p.m.	1:30 p.m to 6 p.m.

Table 4.2: Top Scores Ratio in the future: comparison between the efficiency of the SE_{2hops} score and the one of the Katz measure in predicting new links

Trace	Tracking Period t Length	Prediction Score	TSR in the future
MIT Campus	5 minutes	Katz Measure	13,49%
		SE_{2hops} Score	14,09%
	10 minutes	Katz Measure	13,22%
		SE_{2hops} Score	13,44%
Dartmouth Campus	5 minutes	Katz Measure	9,10%
		SE_{2hops} Score	8,36%
	10 minutes	Katz Measure	9,28%
		SE_{2hops} Score	8,55%
Infocom 2006	5 minutes	Katz Measure	16,96%
		SE_{2hops} Score	18,75%
	10 minutes	Katz Measure	9,82%
		SE_{2hops} Score	9,82%

perform prediction on the three traces and with considering two different tracking periods t (5 and 10 minutes periods). The results obtained for the Katz measure match with the ones analyzed in the previous chapter: the links created between two nodes which have been separated by 2 hops are identified which represent a respectable percentage of all new links occurred. Then, we have thought to exploit the proximity stability and to assess if it contributes in enhancing the prediction of new links. On top of the previous analysis, we have also been motivated by the findings in [110] which highlights that a majority of new links is occurring between two nodes that have been within two hops of each other in a GSM network. Hence, we express the SE_{2hops} score with an estimation of the stability of proximity at 2 hops as we expect that such a stable proximity increases the possibility of a new link occurrence. Considering the results obtained, we clearly remark that using the entropy estimation at two hops improves the prediction of new links when it is performed for the MIT Campus and Infocom 2006 traces. The performance of our proposed metric is slightly worse when the prediction is applied on the Dartmouth Campus trace. Nevertheless, it remains that using the proximity stability looks promising to better predict new links and is worth investigate for future works.

4.6 Conclusion

In this chapter, we highlighted that the prediction efficiency of the tensor-based technique that we have advanced can be improved by taking into consideration other aspects on top of measuring social closeness. We proposed to make the prediction sensitive to the link and proximity stabilities. We showed that a strong relationship between two nodes matches also with a stable link between them. Indeed, when similar intentions are shared for a period of time, the link between the corresponding individuals is expected to be stable. We also outlined that considering proximity stability can also be beneficial to improving the prediction performance. To express the stability itself, we proposed an entropy estimator inspired from data compression and which converges to the expected value of entropy for a time series. In our case, the time series is the sequence of the state of the links in the adjacency matrices (sequence of zeros and/or ones).

To assess the efficiency of our contribution, we tried to join the stability feedback to the tensor-based link prediction framework through proposing new prediction metrics. We assessed that we can improve the performance of the prediction technique especially when the tensor time period tend to be short. In other words, using shorter tensor periods favors more precise tracking of contacts between nodes, which leads to a better estimation of the entropy estimations and then more faithful link prediction. Above all, we identified, according to the length of tracking period, the set of metrics that can be used to enhance, as much as possible, the performance of the tensor-based link prediction technique.

We presented in this first part a tensor-based link prediction framework for human-centered mobile wireless networks and some proposed improvements. These contributions have been advanced as we aspire to propose mechanisms and techniques that can be useful in order to evaluate or improve/design communication protocols dedicated to such networks. In this thesis, on top of proposing link prediction techniques, we have provided two other contributions that match with our aim and that we will present in Part II. Indeed, we have designed a joint model for the IEEE 802.15.4 physical and medium access control layers in order to evaluate more accurately the relative performance when N nodes try to communicate. This joint model is presented in chapter 5. As a second contribution, we design a self-learning repeated game framework inspired by the concept of "The Weakest Link" TV game and dedicated to promote the cooperation among nodes in non-cooperative MANETs. This framework is detailed in chapter 6.

Part II

Other Contributions Proposed in the Thesis

Chapter 5

A Joint Model for IEEE 802.15.4 Physical and Medium Access Control Layers

Contents

5.1 Introduction	86
5.2 Related Work	88
5.3 Developed IEEE 802.15.4 Model for Smart Grid	89
5.3.1 IEEE 802.15.4 PHY Model Description	89
5.3.2 Operation Details for the IEEE 802.15.4 MAC Model and the Interactions with the PHY Model	90
5.4 Simulation and Results	95
5.4.1 Cheking the Validity of the Joint Model	96
5.4.2 Comparison between Combined PHY and MAC Layers and Simple MAC Layer Models	96
5.4.3 Evolution of Node Performance with Growing Densities	98
5.5 Conclusion	99

5.1 Introduction

Wireless sensor networks have been closely studied in recent years. Several studies have investigated behaviors and performances of these networks. Some of them have highlighted such networks properties by relying on empiric results [27, 123, 114, 49,

28] whereas others have focused on reproducing a standard or mechanism functionalities tied to sensors by proposing analytical models [18, 120, 36, 87]. Empirical studies have shown that wireless communication networks are radically different from some simulation models (disc-shaped nodes range for example). Analytical studies have attempted to reproduce mechanisms and technical aspects widely used/seen in these networks in order to track network performances. Among these approaches, those of Zuniga and Krishnamachari [128, 116] stand out. They emphasize the limits of disc-shaped node range models that are used in simulators, and highlight the existence of a transitional region between the connected and disconnected areas. This observation, based on experiments, enables us to understand clearly the reason behind link unreliability in low power wireless networks. Moreover, Zuniga and Krishnamachari underline the impact of asymmetry in transitional region expansion and its negative effect on reliability [116]. Meanwhile, lot of the performance analysis of MAC layer protocol are derived from the Markov model proposed by Bianchi [18] for the IEEE 802.11 standard [1]. This model consists in a Markov chain that reproduces the functionalities of the IEEE 802.11 standard while assuming saturated traffic and ideal channel conditions. This approach has inspired many others, for instance the Park et al. model [87]. It presents itself as a relevant contribution which aims to measure reliability, delay and energy consumption in a wireless network based on IEEE 802.15.4 standard [2]. In this chapter, we develop an IEEE 802.15.4 model that takes into consideration the interactions on PHY and MAC layers. The model, at the PHY layer level, is derived from the Zuniga and Krishnamachari mathematical framework for quantifying link unreliability. The MAC layer model is inspired from an enhanced Park et al. Markov chain. The joint model that combines both PHY and MAC models enables us to consider the causes behind packet losses either at PHY or MAC levels. Indeed, in the Park et al. model, collisions appear to be the only reason for losses, whereas, our model includes constraints posed by SNR (signal-to-noise ratio), modulation and encoding.

The remainder of this chapter is as follows. In Section 5.2, we present the related work which gives an overview of approaches that inspire our model. We focus on our contribution by giving details on the combined PHY and MAC layers models in Section 5.3. In Section 5.4, we compare our proposition to the enhanced Park et al. approach and estimate nodes performances with different network densities. Finally, Section 5.5 concludes the chapter.

5.2 Related Work

Many studies have aimed to understand and to evaluate standards and protocols. The works that tried to identify the properties of these networks mechanisms fall into two categories: i.e. simulations-based [27, 123, 114, 49, 28] relying on empiric observations, or analytical works [18, 120, 36, 87]. Most of analytical studies are based on the Markov model proposed by Bianchi [18] for the IEEE 802.11 standard. This model consists in a Markov chain that mimics the functionalities of the IEEE 802.11 standard while assuming a saturated traffic and ideal channel conditions. Zhai et al. [119] and Daneshgaran et al. [36] exploit the Bianchi model and extend it through more realistic assumptions. These approaches have inspired Griffith and Souryal [7] to develop a model for the IEEE 802.11 MAC layer that adds a frame queue to each node. This contribution enables us to estimate the packet reception rate, the delay, the medium access control layer (MAC layer) service time and the throughput. Similar studies have been developed for the wireless sensor networks, and more especially the IEEE 802.15.4 standard. Hence, we note the models developed by Pollin et al. [89] and Park et al. [87]. The two approaches provide a generalized analysis that allows to measure reliability, delay and energy consumption. In each proposed model, the exponential backoff process is modeled by a Markov chain. Retry limits and acknowledgements in an unsaturated traffic scenario are also taken into consideration.

Park et al. propose a generalized analytical model of the slotted CSMA/CA mechanism with beacon enabled mode in IEEE 802.15.4. This model includes retry limits for each packet transmission. The scenario of a star network in which N nodes try to send data to a sink has been considered and defining the state of a single node through a Markov model has been proposed. Each state of the Markov chain is characterized by three stochastic processes: the backoff stage $s(t)$, the state of the backoff counter $c(t)$ and the state of the retransmission counter $r(t)$ at time t . The described modeling allows us to analyze of the link reliability, delay and energy consumption.

In another context, several works focus on the physical layer modeling. For instance, Zuniga and Krishnamachari [128, 116] have analyzed the major causes behind unreliability [128, 116] and the negative impact of asymmetry on link efficiency [116] in low power wireless links. Instead of the binary disc-shaped model these models reproduce the called *transitional region* [123, 114, 49] in order to model the transmission range. The packet reception rate and the upper-layer protocol reliability are highly instable when a neighbor is located in this region. To understand it, two models have been proposed: a channel model that is based on

the log-normal path loss propagation model [91] and a radio reception model closely tied to the determination of packet reception ratio. Through these models, it is possible to derive the expected distribution and the variance of the packet reception ratio according to the distance.

5.3 Developed IEEE 802.15.4 Model for Smart Grid

Our contribution joins the initiative of Smart Grid [7] to provide tools that evaluate wireless communications standards. The developed model that we propose analyzes an IEEE 802.15.4 PHY and MAC layer channel in which multiple non-saturated stations compete in communicating with a sink. The aim is to combine two relevant models: A PHY model that bypasses the disk shaped node range and takes into consideration the called transitional region and a MAC model that reproduces the CSMA/CA mechanism. The model described enables us to add the impact of PHY layer errors onto the MAC layer and to provide some improvements for the adopted MAC model, in order to obtain more precise output estimations. Our developed model is available at the SGIP NIST Smart Grid Collaboration website [7].

5.3.1 IEEE 802.15.4 PHY Model Description

To model the PHY layer, we have adopted the Zuniga and Krishnamachari approach [128, 116]. The main objective is to identify the causes of the transitional region and quantify their influence on performance without considering interferences (assumption of a light traffic or static interference). To do this, the expressions of the packet reception rate as function of distance are derived. These expressions take into account radio and channel parameters such as the path loss exponent (log-normal shadowing path loss model [91]), the channel shadowing variance, the modulation, the coding and hardware heterogeneity. They describe how the channel and radio influence transitional region growth. We use mathematical frameworks provided to compute packet delivery rate independently of interferences.

The approach followed by Zuniga and Krishnamachari in which we have been interested in describes how the channel and the radio determine the transitional region. On the one hand, for the wireless channel, the log-normal shadowing path loss [91] model is adopted (can be used for small and large coverage systems and its accuracy is demonstrated in comparison with other models). It is given by:

$$PL(d) = PL(d_0) + 10\eta \log_{10} \frac{d/d_0}{+} N(0, \sigma) \quad (5.1)$$

Where d is the transmitter-receiver distance, d_0 is a reference distance, $PL(d_0)$

5.3. Developed IEEE 802.15.4 Model for Smart Grid

is the power decay for the reference distance d_0 , η is the path loss exponent and $N(0, \sigma)$ is a zeromean Gaussian random variable with standard deviation σ .

Hence, for an output power P_t , the received power P_r in dB is expressed by:

$$P_r(d) = P_t - PL(d_0) - 10\eta \log_{10} \frac{d/d_0}{+} N(0, \sigma) \quad (5.2)$$

On the other hand, for the radio, the packet reception rate Ψ of successfully receiving of a packet, using a modulation M is:

$$\Psi(\gamma) = (1 - \beta_M(\gamma))^f \quad (5.3)$$

Where γ is the SNR (Signal-to-Noise Ratio), β_M is the biterror rate and a function of the SNR and f is the frame size. The SNR at a distance d can be expressed from (5.2):

$$\gamma(d)_{dB} = P_r(d) - P_n = N(\mu(d), \sigma) \quad (5.4)$$

Where $N(\mu(d), \sigma)$ is a Gaussian random variable with mean $\mu(d)$, variance σ^2 and P_n is the noise floor. Moreover, the expression of $\mu(d)$ can be determined from (5.2) into (5.4):

$$\mu(d) = P_t - PL(d_0) - 10\eta \log_{10} \frac{d/d_0}{+} N(0, \sigma) - P_n \quad (5.5)$$

Denoting the bit-error rate for the SNR in dB as $BER_M(\gamma_{dB}) = \beta_M(10^{\frac{\gamma_{dB}}{10}})$, the packet reception rate Ψ can be expressed by:

$$\Psi(\gamma_{dB}) = (1 - BER_M(\gamma_{dB}))^f \quad (5.6)$$

5.3.2 Operation Details for the IEEE 802.15.4 MAC Model and the Interactions with the PHY Model

We believe that joining the PHY model described previously with a MAC model is an interesting challenge. Indeed, collisions are the major factor behind frame losses. Nonetheless, considering errors that can happen at the PHY layer includes constraints imposed by SNR (signal-to-noise ratio), modulation, encoding and asymmetry (heterogeneous hardware). Therefore, our contribution allows us to have a more realistic estimation by taking into account the causes of failures at both layers.

The model of IEEE 802.15.4 MAC layer that we have considered is inspired from Park et al. Markov chain [87]. It captures the state of the station backoff stage, the backoff counter and the retransmission counter. We insert into Park et al. model an

M/M/1/K queuing model that endows a finite buffer to a station. On the one hand, the Markov model determines the steady state probability when a station senses the channel in order to transmit a frame and the probability that a frame experiences a failure (due to a collision or to PHY layer failure). On the other hand, the queuing model gives as output some measurements such as the throughput or the probability that the station is idle.

The Park et al. approach, inspired from [89], consists in a generalized analytical model of the slotted CSMA/CA mechanism of beacon enabled IEEE 802.15.4 with retry limits for each packet transmission (the complete description is provided in [87]). The model takes the scenario of N stations that try to communicate with a sink. Park et al. define the probabilities for the following events: a node attempts a first carrier sensing to transmit a frame, a node finds the channel busy during CCA1 or a node finds the channel busy during CCA2. They are denoted by the variables τ , α and β respectively. These three probabilities are related by a system of three non-linear equations that arises from finding the steady state probabilities. Our model, described by the flowchart presented in Fig. 6.2 (the main PHY and MAC inputs are listed in Table I and Table II respectively), aims to solve the non-linear system that expresses these probabilities. In addition, it estimates p_0 , the probability of going back to the idle state by considering the offered load per node parameter λ . In this way, our contribution enables the MAC model to determine this probability, in opposition to [87] (p_0 is taken as an input for the performances analysis).

We start from equations (16), (17) and (18) in [87] and make changes in some of these expressions to enhance the model. The equations (17) and (18) are expressed with probability τ to mention that a node is transmitting. In our mind, this consideration is insufficient because a transmitting node must not be idle, that is why we substitute τ by $(1-p_0)\tau$. Thereby, τ is the probability that a node tries to transmit and $1-p_0$ is the probability that a station has a frame to send. The system considered is given by equations (5.7), (5.8) and (5.9):

5.3. Developed IEEE 802.15.4 Model for Smart Grid

$$\tau = \left(\frac{1-x^{m+1}}{1-x} \right) \left(\frac{1-y^{n+1}}{1-y} \right) b_{0,0,0} \quad (5.7)$$

$$\alpha = \left(L + \frac{N(1-p_0)\tau(1-\tau(1-p_0))^{N-1}}{1-(1-\tau(1-p_0))^N} L_{ACK} \right) (1 - (1 - (1 - p_0)\tau)^{N-1}) (1 - \alpha)(1 - \beta) \quad (5.8)$$

$$\beta = \frac{1 - (1 - \tau(1 - p_0))^{N-1}}{DV} + \frac{N(1 - p_0)\tau(1 - (1 - p_0)\tau)^{N-1}}{DV} \quad (5.9)$$

where $DV = 2 - (1 - (1 - p_0)\tau)^N + N(1 - p_0)\tau(1 - (1 - p_0)\tau)^{N-1}$, $x = \alpha + (1 - \alpha)\beta$ and $y = P_{fail}(1 - x^{m+1})$. The parameter P_{fail} represents the probability of a failed transmission attempt, m is the maximum number of backoffs the CSMA/CA algorithm will attempt before declaring a channel access failure, n is the maximum number of retries allowed after a transmission failure, L is the length of the data frame in slots (a slot has a length of 80 bits), L_{ACK} is the length of an acknowledgement in slots, N is the number of stations and $b_{0,0,0}$ is the state where the state variables of the backoff stage counter, the backoff counter and the retransmission counter are equal to 0 (an approximation is proposed in [87]).

The mechanism that computes these probabilities (using the MATLAB *fsolve* function) allows us to determine the probability of failed transmission P_{fail} , given by:

$$P_{fail} = 1 - (1 - P_{col})(1 - P_e) \quad (5.10)$$

where $P_{col} = 1 - (1 - \tau(1 - p_0))^{N-1}$.

In the above expressions, P_e is the probability of loss due to channel and radio constraints (computed by the PHY model) and P_{col} is the probability of a collision occurring (modified as done with (17) and (18) in [87]).

This mechanism is embedded in a loop that updates p_0 . The developed model solves the system of non-linear equations to determine τ , α , β and therefore P_{fail} . Then, P_{fail} , α and β are used to estimate the mean MAC service time, or the mean time to process a frame, expressed also as Expected Time or *ET* (in [87], Section V.B details how to compute this time. We substitute, of course, P_{col} by P_{fail} to catch errors that can occur at PHY and MAC layers). So, a new value for p_0 is generated and the updated p_0 is used in the next iteration. It is possible to determine p_0 since each device has a buffering capacity. Every node is modeled as an M/M/1/K queue

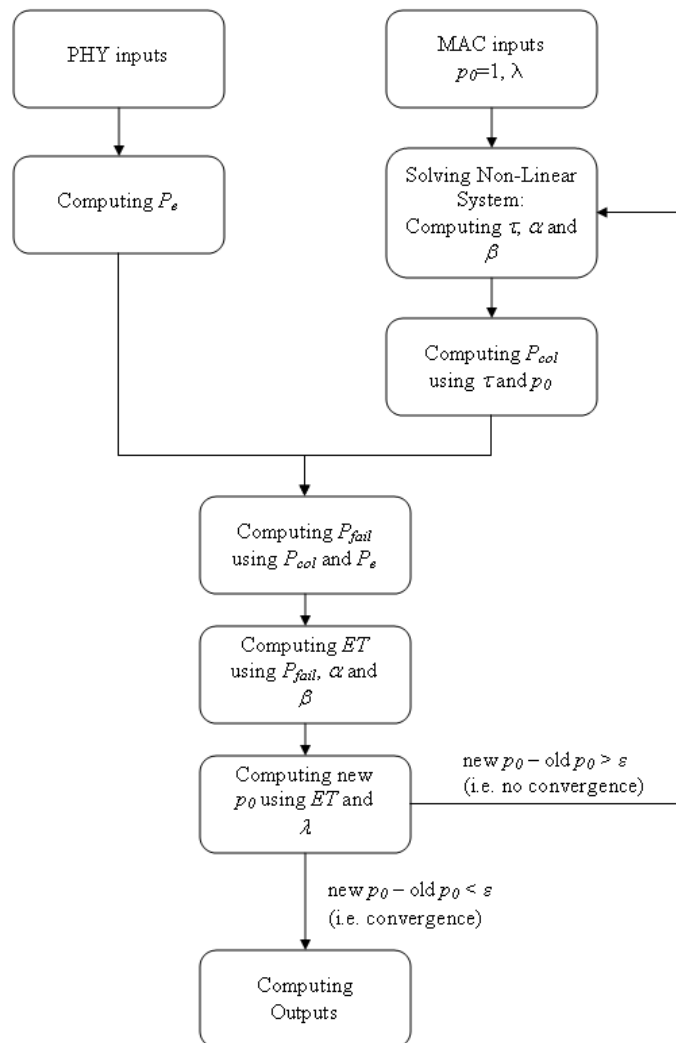


Figure 5.1: IEEE 802.15.4 PHY and MAC model flowchart

5.3. Developed IEEE 802.15.4 Model for Smart Grid

and each queue receives frames following a Poisson arrival process λ frames/s. The queue utilization ρ is the product of the arrival rate λ and the inverse of the mean MAC service time ET . The steady state probability that there are i frames in the queue is:

$$p_i = \rho^i / \sum_{j=0}^K \rho^j \quad (5.11)$$

Hence, the value of p_0 is given by:

$$p_0 = \left[\sum_{j=0}^K \rho^j \right]^{-1} \quad (5.12)$$

The process continues until the value of p_0 converges to a stable value. Once p_0 converges, all outputs concerning queuing analysis can be computed for each value of λ (the per-node load offered). Four outputs are considered in this study: the average waiting time to receive a frame (Eq. (7)), the failure probability (probability of packet loss due to collisions or link constraints)(Eq. (4)), the reliability of a node (the probability of a good frame reception)(Eq. (8)) and the average throughput per node(Eq. (9)).

$$D = \frac{L}{\lambda(1 - p_k)} \quad (5.13)$$

$$R = (1 - p_k)(1 - P_{cf})(1 - P_{cr}) \quad (5.14)$$

$$S_{avg} = \lambda R L_p \quad (5.15)$$

where p_k is the probability of having full buffer, P_{cf} is the probability that the frame is discarded due to channel access failure (Eq. (19) in [87]), P_{cr} is the probability that the packet is discarded due to retry limits (Eq. (20) in [87]), L is the payload size and L_p is the application data size.

