The Live Social Semantics application: a platform for integrating face-to-face presence with on-line social networking

Wouter Van den Broeck[†], Ciro Cattuto[†], Alain Barrat[‡†], Martin Szomszor[*], Gianluca Correndo[*] and Harith Alani[§]
[*] *Intelligence, Agents, Multimedia, School of Electronics and Computer Science, University of Southampton, Southampton, UK*
{msn2, gc3}@ecs.soton.ac.uk
[†] *Complex Networks and Systems Group, Institute for Scientific Interchange (ISI) Foundation, Turin, Italy*
{ciro.cattuto, wouter.vandenbroeck}@isi.it
[‡] *Centre de Physique Théorique, CNRS UMR 6207, Marseille, France*
alain.barrat@cpt.univ-mrs.fr
[§] *Knowledge Media institute, Open University, UK*
h.alani@open.ac.uk

***Abstract*—We describe a novel application that integrates real-world data on the face-to-face proximity of individuals with their identities and contacts in on-line social networks. This application was successfully deployed at two conference gatherings, ESWC09 and HT2009, and actively used by hundreds of people. Personal profiles of the participants were automatically generated using several Web 2.0 systems and semantic data sources, and integrated in real-time with face-to-face proximity relations detected using RFID-enabled badges. The integration of these heterogeneous data sources enables various services that enhance the experience of conference attendees, allowing them to explore their social neighborhood and to connect with other participants. This paper describes the architecture of the application, the services we provided, and the results we achieved in these deployments.**

***Keywords*-sensor networks, active RFID, Web 2.0, social networking services, real-time services, face-to-face presence, spontaneous interaction, enriched awareness, social context awareness, heterogeneous data sources, semantic web, social network cross-mining, deployment**

I. Introduction

A crucial factor that determines the quality of conferences and similar gatherings is the type of social experiences they provide. Consequently, organizers often strive to enhance the social experience of the event by offering activities or technologies that enhance the opportunity for social networking. We strove to significantly further the state of the art by developing an application that integrates (a) the rich social data from *existing* social networking systems, (b) semantic data sources describing collaboration networks and communities of practice, and (c) an infrastructure for sensing face-to-face communication in physical space, based on active radio-frequency identification (RFID). We built several services that use this data to enhance the social experience of conference attendees. The resulting prototypical application was deployed at the 2009 European Semantic Web Conference (ESWC09) and at the 2009 ACM Hypertext conference (HT2009).

The following section summarizes related work. A full description of the Live Social Semantics application is given in section III. Section IV covers the outcomes of the application deployment at ESWC09 and HT2009. Discussion and future work are given in section V, followed by conclusions in section VI.

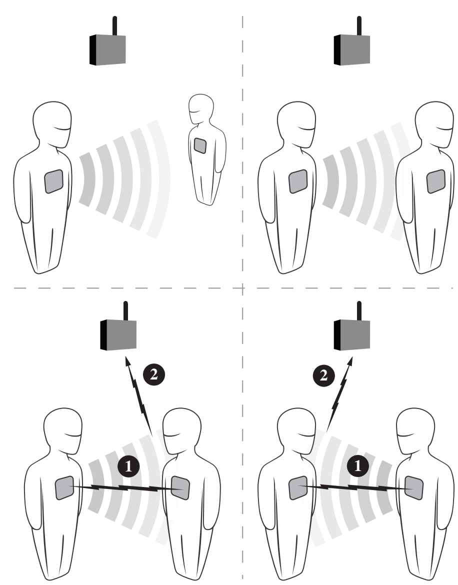

Figure 1. The SocioPatterns platform for distributed sensing of face-to-face proximity. Active RFID tags embedded in conference badges engage in ultra-low-power bidirectional packet exchange (1). Packet exchange is only possible (bottom panels) when two persons are at close range and facing each other, as the body blocks the exchange of low-power packets (top-right panel). Sustained face-to-face interactions are reported (2) to a data collection infrastructure.

