Universidade de Aveiro
Departamento de
Eletrónica, Telecomunicações e Informática
2014

Waldir Aranha
Moreira Junior

Encaminhamento Oportunista baseado em Aspectos Sociais

Social-aware Opportunistic Routing

Universidade de Aveiro
Departamento de
Eletrónica, Telecomunicações e Informática
2014

Waldir Aranha
Moreira Junior

Encaminhamento Oportunista baseado em Aspectos Sociais

Social-aware Opportunistic Routing

Tese apresentada às Universidades de Minho, Aveiro e Porto para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Engenharia Eletrotécnica / Telecomunicações no âmbito do programa doutoral MAP-Tele, realizada sob a orientação científica do Doutor Paulo Jorge Milheiro Mendes, Professor Associado da Escola de Comunicação, Arquitetura, Artes e Tecnologias da Informação da Universidade Lusófona, e da Doutora Susana Isabel Barreto de Miranda Sargento, Professora Auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro.

Apoio financeiro da Fundação para a Ciência e a Tecnologia – FCT através da bolsa SFRH/BD/62761/2009.

o júri / the jury

presidente / president	**Doutor Amadeu Mortágua Velho da Maia Soares** Professor Catedrático da Universidade de Aveiro
vogais / examiners committee	**Doutora Susana Isabel Barreto de Miranda Sargento** Professora Auxiliar da Universidade de Aveiro (Orientadora)
	Doutor Paulo Jorge Milheiro Mendes Professor Associado da Universidade Lusófona de Humanidades e Tecnologias (Coorientador)
	Doutor Nuno Manuel de Carvalho Ferreira Guimarães Professor Catedrático do ISCTE - Instituto Universitário de Lisboa
	Doutor Rui Jorge Morais Tomaz Valadas Professor Catedrático do Instituto Superior Técnico da Universidade de Lisboa
	Doutor Rui Luís Andrade Aguiar Professor Associado com Agregação da Universidade de Aveiro
	Doutor Paulo Manuel Martins de Carvalho Professor Associado da Escola de Engenharia da Universidade do Minho
	Doutora Marília Pascoal Curado Professora Auxiliar da Faculdade de Ciências e Tecnologia da Universidade de Coimbra

agradecimentos / acknowledgments

First of all, I would like to thank my advisors, Prof. Dr. Paulo Mendes and Prof. Dr. Susana Sargento, for their valuable guidance.

Thanks to my colleagues at COPELABS, GERCOM/UFPA, MAP-Tele Doctoral Programme for the great moments shared in the lab and during meetings and classes.

I also acknowledge COPELABS for its important role as hosting institution, and FCT for the financial support during my studies.

Thanks to my mother Matilde, father Waldir, brother Glauber, sister Isabella, and nephew Pedro for the intercontinental support throughout this journey. And to my world-wide friends that provided me with unforgettable and relaxing moments.

And, of course, special thanks to that one person that is always by my side, *slatka moja ljubavi*, my beautiful, loving lady Danica Drpic.

palavras-chave

Redes oportunistas, redes tolerantes a atrasos e disrupções, encaminhamento oportunista, encaminhamento orientado a conteúdo, rotinas diárias do utilizador, dinâmica da rede, duração de contatos, estruturas sociais, proximidade social, métricas de encaminhamento, modelo de avaliação, inclusão social e digital.

resumo

A maior capacidade (e.g., processamento, armazenamento) dos dispositivos portáteis, juntamente com a necessidade constante dos utilizadores de poder obter e enviar informação, introduz uma nova forma de comunicação. Os utilizadores podem trocar dados de uma forma transparente através de contatos oportunistas entre eles, o que caracteriza as Redes Oportunistas. Este tipo de rede permite a comunicação entre utilizadores mesmo quando não existe um caminho fim-a-fim entre eles.

Uma tendência observada nos últimos anos do encaminhamento oportunista refere-se a levar em conta métricas de similaridade social para melhorar a troca de informação. Os relacionamentos sociais, interesses em comum e popularidade são exemplos deste tipo de métrica que tem sido empregue com sucesso no âmbito do encaminhamento oportunista: como os utilizadores interagem com base nos seus relacionamentos e interesses, esta informação pode ser utilizada para decidir sobre quando encaminhar dados.

Esta Tese combina as características dos dispositivos pessoais e que são facilmente encontrados no ambiente urbano com a tendência para uso de similaridade social no contexto de encaminhamento oportunista. Para alcancar este objetivo principal, este trabalho foi dividido em diferentes tarefas mapeadas em objetivos específicos, o que resulta nas seguintes contribuições: i) uma taxonomia atualizada sobre encaminhamento oportunista; ii) um modelo de avaliação universal de encaminhamento oportunista que permite a implementação e teste de novas propostas; iii) três funções sociais que consideram o comportamento dinâmico dos utilizadores e podem ser facilmente utilizadas em outras propostas de encaminhamento; iv) duas propostas de encaminhamento oportunista baseadas nas rotinas diárias dos utilizadores e no conteúdo e interesse dos utilizadores neste conteúdo; e v) uma análise estrutural da rede social formada a partir das abordagens desenvolvidas neste trabalho.

keywords

Opportunistic network, delay/disruption-tolerant networks, opportunistic routing, content-oriented routing, user daily routines, network dynamics, contact duration, social structures, social proximity, routing metrics, assessment model, digital/social inclusion.

abstract

The increased capabilities (e.g., processing, storage) of portable devices along with the constant need of users to retrieve and send information have introduced a new form of communication. Users can seamlessly exchange data by means of opportunistic contacts among them and this is what characterizes the opportunistic networks (OppNets). OppNets allow users to communicate even when an end-to-end path may not exist between them.

Since 2007, there has been a trend to improve the exchange of data by considering social similarity metrics. Social relationships, shared interests, and popularity are examples of such metrics that have been employed successfully: as users interact based on relationships and interests, this information can be used to decide on the best next forwarders of information.

This Thesis work combines the features of today's devices found in the regular urban environment with the current social-awareness trend in the context of opportunistic routing. To achieve this goal, this work was divided into different tasks that map to a set of specific objectives, leading to the following contributions: i) an up-to-date opportunistic routing taxonomy; ii) a universal evaluation framework that aids in devising and testing new routing proposals; iii) three social-aware utility functions that consider the dynamic user behavior and can be easily incorporated to other routing proposals; iv) two opportunistic routing proposals based on the users' daily routines and on the content traversing the network and interest of users in such content; and v) a structure analysis of the social-based network formed based on the approaches devised in this work.

“It is better to have enough ideas for some of them to be wrong,
than to be always right by having no ideas at all.”
— Edward De Bono

To my Brazilian, Serbian,
Montenegrin, Canadian, and
Portuguese families and friends.
To my parents Matilde e Waldir.
To my siblings Isabella and
Glauber, and my nephew Pedro.
To Danica.

Contents

List of Contributions

Book Chapters

- Waldir Moreira and Paulo Mendes. “Social-aware Opportunistic Routing: The New Trend”. In: I. Woungang, S. Dhurandher, A. Anpalagan, A. V. Vasilakos (Eds.), Routing in Opportunistic Networks, Springer Verlag, May, 2013.

- Rute Sofia, Paulo Mendes, Waldir Moreira. “User-centric Networking Living-Examples and Challenges Ahead”. In: A. Bogliolo, A. Aldini (Eds.), User-Centric Networking: Future Perspectives, Springer Lecture Notes in Social Networks, May, 2014.

- Namusale Chama, Antonio Junior, Waldir Moreira, Paulo Mendes, Rute Sofia. “User-centric Networking, Routing Aspects”. In: A. Bogliolo, A. Aldini (Eds.), User-Centric Networking: Future Perspectives, Springer Lecture Notes in Social Networks, May, 2014.

- Carlos Ballester Lafuente, Jean-Marc Seigneur, Rute Sofia, Christian Silva, Waldir Moreira. “Trust Management in ULOOP”. In: A. Bogliolo, A. Aldini (Eds.), User-Centric Networking: Future Perspectives, Springer Lecture Notes in Social Networks, May, 2014.

- Paulo Mendes, Waldir Moreira, Tauseef Jamal, Huseyin Haci, Huiling Zhu. “Cooperative Networking in User-Centric Wireless Networks”. In: A. Bogliolo, A. Aldini (Eds.), User-Centric Networking: Future Perspectives, Springer Lecture Notes in Social Networks, May, 2014.

Journal Papers

- Waldir Moreira, Paulo Mendes, and Susana Sargento. “Assessment Model for Opportunistic Routing”. IEEE Latin America Transactions, 10(3), April, 2012.

- Ronedo Ferreira, Waldir Moreira, Paulo Mendes, Mario Gerla, Eduardo Cerqueira. “Improving the Delivery Rate of Digital Inclusion Applications for Amazon Riverside Communities by Using an Integrated Bluetooth DTN Architecture”. International Journal of Computer Science and Network Security, 14(1), January, 2014.

Internet Draft

- Waldir Moreira, Paulo Mendes, and Eduardo Cerqueira. “dLife: Opportunistic Routing Based on Social Daily Routines”. Internet Draft, draft-moreira-dlife-04, work in progress, May, 2014.

Conference Proceedings

- Waldir Moreira and Paulo Mendes. "Routing Metrics for Delay-Tolerant Networks". In: Proceedings of CRC, Braga, Portugal, November, 2010.
- Waldir Moreira, Paulo Mendes, and Susana Sargento. "Assessment Model for Opportunistic Routing". In: Proceedings of IEEE LATINCOM, Belem, Brazil, October, 2011.
- Carlos Ballester, Jean-Marc Seigneur, Waldir Moreira, Paulo Mendes, Linas Maknavicius, Alessandro Bogliolo, and Paolo di Francesco. "Trust and Cooperation Incentives for Wireless User-Centric Environments". In: Proceedings of IADIS e-Society, Berlin, Germany, March, 2012.
- Waldir Moreira, Paulo Mendes, and Susana Sargento. "Opportunistic Routing Based on Daily Routines". In: Proceedings of IEEE WoWMoM AOC, San Francisco, USA, June, 2012.
- Waldir Moreira, Manuel de Souza, Paulo Mendes, and Susana Sargento. "Study on the Effect of Network Dynamics on Opportunistic Routing". In: Proceedings of ADHOC-NOW, Belgrade, Serbia, July, 2012.
- Alessandro Bogliolo, Paolo Polidori, Alessandro Aldrini, Waldir Moreira, Paulo Mendes, Mursel Yildiz, Carlos Ballester, and Jean-Marc Seigneur. "Virtual Currency and Reputation-based Cooperation Incentives in User-Centric Networks". In: Proceedings of IWCMC, Limassol, Cyprus, August, 2012.
- Waldir Moreira, Ronedo Ferreira, Douglas Cirqueira, Paulo Mendes and Eduardo Cerqueira. "SocialDTN: A DTN implementation for Digital and Social Inclusion". In: Proceedings of ACM MobiCom LCDNet, Miami, USA, September, 2013.
- Waldir Moreira, Paulo Mendes, and Susana Sargento. "Social-aware Opportunistic Routing Protocol based on User's Interactions and Interests". In: Proceedings of AdHocNets, Barcelona, Spain, October, 2013.

Demonstrations

- Carlos Ballester, Jean-Marc Seigneur, Paolo di Francesco, Alessandro Bogliolo, Waldir Moreira, Rute Sofia, Nuno Martins, and Valentino Moreno. "A User-centric Approach to Trust Management in Wi-Fi Networks". In: Proceedings of IEEE INFOCOM, Turin, Italy, April, 2013.

Technical Reports

- Waldir Moreira and Paulo Mendes. "Survey on Opportunistic Routing for Delay/Disruption-Tolerant Networks". Technical Report SITI-TR-11-02, SITILabs, University Lusofona, February, 2011.
- Waldir Moreira and Paulo Mendes. "Social-aware Utility Functions for Opportunistic Routing". Technical Report SITI-TR-12-05, SITILabs, University Lusofona, August, 2012.
- Waldir Moreira and Paulo Mendes. "Social-aware Opportunistic Routing Solutions". Technical Report SITI-TR-13-01, SITILabs, University Lusofona, January, 2013.

- Waldir Moreira and Paulo Mendes. “Structural Analysis of Social-aware Opportunistic Networks”. Technical Report SITI-TR-13-05, SITILabs, University Lusofona, August, 2013.

Software Suites

- Manuel de Souza and Waldir Moreira. “TECD router”. Software SITI-SW-11-03, SITILabs, University Lusofona, April, 2011.
- Waldir Moreira. “dLife v0.1: Opportunistic Routing based on Social Daily Routines”. Software SITI-SW-11-06, SITILabs, University Lusofona, June, 2011.
- Waldir Moreira. “dLife v1.0: Opportunistic Routing based on Social Daily Routines”. Software SITI-SW-12-02, SITILabs, University Lusofona, September, 2012.
- Waldir Moreira. “Social-aware Content-based Opportunistic Routing Protocol (SCORP)”. Software SITI-SW-13-01, SITILabs, University Lusofona, July, 2013.

List of Acronyms

ACK	Acknowledgment
AD	Average Duration
AP	Access Point
CD	Contact Duration
CiPRO	Context Information Prediction for Routing in OppNets
CRAWDAD	Community Resource for Archiving Wireless Data At Dartmouth
dLife	Opportunistic Routing Based on User Social Daily Routine
DM	Decision Maker
DTN	Delay/disruption-Tolerant Networks
EASE	Exponential Age SEarch
EBR	Encounter-Based Routing
EG	Evolving Graph
EID	Endpoint Identifier
FRESH	FResher Encounter SearcH
GREASE	GReedy EASE
GSM	Global System for Mobile
IA	Importance Assigner
ICON	Information Centric Architecture for Opportunistic Networks
ID	Internet-Draft
IRTF	Internet Research Task Force
LEO	Low Earth Orbit
MEED	Minimum Estimated Expected Delay
MIT	Massachusetts Institute of Technology
ONE	Opportunistic Network Environment
OPF	Optimal Probabilistic Forwarding
OppNets	Opportunistic Networks
PREP	PRioritized EPidemic
PROPHET	Protocol using History of Encounters and Transitivity
RAPID	Resource Allocation Protocol for Intentional DTN
SCF	Store-Carry-and-Forward
SCORP	Social-aware Content-based Opportunistic Routing Protocol
SIG	Social Information Gatherer
SIR	Social Information Repository
SW	Social Weighter

TCT	Total Connected Time
TECD	Time-Evolving Contact Duration
TECDi	TECD Importance
TECI	Time-Evolving Contact to Interest
TLV	Type-Length-Value
TTL	Time-To-Live
UCN	User-Centric Networks
UEF	Universal Evaluation Framework
ULOOP	User-centric Wireless Local Loop European project

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

The way communication happens between devices and information is accessed has evolved over time: few years ago the exchange of data took place between fixed devices and data was mostly retrieved from specific repositories. With the advancements in technology, miniaturization of components has made devices more portable. Moreover, such advancements have also made these devices more powerful in terms of processing/storage and with different built-in access technologies (e.g., Wi-Fi, Bluetooth, GSM).

As a consequence, the users of these devices start not only consuming, but also producing data (i.e., becoming prosumers) and building a constant need to be always connected while on-the-go. However, the existing infrastructure (or lack thereof) is not fully compliant with the new communication and information access paradigms. This explains why the existence of different wireless networking approaches, such as mobile adhoc, mesh, cooperative, delay tolerant, opportunistic, and information centric.

All of these approaches are well known, with established architectural design, and have proven their applicability in a wide range of scenarios. But independently of the approach, all of them share a common concern: how to deal with link intermittency. Intermittency can result from node mobility/speed, power-saving mechanisms, distance between communicating parties, physical obstacles, failed/crashed nodes, shadowed areas, weather conditions, among others. Moreover, there are different levels of intermittency: fast (e.g., node reboot), medium (e.g., crossing a infrastructureless area), and long (e.g., waiting for a Low Earth Orbit - LEO - satellite to pass over). Also, the experienced level of intermittency is seen differently by each networking approach: a delay-tolerant interplanetary network may classify an intermittent link, due to a LEO satellite revolution around Earth, as fast in comparison to a link for communication with a Mars rover, while an intermittent link, due to the same revolution, may be seen by an adhoc military network on a hostile situation as too long.

Considering their applicability scenarios and the level of expected intermittency, these networking approaches employ different routing schemes to cope with intermittent connectivity, and therefore with issues such as long/variable delay, asymmetric data rates, high error rates, and network partition that may arise. Under these conditions there is a possibility for the non-existence of end-to-end paths towards a destination [1], which the regular routing solutions cannot properly handle [2].

As mentioned earlier, users' devices have increasing processing power and storage, and with these features they can follow the store-carry-and-forward (SCF) paradigm, where they can keep data until another good intermediate carrier node or the destination is found. This allows for the creation of forwarding opportunities and to mitigate the communication problem that arises from intermittent connectivity even when an end-to-end path is absent.

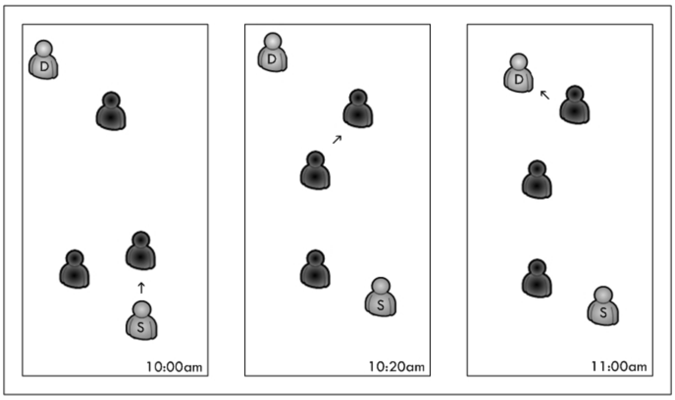


Figure 1.1: Example of data delivery based on SCF paradigm

In Fig. 1.1[1], it is easily seen the absence of an end-to-end path due to the aforementioned problems (e.g., lack of infrastructure) between the source (S) and destination (D) nodes. Based on the SCF paradigm (also employed in Delay-Tolerant Networking as store-and-forward [3]), wireless devices that opportunistically meet can serve as forwarders until the destination is found.

Opportunistic routing has gained attention by taking advantage of contact opportunities to improve forwarding: proposals range from using node mobility to flood the network for fast delivery (e.g., *Epidemic* [4]) up to controlling such flooding in order to achieve the same results based on: encounter history (e.g., *PROPHET* [5, 6]), limited replication (e.g., *Spray and Wait* [7]), prioritization (e.g., *MaxProp* [8]), and encounter prediction (e.g., *EBR* [9]).

Since 2007, a trend has emerged based on different representations of social similarity: i) labeling users according to their work affiliation (e.g., *Label* [10]); ii) looking at the importance (i.e., popularity) of nodes (e.g., *PeopleRank* [11]); iii) using centrality and the notion of community (e.g., *SimBet* [12] and *Bubble Rap* [13]); iv) considering interests that users have in common (e.g., *SocialCast* [14]); v) inferring different levels of social interactions aiming at predicting future social interactions from the users' dynamic behavior found in their daily life routines (e.g., *dLife* [15, 16] and *CiPRO* [17]); vi) combining social information and content knowledge (e.g., *ContentPlace* [18] and *SCORP* [19]).

This Thesis exploits social-aware and content-based opportunistic routing through *dLife* [15, 16] and *SCORP* [19], respectively. They present great potential in what concerns information delivery, since: i) cooperation among users sharing social aspects (i.e., users' relationships, shared interests, common communities) is encouraged, which is beneficial to improve content dissemination [20]; ii) social information is less volatile than human mobility, providing more robust and reliable connectivity graphs, which aids routing [2, 13]; iii) focusing on the content mitigates even more the effects of the lack of an end-to-end path between nodes [18].

[1]Springer and the original publisher (Routing in Opportunistic Networks, v. 1, 2013, p. 27-68, Social-Aware Opportunistic Routing: The New Trend, Waldir Moreira, Paulo Mendes, Figure 2.1, Copyright © 2013, Springer Science+Business Media New York) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media [1].

1.1 Motivation

The challenges faced by routing solutions are further increased by node density. In sparse scenarios, communication suffers with very long delay links (e.g., space communications [21]) or where communication is often disrupted, meaning that contacts do not occur very frequently (e.g., rural areas [22]). The routing challenges imposed by these scenarios are more related to the great distances between the communicating nodes or the time nodes are away from each other, being more of a problem to transport protocols than to routing.

For instance, in space communications the problem is to find a transport protocol able to send packets over wireless links with very long delays. Another example applies to rural areas, where the lack of telecommunications infrastructure leads to the adoption of ferry type of solutions [23], in which the source passes the information to a moving node that carries it directly to the destination(s). Besides space communications and message ferrying, another example of a solution suitable to be applied to sparse scenarios is related with data mules [24], in which specific nodes move around collecting information from several nodes (e.g., sensors) carrying such information to a well-known destination.

Yet, the scenarios considered by the studied proposals (presented in Chapter 2) are more challenging in terms of opportunistic routing, which means looking at dense networks where problems such as long delays in the presence of several routing opportunities and intermittent links (e.g., due to node mobility) intertwine with overlapping spectrum and different sources of interference that increase the routing challenge.

It is important to note that density is defined by node degree (i.e., the unique encounters a node has with others). Thus, node density refers to how well nodes interact in the network. Examples of the envisioned interaction scenarios are the ones found nowadays in urban areas: highly mobile users carrying powerful devices and demanding connectivity in places with poor/without coverage at all (i.e., areas either full of closed APs or with many open APs but with very high interference, subways, public areas).

The routing challenges related to dense scenarios are the main motivation of this work. Thus, social-aware and content-based opportunistic routing approaches are proposed, aiming at forwarding messages towards destinations with high delivery probability, within a time frame that is useful for the lifetime of the message, and with good usage of network resources. It is worth noting that, despite not focusing on sparse scenarios (where node degree would be at most 2, for instance), the proposals presented in this Thesis are devised completely agnostic to the node density of the employing network.

1.2 Objectives

This Thesis aims at investigating social-aware and content-based opportunistic routing solutions. For this purpose, the following set of specific objectives were established:

1. Identify the different types of opportunistic routing approaches.

2. Understand the existing opportunistic routing taxonomies.

3. Study the employed opportunistic routing metrics.

4. Study how nodes can engage in cooperation.

5. Develop social-aware utility functions.

6. Propose social-aware and content-based opportunistic routing approaches.

7. Perform an analysis on the network structure resulting from the devised social-aware approaches.

1.3 Main Contributions

The analysis of related work in the field of opportunistic routing shows that proposals spanning the last thirteen years head towards an approach where mobility is exploited taking advantage of the stochastic encounters that happen between nodes. In addition, social similarity (e.g., relationships, interests) has been used to aid in data forwarding, since topology based in social aspects tend to be less volatile (i.e., vary less) than the ones based on mobility.

It is clear that one recent trend of opportunistic routing solutions is going towards the awareness about social aspects, which are only related to inter-meeting times or actually related to more complex social similarities. It is observed that social ties can indeed aid in data forwarding [13, 15, 16]. Such forwarding becomes even better if user's interests [14, 18, 19] are taken into account and if the social relationship of nodes (carried by people) is explicit [11].

This section briefly presents a description of the contributions of this Thesis to the investigation done in the area of opportunistic routing and to advance the state-of-the-art in the field of social-aware opportunistic routing approaches.

1.3.1 Classification and Evaluation of Opportunistic Routing Approaches

One can easily find different classifications of opportunistic routing approaches. However, such classifications consider aspects (e.g., level of knowledge) that lead to an unbalanced classification, assigning most solutions to a few set of categories, or to very specific classification branches (e.g., by considering information coding or methods to control movement of nodes).

This first contribution covers objectives 1, 2 and 3 as described in Sec. 1.2. It provides an analysis of existing taxonomies regarding opportunistic routing and metrics [25, 26], and proposes a suitable routing taxonomy. The latter includes and highlights the importance of the social-aware category, for an easy classification of current and future proposals [1].

By looking at the state-of-the-art, one can observe that performance evaluation methods consider different aspects (e.g., node number, mobility models) which prevent a more reliable and objective evaluation: proposals are compared in scenarios totally different from the ones they were designed for. In this context, another contribution of this Thesis is the development of a universal evaluation framework [27, 28] to allow a more consistent assessment process, especially when carrying out comparisons between older and new approaches.

1.3.2 Encouraging User Cooperation

Opportunistic routing depends on the willingness of users to cooperate by carrying/storing/relaying information on behalf of others. Cooperation may appear in different levels since it should not be assumed that all nodes are always willing to share their constrained resources on the behalf of others.

Moreover, trust should be applied especially to identify malicious nodes nearby (e.g., DoS attack) and to avoid such nodes for data exchange. The cooperation level between nodes may be related to the information being forwarded. So, this contribution covers objective 4 (cf. Sec. 1.2), investigating how nodes can be encouraged to participate in cooperation [29, 30, 31].

1.3.3 Social-aware Utility Functions

With the gathered knowledge about existing opportunistic routing approaches, especially of those based on social aspects, a set of social-aware utility functions was specified and validated. The proposed functions aim to provide low delivery delay while consuming low number of resources in disruptive scenarios with dense topologies. This contribution covers objective 5 (cf. Sec. 1.2), and presents the devised social-aware utility functions [32, 33].

1.3.4 Social-aware and Content-based Opportunistic Routing Protocols

This contribution covers objective 6 (cf. Sec. 1.2). Based on the created utility functions, the social-aware opportunistic routing protocol based on user's social daily routines, *dLife* [15, 16], was created.

dLife comprises the routing solution of SocialDTN, an instantiation of the DTN architecture and Bundle Layer for Android devices [34]. Both *dLife* and SocialDTN are included in the context of the DTN-Amazon project that aims at promoting the digital/social inclusion of riverside communities of the Amazon region.

One can easily observe that most of the current opportunistic routing proposals support only destination-based (i.e., point-to-point) communications, being *dLife* an example. With that mind, a Social-aware Content-based Opportunistic Routing Protocol, *SCORP* [35, 19] was devised to support point-to-multipoint communications, since this seems to be the basic foundation of applications able to distribute information in challenged networks.

1.3.5 Structure Analysis of Social-based Networks

Understanding the type of graphs (e.g., scale free, small world) formed by routing solutions is imperative to guarantee the suitability of such proposals. One can normally find such analysis based on the social interaction of nodes over the whole duration of the experiments.

The final contribution of this Thesis work covers objective 7 (cf. Sec. 1.2) and provides an analysis of the network structure formed by the social-aware opportunistic routing approaches devised within the context of this work. Such approaches are based on users' social behavior at different periods of time during the users' daily routines. Hence, this analysis presents the dynamic nature of these social structure throughout the users daily interactions [36].

1.4 Outline

The remainder of this Thesis is divided into the following chapters:

Chapter 2 overviews the opportunistic routing proposals related to the development of this Thesis work. It also provides an insight into the existing opportunistic routing taxonomies and metrics as well as existing evaluation frameworks.

Chapter 3 provides an updated opportunistic routing taxonomy along with a universal evaluation framework that aids in assessing the performance of opportunistic routing approaches. It also highlights the role of cooperation in opportunistic networking

Chapter 4 presents the devised social-aware utility functions.

Chapter 5 describes the proposed social-aware and content-based opportunistic routing approaches.

Chapter 6 presents the performance evaluation of the proposed utility functions and routing approaches against well-known opportunistic routing benchmarks.

Chapter 7 provides an analysis on the structure of social-based networks.

Chapter 8 concludes the work and provides future directions for further development of the topics presented in this Thesis.

Chapter 2

Related Work

In order to achieve my goal of proposing a new routing solution, I had to learn about the different opportunistic routing approaches. This led me to propose a new opportunistic routing taxonomy as well as a new universal evaluation framework. This chapter presents the relevant work that helped understanding the different opportunistic routing schemes, how they are related to one another, and the routing metrics they employ.

2.1 Opportunistic Routing Approaches

As I surveyed the state-of-the-art literature, I observed that the existing opportunistic routing proposals either simply ignored social aspects or considered different levels of social awareness when taking forwarding decisions. This section briefly presents the most relevant social-oblivious as well as the social-aware opportunistic routing proposals that aided me carrying out my work.

2.1.1 Social-oblivious Opportunistic Routing

Prior to understanding the importance of social awareness in the context of opportunistic routing and its applicability, I studied proposals covering a 12-year period (2000-2012). To have a better view of the proposals and identify relationships among them, I grouped these social-oblivious proposals into three categories: *single-copy routing*, aiming to improve the usage of network resources; *epidemic routing*, aiming to increase delivery probability; and, *probabilistic-based routing*, aiming to find an optimal balancing between both previous categories.

2.1.1.1 Single-copy Routing

The proposals in this category aim at optimizing the usage of network resources: they refrain from replication and forward messages at every hop based on available connectivity and some form of mobility prediction. This means that only one copy of each message traverses the network towards the final destination. Such copy can be forwarded if the node carrying it decides (i.e., randomly, or based on a utility function) that another encountered node presents a higher probability to deliver the message.

Minimum Estimated Expected Delay (MEED) [37] is an example of a single-copy routing. It uses a contact history (i.e., connection and disconnection times of contacts) metric that estimates the time a message will wait until it is forwarded. A per-contact routing scheme is used to "override" regular link-state routing decision. This allows *MEED* to use any other contact opportunity (i.e., node) that arises prior to what is expected (in terms of computed metric), in order to forward the message.

Spyropoulos et al. (2008) [38] present different single-copy routing algorithms (e.g., *direct transmission*, *randomized routing*, *utility-based routing with 1-hop diffusion*, *utility-based routing with transitivity*, *seek and focus routing*, *oracle-based optimal algorithm*) that simply take advantage of contacts or are based on forwarding probability/utility functions/oracle to forward data.

Single-copy routing is a very interesting approach from a resource point of view, as it keeps the usage of network (e.g., bandwidth) and node (e.g., energy, storage) resources at a low level. However, the experienced delay in message delivery might be quite high. This, in turn, can affect the delivery capability of single-copy approaches that ends up being very low. Another issue is related to the amount of knowledge that needs to be exchanged/available in order to aid forwarding, which in some scenarios generates too much overhead and may be impossible to implement.

To improve delivery, the next category of opportunistic routing approaches relies on replication to spread copies at every contact.

2.1.1.2 Epidemic Routing

Epidemic routing replicates messages at every contact: it takes advantages of any contact to forward a copy of messages to the encountered node as to increase their delivery probability.

A well-known example is *Epidemic* [4], a proposal that employs a full replication strategy, and consequently increases delivery rate. In order to avoid replication of messages already in the buffers of nodes, summary vectors are exchanged between nodes and list the already carried messages. Also, a hop count is maintained as to avoid indefinite message replication.

The proposal indeed increases delivery rate. However, resources of the network/nodes are exploited at their full capacity. This can easily affect the scalability of the proposal, especially in dense, urban scenarios, which are subject of this Thesis work.

The following category aimed at minimizing the consumption of resources by employing a controlled replication approach: either based on the probabilistic choice of next nodes or by using a utility function.

2.1.1.3 Probabilistic-based Routing

Probabilistic-based routing aims at controlling replication and spreading only a few copies of messages by considering nodes' capabilities, message prioritization, encounter history, among others. The main goal is to achieve high delivery rates with low delay and cost.

This category is further divided considering the following routing metrics: *frequency encounters*, *aging encounters*, *aging messages*, and *resource allocation*.

Frequency Encounters: these proposals consider the number of times nodes meet. They are: the *Probabilistic ROuting Protocol using History of Encounters and Transitivity* (PROPHET), *MaxProp*, *Prediction*, and *Encounter-Based Routing* (EBR).

PROPHET [5, 6] uses a probabilistic metric (delivery predictability) that tells how good a node is to deliver a message according to past encounters with its destination. Replication only occurs when nodes that have higher delivery predictability are found.

MaxProp [8] uses a metric called delivery likelihood of messages, by having each node keeping track of a probability of meeting any other peer. Based on the delivery likelihood, a carrier of a message computes a cost for each possible path to the destination, up to n hops. This cost estimation, along with the hop count, are then used to order messages for scheduling and for dropping. In addition, *MaxProp* assigns priority levels to messages to decrease their delivery time, uses a hop list in each message to avoid doubled reception, and uses acknowledgments for the delivered message.

Prediction [39] makes use of contact information to estimate the probability of meeting other nodes in the future. It estimates the contact probability (i.e., timely-contact probability metric) within a period of time by using historical contact information. Upon a contact, replication occurs if the encountered node has a higher contact probability towards the destination of the message.

EBR [9] counts the number of contacts a node has with other nodes, and determines the Encounter Value (EV) that represents the past rate of encounters of the node. With this information, it can predict the rate level of future encounters: the higher EV is, the higher the probability of successful message delivery. Thus, messages are replicated only to nodes with higher EV. EV is also used to determine the number of replicas the encountered node is able to create after receiving a copy.

These proposals are able to increase delivery rate, while reducing the consumption of resources. However, they can still flood the network (in the case of low-predictability nodes only encountering high-predictability ones) and suffer with occasional loops (*PROPHET* [5]), require unlimited storage capabilities (*MaxProp* [8]), experience high delivery delays (*Prediction* [39]), or have great dependence of frequent/long node encounters (*EBR* [9]).

Aging Encounters: these proposals are based on the time elapsed since the last encounter between a given node and the destination. Proposals considering this routing metrics are: *Exponential Age SEarch* (EASE), *FResher Encounter SearcH* (FRESH), and *Spray and Focus.*

EASE [40] considers the history of encounters and geographic position of nodes to make routing decisions based on the time and location of last encounter with the destination. This means that replication is decided considering if the encountered node, either has a recent encounter with the destination, or is physically closer to the destination.

FRESH [41] keeps track of the time elapsed since the last time nodes met. Data forwarding occurs if the time elapsed since the last contact between the encountered node and the destination is smaller than the time elapsed since the last contact between the carrier and the destination.

The most recent example is *Spray and Focus* [42], which spreads a few message copies, and then each copy is forwarded based on a utility function considering the time elapsed since the last contact between the encountered node and the destination.

Since these proposals are based on timers, their performance may be degraded as such timers become obsolete in high mobile environments, resulting in inaccuracies about encounters. Some proposals (*EASE* [40], *FRESH* [41]) depend on mobility patterns and node speed, and may not cope with disconnected clusters, commonly seen in opportunistic communication.

Aging Messages: these proposals control the distance and time message copies can go and stay in the network. *Spray and Wait*, and *Optimal Probabilistic Forwarding* (OPF) are the proposals using this routing metric.

Spray and Wait [7] first sprays copies of a message in an epidemic-like fashion. To avoid flooding, up to L copies of each message can be distributed in the network. Then, the nodes holding a copy of the message enter the wait phase and deliver the messages solely to their destinations.

OPF [43] only replicates messages if such action increases the overall delivery probability of the message. This means that replication maximizes the joint expected delivery probability of the copies to be placed in the system (i.e., in the sender and receiver nodes of the message). *OPF* considers the remaining hop-count and residual lifetime of messages to estimate the effect that message replication may have on the expected delivery rate while satisfying the constant on the number of forwardings per message.

These proposals have problems with stale message removal as well as computational effort to determine message copies that can degrade performance (*Spray and Wait* [7]), and given the different pattern in the mobility of nodes, the needed inter-contact time information may not be readily available (*OPF* [43]).

Resource Allocation: these approaches measure resource availability to perform wise utilization. They are the *Resource Allocation Protocol for Intentional DTN* (RAPID) and *PRioritized EPidemic* (PREP).

RAPID [44] estimates the effect that message replication may have on a predefined performance metric in a network with resource constraints, and replicates messages only if such effect is justifiable.

PREP [45] measures the average fraction of time a link will be available in the future, and defines drop and transmission priorities for messages. By determining the availability of the links for future use, *PREP* can wisely use the resources (e.g., storage and bandwidth).

These proposals may not cope with short-lived contacts that may lead to performance degradation, since global network state is required, by exchanging such information as nodes encounter (*RAPID* [44]), and high level of contact disruption (*PREP* [45]).

