Computer Networks 000 (2016) 1–11

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Pervasive forwarding mechanism for mobile social networks

Kassio Machado[a,b,*], Azzedine Boukerche[b], Pedro O.S. Vaz de Melo[a], Eduardo Cerqueira[c], Antonio A.F. Loureiro[a]

[a] University of Ottawa, Ottawa, Ontario, Canada
[b] Federal University of Minas Gerais, Belo Horizonte, MG, Brazil
[c] Federal University of Para, Belem, PA, Brazil

ARTICLE INFO

Article history:
Received 16 January 2016
Revised 13 August 2016
Accepted 24 August 2016
Available online xxx

Keywords:
Opportunistic networks
Delay tolerant
Mobile social networks
Human behavior

ABSTRACT

In recent years, we have witnessed an increase in the popularity of mobile wireless devices and networks, with greater attention devoted to feasibility of opportunistic computing, sensing, and communication. In Mobile Social Networks (MSNs), communication is provided by spatial proximity and social links between peers, where personal devices carried by users communicate directly in a device-to-device mode. On one hand, human mobility provides encounters between peers and opportunities for communication without additional infrastructure; on the other hand, it introduces intermittent connections, network partitions, and long delay, requiring sophisticated message-forwarding mechanisms to improve network performance. Therefore, socially-inspired approaches which consider network structure and personal user features have been proposed to cope with these challenges. However, many studies disregard adaptive policies of message forwarding capable of dealing with variations of these features. In this paper, we investigated message dissemination in MSNs considering external factors such as temperature and seasonal calendar as environmental features capable of model users'preferences and encounters. We evaluated the time of day, the day of the week, and environmental variables such as weather and geographic position as important factors to the collective behavior and spatiotemporal characteristics of urban scenarios. This paper presents an analysis of real data from weather and human mobility, which depict distinct social interactions and spatial features characterized by changes in thermal conditions. Thus, we propose a socially-aware forwarding mechanism that is adaptable to the seasonality of personal preferences. Our experiments indicated that pervasive data can provide useful information towards the design of the next generation of human-centered Opportunistic Networks.

1. Introduction

The future of computer networks comprises a large variety of applications, composed of different devices and scenarios with many particular features and challenges. Among the new technologies, Opportunistic Networks is an emergent network paradigm focused on direct communication between devices for scenarios independent of infrastructure. Both industry and academia endorse the benefits of opportunistic communication for Delay Tolerant Applications [1], Vehicular Networks [2], Participatory Sensing [3], and Mobile Social Networks (MSN) [4], and, in addition, reinforce the challenges of the area. In these scenarios, regular nodes are mobile and have limited resources; communication occurs based on spatial proximity between peers due to friendship, routine, mobility, or simply by chance. These characteristics provide time-sensitive scenarios with frequent topology changes and lack of end-to-end paths the majority of the time. For this reason, traditional network protocols are neither efficient nor feasible, since they were not designed to deal with intermittent connections and network partitions.

The current ubiquity of portable wireless devices and increasing enhancement of hardware capabilities contribute to the growing interest in applications using this network class. The popularity of personal devices, such as smartphones, has led to significant development of online services focused on user content. Location-Based Social Networks (LBSNs), such as Facebook and Twitter, capture a significant amount of spatiotemporal data about environments and human behavior, turning their applications into repositories of geolocated social information. These online services capture user preferences and urban dynamics [5], and provide highly

* Corresponding author.
E-mail addresses: kmach102@uottawa.ca, kassiolsm@dcc.ufmg.br (K. Machado), azzedine@uottawa.ca (A. Boukerche), olmo@dcc.ufmg.br (P.O.S. Vaz de Melo), cerqueira@ufpa.br (E. Cerqueira), loureiro@dcc.ufmg.br (A.A.F. Loureiro).

http://dx.doi.org/10.1016/j.comnet.2016.08.022
1389-1286/© 2016 Published by Elsevier B.V.

contextualized data by means of real-time streams of observations regarding large sets of features [6]. The large volume of records describes interactions in social media applications, people boarding public transportation systems, and cell phone calls, among others. When chronologically grouped, these observations represent time-series of urban reality. Mining these data sources to formulate mobility models and peer encounters has become an important issue in mobile network scenarios. Moreover, insights about human behavior and its fluctuations have been shown to be relevant aspects for opportunistic networks [4]. Many proposals have studied opportunistic networks as complex systems sensitive to social and spatiotemporal aspects using real data [7–9]. They explore pervasive social context [10], such as social network contacts, personal interests and previously visited venues, in addition to complex network metrics, such as node centrality, betweenness, assortativity and network density.

In MSN scenarios, users are individuals carrying handheld devices with direct connection capabilities such as Bluetooth and Wi-Fi Direct in a device-to-device manner. Due to the relevance of human behavior in MSN applications, social features have been explored to identify communities and nodes with high centrality as a critical issue for improving network performance, since social aspects usually have long-term characteristics. In this direction, many forwarding algorithms have been proposed [11], but only a few consider the temporal changes of these features. For this reason, they are inefficient in front of variations in user mobility and network density, which are common in urban scenarios due to different characteristics of days of the week and time of the day.

This variability in scenarios represents a challenge to the communication method used in opportunistic networks. The Store-Carry-Forward method requires efficient mechanisms for choosing the best nodes and the best time to forward or replicate messages, a non-trivial procedure, considering device constraints such as buffer size, energy consumption and overhead. Usually, the proposed socially-inspired protocols select the relay nodes considering endogenous variables related to social aspects, and disregard environmental variables with potential influence on human behavior, and failing to incorporate mechanisms to adapt to fluctuations. Thus, in this paper, we investigated the following:

- The spatiotemporal variations of urban scenarios, according to several parameters including time of day, weather, and calendar (day of the week, month, seasonal weather, etc.)
- The effects of these variations on human behavior and on the performance of MSNs

The contributions of this paper are threefold: first, we show that the levels of venues'popularity and their visit patterns present distinct behaviors according to seasonal and weather conditions. These findings suggest that environmental variables can support the design of socially-aware and pervasive protocols as additional sources of information. Second, we have designed a simulation of opportunistic communications based on real data from social media applications, incorporating different settings of months, seasons, weather and mobility in New York City. The results present variations in network metrics according to thermal conditions, which evidences the relationship between environmental variables and human mobility, and their effects on the performance of MSN protocols. Finally, we propose a message-forwarding mechanism based on environmental features and node mobility, which applies the insights gained from observing fluctuations in human behavior.

The remainder of this paper is organized as follows. In the next section, we discuss an overview of the related work of message forwarding mechanisms for MSN, including flooding-based and socially-aware protocols. We also present investigations of fluctuations in human behavior characterized by environmental features. In Section 3, we describe the simulation model, the data used, and the combination of weather and social features in our simulations. Section 4 describes the *PervasivePeopleRank* algorithm, our proposal for forwarding messages on MSN-based applications. In Section 5, we present the simulation results, analysis, and findings of environmental effects on opportunistic social communications. Finally, in Section 6, we present our conclusions and future directions.

2. Related work

One of the most significant challenges of communication in opportunistic networks is the design protocol for optimized routing mechanisms. The protocols require sophisticated decision mechanisms to forward messages through the network, using one or more instances of them (replicas). These proposals investigate the use of the personal device capabilities of computing, sensing, communication, and data storage in order to monitor, predict and model entities and events that exist in the physical world, such as the cyber-physical Systems [12]. Therefore, the message forwarding mechanisms should be able to select the best nodes to forward messages and improve main performance indexes, such as delivery ratio and end-to-end delay, taking into account the overhead caused by multiple replicas, hops, and energy.

