Journal of Internet Services and Applications manuscript No.
(will be inserted by the editor)

Middleware for Social Computing: A Roadmap

Licia Capra · Daniele Quercia

Received: date / Accepted: date

Abstract Social computing broadly refers to supporting social behaviours using computational systems. In the last decade, the advent of Web 2.0 and its social networking services, wikis, blogs, and social bookmarking has revolutionised social computing, creating new online contexts within which people interact socially (social *networking*). With the pervasiveness of mobile devices and embedded sensors, we stand at the brink of another major revolution, where the boundary between online and offline social behaviours blurs, providing opportunities for (re)defining social conventions and contexts once again. But opportunities come with challenges: can middleware foster the engineering of social software? We identify three *societal* grand challenges that are likely to drive future research in social computing and elaborate on how the middleware community can help address them.

Keywords Social Computing · Middleware · Ubiquitous Computing

1 Introduction

The advent of Web 2.0 and it social networking services, wikis, blogs, and social bookmarking has created new digital spaces for social interactions, whereby people easily create, gather, process, use, and share a variety of information (e.g., text, pictures, music and video streams) within their individual social circles. It is estimated that an average of 30 billion pieces of content are shared on Facebook every month, with a 40% projected growth in data per year [26].

With the advent of new mobile devices that are increasingly more powerful, networked, sensory rich, and ubiquitous (5 billion mobile phones were in use in 2010, 12% of which were smart-phones, growing at a rate of 20% per year), the boundaries between online and offline social worlds are blurring. If online (web-based) social computing was centered around social *networking* services (e.g., Facebook, Last.fm, Twitter, MySpace) and the sharing of user-generated content within users' *individual* networks, ubiquitous social computing is going to enable *societal* services, where people' actions and dealings will be looked at, in relation to their impact on *common* welfare. At the heart of this transition is the ability to access much *broader* and *bigger* amounts of data, linked to the individuals and the society of which they are the fabric: for example, RFID-based smart-cards give a fine-grained picture of how public transports are being used, with consequent assessment of our impact on society in terms of CO_2 emissions; positioning technology in smartphones offers a detailed record of our movements within a urban setting, with consequent assessment of urban design qualities (e.g., access, enclosure) [2].

The ability to access big and varied amounts of data will result in the development of novel social computing services that will benefit both the individual and society at large (e.g., it is estimated that big data has a potential annual value to the US health care system alone of $300 billion [26]). However, to create value from

This work has been partly supported by the European Community's FP7 (grant agreement 234239) and by the RCUK (Horizon Digital Economy grant EP/G065802/1)

L. Capra
Dept. of Computer Science, University College London
Gower Street, London WC1E 6BT, UK
E-mail: l.capra@cs.ucl.ac.uk

D. Quercia
Computer Laboratory, University of Cambridge
15 JJ Thomson Avenue, Cambridge CB3 0FD, UK
E-mail: dq209@cl.cam.ac.uk

big data, fundamental technical challenges will have to be overcome, for example, in terms of data gathering, processing and sharing.

We first analyse three *social* grand challenges that will likely drive research in the area of social computing (Section 2). We then take a *technical* standpoint, identify the main common threads that transcend these social challenges (Section 3), and propose a research agenda for middleware researchers in support of future social computing applications (Sections 4-6).

2 Social Computing Challenges

The evolution of mobile and ubiquitous technology is creating new opportunities for entire new classes of social computing services. In particular, we identify three main areas that we believe will attract major attention in the coming years: services to *make the world more sustainable* (Section 2.1), services to *promote individual well-being* (Section 2.2), and services to *create a fair digital ecosystem* (Section 2.3).

2.1 Making the World Sustainable

The share of the world's population living in cities has recently surpassed 50 percent. By 2025, we will see another 1.2 billion people living in cities. The world is in the midst of an immense population shift from rural areas to cities, not least because urbanisation is powered by the potential for enormous economic benefits. Economies of scale make concentrated urban centres more productive than rural areas [4]: "clean water and education, for example, can be delivered for 30 to 50 percent less in Indian cities than in rural areas" [14].

Those benefits will be only realised, however, if we are able to manage the increased complexity that comes with larger cities. Rapid urbanisation is currently contributing to the scarcity of vital resources in cities - of energy supply, road capacity, water reserves, and clean air. Without skillful management of resources, cities become centres of decay, crime, urban sprawl, and pollution. However, the decline of weakly managed cities is not unstoppable. Cities can move decisively to tackle resource scarcity by investing in smart urban infrastructures, in which buildings, power lines, gas lines, roadways, and cell phones are all networked together. Wiring cities can neatly improve efficiency, for example, by exposing hidden patterns of waste. As more information becomes available to both city dwellers and businesses, decisions that will make better use of resources will be enabled. The promise is that, by allocating resources more efficiently and offering new urban services, cities will reduce costs, be ready to transition to low-carbon economies, create sustainable environments, and ultimately enhance the citizens' well-being.

