An Autonomous Reputation Framework for Physical Locations based on WiFi Signals

Afra Mashhadi*, Geert Vanderhulst, Utku Günay Acer, Fahim Kawsar
Bell Laboratories
Alcatel Lucent
Antwerp, Belgium
*Dublin, Ireland
firstname(_middlename).surname@alcatel-lucent.com

ABSTRACT

Online reviews are used on the large scale to assess the quality and reputation of urban venues like hotels, restaurants, museums, etc. However, contributing reviews requires manual effort in the digital world, undertaken by only a small fraction of a venue's visitors. In this position paper, we present a framework that automatically assigns an offline reputation score by only relying on the physical presence of a user at a venue. In our approach, we passively capture the list of preferred WiFi networks (PNL) radiating from users smartphone as part of *WiFi Probe requests* in order to anonymously detect similar and recurrent users and to derive a personalised reputation score for an urban venue. By leveraging these ubiquitous WiFi radio signals, we seek to gather participation from a much broader set of visitors than online contributors. In this position paper, we outline our scoring technique, an early prototype architecture and discuss the potential of the proposed framework.

Keywords

Spatial Reputation Score; WiFi Probes; Preferred Network List

Categories and Subject Descriptors

C.2 [**Computer-communication networks**]: Network Architecture and Design

1. INTRODUCTION

As the urban cities grow and expand, the number of venues and Points of Interests (PoIs) in the cities and services that are available to the citizens is also growing rapidly. However, our source of information about the quality and reputation of these PoIs and venues has been confined to online resources such as TripAdvisor[1] and Yelp[2].

Although these sources provide valuable and often very thorough information regarding places and venues, their coupling of physical places' information to the online world imposes various limitations. First, the reviews and scores of these systems are reflective of the opinion of only those customers who have contributed their feedback to these websites, as opposed to the bigger population sample of everyone who visited the place. Secondly, in gathering these feedbacks the social relation between the visitors is not taken into account. That is the user is not aware whether the review score of a place was contributed by her social relations (e.g., her friends or family) or people similar to her. Finally, by design, the usage of online reviews and reputation scores imposes the user to pre-plan and seek this information in advance to visiting a place, taking away the spontaneous facet of discovering new places.

In this paper we overcome these limitations by introducing an offline spatial scoring mechanism, where information from everyone who has visited a venue is anonymously captured and used to derive a *spatial score* - quantification of the reputation of an urban venue based on physical visit patterns of similar and socially intimate users. In so doing we propose a framework which relies on non-obtrusive radio signals originating from the crowds who have visited a place in the past without the need for any application or direct participation. These radio signals, namely WiFi probe requests, contain a list of previously connected Access Points (APs) for each user, referred to as the Preferred Network List (PNL). We then ground our work over the theories of homophily [7] and argue that similar people enjoy similar activities and therefore are more likely to exhibit common APs in their PNL. For example, family members and partners are likely to have a very similar PNL, whereas colleagues would exhibit less commonality in the previously connected APs (i.e., PNLs). However, the homophily is not limited only to the friendship and social relationships that are known to us face-to-face but it also includes the *familiar strangers*, those who we regularly observe in different urban spaces we inhabit without directly interacting with them [8]. Leveraging the captured PNLs, we define a *spatial score* as a personalised similarity score between the visitor and all those who visited the PoI in the past.

Our proposed framework offers an alternative to the existing online spatial reviewing resources, and decouples the PoIs' information from the current Web 2.0 platform. By doing so, we lift the above three obstacles. More specifically, our system leverages input from *everyone* who visited the place carrying a WiFi enabled device to derive this spatial score, thus taking into account a much bigger population sample than the online contributors. Our approach also allows us to discover social relations of those who previously visited the place, enabling us to calculate the spatial score in a personalised way. Furthermore, our framework captures PNLs and communicates back the spatial score to the user without the need for the users device to be connected to the Internet at any

[1] `http://www.tripadvisor.com/`
[2] `http://www.yelp.com/`

WPA'15, May 22, 2015, Florence, Italy.

ACM 978-1-4503-3498-3/15/05 ...$15.00.
http://dx.doi.org/10.1145/2753497.2753498.

point. In this position paper we describe the spatial score, an early architecture and evaluation plan of the envisioned framework.