The Spray-and-Wait [13] (S&W) is one of the most popular algorithms for forwarding messages, using a flooding-based architecture divided into two steps. The split approach enables rapid diffusion of replicas on the network during the first step, in addition to using a customizable utility function for managing the replicas during the second step. Initially, each created message has λ replicas to spread on the network during the *spray* step. A relay node can be any node in the network that meets other nodes with $n > 1$ copies of the original message. As defined by a utility function, the relay node receives $c < n$ copies forwarded by the source or another relay node. When a node has only one replica of the message, it initiates the *wait* step. During this stage, it will not deliver the last replica until it meets the destination node.

Different mechanisms have been proposed for the *spray* and *wait* steps which extend the original algorithm, including Spray-and-Focus [14], which changes the *wait*. The new *focus* step determines that messages with one local replica will be forwarded to their destinations or other relay nodes, based on an evaluation of the time interval since the last two meetings between nodes. The main advantage of this approach is the controlled number of replicas in the network; this is defined by λ, which represents an upper bound to the overhead.

Recent studies have investigated MSNs considering the nodes as users of personal devices such as smartphones, to take advantage of social aspects [8]. These proposals have explored social aspects, such as node popularity [15], social group labeling [16], expected delay and the number of encounters [17], explicit mutual interests [18], and a combination of communities and node centrality [19]. In this direction, Moreira et al. [20] investigated the impact of human behavior on opportunistic social networks. They studied the use of social aspects and data similarity to develop opportunistic forwarding systems for essential services in extreme networking conditions and dense networking scenarios. Furthermore, their work shows suitable types of opportunistic forwarding schemes, according to network density. Their experiments used simulations based on real and synthetic mobility traces, and their findings point to the investigation of self-awareness mechanisms and adaptable forwarding schemes based on network features and the dynamism of user behavior.

Chen et al. [17] proposed a forwarding scheme that considered information from node encounters and time-to-live (TTL) message property. The authors proposed a routing protocol for delay-tolerant applications that distributed multiple replicas between nodes, in proportion to their expected encounter ratio. Their work presented the Expected Encounter-based Routing protocol (EER), using the metrics Expected Encounter Value (EEV) of each node and the minimum Expected Meeting Delay (EMD) between the current node and the destination. Whith similarity to Spray-and-Focus, messages are created with λ replicas and spread on the network proportionally to EEV. Thus, when the number of replicas of a held message is reduced to 1, the single replica is forwarded only to the destination node or a relay node with lower EMD. The experiments used the vehicle-based mobility model, which is part of ONE Simulator [21].

Mtibaa et al. [15] proposed a forwarding mechanism based on node popularity, derived from the *PageRank* algorithm [22]. The *PeopleRank* proposal explores the popularity of nodes using a distributed approach, forwarding new copies of the original message to nodes ranking higher than the current node. The messages are duplicated on demand, and without a specific limit of replicas. The performance evaluation presented results using six datasets of real data, with 27 up to 414 nodes.

Ciobanu et al. [23] explored the social graph from social media applications to provide additional information and support the message forwarding mechanism. The proposed algorithm, OpportuNistic Socially-aware and Interest-based DissEmination (ONSIDE), takes users'interests and contact history into consideration to decrease the congestion and required bandwidth, taking into account the overall network's hit rate and the delivery latency. Similarly, Socievole et al. [24] introduced the multi-layer social network model, which combines social networks based on proximity and online social networks. The authors investigated the relationship between different social network layers regarding node centrality, community structure, link strength, and prediction. Both works discuss the advantages of using social aspects to improve opportunistic dissemination, and the benefits of using online social media applications to obtain the social graph. Nevertheless, these proposals assume an eventual connection to the Internet or to remote servers of social media applications. These assumptions make it difficult to use these proposals in scenarios without infrastructure.

Environmental features can change the social and network variables used by these proposals when a contextual variable (e.g., weather, traffic conditions, day of the year) reaches a critical value, causing changes in the variable of interest (e.g., connection duration, distance traveled, node degree, clustering coefficient). These contextual tipping points, according to the definition of Lamberson et al. [25], can represent symptoms of change in environmental characteristics. Bakhshi et al. [26] discussed how weather conditions can influence people's mood, retail sales, the stock market, among others. The authors argued that many of the effects seen in online communities can be explained using offline theories from experimental psychology. Results showed that during visits to restaurants, user experiences varied according to weather conditions, which also influenced customers'online reviews. Similarly, Bannur et al. [27] studied social media check-in data from the user's perspective, investigating seasonal polarity of check-ins in different regions of the United States. Results showed the seasonal behavior of check-ins for specific categories of venues during the 12 months of 2013 by quantifying the popularity of movies, restaurants, shopping locations, etc., on different days of the week and different months. In addition to seasonal variation in visits, the results showed that ranking of the most popular venues varied during the year.

Considering an urban scenario and social media traces, Cho et al. [28] showed that humans experience a combination of strong, short-range spatially and temporally periodic movement, which is not impacted by the social network structure. Their work showed that, by investigating the Brightkite and Gowalla LBSNs, social relationships can explain about 10%–30% of all human movement, while periodic behavior can explain 50%–70%.

The state of the art of both topics, socially-inspired protocols and social media data mining, classified urban scenarios as dynamic systems and pointed to the influence of social aspects and exogenous variables. Most of the performance evaluations carried by recent studies considered real mobility traces, but the data analyzed represented only a few hundred users, small sets of communities or limited geographic areas, such as universities or conference centers [29–31]. Moreover, existing socially-aware studies have implemented mechanisms based on the history of encounters regardless of their fluctuations and characteristics, which have the potential to deteriorate communication network performance. For this reason, the design of socially-aware forwarding mechanisms with the ability of adapting to different network configurations is a recent challenge, in which prediction of critical points of change can support the pervasive mechanisms in improving the performance in MSNs.

3. Trace-based analysis

In this section, we describe the real data used in simulations, as well as the methodology used to combine weather and social media data. Many papers have explored social media applications to simulate large urban scenarios and investigate their dynamics [7,32,33]. On the face of it, we reinforce the use of real data in our experiments, because environmental conditions are complex to simulate, and their effects on the behavior of users are better observed in situ [34].

3.1. Data description

Many geolocalized data samples about daily life in urban environments are available through urban streams [6], and can be combined as layers of information [3]. Each geolocalized record represents an event limited by a temporal window and spatial area, such as sensing samples of mobility, content interest, venue popularity, etc. We used public data sources in a combined approach to analyze the spatial distribution of users, and encounters between them, in different environmental configurations.

The data collected comprises geolocated data samples of weather conditions and human mobility limited to Manhattan in New York City (NYC) from February to August 2015. The traces of human mobility were built using data from social media applications, specifically geolocalized photos on Instagram[1] and check-ins on Foursquare[2], resulting in a dataset of 1.3 million samples.

By using social media applications as data sources, we obtained real data about venues, users, and encounter routines. Thus, in this work, our simulations consider commutes between real locations, a large number of users with distinct behavior, and areas with time-sensitive agglomerations. According to public data collected from those data sources, we defined a data sample from social media as a 3-tuple $s_m =< u, p, t >$, where u represents an user $u_i \in U$, t is the timestamp of the sample, and p is the u_i's position defined by latitude and longitude coordinates. In addition, we defined the path traveled by u_i within a time window as $u_i^{ts} = \{s_{m1}, s_{m2}, \ldots, s_{mk}\}$.

Weather conditions were collected from the National Weather Service (NWS) and public stations via the Weather Underground[3] service. The service provides data about weather variables with a sensing frequency of up to 60 mi of interval, obtained from 54 weather stations in the area of interest. Weather data samples are

[1] http://www.instagram.com.
[2] http://www.foursquare.com.
[3] http://www.wunderground.com.