Not only old cities are being made smarter, but entire new smart cities are nowadays built from scratch in a matter of a few years - they are often named "instant cities" [23]. The best known example of an instant smart city is Songdo International Business District. This is a new city near Seoul that Cisco System equipped with advanced sensors. It has been designed to be the greenest and most energy-efficient city in the world. It deploys the state-of-the-art in sustainable technologies: in and out flows, whether water or refuse, are measured, monitored, and accordingly managed. The goal is to use 30 percent less water than a city its size, and save 75 percent of its trash from landfills. Cisco has signed deals with additional instant cities in India, Qatar, and Saudi Arabia. Smart city services will thus receive enormous attention in the coming years and promise to contribute to a sustainable world, one in which 6.5 billion people can all enjoy the highest living standards, without penalties to the planet.

2.2 Promoting Individual Well-Being

For humans, simple questions like 'How much money do I spend in a day' or 'What makes me feel happy' are often hard to answer. That is because people have a poor sense of time and cannot reliably remember certain things [51]. Without good time/memory calibrations, people make erroneous judgments and, as such, do not see the consequences of their actions. One way of reducing erroneous judgments is to gather and analyse data, as usually done in the realms of science, business, and (enlightened) governmental sectors.

Only recently, numbering things has entered the realm of personal life [45]. A tiny minority of people (mostly geeks) have started to quantify private aspects. Alexandra Carmichael, for example, tracks 40 things about herself daily, including mood, chronic pain levels, and sexual activity. Since 2004, the philanthropist and entrepreneur Terry Paul has been working on a device that tracks the number of conversational exchanges a child has with adults. The device promises to monitor and predict language development. Former advertising executive Jon Cousins, after being diagnosed with bipolar affective disorder, developed a mood tracking software supplemented by human sympathy in that it automatically sends emails with mood-tracking scores to a few selected friends. Importantly, these examples of self-tracking are not isolated cases. The MedHelp Internet forum reports more than 300,000 new personal tracking projects every month, and the sites of Quantified Self in

USA[1] and HomeCamp in UK[2] host a large number of personal data projects in a variety of areas. Researchers in human-computer interactions have started to take notice as well, and a new discipline called 'Personal Informatics' has emerged. This discipline studies not only self-tracking, but also the corresponding decision-making processes [31].

The future is charged with great potentials. Personal informatics projects enable not just objective research on human subjects in general, but also an understanding of oneself. Furthermore, they allow medical practitioners to appreciate the particulars of one patient's condition and, in doing so, they promise to support two important trends in health care [46]. One is about health care delivery: delivery is becoming more collaborative, and physicians are starting to be seen as advisers within a co-diagnosis and co-care model between themselves and their patients. The second trend is personalised medicine, in which an individual's specific biological characteristics are used to tailor therapies, including drugs and drug dosage.

2.3 Creating Fair Digital Ecosystems

An important aspect, which is orthogonal to the classes of social services discussed above, is that the corresponding systems are not just technical ones but, crucially, *socio-technical*. On one hand, people will use these systems according to norms that will inevitably vary across societies, and we cannot control nor enforce them by law. On the other hand, technology will profoundly alter the control we have over our own identity, giving us access to an unbounded collection of digital records of every single aspect of our lives. In building a fair digital ecosystem, designers will thus face two main challenges: first, how to build digital systems in a way that healthy social norms emerge; and, second, how to regulate access to our digital lives.

To promote the emergence of healthy social norms, system design is crucially important. The way a new system is designed partly impacts which social norms emerge in it [10]. However, once settled, social norms are hard to change, and when companies tell people how they must behave, things go terribly wrong. That is because being forcibly told how to use a service is perceived as a sign of disrespect by users, and disrespect has often caused violence in physical societies [50] and, for now, only public outcries in digital systems. A case in point is Google's launch of a social media service called Google Plus; most of its early adopters were using their real names, but a few were not. Google decided to go after those few with a heavy-handed regulatory policy to enforce the use of real names, which has caused public outcries that are threatening the very existence of the service [11].

In addition to respecting established social norms, a fair digital ecosystem requires mechanisms with which people can regulate their digital identities. The fact that the Internet never forgets is threatening our ability to control our identities. Before the digital age, remembering was costly and hard, and the default for humans was to forget. Forgetting is a good thing for a society, not least because people are willing to engage (they do not fear the recall of trivial past deeds) and take better decisions (forgetting allows human decision-making to generalise and abstract from individual experiences) [28]. In the digital age, the balance has been inverted: remembering is cheaper and easier than forgetting.

