Cyber-Physical Social Networks

CHRISTIAN VON DER WETH, ASHRAF M. ABDUL, and MOHAN KANKANHALLI,
National University of Singapore

In the offline world, getting to know new people is heavily influenced by people's physical context, that is, their current geolocation. People meet in classes, bars, clubs, public transport, and so on. In contrast, first-generation online social networks such as Facebook or Google+ do not consider users' context and thus mainly reflect real-world relationships (e.g., family, friends, colleagues). Location-based social networks, or second-generation social networks, such as Foursquare or Facebook Places, take the physical location of users into account to find new friends. However, with the increasing number and wide range of popular platforms and services on the Web, people spend a considerable time moving through the online worlds. In this article, we introduce cyber-physical social networks (CPSN) as the third generation of online social networks. Beside their physical locations, CPSN consider also users' *virtual locations* for connecting to new friends. In a nutshell, we regard a web page as a place where people can meet and interact. The intuition is that a web page is a good indicator for a user's current interest, likings, or information needs. Moreover, we link virtual and physical locations, allowing for users to socialize across the online and offline world. Our main contributions focus on the two fundamental tasks of creating meaningful virtual locations as well as creating meaningful links between virtual and physical locations, where "meaningful" depends on the application scenario. To this end, we present OneSpace, our prototypical implementation of a cyber-physical social network. OneSpace provides a live and social recommendation service for touristic venues (e.g., hotels, restaurants, attractions). It allows mobile users close to a venue and web users browsing online content about the venue to connect and interact in an ad hoc manner. Connecting users based on their shared virtual and physical locations gives way to a plethora of novel use cases for social computing, as we will illustrate. We evaluate our proposed methods for constructing and linking locations and present the results of a first user study investigating the potential impact of cyber-physical social networks.

CCS Concepts: • **Human-centered computing** → **Social networks**; • **Information systems** → *Web searching and information discovery*; • **Information systems** → Information retrieval

Additional Key Words and Phrases: Cyber-physical social networks, ad-hoc socializing, network creation, data linking, location-based services, social computing

ACM Reference Format:
Christian von der Weth, Ashraf M. Abdul, and Mohan Kankanhalli. 2017. Cyber-physical social networks. ACM Trans. Internet Technol. 17, 2, Article 17 (March 2017), 25 pages.
DOI: http://dx.doi.org/10.1145/2996186

1. INTRODUCTION

17

Socializing is one of the dominant human activities in the offline as well as in the online world [Poushter et al. 2015]. Websites and platforms that enable social computing are

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore, under its IRC@SG Funding Initiative and administered by the Interactive and Digital Media Programme Office (IDMPO).
Authors' addresses: C. von der Weth, A. M. Abdul, and M. Kankanhalli, SeSaMe Centre, Interactive & Digital Media Institute, National University of Singapore, I3 Building, IDMI HQ, #02-02-02, 21 Heng Mui Keng Terrace, Singapore 119613; emails: vonderweth@nus.edu.sg, aabdul@u.nus.edu, mohan@comp.nus.edu.sg.

DOI: http://dx.doi.org/10.1145/2996186

among the most popular online services on the Web. Social networking sites[1] such as Facebook, Google+, LinkedIn, and Twitter allow users to create contacts with others for sharing messages, images, or other information. Groupon and LivingSocial[2] are two popular examples for platforms where users buy goods or services together to drive down prices. Online forums or question-and-answer (Q&A) sites[3] (e.g., Yahoo! Answers, Quora) allow users to discuss and to share typically topic-specific information with each other. Other services that rely and benefit heavily from user participation and/or user-generated content are recommendation sites[4] (e.g., Tripadvisor, Yelp), social new sites[5] (e.g., Reddit, VOAT), and social bookmarking sites[6] (e.g., Delicious, StumbleUpon).

An important aspect of such platforms is how users first get in touch with others. In the offline world, making contact with new friends and networking is often heavily influenced by people's physical context, particularly their current locations. Interactions between people without prior relationships often happen because of shared locations. Everyday examples are people visiting the same museum, exhibition, concert, or sports event or people sharing the same flight or train ride.

Regarding the consideration of users' context and their evolution over time, we can identify different generations of social networking platforms. *First-generation social networks* such as Facebook, Google+, and LinkedIn do not—or initially did not—take users' current context into account. Here, contacts either derive from real-world relationships (family, friends, colleagues etc.) or through explicit seeking and joining behavior on the part of the user. Studies showed that users have only a very small percentage (<10% in all studies) of contacts they have never met offline [Lampe et al. 2006; Hampton et al. 2011; Tang et al. 2016]. This includes the fact that connections that have been exclusively forged online derive from long-term shared interests (e.g., active members in the same online forum or gaming community). *Second-generation social networks, or location-based social networks (LBSN)* consider users' physical context, that is, their current geolocations [Lindqvist et al. 2011; Chen et al. 2013; Bao et al. 2015]. Popular platforms such as Foursquare or Facebook Places[7] allow users to not only share their current locations and activities with their existing contacts but also to find new contacts based on shared locations. Services such as Meetup, Skout, or Tinder[8] focus even more on making new friends or finding dates based on users' geolocations [Xue et al. 2016; Sumter et al. 2017]. However, given the opportunities of global communication, socializing, and collaboration, connecting with others solely due to geographic proximity is rather limiting.

In this article, we now introduce *third-generation social networks* or *cyber-physical social networks*. As a unique feature, cyber-physical social networks (CPSN) take both the physical context as well as the virtual context of users into account. Similarly to location-based social networks, the physical context refers to users' geolocations. Adopting this notion to the Web, we describe the virtual context of users by means of their current *virtual locations* [von der Weth and Hauswirth 2013]. In a nutshell, a virtual location is an individual web page or a set of pages. The underlying intuition is that the web page a user is visiting is a good indicator for the user's current interests, likes, or information needs. Analogously to meeting like-minded people in the physical/offline

[1] https://www.facebook.com/, https://plus.google.com/, https://www.linkedin.com/, https://twitter.com/.
[2] https://www.groupon.com/, https://www.livingsocial.com/.
[3] https://answers.yahoo.com/, https://www.quora.com/.
[4] https://www.tripadvisor.com/, http://www.yelp.com/.
[5] https://www.reddit.com/, https://voat.co/.
[6] http://del.icio.us/, http://www.stumbleupon.com/.
[7] https://foursquare.com/, https://www.facebook.com/places/.
[8] http://www.meetup.com/, https://www.gotinder.com/, http://www.skout.com/.

world, we consider the Web a space not only in which users can navigate but also where users can “meet.” For example, two users watching the same documentary or tutorial on YouTube share common interests, potentially benefiting from interacting with each other, like sharing opinions or experiences. While anectodal evidence indicates that Facebook suggests new friends based on visited and/or liked pages, these suggestions are outside the context of the actual page visits; that is, Facebook suggests new friends without specifying why they have been suggested (except in the case of mutual friends) and potentially long after the page visit. Apart from users’ virtual context, by linking related physical and virtual locations, we enable web users browsing the web and mobile users present in the physical world to connect, thus providing a novel approach to bridge the gap between the offline and online worlds. With the notion of CPSN, we do not aim to replace previous generations of social networks but rather extend or complement them by providing new ways to find interesting or useful connections.

Cyber-physical social networks rely on the notion of virtual locations. While the concept of a physical location is well defined, mapping this notion to the Web is not obvious. Naive solutions like grouping all URLs of a domain are meaningful for, for example, the websites of hotels, restaurants, or shops, but fail for sites with a more dispersed content like recommender or media sharing sites. Using the full Uniform Resource Locator (URL) as a determiner, however, is also in practice often not meaningful. First, different URLs can link to the same content. Second, for many domains, considering different pages as different locations is often unnecessarily fine grained. Thus, we first address the issue of identifying meaningful virtual locations in an automated manner. For our approach, we harness the “wisdom of crowds.” We collect and analyze links shared by users over social media sites. The rationale is that users typically share links that point to the websites’ *units of interest*—that is, users are more likely to share links to a review or a video than links to the start page of the recommender or video sharing site. Based on our analysis, we describe an algorithm that assigns URLs to virtual locations.

As a second challenge, the success and usefulness of CPSN depend on the links between locations, both virtual and physical. Which locations to connect with each other is typically application dependent. For this article, we consider a CPSN offering a live and social recommendation service about venues, including hotels, restaurants, or shops, that goes beyond the traditional channels such as venues’ official website or review sites (e.g., Tripadvisor). We accomplish this by using existing APIs and customized Named Entity Recognition (NER) techniques to connect information about venues stemming from different platforms and media formats. We currently link online resources about venues such as official websites, Tripadvisor reviews, YouTube[9] videos, Flickr[10] and Instagram[11] images, as well as Twitter messages containing the names of those venues. By linking related virtual and physical locations, we enable two novel types of services. First, we enhance users’ online experience by providing tailored information based on their current location. More importantly, we enable ad hoc socializing and networking through presence awareness and instant messaging between users based on shared locations, again both physical and virtual.

Last, to illustrate the benefits of cyber-physical social networks, we present OneSpace, our fully implemented prototype of a live and social recommendation service. Its backend comprises two main components. A data repository maintains the constructed virtual locations and the links between virtual and physical locations. We use an Extensible Messaging and Presence Protocol (XMPP) server to provide presence awareness and instant messaging for the support of socializing between users. OneSpace

[9]https://www.youtube.com/.
[10]https://www.flickr.com/.
[11]https://www.instagram.com/.

users can assume two roles: online users sitting at home or at work using their desktop browser to navigate between sites and mobile users with their smart devices. To cater to both user roles, OneSpace provides two frontend applications. A browser add-on allows for the seamless integration into the browsing experience of web users. Mobile users can access the platform using our mobile application. The use cases for OneSpace particularly relate to collective intelligence, teamwork, collaboration, and entertainment. In our evaluation, we analyze our algorithm for the construction of virtual locations and the linkage between locations, as well as provide a first-user study regarding user behavior and the expected utility of OneSpace. Our results show that taking users' physical and virtual context into account enables novel types of services for social computing.