Therefore, this contribution enables us to enhance Park and al. model at two levels:

- Providing a more precise computation of failure probability by considering possible errors at PHY and MAC layers (link unreliability and collisions).
- Enhancing the MAC model of Park et al. by estimating the probability p_0 for the resolution of non-linear equations (this probability is an input in Park et al. model), modifying some expressions to more efficient estimations and determining outputs relative to a precise scenario (number of nodes and per-

Chapter 5. A Joint Model for IEEE 802.15.4 Physical and Medium Access Control Layers

Table 5.1: Main Parameters used in MAC Layer

Parameter	Value	Parameter	Value
Number of Nodes	5, 10, 50	Queue Size	51 Frames
Smallest Backoff Win	8	Data Rate	250 kbit/s
Max Frame Retries	3	ACK Size	88 bits
Max CSMA Backoff	4	Shadowing STD	4
Max Backoff Exponent	5	IFS	640 μ s
Min Backoff Exponent	3	Max TX-RX Time	192 μ s
MAC Frame Payload	800 bits	PHY Overhead	48 bits

Table 5.2: Main Parameters used in PHY Layer

Parameter	Value	Parameter	Value
Noise Figure	23dB	Bandwidth	30kHz
Pathloss exp	4	STD Tx power	5dBm
Noise	15dB	Preamb. Length	40 bits
Max Tx range	15 m	Min Tx Range	1 m

node load offered).

5.4 Simulation and Results

First of all, we propose to check the validity of our model. Then, we advance two scenarios for the simulations in order to appreciate our joint model. After that, we compare the performances of a node obtained in two different ways. On one hand, we use the Park et al. Markov chain (MAC layer) and on the other hand we test our model. Then, we expose the same performances, using our developed model, for different densities. All the simulations test different values for the offered per-station load, measured in units of frame/s. We choose to start from 0.5 frame/s and increase the offered load to 50 frames/s with a step of 0.5 frame/s (or from 400 bits/s to 40000 bits/s). We select four outputs to illustrate node performances: the average waiting time for a frame reception, the failure probability (probability of frame loss due to collisions or link constraints), the reliability of a node (the probability of a good frame reception) and the throughput. Table 6.1 presents the main inputs at the MAC layer and Table 6.2 enumerates the main ones at the PHY layer.

5.4.1 Cheking the Validity of the Joint Model

As mentioned, The MAC layer model proposed by Park et al. [87] is strongly inspired by the Markov model provided by Pollin et al. [89]. In [89], the authors have proposed to measure the probabilities α and β , through their model, according to a growing number of nodes in a network with saturated traffic. Then, their have compared the obtained values with those provided by simulation. The results have shown that the probabilities obtained in both cases are similar and this observation enable them to validate their model. To do the same for our model, we consider the same simulation scenario and we compare the results obtained by our joint model with the ones obtained by the model of Pollin et al. Table 6.3 details the values obtained for the probabilities α and β with both models according to the growing star network density.

We remark that our model show very similar values of α and β in the case of saturated traffic. This indicates that the two model mimic the behavior of the 802.15.4 standard in the same way. This result enables us to conclude that our model is effective and valid even if the impact of the PHY layer model is not visible due to the saturation traffic condition. To appreciate the contribution of this model, we propose in the following to compare our joint model to the model of Park et al.

5.4.2 Comparison between Combined PHY and MAC Layers and Simple MAC Layer Models

As previously described, when we include the constraints at the physical layer, delivery failures happen more often. There are many reasons for this: weak SNR and modulation and/or encoding errors. We run simulation for a star network with 10 nodes. The results confirm a notable degradation of node performances. In Fig. 5.2(a), the average waiting time is for the the combination of the PHY and MAC models. Inserting link constraints increases the number of retransmissions. Thus, the delay increases. The delay difference between the two approaches is not obvious through this figure. So, we propose to zoom in the figure to show the difference between the two models. We highlight this difference through Fig. 5.3. Fig. 5.2(b) compares the evolution of failure probability for the two approaches. With light offered loads, the impact of the PHY model is conspicuous, especially since the number of collisions is likely to be low. The collisions are more frequent with heavier loads and the probability of occurrences grows quickly, generating network saturation. Meanwhile, the probability of losses due to link conditions remains constant (this probability is determined independently of interferences and computed through an integration over the distance covered by the maximum range and over

Chapter 5. A Joint Model for IEEE 802.15.4 Physical and Medium Access Control Layers

Table 5.3: Comparison of the values for α and β obtained through the joint model and the one of Pollin et al. according to a growing density with saturated traffic

Number of nodes	Joint Model		Model of Pollin et al.	
	α	β	α	β
10	0.8620	0.5054	0.8043	0.4539
12	0.8631	0.5024	0.8173	0.4794
14	0.8636	0.5010	0.8282	0.4941
16	0.8638	0.5004	0.8347	0.5058
18	0.8639	0.5000	0.8413	0.5130
20	0.8639	0.5000	0.8467	0.5166
22	0.8639	0.5000	0.8500	0.5228
24	0.8639	0.5000	0.8543	0.5245
26	0.8639	0.5000	0.8576	0.5294
28	0.8639	0.5000	0.8597	0.5264
30	0.8639	0.5000	0.8630	0.5274
32	0.8639	0.5000	0.8652	0.5274
34	0.8639	0.5000	0.8663	0.5254
36	0.8639	0.5000	0.8684	0.5245
38	0.8639	0.5000	0.8695	0.5225
40	0.8639	0.5000	0.8706	0.5215
42	0.8639	0.5000	0.8728	0.5215
44	0.8639	0.5000	0.8728	0.5196
46	0.8639	0.5000	0.8739	0.5166
48	0.8639	0.5000	0.8750	0.5137
50	0.8639	0.5000	0.8760	0.5127

asymmetry variations). So, the difference between the two approaches is less significant. The same interpretation can be used for reliability evolution, presented in Fig. 5.2(c). Reliability also undergoes the frame discards due to the reaching of maximum frame retries or maximum CSMA backoffs. The rejected frames due to full node queue represent also a possible interpretation with high offered loads. The throughput evolution, presented in Fig. 5.2(d), also undergoes the constraints of the PHY layer, and is logically less significant since it follows the same evolution as reliability (throughput is the product of reliability, offered load and data frame

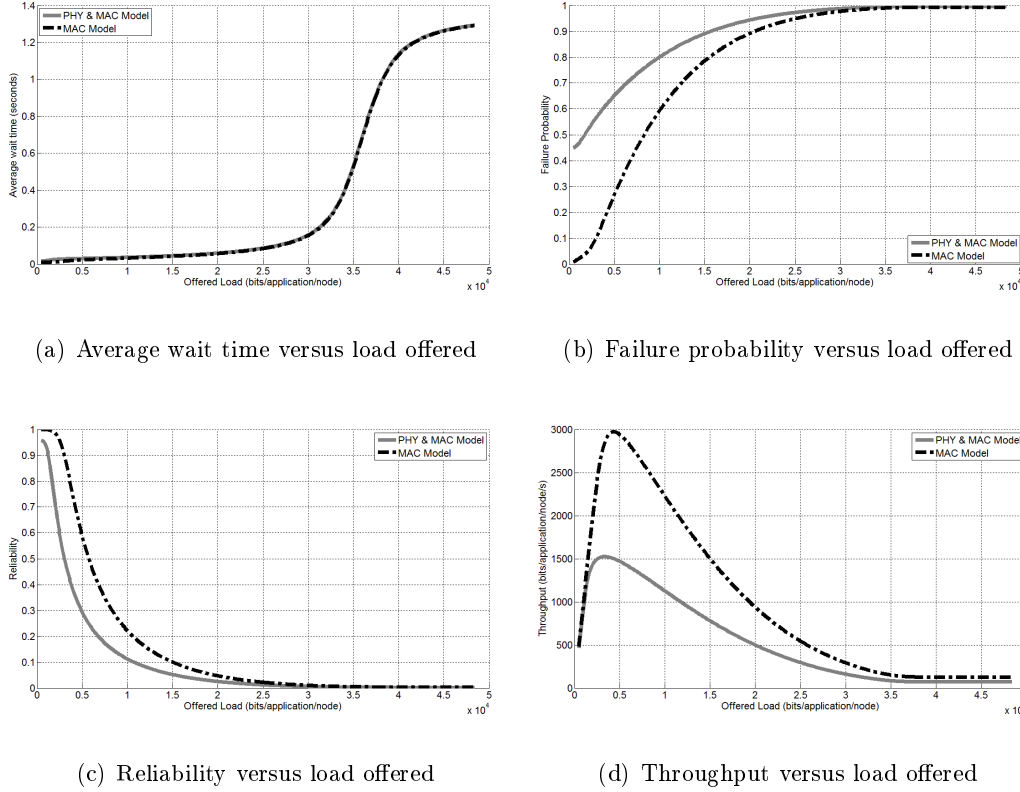


Figure 5.2: Comparison between IEEE 802.15.4 PHY & MAC Model and IEEE 802.15.4 MAC Model

size).

5.4.3 Evolution of Node Performance with Growing Densities

We use our model to compare node performances with three densities. We propose a network with 5 nodes, another with 10 nodes and a third with 50 nodes. We take into account the same outputs cited in the previous section. The major observation is that the IEEE 802.15.4 networks do not support heavy traffic. The denser the network is, the poorer are the performances. We note an increasing delay for denser networks, as observed in Fig. 5.4(a) (Fig. 5.5 proposes a zoom in to better appreciate the delay difference between the three cases). As the number of nodes increases, and with growing offered loads, collisions are more frequent and so the retransmissions are more recurrent. The switching phase to saturated network shows more significant differences between the three network scenarios. Each node queue begin to experience congestion problems; with more retransmission requirements, the queues are busier and the delays are longer. At saturation, the frame losses are

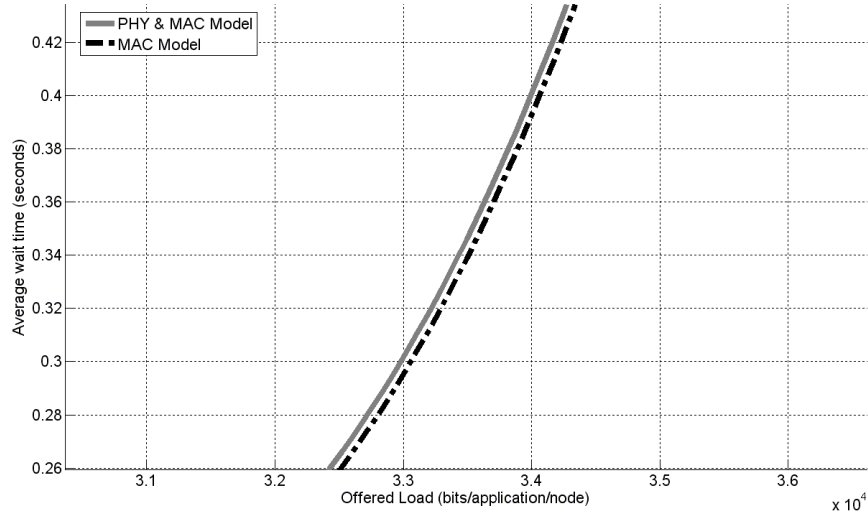


Figure 5.3: Zoom in on the average wait time versus load offered for IEEE 802.15.4 PHY & MAC and IEEE 802.15.4 MAC models

widespread (collisions, link constraints, frames discarded due to retry limits, etc.) for the three scenarios, but the number of nodes still has an impact because it has a negative influence on performances and channel availability (more collisions, more retransmissions, channel congestion, ...). The same reasoning explains a higher failure probability, as presented in Fig. 5.4(b) and a lower reliability as outlined in Fig. 5.4(c) for denser networks and with increasing offered load. The evolution of throughput, shown in Fig. 5.4(d), also matches with the interpretations cited above.

5.5 Conclusion

We presented, in this chapter, a model that mimics the IEEE 802.15.4 functionalities at the PHY and the MAC layers. We aimed to combine two relevant propositions. On the one hand, we modeled constraints that affect link quality using the Zuniga and Krishnamachari mathematical framework: distance, output power, noise, asymmetry and errors related to encoding and modulation. The PHY model bypasses the disk-shaped node range and expresses more precisely the degree of link unreliability. The output of this model represents an important outcome for estimating more faithfully the probability of transmitting frame failure. On the other hand, we enhanced the Park et al. IEEE 802.15.4 MAC layer model to extract the delay and the reliability. Our contribution seeks to improve the Park et al. approach in de-

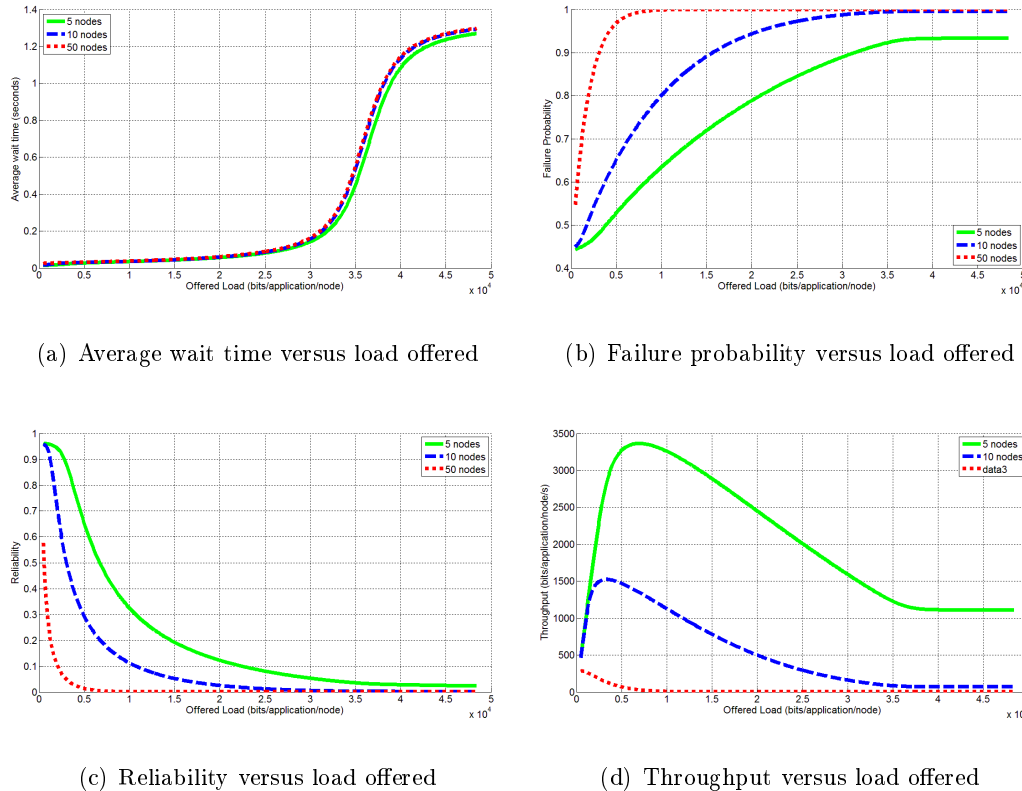


Figure 5.4: Performances evolutions for different densities using IEEE 802.15.4 PHY & MAC Model

termining inherent probabilities (frame transmission, free channel in CCA1/CCA2, failure,...) and combining it with the PHY model to better estimate wireless network parameters. The methodology adopted relies on a Markov chain that follows the flowchart described in Fig.6.2 and on an M/M/1/K queue that we includes with the Park et al. approach. The joint model is available at the SGIP NIST Smart Grid Collaboration website [7] for use.

The simulations showed that more precise estimations are provided by our model versus that by the Park et al. MAC model. The comparison between the two considerations highlights a notable performance deterioration using the combined model. This observation is quite obvious since this combination joins PHY constraints to collisions. Thus, our contribution improves the Park et al. approach by bypassing the assumption that failures are restricted to collisions. The amelioration of the Park et al. approach is not limited to the above description. We tried also to enhance the estimation of inherent probabilities by adjusting some expressions (as for α , β and P_{fail}) and modifying the resolution method to gather new parameters

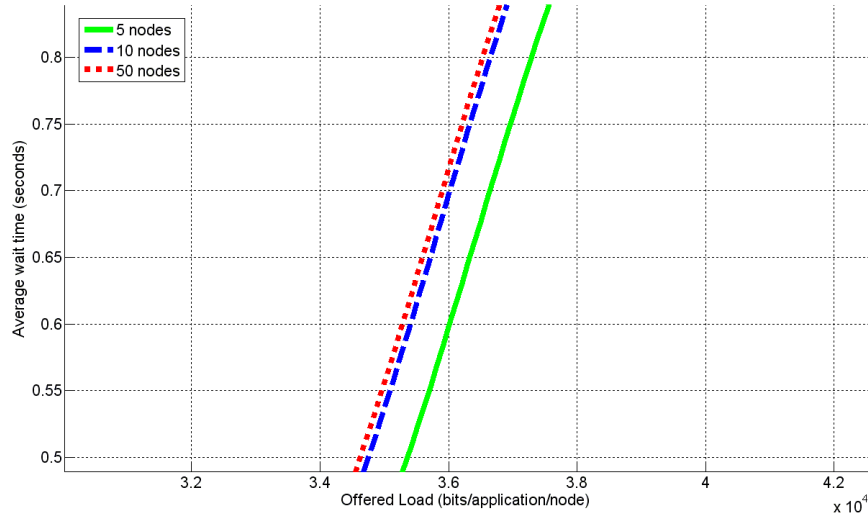


Figure 5.5: Zoom in on the average wait time versus load offered for different densities using IEEE 802.15.4 PHY & MAC Model

(such as p_0 , the probability that a node returns to the idle state, which is considered as an input in Park et al. work).

Our contribution proposes to mimic the IEEE 802.15.4 PHY and MAC layers mechanisms. Nonetheless, it is extensible for reproducing more precise wireless networks standards related to IEEE 802.15.4. It is also adjustable to other standards. Indeed, the considered PHY layer model is quite relevant but assumes that interferences are weak and/or stable. Moreover, the probability of an error at the PHY layer is averaged (through integration over maximum range and maximum asymmetry variation).

Chapter 6

Cooperation Enforcement for Packet Forwarding Optimization in Multi-hop Mobile Ad-hoc Networks

Contents

6.1	Introduction	102
6.2	Related Work	104
6.3	System Model And Problem Formulation	111
6.3.1	"The Weakest Link" TV Game Principle	111
6.3.2	The Proposed Model: Analogy with "The Weakest Link" TV Game	112
6.4	Self-Learning Repeated Game Framework and Punishment Mechanism	113
6.4.1	The Punishment Mechanism	115
6.4.2	The Self-Learning Repeated Game Framework Description	115
6.5	Performance Evaluation and Simulations Results	116
6.6	Conclusion	121

6.1 Introduction

In wireless Ad-hoc networks, nodes are self-organizing and autonomous. They manage their own resources and make their own decisions. In order to maintain network

Chapter 6. Cooperation Enforcement for Packet Forwarding Optimization in Multi-hop Mobile Ad-hoc Networks

connectivity, each node has to forward packets of other nodes. However, since they are known to have limited battery resources, these nodes usually tend to be non-cooperative. Indeed, they might sometimes reject forwarding requests in order to save their proper energy. Thus, nodes are reluctant to participate in routing which may lead the network connectivity to break down. It is then necessary to provide a mechanism that enforces cooperation between nodes and maintains the network connectivity. The problem of forwarding packets in a non-cooperative Ad-hoc network is widely studied, as we highlight afterwards, and many approaches have been proposed. The nodes act selfishly and tend to maximize their own benefits, thus, most of these studies rely on game theory [48] which is a suitable tool to deal with complex interactions between network nodes. From this perspective, the approaches can be classified into two categories depending on the mechanism used to enforce cooperation level between nodes. In the first category, the propositions use the virtual payment scheme. Zhong et al. [125] have proposed Sprite, a credit-based system that makes incentives for nodes to cooperate. Eidenbenz et al. have designed COMMIT [40], a routing protocol based on payment with virtual currencies. The requested intermediate nodes will perceive a compensation that is related to their residual energy level. Ad Hoc-VCG [14] and incentives modeling advanced by Crowcroft et al. [34] also belong to the first category. In the second one, the approaches employ mechanisms to enforce and maintain cooperation between nodes communities. Some works use reputation-based mechanisms. In instance, Kwon et al. [69] who have formulated a Stackelberg game where two nodes sequentially estimate the willingness of each other and decide to cooperate or not according to the opponent reputation score. Also, Buchegger and Le Boudec [24, 25] have defined mechanisms taking into consideration reputation system. Other solutions aim to maintain cooperation by considering punishment threat. In this case, Marti et al. [76] have defined "watchdog" and "pathrater" techniques that improve throughput by excluding misbehaving nodes. Felegyhazi et al. [45] have proposed a scheme that enables the nodes to reach the Nash Equilibrium, under topological conditions (i.e. dependence graph), relying on the "Tit-For-Tat" punishment. Altman et al. [11] have highlighted the "aggressive" punishment in [45] and have proposed milder punishment mechanism which guaranties a Nash Equilibrium and helps nodes to consume less energy. Han et al. [51] have advanced a self-learning repeated game framework based on punishment that determines the nodes optimal packet forwarding probabilities to maintain network connectivity. Pandana et al. [86] have considered the same aim as in [51] and have designed three learning schemes with a punishment mechanism under perfect/imperfect local observation and dependence graph conditions.

In this chapter, we propose a self-learning repeated game framework inspired by "The Weakest Link" TV game. In our approach, the set of nodes forming a route are considered as the candidates of the chains in the TV game. Each node forwarding probability can be seen as good answer probability for each candidate. Indeed, the maximization of global collective gains depends strongly on the cooperation between the candidates involved in the game. Thereby, we adopt the TV game concept to design our scheme with the objective of motivating nodes to create the longest chains and maximizing their utility values. This would increase the probability that the packets are delivered to the destination and then would optimize packets forwarding. Moreover, the framework relevance lies in the repeated game that enforces collaboration between nodes and a learning scheme that tends to reach better cooperation level. We consider also a punishment mechanism that would discourage nodes from acting selfishly. To assess the efficiency of our proposal, we compare our proposal to two other self-learning repeated game schemes proposed in [51] and [86].

The remainder of the chapter is organized as follows: In Section 6.2, we present the self-learning repeated game approaches that have inspired us to design our approach. The proposed model based on "The Weakest Link" TV game principle is illustrated in Section 6.3. The self-learning repeated game framework and punishment scheme are then presented in Section 6.4. Section 6.5 details simulations scenarios used to evaluate our proposal and analyzes the obtained results. Finally, concluding remarks are given in Section 6.6.