2. BACKGROUND

We start by offering a primer on the dynamics of WiFi management frames, that provides the foundation of our framework. We then position our work against past research on WiFi probe based service design.

2.1 Probes and Preferred Network Lists

Apart from data packets, mobile device exchanges overhead messages with APs to manage wireless connections. *Control Frames*, for example, regulate access to an AP and make sure that two devices do not transmit packets at the same time. In addition, devices and APs exchange *Management Frames* that are used for authentication and association with a wireless network.

In our work we are interested in a particular type of management frame, *Probe Request*, which devices use to actively discover available APs. If a probe request is populated with a Service Set Identifier (SSID) field, i.e., the name of the network the AP serves, then it is directed to a particular WiFi network. Otherwise, it is regarded as a broadcast probe request looking for any available network. The destination address and the Basic Service Set ID (BSSID, i.e., the MAC address of the AP) is always set to a broadcast address (i.e., "ff:ff:ff:ff:ff:ff"). Figure 1 shows an example of a directed probe that looks for a specific network called "Hilton NY".

A mobile device sends probe requests for every network it was connected to in the past with the corresponding SSID at the vendor-specific time intervals, depending on the power state of the device and the connection state of the WiFi chip[3] . This list of networks (SSIDs) is denoted as the user's Preferred Network List (PNL). If an AP receives a probe request that is either broadcasted or contains its SSID, it replies with a *Probe Response* frame to inform the device about its proximity. If the SSID of the AP is in the PNL of the device, the device and the AP exchange the authentication and association frames to connect the device to the AP. Unlike a probe request, a probe response is always directed to a specific device. In a probe response, the source address and BSSID are set to the MAC address of the AP whereas the destination address corresponds to the MAC address of the device from which the probe request was received.

Since probe requests are not encrypted, they can easily be intercepted even if they are directed to a particular SSID. In our framework, we capture probe requests originating from users' devices (as a proxy of visitors of a place), and reconstruct PNLs from them (as a proxy of a visitor's familiar places) allowing us to detect homophily. That is, we derive a similarity score from a PNL by comparing it with previously captured PNLs. The result is later communicated back to the user in a probe response. Since this response is directed to a specific device, we can add in personalised information, which the device OS or specific apps can interpret and visualise accordingly.

2.2 Related Work

WiFi probes have been used in the past for various purposes including user tracking [5, 9, 10, 12] and discovering users' social relations [2, 4]. The Preferred Network List of the probes has been exploited by Barbera et al. [2] to discover the social relations between individuals through modelling similarity as a graph. The Adamic-Adar similarity was used to smooth out the influence of frequently used Access Points, through discounting the importance of the APs logarithmically. The authors also studied the other aspects of homophily such as device type and the language of those socially linked users. Similarly, Cunche et al. [4] used PNLs to decide whether two devices potentially belong to socially linked users. While these works focus on detecting and inferring the users' social relationships, we aim to take a rather different perspective and focus on profiling places rather than users. In so doing, we exploit the information provided through the probes of the visitors to derive a spatial score for each place.

[3]When not connected to a WiFi network, mobile devices typically send probe requests every 15 to 60 seconds.

DA	SA	BSSID	SSID
ff:ff:ff:ff:ff:ff	90:68:c3:be:34:9f	ff:ff:ff:ff:ff:ff	HILTON NY

Figure 1: A (partial) Probe Request originating from a device with MAC address "90:68:c3:be:34:9f" that is looking for an AP with SSID "HILTON NY".

3. ARCHITECTURE

In our architecture, we rely on existing wireless networking infrastructure which we augment with two small software modules as depicted in Figure 2: a module responsible for capturing and storing PNLs from WiFi-enabled user devices and a module that computes personalised spatial scores based on these PNLs. For the sake of simplicity, the figure shows our architecture for a place with a single AP. However, it should be noted that it can easily scale up to places with multiple APs by moving shared storage facilities and the spatial scoring module into the cloud. With flexible firmwares like OpenWRT[4], modifications to legacy APs become feasible such that we can leverage open existing hardware. Moreover, we envision that network-centric sensing becomes more and more adopted in future small cells, in order to offer a plethora of location-based values such as place recommendations to users within signal range.