All around the world, policy makers and scholars have run campaigns to promote control of our identities in a digital world that never forgets. A "constitutional right to oblivion" campaign was launched by the French data-protection commissioner Alex Türk; a "reinvent forgetting on the Internet" campaign has been started by the Argentinian writers Alejandro Tortolini and Enrique Quagliano; and "Think B4 U post!" campaign was financed by the European Union to urge young people to consider the "potential consequences" of publishing photos of themselves or their friends without "thinking carefully". However, these campaigns are not definite solutions to the problem of privacy. Users might well 'think carefully' about what to share and what not to, based on reasonable expectations. The problem is that unexpected inferences can often be made from seemingly innocuous social media data. Crandall *et al.* showed that, from publicly available geo-referenced Flickr pictures, one is able to infer several coincidences (e.g., two people taking picture at the same place and at the same time). These coincidences, in turn, reveal "who befriends whom" [7]. The simple act of uploading few pictures on a social media site translates into implicitly and unknowingly disclosing one's private social contacts. Another example is offered by the site *pleaserobme.com*. By combining data from Twitter and Foursquare (a service that lets people share their location so their social contacts can see where they are), *pleaserobme.com* exposes whether users are somewhere other than their home to the entire Internet community, including to burglars. As a final example of unexpected inferences, consider that, from publicly available Twitter profiles (including from privacy-protected ones), one could even infer their users' psychological personality traits [36].

[1] http://www.kk.org/quantifiedself
[2] http://homecamp.org.uk

3 Technical Challenges around Big Social Data

Undermining the realisation of social computing services and applications that address the three grand challenges discussed above, lie the following common technical challenges:

Gathering Social Context. Social computing services require the continuous collection of very fine-grained digital records, both in time and space, of individuals' context and, once aggregated, of the society within which they are embedded. A fundamental question arises as to what constitutes context in these new settings; for example: (1) passively-sensed data (e.g., CO_2 emissions), required to build green computing services to make the world sustainable; (2) user-generated data (e.g., users' tweets), required to build personalised profiling services in support of individual well-being; (3) service usage patterns (e.g., befriending people on Facebook), required to monitor the social norms that emerge when technology is put out of the lab and into society, in order to promote fair digital ecosystems. As context assumes new meanings, novel abstractions and algorithms will be required in support of its gathering.

Democratising Social Software. A distinguishing feature of social computing services is their target user base, that comprises not just a selected few (experts), but virtually all citizens at large. These services should thus be accessible to the people whose lives they affect; this will require, for example, the ability to transform large and heterogeneous raw data streams into knowledge that is presented back to individuals and society, ultimately enabling decision making that is based on data-driven facts rather than the vagaries of human intuition. Also, technologies that work in one social context (e.g., in one city) might not be desirable in others, or might have to be dramatically reworked. New tools that support rapid prototyping and engage citizens in a collective upgrading and problem-solving dimension will thus have to be developed. For example, in the context of smart cities, researchers have been advocating tools that support an"open-source network, where instead of simply having IT workers detect and fix software and code problems as they see them, there would be a collective upgrading and problem-solving dimension involving citizens, a sort of open-source urbanism" [43]. Having those tools in place, social computing services will truly be at the service of their users - and not the other way around.

Governing Social Data Access. In this new digital ecosystem, where all citizens contribute to social knowledge and take responsibility for social actions, key questions arise as to who owns the data being collected, and who owns the services that are being provided based on such data. With so much data linked to the individuals who create and gather it, special care has to be taken in managing access to it, to make sure people will contribute to the digital ecosystem, without, for example, sliding into smart cities where "sensored" becomes "censored" and without contributors fearing their privacy being violated. New frameworks for privacy reasoning and enforcement will have to be developed.

In the next sections, we provide a research roadmap for each of the above technical challenges, briefly discussing what has been achieved so far and, crucially, what remains to be attained by the middleware community.

4 Gathering Social Context

To create an accurate digital footprint of individuals and society, data has to be gathered, both implicitly via sensors, and explicitly via user-generated content. Such data sources are technologically highly heterogeneous. It is a middleware goal to provide application engineers with the right *context* abstractions and associated runtime to instrument the collection of contextual data. Research conducted in the area of context-awareness has tried to tackle a similar goal: starting from [42], where 'context widgets' where first introduced to enable sensing the presence and activity of people, a decade of work has followed. One stream of research has focused on one specific aspect of context, that is, location (e.g., [38], [1]), partly driven by the preponderance of this context facet in many mobile and ubiquitous applications. Another stream has focused on supporting sensing efficiency instead (e.g., [40], [41]), recognising the impact that such task has on local resources (mainly battery) when using mobile phones as sensory platforms. In [34], the two streams of research have come together, using a middleware platform that is capable of providing accurate location sensing, whilst not compromising user experience (that is, battery lifetime).

Work on context-awareness has so far focused on one specific interpretation of context, that is fundamentally *physical*. For example, location is interpreted as space, and thus captured by a pair of coordinates accurately identifying a point in a physical environment. This interpretation of context is very well-suited for a certain class of applications (e.g., car navigation systems), but it becomes rather sterile for social computing. In this domain, for example, location is not just a point in space, but a place to be, defined not only by geographical coordinates, but crucially by the activities we conduct there, and the people with whom we do them [47]. Social computing thus calls for a novel *social* interpretation of context, with new abstractions required to define what is context and how it should be sensed and inferenced. Early work that combines physical data (gathered from sensor-enabled mobile phones) and social data (collected from social networking applications such as Facebook and MySpace), has been presented [30], with the specific goal to sense a user's activity (e.g., being in the gym, in a conversation). Other social facets of context have gone largely unexplored: affective information, for example, is vital for the construction of social computing applications, especially in support of individual well-being, and yet only foundational work on emotion sensing has been conducted thus far [37]. Furthermore, semantically-enriched information about the social network within which a user is embedded (as put forward by the concept of 'circles' in Google Plus[3]) should also be exploited, to define more effective data gathering and inferencing algorithms. To provide application engineers with a 'Social Context Toolkit', middleware researchers should provide a more comprehensive definition of context, together with abstractions and algorithms to sense, relate, monitor and adapt to heterogeneous data streams.

