The Rebirth of Locality: Information, People and Places in a Connected World

Antonio Lima
School of Computer Science
University of Birmingham, United Kingdom

Mirco Musolesi
School of Computer Science
University of Birmingham, United Kingdom

ABSTRACT

One of the most interesting trends in the recent years has been the re-discovery of spatial locality, i.e., of the *here* and *now*. In a world of global spreading of information, location is assuming a fundamental role again. We can consider it as a sort of "rebirth" of locality.

In this position paper, firstly, we present a brief summary of our current work on models, algorithms, applications and tools based on and for spatially-embedded social networks, i.e., networks where the interacting individuals have a well-defined position in space. Our work is focussed on global phenomena that are experienced at a very local scale: we believe that the availability of real-time information about individuals at local, community and global level can be exploited *together* in order to develop and support novel models, applications and systems. In the second part of the paper we outline the challenges and the opportunities in this promising area, discussing key open problems and research directions.

Author Keywords

Location-Based Social Networks, Big Data, Linked Data, Privacy, Personalisation, Predictive Algorithms.

ACM Classification Keywords

H.1.2 User/Machine Systems: Human Information Processing

General Terms

Human Factors; Measurement.

OVERVIEW AND VISION

Advances in communication were expected to cause the so-called "death of distance", i.e., the end of distance as an obstacle to interactions between people [2]. Today, distant interactions are certainly easier than they were in the past, but we certainly cannot declare the death of distance [10]. We are also experiencing an interesting phenomenon, which is not only related to the distance itself, but with the very fact that people access information *locally*. Online services are gradually moving towards a strong focus on what is *close* in time and space to the user (i.e., *hic et nunc*). Some services are designed to give online suggestions about what the user could do or might be interested in at specific times and locations. Indeed, spatial locality is assuming an increasing importance in socio-technological systems. We may consider this as the *rebirth of locality*.

While users are presented with real-time suggestions tailored to their interests, position and time, on the other hand, they are also actually receiving a continuous stream of data associated to global trends, events, products and opportunities. These streams of information are global but they are accessed and consumed locally. Our work is focussed on global phenomena that are experienced at a very local scale: we believe that the availability of real-time information about individuals at local, community and global levels can be exploited *together* in order to develop new models, applications and systems. These rich streams constitute a unique opportunity for practitioners and researchers as their exploitation can lead to novel applications and services and their analysis can provide unique insights about understanding human behaviour and interactions at individual and collective scale.

We are currently working on the design, implementation and evaluation of models, algorithms, applications and tools based on and for spatially-embedded social networks, i.e., networks where the interacting individuals occupy a position in a space. While this space is usually a geographic space, it does not need to be a physical space, but it can also be virtual, representing non-tangible aspects. For instance, the individuals of an online social network can be positioned in a geographic space, according to their usual hangouts, and in a semantic space, according to their interests. In particular, the use of geo-spatial information in social networks allows to analyse individual and collective properties of the social graph concerning individuals (or groups of individuals) in a region or a set of regions.

For example, in [5] we present a set of measures that quantify the effects of social links on the spreading of information in a given geographic area. Every user is assigned a value that indicates how much potential influence he or she has on a particular region. This measure of influence might be used for different purposes, such as political or marketing purposes and for a variety of applications, from promotional campaigns to the management of mass-events or natural and man-made disasters. Intuitively, influence grows with the number of friends that a user has inside the considered region. Key *local influencers* can be identified using these techniques: it is possible to identify users which can spread messages on *global* media in order to address very specific *local* requirements and situations or to reach well-identified geographic areas.

CHI 2013 Workshop GeoHCI, April 2728, 2013, Paris, France.

Another key area of our research work is the exploitation of mobility information in order to augment the design of mobile and networked systems. Recent results also show that human movement is predictable, both in the short [4] and in the long term [9] and it is characterised by specific patterns [8]. In [4] the authors show that user mobility is potentially more predictable when considering mobility of user's social contacts. A reliable prediction of future user position can also be very helpful in improving the applications discussed above, in marketing and disaster-recovery scenarios, and generally for delivering information when and where is needed most.

CHALLENGES AND OPPORTUNITIES

In this section we summarise a series of challenges and opportunities in this research space, outlining a set of open questions for the research community.

Computational Issues. Data coming from popular online services are usually extremely large; these data are not only produced in large quantity at every instant, but they must also be stored for very long periods of time, i.e., they are in a sense *longitudinally* "big data". Additionally, data about individuals are also generated by sensors embedded in smartphones and in the urban fabric [3]. Therefore, nontrivial distributed computational models and systems must be developed and deployed for these services, in order to support run-time computation. Also, as computational power of portable devices is increasing, there is the possibility of distributing computational tasks to user devices. This is also an opportunity for developing privacy-preserving services: smart devices could process sensitive raw-data, possibly use anonymisation or aggregation techniques, and send less sensitive data to the cloud.