Section 2 outlines related work to put our work into context. Section 3 gives a brief overview to OneSpace. Section 4 formalizes the notion of virtual locations and presents our approach for their construction based on websites' units of interests. Section 5 covers the process of creating links between virtual and physical locations. Section 6 details the implementation of OneSpace, both backend architecture and frontend applications. Section 7 presents our evaluation results. Section 8 provides a discussion as well as a roadmap for future work. Section 9 concludes the article.

2. RELATED WORK

Socializing or social networking, being among the dominant human activities, have very successfully found their way into the online world. Websites that enable the interaction and socializing between users are among the most popular online services on the Web [Goel et al. 2012].

Online social networks. The popularity of online social network platforms such as Facebook, Google+, Twitter, and so on, has attracted a lot of research efforts from a variety of fields. Thus, we can only give a brief overview, focusing on landmark and survey articles. We distinguish three main areas: (1) *Structural analyses* investigate the structure and operation of online social networks. This includes the characterization of their topology [Ahn et al. 2007; Wu et al. 2011; Jiang et al. 2013]) and the creation of models to describe the success and evolution of networks [Tang et al. 2010; Xiang et al. 2010]. (2) *Social content analysis* looks at the content generated by users. Main topics include sentiment analyses and opinion mining [Feldman 2013; Hutto and Gilbert 2014], predicting the stock market, elections, and box-office revenues and the detection of trending topics and real-world events [Xie et al. 2013; Abdelhaq et al. 2013]. (3) *Information diffusion in online social networks* investigates how information (news, rumors, etc.) propagates across a population and has been particularly studied in the context of epidemiology and the spread of diseases [Asur and Huberman 2010; Harald et al. 2013]. (4) *Social interactions analysis*, among other sub-topics, covers general user behavior in online social networks [Bond et al. 2012; Jin et al. 2013; Lim et al. 2015], information diffusion [Guille et al. 2013; Bian et al. 2014; Taxidou and Fischer 2014], and the identification of influential users [Cha et al. 2010; Romero et al. 2011]. The long list of existing research works on online social networks highlight their importance and effects on society. However, as various studies have shown, most online connections reflect users' real-world relationships like family members, friends, or colleagues [Lampe et al. 2006; Hampton et al. 2011; Tang et al. 2016]. As such, first-generation online social networks, and hence related research efforts, do not reflect how people make new contacts in the offline world.

Location-based social networks. With the omnipresence of location-aware mobile devices and wireless Internet access, users' physical locations have become a new dimension in the context of social computing and online networking. It has fostered the development of so-called LBSN, where users' physical location is an integral part of the networking aspect. LBSN enable users to establish social connections with others and express their visits to places along with their social profiles. The services provided by LBSN, such as checking in at places, rating them, and commenting about them, are more sophisticated and user-centric as they also bring in their social context into consideration. Platforms such as Facebook Places or Foursquare are some of the most popular ones. Various efforts have analyzed user visits to places and the effect of social ties between users on user movement patterns [Cho et al. 2011; Krueger et al. 2014; Kysela et al. 2015]. Similarly, Backstrom et al. investigated the relationship between geography and friendship and showed that social ties of users can be used to discover approximate locations of users [Backstrom et al. 2010]. Follow-up works showed that not only people's locations but also their mobility patterns allow us to predict social ties between users [Wang et al. 2011; Scellato et al. 2011]. These research efforts show that there exists a strong link between social networks of users and their movements. However, these works focus on already-existing friendships between users. LBSN also facilitate the discovery of new contacts based on users' geolocations. Various studies have confirmed that most people spend most of their time at a rather small number of different places (home, work, leisure, etc.), making geolocations a meaningful parameter to recommend new friends [Lindqvist et al. 2011; Chen et al. 2013]. The success of LBSN in terms of finding new friends based on users' physical locations is further amplified by dedicated friend-finding and online dating platforms such as Skout, MeetMe, or Tinder. However, meeting new people solely due to geographic proximity is in many social computing scenarios unnecessarily restrictive.

Towards cyber-physical social networks. The advent of the Web 2.0, online fora, Q&A systems, and so on, has allowed users to meet and discuss different topics. However, the communication is asynchronous, that is, users typically have to wait for a reply. Thus, the degree of socializing is still rather limited, which is particularly pronounced for questions that need or benefit from a quick answer. But, more importantly, communities on such platforms form in a rather static fashion—users have to join an online forum or have to create and maintain a personal profile that reflects their interests and expertise—which does not adequately reflect the often changing and short-lived interests or information needs of a user. Considering arbitrary web pages as a browsing user's virtual context to enable ad hoc interaction between online users is not a new concept but has only been applied outside the scope of online social networks. The concept of virtual locations originates from the efforts towards collaboratively browsing and searching the Web. TeamSearch [Morris et al. 2006] and CoSearch [Amershi and Morris 2008] are systems providing mechanisms for co-located browsing, that is, where several users gather around one computer. SearchTogether [Morris and Horvitz 2007] and CoScripter [Leshed et al. 2008] enable collaborative browsing between users working with their own computers. They allow any group of users to initiate joint browsing on a website. These systems, however, focus on the collaboration between users with already-established relationships. That includes that users directly inform their contacts to initiate a collaborative browsing session. PlayByPlay [Wiltse and Nichols 2009] describes the need for collaborative browsing platforms to easily and efficiently browse the Web and demonstrates a system that lets the mobile device users and web users collaborate and communicate. In previous works, we first defined the concept of a virtual location [von der Weth and Hauswirth 2013] and motivated the potential benefits of presence awareness and ad hoc socializing across the virtual and physical world by analyzing the distribution of virtual locations in urban areas [von der Weth et al. 2014]. However, these works do not address the construction of virtual locations and the linking of virtual and physical locations, both being fundamental tasks for the development of cyber-physical social networks, and as such do not present a fully implemented prototype.

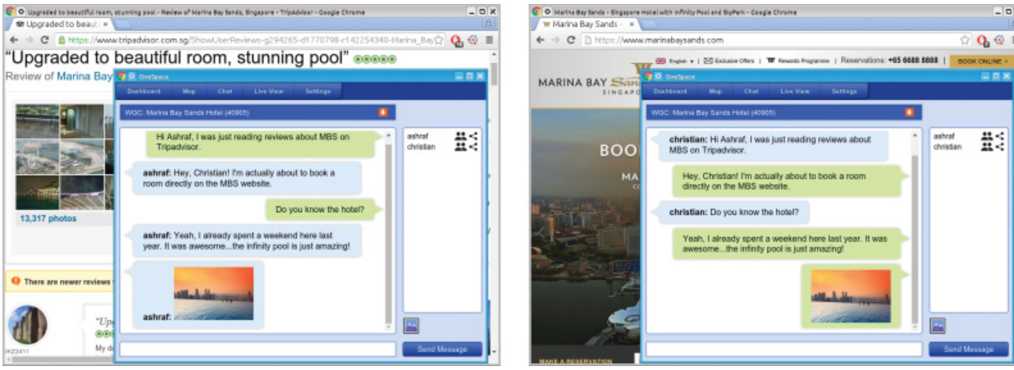


Fig. 1. Use Case I: Surfer Christian (left) is reading reviews about the Marina Bay Sands hotel on Tripadvisor. In the mean time, surfer Ashraf (right) is on the website of the hotel to book a room. Since we connect the website of venues with corresponding reviews, both surfers are in the same group chat, and Christian can get first-hand experiences from Ashraf.

Summing up, current second-generation online social networks consider users' geolocations as a new dimension for making new contacts. Extending this to users' virtual context towards the third generation of social networks—cyber-physical social networks—has been done only on a rather abstract level. First, while previous works motivate the concept of virtual locations and linking physical locations, they do not expand on the *how*. We propose methods to solve both tasks in an automated manner. Second, in contrast to the current state of the art, we present with OneSpace a fully implemented prototype of a cyber-physical social network.

3. ONESPACE: OUR PROTOTYPE OF CYBER-PHYSICAL SOCIAL NETWORK

To illustrate the concept and potential benefits of cyber-physical social networks, we give here a brief overview of OneSpace, our proof-of-concept implementation of a CPSN; Section 6 will detail more on its implementation. As an application scenario, OneSpace offers a live and social recommendation service where users can interact to discuss and share experiences about venues such as hotels, restaurants, or attractions in Singapore. For this, we collected a dataset of venues together their physical locations and connected each venue with related virtual locations. These virtual locations include a venue's official website, Tripadvisor reviews, related Flickr and Instagram images and YouTube videos but also tweets mentioning the venue.

OneSpace users can assume two roles, which we denote as follows: *surfers* are online users at home or at work browsing the Web using a desktop computer; *walkers* are mobile users that access the Web using their smart devices, which we assume can acquire their current geolocations. Given the different devices for surfers and walkers, we provide two application solutions for both roles. Figure 1 shows our browser add-on, illustrating the use case where two surfers visiting different but linked web pages can interact, that is, communicate, with each other. Walkers can download and install the OneSpace mobile app to join the network. As the example use case in Figure 2 illustrates, surfers can identify and connect to walkers on-site, and vice versa, based on their current virtual or physical locations. By linking related locations, apart from connecting users, OneSpace can also provide tailored, location-specific information. For example, a surfer browsing the website of a restaurant can use the add-on to view, for example, Instagram images of the restaurant (often including the food) or read tweets about the venue.