A distinguishing feature of social computing applications, both those aiming to make the world more sustainable and those aiming to achieve individual well-being, is that data gathering is a *continuous* act. In most cases, such act will be performed by battery-powered devices (e.g., smart-phones that people constantly carry with them, sensors embedded in buildings and the environment). Middleware algorithms governing data gathering will thus have to strike the right balance between *accuracy* and *efficiency*. Research in the area of mobile resource management has produced profiling services that gather detailed information about how local resources (especially battery) are being consumed (e.g., [35] [39]), thus offering key information upon which to base run-time adaptation of the sensing act itself (e.g., [16], [24], [20]). These resource management solutions have focused on local, single-device adaptation only. As social computing has re-defined context from a physical entity to a social one, so resource management can be re-defined and take a more social orientation: in [29], for example, inferencing models on co-located mobile devices are pooled together in a sort of mobile cloud, to improve the accuracy of the inferencing engine. A complimentary approach will be to define resource management schemes that load(resource)-balance the act of raw data gathering among the entities that collectively define a social context, for example, by leveraging information about the social network information within which the user is embedded. Indeed, fair participation in data gathering is key not only to provide better user experience (lower impact on local resources), but also to promote social responsibility and participation in an action of collective good.

[3] https://plus.google.com/

5 Democratising Social Software

A common trait of social computing applications is the ability to transform large amounts of data into knowledge upon which citizens can take actions. For example, for green computing applications, one may want to detect historical patterns of CO_2 emissions, so to predict when emissions will exceed safety levels and thus to plan preventive actions; in urban informatics, one may want to study usage patterns of shared bicycle schemes, to quantify the effect that they have on traffic reduction; for quantified-self applications, one may want to see how their fitness levels change over time, in relation to type and amount of exercise, as well as food intake. Developing a social computing application thus requires combining techniques for: data fusion, to integrate data coming from multiple and heterogeneous streams; data analysis and mining, to discover new facts from such data and use them for predictive purposes; machine learning, to match new data against known patterns; and finally knowledge representation, to effectively represent the gathered knowledge back to the user.

At the moment, building a new social computing application is a job for a few, as it requires substantial mathematical expertise to perform any of the steps above. However, citizens with great ideas about useful social computing applications should be empowered to build them (as promoted by the 'Big Society' vision[4]). Middleware practitioners can play a fundamental role in democratizing social software development, by offering libraries and run-time support to conduct the data processing described above. This democratization process is already underway when it comes to a specific, and perhaps more mature, branch of social computing, that is, social networking services: various platforms exist that enable the easy development and deployment of online services containing all key social networking elements (e.g., instant messaging, groups, blogs, music and video sharing, photo albums); although most of these are paid-for platforms (e.g., NING[5], Social En-

[4] http://www.cabinetoffice.gov.uk/big-society

[5] http://www.ning.com/

gine[6], SocialGO[7]), some free open-source platforms are available too (e.g., elgg[8], Dolphin[9], BuddyPress[10]).

The democratization process for social computing software at large is still in its very early days instead. Yet there are many signals that point towards this direction: on one hand, there are efforts to define novel algorithms that can support data analysis geared towards the specific characteristics of social computing applications. For example, in [33], novel adaptive algorithms are being proposed for the automatic identification of community structures in dynamic social networks, with the aim to support the development of more socially-aware networks. On the other hand, novel platforms are being developed in support of data analysis: in [19], for example, novel APIs, data structures, and algorithms have been integrated in a platform to tackle the problem of scalable data analysis of GPS traces; in [5], data stream analytics software is being offered as a cloud service, accessible from mobile devices. Middleware researchers are thus faced with the challenges of developing efficient and effective core data processing techniques, and offer them as services, within a social computing middleware platform, that social application engineers can leverage upon [12]. It is worth noting that knowledge inferred from data streams will be valuable not only for end-users but for the middleware itself: users of social computing applications will exhibit rather different behaviours, as already witnessed on social networking websites like Facebook [15] (e.g., amount of shared content and content quality are not equally distributed across users); different behaviours will result, for example, in different amounts of data being gathered and processed, with direct consequences on QoS parameters, such as network latency and battery consumption. Middleware for social computing should thus dynamically leverage the elicited usage patterns to self-adapt its data gathering and processing services, thus providing higher scalability, robustness and efficiency.

