L

Large Networks

► Pajek

Large-Scale Network

► Actionable Information in Social Networks, Diffusion of

Latent Relationship, Hidden Role

► Role Discovery

Lattices and Ties

► Corporate Online Social Networks and Company Identity
► Social Network Analysis and Company Linguistic Identity

Layered Social Networks

► Multilayered Social Networks

LDA

► Topic Modeling in Online Social Media, User Features, and Social Networks for

Leadership Team

► Top Management Team Networks

Learning

► Human Behavior and Social Networks

Learning Analytics

► Learning Networks

Learning Networks

Caroline Haythornthwaite
School of Library, Archival and Information Studies, University of British Columbia, Vancouver, BC, Canada

Synonyms

Educational data mining; E-learning; Informal learning; Knowledge Networks; Learning analytics; Networked learning

R. Alhajj, J. Rokne (eds.), *Encyclopedia of Social Network Analysis and Mining,*
DOI 10.1007/978-1-4614-6170-8,

Glossary

Formal, Nonformal, and Informal Learning Formal – structured learning associated with degree-granting institutions and credentials; nonformal – structured learning not associated with credentials, for example, learning for hobbies; informal – unstructured learning, such as learning norms and cultural conventions

Homophily Similarity between actors, for example, in race, gender, education, and attitudes

Adaptive Structuration The negotiation and continuous emergence of practices around technology use and group needs

Socio-Technical Capital Social capital associated with managing and prospering through the use of information and communication technologies

Absorptive Capacity Ability of a network to capitalize on new knowledge

Technological Guru An actor who facilitates the recognition and integration of new knowledge into group practice

Broker An actor who sits between and connects two or more networks or cliques within a network

Structural Hole a gap between networks or cliques that could be spanned by an actor, often associated with entrepreneurial opportunity

Transactive Memory Group management of knowledge through understanding and distributing knowledge across the network

Cognitive Social Structure perceptions of actors about the interrelations among actors in the network

Media Multiplexity the tendency for those with stronger ties to use more media to communicate than those with weaker ties

Latent Tie Structure A technical or social structure that puts actors in virtual or physical proximity thus facilitating the initiation of ties

Definition

Learning networks are social networks that support the acquisition of explicit or tacit knowledge in support of the needs and interests of actors in the network. What knowledge reaches an actor depends on the structure of the network and their position in it as well as on the kinds of relations and ties maintained with others that support seeking, gaining, and trusting knowledge from others. Information and knowledge that can be identified, adopted, and incorporated into practices in a network contribute to the social capital of a network and the success as a learning organization or community of inquiry.

Introduction

Social theories of learning emphasize the way interaction builds knowledge as individuals gain information, receive instruction, negotiate meaning, and try out ideas on others. It is this underpinning of engagement between people in the service of learning that lays the foundation for the examination of learning networks. A social network perspective on learning provides a vocabulary, methodology, and a set of analytic techniques for examination of learning phenomena. This entry outlines how learning can be viewed and analyzed from a social network perspective.

Information, Knowledge, and Learning

Learning is often considered as the transformation that occurs as an individual integrates knowledge, information, and ideas into their own experience and perspective. While some knowledge is acquired through personal exploration, much of our learning happens in social contexts such as schools, workplaces, leisure activities, and in social relations with people in known roles, such as teacher and student(s), trainer and trainee, mentor and mentee, master and apprentice, and guru and novice. This kind of learning fits with both *formal learning*, that is, that which happens through educational institutions and associated with formal credentialing, and with *nonformal learning*, that is, learning of subjects and practices not formally accredited by educational institutions, such as learning for hobbies, sports, and subjects of interest. In other contexts, learning happens with those not necessarily in such hierarchical relationships, such as in collaborative, peer-to-peer learning, and in the *informal learning* that happens as language and social norms are recognized and adopted. Formal learning is more associated with the acquisition of explicit knowledge, "know-what" knowledge that is able to be defined and conveyed by rules and procedures, and is often formally written down. Informal learning is more associated with the acquisition of tacit knowledge, "know-how" knowledge about what to do and how to behave in different circumstances, and how to get work done. All these forms of learning can be found supported and expressed through social networks. Learning from and with others is the focus of this entry, with attention to how we acquire and incorporate knowledge in negotiation with others, receiving from some and giving to others.

Of relevance to learning networks are results from social network studies on how information circulates networks and how knowledge is recognized and adopted into practice as well as what is known about how social network structures support learning organizations and communities. The following reviews some of the key aspects of social networks that bear on learning.

Relevant to all social network relationships is the tie maintained between individuals in the network. Birds of a feather do tend to flock together (Smith-Lovin et al. 2001), and thus, *homophily*, the extent to which individuals are similar in race, gender, education, and other socioeconomic characteristics, strongly predicts who is likely to be associated with whom and thus from whom an individual is likely to hear about new ideas and from whom they will ask for and receive help. As Granovetter (1973) outlined in his research on those seeking new jobs, those with whom we are strongly tied tend to travel in the same social circles as we do and thus are exposed to and able to inform us only about the same information we can access. Weak ties, those with people we know less well, and who are more different from us, tend to travel in different social circles. Information they have access to is different from ours and that of our close friends and, hence, is more likely to be new to us. Thus, the *strength of weak ties* rests in their potential for providing exposure to new information. For learning, the idea of collaborative learning has been described in a way that suggests attention to weak ties, for example, in the strength of this kind of learning in providing access to new ideas, to discuss and work out ideas with those with different experience or knowledge, and to gain diverse perspectives.

One downside of weak ties is that these contacts are not committed to helping us and thus not motivated to share information with us. Strong ties, on the other hand, are motivated to share with us and thus more likely to be willing to share what information and knowledge they have. Results from a study of information seeking and learning by Borgatti and Cross (2003) demonstrated the importance of timely and appropriate access as important for the ability to learn from others. Their study found that access encompasses more than just being able to contact someone, including as well that the contact be willing and able to give the time and effort needed to help the individual gain information and/or have it explained to the point of fluency. A strong tie, whether based on a work or friend relationship, is more likely to be willing to put the time in to help. This, and the likelihood that pairs are in more constant contact, may explain why strong ties have also been shown to be a better source of learning about tacit knowledge (Hansen 1999).

In a formal learning context, Haythornthwaite (2002a, b) found that attention to weak ties for access to a diversity of ideas needs to be balanced with the needs of time-limited groups. Where a project outcome is required, strong ties are important because of the need to determine and follow work norms and to commit to achieving a common outcome. Thus, learning environments, like work environments, require attention to support of contact with both weak and strong ties.

Being accessible and able to share information in a way that matches the needs of the recipient extends the idea of homophily to similarity in knowledge. It is easier, for example, to work and learn with those already well versed in the contexts, techniques, and vocabularies of the application area. As groups and disciplines develop, they create shorthands and specialized vocabularies which serve to reduce the mutual work that needs to be done in exchanging knowledge. This kind of mutual way of speaking provides common ground within the group on which new knowledge can be built. Examinations of practices in interdisciplinary teams have suggested the need for the development of such common ground (Klein 1996) but also highlighted differences in readiness to come to such common understanding (Haythornthwaite 2006; Cummings and Kiesler 2008).

As noted, the kind of knowledge exchanged is not just explicit, "know-what" information. As Orlikowski (2002) identified, groups also gain tacit "know-how" knowledge, including knowledge about how "to be" particular kinds of groups. In her cases, organizational members learned "to be" distributed teams operating as globalized companies; in other cases, researchers learned "to be" computer supported collaborative groups (Haythornthwaite et al. 2006), and e-learners learned "to be" online distributed learners (Haythornthwaite and Andrews 2011). Online groups acquire such knowledge through negotiation, in a process of *adaptive structuration* (DeSanctis and Poole 1994), building what Resnick (2002) has called *socio-technical capital*, created and sustained in a group or community's ability to function, manage, and prosper through contemporary information and communication technologies. Such learning about how to use and work through technical means is becoming increasingly evident in tacit learning by groups and in the assemblages of technology, artifacts, and social relations that define groups and communities (e.g., Girard and Stark 2007).

Knowledge held within a network has important consequences for the ability to recognize and act on new knowledge and opportunities, what Cohen and Levinthal (1990) described as the "*absorptive capacity*" of an organization. A well-known role in such acquisition is the "*technological guru*," an individual whose knowledge is such that they can evaluate new opportunities in light of the purpose and goals of the organization to which he or she belongs. Stepping out a bit further is the *broker*, an individual positioned between networks in a way that facilitates or controls the flow of information between networks. Burt (1992) has articulated the benefits of this position in the way the broker spans a "*structural hole*." While the discussion of this role often pertains to their ability to create market opportunities, many teachers are also in the position of brokers who bridge knowledge held by experts and the needs and learning capacities of students. Knowledge translation and transfer thus is expedited by the right kind of individual positioned in a structural hole relating to knowledge. This position has particularly gained attention in the idea of social entrepreneurship, where the individual can expedite positive social benefit by bridging between knowledge and other resource bases.

While brokers may keep it to themselves that they have certain knowledge, an important part of continuous knowledge learning transfer and management within groups is knowing "who knows what," described as the "*transactive memory*" (Moreland 1999) or the *cognitive social structure* (Krackhardt 1987) of a group (including also "who knows who knows what"; Hollingshead et al. 2002). Borgatti and Cross (2003) add that while knowing who has the knowledge is one part of that process, also important is valuing that person positively as a quality resource, an aspect that is bound up in the experience of previous information seeking and learning encounters with the person. Also important is that knowing another network member is the expert on a subject, others can then pass on resources to them, reducing their load but also appropriately allocating resources across the network for efficient information management.

Contemporary information and communication technologies have opened up consideration of how we trust information received online from unknown others and thus whether and how we incorporate that knowledge into our learning (Kelton et al. 2008). For some, the apparent intimacy of an e-mail has led people to trust where they should not, for example, in believing the contents of spam e-mail. We may put this down to the way some computer-mediated communication can appear to originate from a close tie, combined with the newness of norms and understanding of pitfalls of CMC. We can also see that the "community" of CMC users is learning to filter spam, reject offers of easy money, and warn about exposure of private information.

On a more positive note are the crowds and communities that support the growth of knowledge repositories and conversations, including wikis, virtual communities, online worlds, social networking sites, and microblogging as well as institutional repositories, open-access journals, and open digital libraries. A question arises here about why people will contribute to anonymous, centralized knowledge-building efforts. Raymond (1999) and Benkler (2006) each describe a "personal but shared need" as a motivation associated with open-source development. From a social network perspective, crowd motivation appears to be derived from a coorientation to the general ideas associated with the purpose of the project and/or wider societal issues such as open access; in keeping with what is known about group behavior, community motivation appears to include as well interest in how the group will function and to what purpose (Haythornthwaite 2009; Budhathoki and Haythornthwaite 2013). Thus, we might look to motivation and trust of information from online crowds as more like learning from and with weak ties, and trust in online communities as more like learning from and negotiating with strong ties.

Learning Analytics and Learning Networks

Social network analyses of learning are at their formative stage but increasing with recognition that social network principles can provide new insight into learning practice and outcomes. Attention to automated methods for understanding learning and education is growing with the pace of attention to learning outcome assessment in formal learning settings and the volume of online learning data generated from online learning and from the records of institutions of higher education. This is leading to the extension of the application of social network concepts and data mining to educational environments and the rapid adoption of automated techniques for analysis of education, pedagogy, e-learning, and learning as well as analysis of related concepts such as community.

Current work on learning and social networks combines attention to learning practices of teachers and students with the development of tools and measures relating to learning activities, from participation practices to learning trajectories. Data collection methods include questionnaire and interviews, analyses of online discussion transcripts, linguistic analyses, and methods and tools for automated analysis. New tools for learning environments engage with not only analyzing but also visualizing networks and learning activity for participants (e.g., see the SoLAR Open Learning Analytics concept paper, http://solaresearch.org/OpenLearningAnalytics.pdf).

The following provides a brief overview of some current work in learning analytics and learning networks. Other recent work can be found through new and emerging venues addressing learning analytics and learning networks, such as the Society for Learning Analytics Research (http://www.solaresearch.org/), and associated Learning Analytics and Knowledge conferences. Research to date has applied social network analysis to examine teacher and learner network structures, the relational bases of learning communities, entrepreneurship, creativity, academic performance, and media use.

Maarten de Laat has explored teacher networks (De Laat 2011) and their value creation (Wenger et al. 2011) in order to understand the impact and importance of social relations for professional development and organizational learning. De Laat and colleagues developed a "network awareness" tool that visualizes informal professional learning networks in the workplace. This tool is used as a social learning browser, providing and promoting access to learning networks and providing the users with real-time insight in the development of these networks and the actors involved (Schreurs and De Laat 2012). Other works on teacher networks include a collection by Alan Daly (2010) that brings together social network analyses applied to teacher networks, leadership, and educational change and a major study by Fouad Abd-El-Khalick and colleagues that uses a social network approach to examine adoption and dissemination of entrepreneurial practices among science teachers (Entrepreneurial Leadership in STEM Teaching & Learning, http://enlist.mspnet.org/).

Caroline Haythornthwaite has used social network approaches to examine the relational base of networks in online learning communities, both in formal online classes (e.g., Haythornthwaite 2002a) and knowledge communities (Haythornthwaite 2006). Her work examined relations and media use among learners, exploring "who talks to whom, about what, and via which media." Results showed that the number of media used increased with increasing closeness of the tie (*media multiplexity*) and that media used among members of a class conformed to a unidimensional scale: media used class wide by weak ties (e.g., bulletin boards) were supplemented with more private media used by strong ties (e.g., e-mail), and media were added in an order consistent with the strength of the tie. Her work concluded that class-wide media can act as a *latent tie structure* that supports relationship development between those who initially do not know each other. This work also highlights the role of the teacher or authority in such situations as their control of interaction behavior – participation, group projects, and discussion group formation – holds the key to turning latent ties into stronger work and learning ties (Haythornthwaite 2002b, 2005). More recent work extends these ideas into understanding motivators and social network implications for contribution to online crowds and communities (Haythornthwaite 2009; Budhathoki and Haythornthwaite 2013) and exploring network patterns of participation in online discussions (Haythornthwaite and Gruzd 2012).

Shane Dawson and colleagues have applied social network methodologies to specific learning contexts to identify lead indicators of student creativity (Dawson et al. 2011b), learning dispositions (Dawson et al. 2011a), student sense of community (Dawson 2010), and academic performance (Macfadyen and Dawson 2010). Dawson's work also includes the development of the SNAPP tool (Social Networks Adapting Pedagogical Practice); its application in higher education has further demonstrated the capacity for SNA and visualizations to inform pedagogical design and interventions necessary to optimize the student learning environment (Dawson et al. 2011).

Anatoliy Gruzd has developed software tools that automate the extraction of social networks from online discussion. His work uses linguistic analysis and text mining to discover network ties based on content within the online posting. With appropriate permissions and access to the data, the automated methods can circumvent the often burdensome user-based collection of data by questionnaire; Gruzd's methods improve on the use of message headers alone to identify and build "who replies to whom" networks. While headers are appropriate when studying e-mail-centric communication networks where individuals are identified with individual addresses, discussions common in online learning environments tend to be open to the whole group and do not follow a clear one-to-one pattern. To overcome this limitation, Gruzd (2009) proposed and tested an automated method he called "name networks" that uses mentions of personal names in messages to discover social ties between online participants. A study of six online graduate classes at a US university found that this method was twice as likely to reveal the same ties as questionnaire-based self-report networks than the "who replies to whom" networks generated by the use of message header data only. This suggests that social networks discovered by the "name network" method are a viable method for revealing networks that closely reflect users' perceptions of online social interactions.

In another development of methods to reveal networks, Dan Suthers and colleagues use a social network approach to combine examination of interactions among network actors and the media in which these communities are embedded. Their work applies a partitioning algorithm to a bipartite graph of actors and media objects (Suthers and Chu 2012). They have also developed a framework for analyzing the sequential structure of interaction that is distributed across media (Suthers et al. 2010).

Future Directions

A social network perspective on learning can provide new insight into many aspects of learning by individuals, communities, and societies in formal, nonformal, and informal settings. Social network analysis provide a rich set of concepts that can be used to address learning in a number of ways, from understanding the relational basis of learning ties to the influence and effects of found or constructed network structures. Learning itself has multiple interpretations, from the kind of knowledge acquired to the outcomes that result. While learning is often bound up with the notion of education, in formal contexts, social network approaches can provide important perspectives on learning in wider contexts. In brief, learning from an SNA perspective can examine learning as:

- The *relation* that connects people
- A *characterization of the tie* (e.g., based on multiple, context-dependent relations)
- A characterization of the *outcome of relations* (e.g., when a group becomes a learning community)
- The *network outcome of relations* (e.g., the social or learning capital of the network)
- Contact with *ambient influence* (e.g., in informal or latent learning)

New directions include not just analysis of learning cases but also the development of learning environments that provide feedback and visualization of learning processes in real time. An understanding of social networks and learning thus also provides *input for design* of social and technical systems to support learning processes for individuals, groups, and communities.