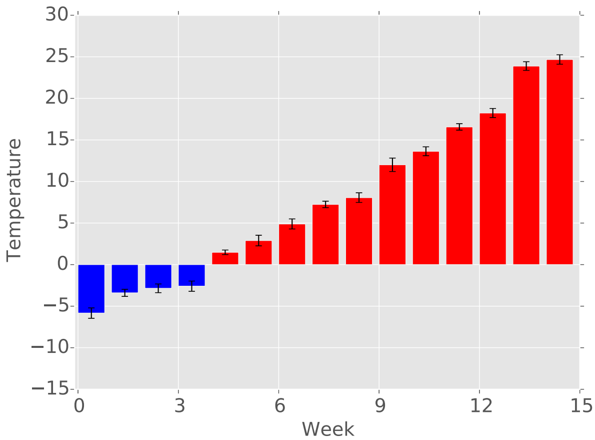


Fig. 1. Temperature average for the selected time series.

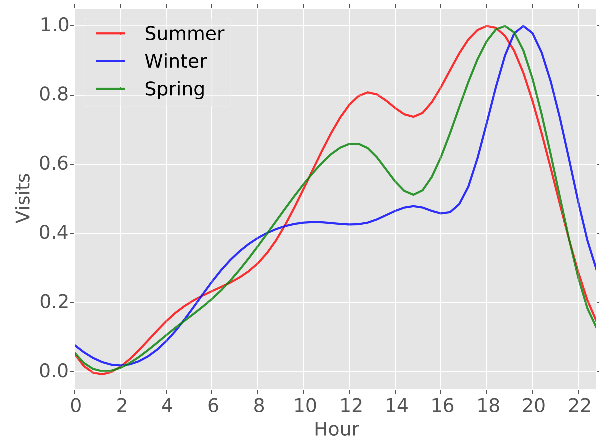
defined as a 3-tuple $w = < t_p, p, t >$, where t_p is the temperature measured in degrees Celsius, p is the position of the weather station, and t is the sample timestamp. The weather conditions of a simulated time series are summarized according to the average of all temperature measures during the selected time window, and classified according to variance.

Using this model, we defined each trace as $T = < U_t, t_p, \Delta t >$, where U_t is a set of u_i^{ts}, t_p is the average of temperature measures, and Δt is the time window of analysis. The set of traces comprises 15 independent time series grouped into seven days, starting on Monday and ending on Sunday, which are subsets of collected data and selected according to the absence of holidays and low variance of temperature. By using this methodology, we defined classes of temperature grouped by intervals of 5 °C, as shown in Fig. 1.

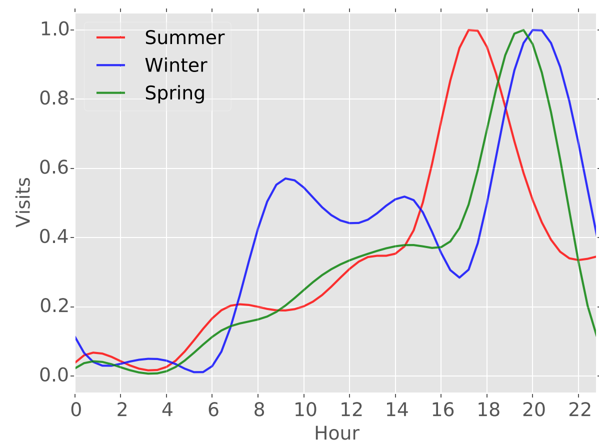
The collected data refers to the period previously mentioned, and is limited by the bounding box of Manhattan defined by geographic coordinates[4]. The social media data samples were collected using the Twitter Stream API[5], and represent data samples obtained at the moment of its online publication, and originally published by mentioned applications; in other words, the samples are collected in real-time and limited to the Foursquare and Instagram applications. The weather data samples are limited according to the geographic position of the weather stations, and are obtained using public API of Weather Underground, which provides queries based on geolocation and date.

3.2. Data combination

Fig. 2 shows the time series of visits for two Points of Interest (POI) in NYC: Central Park (CP) and Times Square (TS). The data represents the normalized average number of visits[6] during daily hours in different seasons and weather conditions. Both places present similar peaks of popularity during the night, but more than one peak occurs in the summer season, specifically at CP, where two similar peaks were registered and did not occur with the same intensity during winter and spring. The difference seen in these time series illustrates how visiting patterns can be influenced by thermal and temporal variations. Note that even popular venues, which can attract crowds any day of the year (such as in well-known POIs), present fluctuations characterized by environmental variables and seasonality.



(a) Central Park



(b) Times Square

Fig. 2. Patterns of check-ins and photos during the seasons (time series temperature).

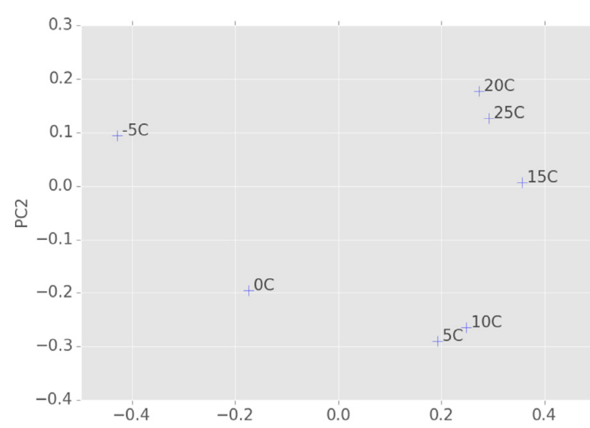
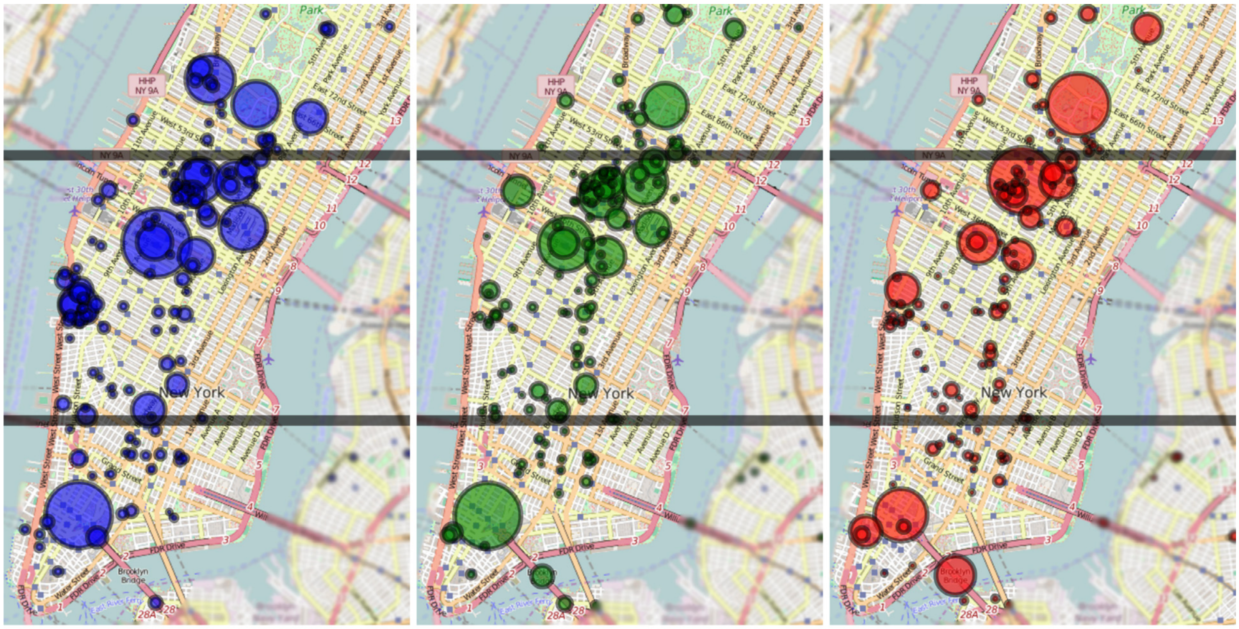


Fig. 3. Principal component analysis of venues'popularity according to temperature.