Apart from offering tools for rapid development of social applications, democratising social software requires addressing another major question, that of the ownership of the data that social computing services operate upon. Current social networking applications (e.g., Facebook, MySpace) are structured so that both the data (e.g., your social network of acquaintances) and the services offered on top of such data (e.g., instant messaging, photo sharing) fall under the same company ownership. As a consequence, users are 'locked' with a single service provider, as the burden of duplicating data across providers is excessive (e.g., for picture sharing, it is either Flickr[11] or Picasa[12]; for location-based services, it is either Foursquare[13] or Gowalla[14]). As we move from social networking to social computing at large, a much broader variety of data *and* service types will become available, so a fundamental question arises as to who owns the data being gathered, and who owns the services being offered on top of it. From a middleware architecture point of view, democratizing social computing application development calls for a separation of the data being gathered, from the services being offered back to the users on top of such data. Such a separation was first advocated in [22], where a web ecosystem was envisaged, grounded on a neat decoupling of data from applications, and with users being in full control of who can access such data using, for example, Personal Containers [32] or (virtual) Droplets [9]. Realising such web eco-system would favour the flourishing of new (and better) services, as they would not have to overcome the major bootstrapping cost involved in (social) data gathering, as they could simply hook up the data already available, provided they get consent. So what steps have been taken, and what remains to be achieved?

One stream of research has been investigating the use of communication protocols to enable the confidential sharing of data via social links (e.g., [3] [44] [48]): rather than locking data in the hand of a service provider, users retain ownership of the data and control who to share it with. In a similar fashion, Prometheus [21] suggests the use of a peer-to-peer architecture, to enable access to multiple sources of social data, where access is dynamically granted based on the strength of the social interactions between users. While suitable for certain classes of (social networking) applications, such approaches are unlikely to scale to the volume of big data that is being gathered by social computing applications. An alternative approach in this domain is being offered by the rise of cloud computing, where the abundance of storage, processing capabilities and power can be exploited in support of efficient data storage and processing (e.g., [13], [18]). Two challenges emerge in this domain that are yet to be explored: on one hand, any single social computing service will need to access data owned by different entities (e.g., to provide a more accurate depiction of a community); on the other hand, a single data provider will give access to its data to dif-

[6] http://www.socialengine.net/
[7] http://www.socialgo.com/
[8] http://elgg.org/
[9] http://www.boonex.com/dolphin
[10] http://buddypress.org/
[11] http://www.flickr.com/
[12] http://picasa.google.com/
[13] https://foursquare.com
[14] https://gowalla.com

ferent types of service providers, for different purposes. Middleware researchers will thus have to address issues of data interoperability, both in terms of semantic interoperability (e.g., how data is defined and stored) and transformation/translation (e.g., how different data is combined).

6 Governing Social Data Access

Supporting the development of social computing applications poses fundamental questions in relation to data access. Indeed, the social data being collected provides a very fine-grained picture of an individual's life. Data privacy is a topic that has long been investigated by the research community in mobile and pervasive computing. However, social computing applications magnify the problem in two main dimensions: zoom and span. On one hand, the act of data gathering surrounds the individual continuously, thus offering a very fine-grained digital footprint of an individual's life, from the places visited, to the people met, from the transportation habits to environmental impact caused. On the other hand, the collected data covers a much wider span, both in space and time: that is, sensitive information can be gathered, not only from individual traces, but also from those collectively contributed by our social communities. Furthermore, as the gathered raw data has an almost infinite lifespan (e.g., to enable studying historical trends and patterns), what may not be inferred today could be inferred and revealed tomorrow, as new data comes through. If individuals perceive their privacy being violated, they may rebel and threaten the grounding principle of social computing. A key element of a middleware for social computing is thus a framework for privacy reasoning and enforcement that tackles the above challenges.

One stream of research has focused on supporting *location* privacy for individuals participating in social networking applications. A fundamental realisation, within these scenarios, is that locations are not perceived by users simply as geographic coordinates, but as places within which they conduct social activities [25]; state-of-the-art privacy preserving schemes for location thus cater for different users' privacy requirements as they vary depending on their sociological interpretation of places [49]. We argue that this transition from physical to social interpretation of location should be broadened to the other aspects of context (Section 4) too: for example, water usage and CO_2 emissions are not simply raw numbers; rather, they carry a social significance (e.g., user's environmental impact on the planet) that should be 'exposed' with care, to avoid stigmatising users. Furthermore, the shift from one source (i.e., one user's device) of a single data(location)-stream to many sources (i.e., social community) of multiple data-streams calls for novel approaches to privacy management, with direct support for: (1) reasoning at different granularities of knowledge (from raw data collected by sensors, to inferred states in the knowledge representation hierarchy), and (2) forecasting potential threats that future data, flowing in the system, may cause.