Cross-References

▶ Community Evolution
▶ Crowdsourcing
▶ Crowdsourcing and Social Networks
▶ Entrepreneurial Networks
▶ Human Behavior and Social Networks
▶ Online Communities
▶ Social Capital

References

Benkler Y (2006) The wealth of networks: how social production transforms markets and freedom. Yale University Press, New Haven

Borgatti SP, Cross R (2003) A relational view of information seeking and learning in social networks. Manag Sci 49(4):432–445

Budhathoki N, Haythornthwaite C (2013) Motivation for open collaboration: crowd and community models and the case of OpenStreetMap. Am Behav Sci 57(5): 548–575

Burt R (1992) Structural holes: the social structure of competition. Harvard University Press, Cambridge

Cohen WM, Levinthal DA (1990) Absorptive capacity: a new perspective on learning and innovation. Adm Sci Q 35:128–152

Cummings J, Kiesler S (2008) Who collaborates successfully? Prior experience reduces collaboration barriers in distributed interdisciplinary research. In: Proceedings of the ACM conference on computer-supported cooperative work CSCW '08, San Diego. ACM, New York

Daly AJ (ed) (2010) Social network theory and educational change. Harvard Education Press, Cambridge

Dawson S (2010) 'Seeing' the learning community: an exploration of the development of a resource for monitoring online student networking. Br J Educ Technol 41(5):736–752

Dawson S, Bakharia A, Lockyer L, Heathcote E (2011) 'Seeing' networks: visualising and evaluating student learning networks. Final Report 2011. Australian Learning and Teaching Council Ltd, Canberra. Available online at: http://research.uow.edu.au/content/groups/public/@web/@learnnet/documents/doc/uow115678.pdf

Dawson S, Macfadyen LP, Lockyer L, Mazzochi-Jones D (2011a) Using social network metrics to assess the effectiveness of broad based admission practices. Australas J Educ Technol 27(1):16–27

Dawson S, Pei-Ling Tan J, McWilliam E (2011b) Measuring creative potential: using social network analysis to monitor and develop learners' creative capacity. Australas J Educ Technol 27(6): 924–942

De Laat MF (2011) Bridging the knowledge gap: using social network methodology for detecting, connecting and facilitating informal networked learning in organizations. In: Proceedings of the 44^{th} IEEE annual Hawaii international conference on system sciences, Kauai. IEEE

DeSanctis G, Poole MS (1994) Capturing the complexity in advanced technology use: adaptive structuration theory. Organ Sci 5(2):121–47

Girard M, Stark D (2007) Socio-technologies of assembly: sense-making and demonstration in rebuilding lower Manhattan. In: Mayer-Schönberger V, Lazer D (eds) Governance and information technology: from electronic government to information government. MIT Press, Cambridge, pp 145–176

Granovetter MS (1973) The strength of weak ties. Am J Sociol 78:1360–1380

Gruzd A (2009) Studying collaborative learning using name networks. J Educ Libr Inf Sci 50(4):243–253

Hansen MT (1999) The search-transfer problem: the role of weak ties in sharing knowledge across organization subunits. Adm Sci Q 44:82–111

Haythornthwaite C (2002a) Strong, weak and latent ties and the impact of new media. Inf Soc 18(5):385–401

Haythornthwaite C (2002b) Building social networks via computer networks: creating and sustaining distributed learning communities. In: Renninger KA, Shumar W (eds) Building virtual communities: learning and change in cyberspace. Cambridge University Press, Cambridge, pp 159–190

Haythornthwaite C (2005) Social networks and internet connectivity effects. Inf Commun Soc 8(2):125–147

Haythornthwaite C (2006) Articulating divides in distributed knowledge practice. Inf Commun Soc 9(6):761–780

Haythornthwaite C (2009) Crowds and communities: light and heavyweight models of peer production. In: Proceedings of the 42nd Hawaii international conference on system sciences, Maui. IEEE Computer Society, Los Alamitos

Haythornthwaite C, Andrews R (2011) E-learning theory and practice. Sage, London

Haythornthwaite C, Gruzd A (2012) Exploring patterns and configurations in networked learning texts. In: Proceedings of the 45th Hawaii international conference on system sciences. IEEE Computer Society, Los Alamitos

Haythornthwaite C, Lunsford KJ, Bowker GC, Bruce B (2006) Challenges for research and practice in distributed, interdisciplinary, collaboration. In: Hine C (ed) New infrastructures for science. Knowledge production. Idea Group, Hershey, pp 143–166

Hollingshead AB, Fulk J, Monge P (2002) Fostering intranet knowledge sharing: an integration of transactive memory and public goods approaches. In: Hinds P, Kiesler S (eds) Distributed work: new research on working across distance using technology. MIT, Cambridge, pp 335–355

Kelton K, Fleischmann KR, Wallace WA (2008) Trust in digital information. J Am Soc Inf Sci Technol 59(3):363–374

Klein JT (1996) Crossing boundaries: knowledge, disciplinarities, and interdisciplinarities. University Press of Virginia, Charlottesville

Krackhardt D (1987) Cognitive social structure. Soc Netw 9:109–134

Macfadyen L, Dawson S (2010) Mining LMS data to develop an “early warning system” for educators: a proof of concept. Comput Educ 54(2):588–599

Moreland R (1999) Transactive memory: learning who knows what in work groups and organizations. In: Thompson L, Levine J, Messick D (eds) Shared cognition in organizations. Lawrence Erlbaum, Mahwah, pp 3–31

Orlikowski WJ (2002) Knowing in practice: enacting a collective capability in distributed organizing. Organ Sci 13(3):249–273

Raymond ES (1999) The cathedral & the bazaar: musings on Linux and open source by an accidental revolutionary. O’Reilly, Cambridge

Resnick P (2002) Beyond bowling together: sociotechnical capital. In: Carroll J (ed) HCI in the new millennium. Addison-Wesley, Boston, pp 247–272

Schreurs B, De Laat MF (2012) Network awareness tool: learning analytics in the workplace: detecting and analyzing informal workplace learning. In: Proceedings of the learning analytics & knowledge conference, Vancouver. ACM

Smith-Lovin L, McPherson M, Cook J (2001) Birds of a feather: homophily in social networks. Annu Rev Sociol 27:415–44

Suthers DD, Chu K-H (2012) Multi-mediated community structure in a socio-technical network. In: Proceedings of the learning analytics and knowledge conference, Vancouver

Suthers DD, Dwyer N, Medina R, Vatrapu R (2010) A framework for conceptualizing, representing, and analyzing distributed interaction. Int J Comput Support Collab Learn 5(1):5–42

Wenger E, Trayner B, De Laat MF (2011) Promoting and assessing value creation in communities and networks: a conceptual framework. Ruud de Moor Centrum, Open Universiteit, Heerlen. Available online at http://wenger-trayner.com/documents/Wenger_Trayner_DeLaat_Value_creation.pdf

Recommended Reading

Gruzd A, Haythornthwaite C (2011) Networking online: cybercommunities. In: Scott J, Carrington P (eds) Handbook of social network analysis. Sage, London, pp 167–179

Haythornthwaite C, De Laat MF (2011) Social network informed design for learning with educational technology. In: Olofsson AD, Lindberg JO (eds) Informed design of educational technologies in higher education: enhanced learning and teaching. IGI Global, Hershey, pp 352–374

Haythornthwaite C, De Laat MF, Dawson S (eds) (forthcoming) Learning analytics. Am Behav Sci, whole issue

Romero C, Ventura S, Pechenizkiy M, Baker RSJd (eds) (2011) Handbook of educational data mining. CRC, Boca Raton/Taylor & Francis

Siemens G, Dragan G (eds) (2012) Learning and knowledge analytics. Educ Technol Soc 15(3), whole issue, pp 1–343

Learning to Rank

▶ Weblog Analysis

Least Squares

Ali Sayfy
Department of Mathematics and Statistics,
American University of Sharjah, Sharjah, UAE

Glossary

LS Least squares
LLS Linear least squares
MF Model function

Definition

Least squares historically grew out of astronomy problems during the turn of the seventeenth century (Nievergelt 2000), in particular, finding trajectories of planetary motions in order to solve ocean's navigation problems. The idea of ***LS*** was developed by Gauss, Legendre, Laplace, and many other mathematicians and scientists (Farebrother 1999). However, the first publication on ***LS*** appeared in 1805 by Legendre; he proposed the idea of minimizing the sum of squares of the errors to obtain the adjusted values of observed quantities (Plackett 1972).

"**Least squares**" is a mathematical procedure to find the best approximate curve or surface to a given set of data points, out of different model functions that approximate the data.

Linear Least Squares

The most common application of the least squares procedure is the ***LLS*** curve fitting, for which the ***MF*** forms

$$y = f(x) = a_1 f_1(x) + a_2 f_2(x) + \cdots + a_m f_m(x), \tag{1}$$

where $f_1, f_2, \cdots, f_m$ are given functions of the independent variable x, called basis functions, and $a_1, a_2, \cdots, a_m$ are the parameters to be found by the ***LS*** procedure, to obtain the best curve that approximates a given set of points:

$$\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)\}.$$

The model function (1) covers variety of functions such as polynomials and trigonometric.

Straight Line Least Squares

The simplest ***LLS*** is the straight line ***LS*** fitting for which the ***MF*** is given by $f(x) = a_1 + a_2 x$. Let d_i be the deviation at x_i, which is the difference between the actual value y_i and the predicted value $f(x_i)$. To find the values of the parameters a_1 and a_2 for the best straight line that approximates a given set of points, such as provided in Table 1 the sum of the squares of the deviations d_i has to be minimized (See Fig. 1 for a typical straight line LS).

Since the sum of squares of the deviations is a function of a_1 and a_2, we need to minimize the function $S(a_1, a_2)$, where

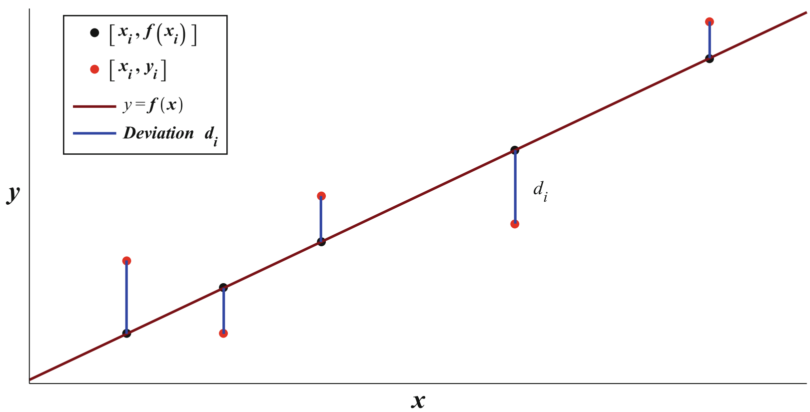
$$S(a_1, a_2) = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (a_1 + a_2 x_i - y_i)^2. \tag{2}$$

From calculus, the condition for $S(a_1, a_2)$ to be minimum is that the first partial derivatives satisfy

$$\frac{\partial S}{\partial a_i} = 0, \; i = 0, 1. \tag{3}$$

Least Squares, Table 1 Sample data for straight line LS

x_i	0.3	0.5	0.8	1.0	1.1	1.4	1.5	1.7	1.9	2.0
y_i	1.0	1.5	2.0	2.7	5.0	7.2	10.8	12.0	15.7	19.9



Least Squares, Fig. 1 Straight line ***LS***

This leads to

$$\begin{aligned}\frac{\partial S}{\partial a_1} &= \sum_{i=1}^{n} 2(a_1 + a_2 x_i - y_i) = 0,\\ \frac{\partial S}{\partial a_2} &= \sum_{i=1}^{n} 2(a_1 + a_2 x_i - y_i)x_i = 0.\end{aligned} \tag{4}$$

Thus, the best line that fits the given set of points can be found by solving the system of the two equations in (4). In particular,

$$a_1 = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2}, \tag{5}$$

$$a_2 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2}, \tag{6}$$

where the summations are from $i = 1$ to $i = n$. For more details, see Wackerly (2008).

Simple Nonlinear Model Functions

Some nonlinear two-parameter ***MF*** can be reduced to linear functions by using simple transformations. Consider the exponential ***MF*** $Y = ae^{bX}$. By taking the natural logarithm of both sides of the equation, we have $lnY = lna + bX$, where ln is the natural logarithm function. Then the model reduces to $y = a_1 + a_2 x$ by using the following transformations:

$$a_1 = lna, \quad a_2 = b, \quad x = X, \quad y = lnY. \tag{7}$$

After finding the values of the new parameters a_1 and a_2 by the ***LLS*** method, the original parameters can be found by $a = e^{a_1}$ and $b = a_2$. The following nonlinear models can be treated similarly:

$$\text{Geometric}: Y = aX^b, \quad a_1 = lna, \quad a_2 = b, \quad x = lnX, \quad y = lnY.$$

$$\text{Hyperbolic}: Y = \frac{1}{a + bx}, \quad a_1 = a, \quad a_2 = b, \quad x = X, \quad y = \frac{1}{Y}.$$

Linear Least Squares Procedures

For the general linear ***MF*** (1), the parameters $a_1, a_2, \cdots, a_m$ are obtained by solving the linear system of m equations:

$$a_1 \sum_{j=1}^{n} f_1(x_j) f_k(x_j) + a_2 \sum_{j=1}^{n} f_2(x_j) f_k(x_j) + \dots$$
$$+ a_m \sum_{j=1}^{n} f_m(x_j) f_k(x_j) = \sum_{j=1}^{n} y_j f_k(x_j), \;\; k = 1, 2, \dots, \tag{8}$$

The system is rewritten as $FA = B$, where

$$F = [f_{ik}]_{m\times m}, \quad A = [a_1,\ a_2,\ \dots,\ a_m]^T,$$
$$B = [b_1, b_2,\ \dots,\ b_m]^T,$$
$$f_{ik} = \sum_{j=1}^{n} f_i(x_j) f_k(x_j), b_k = \sum_{j=1}^{n} y_j f_k(x_j).$$

The efficiency of the fitted curve may be measured by SSE, the sum of squares of the errors (deviations). Smaller value indicates better approximation:

$$SSE = \sum_{i=1}^{n} [y_i - f(x_i)]^2 . \tag{9}$$

For derivations and more details, see (Wackerly et al. 2008) and (Anderson et al. 1994).

Polynomial Least Squares

For the case of polynomial ***MF*** $f(x) = a_1 + a_2 x + a_3 x^2$, the corresponding linear system to be solved is

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}. \tag{10}$$

Example Figure 2 presents three approximate ***LS*** curves for the following data, using different model functions:
The SSE for the three ***MFs***, straight line, exponential, and polynomial of degree two, are 44.5, 10.4, and 6.3, respectively This indicates that the polynomial is better approximation than the other two functions for the given data.

Multiple Linear Least Squares

The procedure for the ***LLS*** can be extended to approximate surfaces. Consider the ***MF*** $z = f(x, y) = a_1 + a_2 x + a_3 y$, which represents planes in three-dimensional space. Following a similar procedure as for the ***LLS***, the parameters for the best plane that approximates a given set of points

$$\{(x_i, y_j, z_{ij}) : i = 1, 2, \dots, n, j = 1, 2, \dots m\}$$

are obtained by minimizing the function:

$$S(a_1, a_2, a_3) = \sum_{i=1}^{n} \sum_{j=1}^{m} [z_{ij} - f(x_i, y_j)]^2 .$$

By setting the partial derivatives with respect to the parameters $a_1,\ a_2$, and a_3 to zero, the following linear system has to be solved to estimate the parameters:

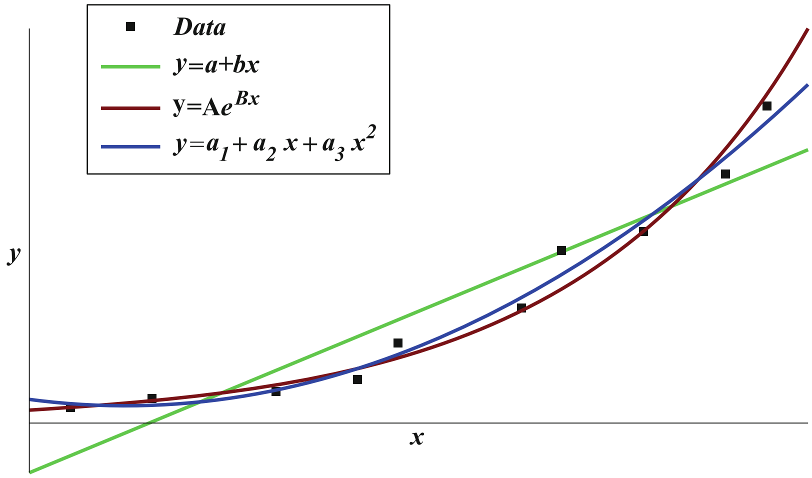
$$\begin{bmatrix} nm & m\sum x_i & n\sum_j^y \\ m\sum x_i & m\sum x_i^2 & \sum\sum x_i y_j \\ n\sum y_j & \sum\sum x_i y_j & n\sum y_j^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$
$$= \begin{bmatrix} \sum\sum z_{ij} \\ \sum\sum x_i z_{ij} \\ \sum\sum y_j z_{ij} \end{bmatrix}. \tag{11}$$

Weighted Least Squares

To obtain a better fitted curve when the deviations are significantly different, the weighted least squares can be used by assigning each data the proper influence on the parameter estimates. In this case, the parameters in the ***MF*** $f(x_i;\ a_1, a_2, \dots,\ a_n)$ are estimated by minimizing the summation:

$$\sum_{i=1}^{n} w_i \left[y_i - f(x_i;\ a_1, a_2, \dots, a_n)\right]^2 ,$$

where w_i's are the selected weights for the data. For more details, see Ryan (1997).