In order to verify whether there are significant differences in the activities done in NYC when the weather changes, we created a $m \times n$ matrix M that represents the places people visit in NYC at different temperatures. Each row $i \in \{1, 2, \ldots, m\}$ of M is a 5 °C temperature range, and each column $j \in \{1, 2, \ldots, n\}$ is the average amount of data samples in place p_j when the temperature was in the range defined by row i. Thus, Fig. 3 shows the Principal Component Analysis (PCA) for matrix M, that is, each

[4] The guidelines for data collection, as well as tools used and their parameters, are available on http://homepages.dcc.ufmg.br/~kassiolsm/comnet.
[5] Application Programming Interface available online on https://dev.twitter.com/streaming/overview.
[6] normalized by the max of individual time series.

Please cite this article as: K. Machado et al., Pervasive forwarding mechanism for mobile social networks, Computer Networks (2016), http://dx.doi.org/10.1016/j.comnet.2016.08.022



(a) Phase negative (b) Phase of transition (c) Phase positive

Fig. 4. Popular venues in New York City in different phases.

point in the graph is a 5 °C temperature range, and the horizontal and vertical axes represent the first and second principal components of *M* according to PCA, respectively. The first two components can explain 74% of the variance seen in the data. The results presented distinct values for the set of temperatures observed, i.e., venue popularity in NYC varies according to the local temperature. The first component (on the horizontal axis) shows the difference between cold and hot temperatures, while the second component (on the vertical axis) apparently measures temperature extremes. Based on these observations, we modeled the popularity of venues in three phases: *negative, transition*, and *positive*. Wherein the *negative* phase comprises time series with average temperatures lower than 0 °C; the *transition* phase includes time series with average temperatures between 0 °C and 10 °C; and the *positive* comprises time series with average temperatures greater than 10 °C.

Fig. 4 shows the analysis of geolocalized data samples, according to the three phases defined in the PCA. The circles represent popular venues in the area of interest, and the size of the circle represents popularity according to the average daily number of visits (for better visualization, we maintained a limit of only 150 of the most popular venues). The results show a variation in popularity during the phases, with new venues observed only in specific phases. For example, during the phase *negative*, three POIs with similar levels of popularity close to Central Park are observable in the North, but their popularity changes during the *transition* and *positive* phases. A similar situation was registered with the Brooklyn Bridge on the South, where at least three POIs were observable in the *positive* phase.

Fig. 5 presents the entropy matrices, grouped according to the phases defined in PCA. Each element of the matrix represents the entropy calculated using $i \in \{1, 2, \ldots, n\}$ that represents the number of data samples at a place p_i observed in intervals of two hours, and according to the days of the week. Entropy values are related to the total number of check-ins observed, where low values indicate few opportunities for encounters between users due to sparse check-ins and their spatial distribution. Hours with lower entropy values occur in periods outside regular business hours in the *transition* and *positive* phases. Entropy begins increasing at 8 h and decreasing at 0 h during the weekdays, a consequence of the routine behavior of the citizens of NYC. The entropy values show critical hours; they are time windows with low mobility, capable of negatively impacting opportunistic communication performance. The phases emphasize the distinct patterns of critical hours, showing the fluctuation of spatial distribution and mutable characteristics of the critical hours set. Few users keep moving according to their particular features; therefore, forwarding mechanisms should pay attention to nodes with high mobility for improving the network performance in critical hours. It is important to note that several particular situations and variables can influence the spatial distribution of people, such as holidays, musical events, traffic jams, and weather conditions. In particular, weather conditions such as snow, rain or severe temperatures can influence personal preferences and urban mobility in the form of traffic jams, inclination to indoor places, and increased demands on public transportation.

4. PervasivePeopleRank

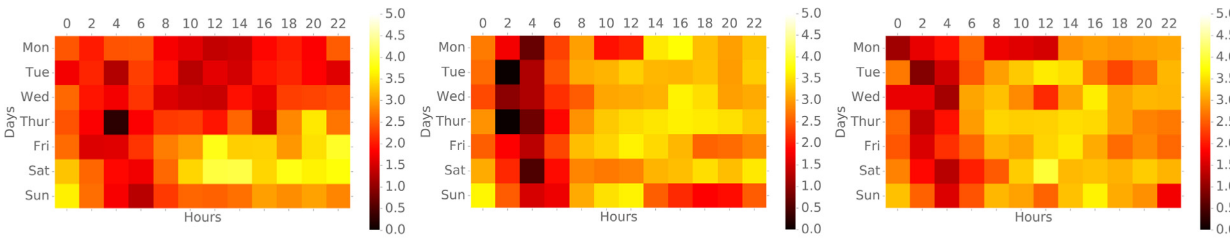
In this section, we present the *PervasivePeopleRank* (PPR), an algorithm designed for forwarding messages in MSN applications, which selects relay nodes based on information about users and the environment.

The PPR extends the previous protocol *PeopleRank* (PeR) proposed by Mtibaa et al. [15], which ranks the nodes according to their social links. When an encounter between two nodes N_i and N_j occurs, the algorithm calculates the individual *PeR* value using the following equation:

$$PeR(N_i) = (1 - d) + d \sum_{N_x \in F_i} \frac{PeR(N_x)}{|F_x|} \tag{1}$$

Eq. 1 describes the *PeR* computation performed on both nodes, where F_i is the set of neighbors connecting to N_i (social links) and d is a damping factor defined as the probability, at any encounter, that the social link between nodes improves the rank of the nodes involved. The damping factor ($0 < d \leq 1$) controls the weight given

Please cite this article as: K. Machado et al., Pervasive forwarding mechanism for mobile social networks, Computer Networks (2016), http://dx.doi.org/10.1016/j.comnet.2016.08.022



(a) Phase negative (b) Phase of transition (c) Phase positive

Fig. 5. Entropy average of encounters grouped by phases.

to the social links on the forwarding decision. The *PeR* value is the metric used for replicating and sending messages towards the central nodes of the network, which have a higher probability of knowing the destination node.

Originally, the social links used in the metric are collected from social media applications. Therefore, the metric eventually requires connection to the Internet or a server capable of providing the users'social graph. Meanwhile, we adapt the protocol to compute the social links using nearby devices close enough to connect directly. The PeR protocol represents a feasible alternative to large scenarios, with a lack of infrastructure and susceptibility to variable features. PeR provides customization of the impact of social links using the damping factor, which provides the adaptability to work in scenarios without additional resources (remote servers and Internet) and the low complexity to compute the main metric *PeR* in distributed form.

The PPR considers seasonal and thermal aspects due to their effects on mobility preferences and node connectivity, taking into account the date, hour and temperature. Algorithm 1 shows the PPR forward decision, in which nodes N_i and N_j share their *PeR* values and the size of their respective sets of social links. The two nodes then update their *PeR* values and replicate messages, if N_j has a greater PeR_j value than PeR_i or the node destination is known by N_j.