As we move from *social networking* applications towards the broader category of *social* applications, initial work in this direction has been conducted. One stream of research has looked into providing architectural support to enable accurate gathering of social data (beyond location), whilst supporting privacy (e.g., [6], [8]). These approaches have a strong engineering focus (e.g., supporting provable privacy with minimum resource overhead); however, by not placing the human at the center of the privacy-preserving scheme design process, they fail to acknowledge the strong dependence between users' *interpretation* of gathered data, and associated privacy requirements. In [27], a middleware framework that supports privacy reasoning in relation to both raw *and* inferred data is proposed; while promising, this framework is limited in two respects: on one hand, utility theory is used to adapt privacy policies. While good for computer-reasoning, utility functions are cognitively difficult to express, thus their practical applicability must be questioned until users' studies are conducted revealing their suitability, or calling for stronger end-user involvement in the privacy design processes. On the other hand, reasoning has been so far limited to the 'here and now', thus not tackling the privacy concerns that may raise over time, as further data is becoming available and is being processed. As discussed in [28], the digital world has cancelled the natural human ability to forget,"the past is ever present", with threatening consequences in terms of our decision-making processes, which may be based on stale and out-of-context information. As social computing middleware aims to provide computational support of human social processes, then the ability to forget our digital past (initially, in forms similar to the solutions proposed by the Vanish project [17]) must be included in our research agenda in support of social data sharing. More generally, future social computing middleware will need to provide usable tools (e.g., policy languages and frameworks) for the expression and enforcement of data access policies that respect human values and cultures, such as trust, reputation, as well as a right to oblivion.

7 Conclusion

The convergence of social and ubiquitous computing is opening vast opportunities for developing novel services that benefit the individual and the society at large. In this paper, we have been calling for community-wide thinking to solve three "grand challenges" that are likely to drive research in social computing (Section 2). These services, from those in support of a sustainable world, to those aiming at achieving individual well-being, rely on big social data to create value to their end-users. We believe middleware will play a key role in enabling the development of fair digital ecosystems, by tackling fundamental issues that underpin all these classes of applications. More precisely, we have highlighted three main areas that middleware researchers will need to address: new abstraction and protocols required to gather social context data (Section 4); new tools necessary for democratising social software development (Section 5); and new frameworks required to govern access to social data (Section 6). By offering abstractions and services that enable rapid prototyping and deployment, middleware can effectively support the swift deployment of new social computing applications, whose actual use and value can only be assessed once they are out of the laboratory and embedded within actual cultural and geographical contexts.

References

1. Martin Azizyan, Ionut Constandache, and Romit Roy Choudhury. Surroundsense: mobile phone localization via ambience fingerprinting. In Kang G. Shin, Yongguang Zhang, Rajive Bagrodia, and Ramesh Govindan, editors, *Proc. of the 15th Annual Intl. Conference on Mobile Computing and Networking*, pages 261–272, Beijing, China, 2009. ACM.
2. Anil Bawa-Cavia. Sensing the urban: Using location-based social network data in urban analysis, June 2011. The 1^{st} Workshop on Pervasive Urban Applications (PURBA).