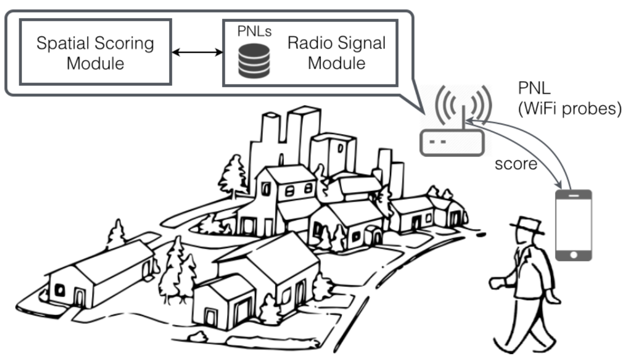

Figure 2: Software modules deployed on urban WiFi APs engaging in a dialog with a user's personal device.

3.1 Radio Signal Module

We use the WiFi interface of an AP to listen to probe requests directed to any AP on a fixed channel. By doing so, we extract two sets of information from the broadcasting device, that is the PNL list of the device, as well as information regarding the presence of the device in the current location. To obtain the PNL of a device, we aggregate captured probe requests based on their source MAC address (which identifies a single device) over a pre-defined time

[4]`http://openwrt.org/`

interval. While we can rely on the MAC address of the device to remain consistent over a short transmission time (e.g., probe requests), we cannot assume the same consistency during the longer time intervals (e.g., hours, days) due to MAC randomisation strategies that are put in place to protect user's privacy. To address this problem, we use the PNL itself as an identification of the device, assuming that the PNL is unique to each individual and along with additional sensing information (e.g., frequency of probing) can serve as the individual's identifier [11] replacing the dependency on the MAC address.

In addition to serving as an identifier, the PNL also presents the previous places that the person had visited and connected to their WiFi APS. However it falls shorts from capturing those places which the user has visited in the past *without* connecting to the WiFi available in the place. To account for these cases, we extract another set of information regarding the device's presence in a venue. We do so by monitoring the time of the first probe and last probe seen from the device, allowing us to infer the duration that the device stayed in the venue. Based on this information we construct a *virtual* list similar to the PNL, which we store in the cloud database. This list contains the device identifier, the SSID along with the BSSID of the APs (that receives the probe request) as well as any additional pieces of information such as frequency or duration of visits, which are not available on the device-maintained PNLs but can be sensed and monitored from the infrastructure. By capturing the duration information, we can filter the short visits (e.g., seconds) that may correspond to a person passing by but not visiting the place. The frequency on the other hand can give us an indication of the topophilia [13] or the attitude and the perception of the user towards the place, that is if the person feels the "sense of belonging" they are more likely to visit the place again. Finally, we construct a list of visited places by extracting the SSIDs out of both the captured device-maintained PNLs as well as the virtually-constructed PNLs, which will later be used to infer the homophily amongst users.

Finally, this module is also responsible for communicating the inferred spatial score back to the user. We do so by relying on "beacon stuffing" [3]. In this approach, extra information is encapsulated in WiFi management frames such that no active WiFi connection needs to be established with the AP (for which a user might not have proper credentials). For instance, a device can ping an AP with a directed probe request which is replied by the AP with a directed probe response in which the inferred spatial score could be embedded in (e.g., in vendor-specific fields that are reserved in WiFi management frames). At the device end, the spatial score can then be extracted from a received probe response and delegated to a notification service (as part of the mobile OS) to which local applications can subscribe and hence leverage the spatial scores as location-based context information.

3.2 Spatial Scoring Module

In order to calculate a personalised spatial score of a PoI for a user we first construct a vector v which includes the captured PNLs from the user as well as the virtually-constructed PNLs. We refer to this union vector simply as *PNL* henceforth. We then compute a pairwise comparison of this vector with all the previously stored PNLs. As each PNL corresponds to a vector (v) of previously connected APs, the pairwise comparison essentially captures whether an element of vector v is also observed in any other PNL (regardless of its position in the vector). However, not all the commonly seen APs carry the same importance. Consider for example a common AP match with SSID "Vodafone Hotspot". The fact that it was seen by two persons in the past, tells us very little about the homophily between them as the widely distributed "Vodaphone Hotspot" APs do not uniquely identify a place. Therefore, we require to give a weight to the elements of a PNL list: private APs such as those at home or at the office are more unique and hence deserve a higher weight. We use TF-IDF [6] to associate a weight w_u to each AP based on its *uniqueness*.