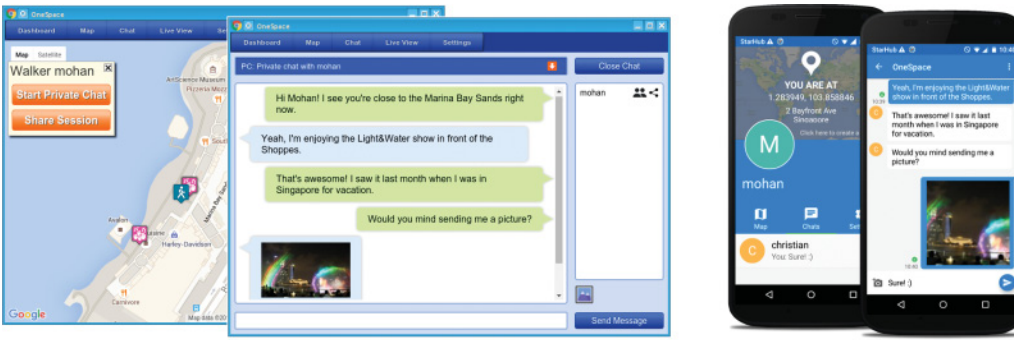


Fig. 2. Use Case II: Surfer Christian is using the map view of the browser add-on to search for users in his area of interest, which is around the Marina Bay Sands hotel in Singapore. Walker Mohan is close to the hotel and has the OneSpace mobile app installed. Christian can now contact Mohan, asking him for information or just to share impressions. While the hotel's website provides the usual information, for example, rates per night, facilities, and so on, Mohan can provide on-the-spot information that can be very valuable to Christian, for example, the current presence of a noisy construction site.

The way OneSpace uses the virtual or physical location of users for socializing illustrates how cyper-physical social networks support use cases and applications beyond the capabilities of first- and second-generation networks. With our current prototype, we focus on connecting like-minded users without prior relationships but based on their common interest, likings, or information needs derived from their shared locations. However, OneSpace can easily be extended by features common to traditional social networks such as keeping a list of permanent contacts, creating and maintaining a personal user profile, sending "offline" messages (i.e., outside group or private chats), enabling and updating privacy settings, and so on. We outline our roadmap from an application perspective for OneSpace as part of Section 8.

4. VIRTUAL LOCATIONS

To describe the virtual context of a user as his or her virtual location is a fundamental requirement for cyber-physical social networks. While the basic idea of considering web pages as virtual locations is rather intuitive, defining and automatically identifying meaningful virtual locations is not trivial.

4.1. Definition

In geographic terms, the most fine-grained way to specify someone's current position is by using geocoordinates. Mapping this concept to the virtual space, the current position of a user is the web page the user is browsing on, specified by a URL. However, deriving a virtual location from a single URL is limiting. First, syntactically different URLs can link to the same page. Second, often not a single page but a set of web pages are of interest to describe a user's location. We therefore extend the definition of a virtual location beyond a single URL.

Definition 1 (*Virtual Location*). A virtual location $loc^v = \{url_1, url_2, \ldots, url_n\}$ is a distinct, non-empty, and finite set of URLs, that is, $\forall loc^v_1, loc^v_2 : loc^v_1 \cap loc^v_2 = \emptyset$.

Which set of URLs form a virtual location is application dependent. On an abstract level, we consider virtual locations as the smallest *unit of interest* of a website. Examples include individual articles in an online newsarticles, as well as images or videos on a media sharing sites. The smallest unit of interest can also be the whole website. For example, we argue that distinguishing between different pages of a hotel or restaurant website is typically not meaningful. In the following section, we present our approach for the automatic construction of meaningful virtual locations.

4.2. Constructing Virtual Locations

Since the notion of a unit of interest differs from website to website and is rather subjective, we address this issue by harnessing the "wisdom of crowds" by collecting and analyzing links that users share over social media sites. The rationale is that users most of the time share links that point to a website's units of interest; that is, users are far more likely to share links to a news article or a video clip than to link to the root domain of the online newspaper or video sharing site. Given a set of shared links with the same domain, we identify which parts of the URLs, particularly which query string keys, are most important to specify the unit of interest.

URL path. Most websites organize their content hierarchically and identify it by means of a URL path structure. This is particularly true for sites that use commonly available Content Management Systems in their backend.

Example 1. nytimes.com uses the following two URL structures to link to their news articles:

```
www.nytimes.com/<YY>/<MM>/<DD>/<cat>/<article>.html
www.nytimes.com/<YY>/<MM>/<DD>/<cat>/<subcat>/<article>.html
```

where <YY>, <MM>, <DD> is the publication year, month and day, respectively. <cat> is the category (e.g., "politics" and "technology"), <subcat> is the optional subcategory, and <article> is the URL-encoded title of the article (e.g., "big-business-must-learn-to-love-apps").

Since the path depths of URLs pointing to a site's units of interests can vary, our goal is to simply catch the exceptional cases of URLs with a very uncommon path depth. In the following, let L_{dom} be the set of shared URLs with domain *dom* and $L_{dom}(pd)$ the set of URLs with domain *dom* and a path with depth *pd*. We then can define

$$freq_{dom}^{path}(pd) = \frac{|L_{dom}(pd)|}{|L_{dom}|} \tag{1}$$

with $freq_{dom}^{path}(pd) \in [0, 1]$, as the relative frequency of shared links with a domain *dom* featuring a path of depth *pd*.

URL query string. Another way for a website to identify its content is the usage of query strings, that is, strings of (key, value)-pairs as part of a URL. Regarding the number as well as the name of the keys, query strings are not standardized. Thus, simply by looking at a URL, it is not obvious which keys are relevant to specify the content and which only affect, for example, the layout.

Example 2. YouTube relies solely on query strings to link to individual videos, for example:

```
http://www.youtube.com/watch?v=<video_id>&hd=1
```

where <video_id> is the unique identifier if the video. If hd=1 is set, then the video is played in high definition. YouTube uses a large variety of query string keys to, for example, skip to a specific time, prevent autoplaying, enable auto loop and to set the size of the embedded video player.

Intuitively, a query string key k is relevant to specify the content of a web page if (a) k occurs in the majority of shared URLs and if (b) its value differs among most URLs containing k. To formalize this, let $L_{dom}(pd, k)$ be set of URLs with domain *dom* and a

Table I. Threshold Parameters Used Within the Algorithm for the Construction of Virtual Locations

t_{freq}^{path}	minimum relative frequency of links with the same domain featuring the same path depth
$t_{freq}^{qs.key}$	minimum relative frequency of the same domain and path depth featuring the same query string key
$t_{uniq}^{qs.val}$	minimum relative uniqueness of a query string value in links with the same domain and path depth

path with depth *pd* and query string containing key *k*. With that, we define

$$freq_{dom}^{qs}(pd, k) = \frac{|L_{dom}(pd, k)|}{|L_{dom}(pd)|} \tag{2}$$

with $freq_{dom}^{qs}(pd, k) \in [0, 1]$ as the relative frequency of shared links with a domain *dom* and path with depth *pd* that contain query string key *k*. Furthermore, let $V_{dom}(pd, k)$ be the set of distinct values for a query string key *k*. Hence,

$$uniq_{dom}^{qs}(pd, key) = \frac{|V_{dom}(pd, k)|}{|L_{dom}(pd, k)|} \tag{3}$$

with $uniq_{dom}^{qs}(pd, key) \in [0, 1]$ defining the relative uniqueness of a query string value for links with domain *dom* and a path of depth *pd*. Finally, we combine both measures into a relevance score:

$$relevance_{dom}^{qs}(pd, k) = freq_{dom}^{qs}(pd, k) \cdot uniq_{dom}^{qs}(pd, key) \tag{4}$$

Note that we do not use the relevance score within our algorithm for constructing virtual locations. The reason is that a high uniqueness value does not compensate for a low relative frequency and vice versa. However, we evaluate the relevance scores in our experiments since the score represents an illustrative measure for analyzing the importance of query string keys.

Constructing virtual locations—algorithm. Based on the three definitions $freq_{dom}^{path}(pd)$, $freq_{dom}^{qs}(pd, k)$, and $uniq_{dom}^{qs}(pd, key)$, we now propose our algorithm for the construction of virtual locations; see Algorithm 1. Besides the URL, the algorithm takes threshold values—as listed in Table I—as additional input parameters. We first extract the domain, path, and query string from the input URL (Line 1). If we do not have any information about the current domain in our repository, then we return the input URL as virtual location (Line 2–5). Otherwise, we check if the depth of the URL's path is sufficiently represented in our repository, that is, whether it contains a sufficient number of URLs of the same domain and path depth according to threshold t_{freq}^{path}. If not, then we again return the URL as virtual location (Line 6–9). Last, we analyze the query string if present. For each (key, value)-pair in the query string we check, for the given domain and path depth, if the relative frequency as well as the uniqueness of the key exceeds the corresponding thresholds $t_{freq}^{qs.val}$ and $t_{uniq}^{qs.val}$. If so, then we add the (key, value)-pair to the newly constructed query string (Line 10–15). Finally, we return as virtual location the concatenation of the domain, path, and the new query string (Line 16). While not necessary to be constructed this way, the resulting virtual location represents the kind of "minimized" version of the input URL, containing only parts relevant to specify the unit of interest. The values for the three threshold parameters we derive from our data analysis (see Section 7).

5. LINKING LOCATIONS

In most application scenarios, connecting users based on shared locations is unnecessarily restrictive with respect to users' contexts. This is particularly true for virtual

ALGORITHM 1: $constructVirtualLocation(url, t_{freq}^{path}, t_{freq}^{qs.key}, t_{uniq}^{qs.val})$

$(dom, path, querystring) \leftarrow \text{parseUrl}(url)$
$pathData \leftarrow \text{getPathDataFromDb}(domain)$
if $pathData = null$ **then**
  return url;
end
$pd \leftarrow \text{calculatePathDepth}(path)$
if $pathData[pd] = null \ \vee \ freq_{dom}^{path}(pd) < t_{freq}^{path}$ **then**
  return url;
end
$s \leftarrow$ "?"
foreach *(key,value)-pair in querystring* **do**
  if $freq_{dom}^{qs}(pd, key) \geq t_{freq}^{qs} \ \wedge \ uniq_{dom}^{qs.key}(pd, key) \geq t_{uniq}^{qs.val}$ **then**
    $s \leftarrow s + key +$ "=" $+ value +$ "&"
  end
end
return $(dom + path + s)$

locations, since the context derived from virtual locations is typically much less unique than the context derived from physical locations. For example, with reviews being the main units of interests on Tripadvisor, each review represents a distinct virtual location. However, the context here is on the level of venues. Enabling only users reading the exact same review to socialize is very counter-intuitive. We therefore link locations, both virtual and physical, into *realms* to share the presence of users across different locations but describing the same context.