3. Marin Bertier, Davide Frey, Rachid Guerraoui, Anne-Marie Kermarrec, and Vincent Leroy. The gossple anonymous social network. In *Proc. of ACM/IFIP/USENIX 11th Intl. Middleware Conference*, pages 191–211, Bangalore, India, 2010.
4. Luis Bettencourt and Geoffrey West. A unified theory of urban living. *Nature*, 467(7318):912–913, October 2010.
5. Qiming Chen and Meichun Hsu. Data stream analytics as cloud service for mobile applications. In Robert Meersman, Tharam S. Dillon, and Pilar Herrero, editors, *OTM Conferences (2)*, volume 6427 of *Lecture Notes in Computer Science*, pages 709–726. Springer, 2010.
6. Cory Cornelius, Apu Kapadia, David Kotz, Daniel Peebles, Minho Shin, and Nikos Triandopoulos. Anonysense: privacy-aware people-centric sensing. In *Proc. of the 6th Intl. Conference on Mobile Systems, Applications, and Services*, pages 211–224, Breckenridge, CO, USA, 2008.
7. D.J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. Inferring social ties from geographic coincidences. *Proc. of the National Academy of Sciences*, 107, December 2010.
8. Emiliano De Cristofaro, Anthony Durussel, and Imad Aad. Reclaiming privacy for smartphone applications. In *Proc. of 9th Annual IEEE Intl. Conference on Pervasive Computing and Communications (PerCom)*, pages 84–92, Seattle, WA, USA,, 2011.
9. J. Crowcroft, A. Madhavapeddy, M. Schwarzkopf, T. Hong, and R. Mortier. Unclouded Vision. In 12^{th} *Intl. Conference on Distributed Computing and Networking (ICDCN)*, January 2011.
10. danah boyd. The Future of Privacy: How Privacy Norms Can Inform Regulation). In 32^{nd} *Intl. Conference of Data Protection and Privacy Commissioners*, October 2010.
11. danah boyd. Designing for Social Norms (or How Not to Create Angry Mobs). http://www.zephoria.org/thoughts/archives/2011/08/05/design-social-norms.html, August 2011.
12. Derek G. Murray, Malte Schwarzkopf, Christopher Smowton, Steven Smith, Anil Madhavapeddy and Steven Hand. Ciel: a universal execution engine for distributed data-flow computing. In 9^{th} *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, 2011.
13. Sabrina De Capitani di Vimercati, Sara Foresti, Stefano Paraboschi, Gerardo Pelosi, and Pierangela Samarati. Efficient and Private Access to Outsourced Data. In *Proc. of the* 31^{th} *IEEE Intl. Conference on Distributed Computing Systems (ICDCS)*, Minneapolis, USA, June 2011.
14. Richard Dobbs and Jaana Remes. What's the biggest limit on city growth? (hint: it's not steel or cement). *McKinsey: What Matters*, February 2011.
15. Hossein Falaki, Ratul Mahajan, Srikanth Kandula, Dimitrios Lymberopoulos, Ramesh Govindan, and Deborah Estrin. Diversity in smartphone usage. In *Proc. of the 8th Intl. Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 179–194, 2010.
16. Roy Friedman and Alex Kogan. Power aware management middleware for multiple radio interfaces. In *Proc. of 10th Intl. Middleware Conference*, pages 288–307, Urbana, IL, USA, 2009.
17. Roxana Geambasu, Tadayoshi Kohno, Amit Levy, and Henry M. Levy. Vanish: Increasing Data Privacy with Self-Destructing Data. In 18^{th} *USENIX Security Symposium*, August 2009.
18. Ioana Giurgiu, Oriana Riva, Dejan Juric, Ivan Krivulev, and Gustavo Alonso. Calling the cloud: Enabling mobile phones as interfaces to cloud applications. In *Proc. of 10th Intl. Middleware Conference*, pages 83–102, Urbana, IL, USA, 2009.
19. Maya Haridasan, Iqbal Mohomed, Doug Terry, Chandramohan A. Thekkath, and Li Zhang. StarTrack Next Generation: A Scalable Infrastructure for Track-Based Applications. In *Proc. of OSDI 2010*, October 2010.
20. Donnie H. Kim, Younghun Kim, Deborah Estrin, and Mani B. Srivastava. Sensloc: sensing everyday places and paths using less energy. In *Proc. of the 8th Intl. Conference on Embedded Networked Sensor Systems (SenSys)*, pages 43–56, 2010.
21. Nicolas Kourtellis, Joshua Finnis, Paul Anderson, Jeremy Blackburn, Cristian Borcea, and Adriana Iamnitchi. Prometheus: User-controlled p2p social data management for socially-aware applications. In *Proc. of ACM/IFIP/USENIX 11th Intl. Middleware Conference*, pages 212–231, Bangalore, India, 2010.