Algorithm 1: PervasivePeopleRank Algorithm

$PeR_i \leftarrow PeR(N_i)$;
$PeR_j \leftarrow send(PeR_i)$;
$F_j \leftarrow send(F_i)$;
$PeR_i \leftarrow update(PeR_j, F_j)$;
for $m \in buffer(i)$ **do**
  if $PeR_j \geq PeR_i$ *OR destination*$(m) \in F_j$ **then**
    forward(N_j, m);
  else
    $\Delta M_j \leftarrow send(\Delta M_i)$;
    if *critical(hour) AND* $\Delta M_j \leq \Delta M_i$ **then**
      forward-ephemeral(N_j, m);

The case of $PeR_j < PeR_i$, PPR applies a time-dependent mechanism which evaluates two features:

- environmental: PPR evaluates whether the current hour is a critical hour of encounters employing the entropy matrices. In our experiments, we defined a critical hour as one that demonstrates lower entropy than the daily average.
- node mobility: the algorithm also evaluates the ΔM_i, which is the daily average of time intervals between mobility events of the node N_i.

We assume the nodes are capable of storing the entropy matrices and the social links locally. The data can be stored in key-value data structures indexed by phases, day of the weeks and time of day the case of the matrices and by the id of user in case of social links. To mitigate the storage cost of social links and the impact of encounters with a single occurrence, we assume that each social link has a lifetime of τ hours. The τ defines the maximum interval between two consecutive meetings of two random users; if the encounter does not happen again before the deadline, the social link is removed. Otherwise, the deadline is renewed.

The environmental and node mobility features are evaluated to cope with hours of low ratio of encounters. Thereby, we assume that in addition to the capability of knowing the day and hour, all nodes are equipped with sensors or other resources for measuring temperature and mobility events. Obtaining the time and calendar information are trivial tasks for modern personal devices. Additionally, these devices have sensors for temperature, luminosity, pedometer, accelerometer, etc., capable of acquiring data about the environment and users'activities, such as weather conditions, walking, and cycling. Therefore, we point out that mobility events can be obtained using alternatives to GPS (Global Positioning System). Thus, the PPR does not enable forwarding based on geographic location; it mitigates privacy issues using the size of the social links set (not the identity of social links) and the time registered for mobility events, instead of users'geographic coordinates.

Urban scenarios can provide a large number of users with different patterns of mobility. The PPR exploits this feature during critical hours, creating *ephemeral* copies of messages, a kind of replica forwarded to nodes with lower *PeR* and higher mobility ($\Delta M_j > \Delta M_i$). Messages flagged as *ephemeral* are forwarded normally, nevertheless with TTL $=$ min (T_m, H_n), where T_m is the original TTL of the message and H_n the end of the critical hour.

5. Performance evaluation

In this section, we present the network model used for simulating the opportunistic communications, the connectivity graph, and the network performance of the *PervasivePeopleRank* algorithm.

5.1. Network model

Node mobility is determined according to the definition given in Section 3.1. Thereby, given two data samples s_{mi} and $s_{mj} \in T$, the settings of opportunistic communication experiments consider an encounter and network connection event between users u_i and u_j when:

- the distance $dt \leq DT_{range}$ between positions p_i and p_j;
- the contact interval $c_{ij} \leq C_{time}$ between time stamps t_i and t_j;

where the DT_{range} is the distance threshold, defined as 50 m (usually reached by Bluetooth or WiFi Direct technologies), and interval C_{time} was experienced as a parameter that varied between 5 mi and one hour. The encounters are formally described as a network contact graph $G(V, E)$, in which the stochastic process of

Table 1
Values for the simulation parameters.

	Parameter	Value
Network	Contact interval (C_{time})	5, 30 and 60 min
	communication range	50 m
	area	25.15 x 24.01 km
	# of nodes	12,854 $\leq n \leq$ 18,315
	Message creation	Each data sample and random $n \leq C_{time}$ minutes
Spray and Wait (S&W)	Replicas (λ)	1000
Expected Encounter Routing (EER)	Replicas (λ)	1000
	Re-encounter time frame	48 h
PeopleRank (PeR)	Damping factor (d)	0.8
PervasivePeoplerank (PPR)	Damping factor (d)	0.8
	Social link lifetime (τ)	48 h
dLife	Re-encounter time frame	48 h

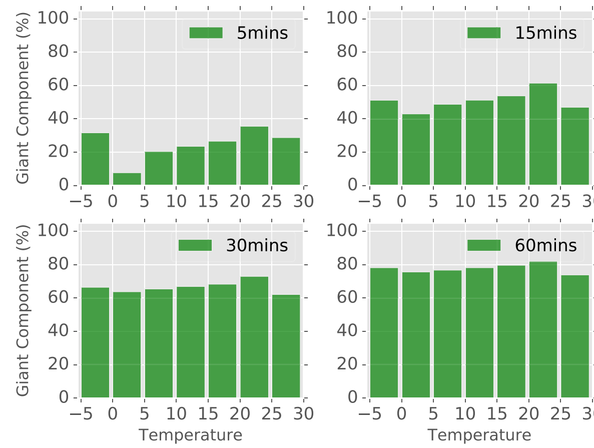
encounter between two nodes $i, j \in V$ is modeled as an edge $e(i, j) \in E$. We assume that the network contact graph is undirected, therefore node i contacts j whenever j contacts i.

The simulation parameters are described in Table 1. The fixed number of replicas used in EER and S&W simulations is enough to compare with related work, as shown in Section 5.3. The damping factor used by PeR and PPR are defined as shown in [15,24] to provide significant relevance to social links. The lifetime of social links defined by τ and the re-encounter time frame were defined considering time series used in simulations composed of seven days.

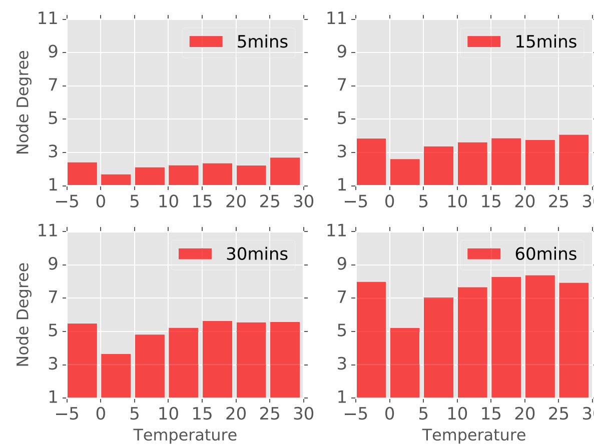
5.2. Contact graph analysis

The network analysis takes into account the contact graphs G_{ct} formed during the trace-based simulations, and the observed environmental temperatures. The graphs are grouped into 4 configurations of C_{time}. Fig. 6a shows the size of the giant component of the contact graph for simulations of different durations and temperatures. Observe that the size of the giant component, when temperatures are inside the *transition* phase, is reduced by up to 19.1% when compared to other phases. The differences in size are noticeable in simulations with C_{time} of 5 and 15 min, which represent 10.1% and 28.3% of all observed encounters in the dataset, respectively. Additionally, results show that C_{time} equal to 15 min is enough to connect more than half of the nodes in the giant component for most scenarios.

Fig. 6b shows the average degree of nodes, according to contact graphs and C_{time} configurations. The results showed that the temperature shift from -5 °C to 0 °C signals the most significant changes in the network structure, where the degree of nodes decreases by an average of 32.2%. Fig. 7 shows the Complementary Cumulative Distribution (CCDF) of the shortest path between any i and $j \in G_{ct}$ using C_{time} as 60 min. The changes in graph structure are characterized by the specific range of temperatures defined in the *transition* phase. The metrics showed the *positive* and *negative* phases as well connected, which provide efficient communication; however, the temperatures of the *transition* phase indicated sparse connectivity and longer paths. Thus, adaptive approaches to forwarding mechanisms are required to deal with the variations of the network structure, in addition, the environment can characterize the changes and provide early-warnings signals [35].