5.1. Defintion

A realm abstracts from individual locations to describe semantically more meaningful contexts like topics or categories. Realms are defined by the set of related virtual and physical locations.

Definition 2 (*Realm*). A realm $\mathcal{R} = \{loc_1^v, loc_2^v, \ldots, loc_m^v, loc_1^p, loc_2^p, \ldots, loc_n^p\}$ is non-empty and finite set of virtual and physical locations.

Based on the definitions of virtual locations and physical locations, each URL or geocoordinate is associated with exactly one virtual or physical location, respectively. However, a location can map to multiple realms, depending on which types of realms are meaningful in the context of a given application or service. Within OneSpace, we mainly focus on realms that directly derive from (mostly touristic) venues in Singapore. Figure 3 shows a example of mapping URLs and geocoordinates to virtual and physical locations and their linking to a common realm. Locations and realms, as the two levels of abstraction, provide a sufficient level of flexibility particularly to combine subsets of URLs and associate them to different topics or categories.

5.2. Creating Links

Similarly to locations, the definition of realms does not specify which virtual and physical locations should be linked into a realm. In the following, we outline three basic methods we applied for automatically creating realms within the application context of OneSpace.

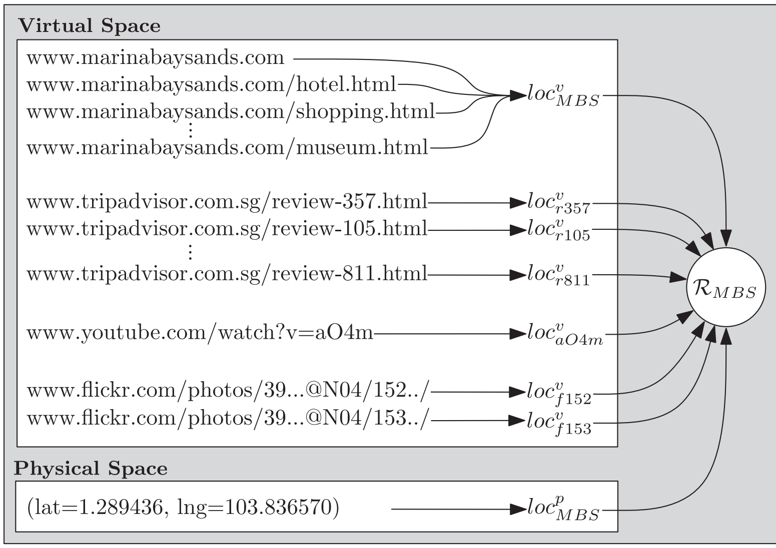


Fig. 3. Example of mapping URLs and geocoordinates to virtual and physical locations, and linking them to a common realm for the Marina Bay Sands (MBS) hotel. First, we argue that all pages of the hotel's official website form a virtual location loc^v_{MBS} since they all reflect a surfer's interest in the hotel. Second, we regard reviews as the unit of interest on Tripadvisor and therefore map each review to its own virtual location. Similarly, we consider YouTube videos and Flickr images about the hotel as virtual locations. Finally, we map all corresponding virtual locations to the realm $\mathcal{R}_{MBS}$. Regarding the physical space, we simply use a single-point geocoordinate to specify the physical location of the hotel and also map this physical location to $\mathcal{R}_{MBS}$.

Exploiting explicit links. For cyber-physical social networks, the links between virtual and physical locations are particularly relevant. Thus, we first collected a dataset of venues with their physical locations. We used the Google Places Application Programming Interface (API) to collect information about locations in the geographic area of Singapore (SG). This set of venues comprises data of 220k+ places, each featuring a geocoordinate. For one, this gives us a comprehensive set $\mathcal{V}_{SG}$ of 190k+ of venue names—note that venues of, for example, restaurant chains feature the same name (*McDonald's*, *KFC*, etc.). Also, 67.2k (30.5%) of the venues are also associated with a URL to a website. Most venues with websites are hotels, restaurants, shops, and tourist attractions but also companies and businesses. We map each venue i to a physical location loc^p_i and define a virtual location loc^v_i by mapping all pages of the venue's website to loc^v_i. Last, we link loc^v_i and loc^p_i to a realm $\mathcal{R}_i$ to represent venue i.

Using public APIs. Many platforms with information about venues provide APIs to access their data. YouTube and Flickr allow us to search for videos and images using a keyword query. For each venue name in $\mathcal{V}_{SG}$, we accessed the APIs using the name + *"singapore"* as a search query. For YouTube, we collected the video URLs of the top 50 results; for Flickr, we collected for the top 50 results the URLs of the image files as well as the URLs of the Hyper Text Markup Language (HTML) pages contaning the images. Both APIs rank the result set according to relevance, a platform-dependent metric that considers title, tags, comments, view count, and so on, of videos and images. We regard each collected URL to a HTML page as a virtual location and connect this virtual location to the realm derived from the physical location. The URL to the Flickr image file we use for an additional feature of OneSpace where we display related images and videos to surfers visiting linked virtual locations.

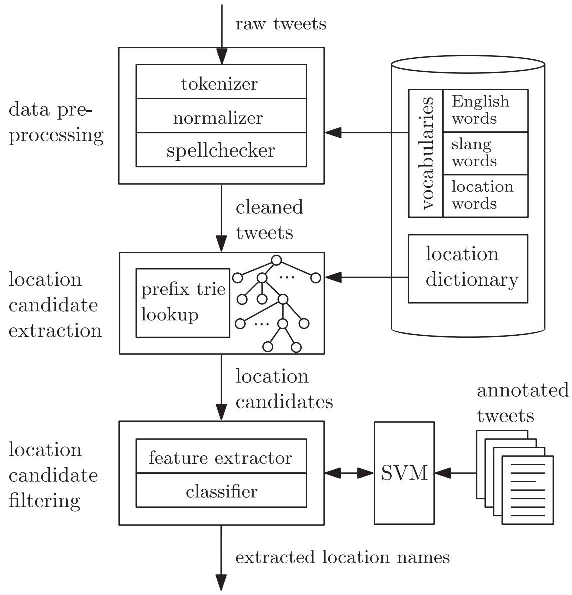


Fig. 4. Architecture of our system for the extraction of venue names from social media messages, highlighting the three main components: data preprocessing/cleaning, location candidate extraction, and location candidate filtering.

Location name matching and detection. For a more general approach to link locations, we extract venue names from text resources to link them to their respective realms. While this refers to the well-established Natural Language Processing task of NER, micro-locations such as venues' names feature characteristics that are very challenging for out-of-the-box NER solutions. For example, names of venues typically contain multiple words, many of them being standard English words. Additionally, we focus on text content shared over social media like Twitter or Facebook. Social media messages often suffer from an unorthodox writing style that existing NER systems cannot sufficiently handle. Here we provide a very brief sketch of our approach and point to von der Weth et al. [2015] for full details.

Tripadvisor reviews. We collected all reviews about hotels, restaurants, attractions, and activities, as well as shopping and nightlife venues, in Singapore. We currently have 261,646 reviews about 5,408 venues in our repository. We map each review j about a venue i to its own virtual location $loc^{v}_{i,j}$ and map each $loc^{v}_{i,j}$ to the already-existing realm $\mathcal{R}_i$. However, even with reliably extracting the name of a venue from a review page, the mapping is not a straightforward task. Naive exact match search often fails to due minor or major alternative spellings of names. Thus, we first normalized all venues names, that is, we transformed all word to lowercase and removed non-alphanumeric characters. We then indexed all venues names in $\mathcal{V}_{SG}$ using a tailored prefix trie that can handle dropped words at the the end of a name. With this approach, we can match the majority of reviews to the corresponding entries $\mathcal{V}_{SG}$. While our matching process is not 100% accurate (see Section 7 for details) a manual inspection shows that many venues are featured in only one of the datasets.

Tweets. Our system for the detection of venue names in tweets comprises three main components, as illustrated in Figure 4. First, we address the characteristics of typically short and informally written social media messages that feature a variety of non-standard tokens (e.g., slang, hashtags, emoticons, URLs). This mainly entails

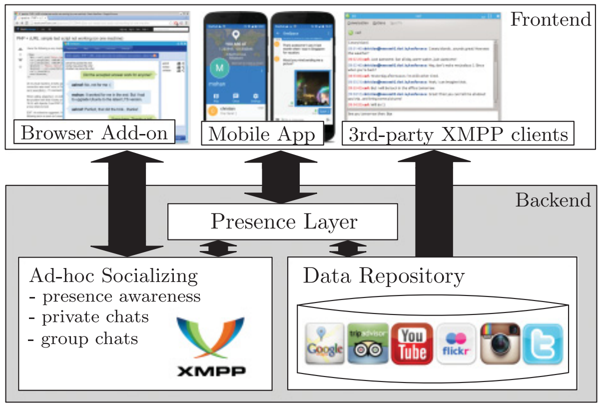


Fig. 5. General architecture of our OneSpace platform.

a tailored pipeline handling basic text processing tasks like tokenizing, normalizing, and spellchecking. Second, we extract location candidates as the longest phrases that match with an entry in a prefix trie indexing all venue names of $\mathcal{V}_{SG}$. This approach results in high recall but typically also in a high number of false positives due to many location names being comprised of standard English words. For example, with a venue name such as *"The Cookie Museum,"* each mentioning of *"cookie"* is a potential reference to a location. Thus, in a final step, we use a supervised learning approach to identify true locations—that is, given a location candidate, we use a binary classification to decide whether the candidate is a true or a false location. We trained a Support Vector Machine (SVM) classifier using cross-validation over a manually annotated dataset of 4,000 tweets, with each annotation being the word of phrase representing a venue name. The accuracy of a classifier generally depends on the set of features considered for the training and classification process. On the other hand, the set of features also affects the performance of both processes. Given the high volume of social media content, we favor a small set of simple features. Therefore, we focus on word-level and list lookup features that can easily be computed, yield a small feature vector, and do not require batches of messages. Our results of a series experiments yield: *precision* = *0.89*, *recall* = *0.7*, and *f1-score* = *0.83*, which outperforms existing approaches.