Least Squares, Fig. 2 Straight line, exponential, and polynomial *LS*

Cross-References

▶ Regression Analysis
▶ Theory of Statistics, Basics, and Fundamentals

References

Anderson DR, Sweeney DJ, Williams TA (1994) Introduction to statistics: concepts and applications. West Group

Farebrother RW (1999) Fitting linear relationships: a history of the calculus of observations 1750–1900. Springer, New York

Nievergelt Y (2000) A tutorial history of least squares with applications to astronomy and geodesy. J Comput Appl Math 121:37–72

Plackett RL (1972) The discovery of the method of least squares. Biometrica 59(2):239–251

Ryan TP (1997) Modern regression methods. Wiley, New York

Wackerly DD, Mendenhall W III, Scheaffer RL (2008) Mathematical statistics with applications, 7th edn. Cengage Learning

References

Lancaster P, Šalkauskas K (1986) Curve and surface fitting: an introduction. Academic, London

Lawson C, Hanson R (1974) Solving least squares problems. Prentice-Hall, Englewood Cliffs

Stigler SM (1986) The history of statistics: the measurement of uncertainty before 1900. Harvard University Press, Cambridge

Todhunter I (1865) A history of the mathematical theory of probability, from the time of Pascal to that of Laplace. Macmillan, London. [New York, Chelsea,1949]

Vos D (2008) Risk analysis: a quantitative guide, 3nd edn. John Wiley & Sons

Legal Implications of Social Networks

Davide Maria Parrilli[1] and G. R. Gangadharan[2]
[1]Deloitte Enterprise Risk Services, Brussels, Belgium
[2]Institute for Development and Research in Banking Technology (IDRBT), Hyderabad, India

Synonyms

Intellectual property rights; Legal issues and implications; Privacy; Regulatory concerns

Glossary

Intellectual Property Rights Intellectual property rights are a set of exclusive rights that are recognized under law for using a number of distinct types of creations of the mind (intangible assets). Common types of intellectual property rights include copyright and industrial property (such as trademarks, patents, industrial design rights)

Defamation Defamation refers to the publication or posting on a social networking site of a false statement (written or oral, for instance, contained in a video) about someone else that causes harm to his/her reputation

Liability Liability refers to the condition of being legally liable for a fact that produces harm to a third party and that has been produced as a consequence of a positive or negative conduct (omission) of the person deemed to be liable. Liability commonly implies the obligation to compensate the third party's damage and/or a criminal sanction

Privacy Privacy refers to the ability of an individual or group to seclude themselves or information about themselves and thereby reveal themselves selectively

Data Protection Under certain jurisdiction, personal data can only be gathered legally under strict conditions for a legitimate purpose. Persons or organizations that collect and manage personal information must protect it from misuse and must respect certain rights of the data owners, defined by the applicable laws of the jurisdiction concerned

Introduction

Social networks are often perceived by users as spaces where their freedom is basically unlimited and where as a consequence everything, or almost everything, can be done and is licit (Peck Pinheiro 2010). The reason for this common perception is basically the fact that users tend to have the feeling that they are anonymous or, at least, that they can cover their real identity under a nickname with no reference to their real identity or, even worse, under a profile that refers to a nonexisting person or to another existing person who usually is not aware that his or her identity is used by somebody without his or her consent.

It is evident that social networks are virtual spaces that are subject to laws and rules like all other social spaces in the offline world (Goldman 2007). In particular, the protection of the following two fundamental values is not (and shall not be) excluded in online social networks: (i) the intellectual creations of the human spirit and (ii) the dignity of all human beings.

In legal terms, the previous statement means that infringements of copyright and acts of defamation that happen online in social networking sites shall be prosecuted, of course when the applicable law states that copyright infringements and defamation can give rise to criminal and/or civil liability. Rules aimed to prevent and punish copyright infringements and acts of defamation are typical of a modern and democratic society, irrespective of where these violations are committed, in the offline or in the online world.

Legal Issues of Social Networking: An Overview

Intellectual Property Rights Infringement

Social networking sites allow individual users and enterprises to post content (including text, graphics, and logos) that may be copyrighted or trademarked. This could lead to severe violations of intellectual property rights law. The use of third party content without owner's permission could result in both criminal and civil liability under the US Copyright Act (the Copyright Act of 1976 and all subsequent amendments contained under Title 17 of the United States Code).

Generally, social networking sites specify and regulate the use of third party content in their terms of use. Following is a fragment of terms of use of Myspace.com related to the use of third party content by users:

> You [users] are solely responsible for the Content that you post on, through or in connection with any of the Myspace Services, and any material or information that you transmit to other Members and for your interactions with other Users.

Furthermore, social networking sites do not take any responsibility for the content posted by users in their sites:

> [A social networking site] assumes no responsibility for the Content, no obligation to modify or remove any inappropriate Content, and no responsibility for the conduct of the User submitting any such Content.

As copyright infringement could happen in any social networking sites, these sites generally inform users of infringement notifications they receive from alleged copyright owners.

Social networking sites differ widely in their policies and practices (Ossian 2009). LinkedIn requires its users to grant LinkedIn a *"nonexclusive, irrevocable, worldwide, perpetual, unlimited, assignable, sublicenseable, fully paid up and royalty-free right [...] to copy, prepare derivative works of, improve, distribute, publish, remove, retain, add, process, analyze, use and commercialize, in any way now known or in the future discovered, any information you provide, directly or indirectly to LinkedIn, including, but not limited to, any user generated content, ideas, concepts, techniques or data to the services, you submit to LinkedIn, without any further consent, notice and/or compensation to you or to any third parties."* On the other hand, Twitter offers a *"personal, worldwide, royalty-free, non-assignable and non-exclusive license to use the software."*

Defamation and Disparagement

Postings on the social networking sites are instantly available as we make and, once posted, they are incapable of "true" deletion (Ossian 2009). Hence, it is necessary to monitor the specification of content regarding any third parties. Generally social networking sites prohibit content that defame any third party and remove any such content, if found. Following is a fragment of terms of service from Twitter regarding the defamation of third party:

> To the maximum extent permitted by applicable law, [a social networking site]will not be liable for any conductor content of any third party on the services, including without limitation, any defamatory, offensive or illegal conduct of other users or third parties.

Social networking sites allow any comments or criticisms posted by users over any person or organization and usually do not control and filter content posted by users. However, statements published in a social networking site that arguably defame a person or an organization could form the basis for legal action against the user.

Data Protection and Privacy Issues

Social networking sites allow users to have a high degree of freedom regarding the information they disclose (Van Alsenoy et al. 2009). As a social networking user reveals the more granular information about him- or herself, he or she would be in the state of losing his or her privacy beyond traditional personally identifiable information (Navetta 2011).

The information disclosed and made public could include geographical location information, photographs and videos (and thus images of the person concerned and/or of other identifiable persons), relation information, online behavioral information, and political view points, including more sensitive information including health-related and sexual orientation-related data (Kosta et al. 2010). The complexity and obtuseness of privacy and permission settings results in the exposure of users' details to unintended recipients (Weir et al. 2011). Following is a fragment of terms of service from Twitter regarding Information Sharing and Disclosure:

> We engage certain trusted third parties to perform functions and provide services to us. We may share your personal information with these third parties, but only to the extent necessary to perform these functions and provide such services, and only pursuant to obligations mirroring the protections of this privacy policy.

In order to protect user privacy, social networking sites keep the layered access control including profile privacy, application privacy, and newsfeed privacy. However, social networking sites often do not make users aware of the dangers of divulging of their personal information (Ai et al. 2009).

Legal Liability for Social Networking Sites Providers

Social networking service providers are deeply affected by potential *copyright* infringements made by their users, who have the de facto possibility of posting illegal material (i.e., material that infringes a third party's copyright, such as music, text, and pictures) on the online social networking platform. It is clear that it may be cumbersome to assess whether or not the material posted or uploaded by the social networking site user violates the third party's copyright and that a complex case-by-case analysis may be necessary. Therefore, it is interesting to assess to what extent service providers are liable for copyright infringements made by their users and subscribers, provided that such infringements are very likely to happen on a massive scale.

From an economical perspective, it may be an interesting solution to allocate responsibilities for users' copyright violations to the service providers. However, it may be cumbersome (and expensive) to find out the identity of the infringer. In some cases, data protection legislation and case law limit the possibilities to investigate about infringers on the Internet. Service providers by definition cannot mask their identity. Hence, it may be in favor of copyright holders to set by law that they are liable as such for copyright infringements made by their users.

From the legal point of view, a person is responsible to the extent that an infringement derives from his/her conduct or omission, when fault by omission is contemplated by the applicable legislation. The legislations of some developed countries tried to find a balance between the abovementioned economic needs and the strict legal principles.

The Copyright Law of the United States, as amended by the Digital Millennium Copyright Act (DMCA) of 1998, goes in this direction. Social networking service providers may be affected by contributory infringement liability, provided that they are likely to potentially receive several daily notification of users' copyright infringements by copyright holders, and therefore, such notices could establish the "knowledge" on the part of the social networking service provider. The contributory infringement liability doctrine may be applicable based on the theory of the social networking sites being facilitators of the direct users' infringement (Darrow and Ferrera 2007). Hence, the US legislation clearly decided to exclude the responsibility of the social networking service providers (and in general of the Internet service providers) not only for vicarious and contributory infringement but also for direct copyright infringements.

In fact according to Section 512 (Limitations on liability relating to material online) of the Copyright Law, a service provider shall not be liable for damages (more precisely, for monetary relief and, as a rule, for injunctive or other equitable relief) for infringements of copyright by a third party if one of the following conducts takes place: (a) transitory digital network communications, (b) system caching, (c) *storage of information residing on systems or networks at direction of users* (the relevant situation in the case of social networking services), and (d) information location tools.

In the field of social networking, under Section 512 (c), social networking site provider enjoys liability if the following three sets of conditions are met:

(A) The service provider does not have actual knowledge that the material (or an activity using the material) on the system/network is infringing. In the absence of such actual knowledge, the service provider is not aware of facts or circumstances from which infringing activity is apparent. Upon obtaining such knowledge or awareness, the service provider acts expeditiously to remove, or disable access to, the material.

(B) The service provider does not receive a financial benefit directly attributable to the infringing activity, in a case in which the service provider has the right and ability to control such activity.

(C) Upon notification of claimed infringement by the copyright holder, the service provider responds expeditiously to remove, or disable access to, the alleged infringing material or to the material claimed to be the subject of infringing activity.

The nature of the financial benefit received by the social networking service provider shall be assessed, taking into account as to whether the fee paid by the user or subscriber depends on the nature of the activity performed by the user or subscriber on the social networking site. If users pay a fixed fee to use the social networking services, regardless of the illegal activity performed by the user, it is less likely to find a direct benefit for the social networking service provider. The Senate Report accompanying the DMCA confirmed that "*receiving a one-time set-up fee and flat periodic payments for service from a person engaging in infringing activities would not constitute receiving a financial benefit directly attributable to the infringing activity, but the direct financial benefit would however, include any such fees where the value of the service lies in providing access to infringing material*." Thus, a case-by-case analysis is necessary for providing an answer to the question whether the user or subscriber pays a fee to access a social networking site whose purpose is to commit copyright infringements. In the vast majority of social networking sites, this is not the case in point, although they theoretically permit such violations.

Furthermore, if the service provider does not have knowledge or awareness of the infringing nature of the user's behavior by itself, but it receives a notification by the copyright holder, the service provider has the obligation to respond expeditiously to remove, or disable access to, the material that is claimed to be infringing or to be the subject of infringing activities.

The limitation of liability pursuant to Section 512 (c) is applicable under the condition that the service provider has designated an agent to receive notifications of claimed infringements by providing contact information (namely, name, address, phone number, and e-mail address of the agent, as well as other contact information which the Register of Copyrights may deem appropriate) to the Copyright Office and by posting such information on the service provider's website or in a location accessible to the public.

The limitations on liability pursuant to Section 512 apply to a service provider if the following conditions are met:

(A) The service provider has adopted and reasonably implemented a policy that provides for the termination in due circumstances of subscribers and account holders who are repeat infringers (and he or she has informed them of the existence of this policy).
(B) The service provider does not interfere with standard technical measures. Such technical measures should be used by copyright owners to identify and protect copyrighted works and have been developed pursuant to a broad consensus of copyright owners and service providers in an open, voluntary, and multi-industry standards process. These technical measures should be available to any person on nondiscriminatory terms and should not impose substantial costs or burdens on service providers.

The non-respect of the conditions set forth under Section 512 (i) will exist when the social networking service provider will be in the conditions, by virtue of a business decision, not to know which specific user posted or uploaded the infringing material since the users are active on the social networking sites in a completely anonymous way. Conversely, the social networking service provider will be eligible for the safe harbor protection, and the condition of reasonable implementation of the policy above subpoint A will be met, as far as he or she diligently responds to infringements notices he or she receives from copyright holders and it consistently terminates the accounts of repeat infringers (Darrow 2007). As regards the question whether the service provider should have a policy aimed to prevent the re-registration of infringing users, there is no evidence in the legislation that such policy should be provided for and implemented by social networking service providers.

Legal Liability for Social Networking Sites Users

Social networking sites enable users to publish content that creates several legal issues including defamation and copyright infringements. In this perspective, users can be liable for what they publish. Assuming that users are unaware of laws for content publishing, users can get themselves into trouble if they are posting any defamatory and copyright-violating content.

Under the US law there are special rules aimed to dramatically limit the liability of users and service providers for the defamatory content published by a third party. This principle, and the exclusion of liability thereof, is set forth by Section 230 of the Communications Decency Act of 1996 as follows: "*[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider*." This provision applies to social networks where users can only post material and to social networks where users have a more active role in terms of user-generated content (for instance, they create pages where other users can post comments, material, etc.). In the latter scenario we have three parties involved: the social networking service provider that put the social networking platform at disposal of users, the user that created the page where other users can post text and material, and the user who materially posted defamatory material on that page.

Section 230 is applicable if following conditions are met:

(i) The defendant (i.e., the person against whom a claim for defamation has been filled by the alleged victim of the defamation) is a provider or user of an interactive computer service.

(ii) The claim treats the defendant as a publisher or speaker of information alleged to be defamatory.

(iii) The challenged communication is information provided by another information content provider, defined as "*any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service*."

Service providers and users shall not be held liable neither when they implement in good faith measures to "*restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected*" or, consequently, measures to "*enable or make available to information content providers or others the technical means to restrict access to material described*" above.

From a different perspective, it is also possible that the users of social networking sites may potentially commit copyright infringements. According to Section 106 of the US Copyright Law, the owner of copyright (usually the creator of the copyrighted work or the licensee thereof), has the exclusive right to do and authorize to reproduce the work, to prepare derivative works based upon the copyrighted work, to distribute copies of records to the public (by sale, transfer of ownership, rental, lease, or lending), to perform the copyrighted work publicly, to display the work publicly, and to perform the work by means of digital audio transmission.

It is clear that many conducts adopted by users on social networking sites constitute violations of copyright holders' rights, namely, when users reproduce, broadcast, or distribute copyrighted material on a social networking site. The legal principles above are reflected, for instance, in the Facebook Statement of Rights and Responsibilities that states as follows: "*You* [the user] *will not post content or take any action on Facebook that infringes or violates someone else's rights or otherwise violates the law*."

The user's liability for copyright infringements may be excluded when he or she falls under the application of the *fair use* rule provided by Section 107 of the US Copyright Law. There is no copyright violation when a person uses copyrighted work (even if the work is unpublished yet) for the purposes such as criticism, comment, news reporting, teaching, scholarship, or research, taking into account the aim and character of the use (commercial vs. nonprofit), the nature of the copyrighted work, the amount and substantiality of the portion of the copyrighted work used (a short extract vs. the entire copyrighted work), and the effect of the use upon the potential market for, or value of, the copyrighted work.

In practice it may be cumbersome to assess if the use of copyrighted work falls under the fair use rule or is an infringement. The 1961 Report of the Register of Copyrights on the General Revision of the US Copyright Law cites some examples of activities and uses that have been considered as fair use in the case law: "*quotation of excerpts in a review or criticism for purposes of illustration or comment; quotation of short passages in a scholarly or technical work, for illustration or clarification of the author's observations; use in a parody of some of the content of the work parodied; summary of an address or article, with brief quotations, in a news report; reproduction by a library of a portion of a work to replace part of a damaged copy; reproduction by a teacher or student of a small part of a work to illustrate a lesson; reproduction of a work in legislative or judicial proceedings or reports; incidental and fortuitous reproduction, in a newsreel or broadcast, of a work located in the scene of an event being reported*." This applies mutatis mutandis also to social networking sites, and therefore, for instance, the quotation of an excerpt for purposes of comment on a social networking site is a clear application of the fair use rule.