(a) Giant component size.



(b) Average degree of nodes.

Fig. 6. Contact graphs.

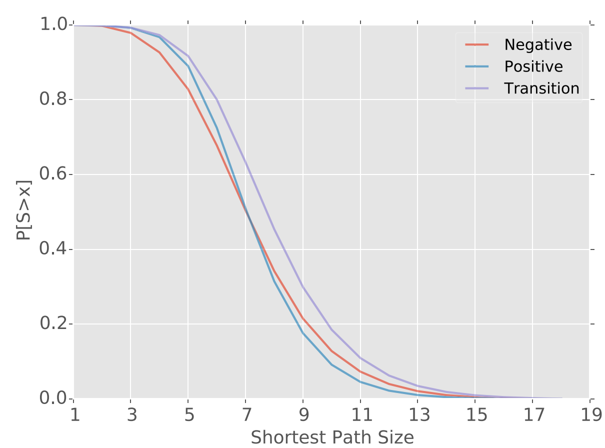
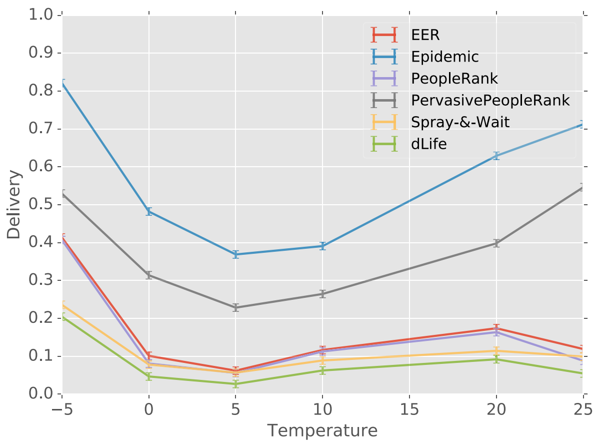


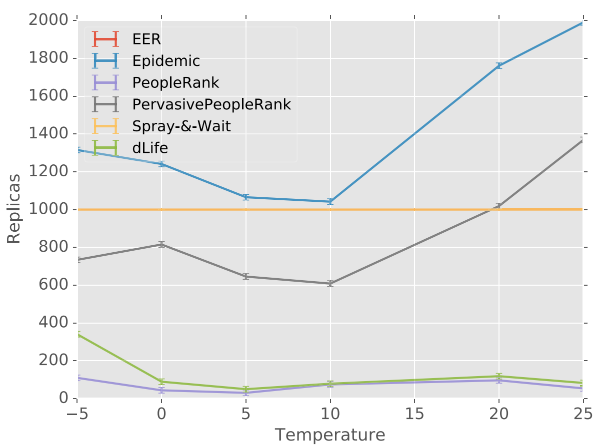
Fig. 7. Distribution of short path.

5.3. Network performance

To evaluate the effects of environment and human behavior on MSN applications and on the proposed forwarding mechanism, we compared *PervasivePeopleRank* (PPR) with 5 other mechanisms: EER [17], PeopleRank [15] (PeR), Spray-and-Wait (S&W) [13], dLife [20], and Epidemic. The opportunistic communication results are presented with a confidence interval of 95% in terms of deliv-

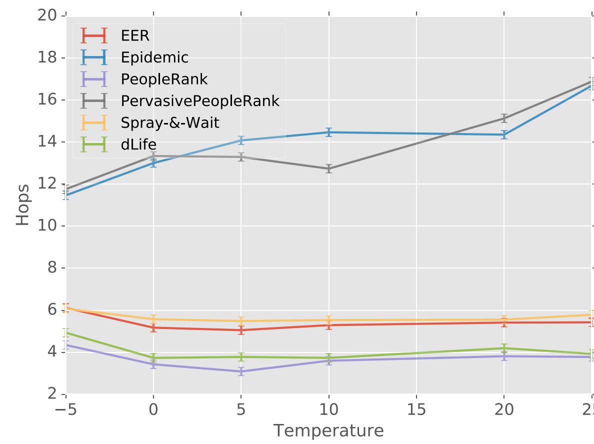


(a) Delivery ratio.

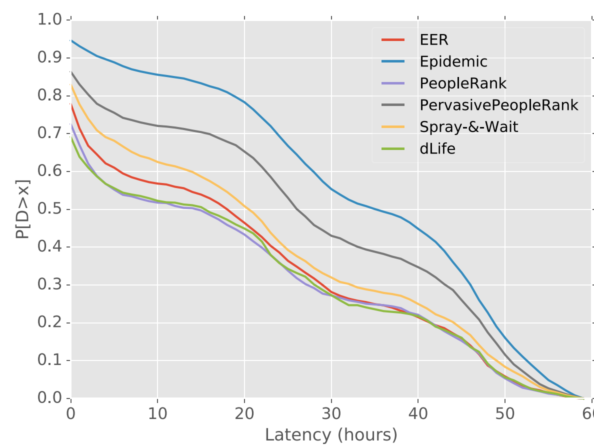


(b) Average of replicas.

Fig. 8. Delivery ratio and average cost according to temperature variation.



(a) Average of Hops



(b) Latency

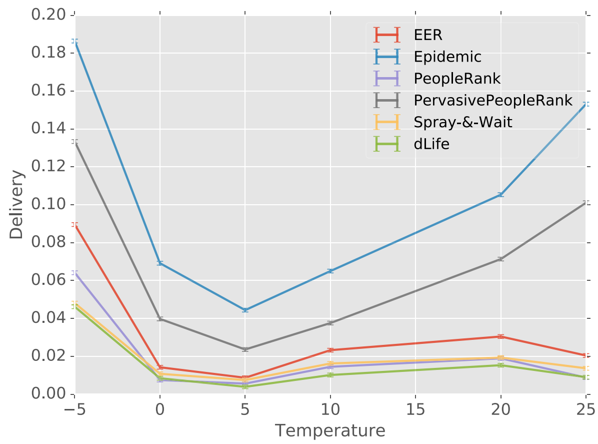
Fig. 9. Average of hops and CCDF of latency.

ery probability (ratio between the number of delivered messages and the number of messages that should be delivered), delay (time elapsed between message creation and delivery), cost (amount of replicas available in the network at the moment of delivery) and hops. The network traffic is generated based on time and mobility. The messages are created for random destinations in two moments: when a node publishes a new data sample (usually changing its position), and after random n minutes since the last published data sample, where $n \leq C_{time}$.

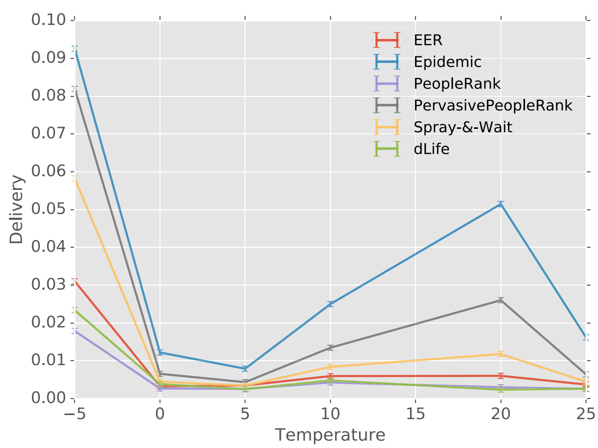
In regards to node buffers, the default TTL of messages is 72 h to attend the usual sparsity nature of opportunistic networks. In addition, we defined messages as generic packets independent of content to focus on message diffusion. Each message represents a unit on buffer, with a capacity for 1000 unique messages. Figs. 8 and 9 present the simulation results using C_{time} as 60 min and λ as 1000 replicas. The delivery results in Fig. 8a show decreasing performance in temperatures inside of the *transition* phase. Nevertheless, PPR algorithm delivered at least 57.8% more messages than the remaining related protocols for the same phase. In the simulations with temperatures corresponding to the *positive* phase, the improvement is 69%. Messages delivered during critical hours of encounters increased 48.2% using PPR. The average number of replicas presented in Fig. 8b shows the constant value for protocols EER and S&W, which are based on the replica limit λ. The increased number of replicas at higher temperatures using PPR occurs as a result of the higher number of contacts provided through mobility. The average interval between mobility events ΔM_n decreases by 11.7% in these temperatures.