As well as accounting for the uniqueness of the APs, we need to consider the *frequency* of individuals visiting a given PoI, as it gives us an indication of the users interest to (approval of) a place. However, this information is not stored on the mobile device or inside WiFi probes, as the PNL only specifies the previously connected APs regardless of how many times a device has been connected to that AP. We account for this reoccurrence weight by keeping a simple counter (as explained in the previous section), representing the number of visits from a given device to a given PoI. We normalise this counter per user and refer to this normalised weight as *recurrence* weight and denote it as w_r.

Finally, the weight w brings together the uniqueness and recurrence weight for each previously connected AP as: $w = \gamma * w_u + \beta * w_r$, where γ and β are constants. We then formulate our spatial score as an average weighted similarity score, w_i if the i^{th} member of the PNL is also observed in v' as follows:

$$f(v, v') = \frac{\sum_{i=1}^{n} w_i}{|v|}$$

where n denotes the number of common APs between v and v' and $|v|$ corresponds to the number of *all* previously connected APs of the query device. This metric has two desirable properties. Firstly, it can capture whether others similar to the user visited the same place. We define these relations in terms of three categories of *self*, *friends* and *familiar strangers* and detect them based on setting a threshold for the similarity score. Secondly, this metric can capture the ageing effect, that is if a user visited the same place in the distant past her similarity score would decay over time as her list of preferred connections (PNL) would have also changed over time. Thus allowing us to discount the spatial score for those places that the user has stopped visiting.

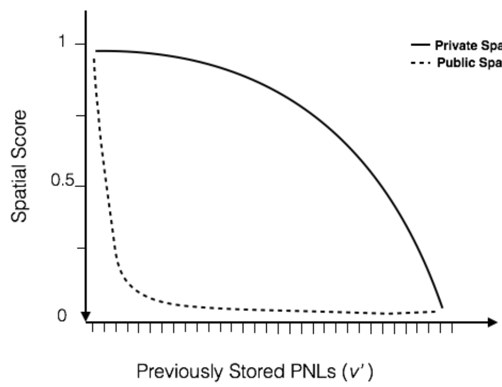

Figure 3: The spatial score distribution for a user at public versus private spaces.

As the result of calculating this score across all v's that are seen in the same venue, we can build a distribution for the user for the specific place. We refer to this distribution as the spatial score distribution. We claim this distribution would take two different shapes depending on the type of the spatial venue. For the private spaces where individuals can inhabit a space by being part of a specific social circle, this distribution would resemble the integral of

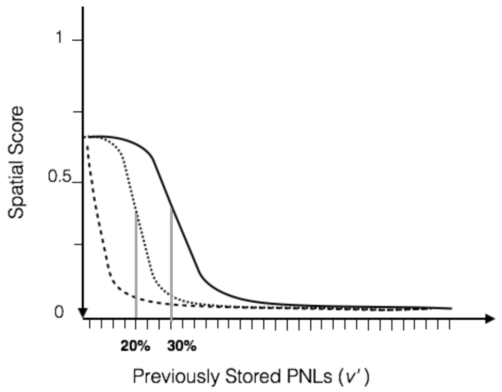

Figure 4: The spatial score modeled as Pareto distribution.

the standard normal distribution as depicted in Figure 3 where the x-axis presents each of the previously seen PNLs (ranked v') and y-axis presents the computed similarity $f(v, v')$. The expected distribution shape of the private places is therefore due to the larger number of similar users (i.e., pairwise similarity comparison). However, for public spaces such as PoIs common to urban cities we expect a long tail distribution (the dashed line in Figure 3) to be observed as it reflects that the place is visited by many strangers. In this paper we are interested in the common public urban venues and thus expect to observe the latter type of the distribution.