6.2 Related Work

We have been interested in some self-learning repeated game frameworks based on a punishment mechanism. We have been inspired by the framework proposed by Han et al. in [51] and those designed by Pandana et al. in [86]. They have helped us to design our own framework. In this section, we propose to give details about these approaches and how the Folk Theorem can solve the problem of the non-cooperative behavior when finding the Nash Equilibrium is not sufficient.

Self-Learning Repeated Game Framework proposed Han et al.

Han et al. propose their framework in order to distributively obtain the optimal forwarding probability in a wireless network. This approach is characterized by two major steps:

- First, the users apply a repeated game to punish misbehaving nodes in order

Chapter 6. Cooperation Enforcement for Packet Forwarding Optimization in Multi-hop Mobile Ad-hoc Networks

to maintain and promote the cooperation.

- Second, the users try to adapt their forwarding probabilities by learning the degree of cooperation in the network.

To build their framework, Han et al. suppose that there is a wireless network of K users and each user k has N_k routes for its packet transmission. Han et al. have assumed that the routes have been determined with a routing protocol, that it exists an end-to-end mechanism which is able to inform the source node that a packet is successfully delivered and that the routes are determined before the forwarding probability optimization. Each user has a utility function and use it to learn the degree of cooperation in the network. The utility function U_k of the user k is expressed by:

$$U_k = \sum_{i=1}^{N_k} P_k^i G \Pi(\alpha_j, j \in I_k^i) - F \alpha_k B_k \quad (6.1)$$

where I_k^i is the set of nodes on the route i of the user k , α_k is the willingness of the user k to forward the packet of another user, B_k is the probability that a node generates a packet, G is the benefit of a successful transmission and F is the cost of forwarding the packet, P_k^i is the probability that the user k transmits its packet through the route i , Π represents the probability of a successful transmission of a packet which depends on the forwarding probabilities of the nodes involved in the requested route.

The first term designates the average benefit of the user k which is closely dependent on the willingness of the users that form the set N_k . The second term points out the cost engendered by forwarding a packet of another user. Given that a user in the network has not a full knowledge about the behavior of others and can deviates from cooperation to optimize its utility, Han et al. formulate the problem as a non-cooperative game where each user tunes its own forwarding probability to maximize its own utility function:

$$\max_{0 \leq \alpha_k \leq 1} U_k(\alpha_k, \alpha_{-k}) \quad (6.2)$$

where $\alpha_{-k} = [\alpha_1, \alpha_2, \dots, \alpha_{k-1}, \alpha_{k+1}, \dots, \alpha_K]$ is the set of forwarding probabilities of all the users in the networks except user k .

In such situation, the Nash Equilibrium enables us to identify the strategy used by each user that maximizes its outcome when the other users are applying their strategies.

Definition 1: The Nash Equilibrium is some strategy set α^* for all nodes, such that for each node i , the following condition is verified:

$$U_i(\alpha_i^*, \alpha_{-i}^*) \geq U_i(\alpha_i, \alpha_{-i}^*), \forall i, \forall \alpha_i \in [0, 1] \quad (6.3)$$

where $\alpha_{-i}^* = (\alpha_1^*, \dots, \alpha_{i-1}^*, \alpha_{i+1}^*, \dots, \alpha_N^*)$.

This condition emphasizes that no node can increase its utility by operating a unilateral change of its forwarding probability, while all other nodes play the Nash Equilibrium strategy. Nevertheless, this equilibrium matches with the strategy where $\alpha_i = 0, \forall i$. Indeed, the outcome of each node is dependent on the cooperation willingness of others meanwhile the cost is tied to its own forwarding probability. Obviously, to maximize its utility, each node would selfishly drop the packets of the others. In such situation, no node has the willingness to cooperate and the successful transmission probabilities tend to be close to zero. Applying a game for one step would motivate nodes to use the Nash Equilibrium strategy which is not convenient. Thereby, a repeated game lead a strategy to be sensitive to the past observations. This fact can build the cooperation between the nodes by considering reputation and/or retribution awareness.

The repeated game can enforce any misbehaving player to be more cooperative. This is due to the punishment that can be applied to encounter selfishness by fading the greediness benefits. Hence, it is preferable to all players to be rather cooperative. To enforce this cooperation and to maintain it, the infinitely repeated game rises as the recourse to use as supported by the Folk Theorem [47]. The average utility outcome of a player k in an infinitely-repeated game is expressed as following:

$$\bar{U}_k = (1 - \beta) \sum_{t=1}^{\infty} \beta^{t-1} U_k(t) \quad (6.4)$$

where $U_k(t)$ is the utility outcome at period t and β is the discount factor.

The Folk Theorem is generally stated as: "Any strictly individually rational and feasible payoff vector of the stage game can be supported as a subgame-perfect equilibrium average payoff of the repeated game." The outcome for an infinitely repeated game can give better payoffs than those that can be obtained with Nash Equilibrium, especially for a discount factor close to 1. This parameter is viewed as the future importance or as the probability that game ends. When it is close to zero, the future is not "significant" and nodes deviations are more attractive against the "forgiveness" of other nodes. On the other hand, when the discount factor is close to one, cooperative nodes are aware of misbehaving ones. A permanent punishment threat corresponds to a discount factor close to 1 and obliges selfish nodes to be

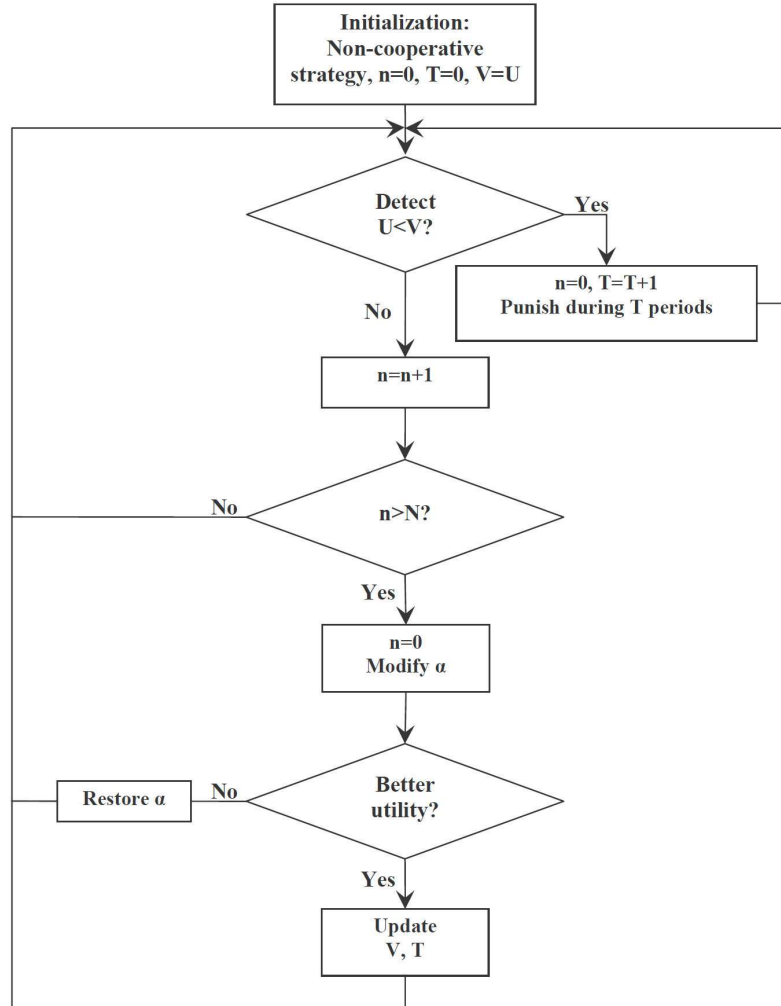


Figure 6.1: The self-learning repeated game flowchart proposed by Han et al.

more cooperative.

Then, to enforce the cooperation in the network, Han et al. design their algorithm to enable the users to learn the optimal packet forwarding probability step by step. The strategy of repeated game is applied to ensure the cooperation by performing a punishment if deviations are detected. The framework proposed by Han et al. for a user k is depicted in Fig. 6.1. The algorithm begins with an initialization phase where all nodes are supposed to be non-cooperative (e.g. their forwarding probabilities are equal to 0). Each user sets a cooperation time counter

n and the punishment time T to 0 and a trigger threshold V to the utility provided by the non-cooperative strategy (obviously equal to 0). Then, the users launch the repeated game strategy. During this phase, at each step, each user compare its utility value U to the threshold V . When the utility value is higher than the trigger threshold, this means that the cooperation is effective in the network and the cooperation time counter n is increased. Otherwise, when a deviation is detected ($V > U$), the punishment phase is engaged during a time T which becomes longer with the persistence of the non-cooperative behavior. The punishment comes down to play the non-cooperative strategy. The long periods of punishment are considered as a threat for nodes that act selfishly. Indeed, if they keep the non-cooperative strategy, their utility values are being forced to be null. Nevertheless, performing such punishment scheme can be restrictive for the network performance as punishment periods can become too long.

When a user detects that the cooperation is getting stable (the timer n reaches a predefined constant N which designates the number periods needed to conclude that the cooperation is enforced), it enters in the self-learning phase through which it tries to adjust its forwarding probability to optimize its own utility. A possible manner to tune this probability is to choose a random value between 0 and 1 and to check if it enables to reach better utilities values. If it is the case, the user updates its trigger threshold V ($V=U$) and its punishment period T as it expects that the other users are willing to cooperate.

Self-Learning Repeated Game Frameworks advanced by Pandana et al.

Pandana et al. have adopted the same problem formulation as Han et al. As cited previously, in a wireless network, the nodes may behave selfishly as they cannot know the degree of cooperation of the other nodes and suggest a self-learning repeated game framework to overcome the non-cooperative situation.

To design their framework, Pandana et al. have chosen to define a utility function based on energy consumption. First of all, they have supposed that the network is consisting of N nodes and have considered the same assumption as Han et al.

As mentioned, Pandana et al. have chosen to express the utility function of each node by considering the amount of energy spent to transmit and to forward packets. For this purpose, the average power expended in a node i to transmit has been defined as:

$$P_s^{(i)} = \sum_{r \in V_i^s} \mu_{S(r)} \cdot K \cdot d(S(r), n(S(r), r))^\gamma \quad (6.5)$$

Chapter 6. Cooperation Enforcement for Packet Forwarding Optimization in Multi-hop Mobile Ad-hoc Networks

where $\mu_{S(r)}$ is the transmission rate of source node $S(r)$, K is the transmission constant, $d(i, j)$ is the distance that separates node i from node j , $n(i, r)$ represents the neighbor of node i that belongs to route r , γ is the transmission path-loss coefficient and V_i^s denotes the set of routes where node i is the source of the packets. The product $K \cdot d(S(r), n(S(r), r))^\gamma$ represents the reliable successful transmission power.

Each node i , in this framework, is characterized by its forwarding probability α_i . Then, the probability of successful transmission from node i to its destination j is closely tied to the forwarding probabilities of intermediates nodes involved in the route that joins node i to node j . The probability of successful transmission is expressed by:

$$P_{Tx,r}^i = \prod_{j \in (r \setminus S(r) = i, D(r))} \alpha_j \quad (6.6)$$

where $D(r)$ is the destination of node i on the route r and $(r \setminus S(r) = i, D(r))$ is the set of intermediate nodes on the route r .

From the two expressions defined above, it is possible to define the *good power* used by a node i which corresponds to the energy consumed to transmit a packet that would be successfully delivered to the destination. Hence, this amount of energy is expressed by:

$$P_{s,good}^{(i)} = \sum_{r \in V_i^s} \mu_{S(r)} \cdot K \cdot d(S(r), n(S(r), r))^\gamma \cdot P_{Tx,r}^i \quad (6.7)$$

As node i can also be an intermediate node, it consumes energy when it forwards the packets of other source nodes. This amount of energy is defined by:

$$P_f^{(i)} = \alpha_i \cdot K \cdot \sum_{r \in W_i} d(i, n(i, r))^\gamma \cdot \mu_{S(r)} \cdot P_{F,r}^i \quad (6.8)$$

where $P_{F,r}^i$ is the probability that node i receives the packet to forward in route r . The expression $\mu_{S(r)} \cdot P_{F,r}^i$ represents the rate of data received by node i from the source in route r . The probability $P_{F,r}^i$ of an intermediate node i in a route r is given by:

$$P_{F,r}^i = \prod_{j \in f_r^1, f_r^2, \dots, f_r^{m-1}} \alpha_j \quad (6.9)$$

where $r = \{S(r), f_r^1, \dots, f_r^{m-1}, f_r^m = i, \dots, f_r^n, D(r)\}$ is the $n + 1$ hops from the source $S(r)$ to the destination $D(r)$. In this expression, node i is the m^{th} intermedi-

ate node designated by f_r^m and it is clear that this probability is depending on the forwarding probabilities of all intermediate nodes that precede node i in the route r .

Relying on the node power used for either transmitting its own packets and forwarding the data of other nodes, Pandana et al. define the *self – transmission efficiency* which is expressed by:

$$U^{(i)}(\alpha_i, \alpha_{-i}) = \frac{P_{s,good}^{(i)}}{P_s^{(i)} + P_f^{(i)}} \quad (6.10)$$

where α_i denotes the forwarding probability of the node i , α_{-i} represents the following set of probabilities $[\alpha_1, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_N]$, $P_{s,good}^{(i)}$ is the energy consumed by node i to transmit a packet that is successfully delivered to the destination (Eq. (6.7)), $P_s^{(i)}$ corresponds to the average power expended in a node i to transmit (Eq. (6.5)) and $P_f^{(i)}$ represents the energy consumed by node i when it forwards the packets of other source nodes (Eq. (6.8)).

As stated in the work of Han et al., the Nash Equilibrium consists to play the selfish strategy and propose a self-learning repeated game with a punishment mechanism to enforce cooperation in the network. The punishment relies on monitoring any selfish behavior and decreasing the outcomes of misbehaving nodes by dropping their packets. In such case, the nodes which apply punishment have to announce it to avoid the interpretation of the punishment as a deviation and thereby the spreading of the selfish behavior in the network. Relying on such monitoring enables then to exploit the Folk Theorem. This monitoring assumes that the nodes that detect a deviation exchange their reports on the identity of the misbehaving node and identify this node by intersecting the exchanged information. This process can be not suitable in a dense network.

Pandana et al. have proposed three self-learning repeated game frameworks based on the utility function defined above. The first one represents a self-learning framework under perfect observability which have been used as a benchmark for performance evaluation. The two others are self-learning frameworks based on local observability, one is based on learning through flooding and the second on learning with utility prediction.

We give some details about the learning through flooding algorithm which we have considered in our evaluation. The only information that a node can observe is the evolution of its own utility function. To learn the forwarding probability, the best way is to gradually increase its forwarding probability and to check if the utility becomes better. If it is the case, the new forwarding probability will be employed.

Chapter 6. Cooperation Enforcement for Packet Forwarding Optimization in Multi-hop Mobile Ad-hoc Networks

The old probability is kept if the utility is not improved. This algorithm enables the nodes to change at the same time their forwarding probabilities. Then, a better cooperation level would be progressively spread. The algorithm is detailed in the following

Algorithm 1 Self-Learning Repeated Game Algorithm based on Learning through Flooding

- 1: Initialization: $t = 0, \alpha_i^t = \alpha_0, \forall i$, small increment ξ , normalization factor η , lowest cooperation level α_{min}
 - 2: **for** Iteration: $t = 1, 2, \dots$ **do**
 - 3: Calculate $U^{(i),t-1}(\alpha_i^{t-1})$ and $U^{(i),t-1}(\alpha_i^{t-1} + \xi)$
 - 4: Calculate $\delta U^{(i),t-1} = U^{(i),t-1}(\alpha_i^{t-1} + \xi) - U^{(i),t-1}(\alpha_i^{t-1})$
 - 5: **for** each i such that $\delta U^{(i),t-1} > 0$ **do**
 - 6: $\alpha_i^t = \alpha_i^{t-1} + \eta \frac{\delta U^{(i),t-1}}{U^{(i),t-1}(\alpha_i^{t-1})}$
 - 7: $\alpha_i^t = \max(\min(\alpha_i^t, 1), \alpha_{min})$
 - 8: **end for**
 - 9: Keep monitoring the deviation
 - 10: **if** a deviation is detected **then**
 - 11: Start the punishment scheme
 - 12: **end if**
 - 13: **end for**
-

In the next section, we describe a self-learning repeated game framework inspired by "The Weakest Link" TV game. We follow the same concept of repeated game used in the described approaches by taking more simpler assumptions than the approaches of Pandana et al. and a better punishment mechanism than the one used in the work of Han et al.

6.3 System Model And Problem Formulation

To optimize packet forwarding in ad-hoc networks, we strongly believe that applying the concept of "The Weakest Link" TV game is an interesting solution. We explain, firstly, the concept of the game that we want to reproduce and then we detail the proposed model.

6.3.1 "The Weakest Link" TV Game Principle

"The Weakest Link" TV game is a game where a group of candidates try to answer correctly, relying on their knowledge, to the questions asked by the TV host. They aim to gather the highest amount of money through successive rounds. In each round, the players try to form a chain of nine correct answers to reach the highest gain. Before answering to a question, a candidate has the possibility to save the collected gain, provided by good answers of precedent candidates, by saying "Bank".

6.3. System Model And Problem Formulation

It is obvious that the longer the chain is, the higher the gain gets. Nevertheless, a player who gives a wrong answer to a question and did not save the collected gains shall break the good answers chain and reset the gain to zero. Secondly, if the candidates save rapidly the collected gains, the chains will be too short to reach important amounts of money. Thus, answering wrongly or saving collected money return the chain counter to zero. The candidates have to avoid being frequently in these situations if they want to maximize their gains. The earnings scale expresses the potential round gain according to correct answers chain length. For example, if there are four good answers and the current player decides to save collected money, the total gains grow with the amount that corresponds to the chain length. Then, the candidates try to create another chain of good answers. Therefore, the key parameter that influences the maximization of earnings is the probability of giving a correct answer.

6.3.2 The Proposed Model: Analogy with "The Weakest Link" TV Game

Our objective is to define an approach that enforces cooperation in a distributed way. The model we propose is inspired by the principle of the TV game. We note interesting analogies between "The Weakest Link" TV game and the Ad-hoc network. The nodes are assumed to be the candidates of the game and forwarding a packet from a node to the next hop is considered as a good answer. We believe strongly that the TV game concept can be used to encourage nodes to cooperate and therefore to optimize packet forwarding. Hence, the nodes, along a route, aim to create the longest chain of successful forwarded packets to get better utility. Despite of the TV game, when a chain of forwarded packet is broken, only the nodes that form the chain and the node which saves the gains will be rewarded. The set of all nodes composing the route is rewarded by the collected gains only if the packet reaches destination. To formulate the expected utility of a node in a route, we propose expressions inspired by [17]. Let α_i and β_i be the forwarding and the saving gain (with chain breaking) probabilities, respectively, for each node i . We assume in our work that $\alpha_i = 1 - \beta_i$. Given a route R with $N-1$ hops (N nodes), we define in Eq. (1) the average gain that a node i can expect when it plays the role of the n^{th} link of the chain. Let $S(i)$ be the next hop of node i in route R . We mean by $C[n]$ the collected earnings when a chain of $(n-1)$ successful transmissions is transformed in currency and by F the cost of forwarding other's packets. Considering the vector of forwarding probabilities α , the utility of a node

i is given by:

$$U_i^R(n, \alpha) = \begin{cases} U_{S(i)}^R(1, \alpha) & \text{if } n = 0 \\ (1 - \alpha_i) \cdot C[n] + \alpha_i \cdot (U_{S(i)}^R(n + 1, \alpha) - F) & \text{if } 0 < n < N \\ C[n] & \text{if } n = N \end{cases} \quad (6.11)$$

When the node i is an intermediate node and given that it wants to maximize its gains, it has to choose between saving the collected currency or increasing the chain length (relying on the cooperation of the successor(s) to maximize benefits despite of the cost of forwarding). We assume that the source of the packet (i.e. $n=0$) will send it with the probability equal to 1. Subsequently, the gains depend on the decisions of the next hops. In addition, when the packet reaches the destination (i.e. $n=N$), the node will save the collected gains with a probability equal to 1. In the latter case, the chain has the length of the route and the gains are maximized.

Therefore, we can formulate this problem as a non-cooperative game where each node will adjust its forwarding probability in order to maximize its own utility. A node i can belong to more than one route, its own utility is then the sum of each route utility, called U_i . To solve this problem, it is necessary to find the Nash Equilibrium of the game.

However, as mentioned in several works as [11, 51, 86], this equilibrium matches with the strategy where $\alpha_i = 0, \forall i$. To avoid a poor network performance, we propose a self-learning repeated game framework, inspired by the concept of "The Weakest Link" TV game that we call The Weakest Link scheme. The objective of this framework is to enforce cooperation between nodes through learning and punishing threat mechanisms.