II. Related Work

The interplay of networking and social contact at a conference gathering was initially investigated in the context of memetic spreading [1] and opportunistic networking for mobile devices [2] using wearable Bluetooth-enabled devices. Subsequent work focused on sensing organisational aspects [3] with Bluetooth-enabled mobile phones, and on characterising some statistical properties of human mobility and contact [4], [5], [6]. All of these early experiments involved a small number of participants, and had no capability of assessing face-to-face human contact in a large-scale setting, as they mostly relied on Bluetooth communication, which only detects co-presence on a few meters scale, or on sophisticated devices whose cost prevents large-scale deployments. RFID-based devices have been also used for Event Data Management (EDM) systems. Recently, the SocioPatterns project[1] investigated face-to-face proximity of individuals at large-scale social gatherings by deploying a distributed and scalable RFID platform [7] (figure 1). The application presented here leverages that platform to mine face-to-face presence in real-time.

To the best of our knowledge, our application is the first where real-world contacts are mashed up in real time with data from on-line social networking systems and with semantic data sources.

III. Live Social Semantics application

The Live Social Semantics application was deployed for 4 days at the 2009 European Semantic Web Conference (ESWC09[2]) [8], and for 3 days at the 2009 ACM Hypertext conference (HT09[3]). More than 500 people attended these conferences, out of which about 300 accepted to use our application. Each participant was issued with a uniquely numbered RFID badge. Users were asked to enter the numeric identifier of their badge on the website of the social application. On this website, users were also asked to declare their on-line identities on Delicious, Flickr, and lastFM[4], and they could activate a Facebook application that collects their Facebook friends.

A. General Architecture

Figure 2 illustrates the global architecture of the Live Social Semantics application. The vertical axis partitions the diagram according to two spaces: the web-based world (i.e. data about individuals available on the web), and the real world (i.e. proximity data from RFID badges). Fusion of data from heterogenerous sources is achieved by representing all the collected data in a semantic (RDF) format. A movie illustrating the user-facing functionality of the application is available at http://www.vimeo.com/6590604 .

[1]http://www.sociopatterns.org
[2]http://www.eswc2009.org
[3]http://www.ht2009.org
[4]http://delicious.com, http://flickr.com, http://last.fm

B. Web-based social networks and profiles of interest

Web data are sourced from social networking systems, to obtain tagging data and friend networks, as well as from the Semantic Web, to obtain information about publications, projects, and communites of practice (via RKBExplorer[5] and semanticweb.org). Data from Web 2.0 systems are imported using APIs or scraping, and subsequently converted to RDF format. Semantic data are collected directly from linked data sites [9], [10], or converted to a linked data representation via the Extractor Daemon. All data are stored in a triple store (center, right of diagram) to have a single service endpoint that provides access and reasoning over data. The Profile Builder (center, top of diagram) processes an individual's tagging activities and link them to DBpedia[6] concepts indicating a user's interests [11]. The assumption is that tags used most often by an individual correspond to the topics, places, events and people they are interested in. Hence we sought to provide a novel dimension to the social interaction at the conference by providing people with a basis to expose their interests, both professional and personal, and see those of others at the conference. Central to this idea is that these profiles can be built automatically. Uses can log in to the application web site to browse and edit the inferred interests, deciding what they want to expose to other participants.

The use of semantic repositories and formats made it relatively easy to automatically integrate heterogenous data from our various sources at run time. All data was represented using an agreed ontology and URI syntaxes, and sent to a triplestore to be put together and networked spontaneously.

C. Real-world face-to-face communication

In order to mine the real-world interactions of conference attendees, we deployed the hardware and software infrastructure developed by the SocioPatterns project [7]. The name badges of those attendees who volunteered to become users of the application were equipped with active RFID badges. The RFID badges engage in multi-channel bi-directional radio communication, and by exchanging low-power signals which are shielded by the human body, they can reliably assess the continued face-to-face proximity of individuals, as illustrated in Fig. 1. We assume continued face-to-face proximity to be a good proxy for a social interaction between individuals.