2.1.1.4 Summary

This section briefly presented the social-oblivious opportunistic routing proposals that I have considered in my state-of-the-art surveying. My choices for these proposals were driven by: i) the number of times they have been cited (i.e., served as benchmarks); and, ii) the number of benchmarks used for their evaluation.

This close analysis helped me identifying the advantages and disadvantages of these proposals regarding their functionalities and application. With this knowledge, I further explored the potential of social metrics to improve routing performance as they allow forwarding decisions to exploit social similarities, which present less volatile characteristics than mobility metrics.

To conclude, it is worth noting that some authors may consider some of the aforementioned solutions as social-aware, for taking into account history of encounters, for instance. However, I believe that social-aware solutions function based on much more elaborate utility functions and/or consider features that can be used to identify/classify individuals or groups of these, i.e., common affiliations, shared interests, social ties, popularity, centrality, among others, which are further discussed next in Sec. 2.1.2.

2.1.2 Social-aware Opportunistic Routing

After gathering knowledge concerning the functioning and applicability of social-oblivious opportunistic routing, I delved into the social-aware versions of such type of routing. As its name suggests, social-aware opportunistic routing is based on social similarity metrics (e.g., common affiliations, shared interests, strength of social ties, popularity, centrality).

These metrics have shown great potential in improving opportunistic routing: the more accurate the social similarity identified among nodes, the more efficient content dissemination is [20]. In addition, this similarity is drawn from the existing social relationships between users, and such relationships are less volatile (i.e., change less often) than the regular considered mobility patterns [2, 13]. This makes social information in the context of opportunistic networks rather interesting since routing decisions can consider strong, well-connected links as opposed to the weak, constantly changing links formed based on mobile behavior.

Thus, one can easily observe that not only do probabilistic-based proposals consider node mobility, but also the social similarity among such nodes. This explains the reason for the appearance, since 2007, of different proposals exploiting social aspects in order to improve delivery rate and decrease delivery costs. This section briefly overviews the following proposals, highlighting their functionalities, advantages and disadvantages: *Label*, *SimBet*, *Bubble Rap*, *SocialCast*, *ContentPlace, PeopleRank*, *CiPRO.* To have a better view of these proposals, I grouped them into four categories: *community-based routing*, that considers user affiliations, communities to which they belong, and their centralities; *interest-based routing*, that relies on users' interests in the content traversing the network as well as content spreading and availability; *popularity-based routing*, which ranks the importance of nodes according to their interaction in the network; and, *dynamic behavior-based routing*, that attempts to reflect the dynamism of node behavior while choosing the next forwarders.

2.1.2.1 Community-based Routing

Community-based routing relies on the clustering of nodes. This clustering can be achieved by simply grouping nodes according to their affiliation [10] or considering elaborate clustering algorithms such as K-clique [46] and weighted network analysis [47]. Additionally, this routing approach can consider different levels of centrality of the nodes (e.g., betweenness [48]) to further improve forwarding decisions. Proposals considering this routing approach are *Label*, *SimBet*, and *Bubble Rap*.

Label [10] was the first to introduce the usage of social information into opportunistic routing. Besides being a rather simple approach, the proposal showed the potential application of social aspects. It consists in labeling the nodes according to their affiliation/groups, and forwarding messages directly to their destinations or to encountered nodes that belong to the same affiliation/group (i.e., labels) as the destinations. Despite of the potential of the applied social information, the delivery rate performance of *Label* depends on message TTL (the higher, the better) and can be easily degraded in scenarios where nodes have a low mixing rate (i.e., nodes do not interact much outside their groups).

SimBet [12] defines a utility function that considers the betweenness centrality of nodes and social similarity at cluster level. Thus, the function identifies nodes that bridge different clusters and the probability of future collaborations of nodes sharing the same communities (i.e., clusters). Nodes compute their *SimBet* utility towards destinations to which they carry messages. Messages are exchanged based on the highest value of the utility function, and are removed from the buffers. This results in a single-copy approach able to have a delivery rate performance close to the one shown by *Epidemic* and very low associated cost (i.e., number of forwardings). However, *SimBet* may experience high delay as it relies on how often and with whom nodes meet. If nodes present a low contact rate, utility values and contact lists may take longer to be updated and diffused. Also short-lived contacts may affect its performance as it requires the exchange of meta-data.

Bubble Rap [49, 13] also employs node centrality, but combines it with the idea of community structure to perform forwarding. Communities are formed considering the number of contacts between nodes and their duration, and centrality is seen from a local (i.e.., inside communities) and global (i.e., whole network) perspective. Messages are replicated based on global centrality until they reach the community of the destination (i.e., a node belonging to the same community). Then, local centrality is used to reach the destination inside the community. To reduce replication, once the message reaches the community of the destination, the node forwarding such message, removes it from its buffer.

By mimicking human behavior through community and node popularity, *Bubble Rap* manages to have good overall delivery rate performance and suitable delivery cost. The main issues *Bubble Rap* needs to cope with are related to the centrality: it may create bottlenecks (i.e., the number of very popular nodes is very small), and a high centrality node may not always have good links to the destination community.

2.1.2.2 Interest-based Routing

Interest-based routing does not focus on the hosts to perform forwarding. Instead, it considers the interest that users have in the content traveling the network to decide whether a copy shall be created. As presented next, proposals in this category can be much more data-aware (i.e., consider how much of the content has been spread and its availability). Proposals based on interests are: *SocialCast* and *ContentPlace*.

SocialCast [14] considers the interest shared among nodes. It devises a utility function that captures the future co-location of the node (with others sharing the same interest) and the change in its connectivity degree. Thus, the utility function measures how good message carrier a node can be regarding a given interest. *SocialCast* functions based on the publish-subscribe paradigm, where users broadcast their interests, and content is disseminated to interested parties and/or to high utility new carriers.

The fact that users' interests are considered allows *SocialCast* to achieve a good delivery rate with very low associated cost and stable latency. However, its performance depends on the co-location assumption (in which nodes sharing interests spend longer times together), which has been proven to not always hold true [11].

Besides taking into account the users' interest on the content, *ContentPlace* [18] also considers information about the users' social relationships to improve content availability. For that, a utility function is computed for each data object by a node considering the access probability of the object and the involved cost in accessing it, and the social strength of the user towards the different communities which he/she belongs to and/or has interacted with. The idea is to have the users to fetch data objects that maximize the utility function with respect to local cache limitations, choosing those objects that are of interest to him/herself and that can be further disseminated in the communities they have strong social ties.

As *SocialCast, ContentPlace* manages to be rather efficient in terms of delivery and resource utilization, and it relies on the assumption that users spend time co-located with their friends, which may not always be true (e.g., strong ties between geographically dispersed nodes, such as good friends living in different countries).

2.1.2.3 Popularity-based Routing

Popularity-based routing takes advantages of the knowledge regarding the popularity of nodes inside the network to decide about forwarding. Some authors may argue that this category should include centrality-based routing as mentioned in Sec. 2.1.2.1 since such algorithms do measure the importance of nodes [49]. However, the popularity (i.e., importance) of nodes here is given by actual social interaction existing among nodes, and not by the number of times these nodes find themselves in the shortest paths between pairs of nodes. As an example of this approach, I highlight *PeopleRank*.

PeopleRank [11] uses social interaction to rank nodes, and forwardings take place based on such ranking. The approach is to consider socially well-connected nodes to choose best next forwarders. The ranking process is analogous to Google's page rank system, in which the relative importance of a Web page is determined according to its links to/from a set of pages.

By considering well-connected nodes, the proposal reaches high delivery rate as well as low cost and delay. However, the proposal does not consider the socially disconnected nodes as they may also be used to improve the routing performance. This suggests that some level of randomness while choosing next hops could also be applied to increase performance.

2.1.2.4 Dynamic Behavior-based Routing

It has been shown that people's routines can be rather useful to determine future behavior [50], and that considering the dynamics of social ties (based on an analysis of contact duration) from different daily routines is important to achieve a correct mapping of real social interactions into a clean social representation able to aid data forwarding [51]. Thus, dynamic behavior-based routing considers what is constantly happening among nodes in terms of social interactions to devise more elaborate forwarding schemes. The *Context Information Prediction for Routing in OppNets* (*CiPRO*) employs such routing approach.

CiPRO [17] considers the time and place where nodes meet throughout their routines. *CiPRO* holds knowledge of nodes (e.g., carrier's name, address, nationality, battery level, memory of the device) expressed by means of profiles that are used to compute the encounter probability among nodes in specific time periods. Nodes that meet occasionally get a copy of the message only if they have higher encounter probability towards its destination. If nodes meet frequently, history of encounters is used to predict encounter probabilities for efficient broadcasting of control packets and messages.

CiPRO has a performance trade-off between delivery probability/cost and delay, and perhaps introducing some randomness, in the occasional contact case, could bring more improvements to the solution.

2.1.2.5 Summary

This section briefly presented the social-aware opportunistic routing proposals considered as I surveyed the state-of-the-art literature. The same approach for social-oblivious proposals was employed regarding my choices for most of these proposals: the number of times they served as benchmarks and their evaluation against other benchmark proposals.

It is clear that social similarity has a great potential in improving opportunistic routing, and indeed it is an interesting direction to consider while devising forwarding solutions. However, one aspect that should be further explored is the dynamism that can be derived from users' daily routines. As observed by Hossmann et al. (2010) [51], if social similarity metrics are able to capture the dynamic behavior of users, the formed connectivity graphs are going to be built reflecting the most important social edges.

Another aspect that is taken for granted is the way communication takes place: it is always based in a source-destination pair. Point-to-multipoint communication has shown its capabilities (i.e., better performance and wise use of resources) when applied to opportunistic routing as it increases the reachability of information [52].

Based on these observations, I develop my work to achieve the goal of proposing novel opportunistic routing solutions.

2.2 Opportunistic Routing Taxonomies

Despite of not existing many classifications, the existing ones look at the opportunistic routing approaches from various perspectives, i.e., based on: i) message delivery strategy (single copy, flooding, replication); ii) employed level of node/network knowledge; iii) usage of oracles; iv) node/network behavior (deterministic and stochastic); v) one-hop/end-to-end information; vi) existence of special devices to aid routing (stationary and mobile).

Another observation is that these classifications consider aspects (e.g., level of knowledge) that lead to an unbalanced classification, assigning most proposals to a few set of categories, or to specific categories (e.g., information coding) that could otherwise be seen as methods and could be applied orthogonally to other categories. In addition, such classifications either do not consider the social trend as a category, or simply see it as mere subcategory under unrealistic scenarios (e.g., with deterministic behavior).

This led me to further analyze these opportunistic routing classifications and consequently propose a more up-to-date taxonomy. This section briefly goes over the most relevant classifications, from 2004 to 2010, that helped me come up with a cleaner and more concise classification that includes social similarity, and strengthen its importance amongst the already well-established routing categories.

The first taxonomy for opportunistic routing was proposed by Jain et al. (2004) [22] encompassing three classifications based on: knowledge from network oracles (contact summary, contacts, queuing, and traffic demand), route computation (proactive and reactive), and determination (source- and hop-based). The former classification is the most important and represents oracles that provide summarized/detailed information about contacts, buffer utilization, and present/future traffic demand. The level of knowledge can be zero or partial/complete (by combining oracles).

Since opportunistic contacts are difficult to accurately predict, considering network knowledge is cumbersome as it needs a central entity (i.e., oracle) in the network and the delay for gathering/processing information into useful knowledge may be too high specially in scenarios where contacts are short lived. Regarding route computation/determination, most of the solutions fall into the reactive and hop-by-hop branches. Thus, this taxonomy does not provide a well-balanced classification for opportunistic routing proposals.

Zhang (2006) [53] provides a classification with two main categories: deterministic (i.e., nodes are aware of network topology), and stochastic (i.e., node/network behavior is random and unknown). In addition, Zhang classifies proposals according to their ability of controlling (or not) node movement. One can observe that Zhang's classification improves the proposal of Jain et al. (2004) as a more realistic branch emerges (i.e., stochastic) and is able to best represent the behavior of opportunistic contacts. But it is important to note that there are some categories in Zhang's classification that can be considered orthogonal (i.e., coding-based approaches, where coding schemes can be applied elsewhere) and do not represent real life deployment (i.e., control movement, since only in specific application scenarios nodes can change other nodes movement patterns simply to answer their needs to send/receive data).

Balasubramanian et al. (2007) [44] present two classification criteria based on routing strategy (i.e., replication or single-copy forwarding) and the effect on performance metrics (i.e., which consider resource constraints - intentional, or not - incidental). Like the proposal by Jain et al. (2004), this classification is not well balanced as most of the solutions fall into the replication and incidental branches.

Song and Kotz (2007) [39] and Nelson et al. (2009) [9] also classify opportunistic routing according to routing strategy (single-copy forwarding/replication). In the former, the replication branch is further divided into considering its effect on resource consumption and on delivery probability. Nelson et al. (2009) further divide replication into flooding- and quota-based (i.e., controlled flooding). This classification is interesting since it highlights that replication is not the cause for network flooding, but the scheme used to perform it. However, these classification models are incomplete as they do not include categories based on metrics such as encounter number, resource usage, or social aspects.

D'Souza and Jose (2010) [54] come up with three major categories: i) flooding-based, where network is flooded with messages to increase their delivery probability, or that apply a controlled flooding by limiting the copies of messages, and by embedding additional information into messages in order to reduce flooding effects; ii) history-based, where past encounters are considered to improve routing decisions; and, iii) special devices-based, where specific devices (stationary or mobile) aid communication among nodes. This latter category may use social interaction among nodes to perform routing decisions.

Just like Zhang (2006), D'Souza and Jose (2010) categorize aspects that can be seen as orthogonal (e.g., network and erasure coding) to any other category. However, this classification is the only one that somewhat considers social aspects, despite the fact that social-aware approaches have been around since 2007. Yet, D'Souza and Jose (2010) include this category under their special devices-based branch, which surely does not comply with the regular behavior (i.e., random and unknown) found in opportunistic networks: the use of special stationary/mobile devices to improve data exchange is not realistic, as the network/nodes will have to present a deterministic behavior in order to correctly place these devices in the system.

Spyropoulos et al. (2010) [55] divide the opportunistic routing proposals based on the employed message exchange scheme: forwarding (similar to single-copy); replication (full and controlled replication based on utility functions); and coding (messages are coded at the source or as they traverse the network). Additionally, the different forwarding and replication utility functions are identified and categorized according to their dependency on the destination. Finally, the authors classify Delay-Tolerant Networks based on a set of characteristics (e.g., connectivity, mobility, node resources, and application requirements) that impact routing and map the suitable routing solutions to the different types of DTNs considering their characteristics. Still, in this classification, categories that can be orthogonal (i.e., coding) continue to appear, and social aspects are seen as a mere destination-dependent function, and not as a potential research direction which includes social relationships, interests, and popularity to improve opportunistic routing.

Fig. 2.1[1] presents a global view of the analyzed classifications proposed between 2004 and 2010. By looking at the existing opportunistic solutions, it is clear that the classification model should focus solely on the stochastic branch. First, because the deterministic one is based on available knowledge from the network state, which is very difficult to get given the nature of opportunistic contacts; and second, solutions are trying to use the least knowledge possible to reduce complexity. Furthermore, the classification model should include categories of routing aiming at achieving an optimum balance between delivery probability (e.g., replication) and resource utilization (e.g., forwarding). This is not seen in current classifications.

Thus, the stochastic branch should include a category for the trend identified since 2007, comprising

[1] Springer and the original publisher (Routing in Opportunistic Networks, v. 1, 2013, p. 27-68, Social-Aware Opportunistic Routing: The New Trend, Waldir Moreira, Paulo Mendes, Figure 2.3, Copyright © 2013, Springer Science+Business Media New York) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media [1].

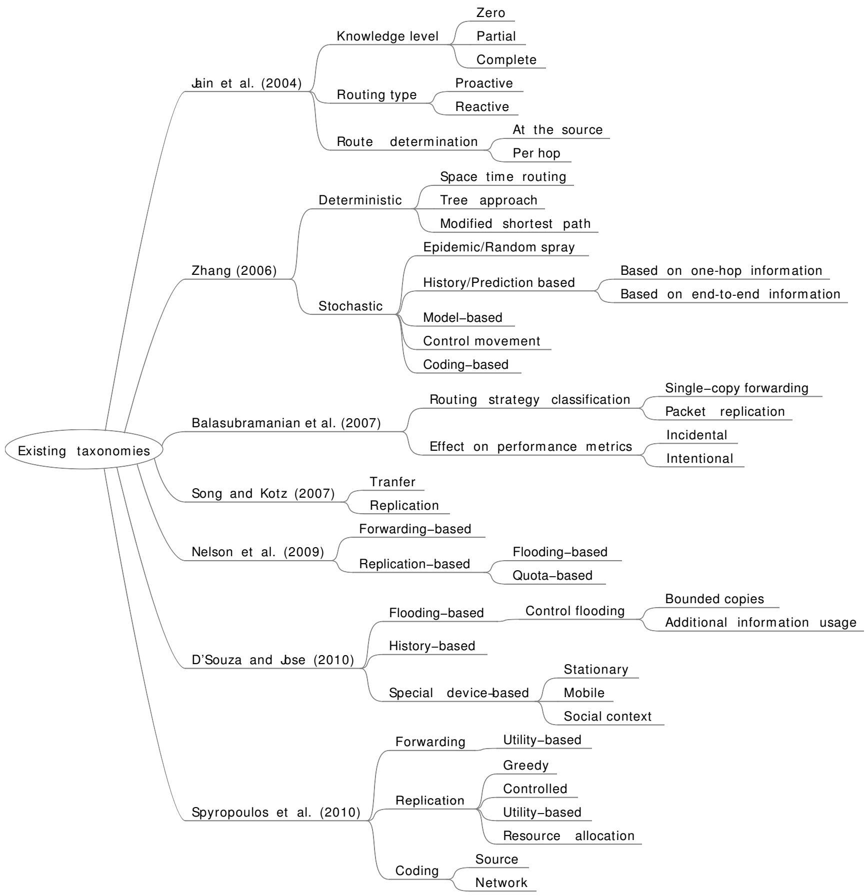


Figure 2.1: Opportunistic routing taxonomies from 2004 to 2010

proposals that aim at achieving a good balance between delivery probability (e.g., message replication) and usage of network resources (e.g., forwarding) by employing the social similarity approach.

Finally, given the vast range of proposals and classifications, another observation is that comparison evaluation does not always consider either a homogeneous set of parameters or comparable experimental setups, which endangers the veracity of conclusions.

2.3 Evaluation Frameworks

The increasing capability of portable devices provide users with new forms of communication. They can quickly form networks by sharing resources (i.e., processing, storage) to exchange information, and even connectivity. This is possible through opportunistic contacts among nodes that can carry/forward information on behalf of other nodes to reach a given destination or connectivity points.

As presented in Sec. 2.1, different opportunistic routing proposals have emerged. Still, there is not a clear understanding of what the best solutions are, given the different application scenarios. Moreover, my analysis of the state-of-the-art literature shows that comparison evaluation between these proposals does not always consider either a homogeneous set of parameters or comparable experimental setups, which endangers the veracity of performance assessment. Thus, even with a stable taxonomy, there is the need to determine an evaluation model based on common performance metrics (e.g., delivery rate and cost, delay, and energy efficiency) and experimental scenarios.

Regarding evaluation frameworks, I highlight the proposal based on *Evolving Graph* (EG) theory to design/evaluate least cost routing protocols. EG provides a formal abstraction for dynamic networks and reflects the different connectivity graphs in the time domain by considering node mobility. The result is that connectivity of links are transcribed into subgraphs for different instant in time.

Thus, Ferreira et al. (2010) [56] take into consideration one of the formalized EG criteria (i.e., foremost) to determine journeys (i.e., future temporary connections between nodes that can form a path over time) in which data can quickly reach its destination. This evaluation framework provides designers with an algorithm that is able to reach good performance in scenarios where connectivity patterns are known beforehand. Additionally, the algorithm can be used as lower bound reference to compare opportunistic routing solutions.

The work of Spyropoulos et al. (2010) [55] also provides principles to help developers in designing routing solutions based on their classification of opportunistic routing, identified utility functions, and DTNs characteristics. The authors show that, by knowing the application characteristics and requirements, the choice/design of routing solutions is eased.

After a thoroughly analysis of different routing solutions under the uni-, multi-, and anycast perspectives, Cao and Sun (2013) [57] define a framework considering one of the following major goals that routing proposals may target: i) effectiveness and efficiency, concerning a balanced tradeoff between delivery/delay and cost; ii) QoS awareness and security, referring to QoS requirements and attack prevention; iii) scalability, aiming at the ability of coping with varying network density.

Still, these works lack a guideline of how performance metrics and experimental setups can be used. Thus, there is the need to provide a set of experimental setups to aid designers in fairly assessing the performance of existing and their yet-to-come opportunistic routing solutions in comparison studies.

With these observations in mind regarding the existing routing proposals, taxonomies and frameworks, I was able to identify the opportunistic routing metrics as well as classify the proposals themselves in a clean and concise taxonomy. Then, I devised an evaluation model based on common performance metrics (e.g., delivery rate and cost, delay, and energy efficiency) and experimental scenarios. My goal is to avoid future performance assessments considering irrelevant performance metrics and biased scenarios. Finally, I could come up with social-aware utility functions and a couple of novel opportunistic routing solutions. Chapters 3 through 7 present the outcomes of my close analysis to the state-of-the-art literature and this work.

Chapter 3

Classification and Evaluation of Opportunistic Routing

As one can observe in Chapter 2, a more up-to-date classification for opportunistic routing proposals is needed in order to suitably include the social-aware trend. Additionally, no guidelines for evaluating opportunistic routing proposals are available in the prior art. Consequently, proposals were being compared under conditions different from which they have been designed for, jeopardizing the assessment of their performance behavior [25, 1].

This chapter goes over the first set of contributions of this Thesis work: an updated taxonomy including the social branch amongst the existing categories of opportunistic routing proposals [1], a Universal Evaluation Framework to allow proposals to be evaluated under the same conditions aiming at a fair comparison assessment [27, 28], and a study done regarding the importance of cooperation when it comes to the support for opportunistic networking [29, 30, 31].

3.1 Taxonomy

Observing the considered opportunistic routing proposals (cf. Sec. 2.1), different trends based on specific goals are identified. Thus, there is the need for a well-balanced and updated taxonomy, which includes the social-aware trend identified over the last years and that focuses solely on the branch (i.e., stochastic) that best reflects opportunistic contacts (cf. Sec. 2.2).

Social similarity metrics [10, 12, 14, 49, 18, 11, 15, 17] brought a new perspective into opportunistic routing: as devices are carried by humans, forwarding decisions are done considering characteristics of the relationships among them (e.g., contacts with other people, time spent with these people, the level of relationship between people, among others). The potential of these metrics are based on the connectivity graphs they form, which are less variable than those based on mobility. Edges represent people's socially meaningful relationships that evolve from node mobility, interaction and social structures.

Generally speaking, every opportunistic routing proposal considers node mobility and the resulting interaction to decide on forwarding: some proposals (e.g., *Epidemic* [4], *Direct Transmission* [38]) are rather simple and use the resulting contacts to reach the destination, while others are more elaborate and consider social aspects in order to find the destination (e.g., *Label* [10], *PeopleRank* [11]). Hence, one shall notice that the proposed taxonomy does not have a mobile-based category (e.g., *Model-* or

Control Movement-based in Zhang (2006) [53] and *Mobile Device-based* in D'Souza and Jose (2010) [54]) as this is an inherent feature of opportunistic proposals.

The proposed taxonomy considers three categories, namely forwarding-, flooding-, and replication-based. The most straightforward category is the forwarding-based, which refers to the single-copy forwarding (e.g., *MEED*, and approaches in Spyropoulos et al. [38]) where only one copy of each message traverses the network towards the destination.

As presented in Sec. 2.2, depending on the level of replication, opportunistic routing proposals can be categorized as: replication [44, 39]; flooding- and quota-based replication [9]; greedy, controlled, utility-based, and resource allocation replication [55]. One can easily observe that these categories are either too generic [44, 39], or end up mixing epidemic (i.e., flooding-based and greedy replication) with other approaches that have different levels of replication [9, 55].

Thus, the proposed taxonomy devotes a specific category to flooding-based proposals, since only proposals that allow every node to spread a copy of each message to every encountered node (e.g., *Epidemic* [4]) are considered. Additionally, this category is left out of the replication class due to the fact that i) having (or not) the quota-based feature (i.e., where the number of created copies does not depend on the number of network nodes [9]) can be found in the different proposals; and ii) proposals (e.g., *PREP* [45], *MaxProp* [8], *PROPHET* [5], and *RAPID* [44]) cannot be simply categorized under the flooding-based category just because certain conditions (e.g., mobility, limited resources) may make them have an epidemic-like behavior [9, 55].

The last category is the replication-based one, which comprises different opportunistic routing proposals and metrics, and is further divided into the following sub-categories: encounter-based, resource usage, and social similarity.

In the encounter-based sub-category, replication takes place based either on: i) *frequency encounters*, where proposals consider the history of encounters with a specific destination to support opportunistic forwarding of messages, or the frequency nodes met in the past, to predict future encounters (e.g., *PROPHET*, *MaxProp*, *Prediction,* and *EBR*); and, ii) *aging encounters*, where proposals consider the time elapsed since the last encounter with the destination to decide about next hops (e.g., *FRESH*, *EASE*, and *Spray and Focus*).

In the sub-category about resource usage, forwarding decisions are made considering the: i) *aging of messages*, where proposals aim to avoid messages to be kept being forwarded in the network by creating metrics that define the age of message copies (e.g., *Spray and Wait*, and *OPF*); and, ii) *resource allocation*, in which knowledge about local resources (e.g., *PREP*, and *RAPID*) are taken into account.

The sub-category about social similarity is more elaborate and exploit the social behavior existing among nodes. It comprises: i) *community detection* proposals that group nodes according to their affiliations, number of contacts, and duration of such contacts (e.g., *SimBet*, *Label, Bubble Rap*); ii) *shared interest* proposals that decide on forwarding based on the interest nodes have in the content (e.g., *SocialCast* and *ContentPlace*); and iii) *node popularity* proposals, where nodes are ranked according to their social interaction and this raking is used to decide on replication (e.g., *PeopleRank*). Fig. 3.1[1] shows the taxonomy proposed in 2011 by Moreira et al. [27, 28].

[1]Springer and the original publisher (Routing in Opportunistic Networks, v. 1, 2013, p. 27-68, Social-Aware Opportunistic Routing: The New Trend, Waldir Moreira, Paulo Mendes, Figure 2.4, Copyright © 2013, Springer Science+Business Media New York) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media [1].

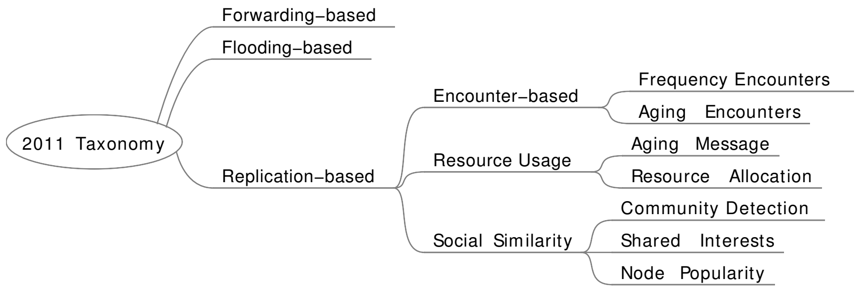


Figure 3.1: Taxonomy proposal for opportunistic routing (2011)

As new social-aware opportunistic routing proposals (i.e., *dLife* [15], *CiPRO* [17]) emerged, this taxonomy had to be updated to include a new sub-category: *user dynamic behavior* as illustrated in Fig. 3.2[2].

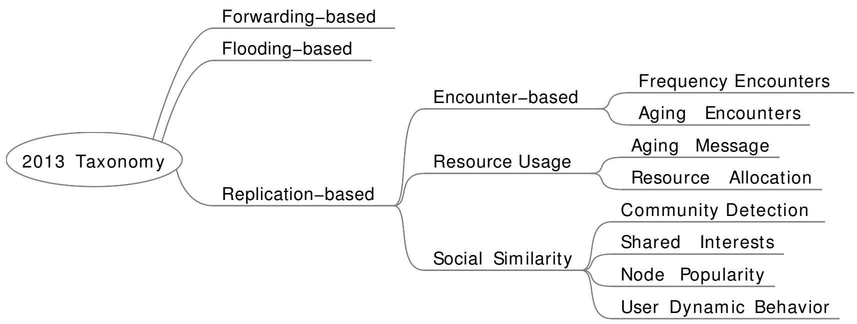


Figure 3.2: Taxonomy proposal for opportunistic routing (2013)

It is important to mention that every category included in the taxonomy has advantages and disadvantages. Yet, the main goal with this new taxonomy was not to identify the most suitable opportunistic routing category, but instead to: i) to emphasize the importance of the new trend based on social similarity; and ii) update a previously proposed taxonomy with a new sub-category, namely the *user dynamic behavior*, based on the latest social-aware opportunistic routing proposals.

3.2 Universal Evaluation Framework

The motivation to devise a Universal Evaluation Framework (UEF) is that current opportunistic routing proposals - no matter if they are based on flooding, forwarding or replication, or if they consider or not levels of social relationship - do not always take into account a similar set of performance metrics and comparable experimental scenarios.

Thus, even with a suitable taxonomy, the UEF aims at a way to fairly evaluate opportunistic routing proposals based upon realistic scenarios (i.e., able to represent the real requirements of applications) and based on common metrics such as delivery rate, cost, and delay. Additionally, with this UEF, designers are expected to avoid the creation of future proposals comprising irrelevant performance metrics, evaluated in specific scenarios and without proper benchmarks.

[2]Springer and the original publisher (Routing in Opportunistic Networks, v. 1, 2013, p. 27-68, Social-Aware Opportunistic Routing: The New Trend, Waldir Moreira, Paulo Mendes, Figure 2.5, Copyright © 2013, Springer Science+Business Media New York) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media [1].

To support the creation of a proper UEF, a close analysis of the state-of-the-art literature (cf. Sec. 2.1) was considered in what concerns evaluation benchmarks, metrics and scenarios. The proposed UEF spans: i) network type, size, and resources; ii) number of nodes, the relationship between them, their interest, available resources, and willingness to participate in communication; and iii) performance metrics, such as delivery rate, cost, and delay [27, 28].

3.2.1 Benchmarks

In order to come up with a suitable UEF proposal, I first identified the benchmark proposals, that is, how often proposals are referenced for comparison purposes by others. Fig. 3.3 shows the proposals in rectangulars, with the outgoing arrows indicating other proposals that were used as benchmark to evaluate them. The incoming arrows illustrate the importance that a proposal has as a benchmark to other related proposals.

From Fig. 3.3 one can see that there are no rules for comparing proposals, whereas authors rely on proposals that have clearly worse performance in relation to their own, given the considered conditions.

It is important to note that the proposals/methods presented in the ellipses are out of the scope of this Thesis work and mostly refer to methods commonly found in the literature (e.g., Random, Wait) or to extensions of the studied proposals (e.g., *GREASE* and *dLifeComm* in the *EASE* and *dLife* proposals, respectively).

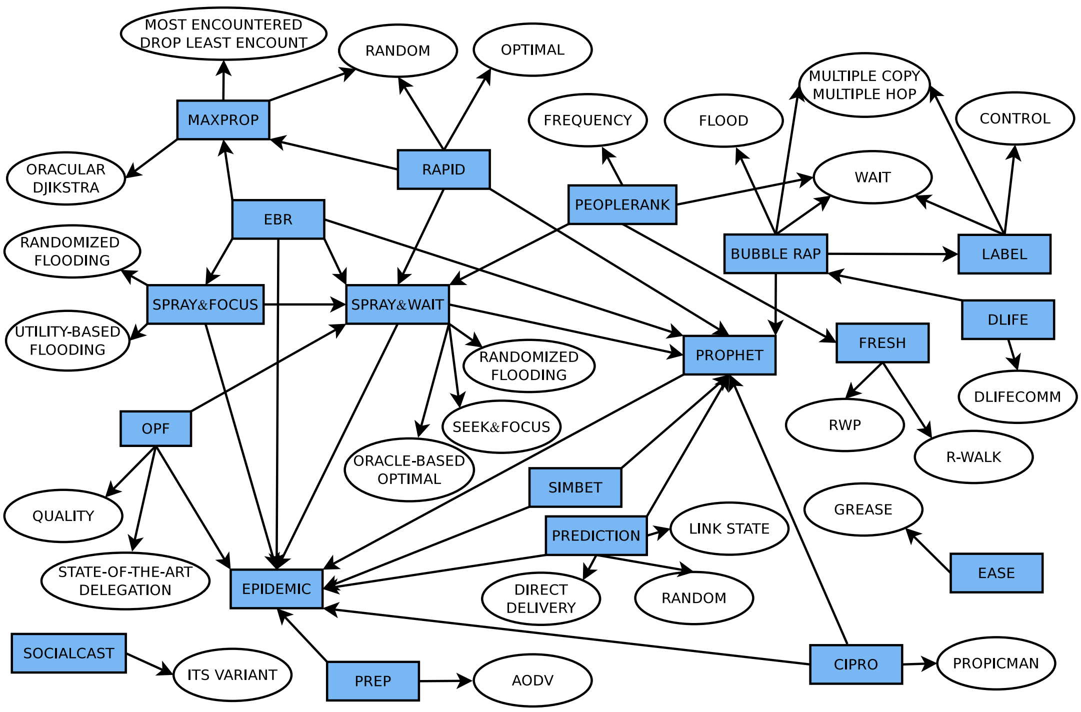


Figure 3.3: Analysis of most common benchmarks

Tables 3.1, 3.2, and 3.3 summarize Fig. 3.3 showing the proposals grouped according to the proposed taxonomy presented in Sec. 3.1. By reading the tables vertically, one can observe how solid is the evaluation of a proposal (naming the column) in what concerns the most relevant related work, while horizontally the tables provide information about how often the proposal has served as benchmark for comparison studies. In order to improve readability, the tables do not include proposals (i.e., in columns) that disregarded the considered related work from their evaluation (i.e., *MaxProp*, *FRESH, EASE, Label*, and *SocialCast*).

From these tables, one can see that the proposals that were compared to the highest number of related work were *EBR* and *RAPID*, compared to five and three other proposals, respectively. For instance, *EBR* is an example of a proposal with a solid evaluation benchmark (cf. Table 3.1), since it was compared to two out of three other related work in its category (*PROPHET* and *MaxProp*), and to two of the three most used benchmarks (*Epidemic,* and *Spray and Wait*).

Epidemic, *PROPHET*, and *Spray and Wait* are the proposals that have been used most often as benchmark, i.e., nine, seven, and five, respectively.

Table 3.1: Taxonomy-benchmark relationship for encounter-based approaches

	Encounter-based			
	Frequency Encounters			Aging Encounters
	PROPHET	Prediction	EBR	Spray and Focus
Epidemic	•	•	•	•
PROPHET		•	•	
Spray&Wait			•	•
MaxProp			•	
Spray&Focus			•	

Table 3.2: Taxonomy-benchmark relationship for resource usage approaches

	Resource Usage			
	Aging Messages		Resource Allocation	
	Spray and Wait	OPF	PREP	RAPID
Epidemic	•	•	•	
PROPHET	•			•
Spray&Wait		•		•
MaxProp				•

Table 3.3: Taxonomy-benchmark relationship for social similarity approaches

	Social Similarity				
	Community Detection		Node Popularity	User Dynamic Behavior	
	SimBet	Bubble Rap	PeopleRank	*dLife*	CiPRO
Epidemic	•				•
PROPHET	•	•			•
Spray&Wait			•		
FRESH			•		
Label		•			
Bubble Rap				•	

With this benchmark analysis, I observed that proposals such as the flooding-based *Epidemic* are compared to more elaborate proposals. In some cases (e.g., *PREP*), *Epidemic* is used only to set the upper bound in terms of delivery rate performance. However, in other cases (e.g., *Prediction*, *EBR* and *CiPRO*), *Epidemic* is exposed to different assumptions and conditions (e.g., different number of nodes, message size, buffer size) from which it has been devised, and considering different performance metrics (e.g., composite metrics in *EBR*, network overhead in *CiPRO*). This consequently leads to unfair performance assessment of this proposal.