6.4 Self-Learning Repeated Game Framework and Punishment Mechanism

As the Nash Equilibrium corresponds to a non-cooperative strategy, it is more suitable to design cooperation under repetitive game. Applying repeated game scheme match perfectly with our proposed model. Indeed, in each game step, along each route, the nodes try to maximize their utilities and would choose to cooperate. In this chapter, we consider an infinite repeated game, where the game duration is unknown to all nodes. Relying on the Folk theorem [47], the outcome of an infinitely

6.4. Self-Learning Repeated Game Framework and Punishment Mechanism

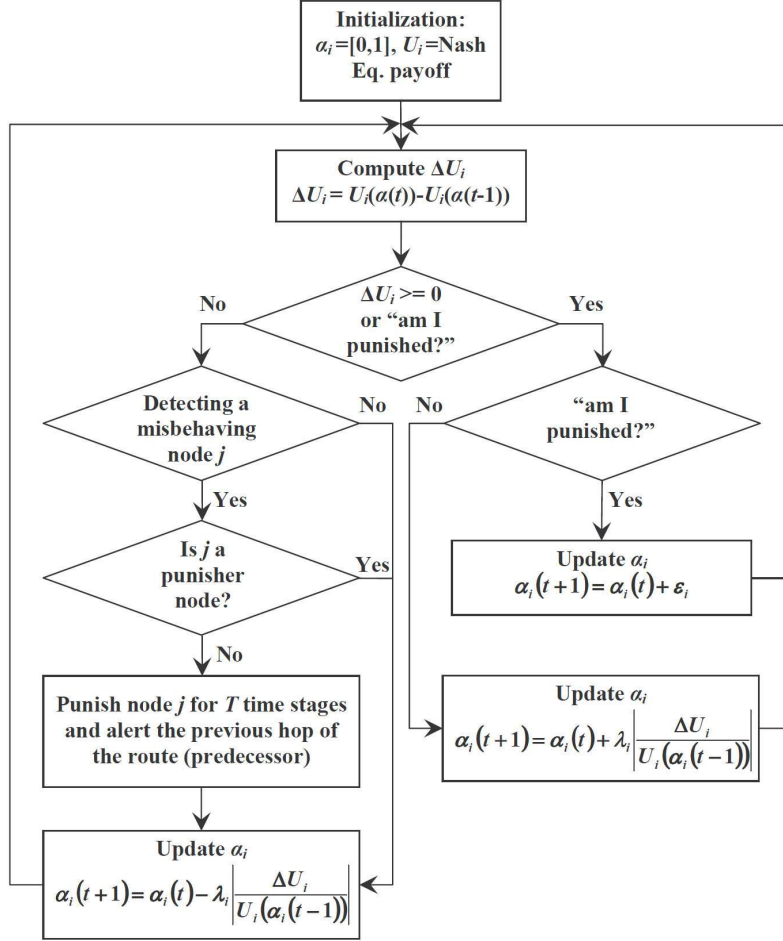


Figure 6.2: The self-learning repeated game flowchart

repeated game can give better payoffs than those that can be obtained with Nash Equilibrium, especially when permanent punishment threat obliges selfish nodes to be more cooperative. Therefore, we propose a repeated game that enforces cooperation and maintains it through a designed punishment mechanism to encounter misbehaving forwarders. The framework we propose is presented by the flowchart in the Fig. 6.2.

In the initialization step, all nodes are more or less selfish. They can set their forwarding probabilities to 0 (the Nash Equilibrium strategy played with one stage game). We assume that the routes are determined with a routing protocol and that each node knows all the routes to which it belongs, as considered in [51, 86]. Then, the nodes start playing repeated game strategy. Note that they are rational and

want to make benefits. Thus, at each step, each node learns through its utility the cooperation level of other nodes and adjusts its forwarding probability following to the others learnt behavior. It is possible that a node deviates from cooperation, then a punishment scheme is designed to discourage misbehaving nodes and to ask them to be more cooperative in the future. It is applied as soon as a selfish behavior is detected and subsequently the framework satisfies the Folk theorem.

6.4.1 The Punishment Mechanism

In this section, we present the punishment mechanism through a simple scenario. We assume that A and B are two successive nodes along a route. We suppose that the node B is a misbehaving node. The node A as one among the "closest" nodes to node B (i.e. the predecessor of B) is designed to punish it (if B rejects the request of A). We assume that the node A is able to detect the lack of cooperation of B (able to distinguish between a packet drop and a packet loss); the node A can conclude, by listening the channel, that node B is misbehaving when it does not forward the packet to the next node. To punish the selfish node B , the node A fixes its forwarding probability to 0 when the packet has to pass through node B . In this case, the node B will not be able to receive any packet from node A . Thereafter, the node B will be excluded from all chains in which its predecessor is in punishment mode. This punishment cancels the node B benefits for a period T and enforces it to cooperate (as its utility decreases). To avoid the propagation of the punishment mode over all nodes, when the node A is designated to punish the node B , the former one informs its predecessor about the execution of the punishment. Then, the punishment act is not interpreted as a deviation. It is important to mention that we assume that the nodes are not malicious.

6.4.2 The Self-Learning Repeated Game Framework Description

At each step of the repeated game, each node compares its current utility value with the former value. If the current utility is better, a cooperation enforcement is concluded. Thereby, the forwarding probability is increased proportionally with the enhancement of the utility to promote the cooperation level. The upgrade of the cooperation level is also led by the coefficient λ_i . It expresses the node sensitivity to the cooperation enforcement. However, when the current utility drops, it is analyzed as a come back to selfishness. Thus, the forwarding probability will decrease (proportionally with the difference). Analogically with "The Weakest Link" TV game, a candidate chooses to break the chain and insures gains if he notices that the following candidate tends to make wrong answers. Hence, cooperative nodes

6.5. Performance Evaluation and Simulations Results

are sensitive to the behavior of the other nodes. As described in the punishment scheme, a node checks if its successor deviates. This deviation can be the result of punishment and an announcement is made to avoid selfishness propagation. The deviation without any notification is considered as a selfish behavior and then the punishment procedure is applied on the misbehaving node. Indeed, during a period T , no packet from its punishing predecessor reaches it. This causes a dramatic utility decrease. Therefore, a punished node is encouraged to be more cooperative in order to avoid longer penalization. Indeed, it increases its forwarding probabilities by a step equal to ε_i . To make possible the cooperation enforcement, another assumption must be considered. In fact, if the maximum gains that can be collected are lower than the forwarding cost, the nodes would not forward any packet even under punishment threat. Finally, it is important to mention that the proposed scheme work well when the mobility is moderate. In other words, transferring all packets on a route must be faster than route breakage due to mobility. We take into consideration this assumption in our simulations.

This framework aims to construct the longest chains on an established route between source and destination nodes, and routes are provided by a routing protocol to all nodes. If a route is not available at any node in the route, the routing tables must be updated. Moreover, the forwarding probability can be used by the routing protocol to determine better routes or to update routes in order to avoid misbehaving nodes to be a part of a route. Also, it is important to mention that this framework is adapted to scenarios where some nodes can either be only sources and/or destinations of packets as a source node forwards its packets to the next hop with a probability equal to 1 (only if the source node apply the punishment mechanism) and a destination node is not involved in the forwarding process.

6.5 Performance Evaluation and Simulations Results

In this section, we evaluate the performances of the proposed approach through two scenarios: the widely used ring network and the random network. We compare our proposal to two other approaches which have inspired us: the scheme proposed by Han et al. [51] and the learning through flooding algorithm designed by Pandana et al. [86]. We implement our framework and do scenarios simulations on MATLAB 7.8.0. For the scheme Han et al., we fix the parameter N to 50 and for the Learning through Flooding algorithm of Pandana et al., we set ξ to 0.001 and η to 1. As remainder, these algorithms and parameters are defined in Section 6.2.

Firstly, we consider a ring network of 25 nodes. The distance between any source and destination is h hops (if a node i is the source, node $\text{mod}(i + h, 25)$ is

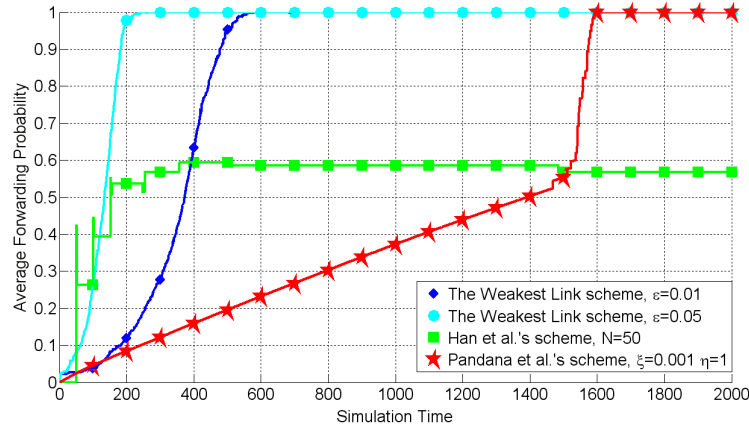


Figure 6.3: The evolution of the average forwarding probability for different self-learning repeated game schemes in the case of the ring network

the destination). For each intermediate node, it is imperative that the forwarding cost must be less important than the maximum gain along a route. By the way, the nodes must have benefits in order to maintain a cooperative behavior. In our simulations, each successful forwarding increments by 1 the gain corresponding to the formed chain. In this scenario, we fix h to 6. We consider that one node can be a source or a destination at most one time and an intermediate node at most five times. The nodes initialize their forwarding probabilities at 0 (i.e. the Nash Equilibrium strategy). We choose also to fix the forwarding cost to 3 (i.e. to make benefits possible) and the period of punishment T to 3 time steps.

We represent in Fig. 6.3 the evolution of the average forwarding probabilities for the considered self-learning repeated game frameworks over 2000 time steps. For our proposal, we depict two different scenarios and each one is characterized by a specific value of ε (where $\varepsilon_i = \varepsilon$ for all nodes and ε takes respectively the values 0.01 and 0.05). The coefficient λ_i is fixed to 0.01 for all nodes. We have also fixed the characteristic parameters of the two other schemes as mentioned in the plot. We invite the reader to refer to [51] and [86] in order to understand the meaning of these parameters (we also assume that each two successive nodes in a route are separated by the same distance in order to simplify the computation of the utility functions when the scheme of Pandana et al. is used).

First of all, we remark that the average forwarding probability observed for the scheme of Han et al. converges to a value that turns around 0.6 and we note that

6.5. Performance Evaluation and Simulations Results

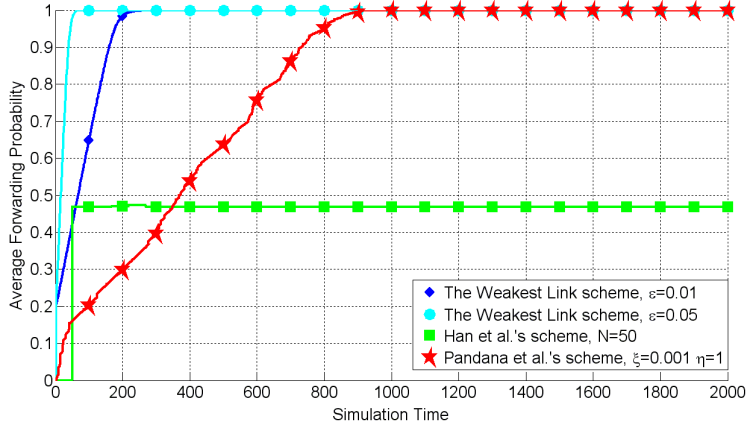


Figure 6.4: The evolution of the average forwarding probability for different self-learning repeated game schemes in the case of a random network

the updates of this average probability become less frequent as time goes on. This finding logically indicates that the cooperation between nodes is limited even if the average cooperation level is rather substantial. The punishment mechanism adopted by the framework of Han et al. is clearly the major cause behind this result and we explain later the reasons behind that. The average forwarding probability obtained for the learning through flooding algorithm designed by Pandana et al. converges to 1 but after too many steps (compared to the Weakest Link scheme result). In their evaluation, Pandana et al. have chosen an initial forwarding strategy for nodes (i.e. all forwarding probabilities initialized to 0.5) different from the Nash Equilibrium strategy (i.e. non-cooperative strategy). This assumption enables to the nodes of the network that adopt this scheme to reach more rapidly high cooperation levels but leads to a skewed evaluation of the algorithm. Regarding the Weakest Link scheme, the average forwarding probability converges to 1, as the scheme of Pandana et al. but needs less time to reach high cooperation levels. We depict the corresponding evolution for two values of ε : 0.01 and 0.05. We remark that the nodes become cooperative faster as they are more reactive to the punishment mechanisms (i.e. higher increase of the forwarding probability when punishment and utility decrease are detected). Then, the scenario where ε is equal to 0.05 highlights a quicker convergence of the average forwarding probability.

To support these conclusions, we determine for each scheme the average packet delivery rate at the destination and the ratio between the forwarded packets and

Chapter 6. Cooperation Enforcement for Packet Forwarding Optimization in Multi-hop Mobile Ad-hoc Networks

Table 6.1: Evaluation metrics in the case of the ring network (all forwarding probabilities initialized to 0)

Schemes	Metrics	Avg. PDR @ Dest.	Fwd. Pkts / Dlv. Pkts
The Weakest Link ($\varepsilon=0.01$)		75.19%	5.1195
The Weakest Link ($\varepsilon=0.05$)		90.30%	5.0452
Pandana et al.		21.29%	5.7362
Han et al.		0.23%	182.7456

Table 6.2: Evaluation metrics in the case of a random network (all forwarding probabilities initialized to 0)

Schemes	Metrics	Avg. PDR @ Dest.	Fwd. Pkts / Dlv. Pkts
The Weakest Link ($\varepsilon=0.01$)		92.94%	4.5709
The Weakest Link ($\varepsilon=0.05$)		98.21%	4.5459
Pandana et al.		67.31%	4.7061
Han et al.		1.62%	8.6341

the delivered packets. Table 6.1 lists the corresponding results for the ring network scenario. It is important to remind that the routes have a length of 6 hops. Then, in the ideal case when the nodes fully cooperate, each packet delivered to the destination needs 5 forwards. The found results reflect the efficiency of our scheme. This efficiency is tuned by the input parameters. As we show, the choice of the parameter ε has an important impact on the convergence speed of the average forwarding probability to 1. The scheme of Pandana et al. shows a limited efficiency over the simulation time because of the low convergence to a satisfying cooperation level.

For the two cited schemes, the nodes become cooperative as time goes on. The two algorithms rely on efficient punishment mechanisms. They share the penalty of the misbehaving node instead of all nodes in the network. The values obtained for the ratios of forwarded packets over delivered packets (slightly higher to 5) prove the effectiveness of these solutions to establish cooperation among nodes (even with some delay for the scheme of Pandana et al.). On the contrary, when a node that uses the framework of Han et al. detects a defection, it punishes all other nodes. This reaction engenders the propagation of the non-cooperative strategies and dramatically falls down the network performance.

In the same way, we consider the scenario of a random network consisting of 100 nodes with 1000 source-destination pairs. Fig. 6.4 depicts the evolution of the average forwarding probability for each scheme and Table 6.2 lists the evaluation metrics for the scenario of the random network (the average number of forwarders per route in this scenario is 4.535 nodes). The previous observations and interpretations match with the results obtained for this scenario. We note also that the performance

6.5. Performance Evaluation and Simulations Results

Table 6.3: Evaluation metrics in the case of the ring network (all forwarding probabilities initialized to 0.5)

Schemes \ Metrics	Avg. PDR @ Dest.	Fwd. Pkts / Dlv. Pkts	Avg. Trans. Eff.
The Weakest Link ($\varepsilon=0.01$)	95.24%	5.0555	0.1617
The Weakest Link ($\varepsilon=0.05$)	97.66%	5.0278	0.1639
Pandana et al.	93.84%	5.0492	0.1581

Table 6.4: Evaluation metrics in the case of the random network (all forwarding probabilities initialized to 0.5)

Schemes \ Metrics	Avg. PDR @ Dest.	Fwd. Pkts / Dlv. Pkts	Avg. Trans. Eff.
The Weakest Link ($\varepsilon=0.01$)	99.77%	4.4674	0.1832
The Weakest Link ($\varepsilon=0.05$)	99.31%	4.4722	0.1829
Pandana et al.	93.54%	4.525	0.1778

of the Weakest Link and Pandana et al.'s schemes are better. This can be explained by higher opportunities to improve the utility function (i.e. nodes belong to a higher number of routes) compared to the case of the ring network.

Pandana et al. have defined the utility function of a node as its transmission efficiency. The transmission efficiency is the ratio of successful self-transmission power over the total consumed power (self-transmission and forwarding). We aim to compare our proposal to the algorithm of Pandana et al. using the cited criterion. We consider the same scenario of the ring network, as previously, and we initialize the forwarding probabilities of nodes at 0.5 to be as close as possible to the simulation inputs considered by Pandana et al.

We list in Table 6.3 the average packet delivery rate, the ratio of forwarded packets over delivered packets and the average transmission efficiency obtained by the Weakest Link scheme and the learning through flooding algorithm for the ring network scenario (the same scenario as previously). We note that the two frameworks highlight high packet delivery rates at destination (i.e. over 93 % with better performance for the Weakest Link scheme) and strong forwarding effectiveness (i.e. ratios of forwarded packets over delivered packets slightly higher than 5). Regarding the comparison based on the transmission efficiency, Table 6.3 emphasizes that the Weakest Link scheme (with the different values chosen for ε) reaches higher average transmission efficiency than the algorithm of Pandana et al. and then allows nodes to use better their energy. This finding can be explained by the upper packet delivery rate at the destination obtained for the Weakest Link scheme.

We can in addition emphasize that the behavior of nodes under the Weakest Link scheme enables nodes to have more "elastic" behavior towards defections. There-

after, it is possible to avoid useless forwards and save energy. In fact, when a node detects a reduction in its utility function, it decreases its forwarding probability and then becomes reluctant to cooperate as the delivery of packets is not accurate. For the algorithm of Pandana et al., the forwarding probability can either increase or remain the same but never decreases. Then, the nodes maintain their cooperation level even if a defection is detected and can uselessly consume their energy.

We compute the same evaluation metrics for the case of a random network which consists of 100 nodes and 1000 source-destination pairs. Each route has at average 4.463 forwarders. We list in Table 6.4 the obtained results for this scenario. We note that the effectiveness of our proposal is verified and that the transmission efficiency provided by the Weakest Link scheme is always better than the one of the algorithm of Pandana et al..

We have to mention that Pandana et al. have also proposed another framework based on utility prediction. It was highlighted in [86] that this latter scheme enables nodes to have better transmission efficiency. Anyway, the behavior of the average forwarding probability follows the same one of the learning through flooding scheme insofar as nodes can only maintain or increase their forwarding probabilities. Nevertheless, a convenient choice of the parameter ε for our scheme enables us to still improve the network performance if necessary.

6.6 Conclusion

In wireless ad-hoc networks, nodes are requested to forward traffic. However, because of limited energy resources, they might refuse to collaborate in order to save their energy. This can lead to a significant amount of lost packets and a deterioration of the network performances.

In order to overcome this problem, we have proposed in this chapter a self-learning repeated game framework that aims to enforce cooperation between nodes. Our framework is inspired by "The Weakest Link" TV game concept. Indeed, the amount of the global collective gains strongly depends on the cooperation degree between the candidates involved in the game. The candidates try to form the longest chain in order to reach the highest gain. Analogically, the nodes, along a route, would tend to achieve the longest sequence of successful packet forwarding and therefore assure that the packet reaches the destination. Our approach is designed as a self-learning repeated game framework that enables nodes to learn each others cooperation levels. Therefore, nodes that are in a same route and that have a high cooperation level may encourage the other nodes of the route to get more cooperative. For this aim, a punishment mechanism has been considered. Thereby,

misbehaving nodes are punished and their utility would dramatically decrease. This allows the network to maintain a relatively satisfying cooperation level.

Simulations have been run and the results have shown that our scheme is efficient for the ring network scenario as well as for the random network scenario. It has been also shown that our proposal outperforms other self-learning repeated game frameworks presented in the literature.

Chapter 7

Conclusion

We conclude this dissertation by summarizing the key contributions of the thesis and proposing some directions for future research.

7.1 Contributions of the Thesis

In this thesis, we have been interested in mobile wireless networks which are structured around humans. We have aimed to propose mechanisms and metrics which are contributing towards improving the evaluation and the design of communication protocols in such networks.

7.1.1 Proposing a Tensor-Based Link Prediction Framework

As the networks of interest are human-centered, we have been motivated by studying and exploiting the social aspect that characterizes such networks. In fact, social interactions highly impact the structure of a network and the social ties define how a network is structured. To extract information about social networks, the Social Network Analysis (SNA) has been exploited in data-mining to analyze them and strongly helped in defining the link prediction problem. Motivated by these findings, we have been the first to propose our own tensor-based link prediction framework to perform link prediction in mobile wireless networks as MANETs and DTNs. Our framework is based on specific salient steps. First, it tracks the contact occurrences during an historical. The tracking information is recorded in a third-order tensor which corresponds to a succession of adjacency matrices. These matrices represent the detected contacts for successive periods of tracking. Then, after collecting the evolution of the topology/neighbors, the collected data is collapsed to obtain a weight for each link according to its lifetime and its recentness. Finally, the Katz

measure, one of the best well-know link prediction metrics, is applied on the weighted graph obtained in order to quantify the strength of social ties between users.

We have assessed the efficiency of our framework through different evaluation techniques and comparisons with similar link prediction techniques applied on real traces. One of the most important findings is the successful adaptation of this framework to the context of mobile networks. In fact, we have proposed performing prediction in a distributed way: each node performs prediction relying on its local information. Such a result is an important contribution which provide the opportunity to implement it in a real context.

7.1.2 Improving Link Prediction Efficiency by Considering Link and Proximity Stabilities

We have wondered if it is possible to improve the performance of our framework and if we can consider a new criterion that characterizes the interactions between nodes. We have derived a heuristic: if a stable contact or proximity (at two hops) is observed during a period of time, this relationship is expected to be persistent in the future. Hence, we have proposed to take into consideration the stability of the relationship between each pair of nodes and to combine such an information with the one provided by our tensor-based framework. Such combination is expected to improve link prediction efficiency. To compute this stability, we have opted for the entropy estimator used in the Lempel-Ziv data compression, as it estimates well the real value of the entropy of a time series, and we have adapted it to the sequences of relationship stored by the third-order tensor. We have designed stability quantifications for links and for proximities at two hops.

From this perspective, we have proposed to combine the stability measures with the tensor-based link prediction framework outputs in order to provide new link prediction metrics. Our aim has been to show that taking into account the stability of relationship can be beneficial to fulfil our aim of improving the efficiency of our link prediction framework. The evaluation led to verify the effectiveness of our intuition has shown that considering the stability estimations enable us to improve the efficiency of our framework.

7.1.3 Other Works Towards the Improvement of the Evaluation and the Design of Communication Protocols in Human-Centered Wireless Networks

As we have stated, the major contribution of this thesis has been to provide efficient link prediction tools for human-centered wireless networks. Nevertheless, we have

been concerned by providing mechanisms or metrics in order to enhance the evaluation of communication protocols in such networks or to propose new mechanisms that are able to overcome some constraints imposed by these networks.