The real-world proximity relations are relayed from RFID badges to RFID readers installed in the conference venue. The readers encapsulate the RFID packets into UDP packets and forward them over a local Ethernet network to a central server. There, the UDP packets from RFID badges are aggregated and fed to a post-processing server that builds and maintains a real-time graph representation of the proximity relations among the tagged attendees. This instantaneous

[5]http://www.rkbexplorer.com
[6]http://dbpedia.org

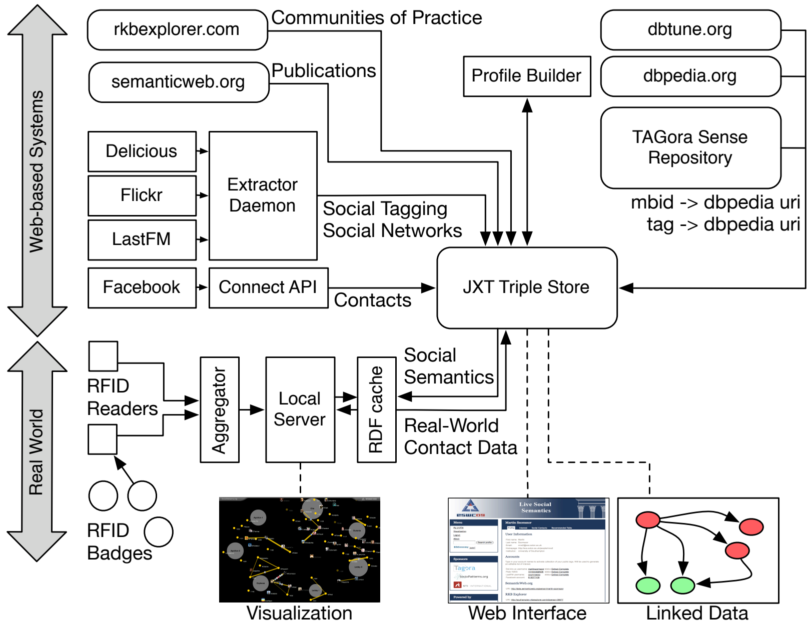

Figure 2. The general architecture of the Live Social Semantics application

contact graph is represented as a time-dependent adjacency matrix A^t_{ij}, such that $A^t_{ij} = 1$ if individuals i and j are in contact at discrete time t, and $A^t_{ij} = 0$ otherwise. The adjacency matrix was updated every 5 seconds.

The post-processing server also maintains a weighted graph representation of their cumulative proximity relations of the tagged attendees over time. The (normalized) adjacency matrix of the cumulative graph during the time interval $[t_1, t_n]$ is defined as $C_{ij}(t_1, t_n) = (1/n) \sum_{k=0}^{n} A^{t_k}_{ij}$. The matrix element $C_{ij}(t_1, t_n) \in [0, 1]$ is the fraction of application time that individuals i and j spent together. Periodically, the cumulated proximity graph is thresholded, and those relations for which $C_{ij} > C_0$ are represented as a set of RDF triples describing the cumulated real-world proximity of attendees, and periodically uploaded to the triple store via RDF / HTTP. A custom *Contact* ontology[7] is used to represent social interactions.

The above graph representations are the central objects describing the real-world component of the application. The instantaneous proximity graph, in particular, is fed to the visualisation clients and to other interactive services available at the conference venue (reactive installations, for example). The real-world proximity relations of the instantaneous proximity graph are mashed up by the server with the web-based attendee relations that it periodically pulls from the triple store. This allows the visualisation clients to display real-world relations in the context of their on-line counterparts.