6. IMPLEMENTATION DETAILS

In this section, we provide deeper insights into the implementation of our OneSpace platform. Figure 5 gives an overview of the general architecture of OneSpace. The complete source code of all components, as well as the current snapshot of our data repository, are publicly available.[12]

6.1. Backend Architecture

Presence Awareness and Communication. For the support of instant messaging between users, we rely on the XMPP.[13] Group chats are the most relevant concepts within OneSpace, since we assign each realm to a group chat. The intuition is that users at the same or related locations are in the same group chat and are therefore aware

[12] https://github.com/chrisvdweth/onespace/.
[13] http://xmpp.org, http://xmpp.org/software/clients.html.

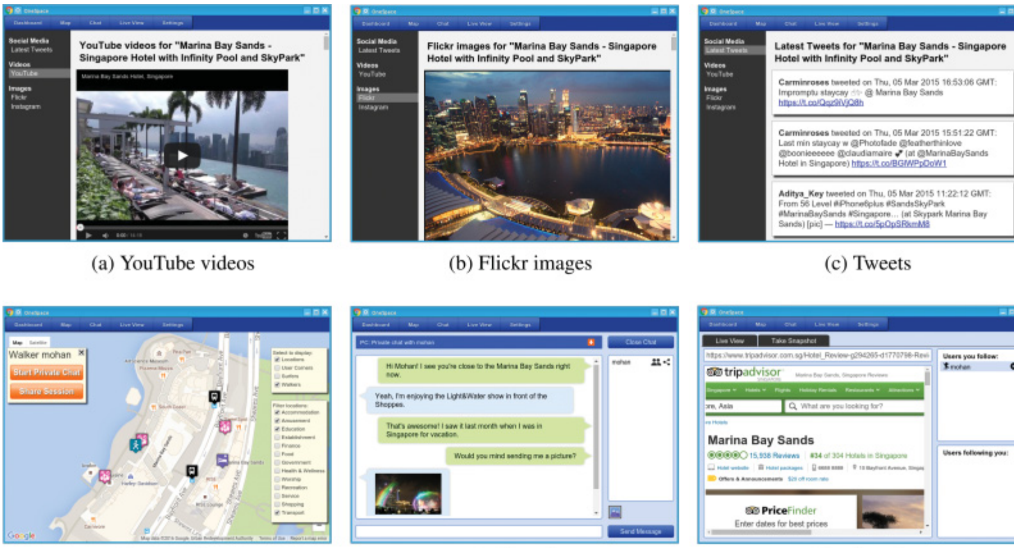
of each other. Surfers implicitly enter the corresponding group chats when browsing the Web. This includes that surfers are automatically in multiple group chats if they are browsing different pages in parallel, either using tabbed browsing or multiple browser windows. Walkers, but also surfers, can discover and explicitly enter group chats that are associated with a geolocation and are therefore displayed on the map (see Section 6.2). As a result, walkers can also participate in different group chats in parallel. Besides group chats, we also support private chats between users.

The use of XMPP as an established, platform-independent, open-standard protocol has several advantages. First, it allows us to access OneSpace using any third-party chat client with XMPP support. For example, the reception of a hotel can keep a constant connection to the group chat associated with that hotel without using the browser add-on or mobile application. Second, it allows third-party developers to easily implement further frontend applications like add-ons for different browsers or mobile applications for different mobile operating system. Last, XMPP is payload-agnostic—that is, one can send any type of string-encoded messages such as JavaScript Object Notation (JSON), Resource Description Framework (RDF), or eXtensible Markup Language (XML). In our implementation, we exchange JSON strings that not only encapsulate normal chat messages but also support image sharing, the live view feature, as well as other control commands shared between users. This makes adding new features to OneSpace very flexible.

Data Repository. The data repository maintains the links between physical and virtual locations. We represent physical locations as (lat,lng)-pairs. We focus on “single-point” locations like hotels, hospitals, museums, pubs, and shops. We also store locations with a larger spatial extent, like parks or golf courses, using a single geocoordinate (cf. Section 8 for a discussion). So far, we integrated the following data collected from different platforms into the data repository: (a) Google Places data containing 220k+ places with their geolocation, of which 67.2k (30.5%) are also associated with a URL to a website, that is, a virtual location; (b) YouTube videos and Flickr images collected using the respective public APIs; (c) tweets mentioning venues extracted from a crawler of geotagged tweets originating from Singapore; and (d) Instagram images as part of tweets mentioning venues.

OneSpace users can share images in group or private chats. All images are stored in the repository, both the raw image files and thumbnail versions. Sending an image by a User A involves the following steps: (a) the image is uploaded and stored in the file system of the backend and indexed in the repository, (b) the backend generates a thumbnail version of the image and also indexes it in the repository, (c) the links to the original image and the thumbnail are sent back to A as response to the upload request, (d) an XMPP messaging containing both image links are sent from A to the group or private chat, and (e) the frontend applications of the recipients display the thumbnail image which can be clicked on to open the original image.

Last, in contrast to the rather static dataset of venues, surfers and walkers can dynamically create and delete their own so-called user corners. A user corner is a physical location and is hence displayed on our map. Each user corner is associated with a public group chat, so any user can join the conversation. This allows, for example, walkers to create corners on the spot in case of an event (disruption of train service, “happy hour” in a bar, street parade, etc.); sharing their experiences while visiting, for example, the zoo, a restaurant or an attraction; or simply to kill some time at the airport while waiting for departure. Right now, we do not link user corners to other physical or virtual locations. We argue, however, that allowing users also to create links is a straightforward and meaningful extension. For example, a user creating a corner to “report” the disruption of the train service can also create links to images showing the goings-on he or she uploaded to Instagram.



(a) YouTube videos (b) Flickr images (c) Tweets

(d) Map view (e) Chat view (f) Live view

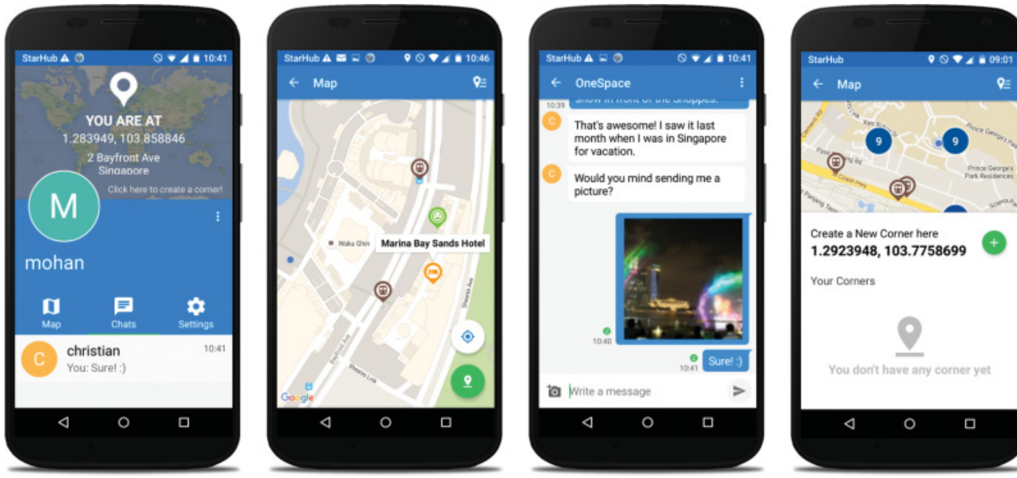
Fig. 6. Different views of the OneSpace browser add-on.

6.2. Frontend Applications

A browser add-on and a mobile app are the two main frontend applications of OneSpace, both providing similar features. With the novelty of considering the virtual context of users for socializing, we put emphasis on the browser add-on to motivate the social computing aspects of our platform.

Web Browser Add-On. The standard application to access the Internet is a web browser. Since we aim for a seamless integration of OneSpace into the normal browsing experience of users, we implemented a browser add-on providing a pop-up window as user interface. The pop-up window approach is very flexible compared to alternative solutions, for example, a sidebar in the browser window or injecting code into a page. Users can organize the pop-up window individually, and it supports tabbed browsing and multiple browser windows. Under the hood, the add-on listens to internal browser events (e.g., "new tab opened," "new page loaded," "new tab selected") and triggers corresponding actions, particularly joining, leaving, or switching between group chats.

The OneSpace browser add-on has various views to display information or to allow users to interact; Figure 6 shows multiple screenshots as examples. The dashboard, Figures 6(a)–(c), displays the tweets, YouTube videos, and Flickr and Instagram images related to the location. The map view, Figure 6(d), uses the Google Maps API to displays all available venues and user corners, as well as surfers and walkers as different markers. Users can click on each marker for further information and to perform a specified action, such as entering the group chat of a venue or user corner or starting a private chat with a walker or another surfer. Via a right-click into the map, surfers can create a new user corner. The chat view, Figure 6(e), has the basic look-and-feel of traditional chat clients. It supports participating in multiple private and group chats in parallel. If a surfer visits a web page, then the add-on connects to all the group chats associated with that virtual location. Apart from normal text messages, surfers can also share images in chats. The chat view displays each image as a thumbnail version. If a surfer clicks on a thumbnail, then the full-sized image opens in a new tab of the main browser window. Last, the live view, Figure 6(f), is a unique feature for surfers.



(a) Start screen (b) Map view (c) Chat view (d) User corner

Fig. 7. Different views of the OneSpace mobile application.

It allows a surfer, the *guide*, to share his or her current browsing session with other surfers, the *followers*. If the guide is browsing a page, then the corresponding URL is encapsulated in an XMPP message and sent to all followers, and displayed within their pop-up windows. The core idea behind this feature is, for example, to enable expert users to help less-tech-savvy users to find relevant information.

Mobile Phone Application. For walkers, we implemented a mobile application, featuring a map view and a chat view as the two core functionalities similar to the browser add-on; see Figure 7. Again, the map view allows walkers to discover places, user corners, surfers or other walkers. We assume that each device can determine its own geolocation, either using GPS or other methods. Thus, walkers using the OneSpace mobile app are also displayed on the map. The basic chat client supports group and private chats. Walkers can send images either by directly taking a picture with the camera or by selecting one from the local gallery. Similarly to surfers, walkers can create user corners but only at their current geolocation (and not anywhere).