Firstly, we have designed a joint model for the IEEE 802.15.4 physical and medium access control layers. This joint model enables us to better estimate the communication performance than the models that are limited to the medium access control layer. This joint model has been proposed for the initiative of Smart Grid to provide tools that evaluate wireless communications standards.

Secondly, we have been interested in the non-cooperative MANETs. In such networks, the nodes can behave selfishly and reject any forwarding request which may fall down the network performance. To overcome this constraint, we have proposed a self-learning repeated game framework inspired by "The Weakest Link" TV game. Based on collecting gains by forming longest chains as possible between each pair of source-destination and on punishment threats which highly penalize misbehaving nodes, we have shown that our scheme outperforms two other self-learning repeated game algorithms presented in the literature.

7.2 Future Research Directions

We present in this section perspectives which can be pursued on the contributions of this thesis.

7.2.1 Implementing the Tensor-based Link Prediction Framework on Real Testbeds

We have stressed that the tensor-based link prediction framework is perfectly adapted to the mobile wireless networks. We have highlighted the fact that the link prediction can be made in a distributed way (using local information instead of global information). It will be challenging to implement this framework on real devices for two reasons.

First, it will be interesting to validate our framework in a real scenario, to measure the additional energy consumption due to recording data and computing Katz scores and to study how the threshold value which determines the decision relative to a score evolves to make better predictions. The threshold value indicates whether the Katz score corresponds to a positive prediction (which means that the score has higher value than the threshold) or a negative prediction (the opposite case).

Second, it is important to prove that the link prediction feedback is really help-

ful to enhance the performance of communication protocols. It will be attractive to design or modify a communication protocol to make it sensitive to such a feedback. This issue seems to be very promising especially for opportunistic networks where identifying suitable contact occurrences may lead to take better decisions to ensure better delivery message for any source-destination pair. It will be also useful to determine the frequency of refreshing the information and computing the link prediction scores for a communication protocol that ensures the tradeoff between energy consumption and delivery rate.

It is also crucial to focus on understanding how new links are created (links that are not observed but keenly expected to occur). Some insights are provided in [103] about characteristics of new links in mobile phone networks. It looks promising to exploit these findings, confirm and extend them to improve the identification of such links.

7.2.2 Investigating the Contributions of Other Expressions for the Stability Measure

To improve the efficiency of our tensor-based link prediction framework, we have proposed to quantify link and proximity stabilities using an adapted entropy estimator of a time series. Investigating other expressions of entropy and study their efficiency can be an attractive perspective. It is possible to design this entropy relying on computing some probabilities. For example, if we consider the evolution of a status of a link, it is possible to compute the entropy based on the probability of having a link status change (going from 0 to 1 or inversely).

We have also pointed out that we have aimed to show that the entropy estimation can be useful to obtain more accurate link prediction results without seeking if our proposed entropy-based metrics provide the best performance. It is challenging to study and to test other combinations between current/new entropy estimations and the outputs of the tensor-based link prediction framework and to assess if it is possible to find the best combination which ensures the highest link prediction performance.

In Chapter 4, we have pointed out the contribution of the proximity stability estimation in enhancing the prediction of new links. We think that working on the design of entropy-based metrics which can provide better prediction performance is promising as we have based our intuition on the findings of Song et al. [103].

7.2.3 Exploiting the Joint Model for IEEE 802.15.4 PHY and MAC Layers for Proposing New Performance Metrics

In the previous subsection, we have proposed to improve the performance of our proposed link prediction framework by quantifying the stability of interactions between nodes. This contribution has also been motivated by our work [118]. In this technical report, we have expressed our motivation to quantify spatiotemporal stability for links and consequently for nodes. Our motivation follows our concern to provide a new metric which is able to improve routing decisions in wireless networks. The metric that we want to express considers both the social relationship between the nodes and the link quality. We have aimed, for the evaluation of entropy-based metrics, to design a comprehensive model which allows us to make accurate simulations. We have opted for the IEEE 802.15.4 to start our investigations and we have proposed, taking into account our willingness to improve evaluation of wireless networks, the joint model of PHY and MAC layers. To extend this model, we have also thought to combine this joint model to the interference model proposed by Qiu et al. [90] to reproduce closest conditions to the reality. Moreover, we have opted for the Time-Variant Community mobility model (TVC model) [57] to create a dynamic topology.

Once efficient metrics are identified, we plan to verify the effectiveness of our findings on real testbeds and using the biased link sampling [121] and the algorithm "Learn On Fly" [122]. The works [121, 122] have proven the efficiency of data-driven link estimation, using biased link sampling, to converge to an efficient routing solution, even with dynamic environment parameters (changing traffic, evolving topologies, ...).

7.2.4 Enhancing the Self-Learning Repeated Game

It would be challenging to test the Weakest Link framework on real testbeds. From this perspective, we can be motivated by enhancing this framework by considering the different channel characteristics among nodes and relaxing the assumption that a node is able to distinguish between a packet drop and a packet loss. We can also take into consideration the residual energy for each node as a parameter in the utility function. All these perspectives would be helpful to design a real cooperation enforcement framework for multi-hop ad-hoc networks.

Appendix A

Thesis Publications

Journal

- Mohamed-Haykel Zayani, Vincent Gauthier, Ines Slama and Djamel Zeghlache, "Tensor-Based Link Prediction in Intermittently Connected Wireless Networks," *Submitted to: The International Journal of Computer and Telecommunications Networking (COMNET)*, August 2011.

International Conferences

- Mohamed-Haykel Zayani, Vincent Gauthier and Djamel Zeghlache, "A Joint-Model for IEEE 802.15.4 Physical and Medium Access Control Layers," *in Proceedings of the 7th International Wireless Communications and Mobile Computing Conference (IWCMC)*, Istanbul, Turkey, 2011.
- Mohamed-Haykel Zayani and Djamel Zeghlache, "Cooperation Enforcement for Packet Forwarding Optimization in Multi-hop Ad-hoc Networks," *in Proceedings of the IEEE Wireless Communication and Networking Conference (WCNC)*, Paris, France, 2012.
- Mohamed-Haykel Zayani, Vincent Gauthier, Ines Slama and Djamel Zeghlache, "Tracking Topology Dynamicity for Link Prediction in Intermittently Connected Wireless Networks," *in Proceedings of the 8th International Wireless Communications and Mobile Computing Conference (IWCMC)*, Limassol, Cyprus, 2012.
- Mohamed-Haykel Zayani, Vincent Gauthier and Djamel Zeghlache, "Improving Link Prediction in Intermittently Connected Wireless Networks by Considering Link and Proximity Stabilities," *in Proceedings of the 13th IEEE Inter-*

Appendix A. Thesis Publications

national Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), San Francisco, United-States of America, 2012.

Technical Report

- Mohamed-Haykel Zayani, Vincent Gauthier and Djamel Zeghlache, "Quantifying Spatiotemporal Stability by means of Entropy : Approach and Motivations," Technical Report RS2M10001, 2010.

Appendix B

Résumé en Français

B.1 Introduction

Les réseaux sans-fil mobiles et particulièrement ceux qui sont centrés sur l'être humain ont connu une expansion importante ces dernières années. Ces réseaux sont dépourvus d'infrastructure et permettent un déploiement dans des environnements où il est difficile voire impossible d'installer une infrastructure centralisée. Dans cette catégorie de réseaux, nous trouvons les réseaux ad-hoc mobiles MANETs [33] et les réseaux tolérants aux délais DTNs [43, 42].

Ces réseaux peuvent être sujets à des contraintes multiples. Nous pouvons citer les ressources limitées des dispositifs ou des nœuds. En outre, ces réseaux sont connus pour avoir des topologies très dynamiques à cause notamment de la mobilité humaine. En effet, il a été démontré dans [29, 59] que la mobilité humaine a un impact important sur le réseau et par conséquent sur sa performance. Ainsi, il s'avère impératif d'étudier et de maîtriser les propriétés de la mobilité humaine afin d'aspirer à concevoir des mécanismes de communications exploitant cette mobilité et à évaluer plus efficacement ces protocoles de communication.

Dans la littérature, nous pouvons trouver un ensemble pertinent de propriétés caractérisant la mobilité humaine: lieux de visite préférés [57], mobilité dépendant du moment de la journée [57] et régularité spatio-temporelle [16, 52, 106, 39, 77]. Nous avons été particulièrement intéressé par une propriété mettant en exergue l'aspect social. En effet, certains travaux à l'instar de [39, 78, 105] ont démontré que la mobilité humaine est étroitement dépendante des interactions sociales entre les entités du réseau et que celle-ci est régie par des intentions à caractère social. C'est ainsi que plusieurs travaux tels que [35, 59, 110] se sont basés sur des concepts, métriques et techniques inspirés de l'analyse des réseaux sociaux connue sous l'abréviation SNA (Social Network Analysis) pour proposer de nouveaux protocoles

de communications ou des mécanismes capables de prédire les liens dans un réseau social.

L'intérêt prononcé pour la SNA ainsi que les interactions sociales qui caractérisent les réseaux centrés sur l'aspect humain nous ont motivés pour proposer un mécanisme capable de prédire les liens dans ces réseaux et qui soit adapté à l'absence d'infrastructure centralisée. Il s'agit de la principale motivation de cette thèse qui sera détaillée dans la section B.2. Nous avons également voulu continuer dans l'optique de proposer des métriques et des techniques capables d'améliorer l'évaluation et la performance des protocoles de communication dans des contextes bien précis. Nous livrons quelques détails sur ces travaux dans la section B.3. Ce résumé sera conclu par la section B.4 dans laquelle nous rappelons les contributions majeures de cette thèse et nous proposons de nouvelles perspectives de recherche.

B.2 Proposition de Méthode et de Métriques pour la Prédiction des liens dans les Réseaux Mobiles Sans-Fil Centrés sur l'Aspect Humain

B.2.1 Mobilité Humaine, Analyse des Réseaux Sociaux et Prédiction des Liens

La mobilité humaine représente un paramètre crucial pour les réseaux sans-fil mobiles. Ainsi, il est primordial d'étudier cette mobilité afin de comprendre ses caractéristiques intrinsèques et exploiter ces propriétés. Cette étude est réalisée dans l'optique de concevoir et/ou évaluer les protocoles de communication dans ces réseaux.

Vers la Compréhension de la Mobilité Humaine

Pour reproduire la mobilité humaine, deux modèles individuels ont été proposés: le Random Walk [41] et le Random Waypoint [62]. Plusieurs autres modèles, décrits notamment dans [26, 79], s'en sont inspirés. Néanmoins, ces modèles restent peu fiables quant à la reproduction fidèle de la mobilité humaine [29, 79]. Par souci d'exhaustivité, plusieurs chercheurs ont procédé à la collecte d'informations sur la mobilité humaine dans des cas réels (campus universitaires, conférences, ...) et notamment les rencontres entre les utilisateurs du réseau. Dès lors, il est possible d'extraire des propriétés de cette mobilité humaine et de mettre en oeuvre des simulations qui se rapprochent indéniablement des cas réels.

Deux catégories de traces ont été considérées. D'une part, nous trouvons les traces dont les rencontres sont déduites à partir des correspondances avec les mêmes

B.2. Proposition de Méthode et de Métriques pour la Prédiction des liens dans les Réseaux Mobiles Sans-Fil Centrés sur l'Aspect Humain

points d'accès [5, 16, 52, 106, 39, 66]. D'autre part, d'autres initiatives se sont fixées l'objectif de recueillir les rencontres directes entre les dispositifs mobiles pour fournir un traçage plus fidèle des interactions entre les noeuds. Ces deux catégories de traces ont été largement utilisées pour extraire des caractéristiques intrinsèques à la mobilité humaine dont nous citons les plus importantes:

- Préférences prononcées pour certains lieux [57].
- Comportement dépendant du temps [57].
- Régularité spatio-temporelle [16, 52, 106, 39, 77].
- Distribution en loi de puissance des temps d'inter-contact entre les noeuds [58, 29, 63].

Outre ces propriétés, certains travaux tels que [39, 78, 105] ont démontré que les interactions sociales entre les utilisateurs du réseau ont un impact sur leur mobilité. Dès lors, ces interactions régissent la structure du réseau. Pour exploiter l'aspect social, extraire des informations structurelles du réseau et concevoir de nouveaux protocoles de communications, plusieurs techniques et métriques ont été empruntées de l'analyse des réseaux sociaux (Social Network Analysis ou SNA) [111].

L'Analyse des Réseaux Sociaux et la Prédiction des Liens

Etant donné que les réseaux sans-fil mobiles sont dépourvus d'infrastructures et que l'aspect opportuniste des communications peut prévaloir, l'analyse des réseaux sociaux a été considérée comme une autre technique de mesure pour les réseaux (les techniques de mesure traditionnelles se penchent sur des mesures de performance telles que les délais et la probabilité de livraison des messages). Ainsi, l'analyse des réseaux sociaux décortique les interactions entre les noeuds dans le réseau pour en extraire des propriétés structurelles (évolution de la topologie, robustesse du réseau, les noeuds centraux, les communautés émergentes, ...). Cet outil a été largement utilisé pour fournir des protocoles de communication s'inspirant de l'aspect social telles que les approches proposées dans [35, 59, 56].

L'analyse des réseaux sociaux a également inspiré Liben-Nowell et Kleinberg [74] pour introduire le problème de prédiction de liens dans les réseaux sociaux. En effet, ils se sont posés la question suivante: étant donné une capture d'un réseau social à un instant t , comment pourrait-on prédire, d'une manière sûre, les liens qui vont s'ajouter au réseau durant l'intervalle de temps $[t' - t]$ où t' représente un instant futur?

Appendix B. Résumé en Français

Pour répondre à cette question, Liben-Nowell et Kleinberg soulignent que les propriétés intrinsèques du réseaux sont capables de donner des indications sur la structure future. Pour argumenter cette hypothèse, ils se sont proposés de prédire les liens dans des réseaux sociaux (réseaux d’auteurs en l’occurrence tels que arXiv [4] et DBLP [6]) à l’aide de métriques se basant sur les voisins communs et sur la structure des liens entre chaque paire de noeuds. À travers leurs simulations, Liben-Nowell et Kleinberg ont démontré que leur hypothèse est fondée. Dans le même esprit, plusieurs travaux ont mis en exergue la pertinence de certaines métriques pour appliquer la prédiction dans les réseaux sociaux à l’instar de [8, 15, 38, 70, 110, 98]. En plus, certains parmi eux ont démontré la capacité de la mesure de Katz [65] à atteindre régulièrement des performances meilleures que les autres métriques présentées dans la littérature.

Inspirés par l’application de l’analyse des réseaux sociaux pour prédire les liens dans ces réseaux, nous nous sommes proposés d’exploiter les interactions qui peuvent exister dans les réseaux mobiles sans-fil comme ils sont bâtis autour d’utilisateurs humains. Dans la section suivante, nous détaillons notre méthode de prédiction de liens basée sur les tenseurs. Cette méthode sera utilisée pour analyser les interactions entre les entités du réseaux, quantifier la force de leurs liens sociaux et exploiter ces mesures pour prédire l’occurrence des liens radios dans le futur.

B.2.2 Méthode de Prédiction des Liens Basée sur les Tenseurs pour les Réseaux Sans-Fil Mobiles

Introduction

Durant les dernières années, plusieurs travaux de recherche se sont intéressés aux défis et challenges posés par les réseaux mobiles sans-fil. Dans ces réseaux, les contacts entre les noeuds ainsi que leurs propriétés s’imposent comme des paramètres clés à prendre en compte pour les protocoles de communication [56]. En effet, les interactions sociales entre les utilisateurs du réseau influent sur leur mobilité et ainsi sur la structure du réseau [57, 105]. Dès lors, il s’avère important de suivre et de comprendre les comportements des utilisateurs pour fournir une base de connaissances profitable aux protocoles de communication. C’est dans cette optique que nous proposons de fournir une méthode qui enregistre des données topologiques du réseau durant un temps de suivi de T périodes, qui extrait à partir de ces données la force des relations sociales entre les différentes entités et utilise ces mesures pour la prédiction des liens pour la période $T + 1$. Nous détaillons dans ce qui suit la méthode de prédiction des liens basée sur les tenseurs que nous proposons.

B.2. Proposition de Méthode et de Métriques pour la Prédiction des liens dans les Réseaux Mobiles Sans-Fil Centrés sur l'Aspect Humain

Description de la Méthode de Prédiction des Liens Basée sur les Tenseurs

La méthode de prédiction des liens que nous proposons s'articule autour d'étapes caractéristiques, comme le présente la figure B.1. Si nous considérons un réseau à topologie variable, la première étape consiste à diviser le temps de suivi de l'évolution de la topologie du réseau en T périodes (dans l'exemple de la figure B.1, $T=4$). À chaque période, la méthode recense tous les contacts dans une matrice d'adjacence (pour une paire de noeuds (i, j) , si un contact a eu lieu entre ces deux noeuds, la matrice d'adjacence assignera la valeur 1 à cette paire et 0 si aucun contact n'est détecté). L'ensemble de ces matrices d'adjacence forme le tenseur \mathcal{Z} . Ainsi, la matrice \mathbf{Z}_p du tenseur \mathcal{Z} traduit les contacts relevés dans le réseau durant la période p et la valeur $\mathbf{Z}_p(i, j)$ indique s'il y a eu contact entre les noeuds i et j durant cette période p . Suite à la phase de collecte des données, il s'agit de compresser toutes ces informations dans une matrice appelée tenseur comprimé pondéré et notée \mathbf{X} . La matrice \mathbf{X} va affecter à chaque lien (i, j) un coefficient relatif à sa récence et à sa fréquence. Pour obtenir les valeurs de la matrice \mathbf{X} , nous appliquons la formule suivante:

$$\mathbf{X}(i, j) = \sum_{p=1}^T (1 - \theta)^{T-p} \mathbf{Z}_p(i, j) \quad (\text{B.1})$$

où \mathbf{X} désigne le tenseur comprimé pondéré du tenseur \mathcal{Z} et θ est un paramètre utilisé pour ajuster le poids relatif à la récence et est compris entre 0 et 1.

Finalement, pour mesurer l'intensité des relations sociales entre les noeuds, nous appliquons la formule de Katz [65] pour obtenir la matrice des scores \mathbf{S} . Dans la littérature, la mesure de Katz a été utilisée pour mesurer la proximité dans le réseau d'une paire de noeuds (i, j) [74, 8, 38, 103] et a été employée pour prédire l'occurrence de liens dans le réseau en se basant notamment sur la proximité (qui exprime une certaine similarité dans la mobilité). Cette mesure sociométrique, pour une paire de noeuds (i, j) , s'exprime par la somme pondérée du nombre de chemins qui connectent le noeud i au noeud j . Ainsi, le score de Katz $\mathbf{S}(i, j)$ pour la paire de noeuds (i, j) est donné par l'équation suivante:

$$\mathbf{S}(i, j) = \sum_{\ell=1}^{+\infty} \beta^\ell |P_\ell(v_i, v_j)| \quad (\text{B.2})$$

où β est un paramètre défini par l'utilisateur, appelé facteur d'amortissement et compris entre 0 et 1 (généralement, ce paramètre est proche de 0 par souci de convergence). L'expression β^ℓ définit le poids d'un chemin de longueur ℓ sauts. Le paramètre $|P_\ell(v_i, v_j)|$ désigne le nombre de chemins, ayant un nombre de sauts ℓ ,

Appendix B. Résumé en Français

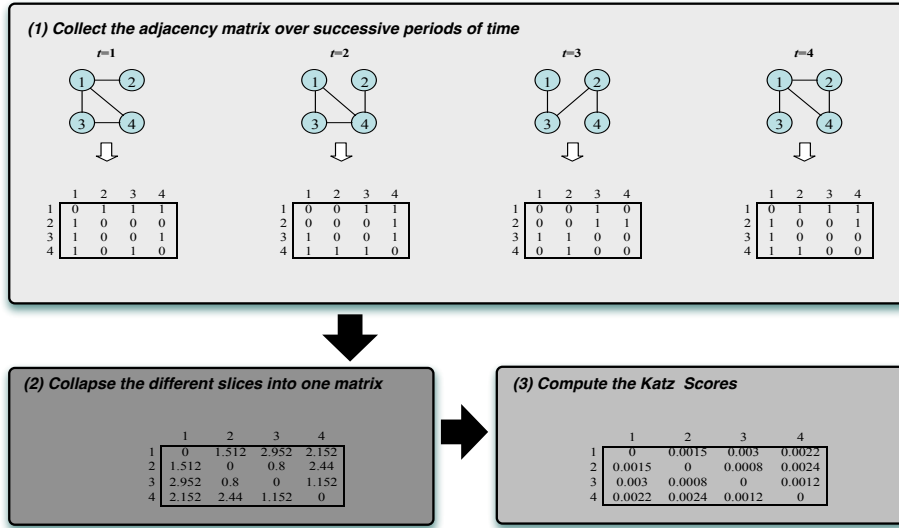


Figure B.1: Exemple de calcul de la matrice \mathbf{S}

connecte v_i à v_j .

Ainsi, plus un chemin connectant deux noeuds est long et plus son poids est faible.

Il est également possible d'appliquer la mesure de Katz sur une matrice d'adjacence. Nous proposons de calculer la matrice \mathbf{S} à partir du tenseur comprimé pondéré \mathbf{X} à l'aide de l'expression suivante:

$$\mathbf{S} = \sum_{\ell=1}^{+\infty} \beta^\ell \cdot \mathbf{X}^\ell = (\mathbf{I} - \beta \cdot \mathbf{X})^{-1} - \mathbf{I} \quad (\text{B.3})$$

où \mathbf{I} représente la matrice identité.

La détermination de la matrice \mathbf{S} permet de quantifier l'intensité des relations sociales entre chaque paire de noeuds (i, j) . Comme définie précédemment, la mesure de Katz privilégie les chemins courts entre une paire de noeuds. Ainsi, plus les chemins sont courts et nombreux, plus le score de Katz est élevé ce qui logiquement exprime une proximité prononcée (un nombre de voisins communs assez important) et par suite d'importantes liaisons sociales et une similarité dans la mobilité. Dès lors, la matrice des scores de Katz \mathbf{S} donnera des indications sur la relation entre chaque paire de noeuds et par conséquent sur la possibilité d'observer un lien entre eux dans le futur.