Moreover, the post-processing server uses the real-world and web-based relations to compute simple recommendation schemes. For example, if two attendees are in contact at a given time, the server provides access to those attendees who may not be present at the same time, but are nevertheless connected to said two users in one of the web-based social networks covered in the application. The visualisation clients (specifically, the user-centered views) can then use this information to enhance the presented information and support browsing of the social network. A simple recommendation scheme is provided by closing triangles that have one social edge grounded in (current or cumulated) physical proximity, and two edges grounded in on-line relationships: this allows to point co-present attendees to a third person who is not present but has on-line connections to both attendees.

D. Visualisation

Two kinds of real-time visualisations were provided. The first, the *spatial view* (figure 3), was publicly displayed on large screens in the main lobby area. The second, the *user focus view* (figure 4), was accessible by means of a web browser on the conference LAN, and is linked to from each user's account page on the application site. Both are dynamic visualisations driven by regular updates received through a TCP socket connection with the local post-processing server.

Spatial view: This view provides an overview of the real-time contact graph. It represents the RFID-badge wearing participants within range of the RFID readers, as well as ongoing social contacts (see section III-C). Each participant is represented by a labelled yellow disc or, when available, by the Facebook profile picture. The contacts are represented by thick yellow edges, whose thickness and opacity reflects

[7] http://tagora.ecs.soton.ac.uk/schemas/LiveSocialSemantics

the weight of the contact. The edges are decorated, where applicable, with small Facebook, Flickr, Delicious, LastFM or COP (community of practice) icons, marking the presence of that relationship in the respective network. (e.g. Facebook friend, Delicious fan, member of same COP). This approach constitutes a projection of said networks onto the real-time contact network.

The SocioPatterns project is primarily concerned with the real-time detection of the contact topology. The precise localisation of the participants in the physical space is of lesser concern. However, a coarse-grained localisation of the participants with respect to the RFID readers is possible. This enabled us to not only represent the contact topology, but also give an indication of which area the participants are in. To this end, the RFID readers were represented by labelled grey shapes, equiangularly laid out on a circumcentric oval, and the participants' shapes are positioned near or in between the readers' marks they are close to. This approach adds spatial structure to the contact graph representation.

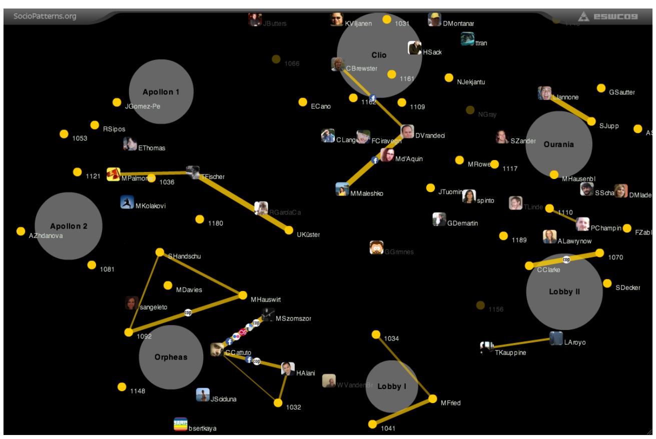

Figure 3. Screenshot of the spatial view grabbed during a session.

User-focus view: This view displays the social neighbourhood of the focussed upon participant. It represents all participants with whom this user has ongoing contact or had significant (cumulative) contact with so far. All physical (proximity-based) interactions between these participants are shown as edges, the current ones in yellow, the historical in grey.

This view furthermore attempts to *close relevant triangles*, by which we mean that all participants that are in some way linked to both the focus participant and any of the initially included participants (i.e. those with whom the focus participant has or had contact), are also included, as well as the concerned links, decorated with the relevant icons like in the spatial view. The objective was to provide the users, after focussing upon themselves, with an overview of that subsection of their social neighbourhood that is relevant for their networking activities at that moment.