Real-time Input, Real-time Output. The need for short processing times also arises from the nature of these data. In fact, sometimes data come in the form of endless streams, which must be processed quickly in order quasi real-time output. One example is the automatic detection of interesting events [6]. Real-time processing of data is the first step towards real-time self-adaptive systems that respond to sudden changes in the environment. Which architectures and algorithms can support real-time event detection? What are the best ways of providing and visualising this information, also considering the context of the user?

Getting Ahead with Prediction. For some application scenarios, it is possible to do even better than real-time, by using prediction technology that is capable of getting *ahead* of time. Research in this area is still at an early stage, but it is already showing promising results [4, 9]. Applications range from targeted information dissemination, based on forecasted context conditions, to law enforcement, such as the identification of future criminal hotspots.

Privacy. The use of data concerning human behaviour raises non-trivial questions about the way they are treated, especially when they include fine-grained information about user location over time [1]. How sensitive is real-time location? Who has the right to store and process this information and for how long? Is it possible to design anonymisation techniques that preserve user privacy while having a small impact on the functionality of services? Answers to these questions are missing, while services are rapidly evolving and ignoring these issues.

Multi-dimensional Linked Big Data The data generated by online social networks and services can be often represented as a network of interactions between entities [7]. Some entities or interactions that are present in a dataset can actually be linked to entities or relations available in others. This fact opens the way to fascinating possibilities, since these multiple sources of linked data can be used as a single rich dataset. What are the best analytical models to represent these massive datasets? Is it possible to design algorithms that are able to extract information from these heterogeneous data? What are the most suitable algorithms for supporting real-time analysis?

Datasets and Benchmarks for the Research Community. Finally, as strange as it might sound in this data-centric era, there is a strong need for novel dataset sharing models. Companies collecting and storing these large-scale datasets are reluctant to share them for understandable reasons, since they represent a core business asset. As a result, scholars struggle to get datasets by crawling small samples or asking them to companies. Either way, they have to comply with Terms of Services and Non Disclosure Agreements, respectively. This data sharing model might undermine the principle of verifiability of science in this field. It also de-incentivises intra- and inter-disciplinary collaboration, constituting a high entry barrier for newcomers.

ACKNOWLEDGMENTS

This work was supported through the EPSRC Grant "The Uncertainty of Identity: Linking Spatiotemporal Information Between Virtual and Real Worlds" (EP/J005266/1).

REFERENCES

1. Atzori, M., Bonchi, F., Giannotti, F., Pedreschi, D., and Abul, O. Privacy-aware knowledge discovery from location data. In *Proceedings of IEEE MDM'07* (May 2007), 283 –287.
2. Cairncross, F. *The death of distance: How the communications revolution is changing our lives*. Harvard Business Press, 2001.
3. Campbell, A. T., Eisenman, S. B., Lane, N. D., Miluzzo, E., Peterson, R., Lu, H., Zheng, X., Musolesi, M., Fodor, K., and Ahn, G.-S. The rise of people-centric sensing. *IEEE Internet Computing Special Issue on Mesh Networks* (June/July 2008).
4. De Domenico, M., Lima, A., and Musolesi, M. Interdependence and predictability of human mobility and social interactions. In *Proceedings of the Nokia Mobile Data Challenge Workshop.* (Oct. 2012).
5. Lima, A., and Musolesi, M. Spatial dissemination metrics for location-based social networks. In *Proceedings of UbiComp'12*, UbiComp '12, ACM (New York, NY, USA, 2012), 972979.
6. Mathioudakis, M., and Koudas, N. TwitterMonitor: trend detection over the Twitter stream. In *Proceedings of SIGMOD'10*, ACM (New York, NY, USA, 2010), 11551158.
7. Newman, M. *Networks: An Introduction*. OUP, 2010.
8. Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., and Mascolo, C. A tale of many cities: Universal patterns in human urban mobility. *PLoS ONE 7*, 5 (May 2012), e37027.
9. Sadilek, A., and Krumm, J. Far out: Predicting long-term human mobility. In *Proceeding of AAAI-12* (2012).
10. Scellato, S., Mascolo, C., Musolesi, M., and Latora, V. Distance matters: geo-social metrics for online social networks. In *Proceedings of WOSN'10*, USENIX Association (Berkeley, CA, USA, 2010).