Dans le cas où l'on suppose qu'il peut y avoir une entité centrale capable de scruter et de suivre la topologie de tout le réseau, le calcul de la matrice des scores \mathbf{S} peut se faire d'une manière **centralisée**. Cependant, dans le contexte de réseaux

B.2. Proposition de Méthode et de Métriques pour la Prédiction des liens dans les Réseaux Mobiles Sans-Fil Centrés sur l'Aspect Humain

sans-fil, mobiles et dépourvus d'infrastructure, il serait judicieux de penser à une alternative pour pouvoir prédire les liens. À partir de cette perspective, nous proposons d'appliquer la prédiction des liens d'une manière **distribuée**. En d'autres termes, nous concevons que chaque noeud prédise ses liens à l'aide de l'information locale sur son voisinage (voisins à un et deux sauts).

Evaluation des performances et Résultats de Simulation

Pour évaluer notre méthode de prédiction des liens et notamment l'approche distribuée de la prédiction, nous considérons trois traces réelles:

- **Première Trace: trace du campus de Dartmouth:** Nous choisissons la trace du 05/01/2006 [5] et nous considérons l'évolution du réseau de 8h à 18h. Nous construisons le tenseur en tenant en compte tous les contacts observés entre 8h et 15h (7 heures). Les rencontres se passant entre 15h et 18h sont considérés comme des évènements futurs. Le nombre total de noeuds détectés dans cette trace est 1018.
- **Seconde Trace: trace du campus de MIT:** Nous nous intéressons à la trace du 23/07/2002 [16] et nous retenons les mêmes considérations horaires que pour la trace précédente. Le nombre de noeuds relevé est 646.
- **Troisième Trace: trace d'Infocom 2006:** Nous considérons les rencontres ayant eu lieu le 25/04/2006 [95] de 8h à 18h entre les 98 noeuds impliqués dans la collecte de contacts (70 d'entre eux sont portés par des personnes). La construction du tenseur se base sur les événements se déroulant de 8h à 14h. Les événements entre 14h et 18h sont considérés comme des évènements futurs.

Pour chaque trace, nous divisons le temps de suivi en périodes et nous envisageons différentes longueurs pour ces périodes: 5, 10 et 30 minutes. Ce paramètre là va fixer la taille du tenseur. Ainsi, Pour la trace du campus de Dartmouth, nous aurons respectivement un nombre de périodes total T égal, respectivement, à 96,48 et 16 pour couvrir les 8 heures de suivi.

Pour analyser les performances de notre proposition, nous proposons dans un premier temps de représenter les courbes ROC (Receiver Operating Characteristic curve) [44] pour la prédiction des liens à la période $T + 1$. En considérant les résultats obtenus après l'application de la prédiction des liens sur la trace du campus de Dartmouth par exemple répertoriés dans la figure B.2 (les résultats sont similaires pour les autres traces), nous soulignons les remarques suivantes:

Appendix B. Résumé en Français

- La méthode de prédiction est efficace vu que les courbures des courbes ROC sont proches du coin supérieur gauche de la figure et qu'ils sont loin de la ligne représentative de la prédiction aléatoire (random guess).
- Le recours à des périodes courtes pour diviser le temps de suivi permet une meilleure appréciation de l'évolution de la topologie et meilleure précision pour effectuer les prédictions.
- L'application de la prédiction des liens, que ce soit pour le cas centralisé ou le cas distribué, atteint les mêmes performances ce qui indique que l'application de la prédictions des liens peut s'effectuer d'une manière distribuée et adaptée aux réseaux sans-fil mobiles.

Dans un second temps, nous évaluons notre approche à l'aide de métriques de performance: l'Area Under the ROC Curve notée AUC [44], le ratio des tops scores à la période $T + 1$ noté TSR, l'exactitude notée ACC [44] et la F-Measure [108]). Pour une évaluation exhaustive, nous comparons notre approche à d'autres métriques utilisées dans la littérature: Voisins Communs, Salton Index [94], Jaccard Index [61], Sørensen Index [104], Hub Promoted Index (HPI) [92], Hub Depressed Index (HDI), Adamic-Adar Index [10], Resource Allocation Index [126] et Rooted PageRank Index [102]. Nous utilisons également les poids obtenus à l'aide du tenseur comprimé pondéré \mathbf{X} pour la comparaison.

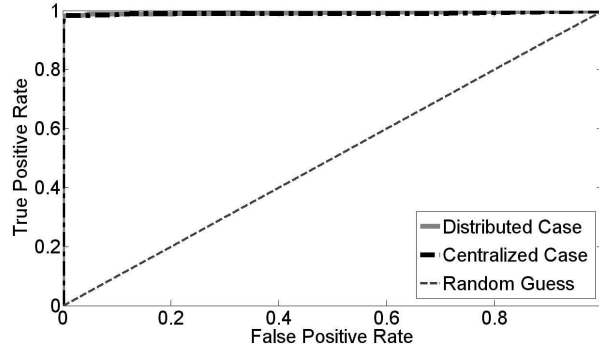
Dans cette évaluation, nous voudrions mettre exergue la capacité de chacune des méthodes à satisfaire deux tâches de prédictions. La première consiste à prédire les liens durant la période $T + 1$ à l'instar de l'évaluation avec les courbes ROC. La seconde s'intéresse à l'identification des nouveaux liens. Ce sont des liens potentiels qui n'ont pas été observés durant le temps de suivi et dont l'occurrence a été constatée durant la période considérée comme future.

Pour chacune des méthodes de prédiction, nous calculons la distance par rapport au cas parfait (AUC=1, TSR=100%, ACC=100% and F-Measure=1). Pour une approche a , cette distance est exprimée comme suit:

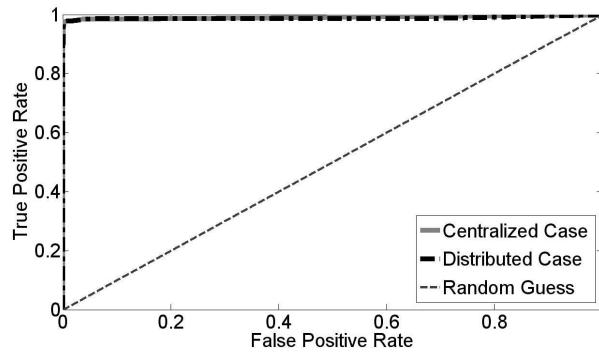
$$Distance(a) = \sqrt{(1 - AUC(a))^2 + (1 - TSR(a))^2 + (1 - ACC(a))^2 + (1 - F1(a))^2} \quad (B.4)$$

La table B.1 présente les résultats d'évaluation obtenus avec les métrique de performances et la comparaison de notre approche par rapport aux autres méthodes de prédiction. Des résultats similaires ont été obtenus pour les autres traces. Ces résultats confirment dans un premier temps les conclusions tirés à l'issue de l'évaluation avec les courbes ROC. Après la vérification des capacités de chacune des méthodes à prédire les liens durant la période $T + 1$ et ceux qui sont potentiels, il s'avère que

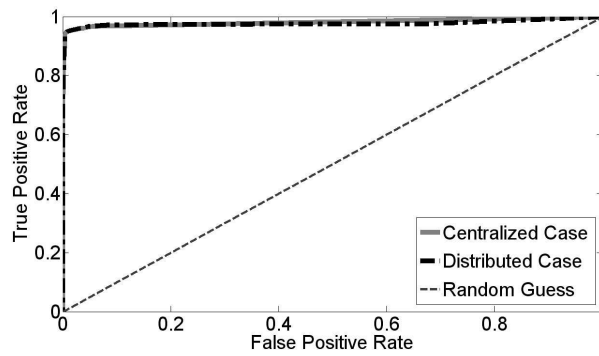
B.2. Proposition de Méthode et de Métriques pour la Prédiction des liens dans les Réseaux Mobiles Sans-Fil Centrés sur l'Aspect Humain



(a) Période du tenseur égale à 5 minutes



(b) Période du tenseur égale à 10 minutes



(c) Période du tenseur égale à 30 minutes

Figure B.2: Courbes ROC pour les différents cas de prédictions de liens appliqués sur la trace du campus de Dartmouth

notre méthode est la plus adéquate pour atteindre les meilleures performances pour les deux cas de prédictions et ceci est clairement souligné par le calcul de distance

Appendix B. Résumé en Français

Table B.1: Métriques d’Évaluation pour la Prédiction de tous les Liens Appliquée sur la Trace du Campus de Dartmouth

Prediction Cases		Metrics		All Links Prediction at $T+1$					New Links Prediction				
		AUC	TSR	ACC	F1	Dist.	AUC	TSR	ACC	F1	Dist.		
$t=5$ mins	Distributed Case	0.9932	93.70%	99.90%	0.9407	0.0868	0.6834	7.14%	99.81%	0.0063	1.3964		
	Centralized Case	0.9905	93.61%	99.90%	0.9416	0.0871	0.5920	7.04%	99.81%	0.0063	1.4206		
	Collapsed Weighted Tensor	0.9932	93.48%	99.90%	0.9428	0.0870	–	–	–	–	–		
	Common Neighbors	0.9911	83.48%	99.57%	0.6946	0.3474	0.6836	7.46%	99.81%	0.0063	1.3942		
	Salt on Index	0.9923	86.96%	99.77%	0.8692	0.1849	0.6834	5.46%	99.81%	0.0042	1.4091		
	Jaccard Index	0.9921	87.38%	99.78%	0.8741	0.1785	0.6830	5.67%	99.81%	0.0042	1.4078		
	Sørensen Index	0.9922	85.88%	99.76%	0.8591	0.1996	0.6834	5.67%	99.81%	0.0042	1.4077		
	Hub Prom. Index	0.9922	86.36%	99.77%	0.8612	0.1948	0.6834	5.57%	99.81%	0.0042	1.4084		
	Hub Depr. Index	0.9919	84.12%	99.73%	0.8404	0.2253	0.6832	5.25%	99.81%	0.0041	1.4107		
	Adamic-Adar Index	0.9923	87.88%	99.79%	0.8787	0.1717	0.6835	8.09%	99.81%	0.0021	1.3931		
	Resource Alloc. Index	0.9926	91.87%	99.85%	0.9176	0.1160	0.6837	7.14%	99.81%	0.0021	1.3993		
Rooted PageRank Index	0.9889	90.53%	99.81%	0.8864	0.1483	0.5638	3.99%	99.80%	0.0041	1.4505			
$t=10$ mins	Distributed Case	0.9915	90.26%	99.84%	0.9093	0.1334	0.6831	6.09%	99.81%	0.0063	1.4035		
	Centralized Case	0.9883	90.19%	99.84%	0.9091	0.1343	0.6097	6.30%	99.81%	0.0164	1.4134		
	Collapsed Weighted Tensor	0.9900	93.48%	99.84%	0.9032	0.1171	–	–	–	–	–		
	Common Neighbors	0.9890	80.41%	99.52%	0.6575	0.3947	0.6837	6.41%	99.81%	0.0063	1.4012		
	Salt on Index	0.9904	84.20%	99.73%	0.8427	0.2232	0.6835	5.36%	99.81%	0.0042	1.4098		
	Jaccard Index	0.9902	83.85%	99.71%	0.8349	0.2312	0.6830	5.46%	99.81%	0.0144	1.4020		
	Sørensen Index	0.9903	83.24%	99.72%	0.8351	0.2353	0.6836	5.46%	99.81%	0.0144	1.4019		
	Hub Prom. Index	0.9903	83.46%	99.73%	0.8317	0.2362	0.6835	5.04%	99.81%	0.0041	1.4120		
	Hub Depr. Index	0.9899	81.17%	99.69%	0.8144	0.2646	0.6834	4.73%	99.81%	0.0041	1.4141		
	Adamic-Adar Index	0.9904	83.87%	99.71%	0.8377	0.2290	0.6838	6.30%	99.81%	0.0063	1.4019		
	Resource Alloc. Index	0.9908	88.34%	99.80%	0.8857	0.1636	0.6838	5.78%	99.81%	0.0063	1.4054		
Rooted PageRank Index	0.9871	87.67%	99.74%	0.8439	0.1994	0.5855	3.57%	99.80%	0.0020	1.4483			
$t=30$ mins	Distributed Case	0.9813	82.31%	99.69%	0.8261	0.2488	0.6860	5.57%	99.81%	0.0042	1.4078		
	Centralized Case	0.9764	82.56%	99.69%	0.8272	0.2467	0.6180	6.51%	99.81%	0.0042	1.4183		
	Collapsed Weighted Tensor	0.9742	81.49%	99.70%	0.8343	0.2498	–	–	–	–	–		
	Common Neighbors	0.9782	71.25%	99.52%	0.5450	0.5387	0.6871	6.83%	99.81%	0.0042	1.3991		
	Salt on Index	0.9799	76.00%	99.57%	0.7535	0.3447	0.6867	5.78%	99.81%	0.0021	1.4077		
	Jaccard Index	0.9796	75.12%	99.55%	0.7454	0.3566	0.6866	5.88%	99.81%	0.0021	1.4071		
	Sørensen Index	0.9797	75.47%	99.56%	0.7471	0.3529	0.6866	5.67%	99.81%	0.0021	1.4085		
	Hub Prom. Index	0.9798	76.02%	99.58%	0.7573	0.3418	0.6865	5.15%	99.81%	0.0021	1.4120		
	Hub Depr. Index	0.9793	73.89%	99.52%	0.7321	0.3747	0.6867	5.78%	99.81%	0.0021	1.4077		
	Adamic-Adar Index	0.9800	74.16%	99.52%	0.7374	0.3690	0.6874	6.83%	99.81%	0.0042	1.3991		
	Resource Alloc. Index	0.9806	80.32%	99.64%	0.7976	0.2830	0.6878	7.14%	99.81%	0.0021	1.3984		
Rooted PageRank Index	0.9750	78.24%	99.56%	0.7826	0.3086	0.6076	4.52%	99.80%	0.0021	1.4358			

par rapport aux cas parfait.

Conclusion

Nous avons proposé une méthode capable de prédire les liens dans les réseaux sans-fil mobile bâtis autour des communautés humaines. Cette méthode suit l’évolution de la topologie ou du voisinage et quantifie l’intensité des relations sociales entre les entités du réseaux. Une telle quantification, obtenue à l’aide de la mesure de Katz, donne des indications intéressantes quant à l’occurrence des liens dans le futur. Les résultats confirment que notre méthode est capable d’atteindre les meilleures performances.

Cependant, nous pensons que la méthode que nous proposons peut être améliorée

B.2. Proposition de Méthode et de Métriques pour la Prédiction des liens dans les Réseaux Mobiles Sans-Fil Centrés sur l'Aspect Humain

en considérant un autre critère qui caractérise les relations entre les entités du réseau. Ce critère est celui de la stabilité. Nous détaillons dans la section suivante comment la considération de la stabilité peut être profitable à l'amélioration de la prédiction des liens, comment la quantifier et comment l'utiliser avec la méthode de prédiction des liens basée sur les tenseurs.

B.2.3 Amélioration de la Méthode de Prédiction des Liens en Considérant les Stabilités de Lien et de Proximité

Introduction

Comme détaillé précédemment, nous avons proposé une méthode de prédiction des liens basée sur le suivi de la topologie du réseau durant un certain nombre de périodes et l'application de la mesure de Katz pour quantifier le degré de similarité des modèles de mobilités des entités du réseau. L'une des principales contributions apportées par notre approche est la possibilité d'adapter cette méthode à un contexte distribué, c'est-à-dire que les noeuds peuvent prédire leurs futurs liens en se basant sur l'information recueillie au niveau du voisinage à un et deux sauts. Cette contribution qui est très importante nous a encore incités à améliorer notre méthode. Pour ce faire, nous nous sommes intéressés à la stabilité des relations entre les entités du réseau en y voyant un indicateur crédible capable de nous mener vers des prédictions plus fiables.

Comment la Stabilité d'un Lien et la Stabilité de Proximité Peuvent Améliorer la Prédiction?

En ce qui concerne la stabilité d'un lien, nous nous sommes basés sur l'heuristique suivante: quand deux entités du réseaux ont des liens sociaux forts, l'apparition d'un lien entre eux est très plausible. En plus, si ce lien existe, il est attendu qu'il soit stable et persévérant. Ainsi, la stabilité d'un lien renforce la possibilité d'observer un tel lien dans le futur. Pour la stabilité de proximité (stabilité de la relation à deux sauts), celle-ci peut renseigner sur l'occurrence d'un lien dans le futur. Si la relation est stable, la mesure de Katz renseigne bien sur la présence du lien, sinon, il est clair que le lien est imprévisible et donc difficile à estimer sa présence dans le futur.

Dès lors, quantifier la stabilité peut s'avérer très intéressant pour améliorer les performances de notre méthode de prédiction basée sur les tenseurs. Nous nous sommes inspirés de la thermodynamique [21] et de la théorie de l'information [97] pour opter pour l'entropie comme métrique capable de quantifier la stabilité. Dans

Appendix B. Résumé en Français

ce qui suit, nous expliquons comment estimer la stabilité d'une relation entre deux noeuds.

Quantifier la Stabilité à l'Aide d'une Estimation de l'Entropie de Séries Chronologiques

Pour notre approche, nous nous sommes intéressés à l'estimateur d'entropie utilisé dans l'algorithme de compression de données de Lempel-Ziv [127, 67]. Cette métrique estime bien l'entropie d'une série chronologique. Pour une séquence de n étapes, l'entropie est estimée à l'aide de l'expression suivante:

$$S^{est} = \left(\frac{1}{n} \sum_i \Lambda_i \right)^{-1} \ln n \quad (\text{B.5})$$

où Λ_i représente la longueur du plus court mot à partir de la position i et qui n'a pas été observé précédemment dans la séquence (entre la position 1 et $i - 1$).

Cet estimateur a été exploité pour trouver le nombre de bits nécessaires pour coder une séquence de lettres alphabétiques. Dans notre cas, nous nous basons sur les états des liens fournis par la construction des tenseurs. Chaque noeud a ainsi deux séquences qui définissent sa relation avec chacun des noeuds qu'il a détecté: une séquence pour l'état du lien et une séquence pour l'état de proximité à deux sauts. Ces séquences sont des successions de uns et/ou de zéros relatives à la succession des périodes de suivi de la topologie. Dès lors, nous proposons l'estimateur d'entropie $E_T^l(i, j)$ qui quantifie la stabilité d'un lien entre les noeuds i and j sur T périodes. Il est donné par l'expression suivante:

$$E_T^l(i, j) = \left(\frac{1}{n} \sum_{t=1}^T \Lambda_t(\mathbf{Z}_t(i, j)) \right)^{-1} \ln n \quad (\text{B.6})$$

où $\Lambda_t(\mathbf{Z}_t(i, j))$ est la longueur du mot le plus court à partir de la période t et qui n'a pas été détecté entre les périodes 1 et $t - 1$.

En suivant la même approche, nous définissons l'estimateur d'entropie $E_T^p(i, j)$ qui quantifie la stabilité de proximité. Cet estimateur est calculé en appliquant l'Eq. B.6 tout en substituant la valeur $\mathbf{Z}_t(i, j)$ par l'état de proximité à deux sauts entre les noeuds i et j à la période t .

L'estimateur d'entropie utilisé scrute une séquence et identifie les mots les plus courts étape par étape. Si les mots les plus courts tendent à devenir longs, ceci explique qu'il y a redondance. La redondance coïncide dans ce cas avec stabilité puisque les mots longs limitent le nombre de mots trouvés et diminue par la suite l'estimation de l'entropie. D'autre part, si les mots identifiés prennent fréquemment

B.2. Proposition de Méthode et de Métriques pour la Prédiction des liens dans les Réseaux Mobiles Sans-Fil Centrés sur l'Aspect Humain

les combinaisons les plus courtes possibles, ceci indique plutôt une relation imprévisible. En effet, identifier des mots qui tendent à être parmi les plus courts possible augmente substantiellement le nombre de mots trouvés et accroît en conséquence l'estimation de l'entropie.

Pour tirer profit de ses mesures de stabilité, il serait judicieux de les exploiter simultanément avec la méthode de prédiction des liens basée sur les tenseurs. Nous proposons dans la suite différentes techniques de combinaisons entre les estimations d'entropie et les métriques fournies par notre méthode de prédiction afin de fournir de nouvelles métriques de prédiction des liens.

Jointure des Mesures de Stabilité avec le Méthode de Prédiction des Liens Basée sur les Tenseurs

À travers nos investigations sur la méthode de prédiction des liens, nous avons fourni des métriques caractérisant les liens entre les entités du réseau à travers leur récence et leur fréquence. Nous avons également mis en exergue le critère de stabilité des relations en proposant des métriques construites autour de l'estimation de l'entropie. Nous pensons fortement que ces deux catégories de métriques sont complémentaires et que leur combinaison permet de concevoir de nouvelles métriques de prédiction encore plus performantes. Nous présentons différents types de combinaisons qui nous ont permis de définir ces nouvelles métriques:

- Combiner le poids fourni par le tenseur comprimé pondéré $\mathbf{X}(i, j)$ avec la stabilité du lien $E_T^l(i, j)$ relatifs aux T périodes de suivi pour calculer le **XE score**.
- Combiner l'intensité des relations sociales exprimée par la mesure de Katz $\mathbf{S}(i, j)$ avec la stabilité du lien $E_T^l(i, j)$ relatifs aux T périodes de suivi pour calculer le **SE score**.
- Combiner le tenseur comprimé pondéré \mathbf{X} avec la matrice exprimant les stabilités des liens E_T^l relatifs à un suivi sur T périodes et appliquer la formulation de Katz sur cette combinaison pour calculer le **XES score**.
- Appliquer la mesure de Katz sur un nouveau tenseur comprimé pondéré \mathbf{X}_{new} pour calculer le **XNS score**. Plusieurs variantes peuvent être envisagées pour cette nouvelle métrique. En effet, le nouveau tenseur comprimé pondéré va prendre en considération, outre l'état du lien:
 - La stabilité du lien au niveau de chaque période (XNS1 score).