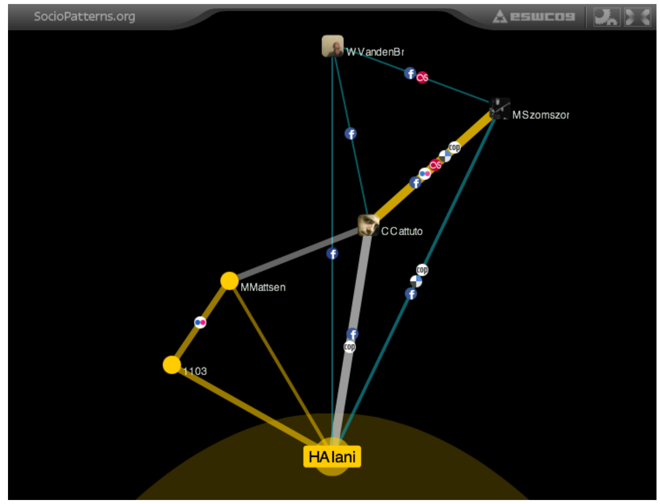

Figure 4. User-focus visualisation in which *HAlani* has the focus. He has ongoing contacts with *MMattsen* and an anonymous user with badge id *1103*, as indicated by the yellow edges. These two users are also in contact, and they are Flickr friends as indicated by the yellow edge and the Flickr icon that decorates it. There has been significant contact between HAlani and *CCattuto*, as indicated by the thick grey line. They are also Facebook friends and share a COP. Both WVandenBr and MSzomszor were included in order to close relevant triangles. The cyan coloured edges indicate that the users are (only) linked in one or more of the social or COP networks.

E. Privacy

Permission was sought from all participants for collecting and using their data. A form was prepared which explained what the data is, how it was going to be used, and for how long. Users were shown how the RFID badges are used, and the geographical limits of where their face-to-face contacts can be detected (conference building). When creating an account on the application site, each user was given the option of destroying their data after the end of the event. As explained in Section III-B, a Profile Of Interest (POI) was generated for each user who declared an account in any of the tagging systems we supported (Delicious, Flickr, and LastFM). To ensure that the users are happy with those interests to be viewed by others, each user was asked to verify and edit their list of interests. These profiles only become visible to other users once their owners verify, edit, and activate them. As an extra security, all data from the RFID badges were encrypted to ensure that they could only be processed by our systems. All the data gathered by the application were stored in a private triple store.

IV. Results

In this section we will report on various results of the two deployments at ESWC09 and HT09: the numbers of participants, their shared networking accounts, interest profile generation, RFID usage, and privacy outcomes.

threshold	fraction of pairs	average degree
1 min	17.1%	14.9
2 min	11.4%	10.2
5 min	5.5%	5.4
15 min	1.7%	2.6
30 min	0.5%	1.4
60 min	0.1%	1.3

Table I
PROPERTIES OF THE CUMULATIVE CONTACT GRAPH

A. Participation

Out of the total of 455 attendees of ESWC09 and HT09, 300 of them took part in Live Social Semantics. Out of these 300 users, 226 of them created an account on the application site, Hence about 25% of the users who collected an RFID badge did not submit any information about themselves (e.g. name, email, social network accounts). Face-to-face contacts of such users were captured, but were not associated with any personal profiles.

B. Social networking accounts

The application site allowed users to declare their accounts on Delicious, Flickr, LastFM, and Facebook. For a total of 226 registered participants, 126 *Facebook* accounts, 87 *Delicious* accounts, 83 *LastFM* accounts, and 75 *Flickr* accounts were declared. 82 Users, or 36% of the 226 registered users, did not declare any social networking accounts, 54 declared one, 44 declared two, 21 declared three, and 24 declared four.