Selected Cases

There is actually a lot of litigation pending concerning the copyright liability of social networking providers, especially in the field of social networks. An interesting and recent case ruled by a US court is *Wolk v. Kodak Imaging Network, Inc.*, 2012 WL 11270 (S.D.N.Y. Jan 3, 2012). The plaintiff, a graphic artist, sued a user-generated content social networking site where users can share pictures and videos. Although the plaintiff sent several notices of infringement to the defendant and access to images was removed by the defendant where the notices listed specific URLs, the defendant left accessible several copies of the infringing images through other URLs.

The court declared that the defendant cannot be held liable for copyright infringement based on the following assumptions: (i) the copying occurred automatically and therefore the defendant did not have the necessary volition to be held liable for the copying and the subsequent copyright infringements, and (ii) automated processes that lead to copyright infringements do not imply the volition required by the law in order to be held liable for copyright infringements.

Furthermore, the knowledge requirements of the copyright legislation were not met, although the social network operator was aware of the fact that infringing material was present on the site. In other words, the court ruled that the standard of knowledge required by the law is very high and is not limited to a passive knowledge. Therefore, it seems that an active conduct of the social networking site is necessary to establish the liability of the social network to facilitate copyright infringements by users.

As regards limitation of service providers' liability in case of defamation perpetrated by a social networking site user, a US upper court ruled that the obligation to remove defamatory content can be enforced only against the user that generated the content but it cannot be extended to and enforced against the website where the defamatory user-generated content has been published, since the website was not in active concert or participation with the user (*Blockowicz v. Williams*, No. 10-1167 (7th Cir. December 27, 2010)).

The limitation of liability for social networking service providers provided for under Section 230 of the Communications Decency Act of 1996 has been applied by US courts in a case involving Google and some comments denigrating a business posted by an anonymous user on a Google's social platform. In *Black v. Google Inc.*, 2010 WL 3746474 (N.D.Cal. September 20, 2010), the US District Court for the Northern District of California dismissed a suit against Google based on the assumption that the defendant just had a passive conduit and could not be held liable for failing to detect and remove the allegedly defamatory content, even if posted by anonymous users, and that "*the website did absolutely nothing to encourage the posting of defamatory content*."

In the case law it has been recalled that the role of the social networking service provided does not have to be absolutely neutral in order to benefit from the limitation of liability provided for under Section 230 of the Communications Decency Act of 1996. In *Reit v. Yelp, Inc.,* – N.Y.S.2d – 2010 WL 3490167 (September 2, 2010) and *Shiamili v. Real Estate Group of New York, Inc.*, 2009 WL 4842470 (N.Y. App. Div. Dec. 17, 2009), different courts stated that an allegation that a social networking provider or a website operator may promote negative content by users does not imply that the social networking provider or the website operator becomes an information content provider, subject to full defamation liability. In other words, the content created by the users does not cease to be data provided by a third party just because the functioning and running of the social network/website may have some influence and effects on the content posted by the user.

A social networking service provider shall not have an active and significant role in creating, developing, or transforming the data provided by the user; thus, he has to maintain a mere passive role in connection with the data posted by the user, as pointed out in *Carafano v. Metrosplash.com, Inc.*, 339 F.3d 1119 (9th Cir. 2003).

Best Practices and Guidelines for Minimizing Risks

An organization can develop a social media policy by determining what the organization wants to accomplish. Then it is important to make employees aware of the social media policy and enforce those policies. This kind of procedures could avoid information leakage in social networking sites by multiple users for or on behalf of an organization (Fayle 2007).

Social networking users are advised to read and review the terms of service and use of the social networking sites. They should understand the ownership and control policies and decide whether to use the site. Before publishing on the social networking sites, users should review the content they are drafting. In this way, users can have control over their own content. Users should be careful and clearly respect the copyright policies of any third party information they intend to use and publish.

The consequence of posting defamatory content on social networking site could serve as the basis for claims such as intentional infliction of emotional distress or interference with advantageous economic relations (Ossian 2009). These risks can be reduced by following the established policies at organizational level for online publishing. At individual level, such policies include attribution disclaimer and the review of content for accuracy and sensitivity before publishing. A social networking site cannot be meaningful if users do not want others to see their information. The level of risks for each user is determined by user behavior and attitudes on social networking sites (Ofcom 2008; Biagi et al. 2010). Information gathered by social networking sites is at the center of legal controversies. The sensitive nature of information that flows through social networking sites should be understood by the providers of social networking sites. The providers should recognize the serious compliance and litigation risks that the collection and distribution of such information entails and consider contractual tools to mitigate these risks including properly drafted privacy policies and terms of use (Pryor et al. 2010).

Cross-References

▸ Legal Implications of Social Networks
▸ Legislative Prediction with Political and Social Network Analysis

References

Ai H, Maiga A, Aimeur E (2009) Privacy protection issues in social networking sites. In: Proceedings of the IEEE/ACS international conference on computer systems and applications, Rabat

Biagi R, Schaffer D, Roe J (2010) Top ten legal issues in social media. Technical report, Neal & McDevitt LLC

Darrow J, Ferrera G (2007) Social networking web sites and the DMCA: a safe-harbor from copyright infringement liability or the perfect storm? Northwest J Technol Intellect Prop 6(1):1–35

Fayle K (2007) Understanding the legal issues for social networking sites and their uses. FindLaw. http://articles.technology.findlaw.com/2007/Sep/18/10966.html

Goldman E (2007) Social networking sites and law. Technical report, Santa Clara University

Kosta E, Kalloniatis C, Mitrou L, Gritzalis S (2010) Data protection issues pertaining to social networking under EU law. Transform Gov People Process Policy 4(2):193–201

Navetta D (2011) The legal implications of social networking. InfoLaw Group

Ofcom (2008) Social networking: a quantitative and qualitative research report into attitude, behaviors, and use. Office of Communications of United Kingdom

Ossian KL (2009) Legal issues in social networking. Technical report, Miller Canfield

Peck Pinheiro P (2010) Direito digital. Editora Saraiva, Saõ Paulo

Pryor G, Rosenbaum J, Wood D, Marcus S (2010) Network interference: a legal guide to the commercial risks and rewards of the social media phenomenon. Reedsmith Publishers

Van Alsenoy B, Ballet J, Kuczerawy A, Dumortier J (2009) Social networks and web 2.0: are users also bound by data protection legislations? Identit Inf Soc 2(1):65–79

Weir G, Toolan F, Smeed D (2011) The threats of social networking: old wine in new bottles? Information security technical report, Issue 16, Elsevier pp 38–43

Legal Issues

► Social Media Policy in the Workplace: User Awareness

Legal Issues and Implications

► Legal Implications of Social Networks

Legislative Prediction with Political and Social Network Analysis

Jun Wang, Kush R. Varshney, and
Aleksandra Mojsilović
Business Analytics and Mathematical Sciences Department, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA

Synonyms

Heterogeneous graph; Legislative vote prediction; Quantitative political science; Random walks; Roll call prediction

Glossary

IPM Ideal point model

IPTM Ideal point topic model

Heterogeneous Graph Refers to a graph with multiple types of nodes and edges

RWHG Random walk over a heterogeneous graph

Political Affinity Refers to the connections such as cosponsorship relations between legislators

Definition

The function of legislatures is to propose and vote on new laws. In some systems of government, including parliamentary governments that follow the Westminster system, the party affiliation of legislators is codified in the constitution, and legislators are bound to vote in lockstep with their party. However, in other systems of government, party affiliation is only one of many factors that influences a legislator's voting *yea* or *nay*. Ideology and political and social relationships are key components in a legislator's voting decision.

A bill is a proposed law under consideration by a legislature that, if passed, becomes a law. The historical record of votes on bills, known as the *roll call*, can be analyzed to obtain many different descriptive statistics but, just as importantly, can be used to learn predictive models of voting behavior. *Legislative prediction*, i.e., prediction of legislator votes on future bills, leads to a better understanding of government and can provide actionable insights to political strategists. The quality and power of legislative prediction is enhanced by using information beyond historical roll call data, especially bill semantics and legislator political and social network data.

Introduction

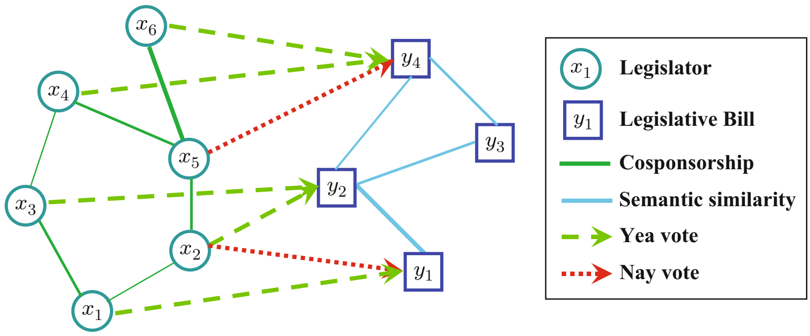
Legislative prediction is difficult because of the myriad factors that affect the decisions of legislators. The historical roll call is one useful data source for the task, but it fails to capture more semantic and relational factors that contribute to legislator vote decisions. By constructing networks of bills based on semantic similarity and networks of legislators based on social and political relationships, and using these networks in conjunction with the roll call, the accuracy of legislative prediction can be significantly improved (Wang et al. 2012).

Many political science studies focus on the federal government of the United States of America, a non-parliamentary system of government, and its legislature, known as the Congress, for which legislative prediction is a meaningful task. Congress is a bicameral legislature composed of the Senate with 100 members known as senators and the House of Representatives with 435 members known as representatives. A session of Congress lasts 2 years, with the current session being the 112th as of 2012. The composition of Congress changes after every session due to elections. Within a session, the only changes are due to death or resignation. There are approximately 700 bills voted upon per session in the Senate and approximately 1,200 in the House of Representatives. Each bill that comes to a vote in Congress is sponsored by at least one legislator and might be cosponsored by other legislators if they coauthored it or if they wish to publicly indicate strong support for it. Thus, frequent cosponsorship of bills reflects collaboration and similarity in ideology between legislators.

Historical Background and Related Work

Legislative roll calls have been studied statistically since the 1920s, if not earlier (Rice 1925). Statistical analysis has identified various factors affecting voting, including representational role (Kuklinski and Elling 1977) and coalition formation (Hinckley 1972). However, the development of predictive models of voting based on the roll call and other legislative data has been initiated this decade, coinciding with the growth in popularity and utilization of data-driven analytics, insight, and decision making in many areas, politics included. Legislative prediction is one aspect of political science that can be investigated with analytics; others include court rulings (Wang et al. 2011) and political campaigns (Luck et al. 2010).

A key model for legislative analysis developed in quantitative political science, the ideal point model (IPM), builds a one-dimensional latent "political space" and places each legislator and bill in that space using only roll call data (Clinton et al. 2004). However, IPM has a limited function for legislative prediction because there is no mechanism to place bills that have not yet been voted upon in the political space. Researchers from the machine-learning and data-mining communities have recently proposed methods that address shortcomings of IPM and that may be used for vote prediction. For instance, the ideal point topic model (IPTM) is constructed using both the historical roll call and the text of the bills (Gerrish and Blei 2011). Since the placement of bills in latent space is based on topic modeling of the semantics of the text (in addition to the roll call), new bills can also be placed. Then, legislative prediction is based on distance between the not yet voted upon bill and legislators.



Legislative Prediction with Political and Social Network Analysis, Fig. 1 Heterogeneous graph representation of roll call data. The heterogeneous graph contains legislator vertices, bill vertices, edges connecting legislators (cosponsorship relation), edges connecting bills (semantic similarity), and directed vote edges from legislators to bills

Similarly, another recent piece of work models the roll call data jointly with a topic model for the bills, adding in time dynamics as well (Wang et al. 2011).

In all of these legislative prediction advances, the semantic similarity of bills is taken into account, but the social and political network of the legislators is not. A few other recent works have modeled influence in the legislator network, but have not included bill semantics (Banerjee et al. 2008; Liu et al. 2010; Netrapalli et al. 2010). The random walk over a heterogeneous graph (RWHG) method is the first to approach legislative prediction by considering the roll call, the bill semantic similarity network, and the legislator political and social network in a joint graph-based way (Wang et al. 2012).

Motivated by the real data generated in different application domains, heterogeneous graph-based formulations have attracted attention recently. For instance, ranking and classifying the vertices of a heterogeneous graph are two well-formulated problems from real applications, such as author and documents co-ranking (Zhou et al. 2007) and identifying research communities from bibliographic data (Ji et al. 2011). A very recent study performs both random walk and propagation over a heterogeneous graph to improve topic modeling of documents (Deng et al. 2011). Compared to these existing approaches, RWHG formulates legislative prediction as a link prediction problem instead of either vertex ranking or classification. Although link prediction has been widely studied for social networks and the Internet, most of the existing methods are developed for graphs with either homogeneous vertices or edges (Backstrom and Leskovec 2011; Hasan et al. 2006; Liben-Nowell and Kleinberg 2007; Lichtenwalter et al. 2010; Lu et al. 2010; Taskar et al. 2004). Recently, Sun et al. proposed to learn optimal weights to combine heterogeneous topological features and then build a logistic regression model to predict coauthorship links (Sun et al. 2011).

Formulation and Methodology

Besides analyzing the semantics of a corpus of legislative bills, it is also critical to explore the social and political connections between legislators to predict votes. A typical formulation to leverage such kind of connection information is based on graph representation. Since the roll call data contains multiple types of entities, e.g., legislators and bills, and edges, e.g., bill similarity, legislator connections, and vote links, it is quite straightforward to use a heterogeneous-structured graph to represent the data.

Heterogeneous Graph for Roll Call Data

The heterogeneous graph converted from the roll call data is illustrated in Fig. 1. Assume there are a total of L legislators and denote the set of legislator vertices as $\mathbf{V}_{(x)} = \{x_l\}_{l=1}^{L}$. These legislators can be connected based on attributes such as party, state, age, gender, and cosponsorship by converting the attributes to a political similarity measure between legislators. Hence, the connections between legislators are represented as $\mathbf{E}_{(x)} = \{e_{(x)_{lm}}\}_{l,m=1}^{L} \subset \mathbf{V}_{(x)} \times \mathbf{V}_{(x)}$, associated with a weight matrix $\mathbf{W}_{(x)} = \{w_{(x)_{lm}}\}$. In addition, we have a set of bills as $\mathbf{V}_{(y)} = \{y_n\}_{n=1}^{N}$. Accordingly, the bills form a graph in the semantic space, where the set of vertices $\mathbf{V}_{(y)}$ represents the bills and the set of edges $\mathbf{E}_{(y)} = \{e_{(y)_{nk}}\}_{n,k=1}^{N} \subset \mathbf{V}_{(y)} \times \mathbf{V}_{(y)}$ connects bills with the weight matrix $\mathbf{W}_{(y)} = \{w_{(y)_{nk}}\}$ as their semantic similarity. The last piece of information we want to leverage into the graph formulation is the initially given set of votes, i.e., the *yea* or *nay* results for the legislators voting on the bills. Since each vote involves two types of vertices, one legislator, and one bill, the vote can be viewed as a special type of directed edge or link across these heterogeneous vertices. This gives the third component of the heterogeneous graph formulation, a bipartite structured vote graph $\mathcal{G}_{(xy)} = \{\mathbf{V}, \mathbf{E}_{(xy)}, \mathbf{W}_{(xy)}\}$, where $\mathbf{V} = \mathbf{V}_{(x)} \cup \mathbf{V}_{(y)}$, $\mathbf{E}_{(xy)} = \{e_{(xy)_{ln}}\} \subset \mathbf{V}_{(x)} \times \mathbf{V}_{(y)}$ $(l = 1, \ldots, L, n = 1, \ldots, N)$, and $\mathbf{W}_{(xy)}$ indicates the binary votes.

Legislators' Social and Political Relations

Social connections among the members of the House and Senate have been well studied in fields like social science and political science because they illuminate information for estimating political relevance and revealing the underlying legislative patterns (Fowler 2006). Different kinds of social connections, such as friendship, family, and acquaintanceship relations, have been identified as important effects on political positions (Beck et al. 2002). However, predicting roll call data is about understanding legislators' ideology more than social relationships between them (Fowler 2006). Therefore, here we use cosponsorship relations as a more robust and direct way to understand the voting behavior and political influence of each legislator. In particular, the pairwise political similarity $w_{(x)_{lm}} \in \mathbb{R}$ refers to the affinity between legislators x_l and x_m, i.e., the weight of the edge $e_{(x)_{lm}}$. Assume that legislators x_l and x_m have a total of c_{lm} cosponsored bills in common and have C_l and C_m individually cosponsored bills. Then the value of $w_{(x)_{lm}}$ computed from the cosponsorship information is $w_{(x)_{lm}} = \frac{c_{lm}}{C_l + C_m}$, where the normalized cosponsorship matrix $\mathbf{W}_{(x)} = \{w_{(x)_{lm}}\}$ presents an intuitive way to estimate the political connectivity of the legislators.