Appendix B. Résumé en Français

- La stabilité de proximité à deux sauts au niveau de chaque période (XNS2 score).
- Les deux mesures de stabilité ensemble (XNS3 score).

À travers ces nouvelles métriques, nous visons à démontrer que considérer le paramètre de stabilité dans les relations entre les entités du réseau participe activement à l'amélioration des performances de prédiction. Nous pratiquons des simulations afin de vérifier l'exactitude de notre intuition.

Evaluation des Performances et Analyse des Résultats

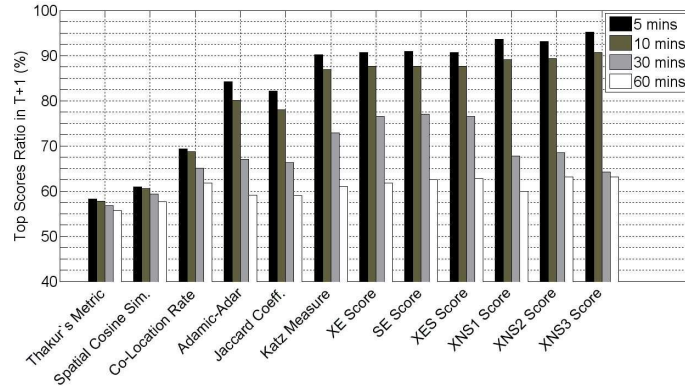
Nous considérons les mêmes traces détaillées dans la section B.2.2 et nous nous proposons de suivre l'évolution de la topologie du réseau pour toutes ces traces de 8h à 16h. Nous divisons ce temps de suivi sur des périodes de longueurs différentes: 5, 10, 30 et 60 minutes. Ainsi, il est nécessaire de considérer respectivement un nombre T de périodes égal à 96, 48, 16 et 8 périodes pour couvrir tout le temps de suivi et construire les tenseurs relatifs. Pour l'évaluation, nous avons recours à des métriques déjà utilisées dans la section B.2.2: le ratio des top scores à la période $T + 1$, l'exactitude des prédictions et la F-Measure. Nous présentons dans la figure B.3 les résultats d'évaluation des métriques basées sur l'entropie pour la prédiction des liens à la période $T + 1$ pour la trace relative au campus de MIT. Les performances de nos propositions sont comparées avec celles de/d'une/des:

- Notre méthode de prédiction basée sur les tenseurs et le calcul de la mesure de Katz.
- Métriques de prédiction de liens exprimant la proximité des noeuds qui sont les mesures d'Adamic-Adar [10] et le coefficient de Jaccard [61, 94].
- Métriques de prédiction des liens se basant sur l'homophilie de la mobilité qui sont le spatial cosine similarity et le co-location rate [110].
- Métrique exprimant la similarité des lieux visités par les entités du réseau [105].

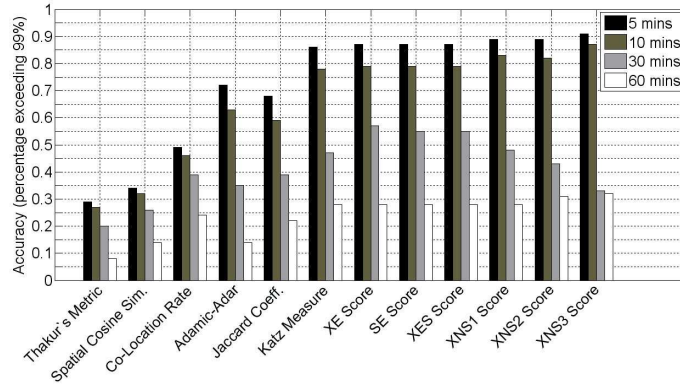
Il est à noter que les résultats obtenus pour les autres traces sont similaires. Les résultats obtenus nous permettent de tirer deux grandes conclusions:

- La comparaison de la performance de la méthode de prédiction des liens basées sur les tenseurs avec les autres techniques présentées dans la littérature renforce notre conviction sur le choix d'utilisation de la mesure de Katz.

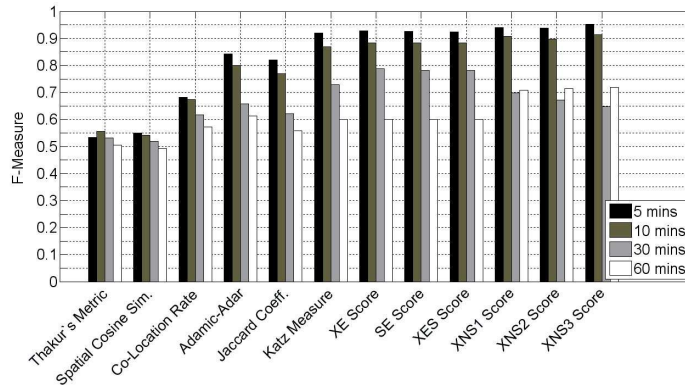
B.2. Proposition de Méthode et de Métriques pour la Prédiction des liens dans les Réseaux Mobiles Sans-Fil Centrés sur l'Aspect Humain



(a) Ratio des Top Scores à la Période $T+1$



(b) Exactitude (pourcentage dépassant 99%)



(c) F-Measure

Figure B.3: Métriques d'Evaluation pour Différentes Techniques de Prédiction Appliquées sur la trace du Campus de MIT

- La comparaison des performances des métriques basées sur l'estimation de la stabilité des relation avec celle de notre méthode souligne le fondement de notre

Appendix B. Résumé en Français

Table B.2: Paramètres de Simulation pour la Vérification de l'Efficacité de la Métrique SE_{2hops} score pour l'Amélioration de la Prédiction des Nouveaux Liens

Trace	Temps de Suivi	Temps durant lequel les contacts sont considérés comme futurs
Campus de MIT (23/07/02)	de 8h à 14h	de 14h à 18h
Campus de Dartmouth (05/01/06)	de 8h à 14h	de 14h à 18h
Infocom 2006 (25/04/06)	de 8h à 13h30	de 13h30 à 18h

intuition et de l'apport indéniable de la considération du critère de stabilité pour améliorer la qualité de la prédiction (au pire des cas, trois métriques sur six basées sur l'estimation d'entropie permettent d'avoir des performances meilleures que celle de notre méthode)

Nous nous sommes également intéressés au potentiel de ces mesures de stabilité pour améliorer la prédiction des nouveaux liens. Motivés par les conclusions avancées dans [110] qui stipulent que la majorité des nouveaux liens se crée entre deux noeuds qui sont régulièrement séparés par deux sauts lors du temps de suivi. Nous nous sommes proposés de fournir une métrique que nous avons appelée SE_{2hops} score. Il s'agit de combiner la mesure de similarité $\mathbf{S}(i, j)$ avec l'estimation de la stabilité de proximité $E_T^p(i, j)$. En prenant en considération les paramètres de simulation résumés dans la table B.2 et deux longueurs différentes pour les périodes de suivi (5 et 10 minutes), nous observons que la métrique SE_{2hops} score est capable d'avoir de meilleures performances par rapport à la méthode de prédiction utilisant la mesure de Katz pour les scénarios relatifs aux traces du campus de MIT et d'Infocom 2006 ce qui indique que le recours à la mesure de stabilité peut également renforcer la qualité de la prédiction des nouveaux liens et que l'investigation sur de futures améliorations s'impose comme une piste de recherche sérieuse.

Conclusion

Nous avons présenté dans cette première partie une méthode capable de prédire les liens entre les entités d'un réseau sans-fil mobile caractérisé par des interactions humaines. Cette méthode s'appuie sur l'identification des relations sociales entre ces entités et exploiter ces connaissances pour estimer la topologie du réseau dans le futur. Cette méthode utilise la mesure de Katz qui s'est avérée être l'une des plus puissantes métriques en terme de prédiction des liens dans des réseaux à caractère social. Nos simulations sur des traces réelles nous ont permis entre autres de démontrer que le choix de cette métrique s'impose. La contribution majeure

B.3. Autres Contributions Avancées dans la Thèse

Table B.3: Ratio des Top Scores dans le Futur: Comparaison entre la performance de la métrique SE_{2hops} score et celle de la mesure de Katz dans la prédiction des nouveaux liens

Trace	Longueur de la Période de Suivi t	Métrique de Prédiction	TSR dans le futur
MIT Campus	5 minutes	Katz Measure	13,49%
		SE_{2hops} Score	14,09%
	10 minutes	Mesure de Katz	13,22%
		SE_{2hops} Score	13,44%
Dartmouth Campus	5 minutes	Mesure de Katz	9,10%
		SE_{2hops} Score	8,36%
	10 minutes	Mesure de Katz	9,28%
		SE_{2hops} Score	8,55%
Infocom 2006	5 minutes	Mesure de Katz	16,96%
		SE_{2hops} Score	18,75%
	10 minutes	Mesure de Katz	9,82%
		SE_{2hops} Score	9,82%

apportée par la proposition d'une telle méthode est sa possible application d'une manière distribuée. Ainsi, les noeuds eux-mêmes peuvent prédire leurs propres liens et l'information locale (relative au voisinage) est suffisante pour réaliser cette tâche.

Nous avons également poursuivi nos recherches sur de possibles améliorations quant à la qualité des prédictions. En effet, nous nous sommes basés sur deux critères caractérisant les relations entre les entités du réseau: la récence et la fréquence des interactions. Nous nous sommes demandés s'il serait possible de considérer un troisième critère capable d'apporter plus de précision aux prédictions et nous nous sommes intéressés à la stabilité des relations (lien ou proximité à deux sauts) pour déceler encore mieux les liens susceptibles d'être observés dans le futur. Nous avons proposé de multiples combinaisons entre les mesures de stabilités et les métriques fournies par notre méthode de prédiction et les résultats de simulation nous ont clairement permis de conclure que la considération de ce critère de stabilité est judicieux vues les améliorations notables apportées à la qualité des prédictions.

B.3 Autres Contributions Avancées dans la Thèse

Outre la proposition de techniques et métriques capables de fournir une information utile aux protocoles de communication, nous nous sommes également penchés sur des travaux dans l'optique d'améliorer l'évaluation et la conception des protocoles de communication dans des réseaux sans-fil où la mobilité peut être considérée.

D'une part, nous nous sommes intéressés à la conception d'un modèle commun pour les couches physique et liaison de données du standard IEEE 802.15.4 [2]. Il s'agit d'un modèle qui joint un sous-modèle de couche physique inspiré des travaux de Zuniga et Krishnamachari [128, 116] (conception de modèles radio et canal) et un sous-modèle construit autour d'une chaîne de Markov [89, 87] et mimant le

Appendix B. Résumé en Français

comportement pouvant être observés au niveau de la couche liaison de données. Nous proposons ce modèle pour mieux estimer les indicateurs de performance car nous avons constaté que les modèles s'intéressant à ce standard se sont limités à une seule couche.

D'autre part, nous avons eu l'opportunité de nous focaliser sur les réseaux ad-hoc non coopératifs et nous avons proposé un mécanisme capable de stimuler la coopération dans de tels réseaux. En effet, nous nous sommes inspirés du célèbre jeu télévisé "Le Maillon Faible" pour inciter les nœuds à construire les plus longues chaînes sur chaque route entre une source et une destination pour améliorer les taux de livraison. Ce mécanisme s'appuie également sur l'apprentissage du niveau de coopération dans le réseau pour adapter les probabilités de forwarding et sur des punitions qui pénalisent fortement les nœuds à caractère égoïste au niveau de leur fonction d'utilité et les obligent à être plus coopératifs. Les résultats de simulations confirment que notre proposition est capable de surpasser quelques autres mécanismes d'apprentissage similaires proposés dans la littérature [51, 86].

B.4 Conclusion

Pour conclure ce résumé, nous rappelons les majeures contributions apportées dans cette thèse et nous proposons quelques perspectives futures. Dans cette thèse, nous nous sommes intéressés aux réseaux sans-fil mobiles centrés sur l'aspect humain. Nous nous sommes fixés l'objectif de proposer des métriques et des mécanismes capables d'améliorer l'évaluation et la conception des protocoles de communications.

D'abord, nous avons été motivés par l'étude et l'exploitation des relations sociales présentes dans ces réseaux. En s'appuyant sur la mesure sociométrique de Katz et une structure en tenseur capable de donner un aperçu sur la dynamique du réseau, nous avons été le premier à proposer une méthode distribuée de prédiction des liens pour des réseaux comme les DTNs et les MANETs. L'application de cette méthode sur des traces réelles et l'évaluation à travers des techniques dédiées n'a fait que confirmer l'efficacité de notre proposition.

Ensuite, nous avons continué à porter un intérêt prononcé pour la prédiction des liens dans ces réseaux. En effet, nous nous sommes demandés s'il est possible d'améliorer les performances de notre méthode. Cette dernière se base sur deux critères pour quantifier les étendues des relations sociales entre les entités du réseau: la récence et la fréquence des interactions. À partir de là, nous nous sommes demandés s'il serait possible de considérer un troisième critère qui participerait activement à l'amélioration de la qualité de la prédiction. En se basant sur une heuristique, nous avons choisi d'intégrer le critère de stabilité des relations dans les interactions.

En quantifiant ce critère à l'aide d'un estimateur d'entropie et en le combinant avec les métriques fournies par notre méthode, nous avons proposé de nouvelles métriques de prédiction. La pertinence de notre proposition a été confirmée par les résultats de simulation et le critère de stabilité s'avère judicieux pour la prédiction des liens.

Enfin, nous avons également eu l'opportunité d'apporter quelques contributions allant dans le sens de notre souci d'apporter des métriques et techniques capables d'améliorer l'évaluation et le design des protocoles de communication. Dans cette optique, nous avons proposé un modèle commun pour les couches physique et liaison de données du standard IEEE 802.15.4 afin d'estimer plus précisément les indicateurs de performances. En outre, nous nous sommes également penchés sur les réseaux ad-hoc non coopératifs et nous avons conçu un mécanisme visant à stimuler les noeuds du réseau à être plus coopératifs et s'inspirant du célèbre jeu télévisé "Le Maillon Faible".

En guise de perspectives, nous songeons à implémenter notre méthode de prédiction des liens dans des cas de déploiements réels et de la joindre à un protocole de communication pour s'assurer que l'information issue de la prédiction est profitable à l'amélioration des performances dans le réseau. Nous pensons également à fournir d'autres formulations et combinaisons pour prétendre à fournir des métriques plus efficaces. Il est à signaler que notre objectif a été de démontrer que l'estimation de la stabilité permet d'améliorer les performances de prédiction sans pour autant chercher quelle est la meilleure formulation possible. Il sera intéressant, d'autre part, de voir d'autres formulations d'entropie et vérifier si nos choix sont perfectibles. Il sera également motivant de mener des recherches plus approfondies pour améliorer la prédiction des nouveaux liens. Nous proposons d'essayer d'identifier de nouveaux indicateurs, outre que les observations faites dans [110], pour prédire plus efficacement l'apparition de ces liens.

Bibliography

- [1] 802.11a-1999 Supplement to IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Spec. 87
- [2] 802.15.4-2006 IEEE Standard for Information Technology- Telecommunications and Information Exchange Between Systems- Local and Metropolitan Area Networks- Specific Requirements Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Spec. 16, 87, 146
- [3] Community resource for archiving wireless data at dartmouth. <http://crawdad.cs.dartmouth.edu/>. 24
- [4] Cornell university library arxiv.org. <http://arxiv.org/>. 15, 30, 40, 133
- [5] CRAWDAD: A community resource for archiving wireless data at dartmouth. <http://crawdad.cs.dartmouth.edu/>. 14, 24, 25, 42, 75, 132, 136
- [6] The DBLP computer science bibliography. <http://www.informatik.uni-trier.de/~ley/db/>. 15, 30, 36, 40, 133
- [7] SGIP NIST Smart Grid Collaboration Site, PAP02: Wireless Communications for the Smart Grid (6.1.5), Call for Input to Task6, 2010. http://collaborate.nist.gov/twiki-sggrid/bin/view/SmartGrid/PAP02Wireless#Call_for_Input_to_Task_6. 88, 89, 100
- [8] Evrim Acar, Daniel M. Dunlavy, and Tamara G. Kolda. Link Prediction on Evolving Data Using Matrix and Tensor Factorizations. In *Proc. of the IEEE International Conference on Data Mining Workshops*, pages 262–269, December 2009. 14, 15, 30, 34, 36, 38, 39, 40, 55, 62, 64, 133, 134
- [9] Utku Günay Acer, Petros Drineas, and Alhussein A. Abouzeid. Random walks in time-graphs. In *Proc. of the Second International Workshop on Mobile Opportunistic Networking, (MobiOpp '10)*, pages 93–100, 2010. 35

-
- [10] Lada A. Adamic and Eytan Adar. Friends and neighbors on the Web. *Social Networks*, 25(3):211–230, July 2003. 29, 36, 49, 62, 79, 137, 143
- [11] Eitan Altman, Arzad Alam Kherani, Pietro Michiardi, and Refik Molva. Non-cooperative forwarding in ad-hoc networks. In Raouf Boutaba, Kevin C. Almeroth, Ramón Puigjaner, Sherman X. Shen, and James P. Black, editors, *NETWORKING*, volume 3462 of *Lecture Notes in Computer Science*, pages 486–498. Springer, 2005. 103, 113
- [12] Beongku An, Joo Sang Lee, and Nam-Soo Kim. An entropy-based routing protocol using mobility in mobile ad-hoc wireless sensor networks. In *Proceedings of the 11th international conference on Advanced Communication Technology - Volume 2*, ICACT'09, pages 949–952, Piscataway, NJ, USA, 2009. IEEE Press. 69
- [13] Beongku An and Symeon Papavassiliou. An entropy-based model for supporting and evaluating route stability in mobile ad hoc wireless networks. *IEEE Communications Letters*, 6(8):328–330, 2002. 69
- [14] Luzi Anderegg and Stephan Eidenbenz. Ad hoc-vcg: a truthful and cost-efficient routing protocol for mobile ad hoc networks with selfish agents. In *Proceedings of the 9th annual international conference on Mobile computing and networking*, MobiCom '03, pages 245–259, New York, NY, USA, 2003. ACM. 103
- [15] Lars Backstrom and Jure Leskovec. Supervised random walks: predicting and recommending links in social networks. In *Proceedings of the fourth ACM international conference on Web search and data mining*, WSDM '11, pages 635–644, New York, NY, USA, 2011. ACM. 15, 30, 36, 133
- [16] Magdalena Balazinska and Paul Castro. Characterizing mobility and network usage in a corporate wireless local-area network. In *Proc. of the 1st international conference on Mobile systems, applications and services, (MobiSys '03)*, pages 303–316, 2003. 14, 24, 25, 42, 75, 130, 132, 136
- [17] Walid Ben-Ameur. Plus de mathématiques pour gagner au “maillon faible”. *Mathématiques et Sciences Humaines = Mathematics and Social Sciences*, pages 27–41, 2006. 112
- [18] G. Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications*, 18(3):535–547, 2000. 87, 88

Bibliography

- [19] Ginestra Bianconi, Paolo Pin, and Matteo Marsili. Assessing the relevance of node features for network structure. *Proceedings of the National Academy of Sciences of the United States of America*, 106(28):11433–11438, July 2009. 70
- [20] Bela Bollobas. *Modern Graph Theory*. Springer, 1998. 15, 35, 61
- [21] Ludwig Boltzmann. Uber die beziehung zwischen dem zweiten hauptsatz der mechanischen warmetheorie und der wahrscheinlichkeitsrechnung respektive den satzen uber des warmegleichgewicht. *Wien. Ber.*, 76:373–435, 1877. 67, 140
- [22] Dan Braha and Yaneer Bar-Yam. From centrality to temporary fame: Dynamic centrality in complex networks: Research articles. *Complex.*, 12:59–63, November 2006. 60
- [23] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998. 29
- [24] Sonja Buchegger and Jean-Yves Le Boudec. Nodes bearing grudges: towards routing security, fairness, and robustness in mobile ad hoc networks. In *Proceedings of the 10th Euromicro conference on Parallel, distributed and network-based processing, EUROMICRO-PDP'02*, pages 403–410, Washington, DC, USA, 2002. IEEE Computer Society. 103
- [25] Sonja Buchegger and Jean-Yves Le Boudec. Performance analysis of the confidant protocol. In *Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing, MobiHoc '02*, pages 226–236, New York, NY, USA, 2002. ACM. 103
- [26] T. Camp, J. Boleng, and V. Davies. A Survey of Mobility Models for Ad Hoc Network Research. *Wireless Communications & Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, 2(5):483–502, 2002. 24, 131
- [27] Alberto Cerpa, Naim Busek, and Deborah Estrin. SCALE: A tool for Simple Connectivity Assessment in Lossy Environments. In *Tech. Rep. CENS*. CENS, ACM Press, 2003. 87, 88
- [28] Alberto Cerpa, Jennifer L. Wong, Miodrag Potkonjak, and Deborah Estrin. Temporal properties of low power wireless links: modeling and implications on multi-hop routing. In *Proceedings of the 6th ACM international symposium*