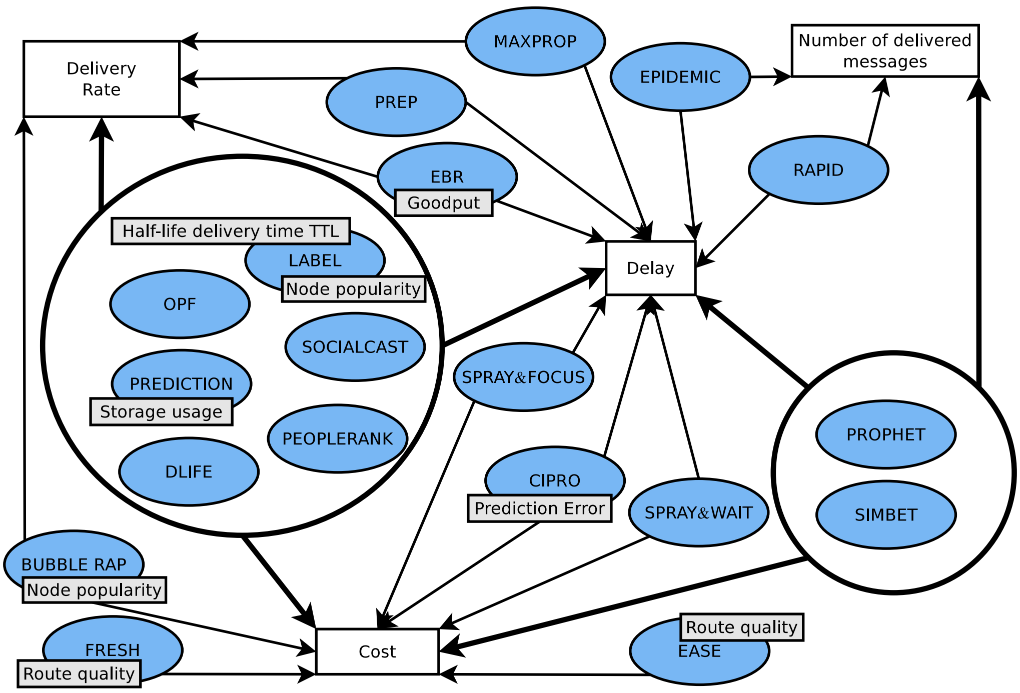


Figure 3.4: Analysis of performance metrics

3.2.2 Performance Metrics

Once the relationship among the benchmark proposals in terms of comparison evaluation was understood, I carefully looked at the performance evaluation metrics employed in their assessment. The goal here is to make sure that evaluations based on the UEF consider similar performance metrics and the same evaluation conditions in order to fairly assess the performance of the proposals.

Fig. 3.4 summarizes the identified performance metrics. To improve readability, the figure uses general terms (i.e., delivery rate, cost, delay, number of delivered messages) to identify the most used metrics (wide rectangulars), since terminology varies according to the proposal (e.g., while *RAPID* considers the delay metric as the worst-case delay, *Epidemic* refers to it as message delivery latency). Additionally, when applicable, proposals have been grouped (e.g., *SimBet*, and *PROPHET*) as they consider the same set of metrics. Finally, the narrow rectangulars depicted in the figure represent other metrics relevant to few of these proposals.

It is important to mention that, since they appeared after this performance metric analysis, *dLife* and *SCORP* already consider the performance metrics as they are defined next, and *dLife* has been included in the presented figures solely for illustration purposes.

Based on this analysis, one can conclude that the most important metrics for opportunistic routing are: delivery rate, cost, delay and number of message delivered. In what concerns the delivery rate, it is defined as the *number of messages that have been delivered out of the total number of messages created.* This metric is used in ten out of the nineteen proposals under different terminology such as: number of messages that have been delivered out of the total of messages created (e.g., *Label*), the fraction of sourced bundles that are delivered to the destination (e.g., *PREP*), the ratio of the number of messages delivered to the number of total messages generated (e.g., *Prediction*), the proportion of messages that have been delivered out of the total unique messages created (e.g., *Bubble Rap*), the ratio between the actual number of messages delivered to the interested subscribers and the ideal one (e.g., *SocialCast*), the message delivery ratio (e.g., *EBR*), simply delivery rate (e.g., *MaxProp* and *OPF*), or the success rate of the algorithm normalized by the success rate of flooding within a delay period (e.g., *PeopleRank*).

The delivery rate is a very important metric to be considered, since it represents the effectiveness of the proposal. Usually, proposals are classified as having good performance if they achieve high delivery rates, i.e., they can delivery as many messages as possible to destinations in a useful time frame.

Depending on the approach, the cost metric can also be associated to the overhead necessary to build a route from a source to a destination (e.g., *search cost* in *FRESH*), the total number of transmissions required to forward/search in order to deliver a message (e.g., *relative cost of routes* in *EASE*), as well as the distribution of the number of hops needed for all the deliveries (e.g., *hop distribution* in *Label* and *Bubble Rap*), and the average number of hops per message (e.g., *SimBet*).

However, the effectiveness of a proposal has to be weighted against the cost associated to the message delivering process.

In this analysis, the cost of opportunistic routing is defined as the *number of replicas per delivered message*. This metric is used in fourteen out of the nineteen analyzed proposals under different instantiations, such as: the number of forwarded messages (e.g., *PROPHET* and *SocialCast*), the number of transmissions per delivered message (e.g., *Spray and Wait*), the number of transmissions (e.g., *Spray and Focus*), the total number of messages transmitted across the air (e.g., *Label* and *Bubble Rap*), the number of times a message copy occurred due to replication (e.g., *Prediction*), the total number of forwards (e.g., *SimBet*), the number of forwardings (e.g., *OPF*), the number of retransmissions (e.g., *PeopleRank*), or the network overhead (e.g., *CiPRO*).

At a first glimpse, looking at delays on scenarios that are susceptible to high delays may sound contradictory. Nevertheless, each message has a time-to-live that is correlated to its utility. Moreover, it is important to remove already delivered messages as soon as possible from the network in order to avoid waste of resources. Thus, making sure that messages reach their destination within a useful time frame is also important from the viewpoint of the performance. Based on Fig. 3.4, it is clear that most of the routing proposals consider delay as a performance metric, more precisely sixteen out of nineteen proposals.

In this analysis, delay is defined as the *time required to deliver all the bytes encompassing a message since its creation*. Delay is also seen differently depending on the proposal, and it can be considered as the message delivery latency (e.g., *Epidemic*), the message delivery delay (e.g., *PROPHET*, *Spray and Wait* and *CiPRO*), the latency of delivered packets (e.g., *MaxProp*), the delay distribution (e.g., *Label* and *PeopleRank*), the delivery delay (e.g., *Spray and Focus*), the fraction of bundles that are delivered within a given delay bound (e.g., *delay CDF* in *PREP*), the average delay and worst-case delay (e.g., *RAPID*), the duration between the generation time and delivery time of a message (e.g., *Prediction*), the average end-to-end delay (e.g., *SimBet*), the average latency (e.g., *SocialCast*), the average delay (e.g., *EBR*), or simply delay (e.g., *OPF*).

In summary, one can conclude that guaranteeing a low cost to maximize delivery rate with low delay seems to be highly desired for opportunistic routing solutions, since this guarantees that end users have access to a significant amount of useful information with a good usage of network resources. This is reflected in the number of proposals that consider delivery rate (ten), cost (fourteen) and delay (sixteen) as performance metrics, as seen in Fig. 3.5.

Besides these three metrics, a small number of proposals (four) also consider the *number of delivered messages*, as a performance metric under different terminology such as: percentage of delivered messages

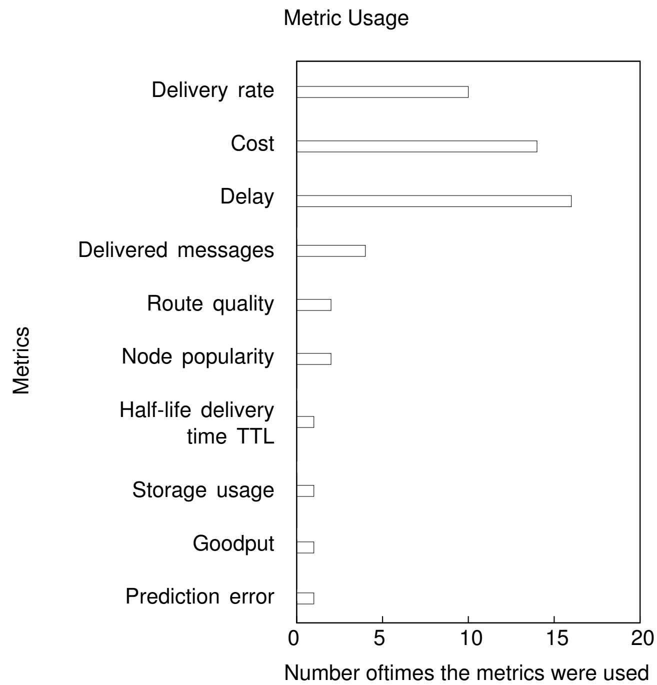


Figure 3.5: Metrics usage for performance evaluation

(e.g., *Epidemic*); number of received messages (e.g., *PROPHET*); number of packets delivered before a deadline (e.g., *RAPID*); total number of messages delivered (e.g., *SimBet*). However, since the number of delivered messages is directly related to the delivery rate, it should not be considered as major evaluation metric.

Also in Fig. 3.5, one can observe six other metrics that are only used to evaluate specific proposals: *route quality* used by *FRESH* and *EASE*; *node popularity* used by *Label* and *Bubble Rap*; *half-life delivery time TTL* used by *Label*; *storage usage* used by *Prediction*; *goodput* used by *EBR*; and *prediction error* by *CiPRO*. The *route quality* metric is defined as the difference between the route found by *FRESH/EASE* and the route with shortest number of hops. The *node popularity* metric is the number of unique contacts between a node and the others in the network. The *half-life delivery time TTL* metric is the TTL value that would allow half of the messages to be delivered. The *storage usage* metric is the maximum storage (in bytes) used across all nodes, and the *goodput* metric is the number of messages delivered divided by the total number of messages transferred (including those transfers that did not result in a delivery). The *prediction error* refers to the percentage of error in predicting encounter probability

From these six metrics that are specific to some proposals, *node popularity*, *half-life delivery TTL*, and *goodput* are the most interesting ones, since they can identify nodes connecting different node clusters, and guarantee message delivery within a useful time frame.

As last remarks, one can easily observe the lack of a convention regarding the terminology used for performance evaluation metrics, and that proposals are seen from different perspectives regarding such metrics. This leads to a difficult evaluation process and provides no fairness when comparing proposals from different categories.

3.2.3 Evaluation Scenario

The lack of convention observed from the analysis of performance metrics is also evident in what concerns the used evaluation scenarios. Therefore, there is the need to create a set of guidelines to be considered when creating experimental scenarios.

Tables 3.4 and 3.5 summarize experimental setup considered by the proposals under study for performance assessment, such as the number of nodes, the number of source and destination pairs, meeting time (i.e., contact time) and time between such meetings (i.e., inter-meeting time), area size (also referred to as network density, and expressed in number of nodes in the surface area), message size, network load (i.e., number of generated messages), message TTL (expressed either in hops or time units), size of node buffer, mobility model and node speed, transmission range, and beacon usage.

From these tables, one can observe that: i) experimental setup can display great differences among proposals; and, ii) few proposals (e.g., *Epidemic*) provide detailed information. Thus, there is the need for defining guidelines to derive a default scenario setup to be used while comparing/evaluating proposals.

Two sets of parameters can be identified, concerning:

- network density (e.g., network area, number of nodes, mobility model, node meeting and inter-meeting times, transmission range and beacon control); and

- traffic (e.g., distribution of sources and destinations, load generation, message size, message TTL, and buffer size).

Table 3.4 shows the different network density parameters considered by some of the studied proposals. This set of parameters allows protocol designers to understand the behavior of proposals in sparse (i.e., sporadic contacts resulting in high delay) and dense (i.e., frequent contacts exploring the ability to choose the best next hops) scenarios.

Network density may be configured by using three parameters: network area, number of nodes, and mobility models. Regarding the number of nodes, one can observe from Table 3.4 that the trend is to consider a roughly average number between 100 and 150 nodes (excluding extreme cases such as *FRESH*, *EASE*, and *Prediction*). This number of nodes may be enough to configure dense or sparse scenarios depending on the considered network area and node mobility model. The network area can span a conference building as well as a city area, so designers must consider these extreme cases in order to better assess the quality of their proposals. By considering different areas, along with different mobility models, the challenge faced by algorithms increases, since different levels of sparseness will emerge as simulations run.

Since a normal assumption to have is that most of the mobile nodes are carried by people, experimental scenarios should consider realistic human mobility models. Random models may not be suitable for that, since it has been proven that humans do have a pattern in mobility behavior [58, 50, 59]. Moreover, people are part of communities [2] that represent their social relationship with others, considering interests and tastes [14, 11]. Hence, considering the relation between mobility models and social interaction within a society seems to be a good method for assessing the performance of a given algorithm.

Table 3.4: Network density parameters considered by the different proposals

Proposals	Area Density (Km^2/# of nodes)	Mobility Model/ Speed (m/s)	Node Meeting/ Inter-meeting Time	Transmission Range (m)	Beacon Control
Epidemic	0.45/50	Node chooses a point and walks there (0 – 20)	Pairs of hosts come in contact periodically and randomly	10 to 250	Internet MANET Encapsulation Protocol
PROPHET	0.45/50 (RWP), and 4.5/50 (CM)	Random Waypoint Mobility (0-20) and Community (10-30)	Pairs of hosts come in contact periodically and randomly	50 and 100	
Spray &Wait	0.25/100 (Traffic load) 0.04/200 (Connectivity)	Random Walk Mobility	Exponentially distributed meeting time	5 to 35	
Spray &Focus	0.04/100	Random Walk, Random Waypoint, and Community based Mobility		5 to 35	Nodes periodically transmit beacons to recognize each other's presence
PREP	9/25	Random Waypoint Mobility (5 – 15)		250	Hello protocol
RAPID	388/20 and 40		Meeting time distributions are exponentially distributed / Inter-meeting time between nodes follow either an exponential or a power law distribution		Scans for other buses 100 times a second
Prediction	0.81/5142	Mobility traces from CRAWDAD		Nodes could discover and connect each other instantly when they were associated with a same AP	
SimBet	Lab and College area/100	Human traces		Bluetooth	
Bubble Rap	Infocom05/41 Hong-Kong/37 Cambridge/54 Infocomo6/98 Reality/97	Human traces	Inter-meeting time follows a power-law distribution	Bluetooth	
EBR	15/26, 51, and 101 (VMM) 9/26, 51, and 101 (REDMM, and RWP)	Vehicular-based Map-driven (2.7 – 13.9) Role-based, Event-driven Disaster Mobility (1 – 20) Random WayPoint (0.5 – 1.5)		250	
OPF	Unknown/300 and 40	Human and Vehicular			
PeopleRank	MobiClique/27 SecondLife/150 Infocom/65, 47, and 62 Hope/414		Median Contact time: 90, 150, 180, and 240s Median Intercontact time (10, 15, 25, and 30mn)	Bluetooth	

Mobility models should also include variations in node speed and pause time in order to describe realistic movement situations. These parameters certainly influence the contact and inter-contact times of nodes. Different proposals [7, 10, 44, 49] have correctly assumed that contacts and inter-contact times are best described as exponentially and power law distributions. However, other proposals [8, 11] simply obtain the contact and inter-contact times from the considered datasets (i.e., human traces).

Another parameter that may have an impact on the density of the network is transmission coverage. Quite a few devices are already equipped with both Bluetooth and Wi-Fi cards, so transmission range must be looked upon. Based on the analyzed proposals, the transmission range should be set to an average value of 100 meters, assuming that devices work with Wi-Fi. This parameter should range from a minimum transmission range of 10 meters, which represents Bluetooth contact between mobile nodes, and a maximum transmission range of 250 meters in cases where fixed entities (i.e., access point) may be used for information relay.

Lastly, the use of beacons is useful to find out more information about potential new neighbors. However, beacons may not be always present since nodes are mobile, and so may go to sleep mode quite often in order to save battery. However, there are some pros and cons. Some of the pros relate to the fact that battery lifetime can last longer. However, this may result in the loss of good contact opportunities. According to the studied proposals, considering beaconing every 100 ms should be enough to achieve a good balance between battery lifetime and network knowledge. However, further investigation is required to validate this assumption.

Table 3.5 presents the traffic parameters considered by the studied proposals. The number of sources and destinations is the first parameter to stand out. In most of the cases, these nodes are chosen randomly and comprise a subset of the total number of nodes. What matters most regarding this parameter is that it remains the same throughout the comparison process in order to guarantee the same conditions to the evaluated proposals.

Regarding load, there are proposals that generate a message per second [4, 5], a number of packets per hour [8], a number of messages uniformly distributed [10, 49], a given number of bytes/messages per second/minute per source [45, 9], as well as proposals that give little [42] or no information about the load used in the network (e.g., *PeopleRank* [11], omitted from Table 3.5 for this reason). This is a setup parameter that must be carefully addressed as it may bring more variation to experiments. For instance, the used load distribution may be related to the mobility model.

Another setup parameter related to traffic is message size, which plays an important role when measuring the consumption of network (e.g., bandwidth) and node (e.g., buffer, and power) resources, since the size of messages may be different, depending on the applications generating them. Also, depending on the average contact duration (especially in highly mobile scenarios) and message size, data exchange may not even happen in each contact. This issue is further worsened if the proposal requires exchange of meta-data information [44, 45] prior to exchanging real data, in which case nodes may waste a portion of a potentially short contact time.

Just a few proposals explicitly mention message size, which can be of 1 KB [4, 45, 44, 39] or vary between 10 to 100 KB [8, 9]. Based on the considered proposals, varying the message size from 1 to 100 KB should provide a more realistic evaluation. For instance, people may write short messages/emails while on the move, so the variation of the size of messages easily represents the applications and time employed when using portable devices on the go.

Another aspect that affects traffic levels in the network is message TTL, which can be represented in number of hops [4, 5, 43] or time units [7, 42]. If messages have high values of TTL, they may end up increasing network/node resource consumption. On the other hand, if messages have a small TTL, they may not even reach their destination. From the considered proposals, TTL normally varies between hours to months or between 3 to 11 hops in average. In the case of time units, based on experimentation it is suggested that varying from 1, 2 days up to 3 weeks is enough to understand the performance behavior of the studied proposals. In the case of hop-based TTL, it has been shown that only 5% of nodes have some level of relationship with the destination in the first hop [49]; however, the interaction values improve (around 35%) for more than 3 hops. Thus, the UEF suggests the number of hops to be longer than/equal to 3 hops at least. Unlimited TTL should also be considered in order to understand the impact of resource consumption of proposals.

Table 3.5: Traffic parameters considered by the different proposals

Proposals	Source/Dest Distribution	# of Generated Messages (Load)	Message Size (Kb)	Message TTL	Buffer Size
Epidemic	45 / 44	One message per second	1	1, 2, 3, 4, and 8 hops	10, 20, 50, 100, 200, 500, 1000, and 2000 messages
PROPHET	45 / 44 (Random) and 2 / 1 (Comm.)	One message/sec (RWP), and 20mgs/sec – 2mgs/5sec (CM)		3 and 11 hops	200 messages
Spray&Wait	1 / 1	Node generate a new message with an inter-arrival time distribution uniform.		4000 – 6000 unit times	
Spray&Focus	1 / 1	Moderate number of CBR traffic sessions		1000 – 10000 time units	
PREP	1 / 1	40 to 200 bytes/sec/node	1		1 to 6MB
RAPID	1 / All	4 packets/hour/node	1		100KB (Power law) 40GB (Trace driven)
Prediction	1 / 1	After each contact event in the contact trace, a message is generated with a given probability	0.08 to 1	Unlimited	Unlimited
SimBet	1 / 1	A single message is generated between each node included in the subset			
Bubble Rap	1 / 1	1000 messages			
EBR		1, 2, and 4 messages/minute/source	25		1MB
OPF	1 / 20 (NUS) and 1/ 1 (UmassDieselNet)			3 and 1 to 5	

Buffer usage can also influence the performance of a given proposal. Considering unlimited buffer space is not realistic at all, since it cannot be assumed that users are willing to share all their storage room with others. It is worth mentioning that this may be different in scenarios where nodes are only there to serve others (e.g., buses) [44]. However, based on experimentation buffer space should be limited (varying according to the device) to a size of around 200 messages (considering 10 KB messages) or 2MB, based on the analyzed proposals.

Table 3.6 summarizes the UEF and its proposed (a) network density and (b) traffic parameters, for fairly assessing the performance behavior of existing and newly-devised opportunistic routing proposals.

Table 3.6: Network and traffic parameters of the proposed UEF

(a) Network Density Parameters				
Area Density (Km^2/ # of nodes)	**Mobility Model/ Speed (m/s)**	**Node Meeting/ Inter-meeting Time**	**Transmission Range (m)**	**Beacon Control**
Span conference building up to city areas/100 to 150	Representing people and different transportation means	Exponentially or Power-law, if it can be determined	Minimum of 10 (Bluetooth) and maximum of 250 (Wi-Fi)	Every 100 ms
(b) Traffic Parameters				
Source/Dest Distribution	**# Generated Messages (Load)**	**Message Size (Kb)**	**Message TTL**	**Buffer Size**
Same subset throughout all evaluation study	Load must be the same	1 to 100	$>$ 3 hops or varying from 1 day to 3 weeks. Unlimited TTL should be also explored	2MB (200 messages of 10 KB)

With this analysis, I could observe that not only are solutions evaluated against others that belong to different categories, but also consider different performance evaluation metrics as well as different conditions to decide on performance. This means that the way proposals are evaluated has an important impact on how they are classified as having or not satisfactory performance, which led me to devise this UEF.

3.3 The Role of Cooperation in the Support to Opportunistic Networking

Even if nodes have different contact opportunities that could lead to an efficient exchange of content, such opportunities are useless if nodes are not willing to cooperate between themselves. Cooperation can take place in different forms (e.g., resource exchange, selling services, and to improve network experience) and cooperative behavior only brings benefits to both users and network.

This section highlights the importance of cooperation to the support for opportunistic networking. Thus, it starts by presenting what users can exchange upon a cooperation opportunity, the incentives employed so users have cooperative behavior, and how cooperation can be seen. Then, it shows the proposed scheme to encourage nodes in engaging in cooperation for resource sharing.

3.3.1 An Overview of Cooperation

The capabilities of users' devices nowadays allow them to easily share their resources (e.g., storage space, bandwidth, processing power), divide efforts (e.g., download of files), and/or provide services (e.g., Internet connectivity, printing capability).

However, despite the fact of the improved capabilities of devices, users may still not be willing to engage in cooperation and share their resources and/or services. There are three major reasons that can affect the user willingness to share [29]: i) trust issues (e.g., interaction with unknown users), ii) not enough local resources (e.g., battery, storage), and iii) the natural egoistic behavior.

Cooperation plays an important role in opportunistic networking, in the sense that it can overcome the lack of trust on other users by considering i) the existing social similarities among cooperating nodes such as communities they belong to [13] and/or interests they share [14]; and ii) the use of virtual crediting used to "pay for" the provided resource/service/effort. If the issue is the availability of local resources, cooperation can increase user willingness by, in turn, providing them with the resources they lack at that moment (e.g., processing power [60], storage [61]) to improve their own operation. In the case of egoistic behavior, cooperation works around that by guaranteeing users that they will have resources available upon their needs.

As discussed in this section, cooperation manifests according to: i) the utilities the users need and have to offer (e.g., resources, services, efforts); ii) the incentives for encouraging users to engage in cooperation (e.g., exchange- and rewarding-based); and, iii) the perspective that cooperation is seen (e.g., utility-sharing and networking-experience).

It is important to mention that, in this section, I am not interested in the target network where the different cooperation utilities, incentives, and perspective are employed and seen. Instead, my sole goal is to show how cooperation can take place independently of the application scenario with few examples found in the literature.

3.3.1.1 Cooperation Utilities

Users may cooperate through sharing the same (or different) type(s) of resources (e.g., storage, processing), services (e.g., Internet connectivity) and/or splitting networking efforts (e.g., obtain the same content).

When it comes to sharing resources, users tend to cooperate by sharing the same type of resources (e.g., storage and processing) [29]. Cooperation can be based on storage sharing [61]: nodes store information based on a coding scheme aiming at a reliable delivery of data in scenarios where nodes sporadically shut down their wireless cards to save energy. Processing is another type of resource that improves the utilization of devices [60]. Nodes cooperate by sharing the processing of tasks among themselves. The approach consists in opportunistically distributing tasks followed by making available the resulting data.

By simply using the Bluetooth Dial-Up Networking, any node can start sharing its Internet service. This can also be achieved by installing applications to this end (e.g., JoikuSpot[3]). However, in these examples of service sharing cooperation, the nodes providing connectivity have no incentives to cooperate. Despite not being deployed in personal devices, Fon spots[4] are a clear example of cooperation among nodes for service sharing. Just by sharing the home connection with other *foneros*, the user has its Internet connectivity experience increased to a worldwide scale.

Effort splitting results from when users want to achieve a common goal (e.g., video download) aiming at sparing resources and time. In effort-sharing schemes, nodes interested in the same video file download different parts (i.e., video descriptors) of it [62]. Then, these parts are shared between the interested parties through a less energy costly medium (e.g., 802.11a). Efforts can also be exchanged by other resources: nodes serve as relays by forwarding messages on behalf of others, and in turn they get a portion of bandwidth for their own benefit [63].

[3]http://www.joiku.com/
[4]http://corp.fon.com

3.3.1.2 Cooperation Incentives

Cooperative behavior can be further encouraged by the use of incentives. Such incentives can be given by means of utilities exchange (i.e., resources, services, and/or efforts [29]) when required, or rewarding-based schemes, where virtual currency [64] may be used to acquire a resource/service/effort in scenarios that trust may still be an issue, and potential providers do not interact with users outside their communities [30, 31].

One can find a classification of cooperation mechanisms considering trust-based (i.e., exchange-based) and trade-based (i.e., rewarding-based) patterns [65]. In the former, trust is imperative so nodes can engage in cooperation, while for the former what matters is the remuneration (done immediately or as a promise) that the potential provider gets.

In summary, for incentives based on the exchange of utilities, the cooperation decision considers the levels of trust between the cooperative nodes. In this case, a trust management system is required and can be centralized and/or decentralized. As for rewarding-based incentives, cooperation only takes place if the provider gets the respective amount of credits (i.e., virtual currency) for the required resource/service/effort [30].

3.3.1.3 Cooperation Perspectives

Cooperation can be seen from two perspectives, namely utility-sharing and networking-experience. The utility-sharing perspective depicts the cooperation process as nodes sharing utilities, based on the fact that they will get the same/other utilities in return, or even virtual currency [30]. This may or not have impact in the wellness (i.e., suitable operation) of the network. Yet, the networking-experience perspective rely on the users themselves to cooperate in order to maintain such wellness.

As network operation is affected by the presence of misbehaving (i.e., non-cooperative, greedy) nodes, overall networking experience can be further improved through cooperative actions of nodes to report their own experiences in certain locations (e.g., sharing information on networking condition to a node going in that direction) [29].

Upon receiving such information, users can react fast and refrain from going to such locations, and interacting with misbehaving nodes. An example of such approach is when nodes are classified as malicious for using more resources than they usually share [66]. This classification happens as nodes interact directly with the malicious node or malicious activity information is received from neighboring nodes. Consequently, malicious nodes have their data refused for forwarding, and the overall experience of cooperative nodes improves.

3.3.2 Cooperation Scheme for Resource Sharing

The proposed cooperation scheme is based on trust and users exchange solely utilities (e.g., resources and/or services) or utilities for credits. The idea is to increase user willingness to engage in cooperation while in trusty (i.e., known, cooperative communities) and untrustworthy (i.e, unknown, uncooperative areas) scenarios.

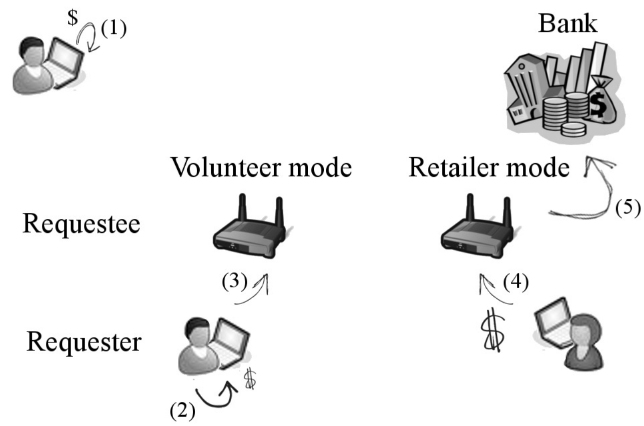


Figure 3.6: Cooperation scheme

As depicted in Fig. 3.6, users can be requesters (i.e., consumers) or requestees (i.e., providers) of resources and services. In the case of requestees, they can operate in two distinctive modes, namely volunteer or retailer. For the former mode, the requestee operates based on the trust it has on requesters, and provides whatever resources and services it has available knowing that it will have resources and services whenever it requires. However, if operating on retailer mode, the requestee only provides resources and services upon an agreement with requesters to pay a corresponding amount of credits.

Upon bootstrapping (1) the user device gets an initial amount of cooperation of credits based on its willingness to cooperate (reflected in terms of trust on others). As this relates to a new user in the network, these credits are provided in order to allow him/her to interact (i.e., pay) with other users specially in untrustworthy scenarios (i.e., requestees do not know this user).

As the cooperation opportunity arises with a volunteer requestee, the requester sends an amount of cooperation credits (2) he/she believes is fair for getting the required resource/service. The requestee accepts the sent credits (3) and cooperation takes places.

If the cooperation involves a retailer requestee, the user sends the cooperation credits to the requestee (4) which engages in cooperation. This process also involves a bank entity which controls the transfer of credits (5) between the accounts of requester and requestee.

The lesson learned from the proposed scheme is that, by engaging in cooperation independently of being exchange- or rewarding-based, contact opportunities indeed become the key aspect of opportunistic networking: content and its handling (e.g., storage, processing) happen easily and appropriately guaranteeing the wellness of network and users.

Independently of the different utilities/incentives/perspectives that cooperation can exchange/be based and seen, within the context of this Thesis work, cooperation takes place from the utility-sharing perspective (i.e., more specifically storage) and can only be exchanged by the same resource. Additionally, users are always operating in volunteer mode as I envision users willingly cooperating among themselves. Trust indeed is of major concern and drives the cooperation process [31]; however, it is out of the scope of this Thesis work, and therefore, it is not taken into account.

It is important to mention that the work done in this section is inserted in the User-centric Wireless Local Loop (ULOOP[5]) European project. This project targets User-Centric Networks (UCN), which has

[5]http://uloop.eu

the user as the focus for determining routing, security, information sharing, and among other networking-related mechanisms. Additionally, UCN nodes have the ability to provide other nodes with services and connectivity independently of the regular access/infrastructure providers [67].

Despite the differences between user-centric and opportunistic networking, both paradigms share one common aspect also found in the context of this Thesis: exploiting the user willingness in participating in the cooperation process.

3.4 Summary of the Chapter

Since 2007, a new trend has led opportunistic routing proposals to consider some level of social interaction and behavior to perform forwarding decisions. Still, previous taxonomies were either not considering such new routing trend, or simply looking at it as yet another destination/device-dependent function (cf. Sec. 2.2). Thus, this chapter introduced a new, updated taxonomy, which gives proper attention to the identified social similarity branch of opportunistic routing proposals, and highlights its potential to improve opportunistic routing based on social relationships, common communities, shared interests, node popularity, and dynamic behavior of users [1].

By reviewing the opportunistic routing state-of-the-art literature, I observed that the evaluation of opportunistic routing proposals does not consider a similar set of performance metrics and comparable experimental scenarios. Hence, this chapter presented a Universal Evaluation Framework (UEF) that allows a fair comparison assessment of opportunistic routing solutions. This UEF considers common performance evaluation metrics (e.g., delivery rate, cost, and delay) and provides a set of guidelines to be considered when creating experimental scenarios, spanning different parameters regarding network density (e.g., network area, number of nodes, mobility model, node meeting and inter-meeting times, transmission range and beacon control) and traffic (e.g., distribution of sources and destinations, load generation, message size, message TTL, and buffer size) [27, 28].

Finally, the chapter is concluded with a study concerning the role of cooperation among users as to support opportunistic networking. The user willingness in engaging in cooperation can be encouraged by means of exchanging resources, or by the provision of incentives such as virtual currency. Such willingness is very important since it allows a better networking experience [29, 30, 31]. Despite this study has been done in the context of user-centric networking, it explores user willingness to encourage cooperation, which is of great interest to this Thesis work.

Chapter 4

Social-aware Utility Functions

After gathering enough knowledge about the functioning of the different opportunistic routing proposals and understanding the importance of cooperation in opportunistic routing, I was able to come up with three utility functions based on the dynamics of user behavior [32, 33, 35, 19], which are then used in the proposals I have designed in the context of this Thesis work.

This chapter starts by introducing the time-evolving property (i.e., behavior in different time periods) that is imperative for social-aware opportunistic networking. Then, it presents the devised social-aware utility functions that capture the level of social relationship between nodes and measure their social importance in the system to improve routing. Next, it shows a set of experimental analysis carried out over the utility functions. The chapter also presents a third social-aware utility function that measures the social relationship among users sharing interests. Finally, the chapter concludes with a scalability study on the devised utility functions.

4.1 Time-evolving Property

It has been proven that information about social relationships is useful for data exchange [10, 12, 14, 18, 11, 13, 17]. As shown in Chapter 2, several opportunistic routing solutions have emerged, attempting to exploit social graphs created considering different contact metrics, such as inter-contact times, similarity (e.g., communities, interests), and betweenness centrality. However, these social metrics are not fully capable of accurately capturing the dynamics of user's behavior [51], resulting in inefficient social graphs. On the one hand, graphs may include edges that do not have significant social meaning. On the other hand, some metrics lead to edges that may end up having the same importance over time producing random-based forwarding solutions. Hence, it is imperative to devise social graphs able to accurately capture significant social ties in scenarios created based on realistic user behavior (e.g., people have different habits in different moments of the day).

It has also been shown how important it is to capture the different connectivity patterns throughout network lifetime [56]. These patterns change as nodes interact in the network, and such feature has a major impact on how social graphs are devised. This means that the performance of forwarding algorithms is directly linked to the used mobility models [68] and how the resulting interactions are mapped into the social graph [51].

In order to build social graphs that reflect dynamic connectivity patterns and social interactions, I introduce the time-evolving property of social ties that allows opportunistic routing protocols to

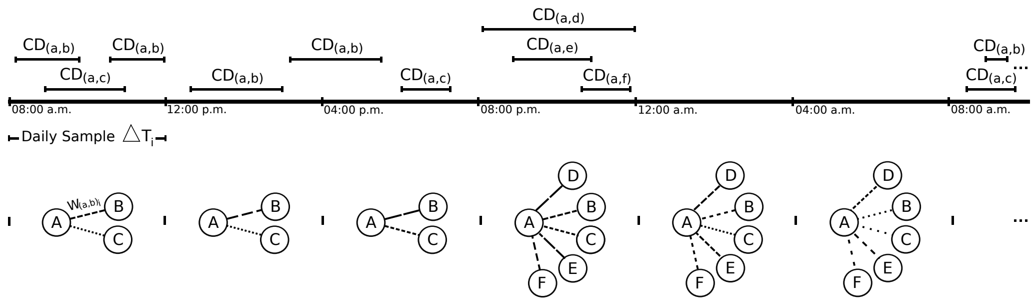


Figure 4.1: Contacts of node A with nodes x $(CD\,(a,x))$ in different time intervals ΔT_i

operate based on social graphs that reflect what is happening during the daily habits in terms of social interactions among people. It is expected that social graphs based on the time-evolving feature may more accurately reflect users' social relationships and contacts than other social metrics such as inter-contact times.