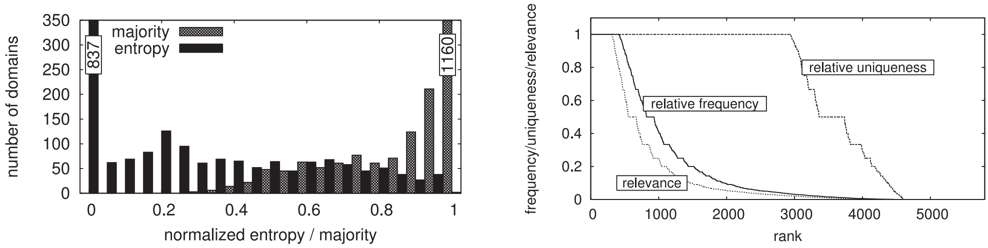
7. EVALUATION

Our evaluation is divided into three parts. In the first two subsections, we analyze our approach for the automated construction of virtual locations and then analyze our data repository as the result of our automated linking between virtual and physical locations. Last, we represent and discuss the responses from a user study regarding first feedback about OneSpace.

7.1. Virtual Locations

Constructing meaningful virtual locations derived from websites' *units of interest* (cf. Section 4) is a fundamental challenge for CPSN and not specific to the application context of OneSpace. To allow us to generalize from our results, we therefore evaluated a context-independent dataset. To this end, we collected our dataset from Reddit, a social news site where users can submit and share links. We used the Reddit API[14] to the request the most recent submitted links every 30s for about 1 week. We then removed all internal links, that is, links point pointed to content on Reddit, and all links

[14]https://www.reddit.com/dev/api.



(a) Distribution of domains regarding norm. entropy and max. frequency of path depths.

(b) Ranking of query string keys according to frequency, uniqueness and relevance.

Fig. 8. Results of the analysis of the shared links dataset for the construction of virtual locations.

with a rare domain ($\leq$20 links per domain), yielding 847k+ links from 2,018 different domains. We then analyzed each link regarding its main URL components: the path and query string. First, for each domain, we calculated the number of overall links and the number of links for each occurring path depth. Second, for each domain and path depth, we analyzed the query string. More specifically, for each (key, value)-pair in a query string, we calculated the key's number of occurrences in the query strings and the number of distinct values for that key.

Path depths. We first calculated the distribution of path depths for each domain. About 40% of all domains have been shared via URLs with the same path depth, that is, with $freq_{dom}^{path}(pd) = 1.0$. The more interesting case are domains of shared links using URLs with different path depths. For these domains, we calculated (a) the *normalized entropy* of the different frequencies and (b) the *maximum frequency* value. The lower the normalized entropy and the higher the maximum relative frequency, the more is one path depth dominating among the URLs of a domain. To give an example, consider 100 shared links of the same domain *dom*. Each link has a path depth *pd* of 1, 2, or 3, with the following frequencies: $freq_{dom}^{path}(1) = 0.8$, $freq_{dom}^{path}(2) = 0.15$, and $freq_{dom}^{path}(3) = 0.05$. Here the normalized entropy is 0.55 and the maximum relative frequency is 0.8. Figure 8(a) shows the result as a histogram. Most domains show a dominating path depth, with a normalized entropy $\leq$0.5 and a maximum relative frequency $\geq$0.66. Only for very few domains users share links that frequently have paths with different depths.

Query strings. Of all 2,018 domains in our dataset, only 843 (41.8%) use any kind of query string as part of their URLs. For these 843 domains, we ranked all 4.6k+ query string keys according to their relative frequency, relative uniqueness, and relevance; see Figure 8(b). As the results show, many of the keys feature a relative uniqueness of 1.0. However, most of these keys are not part of the majority of links of a domain, displayed by the much smaller set of keys with a high relative frequency. Not surprisingly, query string keys with a high relevance score are typically unique identifiers of articles in online newspapers, images or videos on media sharing sites, or user profiles of social networking sites. Regarding the construction of virtual locations, two observations are important. First, only a small number of query string keys are important to specify the content of a web page for a given URL. Second, the sharp drop of all three measures indicate that, in most cases, it is very easy to distinguish important from irrelevant keys.

Threshold settings. With these outcomes, we can now identify practical values to set the three threshold parameters for Algorithm 1 for the construction of virtual locations. To avoid the rare exceptional cases regarding uncommon path depths, we set $t_{freq}^{path} = 0.2$. For a query string key, we require that it is at least present in a

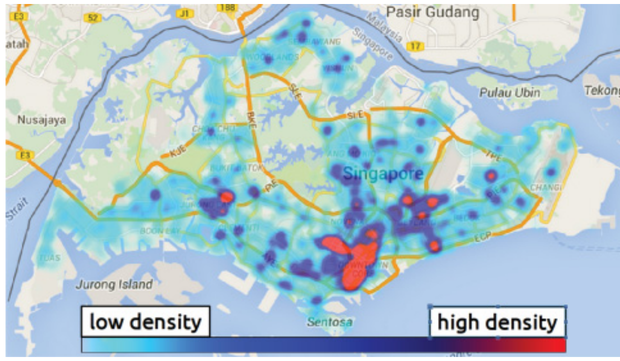


Fig. 9. Heatmap showing the distribution of venues associated with websites. The largest hot spot covers the Central Business District in the mid-southern parts of Singapore. The density of offices, business, hotels, restaurants, and other touristic venues is particularly high here. Other hot spots typically are in areas with at least one large shopping mall that contains a large number of individual shops, each having their own website.

two-thirds majority of links: $t_{freq}^{qs.key} = 0.66$, and that its values are very unique: $t_{uniq}^{qs.val} = 0.9$. Note that the uniqueness of important keys can be less then 100%. First, there are always exceptional cases where users, for example, might share a YouTube link that does not point to a video. Second, a link to a video might be shared multiple times but with (slightly) different URLs. Using these threshold parameters and applying Algorithm 1 on our dataset of 847k+ links, we get 807k+ different virtual locations. Since most Reddit users aim to submit novel content, most links form their own virtual location. There are, however, recurring cases where the same content has been submitted multiple times—particularly news articles or videos—but with varying URLs. A closer inspection showed that in the very most cases URLs pointing to the same content differ in the number and set of used query string parameters. For example the news article `www.cnn.com/2014/08/02/us/us-spy-plane/index.html` has been shared 5 times, either with different query string parameters or different values for the same parameters:

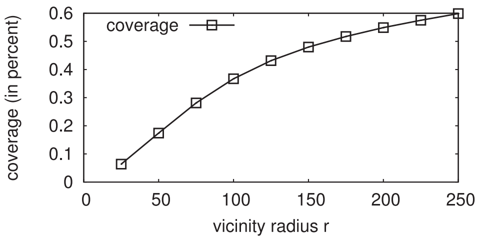
—`http://www.cnn.com/2014/08/02/us/us-spy-plane/index.html`
—`http://www.cnn.com/2014/08/02/us/us-spy-plane/index.html?hpt=hp_t2`
—`http://www.cnn.com/2014/08/02/us/us-spy-plane/index.html?hpt=hp_t1`
—`http://www.cnn.com/2014/08/02/us/us-spy-plane/index.html?c=europe`
—`http://www.cnn.com/2014/08/02/us/us-spy-plane/index.html?hpt=hp_t4`

Summing up, exploiting the wisdom of the crowds by analyzing shared links is an effective way to identify the units of interest of a website. This in turn enables a fully automatic approach to return a meaningful virtual location for a given URL as proposed in Algorithm 1.

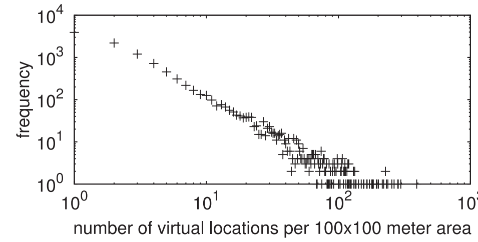
7.2. Linking Locations

We provide an analysis of our data repository of collected and linked physical and virtual locations to motivate the potential benefits of merging the virtual and physical world.

Google Places data. Our current dataset contains 220k+ venues in Singapore. About 67.2k (30.5%) of these venues are associated with websites. Figure 9 shows qualitatively the distribution of venues across Singapore. For a more quantitive analysis, we first calculated their coverage. To this end, we represented each venue as a



(a) Coverage of venues for different vicinity radiuses r_v ($r_v \in \{25, 50, ..., 250\}$).

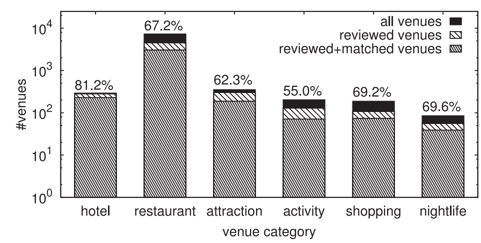


(b) Distribution of squares of size 100x100 meter containing at least one venue.

Fig. 10. Results of the analysis of the collected Google Places data.

	#venues	#reviews	#reviewed
hotels	292	107,276	282 (96.6%)
restaurants	7,292	88,068	4532 (62.2%)
attractions	352	54,656	302 (85.9%)
activities	203	5,955	129 (63.6%)
shopping	187	3,457	107 (57.2%)
nightlife	85	2,234	56 (65.9%)
∑	8,411	261,646	5,408 (64.3%)

(a) Overview of number of venues and reviews collected from Tripadvisor.

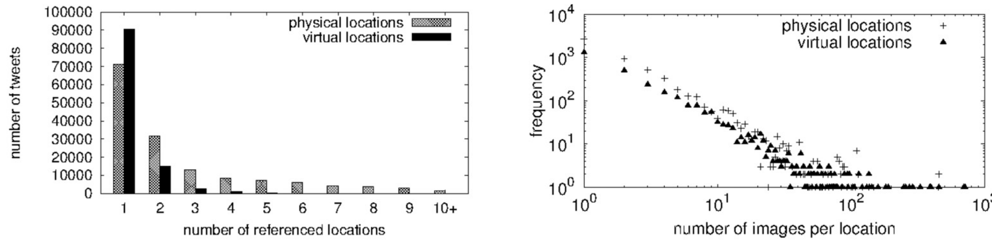


(b) Overlap between the Tripadvisor and Google Places data.