-
- on Mobile ad hoc networking and computing - MobiHoc '05*, page 414, New York, New York, USA, 2005. ACM Press. 87, 88
- [29] Augustin Chaintreau, Pan Hui, Jon Crowcroft, Christophe Diot, Richard Gass, and James Scott. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Trans. on Mobile Computing*, 6(6):606–620, June 2007. 13, 14, 15, 23, 24, 25, 34, 37, 43, 45, 60, 62, 130, 131, 132
- [30] Hua Chen, Baolin Sun, and Yue Zeng. Qos multicast routing algorithm in manet: an entropy-based ga. In *Proceedings of the 2006 international conference on Intelligent computing: Part II, ICIC'06*, pages 1279–1289, Berlin, Heidelberg, 2006. Springer-Verlag. 69
- [31] Fan R. K. Chung. *Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92)*. American Mathematical Society, 1997. 35, 61
- [32] Rudolph Clausius. Ueber die bewegende kraft der wärme und die gesetze, welche sich daraus für die wärmelehre selbst ableiten lassen. *Annalen der Physik*, 155(3):368–397, 1850. 67
- [33] S Corson and J Macker. Mobile ad hoc networking (manet): Routing protocol performance issues and evaluation considerations. *Computing Systems*, 54(2501):1–12, 1999. 12, 22, 130
- [34] Jon Crowcroft, Richard Gibbens, Frank Kelly, and Sven Östring. Modelling incentives for collaboration in mobile ad hoc networks. *Perform. Eval.*, 57:427–439, August 2004. 103
- [35] Elizabeth M Daly and Mads Haahr. Social network analysis for routing in disconnected delay-tolerant MANETs. In *Proc. of the 8th ACM international symposium on Mobile ad hoc networking and computing, (MobiHoc '07)*, pages 32–40, 2007. 14, 15, 27, 35, 61, 130, 132
- [36] F. Daneshgaran, M. Laddomada, F. Mesiti, and M. Mondin. Unsaturated Throughput Analysis of IEEE 802.11 in Presence of Non Ideal Transmission Channel and Capture Effects. *IEEE Transactions on Wireless Communications*, 7(4):1276–1286, April 2008. 87, 88
- [37] A. Doria, M. Uden, and D. P. Pandey. Providing Connectivity to the Saami Nomadic Community. In *Proc. 2nd Int'l. Conf. Open Collaborative Design for Sustainable Innovation (dyd 02)*, Bangalore, India, December 2002. Proc. 2nd. 23

Bibliography

- [38] Daniel M. Dunlavy, Tamara G. Kolda, and Evrim Acar. Temporal link prediction using matrix and tensor factorizations. *ACM Trans. Knowl. Discov. Data*, 5(2):10:1–10:27, February 2011. [14](#), [15](#), [30](#), [34](#), [36](#), [38](#), [39](#), [40](#), [62](#), [133](#), [134](#)
- [39] Nathan Eagle and Alex S. Pentland. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4):255–268, March 2006. [14](#), [15](#), [24](#), [25](#), [130](#), [132](#)
- [40] Stephan Eidenbenz, Giovanni Resta, and Paolo Santi. The commit protocol for truthful and cost-efficient routing in ad hoc networks with selfish nodes. *IEEE Transactions on Mobile Computing*, 7:19–33, January 2008. [103](#)
- [41] A. Einstein. Investigations on the theory of the brownian movement. *Ann. der Physik*, 1905. [24](#), [131](#)
- [42] Vijay Erramilli, Augustin Chaintreau, Mark Crovella, and Christophe Diot. Diversity of forwarding paths in pocket switched networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, IMC '07, pages 161–174, New York, NY, USA, 2007. ACM. [12](#), [23](#), [130](#)
- [43] Kevin Fall. A delay-tolerant network architecture for challenged internets. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, SIGCOMM '03, pages 27–34, New York, NY, USA, 2003. ACM. [12](#), [23](#), [130](#)
- [44] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006. [43](#), [45](#), [49](#), [136](#), [137](#)
- [45] Mark Felegyhazi, Jean-Pierre Hubaux, and Levente Buttyan. Equilibrium analysis of packet forwarding strategies in wireless ad hoc networks - the dynamic case. In *PWC 2003 Personal Wireless Communications*, pages 23–25, 2003. [103](#)
- [46] Massimo Franceschet. Pagerank: standing on the shoulders of giants. *Commun. ACM*, 54(6):92–101, June 2011. [53](#)
- [47] Drew Fudenberg and Eric Maskin. The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica*, 54(3):533–554, 1986. [106](#), [113](#)
- [48] Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, Cambridge, MA, 1991. [103](#)

-
- [49] Deepak Ganesan, Bhaskar Krishnamachari, Alec Woo, David Culler, Deborah Estrin, and Stephen Wicker. Complex Behavior at Scale : An Experimental Study of Low-Power Wireless Sensor Networks. In *Tech. rep. UCLA/CSD -TR 02-0013*, pages 1–11, 2003. 87, 88
- [50] Chao Gui, Baolin Sun, and Chunhua Xia. An entropy-based long-life multipath routing algorithm in manet. In *Proceedings of the 2008 Fourth International Conference on Natural Computation - Volume 05, ICNC '08*, pages 174–178, Washington, DC, USA, 2008. IEEE Computer Society. 69
- [51] Zhu Han, Charles Pandana, and K. J. Ray Liu. A self-learning repeated game framework for optimizing packet forwarding networks. In *in Proc. IEEE Wireless Communications and Networking Conference (WCNC 2005)*, pages 2131–2136, 2005. 17, 19, 103, 104, 113, 114, 116, 117, 147
- [52] Tristan Henderson, David Kotz, and Ilya Ayzov. The changing usage of a mature campus-wide wireless network. In *In Proceedings of ACM MOBICOM*, pages 187–201. ACM Press, 2004. 14, 24, 25, 130, 132
- [53] X. Hong, M. Gerla, G. Pei, and C. Chiang. A Group Mobility Model for Ad Hoc Wireless Networks, 1999. 24
- [54] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990. 26
- [55] Theus Hossmann, Thrasyvoulos Spyropoulos, and Franck Legendre. Know thy neighbor: Towards optimal mapping of contacts to social graphs for dtn routing. In *Proc. of IEEE INFOCOM*, pages 1–9, march 2010. 35, 61
- [56] Theus Hossmann, Thrasyvoulos Spyropoulos, and Franck Legendre. Social network analysis of human mobility and implications for dtn performance analysis and mobility modeling. Technical Report 323, Computer Engineering and Networks Laboratory ETH Zurich, July 2010. 27, 34, 35, 60, 62, 132, 133
- [57] Wei-Jen Hsu, Thrasyvoulos Spyropoulos, Konstantinos Psounis, and Ahmed Helmy. Modeling Spatial and Temporal Dependencies of User Mobility in Wireless Mobile Networks. *IEEE/ACM Trans. on Networking*, 17(5):1564–1577, October 2009. 14, 25, 34, 35, 37, 60, 62, 127, 130, 132, 133
- [58] Pan Hui, Augustin Chaintreau, James Scott, Richard Gass, Jon Crowcroft, and Christophe Diot. Pocket switched networks and human mobility in conference environments. In *Proceedings of the 2005 ACM SIGCOMM workshop*

Bibliography

- on Delay-tolerant networking*, WDTN '05, pages 244–251, New York, NY, USA, 2005. ACM. [14](#), [23](#), [25](#), [132](#)
- [59] Pan Hui, Jon Crowcroft, and Eiko Yoneki. Bubble rap: social-based forwarding in delay tolerant networks. In *Proc. of the 9th ACM international symposium on Mobile ad hoc networking and computing (MobiHoc '08)*, pages 241–250, 2008. [13](#), [14](#), [15](#), [23](#), [26](#), [27](#), [35](#), [61](#), [130](#), [132](#)
- [60] W. Hwang, Taehyong Kim, M. Ramanathan, and A. Zhang. Bridging centrality: Graph mining from element level to group level. In *Proc. of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 336–344, 2008. [15](#), [35](#), [61](#)
- [61] Paul Jaccard. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37:547–579, 1901. [28](#), [36](#), [49](#), [62](#), [79](#), [137](#), [143](#)
- [62] David B. Johnson and David A. Maltz. Dynamic Source Routing in Ad Hoc Wireless Networks. In Imielinski and Korth, editors, *Mobile Computing*, volume 353. Kluwer Academic Publishers, 1996. [24](#), [131](#)
- [63] Thomas Karagiannis, Jean-Yves Le Boudec, and Milan Vojnović. Power law and exponential decay of inter contact times between mobile devices. In *Proc. of the 13th annual ACM international conference on Mobile computing and networking, (MobiCom '07)*, pages 183–194, 2007. [14](#), [25](#), [34](#), [60](#), [132](#)
- [64] Dimitrios Katsaros, Nikos Dimokas, and Leandros Tassiulas. Social network analysis concepts in the design of wireless Ad Hoc network protocols. *IEEE Network*, 24(6):23–29, November 2010. [14](#), [26](#), [35](#), [62](#)
- [65] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, March 1953. [15](#), [18](#), [29](#), [34](#), [36](#), [37](#), [39](#), [62](#), [63](#), [133](#), [134](#)
- [66] M. Kim, D. Kotz, and S. Kim. Extracting a Mobility Model from Real User Traces. In *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, pages 1–13, 2006. [14](#), [24](#), [132](#)
- [67] Ioannis Kontoyiannis, Paul H. Algoet, Yuri M. Suhov, and A. J. Wyner. Nonparametric entropy estimation for stationary processes and random fields, with applications to english text. *IEEE Transactions on Information Theory*, 44(3):1319–1327, 1998. [70](#), [141](#)

-
- [68] P. N. Kugler and M. T. Turvey. *Information, natural law, and the self-assembly of rhythmic movement*. Erlbaum, Hillsdale, New Jersey, 1987. 69
- [69] Hyukjoon Kwon, HyungJune Lee, and John M. Cioffi. Cooperative strategy by stackelberg games under energy constraint in multi-hop relay networks. In *Proceedings of the 28th IEEE conference on Global telecommunications, GLOBECOM'09*, pages 3431–3436, Piscataway, NJ, USA, 2009. IEEE Press. 103
- [70] Linyuan Lü and Tao Zhou. Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, 390(6):1150–1170, 2011. 28, 29, 36, 133
- [71] Kyunghan Lee, Seongik Hong, Seong Joon Kim, Injong Rhee, and Song Chong. Slaw: A new mobility model for human walks. In *Proc. of IEEE INFOCOM*, pages 855–863, April 2009. 35
- [72] Jérémie Leguay, Anders Lindgren, James Scott, Timur Friedman, and Jon Crowcroft. Opportunistic content distribution in an urban setting. In *Proceedings of the 2006 SIGCOMM workshop on Challenged networks, CHANTS '06*, pages 205–212, New York, NY, USA, 2006. ACM. 14, 25
- [73] J. Lian, L. Lin, X. Zhu, and B. Sun. A qos-aware routing protocol based on entropy for mobile ad hoc network. *4th International Conference on Wireless Communications, Networking and Mobile Computing*, pages 1–4, 2008. 69
- [74] David Liben-Nowell and Jon Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7):1019–1031, may 2007. 15, 27, 30, 36, 37, 40, 62, 132, 134
- [75] Jialiang Lu, Fabrice Valois, Mischa Dohler, and Dominique Barthel. Quantifying organization by means of entropy. *IEEE Communications Letters*, 12(3):185–187, 2008. 67
- [76] Sergio Marti, T. J. Giuli, Kevin Lai, and Mary Baker. Mitigating routing misbehavior in mobile ad hoc networks. In *Proceedings of the 6th annual international conference on Mobile computing and networking, MobiCom '00*, pages 255–265, New York, NY, USA, 2000. ACM. 103
- [77] Liam McNamara, Cecilia Mascolo, and Licia Capra. Media sharing based on colocation prediction in urban transport. In J. J. Garcia-Luna-Aceves, Raghupathy Sivakumar, and Peter Steenkiste, editors, *MOBICOM*, pages 58–69. ACM, 2008. 14, 25, 130, 132

Bibliography

- [78] Abderrahmen Mtibaa, Augustin Chaintreau, Jason LeBrun, Earl Oliver, Anna K. Pietilainen, and Christophe Diot. Are you moved by your social network application? In *Proceedings of the First ACM SIGCOMM Workshop on Online Social Networks*, WOSN '08, pages 67–72, New York, NY, USA, August 2008. ACM. [14](#), [15](#), [25](#), [130](#), [132](#)
- [79] Mirco Musolesi and Cecilia Mascolo. Mobility Models for Systems Evaluation. In Benoît Garbinato, Hugo Miranda, Lu'is Rodrigues, Benoît Garbinato, Hugo Miranda, and Lu'is Rodrigues, editors, *Middleware for Network Eccentric and Mobile Applications*, chapter 3, pages 43–62. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. [24](#), [131](#)
- [80] Nicolas Navet and Shu-Heng Chen. On predictability and profitability: Would gp induced trading rules be sensitive to the observed entropy of time series? In Anthony Brabazon and Michael O'Neill, editors, *Natural Computing in Computational Finance*, volume 100 of *Studies in Computational Intelligence*, pages 197–210. Springer, 2008. [62](#), [70](#)
- [81] M E J Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23):8577–82, June 2006. [15](#), [35](#), [61](#)
- [82] Erik Nordström, Christophe Diot, Richard Gass, and Per Gunningberg. Experiences from measuring human mobility using bluetooth inquiring devices. In *Proceedings of the 1st international workshop on System evaluation for mobile platforms*, MobiEval '07, pages 15–20, New York, NY, USA, 2007. ACM. [14](#), [25](#)
- [83] Q. Ou, Y. D. Jin, T. Zhou, B. H. Wang, and B. Q. Yin. Power-law strength-degree correlation from resource-allocation dynamics on weighted networks. *Phys. Rev. E*, 75:021102, 2007. [29](#)
- [84] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford InfoLab., 1999. [15](#), [35](#), [61](#)
- [85] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–8, June 2005. [15](#), [35](#), [61](#)
- [86] C. Pandana, Zhu Han, and K. J. Ray Liu. Cooperation Enforcement and Learning for Optimizing Packet Forwarding in Autonomous Wireless Net-

- works. *Wireless Communications, IEEE Transactions on*, 7(8):3150–3163, 2008. 17, 19, 103, 104, 113, 114, 116, 117, 121, 147
- [87] Pangun Park, Piergiuseppe Di Marco, Pablo Soldati, Carlo Fischione, and Karl Henrik Johansson. A generalized Markov chain model for effective analysis of slotted IEEE 802.15.4. In *2009 IEEE 6th International Conference on Mobile Adhoc and Sensor Systems*, pages 130–139. IEEE, October 2009. 16, 17, 18, 87, 88, 90, 91, 92, 94, 96, 146
- [88] Jim Partan, Jim Kurose, and Brian N. Levine. A survey of practical issues in underwater networks. *SIGMOBILE Mob. Comput. Commun. Rev.*, 11(4):23–33, October 2007. 23
- [89] S. Pollin, M. Ergen, S. Ergen, B. Bougard, L. Der Perre, I. Moerman, A. Bahai, P. Varaiya, and F. Catthoor. Performance Analysis of Slotted Carrier Sense IEEE 802.15.4 Medium Access Layer. *IEEE Transactions on Wireless Communications*, 7(9):3359–3371, September 2008. 16, 88, 91, 96, 146
- [90] Lili Qiu, Yin Zhang, Feng Wang, Mi K. Han, and Ratul Mahajan. A general model of wireless interference. In *MobiCom '07: Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, pages 171–182, New York, NY, USA, 2007. ACM. 127
- [91] Theodore S. Rappaport. *Wireless Communications: Principles and Practice (2nd Edition)*. Prentice Hall, 2002. 89
- [92] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabási. Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586):1551–1555, 2002. 29, 49, 137
- [93] C. Rigano, K. Scott, J. Bush, R. Edell, S. Parikh, R. Wade, and B. Adamson. Mitigating naval network instabilities with disruption toler. In *Military Communications Conference, 2008. MILCOM 2008. IEEE*, pages 1–7, nov. 2008. 23
- [94] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., New York, NY, USA, 1986. 28, 36, 49, 62, 79, 137, 143
- [95] James Scott, Richard Gass, Jon Crowcroft, Pan Hui, Christophe Diot, and Augustin Chaintreau. CRAWDAD trace cambridge/haggle/imote/infocom2006 (v. 2009-05-29). Downloaded from <http://crawdad.cs.dartmouth.edu/cambridge/haggle/imote/infocom2006>, May 2009. 14, 25, 42, 75, 136

Bibliography

- [96] A. Seth, D. Kroeker, M. Zaharia, S. Guo, and S. Keshav. Low-cost communication for rural internet kiosks using mechanical backhaul. In *Proceedings of the 12th annual international conference on Mobile computing and networking*, MobiCom '06, pages 334–345, New York, NY, USA, 2006. ACM. 23
- [97] C. E. Shannon. A mathematical theory of communication. *The Bell Systems Technical Journal*, 27:379–423, 623–656, 1948. 67, 140
- [98] Donghyuk Shin, Si Si, and Inderjit S. Dhillon. Multi-scale link prediction, 2012. 15, 30, 133
- [99] Kim Sneppen, Ala Trusina, and Martin Rosvall. Hide and seek on complex networks. *EUROPHYS.LETT.*, 69:853, 2005. 68
- [100] Kim Sneppen, Ala Trusina, and Martin Rosvall. Measuring information networks. *Pramana (Bangalore)*, 64(6):1121–1125, 2005. 68
- [101] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of Predictability in Human Mobility. *Science*, 327(5968):1018–1021, February 2010. 14, 25, 55, 62, 66, 70
- [102] Han Hee Song, Tae Won Cho, Vacha Dave, Yin Zhang, and Lili Qiu. Scalable proximity estimation and link prediction in online social networks. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, IMC '09, pages 322–335, New York, NY, USA, 2009. ACM. 29, 49, 137
- [103] Libo Song and David Kotz. *Mobile Opportunistic Networks*, chapter Routing in Mobile Opportunistic Networks, pages 1–24. Auerbach Publications, 2011. 25, 43, 126, 134
- [104] Thorwald Sørensen. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol. Skr.*, 5:1–34, 1948. 28, 49, 137
- [105] Gautam S. Thakur, Ahmed Helmy, and Wei-Jen Hsu. Similarity analysis and modeling in mobile societies: the missing link. In *Proc. of the 5th ACM workshop on Challenged networks (CHANTS '10)*, pages 13–20, 2010. 14, 25, 26, 34, 35, 37, 60, 62, 79, 130, 132, 133, 143
- [106] Cristian Tudeuce and Thomas R. Gross. A mobility model based on wlan traces and its validation. In *INFOCOM*, pages 664–674. IEEE, 2005. 14, 24, 25, 130, 132

- [107] H. Van Dyke Parunak and Sven Brueckner. Entropy and self-organization in multi-agent systems. In *Proceedings of the fifth international conference on Autonomous agents*, AGENTS '01, pages 124–130, New York, NY, USA, 2001. ACM. 69
- [108] C. J. van Rijsbergen. *Information Retrieval*. Butterworth, 1979. 49, 137
- [109] Chao Wang, Venu Satuluri, and Srinivasan Parthasarathy. Local Probabilistic Models for Link Prediction. In *Proc. of the Seventh IEEE International Conference on Data Mining, (ICDM '07)*, pages 322–331, October 2007. 15, 36, 37, 62
- [110] Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert L. Barabasi. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, pages 1100–1108, New York, NY, USA, 2011. ACM. 14, 15, 30, 31, 36, 52, 62, 79, 83, 130, 133, 143, 145, 148
- [111] Stanley Wasserman and Katherine Faust. 14, 15, 26, 35, 61, 132
- [112] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998. 27
- [113] Wei Wei, Bing Wang, Chun Zhang, Jim Kurose, and Don Towsley. Classification of access network types: Ethernet, wireless lan, adsl, cable modem or dialup? *Comput. Netw.*, 52(17):3205–3217, December 2008. 69
- [114] Alec Woo, Terence Tong, and David Culler. Taming the underlying challenges of reliable multihop routing in sensor networks. In *Proceedings of the first international conference on Embedded networked sensor systems - SenSys '03*, page 14, New York, New York, USA, 2003. ACM Press. 87, 88
- [115] Shusen Yang, Xinyu Yang, Chao Zhang, and E. Spyrou. Using social network theory for modeling human mobility. *IEEE Network*, 24(5):6–13, 2010. 60
- [116] Marco Zúñiga Zamalloa and Bhaskar Krishnamachari. An analysis of unreliability and asymmetry in low-power wireless links. *ACM Transactions on Sensor Networks*, 3(2):7–es, June 2007. 16, 87, 88, 89, 146
- [117] Mohamed-Haykel Zayani, Vincent Gauthier, Ines Slama, and Djamal Zeghlache. Tensor-based link prediction in intermittently connected wireless networks. *International Journal of Computer and Telecommunications Networking (COMNET)*, submitted. 60, 62, 66, 79, 80, 81

Bibliography

- [118] Mohamed-Haykel Zayani, Vincent Gauthier, and Djamel Zeghlache. Quantifying spatiotemporal stability by means to entropy: Approach and motivations. Technical report, Institut Mines-Télécom, Télécom SudParis, CNRS Samovar UMR 5157, 2010. 127
- [119] H. Zhai and Y. Fang. Physical Carrier Sensing and Spatial Reuse in Multirate and Multihop Wireless Ad Hoc Networks, April 2006. 88
- [120] Hongqiang Zhai, Younggoo Kwon, and Yuguang Fang. Performance analysis of IEEE 802.11 MAC protocols in wireless LANs. *Wireless Communications and Mobile Computing*, 4(8):917–931, December 2004. 87, 88
- [121] Hongwei Zhang, Anish Arora, and Prasun Sinha. Link estimation and routing in sensor network backbones: Beacon-based or data-driven? *IEEE Trans. Mob. Comput.*, 8(5):653–667, 2009. 127
- [122] Hongwei Zhang, Lifeng Sang, and Anish Arora. On the convergence and stability of data-driven link estimation and routing in sensor networks. *ACM Trans. Auton. Adapt. Syst.*, 4(3):18:1–18:29, July 2009. 127
- [123] Jerry Zhao. Understanding Packet Delivery Performance In Dense Wireless Sensor Networks. *Measurement Techniques*, pages 1–13, 2003. 87, 88
- [124] Kun Zhao, Márton Karsai, and Ginestra Bianconi. Entropy of dynamical social networks. January 2012. 70
- [125] Sheng Zhong, Jiang Chen, and Yang Richard Yang. Sprite: A simple, cheat-proof, credit-based system for mobile ad-hoc networks. In *Proc. IEEE Annual IEEE Conference on Computer Communications (INFOCOM)*, pages 1987–1997, 2003. 103
- [126] Tao Zhou, Linyuan Lu, and Yi-Cheng Zhang. Predicting missing links via local information. *The European Physical Journal B*, 71:623, 2009. 29, 49, 137
- [127] Jacob Ziv and Abraham Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, 23(3):337–343, 1977. 16, 18, 61, 62, 70, 141
- [128] M. Zuniga and B. Krishnamachari. Analyzing the transitional region in low power wireless links. In *2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004.*, pages 517–526. IEEE, 2004. 16, 87, 88, 89, 146