4.2 Time-Evolving Contact Duration Utility Function

Prior opportunistic routing protocols have been investigating the utilization of inter-contact times, namely the time elapsed since last contact and the number/frequency of contacts. The expectation is that such metrics allow the creation of accurate social graphs based on the mobility behavior of nodes, and that such social graphs allow the development of less volatile opportunistic routing approaches. The assumption is that social relationships are more stable than sporadic physical contacts. Although routing based on social graphs may lead to more stable forwarding in the presence of intermittent connectivity, I believe that a utility function based on the time or number of contacts is not enough to create graphs reflecting the time-evolving feature of social ties.

This work aims at showing that the accuracy level of social graphs is mainly dependent upon the duration of contacts at different moments in time, instead of the time elapsed since last contact, number/frequency of contacts, or node importance. Contact duration at different moments in time is taken into account, since people have daily habits that lead to a periodic repetition of behaviors [58, 50, 59].

The *Time-Evolving Contact Duration* (*TECD*) utility function is proposed to derive the weight of edges in a social graph based upon the statistical contact duration that nodes have over time. *TECD* encompasses: i) the duration of contacts, representing the intensity of social ties among users; and, ii) time-evolving social ties, reflecting users' habits over different time periods (referred to as daily samples). Fig. 4.1 (Copyright © 2012 IEEE [15]) shows how the social interaction (from the point of view of node A) changes over time reflecting daily routines. As an example, for a period between 8 p.m. and midnight, the social interaction of node A is stronger with nodes D, E, and F than it is with nodes B and C, being the weight of the edges illustrated by the intermittency of the line.

As illustrated in Fig. 4.1, two nodes may have a social weight $w\,(x,y)_i$ that depends on the average total contact duration they have in that same period of time over different days. Within a specific daily sample ΔT_i, node x has n contacts with node y, having each contact k a certain *Contact Duration* $(CD\,(x,y)_k)$. At the end of each ΔT_i, the *Total Connected Time* $(TCT\,(x,y)_i)$ between nodes x and y is given by Eq. 4.1.

$$TCT\,(x,y)_i = \sum_{k=1}^{n} CD\,(x,y)_k \tag{4.1}$$

Each ΔT_i represents a different period of time in the daily routine of a person. Since a behavior pattern can be observed, one can consider that, at each daily sample, a specific social behavior is taking place at work/study place, home, or somewhere else (e.g., out of town, friends' houses). It has been shown that people can have their future behaviors predicted by considering previous ones [50]. Thus, this approach tries to capture such behaviors considering the time that nodes spend together (i.e., *Total Connected Time*) in the same daily samples ΔT_i along different days j.

Through a cumulative moving average (cf. Eq. 4.2), on day j, I determine the *Average Duration* ($AD\,(x,y)_{ji}$) of the *Total Connected Time* between nodes x and y at ΔT_i by considering the current behavior ($TCT\,(x,y)_{ji}$) and the *Average Duration* in the same period of the previous day ($AD\,(x,y)_{(j-1)i}$).

$$AD\,(x,y)_{ji} = \frac{TCT\,(x,y)_{ji} + (j-1)\,AD\,(x,y)_{(j-1)i}}{j} \tag{4.2}$$

I believe that the social strength between nodes in a certain daily sample should also give some indication about the strength between such nodes in subsequent periods of time (referred to as Time Transitive Property). For instance, the social strength of nodes between 8 a.m. and 12 p.m. should provide an expected strength between 12 p.m. and 4 p.m. For routing purposes, such time transitive property helps to select nodes that may provide good delivery probability over consecutive periods of time, where long messages can be fully transmitted using the contacts expected to occur in consecutive periods.

Hence, *TECD* (cf. Eq. 4.3) gives the social weight between any pair of nodes, $w\,(x,y)_i$. Such social weights are a function of the *Average Duration* of the *Total Connected Time* between nodes x and y in ΔT_i ($AD\,(x,y)_i$) and in subsequents $t-1$ daily samples, being t defined by default as the number of configured samples. Thus, k evolves from the current daily sample i to the daily sample $i+t-1$ (e.g., from daily sample three, corresponding to 4 p.m. - 8 p.m. to daily sample two, corresponding to 12 p.m. - 4 p.m., as shown in Fig. 4.1). For values of $k>t$, the value of $AD\,(x,y)$ corresponds to the time slot $k-t$. For instance, for a configuration with six daily samples, for k equals to 9, the value of $AD\,(x,y)$ corresponds to the third sample of the day, which means 4 p.m. - 8 p.m., as illustrated in Fig. 4.1. In Eq. 4.3, the time transitive property of *TECD* is represented by $\frac{t}{t+k-i}$, giving more importance to the average duration of contacts in the current daily sample, and decreasing such importance in subsequent ones.

$$TECD = w\,(x,y)_i = \sum_{k=i}^{i+t-1} \frac{t}{t+k-i}\,AD\,(x,y)_k \tag{4.3}$$

4.3 Importance Utility Function

As social interaction may also be modeled to consider the node importance, I propose a variation of *TECD*, called *TECD Importance* (*TECD*i). *TECDi* computes the *Importance* ($I(x)_i$) of a node x (cf. Eq. 4.4), considering the weights of the edges between x and all the nodes in its neighbor set ($N(x)$) at a specific ΔT_i along with their importance.

$$TECDi = I(x)_i = (1-d) + d \sum_{y \epsilon N(x)} w(x,y)_i \frac{I(y)_i}{N(x)} \tag{4.4}$$

TECDi is based on the *PeopleRank* function [11]. However, it is my understanding that the selection of next hops should consider not only their importance, but also the strength of social ties between message holder and potential next hops. Another difference is that, with *TECDi*, the neighbor set of a node x only includes the nodes which have been in contact with node x within a specific daily sample ΔT_i, whereas in *PeopleRank* the neighbor set of a node includes all the nodes that ever had a link to node x. Note that the dumping factor (d) in the formula is the same used in *PeopleRank* and represents the level of randomness considered by the forwarding algorithm.

4.4 Experimental Analysis of Utility Functions

This section presents further analysis to improve the utility functions considering: i) the inclusion of message TTL while taking a forwarding decision; ii) the level of importance given to the average duration of contacts while determining the social weight; iii) the suitable number of daily samples.

Other improvements (results from code debbuging in the Java implementation for the ONE simulator [68]) done to ensure that the proposed utility functions have satisfactory performance are reported in the format of a technical report [32].

4.4.1 TTL Impact on the Time Transitive Property

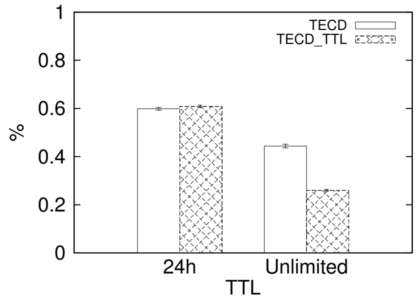
To further explore the Time Transitive Property mentioned in Sec. 4.2, I considered the Time-To-Live (TTL) of the messages while determining the weight to a potential next forwarder.

Thus, for this set of tests, *TECD* was adapted in order to also take into account the message TTL, which is referred to as *TECD_TTL*. Fig. 4.2 shows the results considering the average delivery probability, average cost, and average latency.

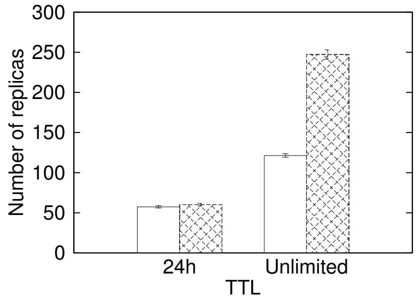
In *TECD_TTL*, message TTL is checked for every message in order to determine how far (in terms of daily samples) it is still alive. Once this in known, the weights to a given destination are then determined considering only the daily samples over which the message is still useful.

Regarding the average delivery probability (cf. Fig. 4.2(a)), *TECD_TTL* has a gain of only 0.96% over *TECD* for a 24-hour TTL. Such gain is dramatically reduced when the message TTL is set to unlimited (*TECD* has performance of more than 18% over *TECD_TTL*).

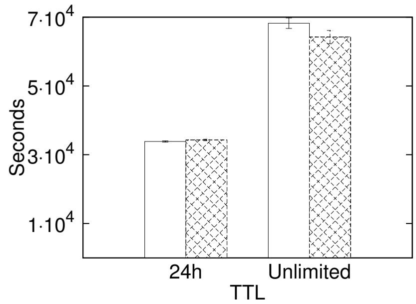
When it comes to the average cost (cf. Fig. 4.2(b)), *TECD_TTL* creates 2.79 and 125.81 replicas more than *TECD* for 24-hour and unlimited TTL, respectively.



(a) Average delivery probability



(b) Average cost



(c) Average latency

Figure 4.2: Effect of message TTL when determining the social weight

Finally, in what concerns the average latency (cf. Fig. 4.2(c)), *TECD_ TTL* has an increase in terms of latency (439.73s) when compared to *TECD* for a 24-hour TTL. For the unlimited TTL case, the lower latency (4041.09s) of *TECD_ TTL* is explained by the fact that the reported latency is a function of the delivered messages, which for this case was low.

In *TECD_ TTL*, messages were able to spend less time in buffer (11156s) than with *TECD* (11379). One can see that, considering the message TTL to determine link weights, *TECD_ TTL* has only a slightly gain compared to its counterpart *TECD* in terms of delivery probability. However, this comes with an increase in the number of created replicas and latency. More replicas are created especially when TTL is close to expiring, since TECD_TTL is expected to increase its delivery probability by forwarding messages as quickly as possible. This latency behavior is justified by the increase in the number of created replicas, since these extra few copies also contribute to the overall latency.

As a conclusion, the inclusion of TTL does not bring improvements, and therefore, it is not considered in the *TECD* computation.

4.4.2 Impact of Importance Level on the Estimation of Average Contact Duration

Social weight between a pair nodes (cf. Eq. 4.3) considers the summation of the different average duration (AD) of contact between these nodes. Additionally, each of these *AD*s is given a level of importance ($\frac{t}{t+k-i}$), and the *AD* in the current daily sample gets the most importance. In order to understand how such level should be set (i.e., how much of each subsequent *AD* should be considered while determining *TECD*), the three levels below were defined, considering the number of daily samples $t = 6$:

- Level A: $w = \frac{6}{21}AD + \frac{5}{21}AD + \frac{4}{21}AD + \frac{3}{21}AD + \frac{2}{21}AD + \frac{1}{21}AD$, given by 4.5.

$$\sum_{k=0}^{t-1} \frac{t-k}{\sum_{j=1}^{t} j} AD \tag{4.5}$$

- Level B: $w = \frac{6}{6}AD + \frac{6}{7}AD + \frac{6}{8}AD + \frac{6}{9}AD + \frac{6}{10}AD + \frac{6}{11}AD$, given by 4.6.

$$\sum_{k=1}^{t} \frac{t}{t+k-1} AD \tag{4.6}$$

- Level C: $w = AD + 0.81AD + 0.62AD + 0.43AD + 0.24AD + 0.05AD$, given by 4.7.

$$\sum_{k=0}^{t-1} 1 - (0.19 \times k)\, AD \tag{4.7}$$

Level A is devised to give very little importance to all *AD*s and to present a subtle importance reduction from one *AD* to another (approximately 0.05), while levels B and C are devised to give much more importance to these *AD*s. However, the importance reduction from one *AD* to another is different between these last two levels: it is rather variable for level B and of 0.19 for level C. Fig. 4.3 shows how the level of importance is given to each *AD* per daily sample.

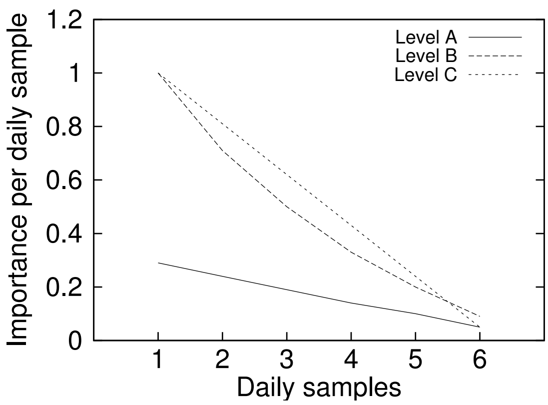
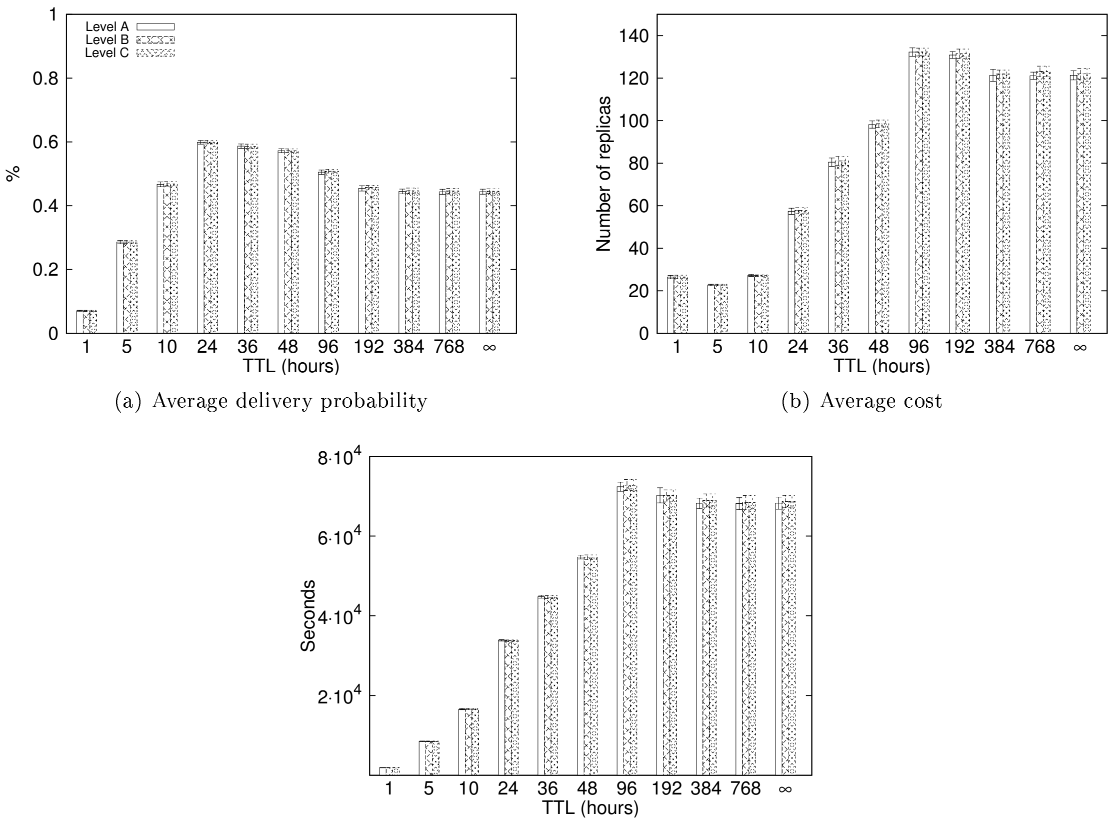


Figure 4.3: Different importance levels

Based on Fig. 4.3, the performance of each level is assessed when determining the social weight for the same performance evaluation metrics, namely average delivery probability, average cost, and average latency as shown in Fig. 4.4.

Regarding average delivery probability (cf. Fig. 4.4(a)), *TECD* with level B managed to have slightly better performance than with levels A and C. This performance of *TECD* with level B comes with a subtle increase in cost (as shown in Fig. 4.4(b)), especially when compared to its version with level A. Regarding latency (cf. Fig. 4.4(c)), *TECD* with level B had an increase in 45% of the cases (i.e., TTL at 10, 96, 384, 768, and infinity).



(a) Average delivery probability

(b) Average cost

(c) Average latency

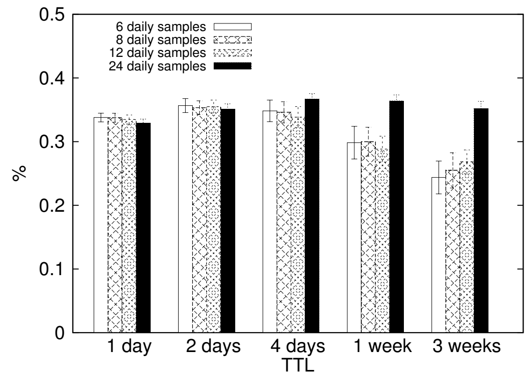
Figure 4.4: Effect of different levels when determining the social weight

One can see that the results are statistically equivalent. Thus, level B is chosen for the *TECD* implementation. The reason is that with this level, *TECD* increases faster as well as decreases faster, which is a good approach especially when penalizing the social weight between nodes to improve routing performance. This is an interesting approach as *TECD* is expected to reflect what is happening in every daily sample between the nodes in terms of social interactions, so only the best ties are considered for data forwarding.

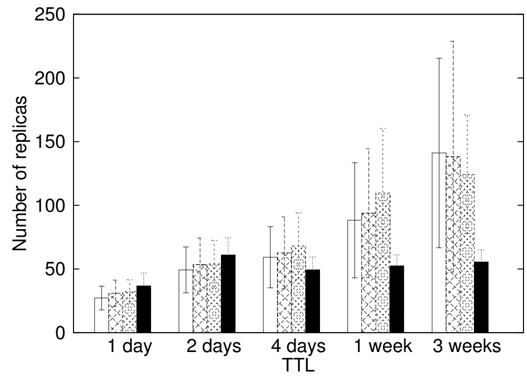
4.4.3 Impact of Daily Sample Size on Social Weights

I also investigated the effect of length of the daily samples when determining the social weights. Thus, I considered 6, 8, 12, and 24 daily samples each with 4, 3, 2, and 1 hour of length.

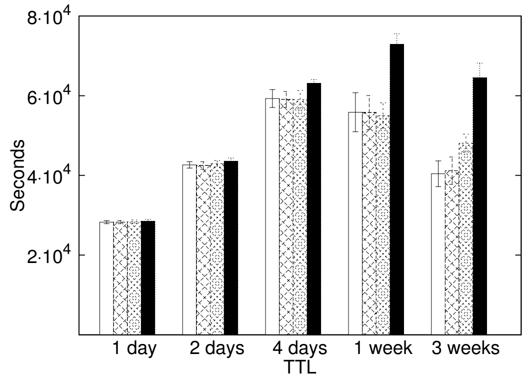
As it can be seen in Fig. 4.5, the performance improves as daily sample length decreases, and is rather stable in terms of average delivery probability and cost (cf. Figs. 4.5(a) and 4.5(b)). However, this improvement comes with higher latency behavior as shown in Fig. 4.5(c), which results from the fact that now next forwarders are carefully chosen, and thus forwarding decisions take more time to happen.



(a) Average delivery probability



(b) Average cost



(c) Average latency

Figure 4.5: Effect of different daily samples when determining the social weight

Still, the 24-daily sample configuration was the one considered, as it provides a much more stable behavior as the TTL increased, and a more refined view of the social weight and importance of nodes in the system.

4.4.4 Suitability of TECD for Community Detection

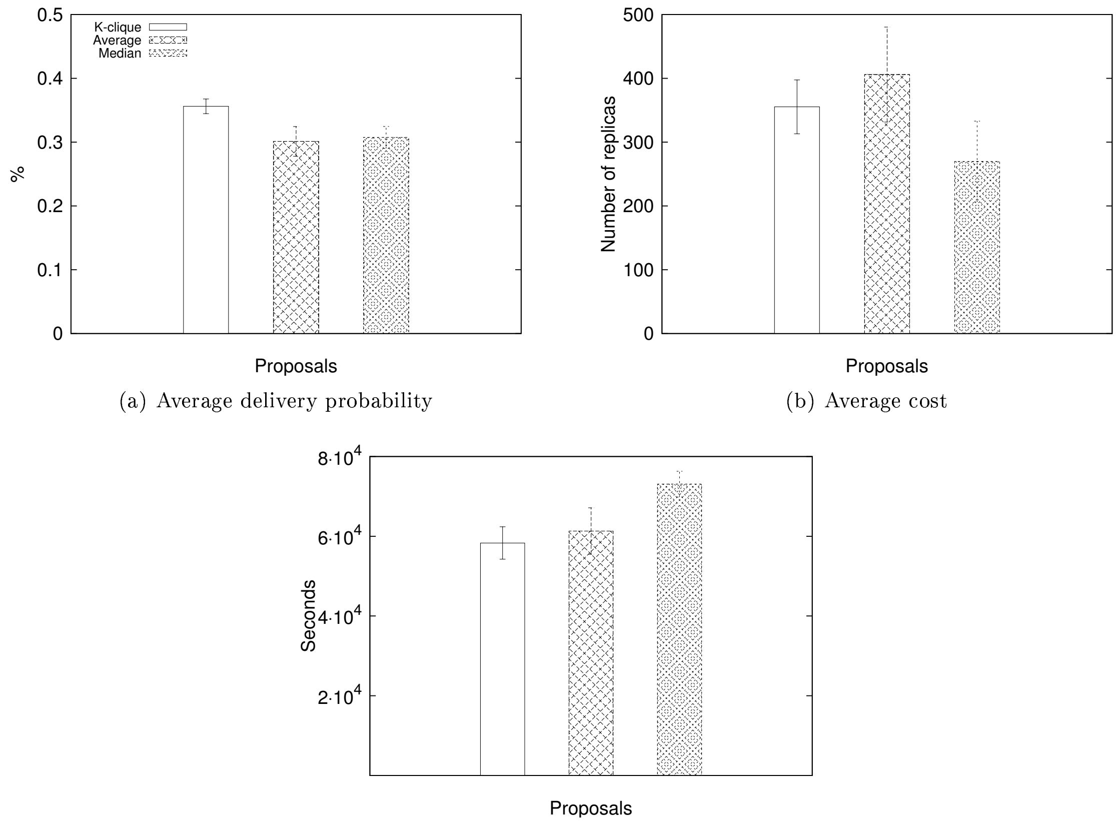
Since social similarity may also involve community formation, I tried to find other uses for the proposed utility functions. Thus, I decided to employ *TECD* when forming communities to understand whether *TECD* could improve community detection.

An algorithm usually considered for community formation is K-clique [46]: it is used by *Bubble Rap* [13], for instance. Thus, I considered the version of K-clique proposed by Hui et al. [69] in which a node A builds a *familiar set* F_i comprising other nodes i that it spends time with (i.e., i is included in F_i, iff the cumulative contact duration with i exceeds a certain threshold T_{th}). Then, node B will add node A to its *local community* C_B if at least $k-1$ nodes in the *familiar set* F_i of node A can also be found in C_B.

The idea is to create alternative algorithms for community detection based on the *TECD* utility function, which is used to decide about including new nodes to the *familiar sets*, and *local communities* considering the weights of nodes that are already part of the community. Additionally, if a node already in the *local community* has a lower weight than the threshold (expressed in terms of average or median of weights), it will be removed from the community.

For instance, a node A has in its *local community* nodes B, C and D with social weights to them of 35, 30, and 40, respectively. Now, node A encounters node E with some regularity. It will only include node E in its *familiar set* and consequently to its *local community*:

- when its weight to E is greater than the average of its weights towards the nodes already in its community ($w(A, E) > average(\, w(A, B),\, w(A, C),\, w(A, D))$;
- or when its weight to E is greater than the median of its weights towards the nodes already in its community ($w(A, E) > median(\, w(A, B),\, w(A, C),\, w(A, D))$;



(a) Average delivery probability

(b) Average cost

(c) Average latency

Figure 4.6: Performance of Bubble Rap with different community detection approaches

For the sake of simplicity, these new community detection algorithms are referred to as Average and Median, respectively. Fig. 4.6 shows the performance of *Bubble Rap* with the original version of K-clique, as well as with the proposed algorithms. It is worth mentioning that the algorithms only differ from the original K-clique when adding nodes to *familiar sets* (and consequently to *local communities*). The common $k - 1$ node approach remains the same for all algorithms in the case of adding solely to *local communities*. To conclude, I defined $k = 5$, since it increased the performance of *Bubble Rap*, and message TTL was set at 24 hours for these experiments.

The results show a different conclusion from what was expected: social weight-based community detection algorithms was to result in communities comprising nodes with strong social connections. Indeed, the utilization of *TECD* resulted in many different close-knit groups, with very low connection between them. The consequence was a lower average delivery probability (cf. Fig. 4.6(a)) with higher latency (cf. Fig. 4.6(c)). Regarding cost (cf. Fig. 4.6(b)), Average produced the highest number of copies while Median produced the lowest. The reason for that is that the Average approach resulted in more nodes in the communities than Median, leading to higher number of replications.

It is important to remember that the goal here was to check whether *TECD*-based community detection algorithms could determine communities reflecting the reality found in the users' daily routines. This could be seen as an extra contribution of this Thesis; however, results have shown that this topic deserves more attention given the ability of *TECD* to cope with the dynamic behavior of nodes and their communities, suggesting a potential future work.

4.5 Time-Evolving Contact to Interest Utility Function

By looking at the opportunistic routing solutions that have emerged, one can conclude that almost all of them aim to transport data from point A to point B considering content forwarding (e.g., single-copy forwarding), or content replication at different levels based on node encounter, resource usage, or social similarity. That is, opportunistic routing solutions operate based on the identification of hosts (source, destination) and are not based on the transported content. However, it has been shown how dynamic scenarios can benefit (performance improvements and wise resource usage) from considering the content while performing routing [52, 70, 71].

Based on this observation and on the stable version of *TECD*, I propose the *Time-Evolving Contact to Interest* (*TECI*) utility function. *TECI* considers the interests of users on the content traversing the network. This means that the social weight given by *TECI* reflects the *probability of meeting nodes with a given interest among those which have similar social daily routine*, while the social weight given by *TECD* refers to the *probability of meeting nodes with same social daily routine*.

To help illustrate how *TECI* is computed, Fig. 4.7[1] shows the social interactions that node A has with others during different daily samples. In order to simplify this example, each encountered node has only one interest (nodes B and F have interest 1, and nodes C, D and E have interests 2, 3, and 4, respectively).

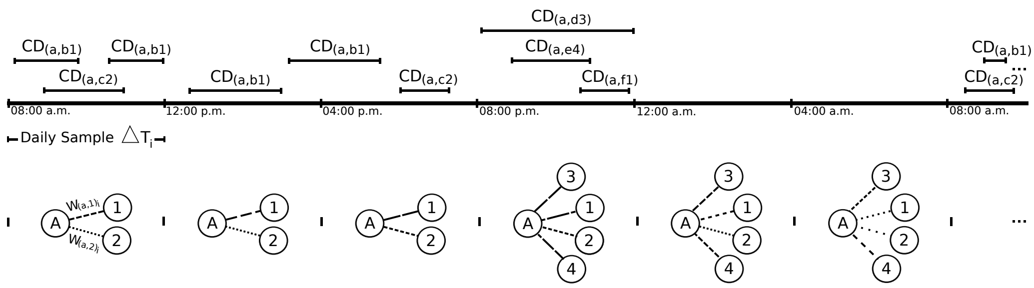


Figure 4.7: Contacts node A has with interests x ($CD(a,x)$) of other nodes in different daily samples ΔT_i.

As for *TECD*, with *TECI* contact duration is measured, but instead of attributing such duration to

[1]Springer and the original publisher (Ad Hoc Networks, v. 129, 2014, p. 100-115, Social-Aware Opportunistic Routing Protocol Based on User's Interactions and Interests, Waldir Moreira, Paulo Mendes, Susana Sargento, Figure 1, Copyright © 2014, Institute for Computer Sciences, Social Informatics and Telecommunications Engineering) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media [19].

nodes, it is attributed to interests that these nodes have (cf. $CD(a, b1)$ in Fig. 4.7). This means that nodes measure the different levels of social interaction (intermittency of lines in graphs) with interests ($w(a, 1)$) of nodes encountered throughout their daily activities. It is based on such social interactions levels (i.e., weights) that nodes decide whether an encountered node is a good forwarder for a message carrying a specific interest.

So, for every daily sample ΔT_i, node A has n contacts with other nodes having an interest x, with each contact k having a certain duration ($CD\,(a, x)_k$); and at the end of each daily sample, node A computes the *Total Connected Time to Interest* x based on Eq. 4.8.

$$TCTI\,(a, x)_i = \sum_{k=1}^{n} CD\,(a, x)_k \tag{4.8}$$

The *Total Connected Time to Interest* x in the same daily sample over consecutive days is used to estimate the average duration of contacts towards this interest x for that specific daily sample. Thus, from node A's perspective, the *Average Total Connected Time to Interest* x during a daily sample ΔT_i in a day j is given by a cumulative moving average of $TCTI$ in that daily sample ($TCTI\,(a, x)_{ji}$) and the $ATCTI$ during the same daily sample ΔT_i in the previous day ($ATCTI\,(a, x)_{(j-1)i}$) as illustrated in Eq. 4.9.

$$ATCTI\,(a, x)_{ji} = \frac{TCTI\,(a, x)_{ji} + (j-1)\,ATCTI\,(a, x)_{(j-1)i}}{j} \tag{4.9}$$

Then, node A computes *Time-Evolving Contact to Interest* x *(TECI)* (cf. Eq. 4.10) to determine its social strength ($w(a, x)_i$) towards x in a daily sample ΔT_i based on the $ATCTI$ in such daily sample and in consecutive $t-1$ samples, where t is the total number of samples. In Eq. 4.10 $\frac{t}{t+\text{k-i}}$ represents the time transitive property as explained in Sec. 4.2.

$$TECI = w(a, x)_i = \sum_{k=i}^{i+t-1} \frac{t}{t+k-i} ATCTI(a, x)_k \tag{4.10}$$

As *TECI* is a variant of the stable *TECD*, no further improvement was required, and the initial goal was achieved: having utility functions that capture the dynamism of social relationships between nodes and between nodes sharing interests.

4.6 Scalability Analysis of Utility Functions

This section presents a scalability analysis for *TECD*, *TECDi*, and *TECI*. For this purpose, it was considered the memory needed for computing these utility functions.

Considering a worst case scenario with k time slots and n nodes, where every node meets all other nodes in each ΔT_i, there are: i) $n \times (n-1)$ variables to store the starting time for every new connection; ii) $n \times (n-1)$ variables to store TCT computations; and iii) $k \times n \times (n-1)$ variables to store AD computations.

If each variable has X bits, *TECD*'s needed resources is given by Eq. 4.11.

$$\text{TECD} = n \times (n-1) \times (k+2) \times X\ bits \tag{4.11}$$

In an example scenario where there are 150 nodes, 6 time slots, and 64 bit double for storing, this results in 1.364 MB of total memory usage in the system, which means that, in average, each node needs up to 4 MB (including the 2 MB of buffer space).

As for *TECDi*, nodes need to store their importance and the importance of nodes they meet. Thus, the amount of needed resources is given by Eq. 4.12.

$$TEDCi = n^2 \times X + \text{TECD}\, bits \tag{4.12}$$

Assuming the illustrated worst case, *TECDi* needs a storage capacity of 1.536 MB, which means that, in average, a node needs to reserve the same 4 MB.

Regarding *TECI*, its scalability is given by the total number of existing interests. Hence, for a worst case scenario with k time slots and m interests, and with every node meeting all other nodes (having at least one interest) in each ΔTi, there are: i) m variables to store every connection; ii) m variables to store $TCTI$ computations; and iii) $k \times m$ variables to store $ATCTI$ computations. If each variable has X bits, TECI's required resources is given by Eq. 4.13.

$$\text{TECI} = m \times (k+2) \times X\, bits \tag{4.13}$$

With 35 interests, 24 time slots, and 64 bit double for storing, *TECI* requires 7.11 KB of storage in each node.

Given the amount of required resources, both *TECD*- and *TECDi*-based proposals can easily scale. In the case where buffer is very limited, a solution is to keep track of the best social weights, eliminating those under a threshold. A configuration with pre-defined thresholds should be considered to investigate which is the most suitable value to be used.

As for *TECI*, content-driven networks are expected to have a high number of interests: in the case where a node meets other nodes that have 1 billion different interests per day, *TECI* requires 193.71 GB of memory, which is still feasible today since nodes (e.g., laptops) do have storage up to 500 GB. However, not all nodes (i.e., smartphones) in dynamic networks have such storage capabilities, and even if they had, owners would probably not share all of it on behalf of others. Thus, *TECI* can scale through reducing its encountered interest space by: i) setting a daily threshold of 2 MB (equivalent to daily meeting nodes with more than 10000 interests); ii) eliminating the interests associated to nodes not well socially connected to them at the end of a day; and iii) if the threshold is reached. These rules set the basics to allow *TECI*-based solutions to scale.

4.7 Summary of the Chapter

One can easily observe that social-aware opportunistic routing has shown great potential. However, the social metrics employed in such routing do not fully capture the dynamic behavior of users [51], which influences its performance. Thus, this chapter started by introducing the time-evolving property of social ties (i.e., behavior in different time periods). This property allows the operation of opportunistic routing over social graphs that reflect the daily social interaction of users. Consequently, opportunistic routing leads to a better usage of network and node resources.

Then, the chapter presented social-aware utility functions, namely: i) the *Time-Evolving Contact Duration* (*TECD*) that determines the social weight among nodes based on their level of interaction; ii) the *TECD Importance* (*TECDi*) that measures node importance based on its neighboring nodes and social weights towards them; and iii) the *Time-Evolving Contact to Interest* (*TECI*) that weighs the social interaction among nodes sharing similar interests.

The chapter also presented an analysis of the proposed utility functions as to improve their performance and to assess their scalability considering worst case scenarios.

Chapter 5

Social-aware and Content-based Opportunistic Routing

With stable versions of the *TECD*, *TECDi* and *TECI* utility functions, I started the implementation of two social-aware opportunistic routing proposals: *dLife*, an opportunistic routing proposal based on the users' social daily routine; and the Social-aware Content-based Opportunistic Routing Protocol, *SCORP*, based on the content traversing the network and the interest of the users in such content.

This chapter starts by presenting *dLife* [15] and its ongoing specification effort [16]. Then, *SCORP* [19] is presented, followed by the differentiating aspects between these proposals [35]. The chapter is concluded by presenting how these proposals fit in the context of the DTN-Amazon project [34].

5.1 Opportunistic Routing Based on User Social Daily Routine

In what concerns solutions based on social similarities, it is important to achieve a correct mapping between real node interaction and the social graph that aids routing. Hossmann et al. (2010) [51] show that the key for successful forwarding is related to the ability of mapping social interaction (resulting from the mobility process) into a clean social representation (i.e., that best reflects the mobility structure), which should capture the daily life routine of nodes. Gonzalez et al. (2008), Eagle and Pentland (2009), and Hsu et al. (2009) [58, 50, 59] show that people have periodical mobility patterns (i.e., routines) and location preferences that can be used to identify future behavior, as well as interaction with people with whom they share similar behavior and potentially the same community.

On the other hand, the identification of social structures encompasses the challenge of detecting and adjusting communities on-the-fly in a useful time frame. Current research efforts show the difficulty of constructing and adjusting social structures in short periods of time [12, 14, 13].

Based on the evidences that opportunistic routing should mimic social behavior (e.g., daily routines) and that the creation of social structures may lead to complex solutions, I propose *dLife* that takes into account the people's daily life routine and identifies strong social ties, aiming at reaching a clean representation of social interactions.

dLife combines *TECD* and *TECDi* to forward messages to nodes that have a stronger social relationship with the destination, or that have greater importance than the current carrier of the message.