C. RFID results

Data from RFID badges were taken for a continuous interval of about 80 hours for ESWC09 and 72 hours for HT09, fully covering both conferences. Proximity data allows us to measure the duration of face-to-face interactions between conference attendees. By aggregating these data over a suitable time window (for example, from the beginning of the event up to the current time) we build a weighted graph representation of the real-world interactions of conference attendees. As an example, in Table I we report the observed fraction of possible pair-wise contacts that involved face-to-face proximity for a time longer than a given threshold. As expected, the cumulative contact graph is dominated by contacts of short duration, and the introduction of a threshold on contact significance makes the graph more and more sparse as the threshold increases. Notice that the high heterogeneity of the contact graph makes it impossible to choose a single threshold for the significance of social contacts. Table I also reports the number of distinct persons that an attendee spent face-to-face time with, as a function of the same contact duration threshold. A more in-depth analysis of the RFID data is reported in Ref. [8].

D. Privacy results

Naturally, privacy is always a concern in such contexts, where personal data is being collected and processed in various ways. As explained in section III-E we took various measures to secure the data and protect privacy, even though most of the data we were gathering was actually in the public domain (e.g., shared tags). Some participants asked for the data to be kept without any anonymisation, to be stored for reuse in coming events, and even to be published so they can link to their profiles and contacts logs from their websites and blogs. On the other hand, some participants were only prepared to take part if the data is anonymised. At ESWC09, users were given two options in the Terms & Conditions form (section III-E) when they registered on the application website: (1) to permit their data to be retained in an anonymized form, or (b) to destroy their data after the conference. 61% of the participants were happy for their data to be kept, while 39% requested the destruction of their data. Sharing data with other participants did not prove to be an issue for most conference attendees. The numbers in section IV-B indicate that the majority were happy to share their social networking accounts.

V. DISCUSSION AND FUTURE WORK

The deployment of the application at ESWC2009 was the first where all components were put together and a good number of participants got to use it. We observed quite a few technical and sociological issues, which we discuss in the following.

The number of available social networking sites on the web is always on the increase, and the popularity of such sites is never constant. In our application, only four of such networking systems were taken into account. Although the ones we selected are currently amongst the most popular ones, several users wished to add other accounts, such as FOAF files and LinkedIn. One approach to increase extendibility and coverage is to use an open architecture to allow external parties to develop and plug applications and services to connect to, and crawl data from, other networking systems, or sources such as FOAF files.

We implemented a privacy and data retention policy in which we pledged to anonymise the resulting data set, and allowed participants to request the complete removal of their data after the end of the event. This approach introduced a number of inconsistencies and ambiguities. Many users expressed their interest in acquiring their data after the conference. However, the anonymisation actually precludes that. Other issues, such as whether a participant holds the right to access information on recorded contacts with participants that choose to have their data fully removed, points to the need to reconsider our privacy and data retention policies for future deployments. We believe it would be important to allow people to retain all their data, including user accounts, profiles, contact logs, etc. This will not only enable them to access their activity log, but it also allow them to carry their accounts across conferences where this application is deployed.

The visualisations were popular points of attraction during the conference. They gave an accurate real-time representation of the social interactions during the conference. People were often coming to those displays, searching for their colleagues, session chairs, organisers, etc. One could further extend these visualisations by encoding the roles of people at the conference, e.g. highlight conference organisers, session chairs, authors. Furthermore, one could introduce support for Twitter, both as another source of on-line social links and as a way of providing additional conference-related content in the visualisations.

In the application deployment reported here, recommendation of attendees based on physical proximity and on links in social networking systems was performed by means of the simple scheme of section III-C. One could consider ranking schemes for suggested attendees, to make the recommendation more useful and serendipitous. Specifically, a representation of the context of the conference and of the attendees' interests can help in ranking suggested social connections in terms of their predictability based on the conference context.

More services will be provided in future application runs, such as a 'search for person', 'I want to meet', and 'find people with similar interests'. Data from RFIDs can be used to identify 'best attended session or talk'. Social contacts from social networking systems and COPs could be used to find out who has made new contacts, especially if we can compare data over several application deployments.