Legislative Prediction via Random Walks

Given some observations of the votes, the objective is to predict the missing votes, especially those for the future new bills. To accomplish such challenging prediction tasks, two major assumptions are made to support the prediction model.

1. **Political affinity connects legislative behavior**. Legislators who have strong affinity, e.g., strong cosponsorship relations, tend to vote similarly on a set of bills.
2. **Legislative behavior is consistent among similar bills**. Semantically similar bills tend to receive the same voting results from legislators.

These two assumptions bring two views of predicting the votes along the column and row directions of the weight matrix $\mathbf{W}_{(xy)}$. Along the row direction, the vote $w_{(xy)_{ln}}$ is estimated based on the known votes of the most similar bills from the same legislator, while along the column direction, the prediction of $w_{(xy)_{ln}}$ is accomplished based on the observed votes of the same bill generated by the most similar legislators.

For the unipartite subgraphs, i.e., legislator cosponsorship graph and bill semantic similarity graph, we consider the random walk with restart (RWR) model (Pan et al. 2004; Tong et al. 2006) to derive the steady-state distributions, which indicate the political and semantic relevance among legislators and bills, respectively. Note that during the process of performing RWR, we break the directed vote links and conduct random walks over these two unipartite graphs independently, as described in Wang et al. (2012). Assume we derive the steady-state probabilities $\mathbf{R}_{(x)}$ and $\mathbf{R}_{(y)}$, representing the political relevance among legislators and semantic relevance of bills, respectively. The next step is to predict the possible link from a legislator x_l to a bill y_n. From the view of random walk, the goal is to estimate the transition chance of a random walker starting from vertex x_l and transiting to y_n. However, different from the random-walk model used in either legislator or bill graph, where the walker only performs random transitions in a unipartite graph, here the random walker has to cross a bipartite graph through the existing vote links. There are two possible paths for a random walker across the vertices of the bipartite graph, i.e., transiting from legislator x_l to bill y_n. For *political relevance-based transition*, the random walker first performs transition from x_l to x_m, where x_m has an observed vote link $e_{(xy)_{mn}}$. Then the walker can easily transit to y_n through the existing vote link. For *semantic relevance-based transition*, the random walker first transits from x_l to y_k based on the existing vote link $e_{(xy)_{lk}}$. Then the random walk is performed within $\mathcal{G}_{(y)}$, resulting in a jump from y_k to y_n based on the semantic relevance.

Note that the above two transition paths are related to the two assumptions we made earlier in this section. Finally, the estimation of the transition chance involves a heterogeneous graph, across two types of vertices through three types of edges. In a summary, the update of transition probability $\mathbf{P}_{(xy)}$ of vote links from time t to time $t + 1$ can be written as

$$\mathbf{P}_{(xy)}(t+1) = \gamma \mathbf{R}_{(x)} \mathbf{P}_{(xy)}(t) + (1-\gamma)\mathbf{P}_{(xy)}(t)\mathbf{R}_{(y)}$$

where the coefficient $0 \leq \gamma \leq 1$ is the probability that the random walker will take the first transition path. Due to the existence of the bipartite graph, the above random walk rule is significantly different than the one over unipartite graph. Since the above formulation involves both transitions within each unipartite graph and across the bipartite graph, the steady-state distributions are more complicated to state. Therefore, we proposed an alternative solution to produce the vote predictions iteratively. In particular, through iteratively performing this random walk, where in each step the bipartite-style vote graph is updated with more links, we can gradually complete all the missing vote links (refer to Wang et al. 2012 for details).

Empirical Case Study

Data and Material For the roll call data, a subset of bills and votes from Gerrish and Blei (2011) are used in our experiments. In particular, we use the data from the two most recent congressional sessions, i.e., the 110th (January 2007 to December 2008) and 111th (January 2009 to December 2010). It contains a total of 1,585 bills, 631 unique legislators who have at least one valid vote, and 638,955 *yea* or *nay* votes. Bill cosponsorship information dates back to the 93rd session of the Congress, and we use a version of the data curated by Fowler (2006). In predicting the votes of the 110th and 111th sessions, we only use cosponsorship data up to the 109th session (December 2006) to avoid overfitting.

Results Similar to the experimental setting in Gerrish and Blei (2011), we perform vote prediction at the session level (a 2-year period). In particular, we design two types of experiments: (a) prediction for random missing votes and (b) sequential prediction. For (a)-type experiments, we randomly partition the roll call data from each session into six folds and use standard cross validation to compute the average prediction accuracy. For (b)-type experiments, we use the votes from the first 21 months to predict the votes in the remaining 3 months of each session. The final reported accuracy is the average accuracy in these two sessions.

We compared with the methods used in Gerrish and Blei (2011), including a simple *yea* model predicting all votes as *yea*, a Lasso-based text-based regression models, and IPTM. Table 1. shows the experimental results on the two prediction tasks. In both tasks, the RWHG method achieved the best performance. The existing methods like the text-based regression

Legislative Prediction with Political and Social Network Analysis, Table 1 The accuracy of two different types of experiments. The compared methods include *yea* model, Lasso (L2), IPTM (Gerrish and Blei 2011), and the random walk over heterogeneous graph (RWHG) (Wang et al. 2012)

Method	*Yea*	Lasso	IPTM	RWHG
(a)-Accuracy %	85.48	89.7	88.7	**91.10**
(b)-Accuracy %	86.98	88.1	87.0	**90.36**

methods learn the prediction model through extracting the semantics of bills, which produced higher performance than the baseline *yea* model. However, after considering the political relationship between legislators, the RWHG method could achieve even higher performance due to its unique strength for combining the cosponsorship information with bill semantics synergistically. For example, in (a)-type tests, the RWHG method improves the baseline performance from 85.48 to 91.10 %, which means it can correctly predict around 42,500 more votes in the two-session period.

Key Findings

This article describes a method to use social and political network to improve the predictions of legislative votes. To leverage such types of network information, the roll call data is converted into a heterogeneous graph, where the homogeneous legislator nodes are connected using the cosponsorship information. A special bipartite graph containing directed edges between legislators and bills represents the vote information. Then a two-stage random-walk model performs vote link prediction over the heterogenous graph.

We performed vote prediction tasks on real roll call data from the 110th and 111th congressional sessions. The experimental results clearly show the superior performance of the proposed RWHG method, compared with the state of the art. In addition, we also conducted an empirical study in Wang et al. (2012) and demonstrated that the semantic information of bills provides relatively stronger clues for predicting the votes than the cosponsorship information between legislators. Note that the bag-of-words (BOW) based semantic similarity could be enhanced based on advanced topic modeling algorithms; it was our intent to consider a new prediction model rather than dwell on the fine points of text analysis. We could also use additional features in defining legislator similarity.

Conclusions and Future Directions

The main contributions of Wang et al. (2012) include the unique heterogenous graph formulation of roll call data and the novel random-walk model for vote link prediction. In particular, we summarize the main contributions below:

1. We proposed a unique formulation of heterogeneous graph for legislative prediction. Specifically, we treat both the legislators and the bills as vertices and accordingly generate three subgraphs. Two unipartite subgraphs with homogeneous vertices and edges connecting legislators and bills separately using the cosponsorship information for legislators and semantic similarity for bills. A special bipartite graph containing directed edges between legislators and bills represents the vote information.
2. Based on the above heterogeneous graph formulation, a two-stage random-walk model is proposed to perform vote link prediction. In the first stage, two independent random walks are launched over the legislator cosponsorship graph and the bill semantic similarity graph to derive the relevance between vertices when achieving steady state. In the second stage, we perform random walk across the directed vote links and iteratively generate new vote links from the most confident predictions.

The described model addresses legislative prediction using political network information, but it is indeed a general formulation and can be extended to other applications with any type of social connections. One possible future direction is to enrich this method by bringing in other social information. Temporal analysis of a longer range of the evolution of such political connections is another interesting direction of investigation.

Cross-References

▶ Arts and Humanities, Complex Network Analysis of
▶ Evolving Social Graph Clustering
▶ Extracting and Inferring Communities via Link Analysis
▶ Human Behavior and Social Networks
▶ Political Networks
▶ Social Networking in Political Campaigns

References

Backstrom L, Leskovec J (2011) Supervised random walks: predicting and recommending links in social networks. In: Proceedings of the fourth ACM international conference on web search and data mining, Hong Kong, pp 635–644

Banerjee O, El Ghaoui L, d'Aspremont A (2008) Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. J Mach Learn Res 9:485–516

Beck PA, Dalton RJ, Greene S, Huckfeldt R (2002) The social calculus of voting: interpersonal, media, and organizational influences on presidential choices. Am Polit Sci Rev 96(1):57–73

Clinton J, Jackman S, Rivers D (2004) The statistical analysis of roll call data. Am Polit Sci Rev 98(2): 355–370

Deng H, Han J, Zhao B, Yu Y, Lin CX (2011) Probabilistic topic models with biased propagation on heterogeneous information networks. In: Proceedings of the ACM SIGKDD conference on knowledge discovery and data mining, San Diego, pp 1271–1279

Fowler JH (2006) Connecting the congress: a study of cosponsorship networks. Polit Anal 14(4): 456–487

Gerrish SM, Blei DM (2011) Predicting legislative roll calls from text. In: Proceedings of the international conference on machine learning, Bellevue, pp 489–496

Hasan MA, Chaoji V, Salem S, Zaki M (2006) Link prediction using supervised learning. In: Proceedings of the SIAM conference on data mining workshops, Bethesda

Hinckley B (1972) Coalitions in congress: size and ideological distance. Midwest J Polit Sci 16(2):197–207

Ji M, Han J, Danilevsky M (2011) Ranking-based classification of heterogeneous information networks. In: Proceedings of the ACM SIGKDD conference on knowledge discovery and data mining, San Diego, pp 1298–1306

Kuklinski JH, Elling RC (1977) Representational role, constituency opinion, and legislative roll-call behavior. Am J Polit Sci 21(1):135–147

Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. J Am Soc Inf Sci Tech 58(7):1019–1031

Lichtenwalter RN, Lussier JT, Chawla NV (2010) New perspectives and methods in link prediction. In: Proceedings of the ACM SIGKDD conference on knowledge discovery and data mining, Arlington, pp 243–252

Liu S, Ying L, Shakkottai S (2010) Influence maximization in social networks: an Ising-model-based approach. In: Proceedings of the Allerton conference on communication, control and computing, Monticello, pp 570–576

Lu Z, Savas B, Tang W, Dhillon I (2010) Supervised link prediction using multiple sources. In: Proceedings of the IEEE international conference on data mining, Sydney, pp 923–928

Luck EM, Beaton J, Moffatt JJ (2010) The social media (r)evolution: Obama's political campaign. In: Proceedings of the Global marketing conference, Tokyo

Netrapalli P, Banerjee S, Sanghavi S, Shakkottai S (2010) Greedy learning of Markov network structure. In: Proceedings of the Allerton conference on communication, control and computing, Monticello, pp 1295–1302

Pan JY, Yang HJ, Faloutsos C, Duygulu P (2004) Automatic multimedia cross-modal correlation discovery. In: Proceedings of the ACM SIGKDD conference on knowledge discovery and data mining, Seattle, pp 653–658

Rice SA (1925) The political vote as a frequency distribution of opinion. J Am Stat Assoc 19(145):70–75

Sun Y, Barber R, Gupta M, Aggarwal CC, Han J (2011) Co-author relationship prediction in heterogeneous bibliographic networks. In: Proceedings of the international conference on advances in social network analysis and mining, Kaohsiung, pp 121–128

Taskar B, Wong MF, Abbeel P, Koller D (2004) Link prediction in relational data. In: Advances in neural information processing systems 16. MIT Press, Cambridge

Tong H, Faloutsos C, Pan JY (2006) Fast random walk with restart and its applications. In: Proceedings of the IEEE international conference on data mining, Washington, pp 613–622

Wang E, Liu D, Silva J, Dunson D, Carin L (2011) Joint analysis of time-evolving binary matrices and associated documents. In: Advances in neural information processing systems 23. MIT Press, Cambridge, pp 2370–2378
Wang J, Varshney KR, Mojsilović A (2012) Legislative prediction via random walks over a heterogeneous graph. In: Proceedings of the SIAM international conference on data mining, Anaheim, pp 1095–1106
Zhou D, Orshanskiy SA, Zha H, Giles CL (2007) Co-ranking authors and documents in a heterogeneous network. In: Proceedings of the IEEE international conference on data mining, Omaha, pp 739–744

Legislative Vote Prediction

► Legislative Prediction with Political and Social Network Analysis

Life Coaching

► Network Analysis in Helping Professions

Linkages

► Social Network Analysis in a Digital Age

Link Analysis

► Origins of Social Network Analysis

Link-Based Classification

► Collective Classification

Link-Based Ranking

► Ranking Methods for Networks

Linked Open Data

Thomas Gottron and Steffen Staab
Institute for Web Science and Technologies, Universität Koblenz-Landau, Koblenz, Germany

Glossary

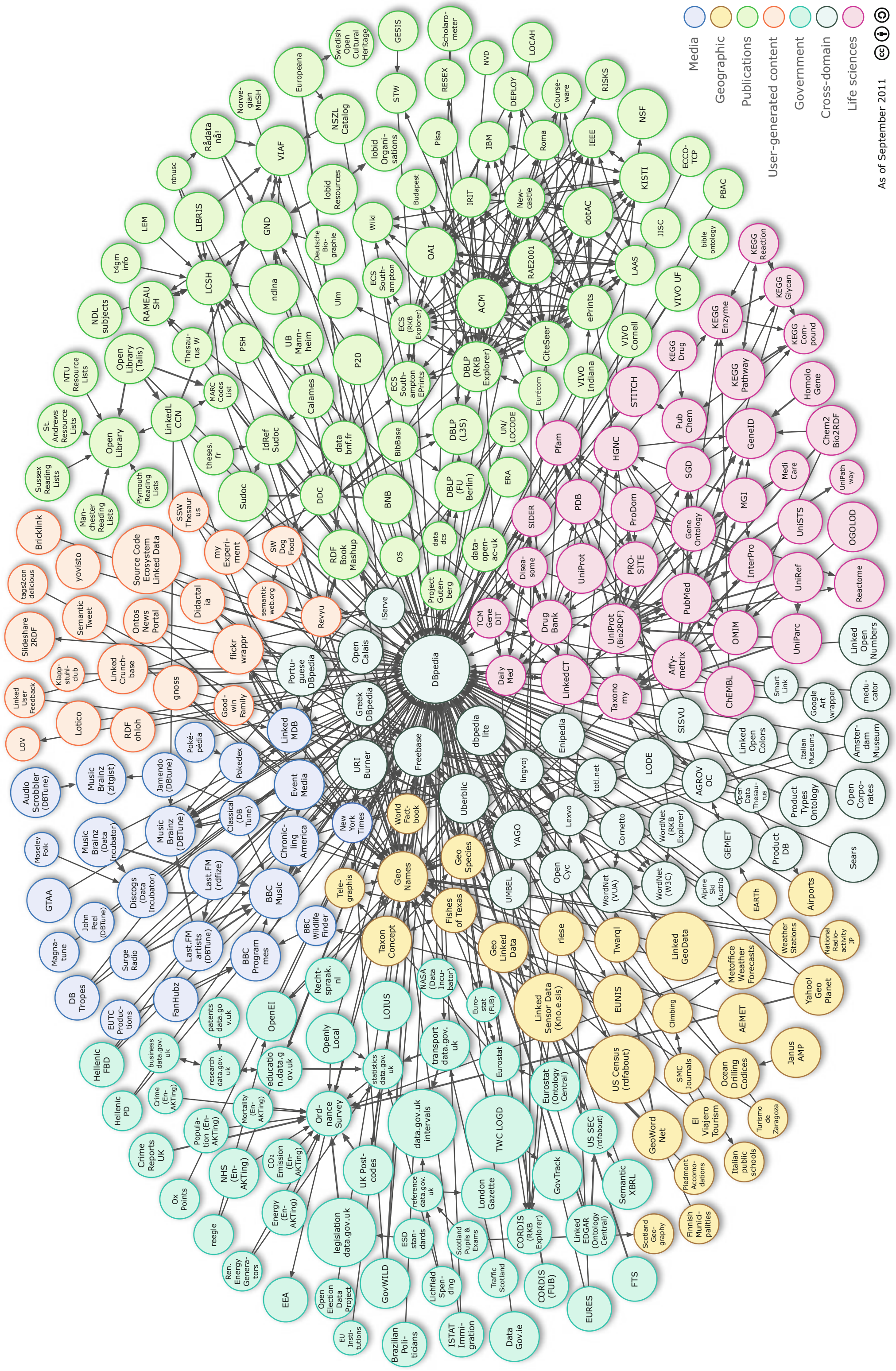
LOD Linked Open Data
URI Uniform Resource Identifier

Definition

Linked Data is the collective term for data following a paradigm based on four simple rules: (a) using URIs to identify entities, (b) using http-URIs allowing to de-reference and look up entities, (c) providing useful information at these URIs based on standard formats (e.g., RDF, SPARQL), and (d) connecting and interlinking to other entities in order to allow for further exploration (Berners-Lee 2007). For the data to fully qualify as Linked *Open* Data (LOD), it should additionally be provided publicly, be available on the web, and be under an open license.