With *TECD* each node computes the average of its contact duration with other nodes during the same set of daily time periods over consecutive days. I assume that contact duration can provide more reliable information than contact history, or frequency when it comes to identifying the strength of social relationships. The reason for considering different daily time periods relates to the fact that users present different behavior during their daily routines [50].

In the case that the carrier and/or encountered node have no social information towards the destination, forwarding takes place based on a second utility function, *TECDi*, where the encountered node gets a message if it has greater importance than the current carrier of the message.

5.1.1 *dLife* Algorithm

As mentioned earlier, *dLife* [15] decides to replicate messages based on the *TECD* and *TECDi* utility functions: if the encountered node has better relationship with the destination in the current daily sample, it receives the copies of the messages (also known as basic strategy). By having higher weight (i.e., high social relationship), there is a much greater chance for the encountered node to meet the destination in the future. If the relationship to destination is unknown, replication only happens if the encountered node has higher importance than the carrier.

The operation of *dLife* happens as follows (cf. Alg. 5.1, Copyright © 2012 IEEE [15]): when the $CurrentNode$ meets a $Node_i$ in a daily sample ΔT_k, it gets a list of all neighbors of $Node_i$ in that daily sample and its weights towards them ($Node_i$.WeightsToAllneighbors). Then, every $Message_j$ in the buffer of the $CurrentNode$ is replicated to $Node_i$ if the weight towards the destination (getWeightTo($Destination_j$)) of the latter is greater than the weight of the $CurrentNode$ towards the same destination. Otherwise, $CurrentNode$ receives the importance of $Node_i$, and messages are replicated if $Node_i$ is more important than the $CurrentNode$ in the current ΔT_k.

Algorithm 5.1 Forwarding with *dLife*

```
begin
 foreach Node_i encountered by CurrentNode do
  receive(Node_i.WeightsToAllneighbors)
  foreach Message_j ∈ buffer.(CurrentNode) & ∉ buffer(Node_i) do
    if (Node_i.getWeightTo(Destination_j) >
                        CurrentNode.getWeightTo(Destination_j))
       then CurrentNode.replicateTo(Node_i, Message_j)
    else
       receive(Node_i.Importance)
        if (Node_i.importance > CurrentNode.importance)
        then CurrentNode.replicateTo(Node_i, Message_j)
end
```

5.1.2 *dLife* Specification

The specification of the *dLife* protocol comes in the format of an Internet-Draft that has been presented to the Delay Tolerant Networking Research Group [16]. This section presents the most relevant parts that are covered in the Internet-Draft: the applicability scenarios, the architecture of a *dLife* node, and the messages and phases involved in the operation of the protocol.

5.1.2.1 Applicability Scenarios

Generally speaking, *dLife* targets scenarios where it is not possible to assume the existence of an end-to-end path between any pair of nodes in any moment in time. The absence of an end-to-end path may be a consequence of node mobility and availability (e.g., nodes may switch off their radios to spare energy), physical obstacles, interference, among others.

As devices are carried by humans who do present a pattern in behavior (i.e., periodical mobility patterns and location preferences [58, 50, 59]), *dLife* can be easily employed to take advantages of the social interactions among the owners of such devices, and to overcome this lack of end-to-end path to deliver content to its due destinations.

With this in mind, *dLife* is expected to operate in urban and mission critical networks. Both scenarios share characteristics such as high node density, mobility and interaction, which are prone to the use of *dLife*.

- Urban Area Networks: this scenario comprises a high number of fixed and mobile devices that could be used to disseminate/exchange information. Unfortunately, this is not feasible due to security reasons (i.e., proprietary access points), egoistic behavior (i.e., lack of incentives for users to cooperate with one another), high cost of providers, among others. Users can benefit from *dLife* in this scenario as the protocol could allow them to disseminate/exchange content any time just by considering their social interactions. There would be no need for accessing expensive services. Additionally, the regular egoistic behavior would lead users to automatically engage in cooperation, as they know that, by carrying content on behalf of others, it is enough for them to have some level of connectivity even in places where no connectivity exists.

- Mission-critical Networks: this scenario involves natural catastrophes or hostile environments (e.g., war zone). Networks formed here cannot rely on the infrastructure as they may be destroyed due to floodings, earthquakes, bombings. However, any Wi-Fi enabled device (e.g., PDAs, cell phones, laptops, APs) could be used, assuming the humanitarian behavior of the owners of such devices. This way, networks could be formed taking into consideration the social interactions among the civilians, police, rescue teams, medics and soldiers to improve the dissemination of relevant information to each of these agents. Consequently, this would result in better disaster-relieving actions in the affected areas.

5.1.2.2 Architecture of a *dLife* Node

The architecture of a *dLife* node comprises five main computational components, as illustrated in Fig. 5.1. Such components compute the social weights and node importance (cf. Sec. 4.2 and 4.3), define the forwarding strategies (cf. Sec. 5.1.1), and implement interfaces to the Bundle Agent (as defined in the Bundle Protocol [72]) and to the lower layers.

The envisioned components are:

- Social Information Gatherer (SIG) - gathers information on the contact duration between the current node and encountered nodes. This is done for each of the daily samples, corresponding to different periods of time in the daily routine of a person. Additionally, SIG obtains social weights and importance of encountered nodes (i.e., potential next forwarders) to be used for forwarding decisions when required.

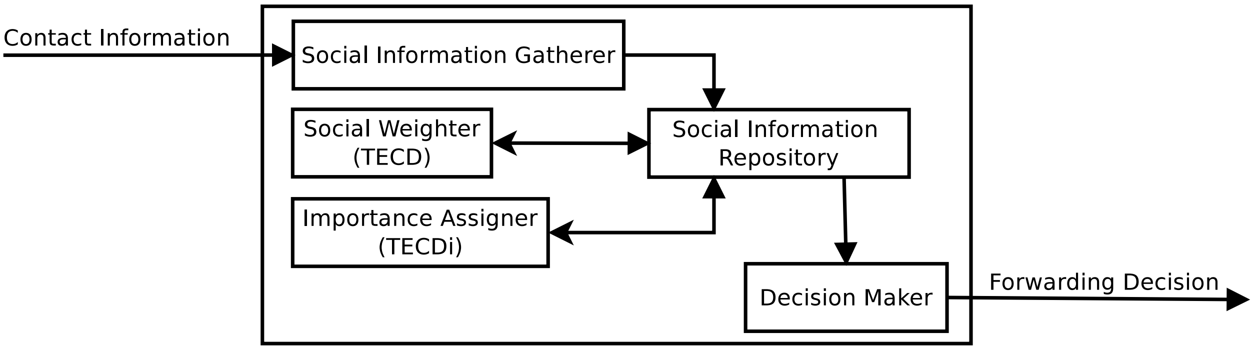


Figure 5.1: Architecture of a *dLife* node

- Social Information Repository (SIR) - stores information regarding contact duration of encounters, social weights and importance of encountered nodes and the current node.

- Social Weighter (SW) - interacts with SIR to get contact duration information at the end of every daily sample. Such information is used to determine the total contact time between the current node and encountered nodes, and the average duration of such contacts in order to compute the social weight between nodes (i.e., *TECD*).

- Importance Assigner (IA) - interacts with SIR to get the social weight between the current node and encountered nodes, and the importance of these encountered nodes at the end of every daily sample. This information is then used to compute the importance of the current node (i.e., *TECDi*).

- Decision Maker (DM) - interacts with SIR upon a new encounter to obtain relevant information in order to decide whether replication should occur.

Besides the forwarding strategy in Sec. 5.1.1 defined in the specification as basic strategy, a second strategy called *prioritized* is defined. It is similar to the basic forwarding strategy, but it prioritizes bundles destined to the encountered node.

Finally, as for the interfaces, *dLife* requires one to the Bundle Agent and another to the lower layers:

- Bundle Agent interface: as expected, *dLife* is only responsible for forwarding decisions, and the Bundle Agent is the one responsible for sending and receiving bundles between peers. Through this interface, the *dLife* routing agent knows about the bundles in the node, and allows it to inform the Bundle Agent about the bundles to be sent to a peering node. These are the interfaces and functionalities that the Bundle Agent is expected to provide to *dLife* routing agent:

  – Get Bundle List: provides the *dLife* routing agent with a list of the stored bundles and their attributes.

  – Send Bundle: notifies the Bundle Agent to send a specific bundle.

  – Drop Bundle Advice: advises the Bundle Agent that a specific bundle may be dropped if appropriate.

  – Acked Bundle Notification: informs *dLife* routing agent whether a bundle has been delivered to its final destination and time of delivery.

- Lower Layers interface: *dLife* needs to be aware of the presence of neighboring nodes, as well as when these nodes are not anymore in the vicinity in order to start accounting for the contact duration between nodes. Thus, these are the interfaces and functionalities that the lower layers are expected to provide to *dLife* routing agent:
  - New Neighbor: informs the *dLife* routing agent of the presence of a new node currently within communication range of the current node, based on the used wireless networking technology.
  - Neighbor Gone: informs the *dLife* routing agent that one of its neighboring nodes is out of communication range.

5.1.2.3 *dLife* Messages

dLife messages comprise a header followed by one or more Type-Length-Value (TLVs) components. For the sake of simplicity, this section briefly explains the purpose of each TLV that may be included in *dLife* messages. For further details on the fields in the header and TLVs, the reader is encouraged to refer to the *dLife* Internet-Draft [16].

- Hello TLV: sets up a link between two *dLife* nodes. Hello messages are the first messages to be exchanged between neighboring nodes and contain the Endpoint Identifier (EID), storage capacity, and current time of the node and a timer value. Nodes store the information in this message, and acknowledge by signaling that the communication has been established. This is done by means of an ACK that, if failed to be received, disconnection occurred and link is assumed broken.
- ACK TLV: informs the peering node that i) the connection must be broken due to storage constraints; ii) the Social TLV was correctly received; and iii) there are discrepancies/errors with EIDs.
- EID Dictionary TLV: contains the list of EIDs used by the peering nodes to make routing decisions, and it is used to synchronize this information between the peering nodes.
- Social TLV: contains the list of nodes that the peering node has encountered, its social weight towards them, and the importance of the peering node (SWNI), as well as a list of bundles carried by the peering node (bundleList), and a list of acknowledged bundles (ackedBundleList). This TLV counts with a "More Social TLVs" flag that informs whether (1) or not (0) the social message requires more TLVs to be sent in order to be fully transferred.

5.1.2.4 Protocol Operation

The operation of the *dLife* protocol follows the states depicted in Fig. 5.2. All nodes start at the *sensing* state. Upon a new contact (1), notified by the lower layer, the *dLife* node starts counting the contact duration with this peering node and the hello procedure is initiated. As the protocol is being worked to support multiple contacts, it also remains at the *sensing* state to detect other contact opportunities. If by any reason (e.g., peer moved away or ran out of battery) the peer is out of communication range, the lower layer notifies (2) the *dLife* module, which in turn stops counting contact duration towards the peer and remain in the *sensing* state awaiting for other peers.

Once in the *hello* state, the *dLife* node awaits for the Hello TLV. Upon the reception of this TLV (3), the *dLife* node acknowledges the successful reception and saves the information in this message.

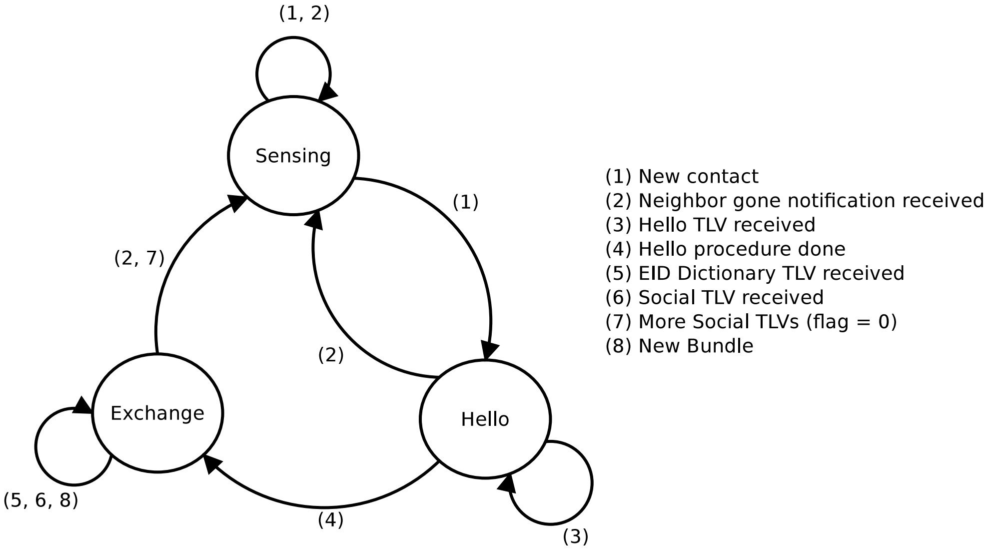


Figure 5.2: Operation of *dLife* protocol

This means that the hello procedure is done (4), and the *dLife* node can shift to the *exchange* state. If the lower layer notifies that the peer is not in the vicinity anymore (2), the *dLife* node stops counting contact duration towards the peer and shifts back to the *sensing* state.

As soon as the *dLife* node enters the *exchange* state, it sends its EID dictionary and SWNI information to the peer. The *dLife* node remains in this state until: i) the EID dictionary of the peer is received (5), which is used to update its local dictionary; ii) the Social TLV of the peer is received (6), that is used by its routing agent to inform the Bundle Agent about the bundles to be forwarded to the peer; and iii) there are new bundles (8) to be sent to the peering node. The *dLife* node shifts back to the *sensing* state if i) the lower layer notifies that the peer is gone (2), or ii) the More Social TLVs flag is set to 0 (7).

It is important to note that, for the sake of simplicity, the inner states for both *hello* and *exchange* states have been omitted. These inner states make sure that the link is established between the peering nodes prior to the *exchange* state and whether the EID dictionaries have no errors, *ackedBundleList* is updated, among others further detailed in [16].

5.2 Opportunistic Routing Based on Content

As one can observe, *dLife* makes use of social weights among nodes and their importance to deliver content between a specific pair of nodes. Additionally, one can easily notice that *dLife* and *SCORP* differ in what social weight reflects.

The differentiating aspect between these social-aware opportunistic routing solutions concerns the type of information abstracted from the computed social weights. In the context of *dLife*, the social weight provides information about the *probability of encountering nodes with similar daily social habits*, while in its content-based counterpart, *SCORP*, social weight is understood as the *probability of encountering nodes with a certain interest among the ones that have similar daily social habits*.

It has been shown that focusing on the content, and not on the host, can improve the performance of challenged networks [14, 18] by allowing an efficient direct communication between producers and consumers of information. In addition, exploiting nodes' social interactions and structure (i.e., communities [13], levels of social interaction [15, 17]) has increased opportunistic routing performance. Thus, combining content knowledge (i.e., information type, interested parties) with social proximity shall bring benefits (faster, better content reachability) in challenged networks.

Reasons that motivate bringing content knowledge and social awareness together are: i) nodes with a similar daily habits have higher probability of having similar (content) interest [14]; and, social awareness allows a faster dissemination of data taking advantage of the more frequent and longer contacts between closer nodes. Thus, this section introduces how the Social-aware Content-based Opportunistic Routing Protocol (*SCORP*) functions.

As explained in Sec. 5.1, *dLife* considers how nodes interact in their daily routines. Based on that, this routing solution can determine the social strength existing between nodes based on the duration of their contacts (by means of *TECD*) and compute the importance of nodes in specific periods of time (by means of *TECDi*).

The *SCORP* proposal adapts the *TECD* utility function to shift focus from the hosts to the content, thus introducing content knowledge and becoming able to perform point-to-multipoint delivery. This resulted in the *TECI* utility function presented in Sec. 4.5.

SCORP then exploits social proximity and content knowledge to augment the efficiency of data delivery in urban, dense scenarios.

5.2.1 *SCORP* Algorithm

Alg. 5.2[1] shows the simple operation of *SCORP*: when the $CurrentNode$ meets a $Node_i$ in a daily sample ΔT_k, it gets a list with all interests $Node_i$ had contact in that daily sample and the social weights towards the nodes having these interests ($Node_i$.weightsToAllinterests computed based on Eq. 4.10). $Node_i$ also sends a list of the messages it is carrying ($Node_i$.carriedMessages). Then, every $Message_j$ in the buffer of $CurrentNode$ is replicated to $Node_i$ if:

- $Node_i$ has interest ($Node_i$.getInterests) in the message's content ($Message_j$.getContentType); or
- the social weight of $Node_i$ towards an interest (i.e., $Message_j$.getContentType) is greater than the weight of the $CurrentNode$ towards this same interest.

Algorithm 5.2 Forwarding with *SCORP*

begin
 foreach $Node_i$ encountered by $CurrentNode$ **do**
  receive($Node_i$.weightsToAllinterests and $Node_i$.carriedMessages)
  foreach $Message_j \in$ buffer.($CurrentNode$) & $\notin$ buffer($Node_i$) **do**
   if ($Message_j$.getContentType $\in Node_i$.getInterests)
    then $CurrentNode$.replicateTo($Node_i$, $Message_j$)
   else if ($Node_i$.getWeightTo($Message_j$.getContentType) >
      $CurrentNode$.getWeightTo($Message_j$.getContentType)
    then $CurrentNode$.replicateTo($Node_i$, $Message_j$)
end

[1]Springer and the original publisher (Ad Hoc Networks, v. 129, 2014, p. 100-115, Social-Aware Opportunistic Routing Protocol Based on User's Interactions and Interests, Waldir Moreira, Paulo Mendes, Susana Sargento, Algorithm 1, Copyright © 2014, Institute for Computer Sciences, Social Informatics and Telecommunications Engineering) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media [19].

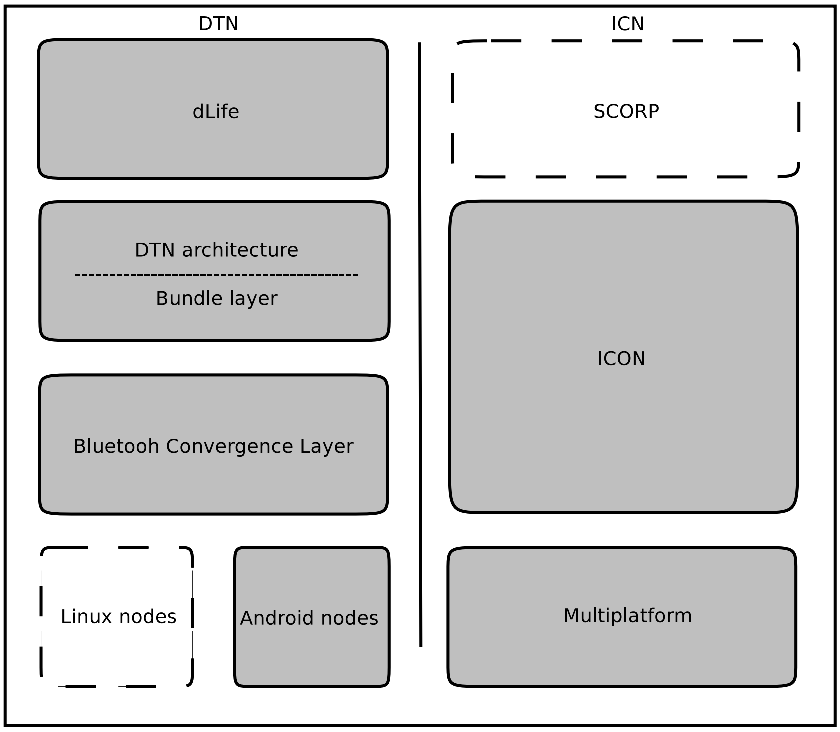


Figure 5.3: DTN-Amazon project

Based on this, *SCORP* is expected to replicate only to nodes that do have interest in the content in the message to be forwarded, or that are socially well connected to other nodes that have that specific interest. Consequently, improvements to resource usage and delivery latency are expected due to the creation of less replicas.

5.3 Implementation Context

It is worth mentioning that the work done in this chapter is part of the DTN-Amazon project between COPELABS and Federal University of Pará (UFPA). This project aims at mitigating the effects of digital divide and social exclusion in the riverside communities close to the UFPA campus in Belém, Pará, Brazil.

The project comprises two main streams following the Delay-Tolerant and Information-Centric Networking paradigms as illustrated in Fig. 5.3. Both proposals devised in the context of this Thesis work, *dLife* and *SCORP*, comprise the routing mechanisms to be employed in the project. It is important to note that the shaded boxes represent the parts that are currently under development or finetuning. The boxes with dashed lines have either been tested in simulator (*SCORP*) or are to be developed.

The DTN stream has resulted in what is called the SocialDTN [34], an Android application based on the DTN architecture [3], Bundle Protocol [72], and *dLife* Internet-Draft [16]. With SocialDTN, the idea is to exploit social proximity and interactions between agents (e.g., health and teachers) acting in these communities, and be independent of any network infrastructure. This is to facilitate the seamless exchange of content (i.e., videos, photos) between these agents in order to improve their actions. Additionally, SocialDTN is expected to allow the asynchronous communications between the agents without relying on infrastructure, which is in most cases inexistent in these areas.

SocialDTN comprises a Bluetooth Convergence Layer (BCL) that allows nodes to communicate over the Bluetooth technology and to exploit the aforementioned social proximity and interactions. Regarding routing, *dLife* is the choice as it fits in the context of social interaction.

Still in DTN stream, another work extended the IBR-DTN [73] to also support the BCL but with a bundle compression control scheme. The idea with this extension [74] is to improve the exchange of data over the short-lived contact opportunities that may happen between nodes acting in the Amazon riverside communities.

As for the ICN stream, an Information and Content-oriented Opportunistic Networking (ICON) [75] approach has been developed based on the content/information-centric paradigm. This means that the focus is not on the host, but instead in the content traversing the network. As *SCORP* has shown great potential for content-based routing, it will be part of ICON as routing choice.

It is important to mention that both SocialDTN and ICON are being developed in a modular way in other to comprise different operating systems, as well as to allow the development of other schemes for routing, naming, among others.

5.4 Summary of the Chapter

This chapter presented the social-aware opportunistic routing *dLife*, which considers the social daily routine of users to make routing decisions. By combining *TECD* and *TECDi*, *dLife* aims at choosing only socially well-connected nodes to perform forwarding. Its algorithm as well as its specification (i.e., application scenarios, node architecture, relevant messages, and operation) are discussed throughout this chapter.

Then, the Social-aware Content-based Opportunistic Routing Protocol (*SCORP*) and its algorithm are introduced. This second opportunistic routing approach is based on *TECI* and takes into account the content traversing the network and the interest that the users have in such content. *SCORP* emerged given the potential observed in terms of performance improvements and wise resource usage when considering content to forward data [52, 70, 71].

To conclude, the chapter presented the context of implementation of both *dLife* and *SCORP*, which is part of DTN-Amazon project. This project is a partnership between COPELABS and Federal University of Pará (UFPA), and aims to reduce the digital divide and social exclusion observed in isolated Amazon regions.

Chapter 6

Performance Evaluation

The focus of this chapter is on the performance evaluation of the social-aware utility functions as well as the social- and content-based opportunistic routing proposals developed in the context of this Thesis work. For that, different opportunistic routing benchmarks were taken into account to help assessing the performance of the proposed work.

To facilitate presentation and understanding, the chapter is divided into three sets of experiments dedicated to the performance analysis of the: i) social-aware utility functions (Sec. 6.3); ii) opportunistic routing based on user social daily routine, *dLife* (Sec. 6.4); and iii) opportunistic routing based on content, *SCORP* (Sec. 6.5).

The chapter starts by presenting the evaluation methodology employed throughout the experiments. Then, a section is dedicated to the common setup used in these sets of experiments. It is important to note that the experimental setup may vary according to the goal in each set of experiments, but generally speaking, the setup comprises scenarios based on synthetic mobility models and human traces, and exposes all the proposals to the same conditions (e.g., generated load, pair of communicating nodes, transmission rate and range, number of nodes, among others). The differences in setups are mentioned in each corresponding experimental result section when applicable.

6.1 Evaluation Methodology

Performance analysis is carried out on Opportunistic Network Environment (ONE) [68] simulator, and all results are presented with a 95% confidence interval.

Proposals are assessed in terms of the following performance metrics: average delivery probability (i.e., *ratio between the number of delivered messages and total number of created messages*), average cost (i.e., *number of replicas per delivered message*), and average latency (i.e., *time elapsed between message creation and delivery*).

It is worth mentioning that, for the experiments involving content-oblivious and content-oriented proposals (cf. Sec. 6.5), the average delivery probability metric is defined as the *ratio between the number of delivered messages and the total number of messages that should have been delivered* for fairness purposes while evaluating the considered proposals.

6.2 Common Experimental Setup

The synthetic mobility model comprises different mobility patterns. It simulates a 12-day interaction in the city of Helsinki between 150 nodes divided into 8 groups of people and 9 groups of vehicles. Each node may have a 250-kbps Bluetooth interface with 10-meter communication range and/or 11-Mbps Wi-Fi interface with 100-meter communication range.

One vehicle group (10 nodes) follows the *Shortest Path Map Based Movement* mobility model, and represents police patrols that randomly choose destinations and use the shortest path to reach them with waiting times ranging from 100 to 300 seconds. The remaining 8 vehicle groups (each with 2 nodes) represent buses following the *Bus Movement* mobility model with waiting times ranging from 10 to 30 seconds. The speed of vehicles range from 7 to 10 m/s.

The groups of people have different number of nodes: group A has 14 nodes; groups C, E, F, and G have 15 nodes each; groups B and D have 16 nodes each; and group H has 18 nodes. People have walking speeds between 0.8 to 1.4 m/s following the *Working Day Movement* mobility model, and may use the bus to move around. Each group was configured to have different offices, meeting spots, and home locations. Each person has an average of 8 daily working hours, and walks around the office with pause times between 1 minute and 4 hours. These people also have a 50% probability of having a leisure activity after work, which may be done alone or in group and last up to 2 hours.

For the trace-based scenario, the CRAWDAD human traces [76] are used. It represents a period of two months while 36 Cambridge University students moved throughout their daily routines.

The load generated in these scenarios are equivalent to 6000 messages being created among the same source/destination pairs across all experiments. The size of messages ranges from 1 kB to 100 kB, as to represent the different applications running over opportunistic networks (e.g., asynchronous chat messages, e-mails). The available buffer space is also limited, as users may not be willing to share all of the storage capacity of their devices. Regarding message TTL, it may vary from days/weeks up to unlimited, as to observe the performance behavior (i.e., buffer consumption, number of replicas) of the studied proposals.

6.3 Evaluation of Social-aware Utility Functions

This section presents the performance evaluation of *TECD* and *TECDi* utility functions against contact-based and social-based proposals.

The choices for the contact-based *PROPHET* [5] and *Epidemic* [4] rely on the fact that they are the most used benchmarks (cf. Sec. 3.2), with the former being widely recognized within the DTN research community, and the latter for representing a flooding-based approach that reaches high delivery rates. Regarding the social-based *Bubble Rap* [13] and *PeopleRank* [11], their selection relates to the fact that they are good representatives of proposals based on social structures and node popularity.

It is important to mention that *PeopleRank* is represented by *Rank* that, instead of determining the overall importance of a node at each encounter, as *PeopleRank* does, it estimates the node importance at the end of each daily sample [33].

The section starts by presenting the experimental setup. Then it is further divided showing the evaluation of *TECD* against *PROPHET* and *Epidemic*; and the evaluation of *TECDi* against *Bubble Rap* and *Rank*. The results on this section have been published in [33].

6.3.1 Experimental Setup

For this set of experiments, only the synthetic mobility model scenario is considered. Moreover, all nodes are equipped with Bluetooth (250 kbps/10 m), and the buses also have Wi-Fi (11 Mbps/250 m) to facilitate bus-to-bus data exchange. Due to limitations of the *Bubble Rap* implementation in ONE, all nodes are equipped with only Wi-Fi (11 Mbps/100 m) interface for the experiments with this proposal.

The load generated and message buffer sizes follow the common setup (cf. Sec. 6.2). Message TTL is set at 1 day (i.e., 24 hours) and unlimited, as the goal is: i) to observe the impact on delivery and cost according to the life time of messages in the experiments; and, ii) to assess the performance of the devised social-based utility functions.

Regarding the proposals, both *Rank* and *TECDi* have the dumping factor set to 0.8, as this value lies among the ones in which *PeopleRank* [11] had the best success rates. *Bubble Rap* [13] considers the K-clique and single window algorithms for community formation and node centrality computation, respectively. Also, its parameter k is set to 5, as *Bubble Rap* presents the best overall performance in terms of delivery probability, cost and latency.

6.3.2 Evaluation against Contact-based Algorithms

The *TECD*-based routing proposal only forwards a message to a new forwarder if it has a strong social relationship to the destination of the message (cf. Fig. 6.1[1]). This is of advantage to *TECD* (cf. Fig. 6.1(a)), which has a 16.17% and 23.77% gain over *PROPHET* and *Epidemic*, respectively, for the 24-hour TTL case. For the unlimited TTL case, the advantage of the *TECD*-based routing proposal remains with a 18.41% and 23.26% gain over *PROPHET* and *Epidemic*, respectively.

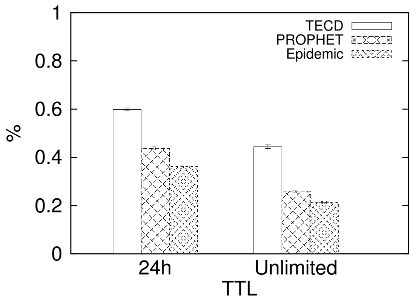
As *TECD* captures the level of social ties, messages are only forwarded to nodes that are socially well connected to the destinations, even if that means that the carrier has to hold to the messages for longer times. *PROPHET* experiences occasional loops which affect its performance by taking the opportunity of exchanging other messages, thus being reflected in this proposal's delivery probability. As *Epidemic* relies on an extreme replication approach, it experiences a quick exhaustion of buffer space which is set to 2 MB.

Regarding the average cost (cf. Fig. 6.1(b)), *TECD* produces much less replicas (57.37) to have a successful delivery for the 24-hour TTL case, when compared to *PROPHET* (272.35) and *Epidemic* (454.12). As for unlimited TTL, *TECD* presents an increase in the number of replicas (121.30) to perform a delivery, but it still remains lower than the number of replicas created by *PROPHET* (247.11) and by *Epidemic* (535.04).

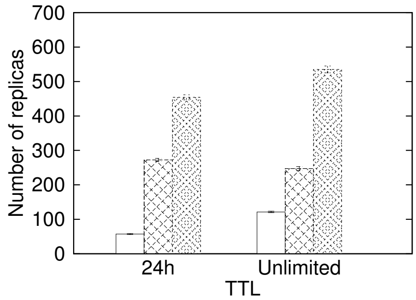
TECD overcomes *PROPHET* as its forwardings take place based on the social strength between nodes. This reduces the number of replicas, since carrier nodes keep messages for longer times until a socially well-connected next hop comes within communication range. The observed increase in the number of replicas of *TECD* for the unlimited TTL case was expected, since messages live longer in the system. *PROPHET* presents a lower cost for the unlimited case, as the long-lived messages contribute for the exhaustion of buffer space. Finally, *Epidemic* is expected to have an increase in cost as TTL increases, leading to an approximately 18% increase for the unlimited TTL case.

[1] Springer and the original publisher (Ad-hoc, Mobile, and Wireless Networks, v. 7363, 2012, p. 98-111, Study on the Effect of Network Dynamics on Opportunistic Routing, Waldir Moreira, Manuel de Souza, Paulo Mendes, Susana Sargento, Figure 2, Copyright © 2012, Springer-Verlag Berlin Heidelberg) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media [33].

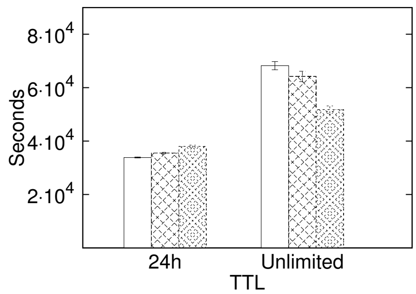
Regarding average latency (cf. Fig. 6.1(c)), *TECD* takes less time to deliver its messages for the 24-hour TTL case: 1672.82 s and 4173.53 s lower than *PROPHET* and *Epidemic*, respectively. Despite of taking more time to decide on the next hop, *TECD* chooses nodes that have a strong social relationship with destinations, and this reduces its number of hops (3.23) towards such destinations when compared to *PROPHET* and *Epidemic* (3.66 and 11.10, respectively) and improves delivery time. While the latency behavior of *Epidemic* is mostly due to the paths used to reach the destination, *PROPHET*'s delivery time is increased also due to the identified occasional loops.



(a) Average delivery probability



(b) Average cost



(c) Average latency

Figure 6.1: Evaluation of contact-based metrics

As for the unlimited TTL case, *TECD* holds messages longer as to wait for a better forwarding opportunity in future contacts. It is also observed that *PROPHET* behaves similarly to what is reported in its original paper [5]: as TTL increases, so does its latency. This relates to the fact that *PROPHET* does not clear already delivered messages from the system. This in turn occupies buffer and takes the opportunity of undelivered messages. Regarding *Epidemic*, since it floods the network with many copies of messages, this also reduces the delivery time of few messages. Consequently, it experiences a latency reduction, yet with a very high associated cost (cf. Fig. 6.1(b)).

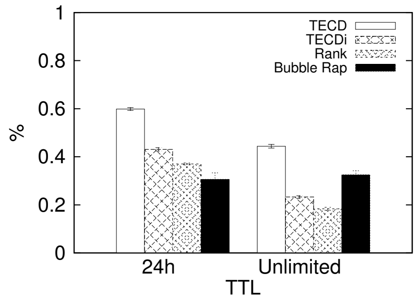
6.3.3 Evaluation against Social-based Algorithms

In this section, the performance of *TECD* and *TECDi* is analyzed against *Bubble Rap* and *Rank* (cf. Fig. 6.2[2]). The goal is: i) to analyze the advantages of considering social weight as to best compute

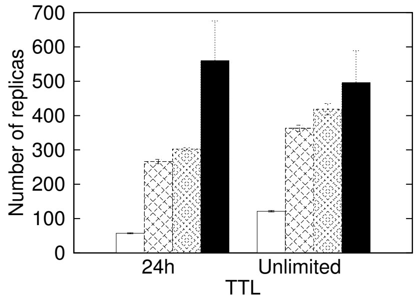
[2]Springer and the original publisher (Ad-hoc, Mobile, and Wireless Networks, v. 7363, 2012, p. 98-111, Study on the Effect of Network Dynamics on Opportunistic Routing, Waldir Moreira, Manuel de Souza, Paulo Mendes, Susana Sargento, Figure 2, Copyright © 2012, Springer-Verlag Berlin Heidelberg) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media [33].

the importance of nodes (used in *Rank*); and, ii) to analyze how time-evolving social-based approaches perform against proposals that only consider the identification of social structures such as *Bubble Rap*.

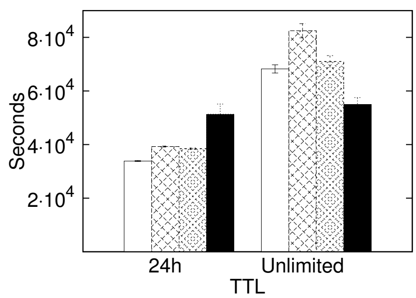
Regarding the average delivery probability (cf. Fig. 6.2(a)), *TECD* outperforms *Rank* and *Bubble Rap* by 22.84% and 29.26%, respectively, for the 24-hour TTL case. As for the unlimited TTL case, one can observe a reduction in the delivery performance of *TECD*, *TECDi*, and *Rank*, with *Bubble Rap* being the only one to improve as TTL increases.