VI. Conclusions

The Live Social Semantics application was a demonstration of how semantics from several different sources can be harnessed and used to enhance the real-world interactions of people at a social gathering. In particular, the combination of semantic data from social media with the real-world encounters of attendees provides a new way of connecting to people, both in the real world and on-line.

Exposing real-world encounters in digital form facilitates mining interesting and serendipitous social connections, and greatly facilitates the process of establishing new on-line connections to encountered people. On the other hand, connections in social networking systems such as Facebook can be used to stimulate real-world encounters on the basis of shared acquaintances and interests. All of these opportunities were explored by the participants of ESWC09 and HT09, and their reactions, observations, and responses provided valuable input on the future evolution of the platform.

In general, this application goes in the direction of making the co-evolution of real-world social networks and on-line social networks more transparent to users, more lightweight, and more usable. The Live Social Semantics application also provided a first opportunity to expose the semantics of social encounters, and investigate recommendation schemes in bodies of data that mix links from social media with links from real-world encounters.

Acknowledgements

This research was supported by the SocioPatterns.org project and by the TAGora project (Future and Emerging Technologies program of the European Commission, contract IST-34721). The authors acknowledge technical support from Milosch Meriac and Brita Meriac from Bitmanufaktur.

References

[1] R. Borovoy, F. Martin, S. Vemuri, M. Resnick, B. Silverman, and C. Hancock, "Meme tags and community mirrors: Moving from conferences to collaboration," in *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*, 1998.

[2] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *WDTN '05: Proc. 2005 ACM SIGCOMM workshop on Delay-tolerant networking*. New York, NY, USA: ACM, 2005.

[3] N. Eagle and A. (Sandy) Pentland, "Reality mining: sensing complex social systems," *Personal Ubiquitous Comput.*, vol. 10, no. 4, pp. 255–268, 2006.

[4] P. Thibodeau, "IBM uses RFID to track conference attendees," 2007, http://pcworld.about.com/od/businesscenter/IBM-uses-RFID-to-track-confere.htm.

[5] A. Scherrer, P. Borgnat, E. Fleury, J.-L. Guillaume, and C. Robardet, "Description and simulation of dynamic mobility networks," *Comput. Netw.*, vol. 52, no. 15, pp. 2842–2858, 2008.

[6] E. O'Neill, V. Kostakos, T. Kindberg, A. F. gen. Schieck, A. Penn, D. S. Fraser, and T. Jones, "Instrumenting the city: Developing methods for observing and understanding the digital cityscape," in *Ubicomp*, ser. Lecture Notes in Computer Science, P. Dourish and A. Friday, Eds., vol. 4206. Springer, 2006, pp. 315–332.

[7] A. Barrat, C. Cattuto, V. Colizza, J.-F. Pinton, W. Van den Broeck, and A. Vespignani, "High resolution dynamical mapping of social interactions with active rfid," 2008, http://arxiv.org/abs/0811.4170.

[8] H. Alani, M. Szomszor, C. Cattuto, W. Van den Broeck, G. Correndo, and A. Barrat, "Live social semantics," in *8th International Semantic Web Conference (ISWC)*, October 2009.

[9] K. Mller, T. Heath, S. Handschuh, and J. Domingue, "Recipes for semantic web dog food - the eswc and iswc metadata projects," in *Recipes for Semantic Web Dog Food - The ESWC and ISWC Metadata Projects*, Busan, Korea, 2007.

[10] H. Glaser, I. Millard, and A. Jaffri, "Rkbexplorer.com:a knowledge driven infrastructure for linked data providers," in *Proc. European Semantic Web Conference*, Tenerife, Spain, 2008.

[11] M. Szomszor, H. Alani, I. Cantador, K. O'Hara, and N. Shadbolt, "Semantic modelling of user interests based on cross-folksonomy analysis," in *Proc. 7th Int. Semantic Web Conf. (ISWC)*, Karlsruhe, Germany, 2008.