Fig. 9 shows the average number of hops and the CCDF of latency. With respect to these results, it is worth emphasizing that simulations of urban areas, such as NYC, can provide a large number of single encounters (in other words, encounters with just one occurrence). In addition, these application scenarios provide subsets of nodes with few connections or low mobility, that is, nodes walking in small sub-areas or visiting unpopular places. Nodes with these features are accessible mainly through long paths or specific nodes, such as bridge nodes, responsible for connecting different communities and areas [20]. For this reason, Epidemic with the simple flood technique provides the best performance of delivery ratio and high average of hops. Actually, the related protocols select relay nodes primarily considering centrality and social aspects, in an attempt to use short paths and lower delay. However, in large geographic areas these approaches limit the number of feasible encounters to message transfer to a set of low-frequency events, and negatively affect delivery. That is, the related protocols quickly reach the well-connected nodes (Fig. 9a); nonetheless, the messages are replicated or forwarded to another node with higher centrality, another node that had previously met the destination, or directly to the destination. In case of few connected destinations or low relay node mobility, more time may be required before a more suitable candidate for relaying the node is encountered, or a node from the destination social group is found.

Please cite this article as: K. Machado et al., Pervasive forwarding mechanism for mobile social networks, Computer Networks (2016), http://dx.doi.org/10.1016/j.comnet.2016.08.022



(a) 30 minutes.



(b) 5 minutes.

Fig. 10. Delivery ratio considering different C_{time}.

dLife and PeR use 85.8% fewer replicas because the messages are usually forwarded to the high centrality nodes, but the infrequent encounters with feasible relay nodes, according to their respective decision mechanisms, stops the diffusion. Hence, the delivery rate is 76.4% less than PPR. The greedy approach of PPR reaches distant nodes and improves the delivery ratio, but naturally increases the overall number of hops. Nevertheless, PPR delivers 20.1% more messages using 15% fewer replicas than EER and S&W.

Fig. 10 presents the delivery results, using C_{time} as 30 and 5 min, and the performance is proportionately similar. Observe that the delivery rate in these scenarios decreases as temperatures fall in the *transition* phase. Nevertheless, the delivery rate using the proposed protocol is 54.4% and 47.9% better than related proposals in these scenarios, respectively. Considering all scenarios, the delivery rate is improved by at least 54.1% and 61.4%.

6. Conclusion

In this paper, we investigated the seasonal patterns of urban mobility and their features facing thermal variation. Our observations indicated some effects of spatiotemporal features in human mobility and encounters in a MSN application. The social media data used in our investigation presented a fluctuation in venue popularity and of probable encounters between peers. Results showed that temperature can explain 74% of the variance in the popularity of venues. Moreover, we showed that distinct patterns of encounters can be characterized by 3 ranges of temperatures. The changes in environmental variables provided the identification of distinguished behaviors observable by the spatial distribution of users, an important feature for the design of message forwarding mechanisms for people-centric approaches and large geographic areas.

In addition, we used the spatiotemporal insights to propose the *PervasivePeopleRank*, a cyber-physical message forwarding mechanism for Mobile Social Networks. The mechanism improves delivery by an average of 57.8% by distributing multiple replicas of messages according to node centrality, mobility and seasonal aspects.

Finally, our results indicate that environmental factors can characterize the state of the network, providing insights about the dynamism of urban scenarios. Specifically, temperature was shown to be a relevant feature in assisting the forwarding decision process for networks based on physical proximity and susceptible to human behavior.

Acknowledgments

This work was partially supported by the Canada Research Chairs program, NSERC Strategic Project Program and NSERC DIVA Network Research Program, and ORF/MRI Program. The authors wish to also thank the Brazilian research departments CNPq and CAPES.

References

[1] K. Wei, X. Liang, K. Xu, A survey of social-aware routing protocols in delay tolerant networks: applications, taxonomy and design-related issues, Commun. Surv. Tut., IEEE 16 (1) (2014) 556–578.
[2] S. Al-Sultan, M.M. Al-Doori, A.H. Al-Bayatti, H. Zedan, A comprehensive survey on vehicular ad hoc network, J. Netw. Comput. Appl. 37 (2014) 380–392.
[3] T.H. Silva, P.V. Melo, J. Almeida, A. Viana, J. Salles, A. Loureiro, Participatory sensor networks as sensing layers, in: Proceedings of the IEEE Conf. on Social Computing and Networking (SocialCom'14), 2014.
[4] K. Yang, X. Cheng, L. Hu, J. Zhang, Mobile social networks: state-of-the-art and a new vision, Int. J. Commun. Syst. 25 (10) (2012) 1245–1259.
[5] T. Silva, P. Vaz De Melo, J. Almeida, A. Loureiro, Large-scale study of city dynamics and urban social behavior using participatory sensing, Wireless Commun., IEEE 21 (1) (2014) 42–51.
[6] V. Moosavi, L. Hovestadt, Modeling urban traffic dynamics in coexistence with urban data streams, in: Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, ACM, 2013.
[7] T. Hossmann, T. Spyropoulos, F. Legendre, Putting contacts into context: Mobility modeling beyond inter-contact times, in: Proceedings of the Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing, ACM, 2011.
[8] X. Hu, T. Chu, V. Leung, E. Ngai, P. Kruchten, H. Chan, A survey on mobile social networks: applications, platforms, system architectures, and future research directions (2015).
[9] P. Bellavista, R. Montanari, S.K. Das, Mobile social networking middleware: a survey, Pervasive Mobile Comput. 9 (4) (2013) 437–453.
[10] D. Schuster, A. Rosi, M. Mamei, T. Springer, M. Endler, F. Zambonelli, Pervasive social context: taxonomy and survey, ACM Trans. Intell. Syst. Technol. (TIST) (2013).
[11] W. Moreira, P. Mendes, Social-aware opportunistic routing: the new trend, in: Routing in Opportunistic Networks, Springer, 2013, pp. 27–68.
[12] R.H. Rawung, A.G. Putrada, Cyber physical system: paper survey, in: ICT For Smart Society (ICISS), 2014 International Conference on, IEEE, 2014, pp. 273–278.
[13] T. Spyropoulos, K. Psounis, C.S. Raghavendra, Spray and wait: an efficient routing scheme for intermittently connected mobile networks, in: Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking, ACM, 2005, pp. 252–259.
[14] T. Spyropoulos, K. Psounis, C.S. Raghavendra, Spray and focus: Efficient mobility-assisted routing for heterogeneous and correlated mobility, in: Pervasive Computing and Communications Workshops, 2007. PerCom Workshops'07. Fifth Annual IEEE International Conference on, IEEE, 2007, pp. 79–85.
[15] A. Mtibaa, M. May, M. Ammar, C. Diot, Peoplerank: combining social and contact information for opportunistic forwarding, in: Proceedings of INFOCOM, San Diego, USA (March 2010), 2012.
[16] P. Hui, J. Crowcroft, How small labels create big improvements, in: Pervasive Computing and Communications Workshops, 2007. PerCom Workshops'07. Fifth Annual IEEE International Conference on, IEEE, 2007, pp. 65–70.
[17] H. Chen, W. Lou, Contact expectation based routing for delay tolerant networks, Ad Hoc Netw. 36 (2016) 244–257.