Making use of semantic web formats, LOD implements the vision of a web of data. The underlying technologies allow on the one hand for a globally unique identification of entities via URIs as well as clear semantics of the relations modelled by the links between the entities. As a consequence, LOD should be easier to integrate and reuse in multiple contexts and applications. On the other hand, the underlying technologies are well established in a web context. This means that there are many mature tools, programs, and libraries available to interact and operate with LOD.

While the four high-level Linked Data principles formulated are quite unspecific from a technical point of view, there are various established best practices for publishing and providing LOD on the web. These best practices have evolved mainly around the questions of, for instance, how to implement de-referencing of the http-URIs for



Linked Open Data, Fig. 1 Illustration 1: The Linked Open Data cloud (Cyganiak and Jentzsch 2011) entities using existing web technologies, which useful data to provide when a URI is looked up, how to model data, and what are suitable and best ways to interlink and cross-reference data sets on the web Heath and Bizer (2011).

LOD has seen a tremendous growth in the last years. The resulting distributed graph of interlinked entities on the web is commonly referred to as the LOD cloud (see Fig. 1) and spans hundreds of data sources providing more than 30 billion RDF triples (Weikum 2013). Thereby, LOD covers various domains ranging from media-related content, social networks and user-generated contents, bibliographic data, life sciences, medicine, biology, geographic data, to governmental data. Furthermore, some data sources such as DBPedia and YAGO provide general, cross-domain information and thereby play a pivotal role in connecting also data from very different domains.

Cross-References

▸ RDF

References

Berners-Lee T (2007) Linked Data. http://www.w3.org/DesignIssues/LinkedData.html. Accessed 19 Aug 2013

Cyganiak R, Jentzsch A (2011) Linking Open Data cloud diagram. http://lod-cloud.net/

Heath T, Bizer C (2011) Linked Data: evolving the web into a global data space. Synthesis lectures on the semantic web: theory and technology, vol 1(1), 1st edn. Morgan & Claypool, San Rafael, pp 1–136

Weikum G (2013) Where's the data in the big data wave? http://wp.sigmod.org/?p=786. Accessed 19 Aug 2013

Recommended Reading

Heath T, Bizer C (2011) Linked Data: evolving the web into a global data space. Synthesis lectures on the semantic web: theory and technology, vol 1(1), 1st edn. Morgan & Claypool, San Rafael, pp 1–136

Link: Edge

▸ Community Detection, Current and Future Research Trends

▸ Networks in the Twenty-First Century

Link Prediction

▸ Imputation of Missing Network Data: Some Simple Procedures

▸ Inferring Social Ties

Link Prediction: A Primer

Nitesh V. Chawla[1] and Yang Yang[2]

[1]Department of Computer Science and Engineering, Interdisciplinary Center for Network Science and Applications (iCeNSA), University of Notre Dame, Notre Dame, IN, USA

[2]Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, USA

Synonyms

Edge prediction; Relationship extraction; Social ties inferring

Glossary

Common Neighbors For two nodes, u and v, the set of their common neighbors is defined as $N = \Gamma(u) \cap \Gamma(v)$, and correspondingly the size of set N is $|\Gamma(u) \cap \Gamma(v)|$ (Newman 2001)

Jaccard Coefficient *Jaccard coefficient* (Liben-Nowell and Kleinberg 2007) is a normalization of the *common neighbors* metric, which is defined as $JC(u, v) = \frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}$

Adamic/Adar The *Adamic/Adar* (Adamic Lada and Adar 2003) metric is defined as $AA(u, v) = \sum_{n \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\log|\Gamma(n)|}$, where $n \in N, N = \Gamma(u) \cap \Gamma(v)$ is the set of common neighbors of u and v

Preferential Attachment The *preferential attachment* (Barabasi et al. 2002) metric is the multiplication of nodes u and v's degrees, $PA(u, v) = \Gamma(u).\Gamma(v)$

Katz Leo Katz proposed this metric in Katz (1953); *Katz* metric sums all paths that exist between nodes u and v and penalizes the contribution of long paths exponentially by a factor of β^l, where l is the length of the path. The equation of *Katz* is $katz(u, v) = \sum_{l=1}^{+\infty} \beta^l.|\text{paths}_{u,v}^l|$

Graph Distance The shortest path length between two given nodes u and v

Class Imbalance In the link prediction problem, the class imbalance refers to the inherent disproportion of links that can form to links that do form

ROC A receiver operating characteristic (ROC) represents the performance trade-offs between true positives and false positives by thresholding at different decision boundaries for relative true positive rate and false positive rates

AUROC Area under receiver operating characteristic (ROC) curve (Clause et al. 2008)

Definition: The Link Prediction Problem

Link prediction, that is, predicting the formation of links in a network in the future or predicting the missing links in a network, is an active topic of research. Generally, the link prediction problem can be classified into two categories: (1) predict the likelihood of a future link or missing link between two nodes, knowing that there is no link between them in the current network, and (2) predict whether there will be a future interaction between two nodes that are observed to have an association before. The specific problem discussed in this entry falls into the latter category. Formally, the link prediction problem can be formulated as below:

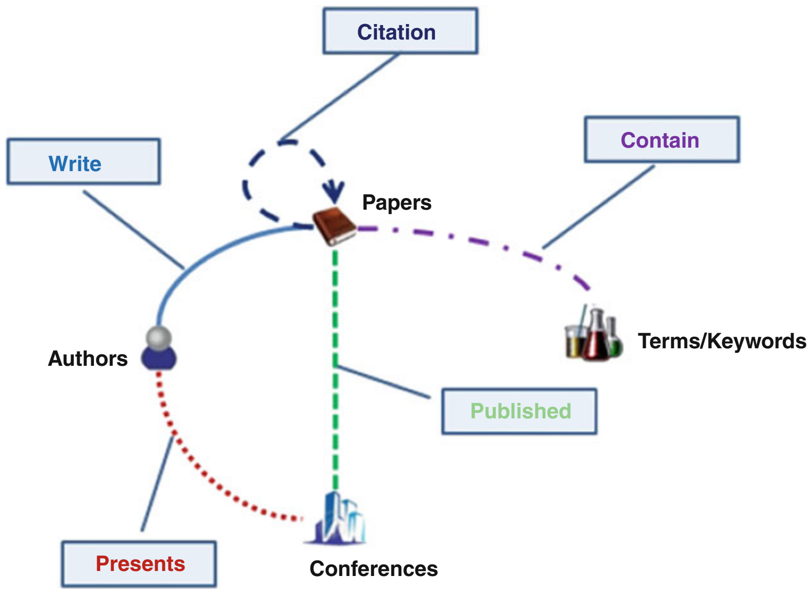
Definition 1 (Link Prediction) In a network G=(V, E), the link prediction task in such network is to predict whether there will be a link between a pair of nodes u and v, where $u, v \in V$ and $e(u, v) \notin E$.

In real world there are two kinds of networks:

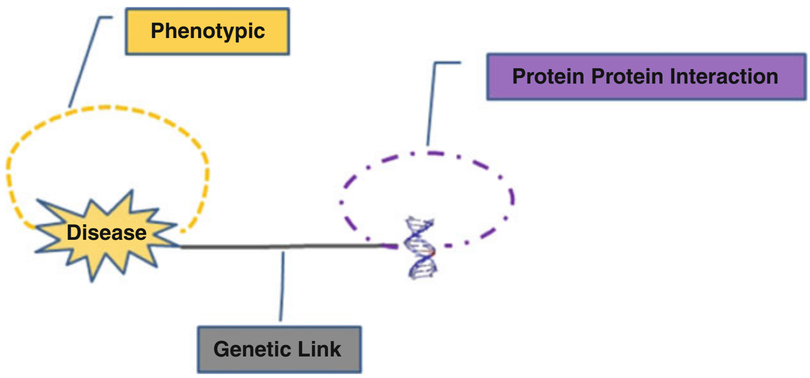
Definition 2 (Heterogeneous Network and Homogeneous Network) A network is defined as a graph $G = (V_1 \cup V_2 \ldots \cup V_I, E_1 \cup E_2 \ldots \cup E_J)$, where $V_i (i \in \{1, 2, \ldots, I\})$ represents the set of nodes of the same object type i and $E_j (j \in \{1, 2, \ldots, J\})$ represents the set of links with the same type j. When the types of nodes $I > 1$ or the types of links $J > 1$, the network is called a **heterogeneous network**; otherwise, it is a **homogeneous network**.

Introduction

Link prediction is an important task in network analysis, benefiting researchers and organizations in a variety of fields. There are a variety of techniques for link prediction, ranging from feature-based classification to probabilistic models and matrix factorization. In this entry, we mainly discuss how to solve the link prediction problem as a *supervised classification* task. Generally, we can categorize link prediction methods into two classes: (1) unsupervised methods, which extract features, such as *common neighbors*, to directly estimate the likelihood of the link, and (2) supervised methods, which train a binary classification model by extracting a set of features. Most of these methods are designed for homogeneous networks; however, many important real-world networks are heterogeneous (Viswanath et al. 2009), such as DBLP bibliographic networks and human disease-gene networks (Figs. 1 and 2) (Deng et al. 2011). The complexity of structural dependency and heterogeneity of links produces obstacles for link prediction in heterogeneous networks; however, it also provides us abundant information to learn from. Well-known topological features designed for homogeneous networks are difficult to apply to heterogeneous networks, such as *common neighbors* and *Adamic/Adar*;



Link Prediction: A Primer, Fig. 1 DBLP bibliographic network



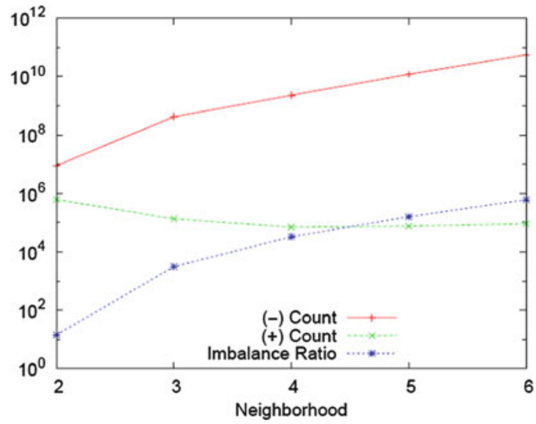
Link Prediction: A Primer, Fig. 2 Disease-gene network

inappropriate solutions lead to a loss of information. A few studies have worked on the link prediction in heterogeneous networks, from the early work (Taskar et al. 2003) to the recent work of Sun et al. (2011), Davis et al. (2011), Lichtenwalter and Chawla (2012), Yang et al. (2012), and Dong et al. (2012).

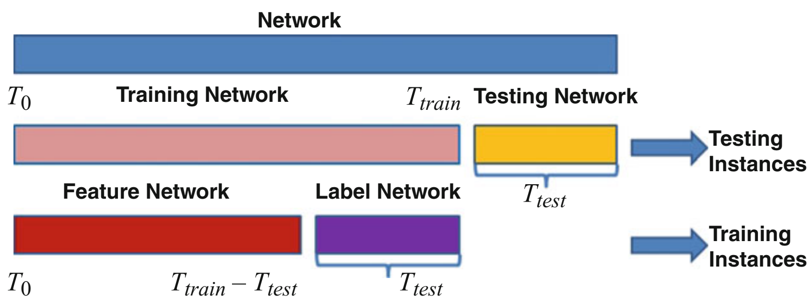
Key Points

In this entry, we mainly discuss the solutions that model the link prediction problem as a *supervised* *classification* task. Certain key steps are required and determine the final prediction performance (Fig. 5):

- Select effective features that describe the link likelihood appropriately. Most of the features are extracted from the network topology; sometimes non-topological features are also included into the set of features, such as location information (Scellato et al. 2011).
- Develop methods of solving class imbalance issue. For example, in Fig. 3, we can see the proportion of (−) count (potential links that do not form) to (+) count (potential links that do form) ranges from 10 to 10^6



Link Prediction: A Primer, Fig. 3 Class imbalance (cell phone network) (Lichtenwalter et al. 2010)



Link Prediction: A Primer, Fig. 4 Experimental setup

in different neighborhoods, which means the ratio between positive instances and negative instances in the classification problem is inherently disproportional.

- Appropriate experimental setup for the extraction of training set and testing set. One of the commonly used configurations is presented in Fig. 4; we use the link information from the training interval (purple area) to make predictions of future links in the test interval (yellow area).

Historical Background

Unsupervised Solutions of the Link Prediction Problem

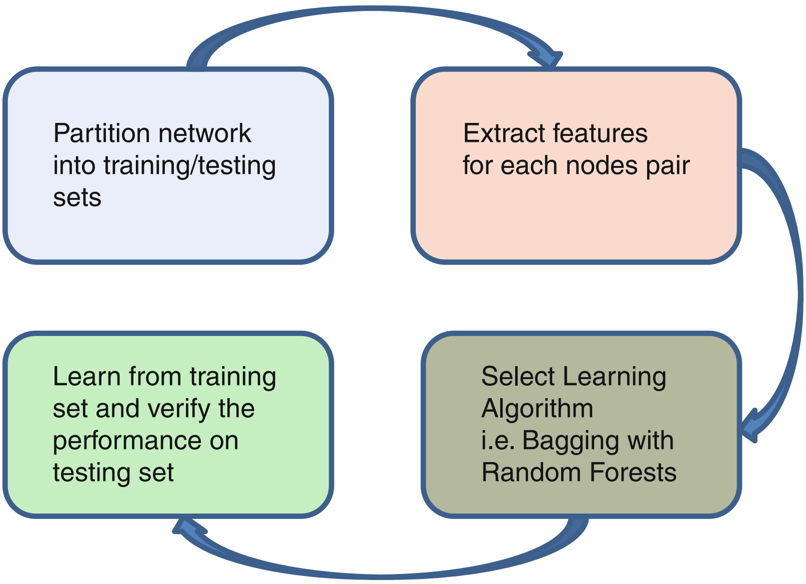
Liben-Nowell and Kleinberg (2007) proposed one of the earliest link prediction methods in social networks. Their approach is based on measures for analyzing the "proximity" of nodes in a network, including *common neighbors*, *preferential attachment*, *Adamic/Adar*, *Jaccard coefficient*, *Katz*, and *graph distance*. In the work of Liben-Nowell and Kleinberg (2007), they proposed there could be better solutions by using machine learning technologies.

Link Prediction as a Supervised Task

As we have discussed, the techniques for the link prediction task range from feature-based classification to probabilistic models and matrix factorization:

Probabilistic Models for Link Prediction: There is a stream of research that focus on modeling the posterior probability of the co-occurrence of the node pairs, such as *Markov random field*-based probabilistic model proposed by Wang et al. (2007) and Kashima's work Kashima et al. (2006) on network evolution-based probabilistic model.

Matrix Factorization for Link Prediction: In the work of (Menon and Elkan 2011; Kunegis and Lommatzsch 2009), the observed graph



Link Prediction: A Primer, Fig. 5 Supervised learning process

is represented as a matrix $\{0, 1, ?\}^{n \times n}$, where 0 denotes a known absent link, 1 denotes a known present link, and ? denotes an unknown status link. The goal is to make predictions for the ? entries by using the *matrix factorization* technique.

Feature-Based Classification for Link Prediction: The link prediction problem can be modeled as a *supervised classification* task (Murata and Moriyasu 2007; Lichtenwalter et al. 2010; Scellato et al. 2011; Sun et al. 2012; Qiu et al. 2011), where each instance corresponds to a pair of nodes in the network. One of the representative works that model the link prediction task as *classification* problem is a general supervised framework called HPLP developed by Lichtenwalter et al. (2010) (Fig. 5). By using the ability of supervised frameworks and oversampling strategies, the HPLP framework was demonstrated to outperform most unsupervised link predictors.

Link Prediction in Social Networks

In the above section we have introduced main streams of the link prediction techniques, which work in *homogeneous* networks. However, many real-world networks are heterogeneous (Figs. 1 and 2), which motivates the research of link prediction in heterogeneous networks. Most topological features for the link prediction in homogeneous networks are difficult to apply in heterogeneous networks. The key idea of a successful heterogeneous link prediction framework is *how to capture the interrelation between different types of links and employ this information effectively and precisely*. In the following sections, we will discuss the most recent solutions of the link prediction task in heterogeneous networks.

Meta Path: Sun et al. (2011) proposed the concept of **meta path**, which is a path defined on the network schema, where nodes are object types and edges are relations between object types. Table 1 lists examples of several meta paths of length 3 between authors in the DBLP network (Fig. 1). In Table 1, two types of *meta paths* between node pairs have different *p value* in describing their latent links. The *meta path*-based features are systematically extracted from the network and used in a supervised model to predict coauthor relationship in a complex DBLP network.

MRLP: Davis et al. (2011) proposed a probabilistic method (**MRLP** (multi-relational link prediction)) that performs a triad census of two nodes in heterogeneous networks, which conveniently translates to a nonarbitrary, data-justified weighting scheme (Fig. 6).