Fig. 11. Results for the analysis of the collected Tripadvisor data.

circle with its geolocation as center and a vicinity radius r_v. Radius r_v describes the vicinity around a venue in which we consider a user being present at that venue. With this, the coverage derives from the accumulated area of all circles; we report the coverage as ratio compared to the total area of Singapore. Naturally, the coverage increases for larger r_v, resulting in up to 60% coverage for $r_v = 250m$. Note that the other 40% include large, undeveloped areas or purely industrial areas, as well nature reserves. Regarding the distribution of venues, we divided the city area into squares with side lengths of 100m and counted the number of venues within each square. Figure 10(b) shows the distribution of all non-empty squares. As one would expect, the distribution of venues shows a power-law relationship: While most squares contain only a small number of locations, few squares contain a very large number of them. These results show that readily available data already provide a significant overlap between the virtual and physical space—that is, the number of venues that feature both a physical and virtual location is very high, even without additional links to related locations. As a consequence, in urban areas, each walker has a very high likelihood to be “close” to at least one website at any given time. In particular, the connections of venues such as hotels, restaurants, bars, shops, or tourist attractions with their respective official websites are most useful in the context of OneSpace.

Tripadvisor data integration. We crawled 261k+ reviews about 8,411 venues in Singapore from Tripadvisor. The table in Figure 11(a) shows basic statistics; note that not every venue has been reviewed. Regarding the mapping of reviews to their corresponding virtual location, we calculated the overlap between the Tripadvisor and Google Places dataset using our name-matching approach as described in Section 5 (see Figure 11(b)); the number above the bar is the ratio of matched and reviewed venues. The results show that we can match the majority of reviews, but not all, to the corresponding Google Places venues. While our matching process is not 100% accurate,



(a) Number of reference of physical and virtual locations in tweets.

(b) Distribution of the number of Instagram images linked to locations.

Fig. 12. Results for the analysis of the Twitter and Instagram data.

a manual inspection shows that most missing matches are caused by venues that are featured in only one of the two datasets.

Tweets and Instagram images. We applied our location name extraction method on a dataset of 1.3 million tweets. We observed that often an extracted name can refer to multiple physical locations (e.g., names of restaurant chains such as *"McDonald's"* or *"KFC"*). We therefore ignored all extracted names that potentially point to more than 10 actual locations. About 11.2% of all tweets refer to at least one physical location. Figure 12(a) shows the distribution of tweets with respect to the number of referenced physical and virtual locations. In most case, an extracted name refers to one or only a very small number locations. In a subsequent step, we extracted for each tweet sent via Instagram the image URLs to both the web page and media file. Using the extracted locations, we finally linked the Instagram images to the respective realm. Figure 12(b) shows that the distribution of the number images linked to locations follows a power-law relationship. Particularly at popular tourist attractions, people share images on Instagram.

YouTube videos and Flickr images. We used the keyword-based search API of both platforms to collect videos and images. As search queries we used all location names from our Google Places dataset and added the term *"singapore"* to minimize the number of irrelevant hits. For each location, we collected the URLs of up to 50 videos and images; in the case of Flickr we consider both the URL to the web page as well as the image URL. Each of these URLs we consider as virtual location and link them to the existing and related virtual and physical locations of the same realm. Our data repository is currently composed of over 1.9 million YouTube URLs and over 44k Flickr URLs. The reason for the large difference between these two numbers is that the YouTube API is far more "generous." As result, apart from very popular locations, many returned videos do not adequately reflect the intended search request. Filtering and ranking videos (but also images) using different methods including media content analysis is part of our future work.

7.3. User Study

A full evaluation of the effectiveness of our platform is extremely challenging and beyond the scope of this article; see a discussion in Section 8. For this, we invited 20 participants (who were students) and introduced OneSpace to them in the form of a presentation. We then let them try and test both frontend applications of OneSpace. After that, we conducted an anonymous survey asking about their impressions. In the following, we present the main results of this survey.

Recommendation discovery. We first asked the participants if and how they use different channels to choose new venues such as restaurants or attractions; see Table II (top). The results show that users regularly check online sources like recommendation

Table II. Results of User Study

"How often do you use different sources to choose a venue (restaurants, hotels, etc.)?"

	Never	**Rarely**	**Sometimes**	**Often**	**Always**
Recommender sites	10%	10%	45%	25%	10%
Family, friends, etc.	0%	0%	40%	45%	15%

"How often do you share your visits of venues and your experiences with others?"

	Never	**Rarely**	**Sometimes**	**Often**	**Always**
Social media	55%	15%	15%	15%	0%
Media sharing	90%	5%	0%	5%	0%
Recommender sites	40%	30%	25%	5%	0%
Emails, chat, etc.	15%	15%	30%	30%	10%

"How do you rate the helpfulness of the OneSpace frontend applications and services"

	Not at all	**A little**	**Somewhat**	**Very**	**Extremely**
Browser add-on	10%	10%	25%	40%	15%
Mobile application	5%	0%	40%	50%	0%
Tailored data	5%	5%	25%	55%	10%
Instant messaging	5%	20%	35%	30%	10%

sites and ask for opinions from family members or friends even more so. This suggests that users prefer immediate and first-hand information, particularly from known and trusted others. This indicates the potential added value of OneSpace to users but also the need for effective trust and reputation management (cf. Section 8).

Recommendation sharing. We were also interested in how users share their own experiences with others online. Table II (middle) shows how our 20 survey participants use various distribution channels. Again, the most commonly used are emails or chats as a more personalized and direct way to share experiences. A bit surprisingly, more than half of the participants stated that they write reviews on recommendation sites at least occasionally. Also common is using social media such as Facebook or Twitter to comment on visited venues. OneSpace enables a new way for personalized recommendation sharing to other users that have an immediate interest by teaming-up users based on shared locations, both virtual and physical.

Assessment of OneSpace. Last, the participants were asked to rate the different aspects of OneSpace—the two frontend applications and the two types of services—with respect to their helpfulness; see Table II (bottom). In general, most participants see the potential of cyber-physical social networks like OneSpace. Interestingly, the helpfulness of the browser add-on is more disputed. Our explanation is that the mobile application is more familiar to users since it is used like commonly available location-based service applications. The browser add-on, which particularly reflects the ideas behind CPSN, is a rather novel and untested concept. Regarding the helpfulness of the two types of services, the results are more consistent. At least 3/4 of all participants find tailored information and the support of instant messaging at least somewhat helpful.

8. DISCUSSION AND ROADMAP

With OneSpace, we presented our current research prototype of a cyber-physical social network. It mainly acts as proof of concept to motivate the potential benefits of CPSN. In this section, we outline our long-term efforts to make OneSpace as a CPSN a successful concept outside the lab.

Permanent connections and extended features. Compared to other established social networking platforms, we currently focus on making new connections between users based on their physical and virtual locations. As a result, for now, the main features of the frontend applications revolve around presence awareness and ad hoc socializing. In our future work, we plan to provide a contact list that allows users to establish long-term connection with other users, other features commonly found on social networking platforms such as the logging of chat histories, or the possibility to explicitly search for other users. Other meaningful extensions of OneSpace may include leveraging connections from user's existing social networks, for example, Facebook or Google+. Depending on the expected added values of different solutions, we aim to support long-term connections between users to strengthen the networking component of our platform.

Privacy concerns. Like the physical location of a user, his or her virtual location is privacy-sensitive information. For example, surfers do not want to share all web pages they are browsing on. For the time being, users can go "offline" by simply closing the pop-up window of the browser add-on or by closing the mobile application. The importance of privacy-preserving techniques has long been acknowledged in the context of location-based services, including location-based social networks. The basic approach is not disclosing one's exact location through, for example, data obfuscation or anonymization techniques. The main challenge here is to identify a meaningful tradeoff between the level of privacy and the quality of the provided service. It needs to be investigated how existing privacy-preserving techniques for traditional location-based services can be applied in the context of virtual locations. Suitable solutions might include mechanisms that enable users to automatically blacklist places based on, for example, their associated set of tags. Other solutions might involve the formulation of privacy polices known from online social networks—that is, users can explicitly formulate to which groups (family, friends, colleagues, etc.) which virtual locations they want to disclose.

Locations with extended dimensions. So far, we have limited ourselves to describing a physical locations by means of a (lat,lng)-pair, similar to many common location-based services. However, considering also the height is potentially important in case multistoried buildings (e.g., shopping malls or offices) facilitating many venues but is beyond the average accuracy of GPS. Reliably measuring the height within buildings calls for suitable indoor location solutions (e.g., using Wi-Fi signal strength) or alternative sensors, like a barometer, found in some mobile devices. Furthermore, many venues like parks and golf courses but also large buildings inherently feature spatial extents. Representing such venues as, for example, polygons, will require us to handle spatial relations between overlapping, intersecting, covering, and so on, polygons [Egenhofer and Franzosa 1991]. This in turn affects the linking between virtual and physical locations. For example, consider a restaurant inside a shopping mall with both restaurant and mall featuring dedicated websites. By exploiting the spatial relation, we can automatically link the geolocation of the shop with the website of the shopping mall.

Extended mapping and linking. CPSN benefit from linking virtual and physical locations into realms. We presented different approaches for how the linking can be done automatically in the context of OneSpace. Besides automatic linking approaches, solutions in the spirit of the Web 2.0, that is, letting users do the work, are also conceivable. For example, the OneSpace frontend applications can offer features that allow users to manually create links between virtual and physical locations. The challenges here are less on the algorithmic side but rather to incentivize users to create meaningful links. In general, users are willing to contribute if their perceived benefit (greatly) outweighs their perceived costs. Thus, creating links must be an easy task and add value to users' experience, for example, by improved services or reputation points. On the other hand, the easier it is to create links, the easier it can be exploited by malicious party. For example, a user selling a shady product might link popular web pages to the product's page to lure users into visits. These are general challenges for crowdsourcing approaches and often require solutions customized to the scenario at hand.