Please cite this article as: K. Machado et al., Pervasive forwarding mechanism for mobile social networks, Computer Networks (2016), http://dx.doi.org/10.1016/j.comnet.2016.08.022

[18] P. Costa, C. Mascolo, M. Musolesi, G.P. Picco, Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks, Sel. Areas Commun., IEEE J. 26 (5) (2008) 748–760.

[19] P. Hui, J. Crowcroft, E. Yoneki, Bubble rap: social-based forwarding in delay–tolerant networks, Mobile Comput., IEEE Trans. 10 (11) (2011) 1576–1589.

[20] W. Moreira, P. Mendes, Impact of human behavior on social opportunistic forwarding, Ad Hoc Netw. 25 (2015) 293–302.

[21] A. Keränen, J. Ott, T. Kärkkäinen, The one simulator for dtn protocol evaluation, in: Proceedings of the 2nd international conference on simulation tools and techniques, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009, p. 55.

[22] S. Brin, L. Page, Reprint of: The anatomy of a large-scale hypertextual web search engine, Comput. Netw. 56 (18) (2012) 3825–3833.

[23] R.-I. Ciobanu, R.-C. Marin, C. Dobre, V. Cristea, C.X. Mavromoustakis, Onside: Socially-aware and interest-based dissemination in opportunistic networks, in: 2014 IEEE Network Operations and Management Symposium (NOMS), IEEE, 2014, pp. 1–6.

[24] A. Socievole, E. Yoneki, F. De Rango, J. Crowcroft, Ml-sor: Message routing using multi-layer social networks in opportunistic communications, Comput. Netw. 81 (2015) 201–219.

[25] P. Lamberson, S.E. Page, et al., Tipping points, Q. J. Polit. Sci. 7 (2) (2012) 175–208.

[26] S. Bakhshi, P. Kanuparthy, E. Gilbert, Demographics, weather and online reviews: a study of restaurant recommendations, in: Proceedings of the 23rd international conference on World wide web, ACM, 2014, pp. 443–454.

[27] S. Bannur, O. Alonso, Analyzing temporal characteristics of check-in data, in: Proceedings of the companion publication of the 23rd international conference on World wide web companion, International World Wide Web Conferences Steering Committee, 2014, pp. 827–832.

[28] E. Cho, S.A. Myers, J. Leskovec, Friendship and mobility: user movement in location-based social networks, in: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2011, pp. 1082–1090.

[29] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, A. Chaintreau, CRAWDAD dataset cambridge/haggle (v. 2009-05-29), 2009, (Downloaded from http://crawdad.org/cambridge/haggle/20090529). 10.15783/C70011.

[30] A.-K. Pietilainen, C. Diot, CRAWDAD dataset thlab/sigcomm2009 (v. 2012-07-15), 2012, (Downloaded from http://crawdad.org/thlab/sigcomm2009/20120715). 10.15783/C70P42.

[31] N. Eagle, A.S. Pentland, CRAWDAD dataset mit/reality (v. 2005-07-01), 2005, (Downloaded from http://crawdad.org/mit/reality/20050701). 10.15783/C71S31.

[32] T.H. Silva, P.O.V. de Melo, A.C. Viana, J.M. Almeida, J. Salles, A.A. Loureiro, Traffic condition is more than colored lines on a map: Characterization of waze alerts, Social Informatics, Springer, 2013.

[33] A.I.J. Tostes, T.H. Silva, F. Duarte-Figueiredo, A.A. Loureiro, Studying traffic conditions by analyzing foursquare and instagram data (2014).

[34] A.-K. Pietiläinen, E. Oliver, J. LeBrun, G. Varghese, C. Diot, Mobiclique: middleware for mobile social networking, in: Proceedings of the 2nd ACM workshop on Online social networks, ACM, 2009, pp. 49–54.

[35] M. Scheffer, J. Bascompte, W.A. Brock, V. Brovkin, S.R. Carpenter, V. Dakos, H. Held, E.H. Van Nes, M. Rietkerk, G. Sugihara, Early-warning signals for critical transitions, Nature 461 (7260) (2009) 53–59.

Kassio L. S. Machado is a Computer Engineer (2010) with Masters degree (2012), and currently a Ph.D. candidate at the Federal University of Minas Gerais, Brazil. He is currently visiting the PARADISE Research Laboratory at the University of Ottawa, Canada, where is conducting research in the area of Wireless Networks. His publications and interests include Pervasive approaches and mechanisms for People and Content Centric Networks, Opportunistic Networks and Internet-of-Things.

Azzedine Boukerche is a Full Professor and holds a Canada Research Chair position at the University of Ottawa. He is the founding director of PARADISE Research Laboratory at Ottawa and NSERC-DIVA Research of Excellence Centre. He is a Fellow of the Canadian Academy of Engineering. His current research interests include wireless ad hoc and sensor networks, wireless networks, mobile and pervasive computing, wireless multimedia, QoS service provisioning, performance evaluation and modeling of large-scale distributed systems, distributed computing, large-scale distributed interactive simulation, and parallel discrete event simulation. He is a holder of an Ontario Early Research Excellence Award (previously known as Premier of Ontario Research Excellence Award), Ontario Distinguished Researcher Award, and Glinski Research Excellence Award.

Pedro O. S. Vaz de Melo is an assistant professor in the Computer Science Department (DCC) of Federal University of Minas Gerais (UFMG). He has degree (2003) and Masters (2007) in Computer Science from the Pontifical Catholic University of Minas Gerais (2003). He got his Ph.D. at Federal University of Minas Gerais (UFMG) with a one year period as a visiting researcher in Carnegie Mellon University and a five months period as a visiting researcher at INRIA Lyon. His research interest is mostly focused on knowledge discovery and data mining in complex and distributed systems.

Eduardo C. Cerqueira received Master in Computer Science at Federal University of Santa Catarina, Brazil (2003) and his PhD in Informatics Engineering from the University of Coimbra, Portugal (2008). He was an invited auxiliary professor at the Department of Informatics Engineering of the University of Coimbra (2008–2009). He is an associate professor at the Faculty of Computer Engineering of the UFPA in Brazil and now researcher at Network Research Las at UCLA/USA and Centre for Informatics and Systems of the University of Coimbra (CISUC)/Portugal. His publications include 5 edited books, 5 book chapters, 4 patents, 1 IETF Internet Draft and over than 140 papers in national/international refereed journals/conferences. He is involved in the organization of several international conferences and workshops, including Future Multimedia Networking (IEEE FMN), Future Human-centric Multimedia Networking (ACM FhMN), ICST Conference on Communications Infrastructure, Systems and Applications in Europe (EuropeComm), Latin America Conference on Communications (IEEE LATINCOM) and Latin American Conference on Networking (IFIP/ACM LANC). He has been serving as a Guest Editor for 5 special issues of various peer-reviewed scholarly journals. His research involves Multimedia, Future Internet, Quality of Experience, Mobility and Ubiquitous Computing.

Antonio A. F. Loureiro received his B.Sc. and M.Sc. degrees in computer science from the Federal University of Minas Gerais (UFMG), Brazil, and the Ph.D. degree in computer science from the University of British Columbia, Canada. Currently, he is a Full Professor of computer science at UFMG, where he leads the research group in wireless sensor networks and ubiquitous computing. His main research areas are wireless sensor networks, urban sensing, ubiquitous computing, and distributed algorithms. In the last 15 years he has published extensively in international conferences and journals related to those areas, and also presented tutorials and keynote talks at international conferences.