22. Maxwell Krohn, Alex Yip, Micah Brodsky, Robert Morris, and Michael Walfish. A World Wide Web Without Walls. In *6th ACM Workshop on Hot Topics in Networking (Hotnets)*, Atlanta, GA, November 2007.
23. Greg Lindsay and John D. Kasarda. At home in the aerotropolis. *McKinsey: What Matters*, February 2011.
24. Hong Lu, Jun Yang, Zhigang Liu, Nicholas D. Lane, Tanzeem Choudhury, and Andrew T. Campbell. The jigsaw continuous sensing engine for mobile phone applications. In *Proc. of the 8th Intl. Conference on Embedded Networked Sensor Systems (SenSys)*, pages 71–84, 2010.
25. Clara Mancini, Keerthi Thomas, Yvonne Rogers, Blaine A. Price, Lukasz Jedrzejczyk, Arosha K. Bandara, Adam N. Joinson, and Bashar Nuseibeh. From spaces to places: emerging contexts in mobile privacy. In Sumi Helal, Hans Gellersen, and Sunny Consolvo, editors, *UbiComp*, ACM Intl. Conference Proceeding Series, pages 1–10. ACM, 2009.
26. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. Big data: The next frontier for innovation, competition, and productivity. Technical report, McKinsey Global Institute, 2011.
27. Daniel Massaguer, Bijit Hore, Mamadou H. Diallo, Sharad Mehrotra, and Nalini Venkatasubramanian. Middleware for pervasive spaces: Balancing privacy and utility. In *Proc. of 10th Intl. Middleware Conference*, pages 247–267, Urbana, IL, USA, 2009.
28. Viktor Mayer-Schönberger. *Delete: the virtue of forgetting in the digital age.* Princeton University Press. Princeton University Press, 2009.
29. Emiliano Miluzzo et. al. Darwin phones: the evolution of sensing and inference on mobile phones. In *Proc. of the 8th Intl. Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 5–20, San Francisco, California, USA, 2010.
30. Emiliano Miluzzo et. al. Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application. In *Proc. of the 6th Intl. Conference on Embedded Networked Sensor Systems*, pages 337–350, Raleigh, NC, USA, 2008. ACM.
31. Neema Moraveji, Ryo Akasaka, Roy Pea, and B.J. Fogg. The role of commitment devices and self-shaping in persuasive technology. In *Proc. of the ACM Conference Extended abstracts on Human Factors in Computing Systems (CHI)*, pages 1591–1596, May 2011.
32. R. Mortier, C. Greenhalgh, D. McAuley, A. Spence, A. Madhavapeddy, J. Crowcroft, and S. Hand. Personal Containers or, Your Life in Bits. In *Digital Futures Workshop*, October 2010.
33. Nam P. Nguyen, Thang N. Dinh, Ying Xuan, and My T. Thai. Adaptive algorithms for detecting community structure in dynamic social networks. In *INFOCOM*, pages 2282–2290. IEEE, 2011.
34. Jeongyeup Paek, Joongheon Kim, and Ramesh Govindan. Energy-efficient rate-adaptive gps-based positioning for smartphones. In *Proc. of the 8th Intl. Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 299–314, San Francisco, California, USA, 2010.
35. Feng Qian, Zhaoguang Wang, Alexandre Gerber, Zhuoqing Morley Mao, Subhabrata Sen, and Oliver Spatscheck. Profiling resource usage for mobile applications: a cross-layer approach. In *Proc. of the 9th Intl. Conference on Mobile Systems, Applications, and Services*, pages 321–334, Bethesda, MD, USA, 2011.
36. Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcroft. Our Twitter Profiles, Our Selves: Predicting Personality with Twitter. In 3^{rd} *IEEE Conference of Social Computing (Socialcom)*, October 2011.
37. Kiran K. Rachuri, Mirco Musolesi, Cecilia Mascolo, Peter J. Rentfrow, Chris Longworth, and Andrius Aucinas. Emotionsense: a mobile phones based adaptive platform for experimental social psychology research. In *Proc. 12th Intl. Conference on Ubiquitous Computing*, pages 281–290, Copenhagen, Denmark, 2010.
38. Anand Ranganathan, Jalal Al-Muhtadi, Shiva Chetan, Roy H. Campbell, and M. Dennis Mickunas. Middlewhere: A middleware for location awareness in ubiquitous computing applications. In Hans-Arno Jacobsen, editor, *Middleware*, volume 3231 of *Lecture Notes in Computer Science*, pages 397–416. Springer, 2004.
39. Andrew Colin Rice and Simon Hay. Decomposing power measurements for mobile devices. In *PerCom*, pages 70–78. IEEE Computer Society, 2010.
40. Oriana Riva. Contory: A middleware for the provisioning of context information on smart phones. In Proc. of Intl. Conference on Middleware, pages 219–239. 2006.
41. Nirmalya Roy, Archan Misra, Christine Julien, Sajal K. Das, and Jit Biswas. An energy-efficient quality adaptive framework for multi-modal sensor context recognition. In *Proc. of 9th Annual IEEE Intl. Conference on Pervasive Computing and Communications (PerCom)*, pages 63–73, Seattle, WA, USA,, 2011.
42. Daniel Salber, Anind K. Dey, and Gregory D. Abowd. The context toolkit: aiding the development of context-enabled applications. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, pages 434–441, 1999.
43. Saskia Sassen. Talking back to your intelligent city. *McKinsey: What Matters*, February 2011.
44. Valerio Schiavoni, Etienne Riviere, and Pascal Felber. Whisper: Middleware for Confidential Communication in Large-Scale Networks. In *Proc. of the* 31^{th} *IEEE Intl. Conference on Distributed Computing Systems (ICDCS)*, Minneapolis, USA, June 2011.
45. Doc Searls. Eof: Now Data Gets Personal. *Linux Journal*, January 2010.
46. Melanie Swan. Emerging Patient-Driven Health Care Models: An Examination of Health Social Networks, Consumer Personalized Medicine and Quantified Self-Tracking. *Intl. Journal Environ. Research Public Health*, pages 492–525, February 2009.
47. Karen P. Tang, Jialiu Lin, Jason I. Hong, Daniel P. Siewiorek, and Norman M. Sadeh. Rethinking location sharing: exploring the implications of social-driven vs. purpose-driven location sharing. In *Proc. 12th Intl. Conference on Ubiquitous Computing*, pages 85–94, Copenhagen, Denmark, 2010.
48. Jin Teng, Boying Zhang, Xinfeng Li, Xiaole Bai, and Dong Xuan. E-Shadow: Lubricating Social Interaction using Mobile Phones. In *Proc. of the* 31^{th} *IEEE Intl. Conference on Distributed Computing Systems (ICDCS)*, Minneapolis, USA, June 2011.
49. Eran Toch, Justin Cranshaw, Paul Hankes Drielsma, Janice Y. Tsai, Patrick Gage Kelley, James Springfield, Lorrie Faith Cranor, Jason I. Hong, and Norman M. Sadeh. Empirical models of privacy in location sharing. In *Proc. 12th Intl. Conference on Ubiquitous Computing*, pages 129–138, Copenhagen, Denmark, 2010.
50. Richard Wilkinson and Kate Pickett. *The Spirit Level: Why Greater Equality Makes Societies Stronger.* Bloomsbury Press, December 2009.
51. Gary Wolf. The data-drive life. *The New York Times*, April 2010.