VCP: In the work of Lichtenwalter and Chawla (2012), proposed the concept of a vertex collocation profile (**VCP**) for the purpose of link analysis and prediction. In their definition, $VCP^{n,r}_{v,u}$ is a vector describing the relationship between two nodes v and u, in terms of their common membership in all possible subgraphs of n vertices over r relations. Figure 7 gives all possible elements that are included in $VCP^{3,2}_{s,t}$.

All of these methods are designed to describe the "proximity" between two nodes by capturing as much information as possible in heterogeneous networks. **Meta path** statistically measures the significance level of each meta path in supporting the likelihood of coauthorship for two authors. *MRLP* and *VCP* employ local substructures information to depict the link likelihood of two nodes in the future.

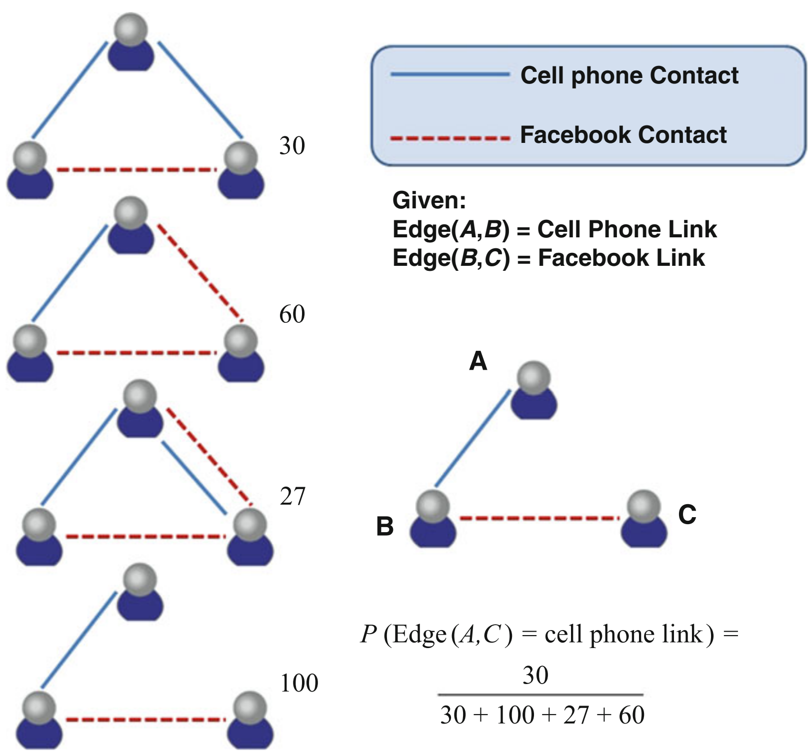
Link Prediction: A Primer, Table 1 Examples of meta path

Meta path	Description	Sig. level
$A - P \rightarrow P - A$	a_i cites a_j	$**$
$A - P \leftarrow P - A$	a_j cites a_i	$***$

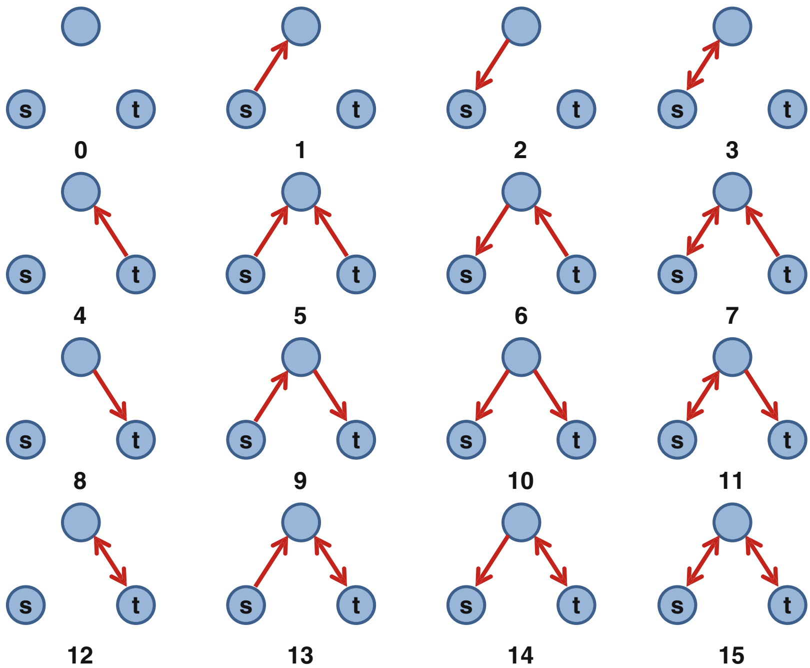
$*p < 0.1$; $**p < 0.05$; $***p < 0.01$; $****p < 0.001$

Key Applications

Beside friendship recommendation (Dong et al. 2012) in social networks, the link prediction has applications in many other domains. In the healthcare domain, the identification of



Link Prediction: A Primer, Fig. 6 MRLP toy example



Link Prediction: A Primer, Fig. 7 16 possible elements in $VCP^{3,2}_{s,t}$

interactions between drugs and target proteins or between diseases and proteins is a key area (Davis and Chawla 2011). In the area of e-commerce, one of the important usages of link prediction is to construct recommendation systems (Liu and Kou 2007).

Future Directions

Link prediction as a research area continues to evolve and is starting to play a key role in understanding and modeling network dynamics. Some of the recent work has included temporality in link prediction. Sun et al. (2012) developed the methodology for predicting when a link will form, not just whether it will form. Yang et al. (2012) showed that considering the temporal information significantly improves the performance of predicting new links. Recently cross-network prediction and transfer learning (Dong et al. 2012; Horvat et al. 2012) become hot topics in the link prediction field. Some of the open challenges remain in scaling the problem to include a much richer attribute space of links and nodes.

Cross-References

▸ Anonymization and De-anonymization of Social Network Data
▸ Distance and Similarity Measures
▸ Inferring Social Ties
▸ Social Recommendation in Dynamic Networks

References

Adamic LA, Adar E (2003) Friends and neighbors on the web. Soc Netw 25(3):211–230

Barabasi A-L, Jeong H, Neda Z, Ravasz E (2002) Evolution of the social network of scientific collaboration. Phys A 311(3–4):590–614

Clause A, Moore C, Newman MEJ (2008) Hierarchical structure and the prediction of missing links in network. Nature 453:98–101

Davis D, Chawla NV (2011) Exploring and exploiting disease interactions from multi-relational gene and phenotype networks. PLoS One 6(7):e22670

Davis D, Lichtenwalter R, Chawla NV (2011) Multi-relational link prediction in heterogeneous information networks. In: ASONAM'11, Kaohsiung

Deng H, Han J, Zhao B, Yu Y, Lin CX (2011) Probabilistic topic models with biased propagation on heterogeneous information networks. In: KDD'11, San Diego

Dong Y, Tang J, Wu S, Tian J, Chawla NV, Rao J, Gao H (2012) Link prediction and recommendation across heterogeneous social networks. In: ICDM'12, Berlin

Horvat E-A, Hanselmann M, Hamprecht FA, Zweig KA (2012) One plus one makes three (for social networks). PLoS One 4:121–134

Kashima H, Abe N (2006) A parameterized probabilistic model of network evolution for supervised link prediction. In: ICDM'06, Leipzig

Katz L (1953) A new status index derived from sociometric analysis. Psychometrika 18(1):39–43

Kunegis J, Lommatzsch A (2009) Learning spectral graph transformations for link prediction. In: ICML'09, Montreal

Liben-Nowell D, Kleinberg J (2007) The link prediction problem for social networks. J Am Soc Inf Sci Technol 58(7):1019–1031

Lichtenwalter R, Chawla NV (2012) Vertex collocation profile: subgraph counting for link analysis and prediction. In: WWW'12, Lyon

Lichtenwalter R, Lussier J, Chawla NV (2010) New perspectives and methods in link prediction. In: KDD'10, Washington

Liu Y, Kou Z (2007) Predicting who rated what in large scale datasets. In: SIGKDD Exploration Newsletter 9:62–65

Menon AK, Elkan C (2011) Link prediction via matrix factorization. In: ECML-PKDD'11, Athens

Murata T, Moriyasu S (2007) Link prediction of social networks based on weighted proximity measures. In: IEEE/WIC/ACM ICWI'07, Vila Real

Newman MEJ (2001) Clustering and preferential attachment in growing networks. Phys Rev Lett E 64:025102

Qiu B, He Qi, Yen J (2011) Evolution of node behavior in link prediction. In: AAAI'11, San Francisco

Scellato S, Noulas A, Mascolo C (2011) Exploiting place features in link prediction on location-based social networks. In: KDD'11, San Diego

Sun Y, Barber R, Gupta M, Aggarwal CC, Han J (2011) Co-author relationship prediction in heterogeneous bibliographic networks. In: ASONAM'11, Kaohsiung

Sun Y, Han J, Aggarwal CC, Chawla NV (2012) When will it happen? Relationship prediction in heterogeneous information networks. In: WSDM'12, Seattle

Taskar B, Wong M-F, Abbeel P, Koller D (2003) Link prediction in relational data. In: NIP, Acapulco, Mexico

Viswanath B, Mislove A, Cha M, Gummadi KP (2009) On the evolution of user interaction in facebook. In: WSON'09, New York, NY, USA

Wang C, Satuluri V, Parthasarathy S (2007) Local probabilistic models for link prediction. In: ICDM'07, Omaha

Yang Y, Chawla NV, Sun Y, Han J (2012) Predicting links in multi-relational and heterogeneous networks. In: ICDM'12, Brussels

Link Rank

► Community Detection in Social Network: An Experience with Directed Graphs

Links

► Mapping Online Social Media Networks

Location

► Spatiotemporal Footprints in Social Networks

Location-Based Online Social Networks: Location-Based Online Social Media, Location-Based Online Social Services

► Privacy Preservation and Location-Based Online Social Networking

Location-Based Recommendation

► Spatiotemporal Personalized Recommendation of Social Media Content

Location-Based Social Networking Services

► Location-Based Social Networks

Location-Based Social Networks

Wang-Chien Lee and Mao Ye
Department of Computer Science and Engineering, Pennsylvania State University, University Park, PA, USA

Synonyms

Geo-social networking services; Location-based social networking services

Glossary

Check-In A record of a user in location based networking service announcing her/his visit to a physical place and sharing this information to her/his friends

GPS Global Positioning System

LBSNs Location Based Social Networking Services

Location Based Services (LBS) Online services that facilitate search and/or access of specific data objects based on the use of location of the user and objects

Social Links The specified connections among users in social networking systems

Social Networking Services Online services that facilitate their users to communicate and to share personal information/opinions with friends

UGGSD User-Generated Geo-Social Data

Definition

Social networking services are online services, platforms, or websites that facilitate their users to communicate and to share information such as interests, ideas, opinions, news, articles, activities, and events photographs with friends. In addition to *personal profiles*, each user on a social networking service specifies a number of *social links* to denote her/his friends and thus forming a large-scale *online social network*. Facebook (http://www.facebook.com), LinkedIn (http://www.linkedin.com), Twitter (http://www.twitter.com), and Google+ (http://plus.google.com) are some well-known examples of social networking services.

Location-based social networking services (LBSNs) are a new type of social networking service emerging in recent years. Owing to the growing availability of positioning technology (e.g., Global Positioning System), various kinds of smart devices (e.g., phones, navigators, cameras) can now easily acquire locations of their users' activities and events. As a result, these users can easily generate and upload various forms of information (e.g., photographs, videos, tweets) associated with locations for sharing and publishing of their happenings, activities, and events. Usually through interesting and easy-to-use "apps" on popular mobile devices, LBSNs add a very important "location" dimension to the conventional online social networks. This addition significantly enriches the information shared on LBSNs and thus allows the service providers to develop more advanced functionalities, e.g., allowing their users to retrieve location-enriched information such as friends nearby based on their current locations. Foursquare (http://www.foursquare.com), Facebook Places (http//www.facebook.com/about/location), Bikely (http://www.bikely.com), Yelp (http://www.yelp.com), and Google Latitude (http://www.facebook.com/latitude) are some representative location-based social networking services.

Users and *locations* are two important components in LBSNs. Users are important because LBSNs allow users to connect with friends, organize activities, share their opinions, and upload photos, videos, and blogs. Different from conventional social networking services that connect users merely in the cyber world, the LBSNs bring people together via both cyber connections and "physical" interactions with locations (or objects associated with locations such as point of interests (POIs) or generally places). Obviously, this new feature is enabled by wireless networks and positioning technologies, which allows users to track and share their personal location and other location-related information on the fly with mobile devices, any time anywhere. Most importantly, it has brought social networking activities from the virtual world back to real life and allow our real-life experiences be shared effortlessly in the virtual world.

As location is one of the most essential elements in our daily life, a lot of novel applications can be supported by LBSNs, e.g., recommending POIs to users. As mentioned, LBSNs have brought people together via both cyber connections and physical interactions with locations. Moreover, along with user and location information, the *interactions* between users and locations have been well captured in LBSNs, which represent a wealth of rich information resources valuable for innovative location-based applications.

Introduction

Rapid technological advances towards "Web 2.0" have led to the growth of *user-generated content* in recent years. This trend has appeared in a wide range of new Web applications such as Wikis, forums, blogging sites, and content-sharing services. Among user-generated data, e.g., blogs, commentaries, photos, and videos, many of them contain *location information*, explicitly or implicitly, along with other rich semantic information. For example, Flickr (http://www.flickr.com) allows users to semantically and geographically tag their photographs and organize them according to locations. EveryTrail (http://www.everytrail.com) allows travelers to share photos and journals along with GPS trajectories. Yelp lets users exchange reviews of POIs, such as merchants and restaurants. Also as part of this Web 2.0 trend, various social and scientific activities, e.g., participatory sensing (Goldman et al. 2009), citizen scientists (NASA Science – Citizen Scientists 2012), and volunteered geographic information (VGI) (Goodchild 2007), have started to make impacts on scientific studies and our society. Among them, VGI has received much attention from the GIS and spatial database communities due to the large volume of valuable *user-generated spatial data* being made available on the Web. Some well-known VGI examples are OpenStreetMap (http://www.openstreetmap.org), Wikimapia (http://www.wikimapia.com), and Google MyMaps (http://maps.google.com/).

Recently, a growing number of LBSNs have emerged. Enabled by the wide availability of smartphones equipped with GPS receivers, these LBSNs allow users to connect with friends, explore places (e.g., restaurants, stores, theaters), share their locations, and upload photos, videos, and blogs. Through these services, users can comment on a restaurant, let friends know where they are, or just to find out if some friends happen to be nearby and want to get together for a cup of tea. Take Foursquare as an example. It annotates a place/location with address, map, visitors, categories, and tags (note that we use locations and places interchangeably in this article while not causing confusions). It allows users to share their visited locations, write tips, and comment on their experiences. Moreover, it provides personalized recommendations and deals to a user based on the places the user and her/his friends like.

Different from conventional social networking services that connect people merely in the cyber world, these LBSNs bring people together via both "cyber connections" and "physical interactions" with places. More importantly, the interactions between users and places have been naturally captured as the user-generated geo-social data (UGGSD) in these services. The aforementioned UGGSD represents a wealth of rich information resources which are valuable for scientific research and innovative applications. For example, OpenStreetMap provides rich information about road map and POIs; Foursquare and Gowalla contain plenty of the footprints of socially connected users, and Yelp provides a good platform for users to share reviews and ratings of places. In addition to the rich location information and social connections noted above, one important trend is the sharing of personal views, opinions, observations, and experiences expressed in various forms (e.g., liked/disliked votes, ratings, commentary articles, and photographs). The collective intelligence embedded in the individual views shared on the social Web is a very valuable resource for many innovative and realistic applications. In the past few years, an increasing number of research on analyzing and mining USGSD in support of location-based services have been conducted and reported in the literature.

This article aims to introduce the state-of-the-art research on analysis and mining of user-generated geo-social data in location-based social networks. The rest of this article is organized as follows. In the section of Historical Background, we briefly trace the LBSNs back to advances in location systems and social networking services. Additionally, we add information about conferences where research on LBSNs are actively pursued. In the section of User-Generated Geo-Social Data, we categorize various forms of data used in analysis and data mining of LBSNs. Next, we discuss some potential LBSN applications in the section of Key Applications. Then, we discuss the ongoing data analysis and mining research on LBSN in the section of Research on Location-Based Social Networks. Finally, we point out some potential research topics in the section of Future Directions.

Historical Background

As its name suggests, location-based social networking services are rooted from location-based services and social networking services. Due to the need to know the physical location of things/persons in many applications, location systems and localization techniques have been developed for both indoor and outdoor usages. The Global Positioning System (GPS), developed in 1970s, is perhaps the most widely used outdoor location-sensing system today, while the research on indoor location systems could be dated back to Active Badge system (Want et al. 1992) in 1989. On the other hand, early efforts to support social networking via Internet appeared in form of various online services (such as emails, Usenet, and bulletin board services). Early social networking on the World Wide Web began in the form of generalized online communities (such as chat rooms and messengers) to bring people together and to interact with each other. Some communities allows people to link to each other and share information. As the Internet and Web technologies evolved, many sites began to develop more advanced features for users to search and manage friends. These early efforts and devolvement all contributed to the success and popularity of today's social networking sites. The technological advances in both location systems and social networking sites naturally spark the emergence of location-based social networking services, which facilitate sharing of user-generated location data (and geo-tagged data objects) to connect users with each other, as well as places and events of their interests.