Large-scale user study. In Section 7.3, we investigated users online behavior and users' assessment in term of a user study with 20 participants. We acknowledge that, due to the small scale of our study, the generalizability of our results is limited. We argue that a full-fledged user study to evaluate the effectiveness of OneSpace is beyond the scope of this article and merits its own considerations given the involved challenges and required scale. First, the definition of meaningful metrics is not obvious since the benefits includes intangible factors such as the enjoyment due to socializing. Second, we argue that the results of a classic lab study to evaluate the usefulness of OneSpace have their principle limitations. We therefore aim for a field study or a similar setting where volunteers use OneSpace over a prolonged period of time.

9. CONCLUSIONS

Cyber-physical social networks represent a new paradigm for social computing. It allows users to connect and interact with each other based on their physical as well as virtual locations. Compared to physical locations, the notion of a virtual location is typically less well defined and a less unique indicator for a user's (virtual) context. We addressed both issues by presenting algorithms for the construction of virtual locations and the linking of related locations, both virtual and physical. As an example implementation of a CPSN, we presented OneSpace, a live and social recommendation platform. OneSpace supports presence awareness and ad hoc socializing between both web and mobile users based on shared locations. While being first and foremost a research prototype, our platform explores and also suggests novel use cases of social computing. We argue that the paradigm of CPSN will boost online networking across a wide range of application scenarios including geo-social search, ad hoc online collaboration, social journalism of real-world events, as well as novel approaches towards online advertising or online gaming and entertainment.

REFERENCES

Hamed Abdelhaq, Christian Sengstock, and Michael Gertz. 2013. EvenTweet: Online localized event detection from twitter. *Proc. VLDB Endow.* 6, 12 (2013), 1326–1329.

Yong-Yeol Ahn, Seungyeop Han, Haewoon Kwak, Sue Moon, and Hawoong Jeong. 2007. Analysis of topological characteristics of huge online social networking services. In *WWW'07*. ACM, 835–844.

Saleema Amershi and Meredith Ringel Morris. 2008. CoSearch: A system for co-located collaborative web search. In *SIGCHI'08*. ACM, 1647–1656.

Sitaram Asur and Bernardo A. Huberman. 2010. Predicting the future with social media. In *WI-IAT'10*. IEEE, 492–499.

Lars Backstrom, Eric Sun, and Cameron Marlow. 2010. Find me if you can: Improving geographical prediction with social and spatial proximity. In *WWW'10*. ACM, 61–70.

Jie Bao, Yu Zheng, David Wilkie, and Mohamed Mokbel. 2015. Recommendations in location-based social networks: A survey. *GeoInformatica* 19, 3 (2015), 525–565.

Jingwen Bian, Yang Yang, and Tat-Seng Chua. 2014. Predicting trending messages and diffusion participants in microblogging network. In *SIGIR'14*. ACM, 537–546.

Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. A 61-million-person experiment in social influence and political mobilization. *Nature* 489, 7415 (2012), 295–298.

Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna P. Gummadi. 2010. Measuring user influence in twitter: The million follower fallacy. In *ICWSM'10*, Vol. 10. AAAI Press, 10–17.

Terence Chen, Mohamed Ali Kâafar, and Roksana Boreli. 2013. The where and when of finding new friends: Analysis of a location-based social discovery network. In *ICWSM'13*. AAAI Press.

Eunjoon Cho, Seth A. Myers, and Jure Leskovec. 2011. Friendship and mobility: User movement in location-based social networks. In *SIGKDD'11*. ACM, 1082–1090.

Max J. Egenhofer and Robert D. Franzosa. 1991. Point-set topological spatial relations. *Int. J. Geogr. Inf. Syst.* 5, 2 (1991), 161–174.

Ronen Feldman. 2013. Techniques and applications for sentiment analysis. *Commun. ACM* 56, 4 (2013), 82–89.

Sharad Goel, Jake M. Hofman, and M. Irmak Sirer. 2012. Who does what on the web: A large-scale study of browsing behavior. In *ICWSM'12*.

Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A. Zighed. 2013. Information diffusion in online social networks: A survey. *SIGMOD Rec.* 42, 2 (2013), 17–28.

Keith N. Hampton, Lauren Sessions Goulet, Lee Rainie, and Kristen Purcell. 2011. Social Networking Sites and Our Lives. (2011). Pew Internet & American Life Project, May 6.

Schoen Harald, Gayo-Avello Daniel, Takis Metaxas Panagiotis, Mustafaraj Eni, Strohmaier Markus, and Gloor Peter. 2013. The power of prediction with social media. *Internet Res.* 23, 5 (2013), 528–543.

Clayton J. Hutto and Eric Gilbert. 2014. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *ICWSM'14*.

Jing Jiang, Christo Wilson, Xiao Wang, Wenpeng Sha, Peng Huang, Yafei Dai, and Ben Y. Zhao. 2013. Understanding latent interactions in online social networks. *ACM Trans. Web* 7, 4, Article 18 (2013), 18:1–18:39.

Long Jin, Yang Chen, Tianyi Wang, Pan Hui, and A. V. Vasilakos. 2013. Understanding user behavior in online social networks: A survey. *IEEE Commun. Mag.* 51, 9 (2013), 144–150.

Robert Krueger, Dennis Thom, and Thomas Ertl. 2014. Visual analysis of movement behavior using web data for context enrichment. In *PacificVis'14*. IEEE, 193–200.

Jiří Kysela, Josef Horálek, and Filip Holík. 2015. Measuring information quality of geosocial networks. In *New Trends in Intelligent Information and Database Systems*. Springer, 171–180.

Cliff Lampe, Nicole Ellison, and Charles Steinfield. 2006. A Face(Book) in the crowd: Social searching vs. social browsing. In *CSCW'06*. ACM, 167–170.

Gilly Leshed, Eben M. Haber, Tara Matthews, and Tessa Lau. 2008. CoScripter: Automating & sharing how-to knowledge in the enterprise. In *CHI'08*. ACM, 1719–1728.

Bang Hui Lim, Dongyuan Lu, Tao Chen, and Min-Yen Kan. 2015. #Mytweet via instagram: Exploring user behaviour across multiple social networks. In *ASONAM'15*. ACM, 113–120.

Janne Lindqvist, Justin Cranshaw, Jason Wiese, Jason Hong, and John Zimmerman. 2011. I'm the mayor of my house: Examining why people use foursquare—a social-driven location sharing application. In *CHI'11*. ACM, 2409–2418.

Meredith Ringel Morris and Eric Horvitz. 2007. SearchTogether: An interface for collaborative web search. In *UIST'07*. 3–12.

Meredith Ringel Morris, Andreas Paepcke, and Terry Winograd. 2006. TeamSearch: Comparing techniques for co-present collaborative search of digital media. In *TABLETOP'06*. IEEE Computer Society, 97–104.

Jacob Poushter, Jill Carle, James Bell, and Richard Wike. 2015. Internet Seen as Positive Influence on Education but Negative on Morality in Emerging and Developing Nations. Pew Internet & American Life Project, March 19.

Daniel M. Romero, Wojciech Galuba, Sitaram Asur, and Bernardo A. Huberman. 2011. Influence and passivity in social media. In *WWW'11 Companion*. ACM, 113–114.

Salvatore Scellato, Anastasios Noulas, and Cecilia Mascolo. 2011. Exploiting place features in link prediction on location-based social networks. In *SIGKDD'11*. ACM, 487–501.

Sindy R. Sumter, Laura Vandenbosch, and Loes Ligtenberg. 2017. Love me tinder: Untangling emerging adults motivations for using the dating application tinder. *Telemat. Inf.* 34, 1 (2017), 67–78.

John Tang, Mirco Musolesi, Cecilia Mascolo, and Vito Latora. 2010. Characterising temporal distance and reachability in mobile and online social networks. *SIGCOMM Comput. Commun. Rev.* 40, 1 (2010), 118–124.

Jih-Hsin Tang, Ming-Chun Chen, Cheng-Ying Yang, Tsai-Yuan Chung, and Yao-An Lee. 2016. Personality traits, interpersonal relationships, online social support, and Facebook addiction. *Telemat. Inf.* 33, 1 (2016), 705–714.

Io Taxidou and Peter M. Fischer. 2014. Online analysis of information diffusion in twitter. In *WWW'14 Companion*. ACM, 1313–1318.

Christian von der Weth, Lekha Chaisorn, and Mohan Kankanhalli. 2015. *Micro-Location Detection in Tweets*. Technical Report. Interactive & Digital Media Institute, National University of Singapore.

Christian von der Weth and Manfred Hauswirth. 2013. Finding information through integrated ad-hoc socializing in the virtual and physical world. In *WI-IAT'13*. IEEE, 37–44.

Christian von der Weth, Vinod Hedge, and Manfred Hauswirth. 2014. Virtual location-based services: Merging the physical and virtual world. In *ICWS'14*. IEEE, 113–120.

Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert-Laszlo Barabasi. 2011. Human mobility, social ties, and link prediction. In *SIGKDD'11*. ACM, 1100–1108.

Heather Wiltse and Jeffrey Nichols. 2009. PlayByPlay: Collaborative web browsing for desktop and mobile devices. In *CHI'09*. ACM, New York, NY, 1781–1790.

Shaomei Wu, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. 2011. Who says what to whom on twitter. In *WWW'11*. ACM, 705–714.

Rongjing Xiang, Jennifer Neville, and Monica Rogati. 2010. Modeling relationship strength in online social networks. In *WWW'10*. ACM, 981–990.

Wei Xie, Feida Zhu, Jing Jiang, Ee-Peng Lim, and Ke Wang. 2013. TopicSketch: Real-time bursty topic detection from twitter. In *ICDM'13*. 837–846.

Minhui Xue, Limin Yang, Keith W. Ross, and Haifeng Qian. 2016. Characterizing user behaviors in location-based find-and-flirt services: Anonymity and demographics. *Peer-to-Peer Networking and Applications* (2016), 1–11.

Received February 2016; revised July 2016; accepted September 2016