(a) Average delivery probability



(b) Average cost



(c) Average latency

Figure 6.2: Evaluation of social-based metrics

TECDi has a 6.06% and 4.88% gain over *Rank* for the 24-hour and unlimited TTL cases, respectively. When compared to *Bubble Rap*, *TECDi* has a 12% gain for the 24-hour TTL case, given the fact that *Bubble Rap* has to form communities affecting its delivery. This advantage ceases for the unlimited TTL case in which *Bubble Rap* has a 9% gain over *TECDi*. With messages staying longer in the system, *Bubble Rap* takes this opportunity to properly form communities, and uses such information to reach destinations as reported in [13].

Bringing social strength and node importance together has benefits: *TECDi* determines the importance of nodes considering not only their importance but also the social strength among them. This means that a node with socially weak neighbors has a lower importance with *TECDi* than with *Rank*. And consequently, important nodes with *TECDi* are those with strong social ties to other important nodes in the network.

Regarding the average cost (cf. Fig. 6.2(b)), *TECD* produces the least number of replicas (57.37) for the 24-hour TTL case by considering social strength towards destinations. *TECDi* has the second best cost performance behavior, since next hops are chosen based not only on their degree, but also on the strength of the relationships towards their neighbors. For the unlimited TTL case, the cost for *TECD*, *TECDi*, and *Rank* increases as messages live longer and therefore can be further replicated. *Bubble Rap* replicates more as it relies on the global centrality (while communities form) to do it so. Thus, high centrality nodes (e.g., buses and police patrols) are always receiving content contributing to high cost values. The subtle cost reduction of *Bubble Rap* for the unlimited TTL case refers to buffer exhaustion, leading to less replications.

As for the average latency (cf. Fig. 6.2(c)), for the 24-hour TTL case, *TECD* delivers its messages in shorter time than *Rank* and *Bubble Rap* (less 4598 s and 17476 s, respectively), while *TECDi* takes more time than *Rank* (818.37 s) but still lower than *Bubble Rap* (12059 s). Regarding the unlimited TTL case, *TECD* still takes less time (2752 s) to deliver its messages than *Rank*, but loses its advantage to *Bubble Rap*, taking an average of 13299 s for a delivery. As mentioned before, the latency increase of *TECD* is due to its attempt in finding better next hops by holding the messages longer. With *TECDi*, once messages reach high importance nodes, they may take longer to reach destinations if these top-ranked nodes seldom interact outside their social groups.

The latency experienced by *Bubble Rap* includes the time to form communities and the time to find high centrality nodes for the 24-hour TTL case. Since latency is a function of the delivered messages and its delivery rate increased for the unlimited TTL case, the increase in the time for delivery messages of *Bubble Rap* was expected.

Both *Rank* and *TECDi* present higher latency than *TECD*, since messages are held longer especially when the carrier node has a high importance factor in the network. The importance factor explains the reason why *TECDi* takes longer to deliver its messages when compared to *Rank*, as it also considers the social strength among nodes.

6.3.4 Summary

This section presented the potential of considering the user dynamic while computing social functions. The proposed social-aware utility functions, *TECD* and *TECDi*, are based on the daily routine of users, and weigh the social ties among these users.

As presented, network dynamics have its impact on the performance of opportunistic routing solutions: *TECD* presents the best overall performance amongst the studied contact- and social-based proposals. It also outperforms *TECDi* with delivery gains up to 21.1%, and lower cost (242 less replicas) and latency (approximately 17.3%). This is explained by the fact that *TECD* uses social strength for message forwarding, which is much more reliable than node importance (message may be stuck with high importance nodes that have little social interaction with destinations). Consequently, *TECDi* has unnecessary cost increase that affects its delivery capability and time. Yet, when compared to the remaining proposals, *TECDi* has shown potential in improving forwarding, and the combination of this approach with *TECD* is further investigated in the next section.

6.4 Evaluation of Opportunistic Routing Based on User Social Daily Routine

This section presents the performance evaluation of *dLife*, which emerged from the combination of *TECD* and *TECDi*. *dLife* is evaluated against its community-based variant *dLifeComm* [15] and also social, community-based *Bubble Rap* [13].

The goal is to show the potential of the social-aware *dLife* proposal independently of social structures and node centrality. Additionally, the impact of centrality metrics is analyzed as they can lead to the appearance of bottlenecks.

The section starts by describing the experimental setup. Then, the performance results of *dLife*, *dLifeComm*, and *Bubble Rap* are further divided according to the scenario used, namely synthetic mobility and human trace. The results on this section have been published in [15].

6.4.1 Experimental Setup

For this set of experiments, both scenarios based on synthetic mobility models and human traces are considered. All nodes are equipped with a Wi-Fi interface (11 Mbps/100 m).

The load generated and message buffer sizes follow the common setup as before (cf. Sec. 6.2). Message TTL varies between 1, 2, 4 days and 1, 3 weeks. These values were chosen based on the ones in which *Bubble Rap* was reported to have the best performance behavior in terms of delivery probability and cost [13].

Regarding the proposals, both *dLifeComm* and *Bubble Rap* consider K-Clique and cumulative window algorithms for community formation and node centrality computation, respectively [13]. The parameter k is set to 5. As *dLife* measures social weights considering different time periods (cf. Sec. 5.1), the number of daily samples is set to 24 (i.e., each of one hour).

6.4.2 Evaluation over Synthetic Mobility Scenario

This section presents the evaluation of *dLife*, *dLifeComm*, and *Bubble Rap* over the synthetic mobility scenario (cf. Fig. 6.3, Copyright © 2012 IEEE [15]). Regarding the average delivery probability as shown in Fig. 6.3(a), *dLife* and *dLifeComm* have performances up to 39.5% and 31.2%, respectively, better than *Bubble Rap*.

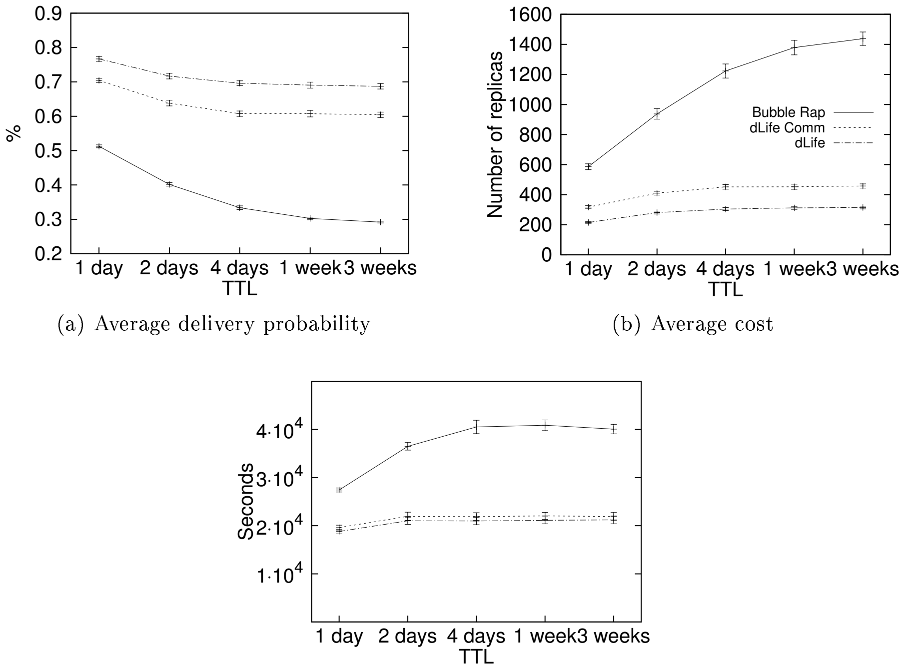
As communities are not readily available (they are formed as nodes interact), *Bubble Rap* takes forwarding decisions based on the node global centrality. However, nodes in this scenario present a very heterogeneous centrality: only approximately 17% of the nodes have very high centrality, while the remaining nodes have either mid or low centrality. As a great part of the messages have mid/low centrality sources, message replication is increased, as *Bubble Rap* replicates upon encountering a high centrality node. This impacts its delivery capability, since the buffer exhausts quickly and is further degraded as TTL increases (messages replicate more as they are allowed longer in the network).

dLife and *dLifeComm* also experience some level of buffer exhaustion with TTL increase, and the performance of *dLifeComm* is further degraded due to the overhead with community formation. However, such performance penalty is lessened as *dLifeComm* is able to capture the dynamism of the behavior of nodes.

As for the average cost (cf. Fig. 6.3(b)), *dLife* and *dLifeComm* require a much lower number of replicas (up to 78% and 68%, respectively) when compared to *Bubble Rap* to perform a successful delivery. This performance behavior is a result of the wise forwarding decisions made by these proposals based on *TECD* and *TECDi*. Thus, such utility functions allow *dLife* and *dLifeComm* to have a social graph encompassing strong edges and very important vertices.

Although employing a scheme to discard messages that reach the destination's community, the cost of *Bubble Rap* is expected to increase with TTL, since new replicas of these messages are still being created by other carriers. This explains the high cost of this proposal.

Fig. 6.3(c) presents the average latency in which *dLife* and *dLifeComm* take less time to deliver messages (48.3% and 46.1%, respectively) than *Bubble Rap*. One can easily see how taking forwarding



(a) Average delivery probability

(b) Average cost

(c) Average latency

Figure 6.3: Evaluation over synthetic mobility scenario

decisions based on the dynamic social behavior of nodes is beneficial: with *TECD* and *TECDi*, forwarding takes place based on social strength between nodes and their importance in specific time periods. Consequently, messages reach their destinations faster, as both utility functions reflect the ability of encountering such destinations in the near future.

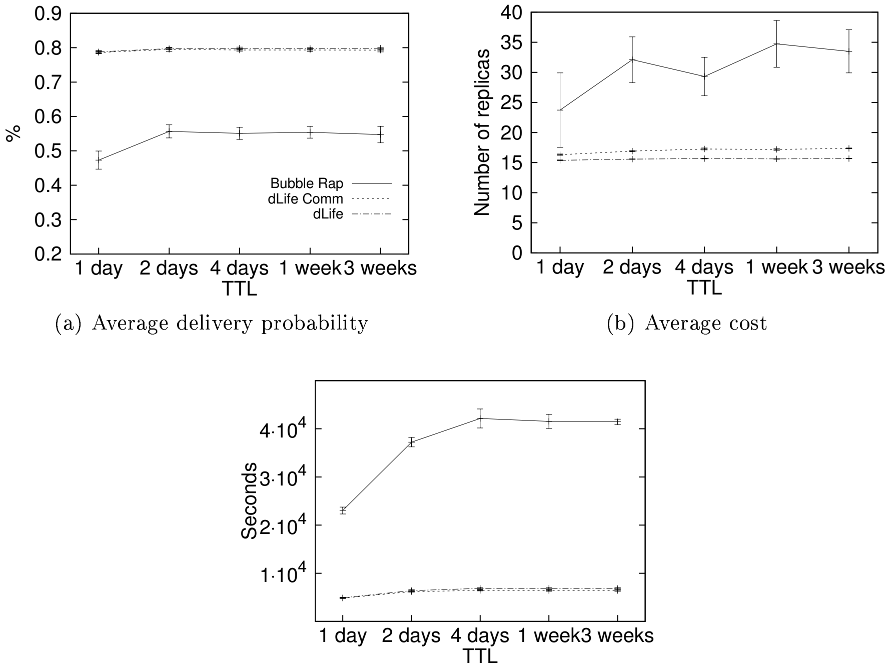
By not capturing such dynamism, replicas created with *Bubble Rap* take longer to be delivered, since this proposal is not aware of the strength of the social ties between the potential next hops for the message and its destination.

6.4.3 Evaluation over Trace-based Scenario

This section presents the evaluation of *dLife*, *dLifeComm*, and *Bubble Rap* over the trace-based scenario (cf. Fig. 6.4, Copyright © 2012 IEEE [15]). Fig. 6.4(a) shows the average delivery probability, where *dLife* and *dLifeComm* reach up to 31.5% and 31.3%, respectively, better performance than *Bubble Rap*.

It is observed that the delivery performance of *Bubble Rap* increases with TTL, as reported in Hui et al. (2011) [13]. For the 2-day TTL case, *Bubble Rap* is able to reach more destinations. Yet, its delivery increasing trend halts for the the remaining TTL cases, since it relies on high centrality nodes to deliver inside the destination community. The issue resides on the fact that the high centrality nodes might not be socially well connected to the wanted destinations, and the benefit of higher TTLs is not exploited any further by *Bubble Rap*.

Being able to capture the dynamics of user social behavior in specific time periods lead both *dLife* and *dLifeComm* to good performance. In the case of *dLifeComm*, its performance is further improved when compared to *Bubble Rap*, since it also relies on node importance to find destinations instead of simply considering the communities that nodes belong to, and their cumulative centrality which does



(a) Average delivery probability

(b) Average cost

(c) Average latency

Figure 6.4: Evaluation over trace-based scenario

not reflect reality.

As the number of formed communities is small (approximately 6.7) and most of the nodes are encompassed by these communities, this explains the similar behavior of *dLife* and *dLifeComm*, as the latter proposal does not suffer as much with community formation overhead, as seen in Sec. 6.4.2. Interestingly, from these results one can observe that centrality has a greater impact than the notion of community formation on social-based opportunistic routing. Centrality leads to the appearance of bottlenecks that can be seen when comparing *dLifeComm* that combines the notion of community and *TECD*/*TECDi*, and *Bubble Rap* that combines the notion of community and centrality.

As for the average cost presented in Fig. 6.4(b), *dLife* and *dLifeComm* produce up to 55% and 50.5% less replicas than *Bubble Rap*. For the 4-day TTL case, *Bubble Rap* displays a cost reduction that is linked to the sporadicity of contacts in the scenario: messages are created during periods of no contacts, which result in a lower number of replicas.

Regarding average latency, *dLife* and *dLifeComm* are able to deliver messages in 83.7% and 84.7%, respectively, less time than *Bubble Rap* for this scenario. Although nodes in such scenario have sporadic contact, *TECD* and *TECDi* are capable of distinguishing which nodes have strong social relationships and importance. This allows both *dLife* and *dLifeComm* to reach destinations in a few number of hops and, consequently, in less time.

Despite of experiencing also a reduction in the number of hops to reach destinations, *Bubble Rap* still presents almost similar latency behavior as presented in Sec. 6.4.2. Since communities are very few and almost all nodes belong to them, *Bubble Rap* resorts to centrality, which does not reflect reality. Thus, nodes receive messages based on their high centrality values, but the message takes more time to reach destinations, as such nodes may not be the best option to perform the expected delivery at that specific moment.

6.4.4 Summary

This section presented how the combination of the *TECD* and *TECDi* utility functions, comprising a new social-aware opportunistic routing proposal, is useful for data forwarding in opportunistic networks.

dLife shows that the dynamism of users' social daily behavior brings benefits to opportunistic routing, resulting in wiser forwarding decisions with suitable delivery probability, cost and latency performance. Additionally, it is clear that, when comparing *Bubble Rap* and *dLifeComm*, centrality ends up negatively impacting the system performance much more than the identified overhead associated to the notion of community formation.

6.5 Evaluation of Opportunistic Routing Based on Content

This section presents the evaluation of *SCORP* against *dLife* [15], *Bubble Rap* [13], and *Spray and Wait* [7]. Despite of being a social-oblivious solution, *Spray and Wait* is considered in this set of experiments as a lower bound in what concerns delivery cost.

The goal is to show the potential of combining social awareness with content information to improve routing in opportunistic networks.

The section starts by presenting the experimental setup, followed by the results obtained based on synthetic mobility models and trace-based scenarios. The results on this section have been published in [19].

6.5.1 Experimental Setup

For this set of experiments, the synthetic mobility models and human traces scenarios are considered. Also, all nodes are equipped with a Wi-Fi interface (11 Mbps/100 m).

It is worth remembering that, for these experiments, the average delivery probability metric is defined as the *ratio between the number of delivered messages and the total number of messages that should have been delivered*.

The load generated and message buffer sizes follow the common setup as before (cf. Sec. 6.2), and message TTL varies between 1, 2, 4 days and 1, 3 weeks for the synthetic mobility model scenario. As *SCORP* is content-oriented, the groups of people in this scenario have 10 different and randomly assigned interests that may overlap fully or partially with the interests of other groups. So, to achieve the same 6000-message load of the simulations of *Spray and Wait, Bubble Rap* and *dLife,* 170 messages with content matching attributed interests are enough to generate the same 6000 messages to be delivered throughout the simulation of *SCORP*.

Regarding the trace-based scenario, message TTL is set to 1 day, and the load varies according to the number of messages generated per destination (1, 5, 10, 20 and 35 from *Spray and Wait, Bubble Rap* and *dLife* sources) or to the number of interests nodes have (1, 5, 10, 20, and 35 for *SCORP* destinations). This results in a total of 35, 175, 350, 700, and 1225 messages being expected to be received throughout the simulations of the studied proposals. The msg/int notation in the figures denotes the number of different messages sent by *Spray and Wait, Bubble Rap* and *dLife* sources, or the number of different interests of each of the *SCORP* receivers.

To guarantee fairness for *Spray and Wait, Bubble Rap* and *dLife* in the human trace scenario, node 0 has no buffer size restriction to avoid message discarding, due to buffer constraint given the number of messages it has to generate. Additionally, the rate of message generation varies with the load: when the load is 1, 5, and 10 messages generated to each node, they are generated at a rate of 35 messages per day. As for the load with 20 and 35 messages, the rates are of 70 and 140 messages per day, respectively. This is done to allow *Bubble Rap* and *dLife* messages to be exchanged/delivered given the message TTL (i.e., 1 day).

Regarding the proposals, *Spray and Wait* runs in binary mode with the number of copies L set to 10. *Bubble Rap* uses algorithms for community formation and node centrality computation (K-Clique and cumulative window) [13]. *dLife* and *SCORP* consider 24 daily samples of one hour as mentioned in Sections 5.1 and 5.2.

6.5.2 Evaluation of TTL Impact

This section presents the impact of TTL (cf. Fig. 6.5[3]). The synthetic mobility model is used with varying message TTL to observe the impact of message TTL on the studied opportunistic routing proposals, and to choose the TTL value in which these proposals have the best overall performance. As a general remark, the average number of contacts per hour in this scenario is of 962, happening in a homogeneous manner.

Fig. 6.5(a) shows the average delivery probability. As *Bubble Rap* relies mostly on global centrality to perform forwardings, its performance is degraded due to the fact that very few nodes have high centrality (20%), and a high number of the messages is generated in low centrality nodes. This leads to more replications and consequently buffer exhaustion. This issue is further worsened as TTL increases.

dLife has a 21% advantage over *Bubble Rap*, since it captures the dynamism of node behavior. However, *dLife* takes longer to have a stable view of the network in terms of social weights due to the high number of contacts and their frequency. This results in unwanted replications, also leading it to experience buffer exhaustion and preventing other message to be delivered.

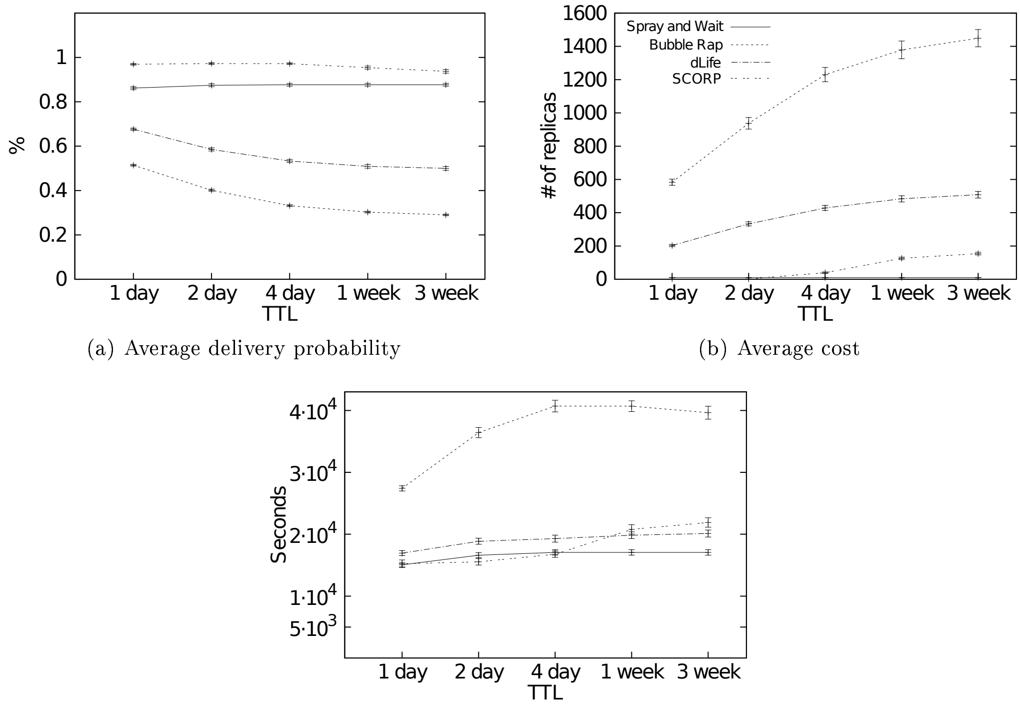
Spray and Wait overcomes *Bubble Rap* and *dLife* by up to 58.6% and 37.7%, respectively. This is due to the fact that the random replicas created by *Spray and Wait* reach nodes (i.e., buses and police patrols) covering most of the simulated area, and that have higher chance of coming into contact to the destination of messages.

Compared to the other proposals, *SCORP* reaches up to 64.7%, 44.5%, and 10.7% over *Bubble Rap*, *dLife* and *Spray and Wait*, respectively, by taking advantage of the interests that nodes share. This results in a quick message dissemination. Yet, *SCORP* experiences a subtle decrease in its delivery rate as message TTL increases. Since messages live longer in the network, few messages are discarded due to the number of forwardings and resulting buffer exhaustion.

Regarding the average cost (cf. Fig. 6.5(b)), *Bubble Rap* replicates more than the other proposals to perform a successful delivery, since messages are allowed to live longer in the network [13].

dLife relies on the social strength and node importance to replicate [15], which leads it to create up to 65.2% less replicas than *Bubble Rap* for the simulated TTLs. This relates to the fact that the social weights are more accurate than community formation. Given the subjective nature of the latter, communities are formed based on pre-defined and static contact duration, when in reality people consider much more than this to define their communities.

[3]Springer and the original publisher (Ad Hoc Networks, v. 129, 2014, p. 100-115, Social-Aware Opportunistic Routing Protocol Based on User's Interactions and Interests, Waldir Moreira, Paulo Mendes, Susana Sargento, Figure 2, Copyright © 2014, Institute for Computer Sciences, Social Informatics and Telecommunications Engineering) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media [19].



(a) Average delivery probability

(b) Average cost

(c) Average latency

Figure 6.5: Evaluation over synthetic mobility scenario

By considering interests, *SCORP* can easily identify the potential destinations for the carried messages. Consequently, *SCORP* produces up to 99.8% and 99.4% less replicas than *Bubble Rap* and *dLife*, respectively.

Spray and Wait is less costly as it limits the number of replicas ($L = 10$) that should be created per each message. This allows this proposal to have the best cost behavior among the studied solutions, displaying an average of 10.14 replicas across the TTL configurations. However, for the 1- and 2-day TTL cases, SCORP manages to have a lower cost (8.6 and 8.3 less replicas, respectively) when compared to *Spray and Wait*, showing the potential of *SCORP* for applications generating messages with a timely limited utility (i.e., low TTL).

Fig. 6.5(c) shows that, in terms of average latency, *Bubble Rap* takes more time (up to 58.1%, 52.6% and 58.8%) to deliver messages than *Spray and Wait*, *dLife* and *SCORP*, respectively. Due to the fact that communities are outdated and few nodes have high centrality, messages reach nodes which are weakly connected to destinations

Both *dLife* and *SCORP* experience lower latencies, since messages reach only nodes that are socially well connected to the destination or that share specific interests. This increases the probability of these proposals in delivering messages in less time.

Despite of having a small advantage over *Spray and Wait* and *dLife* (up to 6.4% and 17.6% less latency, respectively), *SCORP* ends up taking more time to deliver some messages (1- and 3-week TTL) since messages can stay longer in the network. This means that *SCORP* shall take longer to choose the best next forwarders.

It can be observed that message TTL has little impact in the social-oblivious *Spray and Wait*. However, such impact varies over the social-aware proposals. Additionally, considering content information (i.e., users' interests) is advantageous. *SCORP* has delivery performance up to 97.2% with very little associated cost and low latency.

With this evaluation, the 1-day message TTL value was considered for the next set of experiments, as it allows the proposals to deliver a fair amount of the created messages with less associated cost and latency.

6.5.3 Evaluation of Network Load Impact

This section presents the impact of network load (cf. Fig. 6.6[4]). The human trace-based scenario is used with varying network load to observe the behavior of the proposals considering the exchange of data independently of the existing levels of disruption/intermittency. As a general remark, the average number of contacts per hour in this scenario is of 32; and *Bubble Rap* forms an average of approx. 6.7 communities, where most of them comprise almost all nodes. Finally, all results in this section are presented with an increasing number of messages/interests (msg/int) per node.

Fig. 6.6(a) presents the average delivery probability. For the 1 msg/int configuration, *Bubble Rap* has better performance than *Spray and Wait* and *dLife*/*SCORP* (delivering 4.9% and 24.8%, respectively). As most of the communities comprise almost all nodes and replication is done within those communities, this is advantageous to *Bubble Rap* that creates more replicas, increasing its probability of delivering content.

Spray and Wait presents a 20% advantage over *dLife* and *SCORP*. However, such advantage is degraded when considering the results described in Sec. 6.5.2. Different from the synthetic mobility model scenario, the human trace scenario has nodes following routines, and there are no nodes covering the whole extension of the simulated area. This results in replicas reaching nodes that never encounter the desired destinations.

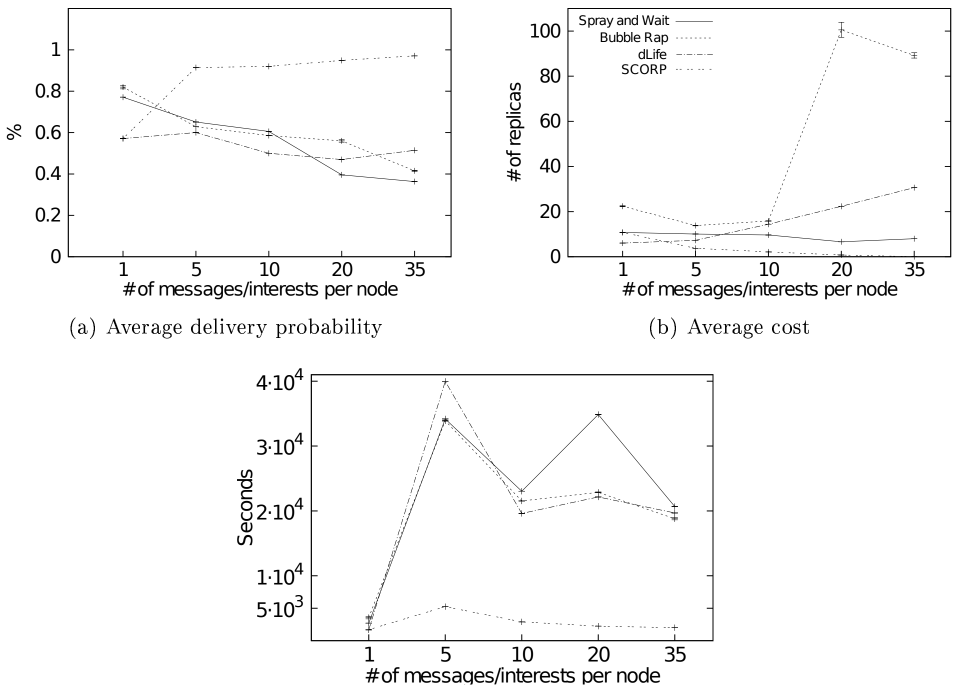
Both *dLife* and *SCORP* display similar performance, since what guides their forwarding is the social weight or node importance (*dLife*) or social weight to interests (*SCORP*). As contacts are very little (average 32 contact/hour) and happen in a sporadic manner, *dLife* and *SCORP* replicate less and thus deliver less content.

For 5 and 10 msg/int configurations, the reduced TTL and sporadic contacts directly affect the delivery performance of both *Spray and Wait* and *Bubble Rap*. As messages may be created during periods of no contacts, TTL expires prior to the proposals delivering messages to their destination.

For the 20 and 35 msg/int configurations, the performance of *Bubble Rap* is further degraded by buffer exhaustion. To support this claim, buffer occupancy is estimated for the 20 msg/int configuration: *Bubble Rap* performs an average of 39240 forwardings during the simulation; by dividing this by the number of days (roughly 12) and by the number of nodes (35, source not included), there is an average of 3270 replicas created per node. By multiplying this by the average message size (52275 bytes), the result is a buffer occupancy of 4.88 MB per node, exceeding the 2MB allowed. It is important to mention that this is just an estimation for the worst case scenario with *Bubble Rap* spreading copies to all nodes. As this does not happen, since *Bubble Rap* employs centrality to control replication, buffer exhaustion worsens as replication occurs to few nodes and not all as in the estimation. Since message generation rate

[4]Springer and the original publisher (Ad Hoc Networks, v. 129, 2014, p. 100-115, Social-Aware Opportunistic Routing Protocol Based on User's Interactions and Interests, Waldir Moreira, Paulo Mendes, Susana Sargento, Figure 3, Copyright © 2014, Institute for Computer Sciences, Social Informatics and Telecommunications Engineering) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media [19].

increases with load, messages can easily take over forwarding opportunities of other messages, reducing the delivery probability of the newly created messages.



(a) Average delivery probability

(b) Average cost

(c) Average latency

Figure 6.6: Evaluation over trace-based scenario with a 1-day TTL

dLife presents a more stable behavior than *Bubble Rap*, as it considers social strength or node importance to reach destinations. Yet, due to its design choices, *dLife* is affected by the rate of contacts. It also experiences buffer exhaustion (approximately 24% more buffer occupancy than the allowed) for the 35 msg/int configuration.

With *SCORP*, one can observe how content awareness can improve opportunistic routing. The delivery probability of *SCORP* increases, since the ability to deliver content of nodes increases (i.e., the more interests a nodes has, the better forwarder it becomes).

Fig. 6.6(b) presents the average cost. For the 1 msg/int configuration, *Bubble Rap* presents the highest cost, since its forwardings are based in the formed communities. *Bubble Rap* has an average of 671.4 forwardings against the 317, 141 and 236 forwardings by *Spray and Wait*, *dLife* and *SCORP*, respectively, to perform a successful delivery.

As expected, *Spray and Wait* presents a stable cost, since it limits the number of create replicas. By considering well socially-connected nodes or nodes interested in the content of the messages traversing the network, *dLife* and *SCORP* tend to replicate less.

Due to a particularity in its implementation, *SCORP* replicates more than *dLife*: nodes with a certain interest, not only process that message, but also keep a copy for further replication as they may have a chance to find nodes with this same interest, or that met other nodes with such interest. In this latter case, a node receiving a message with content matching its interest, also replicates it (unnecessary and unwanted replicas) to nodes that often have encountered it (and have a greater social weight to that specific interest).

For the 5, 10, 20 and 35 msg/int configurations, the number of forwardings increases with the load. Still, this is not enough to increase the delivery performance of *Bubble Rap* and *dLife*, and just contributes to their associated cost in delivering content.

With more interests, a *SCORP* node becomes a potential forwarder to other nodes. Hence, the unwanted replicas observed in the 1 msg/int configuration have a positive effect in what concerns spreading content. Furthermore, by considering only interested nodes or nodes that interact with others interested in the carried message, *SCORP* reduces cost: it produces an average of approximately 3.5 replicas across the msg/int configurations against an average of 9, approximately 48.4 and 16.1 replicas of *Spray and Wait*, *Bubble Rap* and *dLife*, respectively. Finally, *SCORP* uses less resources (i.e., buffer): content awareness leads to a buffer occupancy varying between approximately 0.03 MB (1 msg/int) and 0.15 MB (35 msg/int).

The average latency is presented in Fig. 6.6(c). For the 1 msg/int configuration, *Bubble Rap* messages take 24% and 52% longer to be delivered when compared to *Spray and Wait* and *dLife*/*SCORP*, respectively. As most of the formed communities comprise almost all nodes, messages may be exchanged between nodes that take longer to reach destinations (despite the fact that node do share the same community).

In the case of *dLife* and *SCORP*, their lower latency is explained by the fact that most of their deliveries (90%) are performed directly within the first two hours of simulation between source and destinations nodes. The same behavior is observed for *Spray and Wait*, which delivers 85% of its messages directly up to the second hour of simulation; however, 17% of such deliveries are performed directly, reflecting the power behind random replications.

For the 5, 10, 20 and 35 msg/int configurations, the peak at the 5 msg/int configuration is due to the time messages are created. By looking at the contact distribution, one can observe that messages are created in periods of little to no contacts with followed by long periods (between 12 and 23 hours) of almost no contact. This contributes to the observed increase in latency as messages are stored longer. As load increases, messages are generated before or during periods of high number of contacts, which reduces the experienced latency. Since latency is determined based on the delivered messages, this explains the decrease and variable behavior for the 10, 20 and 35 msg/int configurations experienced by *Spray and Wait*, *Bubble Rap* and *dLife*: the increase and decrease of delivery rate are influenced by chosen forwarders that take longer to encounter the destination and deliver content.

SCORP delivers its messages up to 93.61%, 90.25% and 89.94% less time than *Spray and Wait*, *Bubble Rap* and *dLife*, respectively. A *SCORP* node can receive more information since it is interested in the content being replicated, and becomes a better forwarder as the chance of meeting nodes sharing the same interests is high. It is observed that almost all communities comprise almost all nodes. Despite the idea of community formation is of no importance to *SCORP*, this observation suggests that nodes have a high number of contacts, which is of great advantage for *SCORP,* since it can find destinations (i.e., interested nodes) faster. To support this claim, by looking at the delivered messages, one can observe that shared interests account for 46%, 53%, 59% and 66% of deliveries in the 5, 10, 20, and 35 msg/int configurations, respectively. Other destinations are reached through the ability of *SCORP* in identifying interested parties, which further improves its performance.

6.5.4 Summary

This section presented the benefits of combining social awareness and content information (i.e., information type, interested parties) to improve data dissemination in urban, dense scenario.

The *SCORP* approach is based on users' daily interactions and interests, and shows that data dissemination can be further improved in challenged networks when routing is devised considering content knowledge and social proximity. *SCORP* presents better performance than social-aware and content-oblivious opportunistic routing proposals, such as *Bubble Rap* and *dLife*. The delivery capability of *SCORP* reaches up to 97% with an average of 46.9 minutes against the 335.5 and 343.7 minutes required by *Bubble Rap* and *dLife*, respectively, to perform a delivery. Moreover, *SCORP* replicates up to approximately 13.9 and 4.7 times less than *Bubble Rap* and *dLife*, respectively.

6.6 Summary of the Chapter

This chapter presented the evaluation carried out to assess the performance of the proposed social-aware utility functions as well as the social- and content-based opportunistic routing approaches. The evaluation took into account different opportunistic routing benchmarks for comparison purposes over scenarios comprising synthetic mobility models and human traces. Every experiment was set to expose the studied proposals to the same conditions (e.g., generated load, pair of communicating nodes, transmission rate and range, number of nodes, among others).

The chapter started with the evaluation of the social-aware utility functions, highlighting the performance of each of them and discussing on their potential when working separate and when combined. Then, the combination of these utility functions resulted in the new social-aware opportunistic routing *dLife*, which relies on the dynamics of user behavior to take forwarding decisions. As presented, it is clear how opportunistic routing can profit when focusing on the dynamism of social interactions.