However, depending on the similarity of the past visitors the long-tail distribution is expected to exhibit different properties. For example, the majority of the spatial scores (that is the top 80%) could have resulted from only 20% of the past visitors, resulting in a Pareto distribution, or coming from even a lesser percentage of the past visitors. Figure 4 illustrates variations of the spatial score (long tail) distribution for three cases. While the dotted line presents the Pareto principle of 80-20, the dashed line presents the scenarios where the high spatial similarity score is only contributed by a few. The solid line on the other hand presents the cases where the top spatial score is contributed by 30% of the past visitors, thus indicating an increased likelihood of the users' homophily.

We can capture these differences in the distributions through the shape parameter α of the distribution. The smaller the α, the more percentage of the past visitors were found to be similar with the current user. As α grows bigger, the power law property of the distribution becomes more announced corresponding to the cases where very little similarity with all the past visitors exists. The final result for each user for a given place is thus a combination of the mean (μ) of the distribution and the shape (α) parameter, which is communicated to the user's device through probe response. This personalised reputation score can then be presented to the user in a semantically rich way for example through icons.

4. CONCLUDING REMARKS

In this position paper, we propose a spatial scoring technique together with an early prototype architecture to quantify the reputation of an urban venue using everyday WiFi radio signals that emit from users smart phones. As a future avenue of this work, we are currently implementing our framework for a real world evaluation. Our implementation is based on Meshlium devices from Libelium [1] which have all the properties of a real WiFi AP including the ability to easily program them. A Meshlium features a 500 Mhz x86 processor, two WiFi interfaces (one of which acts as an AP by default) and runs an embedded Debian Linux operating system. Furthermore, in order to communicate and visualise the spatial score to the user, we are currently developing an Android application. This application primarily acts as a widget and displays the received spatial score in different colours with semantically rich labelling for easy user interpretation. We plan to deploy this framework and its accompanied application starting in spring 2015, with a pool of 50 users. To avoid unmanageable data sparsity, users will be selected so that their mobility mainly covers a restricted area in the city, with good overlaps (e.g., university staff and students). We will assess the usability of our system through surveys and structured interviews.

5. REFERENCES

[1] `http://www.libelium.com/`.

[2] Marco V Barbera, Alessandro Epasto, Alessandro Mei, Vasile C Perta, and Julinda Stefa. Signals from the crowd: uncovering social relationships through smartphone probes. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 265–276. ACM, 2013.

[3] Ranveer Chandra, Jitendra Padhye, Lenin Ravindranath, and Alec Wolman. Beacon-stuffing: Wi-fi without associations. In *HOTMOBILE'07*, 2007.

[4] Mathieu Cunche, Mohamed Ali Kaafar, and Roksana Boreli. I know who you will meet this evening! linking wireless devices using wi-fi probe requests. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2012 IEEE International Symposium on a*, pages 1–9. IEEE, 2012.

[5] Mathieu Cunche, Mohamed-Ali Kaafar, and Roksana Boreli. Linking wireless devices using information contained in wi-fi probe requests. *Pervasive and Mobile Computing*, 11:56–69, 2014.

[6] Karen Sparck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1), 1972.

[7] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.

[8] Stanley Milgram, John Ed Sabini, and Maury Ed Silver. *The individual in a social world: Essays and experiments* . Mcgraw-Hill Book Company, 1992.

[9] Wicher Minnaard. Out of sight, but not out of mind: Traces of nearby devices' wireless transmissions in volatile memory. *Digital Investigation*, 11:S104–S111, 2014.

[10] ABM Musa and Jakob Eriksson. Tracking unmodified smartphones using wi-fi monitors. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, pages 281–294. ACM, 2012.

[11] Jeffrey Pang, Ben Greenstein, Ramakrishna Gummadi, Srinivasan Seshan, and David Wetherall. 802.11 user fingerprinting. In *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, pages 99–110. ACM, 2007.

[12] Arvind Thiagarajan, James Biagioni, Tomas Gerlich, and Jakob Eriksson. Cooperative transit tracking using smart-phones. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, pages 85–98. ACM, 2010.

[13] Yi-Fu Tuan. *Topophilia: A study of environmental perceptions, attitudes, and values*. Columbia University Press, 2013.