Finally, the chapter presented the improvements for opportunistic routing when combining social awareness and content information (i.e., information type, interested parties). This combination led to the Social-aware Content-based Opportunistic Routing Protocol (*SCORP*), which has outperformed social-aware and content-oblivious opportunistic routing proposals. Given the fact that *SCORP* considers not only the social interactions of users, but also their interest in the content traversing the network, it is able to have almost optimum delivery with less time and associated cost.

Chapter 7

Structure Analysis of Social-based Networks

The analysis of the structure of opportunistic networks has shown that this type of networks may exhibit different characteristics (e.g., small world, scale free, power-law). Interestingly, such networks may be even modeled to display a specific characteristic (e.g., scale free [77]) based on the application scenario to which they are subject.

Independently of the displayed feature, the satisfactory functioning of opportunistic routing solutions is directly related to the fact that such solutions can cope with the structural properties coming from these characteristics [78], which in turn result from the underlying user mobility patterns [51].

Moreover, given its observed dynamic behavior, opportunistic networking displays the time-evolving (or time-varying) behavior [79], in which users' behavior and the links between them vary in different instants in time.

This chapter does not aim at showing whether the structure of opportunistic connectivity graphs have the time-evolving/varying behavior, as there are already different works that address such topic [80, 81, 82, 83, 79]. Instead, this chapter analyzes the structure of the network of social relationships formed based on social awareness, and shows how such structure can be characterized and whether the featured characteristics vary in different time periods [36].

It is important to note that the structure analysis throughout this chapter was carried out on the Gephi v0.8.2 [84] and Cytoscape v2.8.3 [85] analysis tools. Finally, the different/increasing sizes and shades of gray (up to black) in the figures indicate the degree of nodes and social strength of edges.

7.1 Introduction to Network Structure Characterization

Opportunistic networks have shown (or even have been modeled to display) features compatible with small world [78], scale free [77], and contact/inter-contact time power law distribution [86] properties. This section briefly presents each of these properties with few examples that may apply to the structure of opportunistic networks.

7.1.1 Small-World Networks

In small-world networks, most of the nodes may be reached through a small series of hops. That is to say that, the average path length between nodes is shorter than the network diameter. Moreover, besides the presence of short paths between nodes, small-world networks also present a high clustering coefficient feature [87, 88].

As an example of small-world features, Fig. 7.1 illustrates a network of airline flights[1].

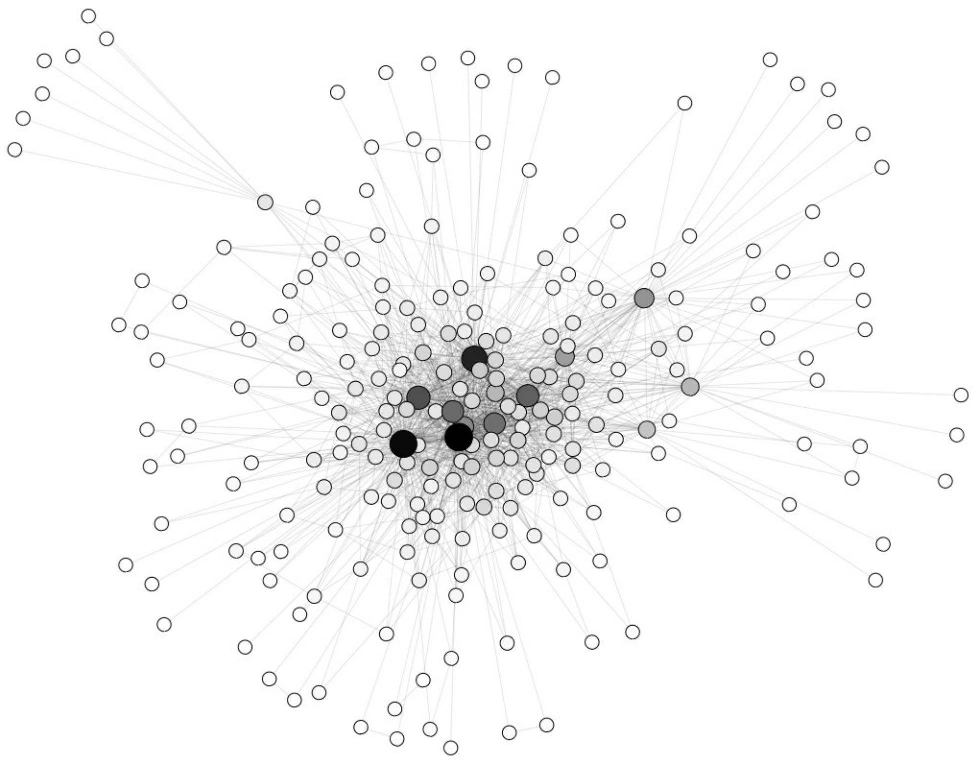


Figure 7.1: Airline flight network

This network has 235 nodes (i.e., vertices) with 1297 edges between them. It presents an average path length of 2.318, a network diameter of 4, and an average clustering coefficient (not considering nodes with degree ≤ 1) of 0.652, which indicate small-world features comparative to the ones reported by Hossmann et al. (2011) [78].

The degree distribution could be another feature used to identify small-world networks. However, such feature is unable to reflect real-world behavior, since the small-world model was not meant to capture real-world degree distributions [87, 88].

It has been shown that the structure of the network of contacts of different human traces (e.g., Wi-Fi associations at Dartmouth and ETH Zurich campuses, MIT Bluetooth contacts, Gowalla check-ins) formed during the experiments does have such features [78]. These findings are of great interest for this Thesis, as the goal of this chapter is to check whether such features remain in the structure of the network of social relationships formed in different periods of time.

[1] https://gephi.org/datasets/airlines.graphml.zip

7.1.2 Scale-Free Networks

A scale-free feature is related to the network degree distribution following a power law [87, 88]. This means that scale-free networks have a high number of nodes (i.e., hubs) with degree higher than the average degree found in the network. Additionally, a large part of the nodes is connected, comprising the giant component (i.e., a considerable number of connected vertices) [89].

In Fig. 7.2 an example of scale-free network is given: the Protein-Protein interaction network in yeast[2]. It is important to mention that this network was treated as undirected graph in both tools.

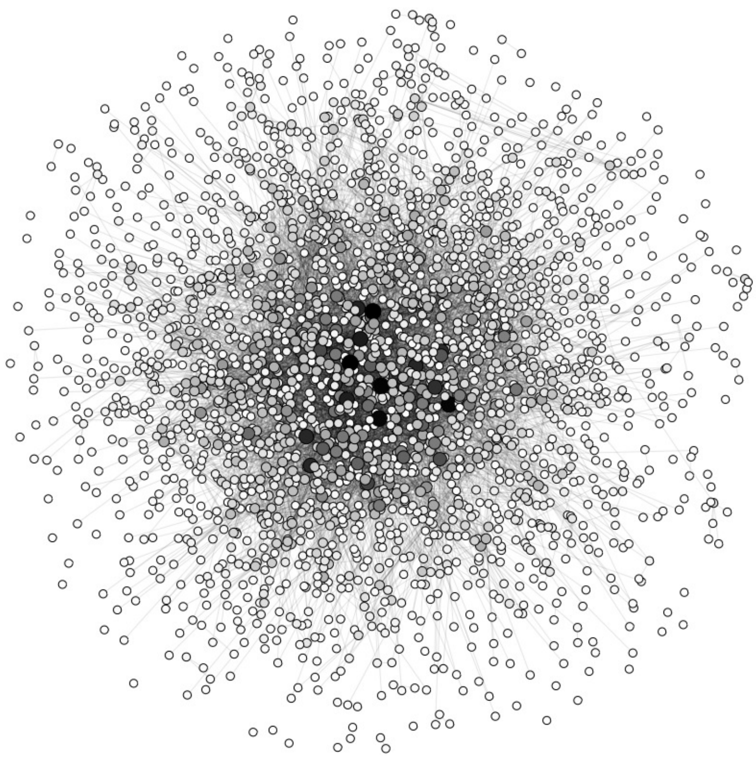


Figure 7.2: Protein interaction in yeast

This network has 2361 nodes with 7182 edges between them from which 536 edges are self-loops. Fig. 7.3 shows the degree distribution of the network, which indeed approximates a power law ($y = ax^b$, with $a = 2463.4$ and $b = -1.861$). Regarding the average degree, it is of 5.63 and 694 (29.39%) nodes have degree higher than such average. Finally, 2224 (94.2%) nodes are in the giant component. These features are compatible with scale-free networks.

[2] https://gephi.org/datasets/yeast.gexf.zip

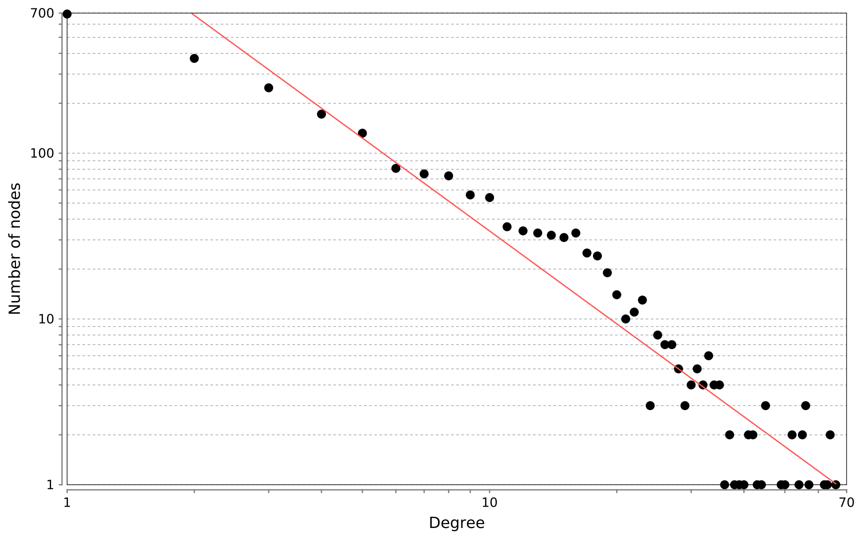


Figure 7.3: Scale-free degree distribution

As observed in Sec. 7.1.1, the structure of the network of contacts displays small-world features [78]. However, it is not known whether these features remain in different periods of time, as opportunistic networks are very dynamic.

Besides that, these networks have shown satisfactory functioning when modeled with scale-free features [77]. Thus, this section introduces these features that are used later in the chapter to check if they can be seen when looking at the structure of the network of social relationships in different time periods.

7.1.3 Power Law

Opportunistic networking takes advantages of the contacts taking place between users for data exchange. That is, this networking paradigm depends on the underlying human mobility [90].

This means that the contact time (i.e., duration) and inter-contact time (i.e., time elapsed between such contacts) are of great importance to decide whether data should be exchanged to reach a specific destination.

It has been shown that, not only can a power law characterize the degree distribution of scale-free networks, but it can also characterize the distribution of contact and inter-contact times in opportunistic networking environment [91]. More interestingly, this power-law property has been found in different human mobility traces with experiments spanning indoor to outdoor and campus to city-wide areas [86].

By knowing how contact/inter-contact times are characterized, opportunistic routing solutions can be tuned to make the most out of whichever length the contacts have, and the delay involved in the data exchange, for instance. That is, to have proposals that can cope with the dynamism found in the user behavior [1].

7.2 Timely Fashion Analysis

From the different properties seen in Sec. 7.1 that the network structure of social relationships resulting from opportunistic contacts may have, this chapter aims at understanding which properties are found in the social-aware approach devised in this Thesis work.

This section starts first with the structure characterization considering the social and contact networks (resulting from user interaction) formed throughout the experiment (i.e., whole-time network structure). Then, given the time-varying feature of opportunistic contact graphs [80, 81, 82, 83, 79], the section presents what properties the network structure has in different time periods (each of one hour), and checks whether these properties remain or change as time evolves.

To achieve such goal, two CRAWDAD human traces are considered: i) Cambridge [76], which comprises a group of 36 students carrying iMote devices while in their daily routines during a two-month period in the city of Cambridge, UK; and, ii) MIT [92], which comprises 97 Nokia 6600 smart phones distributed among the students and staff of this institution. This dataset is worth of approximately 40 years of information.

As the main interest here is on the structure of the resulting social network, I only consider the *TECD* utility function, which leads to a connectivity graph based on the social weights among users.

Since the social weight universe can comprise very low weights (resulting from short contacts), I define a threshold to eliminate few edges, which stand below the lower bound considering a 95% confidence interval of such social weights. This is analogous to the approach used by Hossmann et al. (2011) [78] to improve the interpretability of weighted graphs; however, their analysis is done over the entire experiment, while I also want to understand the social network structure in the different time periods (i.e., time period-based structure).

7.2.1 Whole-time Network Structure

This section analyzes the properties of the structures of the social and contact networks formed from the Cambridge and MIT traces, considering the full network created throughout the experiment. Thus, Table 7.1 presents the identified small-world and scale-free properties of these formed graphs.

Table 7.1: Small-world and scale-free properties

		Social-based		**Contact-based**	
		Cambridge	**MIT**	**Cambridge**	**MIT**
	# of Edges	143.16	387.10	483	2280
Small World	**Avg Clustering Coefficient**	0.69	0.66	0.84	0.74
	Avg Path Length	2.27	2.64	1.23	1.51
	Network Diameter	5.2	6.2	2	3
Scale Free	**Avg Degree**	8	7.98	26.8	47
	# of Nodes with Degree Higher than Avg Degree	21.22	41.57	22	52
	# of Nodes in the Giant Component	33.70	72.33	36	96

For the sake of simplicity, the table displays two graph types based on: i) social aspects considering the *TECD* utility function (i.e., *social-based*), and ii) node contacts (i.e., *contact-based*). That is, the former is a graph where edges are socially weighted with such weights changing as nodes interact; while the latter is an incremental graph with new edges emerging as new contacts take place among nodes.

As the goal is to characterize the structure of the network (be it social- or contact-based), checking whether the contact/inter-contact times follow a power law is out of the scope of this work. Finally, as *TECD* works based on daily samples, the *social-based* columns present the average of each property over the entire simulation.

As presented in Sec. 7.1, to display small-world properties, networks shall have an average path length smaller than the network diameter and a high clustering coefficient [87, 88]. From Table 7.1, one can observe that the structure of social- and contact-based networks is compatible with small-world networks for both datasets.

When social awareness is considered, the connectivity graph has a much lower number of edges than the contact-based version. This is expected since, as nodes interact, social-based approaches select edges that best reflect the existing social interaction among nodes to decide on forwarding (e.g., to reach a brother, it is better to consider another family member than choosing a classmate).

Additionally, with social awareness nodes i) are not as clustered as in contact-based networks ($0.69 < 0.84$ for Cambridge, and $0.66 < 0.74$ for MIT); and ii) need more hops to be reached ($2.27 > 1.23$ for Cambridge, and $2.64 > 1.51$ for MIT). However, social-based networks consider only very good social links (i.e., edges), which has a positive effect for routing (e.g., less resource consumption and latency, higher delivery probability as discussed in Chapter 6).

Despite the potential of weak ties [51], contact-based network routing may end up consuming much more resources and having negative effect in delivery and experienced latency.

One may question this observation, as contact-based networks have more edges and small average path length and diameter, which could lead to better delivery and low latency; however, the network does not start with such number of edges, and the chosen next forwarders may not be the best options, since the edge in the connectivity graph can represent a contact that happened a long time ago, and does not reflect reality given the fact the graph is incrementally formed.

Interestingly, the formed social- and contact-based networks have few properties compatible with scale-free networks. From a total of 36 nodes in the Cambridge traces, both networks present a considerable number of nodes with degree higher than the average (21.22 and 22) and that are included in the giant component (33.7 and 36). However, when observing the degree distribution depicted in Figs. 7.4 and 7.5, one can conclude that the degree distributions of the formed social- and contact-based networks do not follow a power-law distribution.

In the case of the social-based network in Fig. 7.4, the last daily sample of the last day is used to illustrate the degree distribution. The line represents a power law ($y = ax^b$) with $a = 1.291$ and $b = 0.264$. Fig. 7.5 shows the degree distribution of the contact-based network, where the line represents a power law ($y = ax^b$) with $a = 0.056$ and $b = 1.123$.

Regarding the MIT traces, both social- and contact-based networks display scale-free properties. Out of the 97 nodes, a good part of these nodes have degree higher than the average (41.57 and 52 nodes). Also, great part of the nodes are in the giant component (72.33 and 96 nodes). Yet, by looking at the degree distribution (cf. Figs. 7.6 and 7.7), one can see that it does not fit a power law.

For the social-based network in Fig. 7.6, the last sample of the last day was also considered to illustrate the degree distribution, in which the line representing the power law ($y = ax^b$) has $a = 7.5$ and $b = -0.388$. As for the contact-based formed from the MIT traces, the line in Fig. 7.7 represents a power law ($y = ax^b$) with $a = 0.932$ and $b = 0.145$.

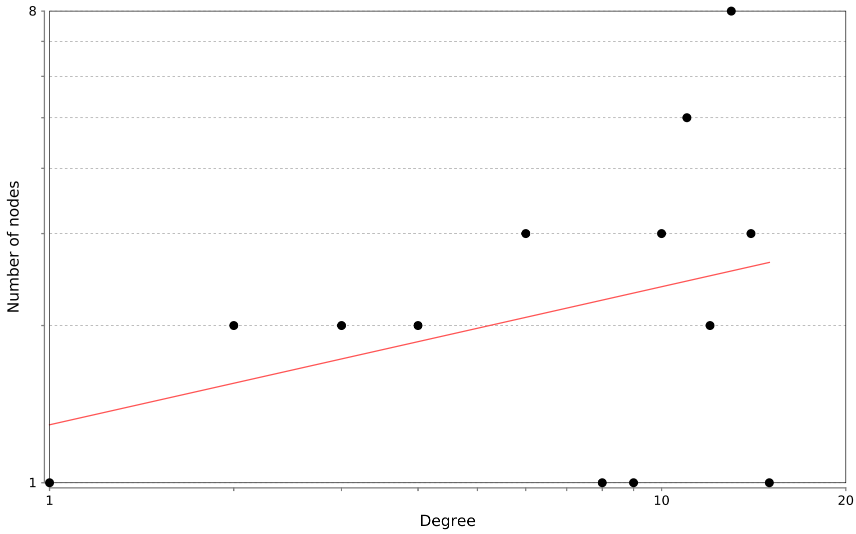


Figure 7.4: Degree distribution of whole-time social-based network (Cambridge traces)

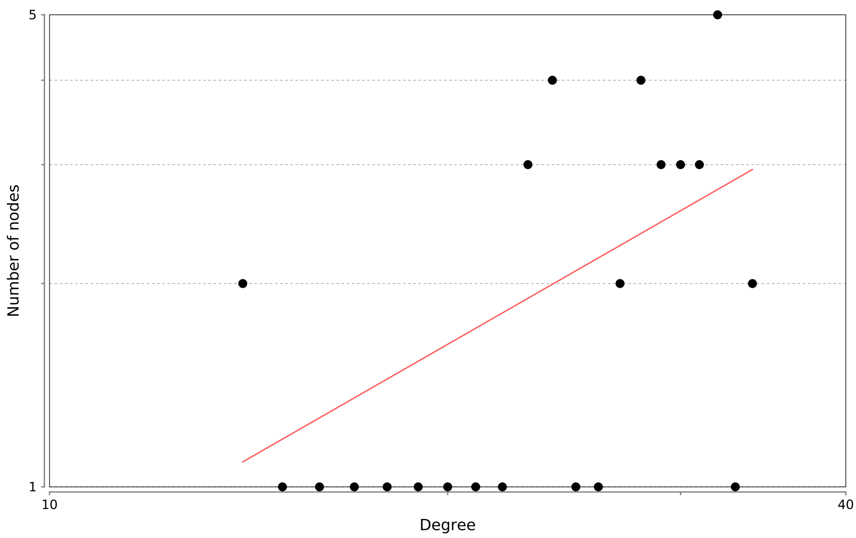


Figure 7.5: Degree distribution of whole-time contact-based network (Cambridge traces)

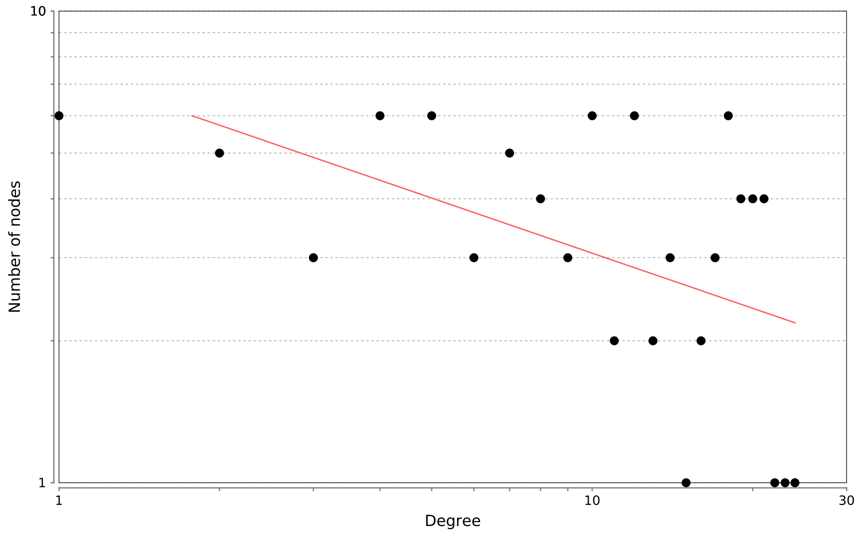


Figure 7.6: Degree distribution of whole-time social-based network (MIT traces)

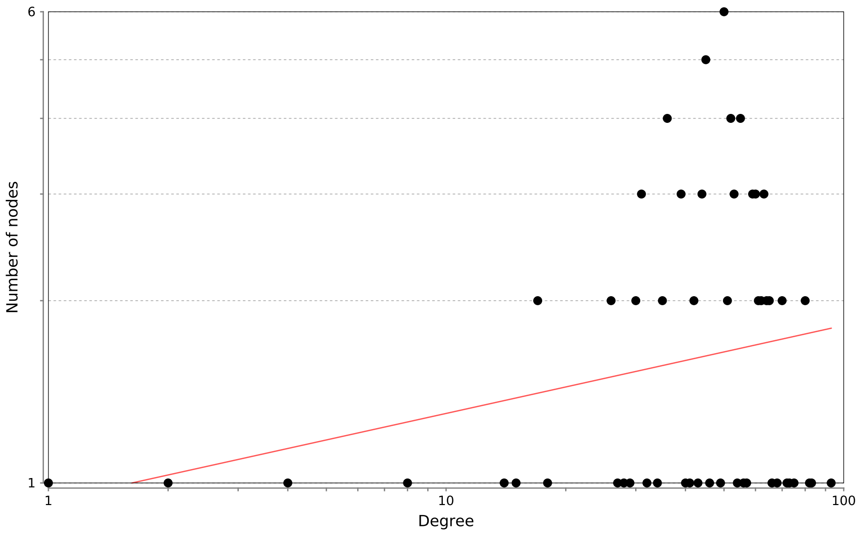


Figure 7.7: Degree distribution of whole-time contact-based network (MIT traces)

In summary, regarding the analysis carried out over the entire network behavior, one can conclude that, independently of being based on social awareness or solely on contacts, the connectivity graph displays features of small-world networks. This is also observed by Hossmann et al. (2011) [78] in their analysis, where different human traces have small-world features. In what concerns opportunistic routing, this is interesting as short path lengths and high clustering coefficients help social-based approaches in identifying features such as communities, levels of social interaction, node centrality, which can improve routing performance. As for contact-based routing, this is irrelevant, as these proposals are more interested in getting information disseminated, and surely have their performance degraded for not considering the social nature of contacts.

Moreover, both social- and contact-based networks display few scale-free properties, but not enough to be classified as such. Still, the identified properties may indicate that destination nodes in the formed networks can be quickly found, since a considerable number of nodes is well connected.

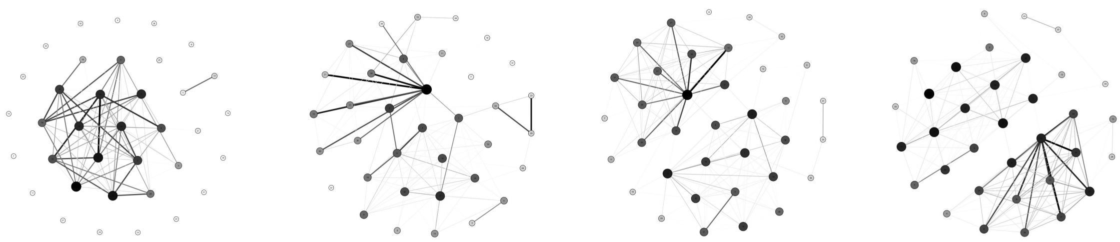
7.2.2 Time Period-based Structure

This section analyzes the properties of the structure of social and contact networks considering the network structure in different time periods. The goal in this section is to observe whether the network structure changes in different periods of time.

As *TECD* considers daily samples of one hour each, information about the structure of the social- and contact-based networks was collected in different days and samples considering the CRAWDAD Cambridge and MIT traces. It is important to note that the Cambridge dataset is worth of two months of data. However, when simulated it is worth almost 12 day of communications. Regarding the MIT dataset, given the amount of data, the equivalent to 196.5 days (i.e., 6.5 months) was considered.

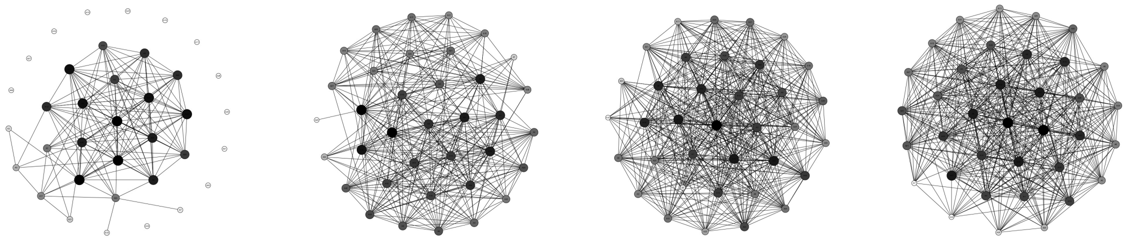
Figs. 7.8 and 7.9 show how the structure of the social- and contact-based networks formed from the Cambridge traces changes from the first daily sample of the first day to the last sample of the last day. It is worth mentioning that the Fruchterman Reingold layout, available in Gephi [84], was used solely to improve the presentation of the formed networks.

With social awareness, only the links (i.e., edges) representing socially well-connected nodes are present in specific time periods. This allows for a better view of the dynamicity of user behavior, since it reflects the different levels of social interaction throughout the users' daily routines. Consequently, opportunistic routing proposals based on social awareness can take better forwarding decisions, since nodes shall only forward content to others which are socially well connected to destinations [15].



(a) Day 1-Sample 0 (b) Day 4-Sample 8 (c) Day 8-Sample 3 (d) Day 12-Sample 23

Figure 7.8: Social-based network (time period-based structure)



(a) Day 1-Sample 0 (b) Day 4-Sample 8 (c) Day 8-Sample 3 (d) Day 12-Sample 23

Figure 7.9: Contact-based network (time period-based structure)

As one could expect, the contact-based network evolves in the sense that, as nodes interact, new edges emerge. While with social awareness nodes can be distinguished in terms of their social interactions (cf. node degree and social strength of edges in Fig. 7.8(d)), this does not apply to the structure of contact-based networks. Fig. 7.9(d) shows a homogeneity trend in terms of node degree for the contact-based network. This feature of contact-based networks may be costly in terms of resource utilization in opportunistic networks, since routing based on such connectivity graphs might become a mere flooding solution (i.e., nodes cannot be distinguished, thus replication may occur unwisely) [33].

Table 7.2 presents the small-world and scale-free properties for different time periods considered in Figs. 7.8 and 7.9 for the Cambridge traces. In order to classify how the structure of social- and contact-based networks changes in different time periods, each of these periods are considered; however, for the sake of simplicity, only four specific periods of different days are displayed to facilitate reading.

The small-world properties remain for the structure of the social-based network. When compared to the whole-time structure analysis (cf. Table 7.1), one can observe that the structure changes in terms of node degree and social weights of edges, but the characteristics are still compliant with small-world networks. Moreover, the number of edges stabilize, showing how *TECD* can indeed capture the dynamic behavior of social interactions. The contact-based network also remains with small-world features as identified in the whole-time structure analysis.

Table 7.2: Small-world and scale-free properties found in the Cambridge traces

			Small World			**Scale Free**		
Graph Type	**Time Period**	**# of Edges**	**Avg Clustering Coefficient**	**Avg Path Length**	**Network Diameter**	**Avg Degree**	**# of Nodes with Degree Higher than Avg Degree**	**# of Nodes in the Giant Component**
Social-based	Day 1 - Sample 0	78	0.60	1.49	3	4.33	15	17
	Day 4 - Sample 8	125	0.67	2.23	4	6.94	21	34
	Day 8 - Sample 3	162	0.75	2.19	5	9	23	35
	Day 12 - Sample 23	162	0.77	2.41	6	9	22	34
Contact-based	Day 1 - Sample 0	139	0.72	1.61	3	7.72	19	24
	Day 4 - Sample 8	347	0.77	1.46	3	19.28	19	36
	Day 8 - Sample 3	462	0.82	1.27	2	25.67	18	36
	Day 12 - Sample 23	483	0.84	1.23	2	26.83	22	36

When comparing the social-based network to the contact-based one, social-aware nodes are less clustered and require more hops to be reached in the former for the different time periods. However, the social-based network has the advantage of displaying only socially well-connected nodes with a lower, and yet useful number of edges. In terms of routing, this is rather interesting since social-based approaches only select forwarders that can increase delivery probability.

One could argue that such features could also be advantageous to contact-based routing approaches (i.e., more clusters and smaller path length and diameter). However, contact-based solutions fail in capturing the changes in the connectivity network as they are incrementally formed (cf. Sec. 7.2.1). Despite of the potential in quickly reaching destination nodes due to the small-world features, contact-based routing may experience longer delays, affecting its delivery probability. Additionally, given the identified homogeneity trend in terms of node degree, cost (i.e., number of replications to achieve a successful delivery) may easily become an issue.

Regarding the scale-free features, both social- and contact-based networks still display a considerable number of nodes with degree higher than the average degree, as well as most (if not all) nodes in the giant component.

Figs. 7.10 and 7.11 present the degree distribution of social- and contact-based networks for day 4 - sample 8 for the Cambridge traces. Still, the degree distribution of both networks does not follow a power law, so they cannot be characterized as scale-free networks. The lines representing the power law ($y = ax^b$) have $a = 3.211$ and $b = -0.235$ for the social-based network, and $a = 0.930$ and $b = 0.234$ for the contact-based network.

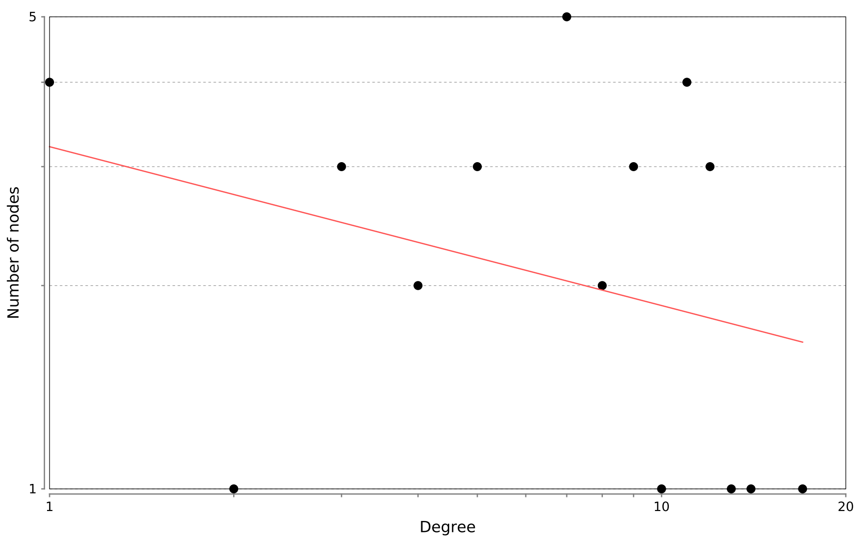


Figure 7.10: Degree distribution of time-period social-based network (Cambridge traces)

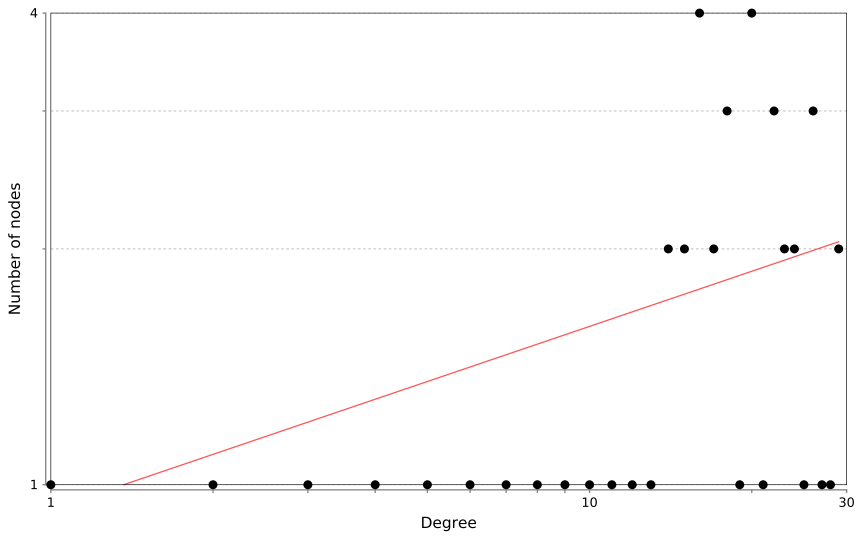


Figure 7.11: Degree distribution of time-period contact-based network (Cambridge traces)

Now the properties of the social- and contact-based networks formed from the MIT traces are presented. Given the number of links existing between the nodes, this section does not show how the structure of both networks changes. However, by looking at the Table 7.3, one can still observe the trend found with the Cambridge traces: small-world structures with a few scale-free characteristics.

Comparing to the structures formed based on the Cambridge traces, one can observe from the MIT traces that, in almost all samples, the number of nodes with degree higher than the average has decreased (below 50% of the total number of nodes) for the social-based structure. As MIT experiments involve more individuals that belong to various groups and interact in different areas, the number of hubs must represent those who really can connect the different existing groups.

Table 7.3: Small-world and scale-free properties found in the MIT traces

			Small World			Scale Free		
Graph Type	**Time Period**	**# of Edges**	**Avg Clustering Coefficient**	**Avg Path Length**	**Network Diameter**	**Avg Degree**	**# of Nodes with Degree Higher than Avg Degree**	**# of Nodes in the Giant Component**
Social-based	Day 1 - Sample 23	43	0.67	2.60	7	0.89	20	18
	Day 60 - Sample 0	426	0.66	2.48	6	8.78	47	76
	Day 130 - Sample 0	450	0.68	2.45	5	9.28	41	75
	Day 197- Sample 13	464	0.67	2.84	6	9.57	46	87
Contact-based	Day 1 - Sample 23	113	0.69	1.94	4	2.33	23	29
	Day 60 - Sample 0	1817	0.69	1.61	3	37.46	48	95
	Day 130 - Sample 0	2126	0.73	1.54	3	43.84	51	96
	Day 197 - Sample 13	2280	0.74	1.51	3	47.01	52	96

Besides having a reasonable number of nodes as hubs and being part of the giant component, as it can be seen in Figs. 7.12 and 7.13, the degree distribution (for day 60 - sample 0) of both social- and contact based structures still does not follow a power law. The lines representing the power law ($y = ax^b$) have $a = 7.956$ and $b = -0.413$ for the social-based network, and $a = 1.650$ and $b = -0.013$ for the contact-based network.

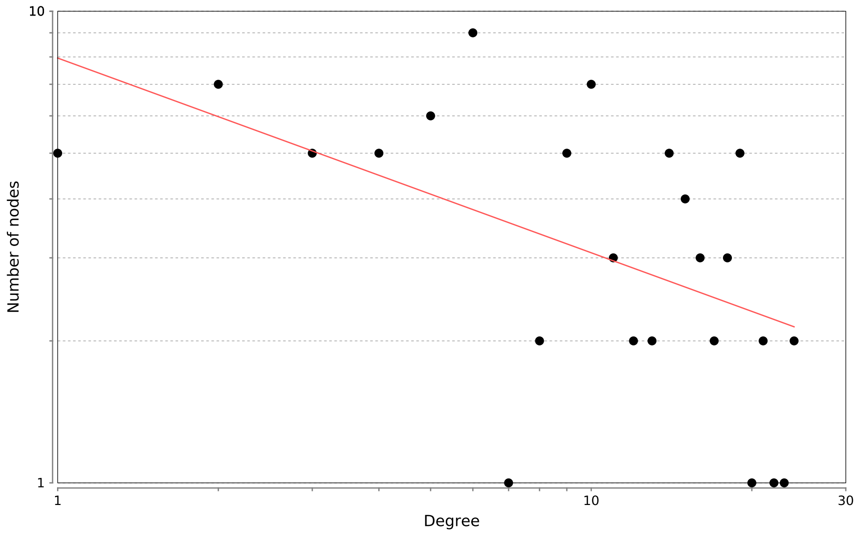


Figure 7.12: Degree distribution of time-period social-based network (MIT traces)

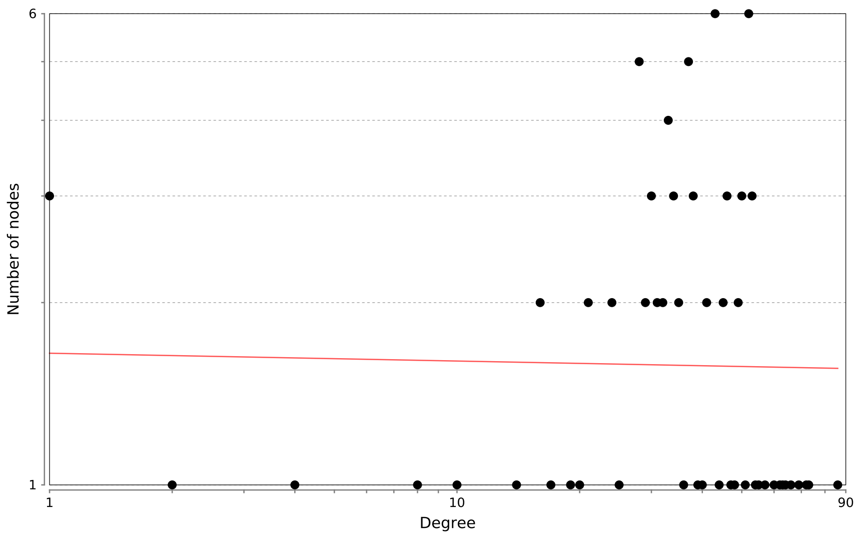


Figure 7.13: Degree distribution of time-period contact-based network (MIT traces)

Since user behavior is rather dynamic, in the sense that social interaction changes in different time periods, with this study I achieved the goal of observing whether the features of structure of social-based networks remained unchanged in such periods of time.

7.3 Summary of the Chapter

This chapter presented that the structure of social-aware opportunistic networks indeed changes in terms of node degree and social strength of edges, but its structure still presents small-world properties (average path length smaller than the network diameter and high clustering coefficient [87, 88]) for the different time periods. This finding is interesting as it highlights the importance of suitably identifying socially well-connected nodes for the better performance of social aware opportunistic routing.

Moreover, the social networks studied also displayed features compliant with scale-free networks (high number of nodes with degree higher than the average and comprising the giant component [87, 88]). However, the identified degree distributions did not follow a power law. Scale-free properties are interesting in the context of opportunistic networking as they present low network diameter and high fault resilience [77]. In the context of social-aware opportunistic routing, such features can aid in delivery, given how fast destinations can be reached and the increased number of alternate paths to reach such destinations.

Social networks have been reported to have small-world features in their structure [78]. However, previous analysis has been done over the whole network behavior, whereas the analysis done in this chapter looked at the structure in different time periods. The findings provide clues that social-aware opportunistic routing cannot be oblivious to the existing dynamicity of user behavior, that should be captured in a timely fashion.

Chapter 8

Conclusions and Future Directions

This chapter summarizes the work done in the context of this Thesis, and presents ongoing efforts and future research directions.

8.1 Conclusions

The advancements in the manufacturing of portable devices have introduced a new way of communication: users nowadays carry very powerful devices (in terms of processing, storage, and access technologies), and wish to exploit their devices at maximum. This means consuming and producing information while they are on the move.

Opportunistic networking has shown great potential in such scenario, by exploring the increased capabilities of users' devices to allow the exchange of information upon contact opportunities between these devices.

Since these devices are carried by people who naturally have different levels of interactions with others, a new form of opportunistic routing emerged, which considers the existing social similarity among those carrying such devices. Same work/educational affiliation, shared interests, communities in common, and node popularity are some examples of social similarity metrics considered in the context of opportunistic routing.

The work in this Thesis is motivated by the routing challenges (i.e., high mobility rate, lack of infrastructure, intermittent connectivity) found in the environment where these users want to communicate. Thus, this Thesis exploits the dynamic social interactions among users within the context of social-aware and content-based opportunistic routing.

For this purpose, a set of specific objectives was defined to help achieving the goal of proposing new opportunistic routing proposals. Among these specific objectives were the identification of the different types of existing opportunistic routing proposals; understanding of the existing opportunistic routing classifications (i.e., taxonomies); the detailed study of the employed opportunistic routing metrics; the understanding of cooperation among users; the development of social-aware utility functions; the design and specification of social-aware and content-based opportunistic routing proposals; and the analysis of the network structure resulting from the devised social-aware approaches.

Each defined specific objective led to a contribution to the opportunistic routing research community. The following are the main contributions achieved with this work.

Classification and evaluation of opportunistic routing approaches: by looking at 12-year period worth of opportunistic routing protocols, their metrics and their existing classifications [25, 26], it was clear that the new trend based on social similarity metrics needed its duly recognition.

Thus, I devised a new taxonomy for opportunistic routing proposals which included the social similarity branch, later being updated with a new subcategory that includes the user dynamic behavior, subject of this Thesis work [1].

With the analysis done to propose the new taxonomy, I soon realized that no guidelines were available, so proposals (existing and newly proposed) could have their performance fairly assessed. Thus, different proposals were analyzed according to the identified trends, information was collected on their evaluation process, and common properties were found (i.e., routing strategy and metrics), which led me to the Universal Evaluation Framework (UEF). This UEF comprises a set of performance parameters and experimental setup to aid designers fairly assessing the performance of opportunistic routing solutions [27, 28].

Encouraging user cooperation: the aim here is showing: i) how the cooperative behavior of users is of great importance in opportunistic routing; and ii) how it can be achieved by simply offering users other resources they might need, or even through a simple and yet safe virtual crediting system [29, 30, 31].

Indeed, the increased capabilities of devices are of no use for routing if their owners are not willing to cooperate by carrying, storing, and/or relaying information on behalf of others. Thus, cooperation is a must when it comes to guaranteeing the wellness of the network and improved networking experience for the users.

In the context of this Thesis, cooperation happens through the sharing of node resources (i.e., buffer) and users are considered to be always willing to cooperate. The work developed to achieve this contribution is inserted in the context of the ULOOP European project.

Social-aware utility functions: prior to designing and specifying new opportunistic routing proposal, I was required to come up with social-aware utility functions capable of coping with the dynamic behavior of users.

This contribution comprises three utility functions: *TECD*, that captures the level of social interaction among nodes; *TECDi*, that measures the importance of node based on its social interactions; and *TECI*, that quantifies the social interactions among nodes sharing interests [32, 33, 35].

The developed utility functions have been properly tested to reach stability. As the developed social-aware utility functions are able to capture the dynamism of user behavior, they have shown satisfactory performance (i.e., delivering more with less associated cost and latency) when compared to the utility functions found in the prior art.

Social-aware and content-based opportunistic routing protocols: once I have reached a satisfactory performance behavior with the utility functions, it was time to design and implement the proposed opportunistic routing protocols. First, this contribution starts with the social-aware opportunistic routing protocol based on user's social daily routines, *dLife* [15], which is based on the *TECD* and *TECDi* utility functions.

The first finding shows that opportunistic routing can indeed benefit by considering the dynamism of users' social daily behavior. Hence, *dLife* could perform wiser forwarding decisions with better delivery probability, cost and latency performance than *Bubble Rap* and *dLifeComm*. Another finding concerns centrality, which has shown a more negative impact than the notion of community formation.

Currently, *dLife* is being specified within the IRTF DTN Research Group [16]. The *dLife* proposal is also inserted in the context of the DTN-Amazon project as the routing solution for SocialDTN, a DTN architecture implementation for Android devices [34].

Since focusing on the content, and not on the host, has shown great potential in opportunistic network routing, another contribution is the design of point-to-multipoint routing solution: the Social-aware Content-based Opportunistic Routing Protocol (*SCORP*) [19]. *SCORP* is a content-oriented opportunistic proposal that allows the exchange of data based on the content and the interest of users in it.

A third finding refers to the fact that, by combining social awareness to content information (i.e., information type, interested parties), opportunistic routing can improve significantly. *SCORP* is able to almost reach optimum delivery rate with very little associated cost and delay when compared to *dLife* and *Bubble Rap*. Results suggest that content-oriented solutions are much more interesting than content-oblivious ones in the context of social-aware opportunistic routing.

Structure analysis of social-based networks: this contribution aims at understanding the network structure formed by the social-aware approaches proposed in this work.

What is observed in the prior art is that normally, such type of analysis is carried out over the whole network behavior. Instead, the interest here is to identify which properties (e.g., scale free, small world) are displayed on the dynamic behavior of users and whether such properties remain in different periods of times.

First findings indicate that both social- and contact-based connectivity networks display small-world and few scale-free properties over the whole network behavior and in different periods of time. Additionally, independently of being or not social-aware, opportunistic routing solution must not be agnostic to the existing user dynamic behavior. This contribution is yet to be submitted to a high-impact journal.

8.2 Ongoing Efforts and Future Research Directions

This Thesis comprises different topics related to social-aware opportunistic routing, such as classification, evaluation, utility functions, protocols, and structure analysis, and each of such topics can be further updated and improved. This is referred to as ongoing efforts or future research directions.

Amongst the ongoing efforts are the constant update of the proposed taxonomy [1], evaluation framework [27, 28] and *dLife* protocol specification [16]. As new opportunistic routing proposals emerge, the taxonomy and framework have to be updated following the new trends. The same is expected with the specification of *dLife*: as the real-world implementation evolves, this must be reflected in the draft being proposed.

Similarly to *dLife* and given its potential, *SCORP* [19] shall be further exploited in what concerns also having its specification as a protocol with the IRTF DTN Research Group.

As the functioning of opportunistic routing is related to the properties displayed in the formed network structure, a future research direction includes the exhaustive analysis of such structure to draw concrete conclusions. So far, I have identified that small-world and scale-free properties emerge in the structure of social-aware networks in different periods of time. However, the more one understands such structures, the more one knows that they can be further used to improve the social-aware and content-based opportunistic utility functions and protocols devised in this Thesis work.

Another future direction involves the shift from the simulator to the real-world of the concepts learned in this work. The devised social-aware and content-based opportunistic utility functions and protocols have shown great potential to be employed in today's urban scenario: there are plenty of powerful devices that are wandering around and could be used to store, carry, and forward information upon the need to do it so. Moreover, user willingness to engage in such cooperative behavior would be boosted as soon as users know they will always have a means for opportunistically sending and retrieving content with associated cost of just carrying data on behalf of others.

8.3 Deviations from the Thesis Proposal

The only point that this Thesis work left unanswered was regarding the real-world experimentation. It was planned to have a prototype built with the findings of this Thesis.

This work was inserted in the context of DTN-Amazon as mentioned in Chapter 5. However, it is important to note that the project is done without any sort of funding and so far there is a struggle to find the proper manpower.

Despite these drawbacks, the team has managed to propose an initial implementation of the DTN architecture based on Android platform, the SocialDTN [34]. But still there is work to be done in order to have the real-world implementation ready for deployment and testing. This makes it a future research topic, which is also related to the ongoing standardization effort of *dLife*.

References

[1] W. Moreira, P. Mendes, "Social-aware Opportunistic Routing: The New Trend", in: I. Woungang, S. Dhurandher, A. Anpalagan, A. V. Vasilakos (Eds.), *Routing in Opportunistic Networks*, Springer Verlag, May, 2013. doi:10.1007/978-1-4614-3514-3_2.

[2] P. Hui, "People are the Network: Experimental Design and Evaluation of Social-based Forwarding Algorithms", Tech. Rep. UCAM-CL-TR-713, University of Cambridge, Computer Laboratory, March, 2008.

[3] V. Cerf, S. Burleigh, A. Hooke, L. Torgerson, R. Durst, K. Scott, K. Fall, H. Weiss, "Delay-Tolerant Networking Architecture", RFC 4838 (Informational), April, 2007.

[4] A. Vahdat, D. Becker, "Epidemic Routing for Partially Connected Ad Hoc Networks", Tech. Rep. CS-200006, Duke University, April, 2000.

[5] A. Lindgren, A. Doria, O. Schelén, "Probabilistic Routing in Intermittently Connected Networks", ACM SIGMOBILE Mob. Comput. Commun. Rev., vol. 7, no. 3, pp. 19–20, Association for Computing Machinery, July, 2003. doi:10.1145/961268.961272.

[6] S. Grasic, E. Davies, A. Lindgren, A. Doria, "The Evolution of a DTN Routing Protocol - PRoPHETv2", in: Proceedings of ACM MobiCom CHANTS, Las Vegas, USA, September, 2011. doi:10.1145/2030652.2030661.

[7] T. Spyropoulos, K. Psounis, C. S. Raghavendra, "Spray and Wait: An Efficient Routing Scheme for Intermittently Connected Mobile Networks", in: Proceedings of ACM SIGCOMM WDTN, Philadelphia, USA, August, 2005. doi:10.1145/1080139.1080143.

[8] J. Burgess, B. Gallagher, D. Jensen, B. N. Levine, "MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks", in: Proceedings of IEEE INFOCOM, Barcelona, Spain, April, 2006. doi:10.1109/INFOCOM.2006.228.

[9] S. Nelson, M. Bakht, R. Kravets, "Encounter-Based Routing in DTNs", in: Proceedings of IEEE INFOCOM, Rio de Janeiro, Brazil, April, 2009. doi:10.1109/INFCOM.2009.5061994.

[10] P. Hui, J. Crowcroft, "How Small Labels Create Big Improvements", in: Proceedings of PerCom Workshops, White Plains, USA, March, 2007. doi:10.1109/PERCOMW.2007.55.

[11] A. Mtibaa, M. May, C. Diot, M. Ammar, "Peoplerank: Social Opportunistic Forwarding", in: Proceedings of IEEE INFOCOM, San Diego, USA, March, 2010. doi:10.1109/INFCOM.2010.5462261.

[12] E. M. Daly, M. Haahr, "Social Network Analysis for Routing in Disconnected Delay-Tolerant MANETs", in: Proceedings of ACM MobiHoc, Montreal, Canada, September, 2007. doi:10.1145/1288107.1288113.

[13] P. Hui, J. Crowcroft, E. Yoneki, "Bubble Rap: Social-Based Forwarding in Delay-Tolerant Networks", IEEE Transactions on Mobile Computing, vol. 10, no. 11, pp. 1576–1589, IEEE Educational Activities Department, November, 2011. doi:10.1109/TMC.2010.246.

[14] P. Costa, C. Mascolo, M. Musolesi, G. P. Picco, "Socially-aware Routing for Publish-Subscribe in Delay-tolerant Mobile Ad Hoc Networks", IEEE J.Sel. A. Commun., vol. 26, no. 5, pp. 748–760, IEEE Press, June, 2008. doi:10.1109/JSAC.2008.080602.

[15] W. Moreira, P. Mendes, S. Sargento, "Opportunistic Routing Based on Daily Routines", in: Proceedings of IEEE WoWMoM AOC, San Francisco, USA, June, 2012. doi:10.1109/WoWMoM.2012.6263749.

[16] W. Moreira, P. Mendes, R. Ferreira, D. Cirqueira, E. Cerqueira, "Opportunistic Routing Based on Users Daily Life Routine", Internet Draft, draft-moreira-dlife-04 (work in progress), May, 2014.

[17] H. A. Nguyen, S. Giordano, "Context Information Prediction for Social-based Routing in Opportunistic Networks", Ad Hoc Netw., vol. 10, no. 8, pp. 1557–1569, Elsevier Science Publishers B. V., November, 2012. doi:10.1016/j.adhoc.2011.05.007.

[18] C. Boldrini, M. Conti, A. Passarella, "Design and Performance Evaluation of ContentPlace, a Social-aware Data Dissemination System for Opportunistic Networks", Comput. Netw., vol. 54, no. 4, pp. 589–604, Elsevier North-Holland Inc., March, 2010. doi:10.1016/j.comnet.2009.09.001.

[19] W. Moreira, P. Mendes, S. Sargento, "Social-aware Opportunistic Routing Protocol Based on User's Interactions and Interests", in: Proceedings of AdHocNets, Barcelona, Spain, October, 2013. doi:10.1007/978-3-319-04105-6_7.

[20] E. Jaho, M. Karaliopoulos, I. Stavrakakis, "Social Similarity as a Driver for Selfish, Cooperative and Altruistic Behavior", in: Proceedings of IEEE WoWMoM, Montreal, Canada, June, 2010. doi:10.1109/WOWMOM.2010.5534930.

[21] S. Farrell, V. Cahill, "Delay- and Disruption-Tolerant Networking", Artech House Inc., Norwood, USA, September, 2006.

[22] S. Jain, K. Fall, R. Patra, "Routing in a Delay Tolerant Network", in: Proceedings of ACM SIGCOMM, Portland, USA, August, 2004. doi:10.1145/1015467.1015484.

[23] W. Zhao, M. Ammar, E. Zegura, "A Message Ferrying Approach for Data Delivery in Sparse Mobile Ad Hoc Networks", in: Proceedings of ACM MobiHoc, Roppongi Hills, Japan, May, 2004. doi:10.1145/989459.989483.

[24] R. Shah, S. Roy, S. Jain, W. Brunette, "Data Mules: Modeling a Three-tier Architecture for Sparse Sensor Networks", in: Proceedings of IEEE ICC SNPA, Anchorage, USA, May, 2003. doi:10.1109/SNPA.2003.1203354.

[25] W. Moreira, P. Mendes, "Routing Metrics for Delay Tolerant Networks", in: Proceedings of CRC, Braga, Portugal, November, 2010.

[26] W. Moreira, P. Mendes, "Survey on Opportunistic Routing for Delay/Disruption Tolerant Networks", Tech. Rep. SITI-TR-11-02, SITI, University Lusofona, February, 2011.

[27] W. Moreira, P. Mendes, S. Sargento, "Assessment Model for Opportunistic Routing", in: Proceedings of IEEE LATINCOM, Belem, Brazil, October, 2011. `doi:10.1109/LatinCOM.2011.6107393`.

[28] W. Moreira, P. Mendes, S. Sargento, "Assessment Model for Opportunistic Routing", IEEE Latin America Transactions, vol. 10, no. 3, pp. 1785–1790, IEEE Region 9, April, 2012. `doi:10.1109/TLA.2012.6222585`.

[29] C. Ballester, J.-M. Seigneur, W. Moreira, P. Mendes, L. Maknavicius, A. Bogliolo, P. di Francesco, "Trust and Cooperation Incentives for Wireless User-Centric Environments", in: Proceedings of IADIS e-Society, Berlin, Germany, March, 2012.

[30] A. Bogliolo, P. Polidori, A. Aldini, W. Moreira, P. Mendes, M. Yildiz, C. Ballester, J. Seigneur, "Virtual Currency and Reputation-based Cooperation Incentives in User-Centric Networks", in: Proceedings of IWCMC, Limassol, Cyprus, August, 2012. `doi:10.1109/IWCMC.2012.6314323`.

[31] C. Ballester, J.-M. Seigneur, P. di Francesco, A. Bogliolo, W. Moreira, R. Sofia, N. Martins, V. Moreno, "A User-centric Approach to Trust Management in Wi-Fi Networks", in: Proceedings of IEEE INFOCOM, Turin, Italy, April, 2013.

[32] W. Moreira, P. Mendes, "Social-aware Utility Functions for Opportunistic Routing", Tech. Rep. SITI-TR-12-05, SITI, University Lusofona, August, 2011.

[33] W. Moreira, M. de Souza, P. Mendes, S. Sargento, "Study on the Effect of Network Dynamics on Opportunistic Routing", in: Proceedings of ADHOC-NOW, Belgrade, Serbia, July, 2012. `doi:10.1007/978-3-642-31638-8_8`.

[34] W. Moreira, R. Ferreira, D. Cirqueira, P. Mendes, E. Cerqueira, "SocialDTN: A DTN Implementation for Digital and Social Inclusion", in: Proceedings of the 2013 ACM MobiCom LCDNet, Miami, USA, September, 2013. `doi:10.1145/2502880.2502892`.

[35] W. Moreira, P. Mendes, "Social-aware Opportunistic Routing Solutions", Tech. Rep. SITI-TR-13-01, SITILabs, University Lusofona, January, 2013.

[36] W. Moreira, P. Mendes, "Structural Analysis of Social-aware Opportunistic Networks", Tech. Rep. SITI-TR-13-05, SITILabs, University Lusofona, August, 2013.

[37] E. P. C. Jones, L. Li, P. A. S. Ward, "Practical Routing in Delay-Tolerant Networks", in: Proceedings of ACM SIGCOMM WDTN, Philadelphia, USA, August, 2005. `doi:10.1145/1080139.1080141`.

[38] T. Spyropoulos, K. Psounis, C. S. Raghavendra, "Efficient Routing in Intermittently Connected Mobile Networks: the Single-copy Case", IEEE/ACM Trans. Netw., vol. 16, no. 1, pp. 63–76, IEEE Press, February, 2008. `doi:10.1109/TNET.2007.897962`.

[39] L. Song, D. F. Kotz, "Evaluating Opportunistic Routing Protocols with Large Realistic Contact Traces", in: Proceedings of ACM MobiCom CHANTS, Montreal, Canada, September, 2007. `doi:10.1145/1287791.1287799`.

[40] M. Grossglauser, M. Vetterli, "Locating Mobile Nodes with EASE: Learning Efficient Routes from Encounter Histories Alone", IEEE/ACM Trans. Netw., vol. 14, no. 3, pp. 457–469, IEEE Press, June, 2006. doi:10.1109/TNET.2006.876204.

[41] H. Dubois-Ferriere, M. Grossglauser, M. Vetterli, "Age Matters: Efficient Route Discovery in Mobile Ad Hoc Networks Using Encounter Ages", in: Proceedings of ACM MobiHoc, Annapolis, USA, June, 2003. doi:10.1145/778415.778446.

[42] T. Spyropoulos, K. Psounis, C. S. Raghavendra, "Spray and Focus: Efficient Mobility-Assisted Routing for Heterogeneous and Correlated Mobility", in: Proceedings of PerCom Workshops, White Plains, USA, March, 2007. doi:10.1109/PERCOMW.2007.108.

[43] C. Liu, J. Wu, "An Optimal Probabilistic Forwarding Protocol in Delay Tolerant Networks", in: Proceedings of ACM MobiHoc, New Orleans, USA, May, 2009. doi:10.1145/1530748.1530763.

[44] A. Balasubramanian, B. Levine, A. Venkataramani, "DTN Routing as a Resource Allocation Problem", in: Proceedings of ACM SIGCOMM, Kyoto, Japan, August, 2007. doi:10.1145/1282380.1282422.

[45] R. Ramanathan, R. Hansen, P. Basu, R. Rosales-Hain, R. Krishnan, "Prioritized Epidemic Routing for Opportunistic Networks", in: Proceedings of ACM MobiSys MobiOpp, San Juan, Puerto Rico, June, 2007. doi:10.1145/1247694.1247707.

[46] G. Palla, I. Derényi, I. Farkas, T. Vicsek, "Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society", Nature, vol. 435, no. 7043, pp. 814–818, Nature Publishing Group, June, 2005. doi:10.1038/nature03607.

[47] M. E. J. Newman, "Analysis of Weighted Networks", Phys. Rev. E, vol. 70, no. 5, American Physical Society, November, 2004. doi:10.1103/PhysRevE.70.056131.

[48] L. C. Freeman, "A Set of Measures of Centrality Based on Betweenness", Sociometry, vol. 40, no. 1, pp. 35–41, American Sociological Association, March, 1977. doi:10.2307\%2F3033543.

[49] P. Hui, J. Crowcroft, E. Yoneki, "Bubble Rap: Social-based Forwarding in Delay Tolerant Networks", in: Proceedings of ACM MobiHoc, Hong Kong, China, May, 2008. doi:10.1145/1374618.1374652.

[50] N. Eagle, A. Pentland, "Eigenbehaviors: Identifying Structure in Routine", Behavioral Ecology and Sociobiology, vol. 63, no. 11, pp. 1689–1689, Springer-Verlag, September, 2009. doi:10.1007/s00265-009-0830-6.

[51] T. Hossmann, T. Spyropoulos, F. Legendre, "Know thy Neighbor: Towards Optimal Mapping of Contacts to Social Graphs for DTN Routing", in: Proceedings of IEEE INFOCOM, San Diego, USA, March, 2010. doi:10.1109/INFCOM.2010.5462135.

[52] W. Zhao, M. Ammar, E. Zegura, "Multicasting in Delay Tolerant Networks: Semantic Models and Routing Algorithms", in: Proceedings of ACM SIGCOMM WDTN, Philadelphia, USA, August, 2005. doi:10.1145/1080139.1080145.

[53] Z. Zhang, "Routing in Intermittently Connected Mobile Ad Hoc Networks and Delay Tolerant Networks: Overview and Challenges", IEEE Commun. Surveys Tuts., vol. 8, no. 1, pp. 24–37, IEEE Press, January, 2006. doi:10.1109/COMST.2006.323440.

[54] R. J. D'Souza, J. Jose, "Routing Approaches in Delay Tolerant Networks: A Survey", International Journal of Computer Applications, vol. 1, no. 17, pp. 8–14, Foundation of Computer Science, February, 2010. doi:10.5120/370-557.

[55] T. Spyropoulos, R. N. Rais, T. Turletti, K. Obraczka, A. Vasilakos, "Routing for Disruption Tolerant Networks: Taxonomy and Design", Wirel. Netw., vol. 16, no. 8, pp. 2349–2370, Kluwer Academic Publishers, November, 2010. doi:10.1007/s11276-010-0276-9.

[56] A. Ferreira, A. Goldman, J. Monteiro, "Performance Evaluation of Routing Protocols for MANETs with Known Connectivity Patterns Using Evolving Graphs", Wirel. Netw., vol. 16, no. 3, pp. 627–640, Kluwer Academic Publishers, April, 2010. doi:10.1007/s11276-008-0158-6.

[57] Y. Cao, Z. Sun, "Routing in Delay/Disruption Tolerant Networks: A Taxonomy, Survey and Challenges", IEEE Commun. Surveys Tuts., vol. 15, no. 2, pp. 654-677, IEEE Press, May, 2013. doi:10.1109/SURV.2012.042512.00053.

[58] M. C. Gonzalez, C. A. Hidalgo, A.-L. Barabasi, "Understanding Individual Human Mobility Patterns", Nature, vol. 453, no. 7196, pp. 779–782, Nature Publishing Group, June, 2008. doi:10.1038/nature06958.

[59] W.-J. Hsu, T. Spyropoulos, K. Psounis, A. Helmy, "Modeling Spatial and Temporal Dependencies of User Mobility in Wireless Mobile Networks", IEEE/ACM Trans. Netw., vol. 17, no. 5, pp. 1564–1577, IEEE Press, October, 2009. doi:10.1109/TNET.2008.2011128.

[60] D. G. Murray, E. Yoneki, J. Crowcroft, S. Hand, "The Case for Crowd Computing", in: Proceedings of ACM SIGCOMM MobiHeld, New Delhi, India, August, 2010. doi:10.1145/1851322.1851334.

[61] T. Madsen, A. Grauballe, M. Jensen, A. Paramanathan, J. Rasmussen, F. Fitzek, "Reliable Cooperative Information Storage in Wireless Sensor Networks", in: Proceedings of ICT, St. Petersburg, Russia, June, 2008. doi:10.1109/ICTEL.2008.4652636.

[62] Q. Zhang, F. Fitzek, M. Katz, "Evolution of Heterogeneous Wireless Networks: Towards Cooperative Networks", in: Proceedings of CICT, Copenhagen, Denmark, November, 2006.

[63] D. Zhang, O. Ileri, N. Mandayam, "Bandwidth Exchange as an Incentive for Relaying", in: Proceedings of CISS, Princeton, USA, March, 2008. doi:10.1109/CISS.2008.4558621.

[64] S. Greengard, "Social Games, Virtual Goods", Commun. ACM, vol. 54, no. 4, pp. 19–22, Association for Computing Machinery, April, 2011. doi:10.1145/1924421.1924429.

[65] P. Obreiter, J. Nimis, "A Taxonomy of Incentive Patterns - The Design Space of Incentives for Cooperation", in: Proceedings of AP2PC, Melbourne, Australia, July, 2003.

[66] S. Buchegger, J.-Y. Le Boudec, "Performance Analysis of the CONFIDANT Protocol", in: Proceedings of ACM MobiHoc, Lausanne, Switzerland, June, 2002. doi:10.1145/513800.513828.

[67] R. Sofia, P. Mendes, "User-Provided Networks: Consumer as Provider", IEEE Communications Magazine, vol. 46, no. 12, pp. 86-91, IEEE Communications Society, December, 2008. doi:10.1109/MCOM.2008.4689212.

[68] A. Keränen, J. Ott, T. Kärkkäinen, "The ONE Simulator for DTN Protocol Evaluation", in: Proceedings of SIMUTools, Rome, Italy, March, 2009. doi:10.4108/ICST.SIMUTOOLS2009.5674.

[69] P. Hui, E. Yoneki, S. Y. Chan, J. Crowcroft, "Distributed Community Detection in Delay Tolerant Networks", in: Proceedings of ACM SIGCOMM MobiArch, Kyoto, Japan, August, 2007. doi:10.1145/1366919.1366929.

[70] C. Boldrini, M. Conti, A. Passarella, "Context and Resource Awareness in Opportunistic Network Data Dissemination", in: Proceedings of IEEE WoWMoM, Newport Beach, USA, June, 2008. doi:10.1109/WOWMOM.2008.4594890.

[71] C. Boldrini, M. Conti, A. Passarella, "Modelling Data Dissemination in Opportunistic Networks", in: Proceedings of ACM MobiCom CHANTS, San Francisco, USA, September, 2008. doi:10.1145/1409985.1410002.

[72] K. Scott, S. Burleigh, "Bundle Protocol Specification", RFC 5050 (Experimental), November, 2007.

[73] M. Doering, S. Lahde, J. Morgenroth, L. Wolf, "IBR-DTN: An Efficient Implementation for Embedded Systems", in: Proceedings of ACM MobiCom CHANTS, San Francisco, USA, September, 2008. doi:10.1145/1409985.1410008.

[74] R. Ferreira, W. Moreira, P. Mendes, M. Gerla, E. Cerqueira, "Improving the Delivery Rate of Digital Inclusion Applications for Amazon Riverside Communities by Using an Integrated Bluetooth DTN Architecture", International Journal of Computer Science and Network Security, vol. 14, no. 1, pp. 17–24, January, 2014.

[75] B. Batista, P. Mendes, "ICON - An Information Centric Architecture for Opportunistic Networks", in: Proceedings of IEEE INFOCOM NOMEN, Turin, Italy, April, 2013.

[76] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, A. Chaintreau, "CRAWDAD trace cambridge/haggle/imote/content (v. 2006-09-15)", September, 2006", Downloaded from http://crawdad.cs.dartmouth.edu/cambridge/haggle/imote/content.

[77] S. Ferretti, V. Ghini, F. Panzieri, "Scale-free Opportunistic Networks: Is It Possible?", in: Proceedings of PerCom Workshops, Lugano, Switzerland, March, 2012. doi:10.1109/PerComW.2012.6197590.

[78] T. Hossmann, T. Spyropoulos, F. Legendre, "A Complex Network Analysis of Human Mobility", in: Proceedings of IEEE INFOCOM NetSciCom, Shanghai, China, April, 2011. doi:10.1109/INFCOMW.2011.5928936.

[79] A. Casteigts, P. Flocchini, W. Quattrociocchi, N. Santoro, "Time-varying Graphs and Dynamic Networks", in: Proceedings of ADHOC-NOW, Paderborn, Germany, July, 2011. doi:10.1007/978-3-642-22450-8_27.

[80] F. Harary, G. Gupta, "Dynamic Graph Models", Mathematical and Computer Modelling, vol. 25, no. 7, pp. 79–87, Elsevier, April, 1997. doi:10.1016/S0895-7177(97)00050-2.

[81] D. J. Watts, S. H. Strogatz, "Collective Dynamics of Small-World Networks", Nature, vol. 393, no. 6684, pp. 440–442, Nature Publishing Group, June, 1998. doi:10.1038/30918.

[82] S. Dorogovtsev, J. Mendes, "Evolution of Networks", Advances in Physics, vol. 51, no. 4, pp. 1079–1187, Taylor & Francis Ltd, 2002. doi:10.1080/00018730110112519.

[83] V. Kostakos, "Temporal Graphs", Physica A: Statistical Mechanics and its Applications, vol. 388, no. 6, pp. 1007–1023, Elsevier, March, 2009. doi:10.1016/j.physa.2008.11.021.

[84] M. Bastian, S. Heymann, M. Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks", in: Proceedings of AAAI ICWSM, San Jose, USA, May, 2009.

[85] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks", Genome research, vol. 13, no. 11, pp. 2498–2504, Cold Spring Harbor Lab, November, 2003. doi:10.1101/gr.1239303.

[86] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, J. Scott, "Impact of Human Mobility on Opportunistic Forwarding Algorithms", IEEE Transactions on Mobile Computing, vol. 6, no. 6, pp. 606–620, IEEE Educational Activities Department, June, 2007. doi:10.1109/TMC.2007.1060.

[87] M. Newman, "The Structure and Function of Complex Networks", SIAM Review, vol. 45, no. 2, pp. 167–256, Society for Industrial and Applied Mathematics, 2003. doi:10.1137/S003614450342480.

[88] M. Newman, "Networks: An Introduction", Oxford University Press Inc., New York, USA, May, 2010.

[89] G. D'Angelo, S. Ferretti, "Simulation of Scale-Free Networks", in: Proceedings of SIMUTools, Rome, Italy, March, 2009. doi:10.4108/ICST.SIMUTOOLS2009.5672.

[90] M. Conti, S. Giordano, M. May, A. Passarella, "From Opportunistic Networks to Opportunistic Computing", IEEE Communications Magazine, vol. 48, no. 9, pp. 126–139, IEEE Communications Society, September, 2010. doi:10.1109/MCOM.2010.5560597.

[91] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, C. Diot, "Pocket Switched Networks and Human Mobility in Conference Environments", in: Proceedings of ACM SIGCOMM WDTN, Philadelphia, USA, August, 2005. doi:10.1145/1080139.1080142.

[92] N. Eagle, A. S. Pentland, "CRAWDAD data set mit/reality (v. 2005-07-01)", July, 2005, Downloaded from http://crawdad.cs.dartmouth.edu/mit/reality.