Research on location-based social networks has also received significant attention from the academia and industry. The First International Workshop on Location-Based Social Networks took off in Seattle in 2009. Meanwhile, the topic of location-based social networks and relevant research issues in various aspects have been actively pursued in major conferences such as International Conference on World Wide Web (WWW), International Conference on Advances in Geographic Information Systems (GIS), International Conference on Research and Development in Information Retrieval (SIGIR), International Conference on Knowledge Discovery and Data Mining (KDD), and International Conference on Data Engineering (ICDE).

User-Generated Geo-Social Data

The user-generated geo-social data (UGGSD) represent a wealth of rich information resources for various applications. For example, OpenStreetMap provides a user-generated map of the whole world for free use; Foursquare and Gowalla allow mobile users to mark their visits to places as "check-ins," and Yelp provides reviews and ratings about local business. Various opinions, experiences, ratings, and check-ins of users have been packaged as information products for user consumption. However, besides these simple information services, hidden collective intelligence of users in various forms of UGGSD could be exploited for supporting advanced applications. Generally speaking, the UGGSD used in research works can be classified as (i) geo-tagged multimedia data, (ii) GPS data, and (iii) user-user/user-place interactions and activity data.

Geo-Tagged Multimedia Data

Many Web 2.0 sites (e.g., Flickr) allow users to easily tag multimedia data objects such as blogs, photographs, and videos with locations. By annotating these multimedia objects with locations and other semantic labels, many useful functions and services are provided in these sites. For example, Panoramio (http://www.panoramio.com), a geolocation-oriented photo-sharing website, allows photos uploaded to the site be accessed as a layer in Google Earth (http://www.google.com/earth/index.html) and Google Maps. As a result, its users can learn more about a given area by viewing the photos via Google Earth and Google Maps. Additionally, many sites facilitate their users to discuss or comment on the shared objects associated with some activities (e.g., touring in Florida or trying out a new restaurant at Davis Square in Boston). Moreover, Weblogs, forums and social communities for travels, (e.g., TravelPod (http://www.travelpod.com), IgoUgo (http://igougo.com), and TravelBlog (http://www.travelblog.com)) make sharing of travel experiences as travelogues easy. These travelogues, usually containing rich travel information such as the tours, lodging, meals, expenses, and weather conditions, contain valuable spatial information or location entities (e.g., the words "New York City") in textual form. In this category of UGGSD, the tagged or extracted locations are mainly considered as an additional attribute associated with the multimedia objects.

GPS Data

Global Positioning System (GPS) plays an important role in bringing location information to many location-based applications and services. For example, users in Foursquare and Facebook Places share their (current) locations by "checking in" to places such as a shopping mall or a coffee shop. As a result, a user can find friends around her/his location and thus is able to initiate some activities, e.g., go shopping or have a coffee together. In this scenario, the GPS coordinates of user's check-ins denote the point of events that happened. In addition to the aforementioned point data, sequences of GPS coordinates are also often collected as *trajectories* to track the movements of users (or other moving objects such as animals or cars). A trajectory consists of a series of sampled GPS points, where each point logs the user's visited locations along with corresponding time stamps. Many sites such as EveryTrail and Bikely, by providing mobile apps on smartphones, encourage users to log and upload their movement trajectory, along with photographs and commentaries/travelogues. Notice that the trajectories do not only provide the basic information in terms of sequences of points but also other information such as the distance, duration, velocity, and time. Additionally, when coupled with sensors and maps, many valuable information may be associated with trajectories to support various applications in health care, transportation, and sports. In this category of UGGSD, the locations (in forms of GPS coordinates) are an inherent attribute associated with points or trajectories.

Interactions/Activity Data

As mentioned earlier, many LBSNs aim to bring their users together via both cyber connections and physical interactions with places. In addition to the check-ins in the aforementioned LBSNs such as Foursquare, Gowalla, and Facebook Places, newly emerged online social services, such as Meetup (http://www.meetup.com), Plancast (http://www.plancast.com), Yahoo! Upcoming (http://upcoming.yahoo.com), and Eventbrite (http://www.eventbrite.com) have provided convenient online services for their users to organize social events and activities, ranging from movie night and dining out to technical conferences and business meetings. Unique from the conventional LBSNs, these new services also promote physical social interactions in groups. More importantly, the events and activities are held in particular places and thus represent interactions among the users and also with places. The interactions between users and places have been captured as UGGSD in these services, in addition to the social links among the users captured as social graph data.

Key Applications

Various opinions, experiences, ratings, and check-ins of LBSN users have been packaged as information products for user consumption. Thus far, most of these services are mostly simple information services that explicitly use user-generated data to allow mobile users marking their visits to places as "check-ins" (e.g., Foursquare), to provide reviews and ratings about local business (e.g., Yelp), to provide a user-generated map of the whole world for free use (e.g., OpenStreetMap), to organize group activities and events (e.g., MeetUp), and to retrieve and display geo-tagged photographs in Google Earth (e.g., Panoramio). However, besides these simple information services, many advanced LBSN services and applications have also been proposed or suggested in the literature (Zheng 2011). In the following, we first list some of these advanced applications and then discuss some approaches proposed to realize these services.

- **Local Expert Search or Recommendation** these services aim to find and recommend experts who are very knowledgeable about certain location-dependent information (e.g., restaurants, events, trails, travel tips) by exploiting the user-generated data shared by users.
- **Friend's Recommendation** this service aims to recommend users who have similar interests, behavior, preferences, and experiences (e.g., visited the same places or taking similar kinds of trails) to the targeted user as friends. This could be done by exploiting user similarity in terms of locations and trips/trajectories taken previously as well as comments and blogs about the locations/trips.
- **Group Activity Initiation and Organization** this service may help to identify groups of users with similar interests when new group activities are organized. This service could be realized by clustering similar users based on common visited places, trips, activities, interests, etc. Additionally, factors such as acquaintance, travel distance, and time availability can be considered in the service (Yang et al. 2012).
- **Personalized POIs/Places Search and Recommendations** these services aim to identify interesting POIs/places for a user who is interested in exploring places and happenings in her/his own city or who is visiting a new city. This kind of services can be supported by recommending the visited locations of similar users (or general public) by mining the user-generated data in LBSNs.
- **Trip Planning** several services may assist LBSN users in planning a trip based on user requirements and preferences. By mining the collective intelligence from various types of UGGSD, such as travelogues, trajectories and check-ins, services may find travelogues containing locations, themes, and activities of interests to the user and plan for a sequence of POIs (or in form of an itinerary) for the user.
- **Events/Activities Discovery, Search, and Recommendation** these services may support discovery of events and activities (e.g., accidents, earthquakes, traffic jams, sports games) which may in turn facilitate search and personalized recommendation of activities and events to users who are looking for activities in certain locations or regions. This kind of services can be supported by aggregating geo-tagged multimedia data objects (such as photographs or tweets) generated by users. Moreover, visited locations of similar users to the targeted user may be recommended by mining the patterns of similar users from user-generated data in LBSNs.
- **Tag Recommendation** the service may assist users in tagging multimedia objects shared in LBSNs. To facilitate efficient search and recommendation in LBSNs, semantic tags that annotate data objects to be searched play a critical role. However, a challenging issue, especially for new LBSN sites, is that a lot

of objects are missing tags. Thus, many incentives or techniques have been provided to encourage users to tag the objects. Tag recommendation service can be developed to make the tagging process easy. This kind of services can be supported by mining the patterns and phenomenons in UGGSD in LBSNs.

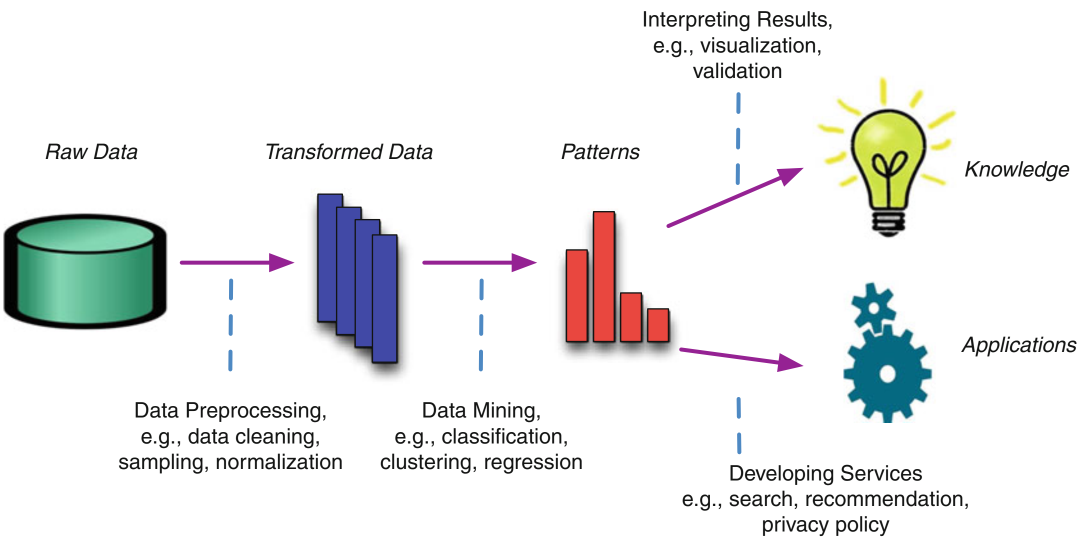
Research on Location-Based Social Networks

As indicated in our discussions on various types of data and applications, analysis and mining of the UGGSD provide important technical foundation for provisioning of many key applications and services in LBSNs. The process of mining UGGSD usually consists of four or five steps as shown in Fig. 1. As shown, the collected raw data are first preprocessed (e.g., by data cleaning, sampling, and normalization) and transformed into a better form. Then data mining algorithms are developed and applied over the transformed data to uncover patterns and hidden structures in the data. Via visualization and validation, the results are analyzed and distilled for knowledge discovery. Meanwhile, the discovered patterns and/or knowledge are employed to develop applications.

Notice that different types of data may require different data preprocessing methods for data transformation. Additionally, various application scenarios may have to deal with different UGGSD and face unique research problems. Indeed, research on analysis and mining of UGGSD in LBSNs have received significant attention from the research community in the past several years. A rich set of data mining algorithms and techniques have been devised to address many interesting and challenging research problems. In this section, we provide a general overview of related works in spatio-social analysis and data mining in LBSNs. Meanwhile, we introduce some selected research studies in more details.

Spatio-Social Data Analysis

The recent emergence of LBSNs such as Foursquare, Gowalla, and Facebook Places provides a unique opportunity for analytical studies of the social and spatio-temporal characteristics exhibited in LBSN users' behaviors and mobility patterns. In Backstrom et al. (2010), a spatio-social property is observed from Facebook dataset, i.e., the probability of friendship between two socially linked users is roughly inversely proportional to the distance between their home locations. Based on this observation, an effective algorithm for predicting a user's home location is proposed. Scellato et al. also observe some spatial heterogeneous characteristics existing in the social relationships (e.g., social ties and social triangles) of users (Scellato 2011). For example, about 10 % of users have social ties with an average friend distance of 10 km, whereas about 20 % of users have social ties with an average friend distances over 2,000 km. On the other hand, at least 20 % of users have social triangles with an average triangle length less than 100 km, whereas the 20 % of users have social triangles with an average triangle length over 2,000 km. Particularly, users with more friends tend to create triads with individuals located farther away than expected by random. In 2010, Ye et al. conduct a spatio-social analysis on real data collected from Foursquare and observe that there exist strong social and geospatial ties among users and their visited locations in the system. In 2011, McGee et al. propose to investigate the relationship between geographical distance and tie strength between two users by analyzing data collected from Twitter. Activities such as following, mentioning, and actively engaging in conversations with other users are taken as strong signals to determine the geographical distance between a pair of users. In 2011b, Cheng et al. analyze human mobility patterns by collecting data from Foursquare and Gowalla. They find that users follow the LevyFlight mobility pattern and also exhibit regular behaviors (Brockmann et al. 2006). Furthermore, they find that in addition of geographical and economical constraints, social network structure also has an impact on the mobility patterns. In 2011, Cho et al. find that user's movement in LBSNs is a combination of periodic movement and random jump, where periodic movement is due to the constraint of geographical proximity and random



Location-Based Social Networks, Fig. 1 A process of data mining and knowledge discovery for LBSN applications

jump is correlated with the social activities. More specifically, short-ranged travel exhibits both spatial and temporal regularity and thus is not affected by social relationships, while long-distance travel is more easily affected by social ties.

As hinted in the above studies, correlations exist between user's spatial activities and their social relationships. Research in Onnela and Saramäki (2007) shows that there is a strong correlation between physical space and social network structures in many diverse settings. Moreover, based on observations of users' spatial activities, it is possible to infer the social tie among users (Eagle et al. 2009; Wang et al. 2011). Authors in Eagle et al. (2009) show that there are distinctive temporal and spatial patterns in physical proximity and calling patterns for those people who have close social connection. Thus the temporal and spatial patterns are strong signals for social tie inference. In 2011, Wang et al. find that human mobility pattern could serve as a good predictor for formation of new social ties. Moreover, the combination of both mobility and network measures can significantly improve the social tie inference accuracy. Finally, Crandall et al. (2010) conduct similar research by using a Flickr dataset. They investigate the extent to which social ties between people can be inferred from co-occurrence in time and space.

Data Mining Techniques for Knowledge Discovery and Applications

In addition to data analysis, many researchers have ridden on the analytical findings and observations to mine the UGGSD in order to support knowledge discovery and various applications and services. As UGGSD tend to capture various kinds of implicit information about users and their interactions with places, many research work focus on developing data mining techniques for user profiling, such as detecting users' locations and modeling user preferences in places/point of interests. These research works are particularly useful for supporting location-based services such as search or recommending places or activities to users. Additionally, research on automatic extraction of location names and suggestion of semantic tags from UGGSD in LBSNs have been actively pursued as they are also very important and useful for provisioning of services in LBSNs. In the following, we provide overviews on research works in these two areas and introduce two selected studies, correspondingly.

User Location Detection and Place Recommendation

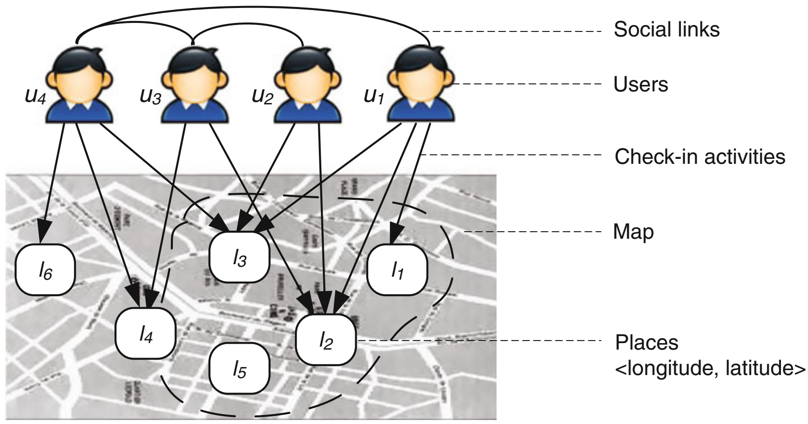
User profiling is essential for personalized search and recommendation services. Thus, detecting user locations is very important for profiling users in LBSNs. In 2011, Hecht et al. study the behavior of users with respect to the location attribute in their Twitter profiles. They find that 34 % of users do not provide real location information, e.g., frequently incorporating fake locations. Nevertheless, they also find user tweets can help to infer the user's location with decent accuracy. In 2010, Cheng et al. use similar ideas to model the spatial distribution of words in tweets to predict the user's location. With the information about user's location, better personalized location-based services can be provided for users.

Personalized recommendation techniques for assisting users to explore new places/point-of-interests have also received a lot of interests from the research community (Ye et al. 2010, 2011c; Zheng et al. 2011; Chow et al. 2010). In 2011, Zheng et al. propose to provide two types of location recommendation services to individual users. The former is generic as it basically recommends to users the most popular locations and travel itinerary within a given spatial region. The latter stresses on personalization as it makes recommendations to a targeted user based on the user's travel preferences (i.e., via the idea of collaborative filtering). In 2010 and 2011c, Ye et al. propose to incorporate the factors of social influence into recommendation algorithms, as users may turn to their friends for opinions sometimes and thus get influenced by friends. Particularly, in Ye et al. (2011c), geographical influence, in addition of social influence, has been considered to improve performance of place recommendation. In 2010, Chow et al. introduce GeoSocialDB to support scalable location-based social networking services, such as location-based news feed and ranking as well as location-based place recommendation.

Here we introduce an approach taken in Ye et al. (2011c) to support a *place recommendation* service in LBSNs which aims at recommending new places/POIs to users in order to help them explore new places and know their cities better. This work is innovative because the authors exploit the unique geographical implications embedded in users' interactions with locations, in addition to applying the social influence from users' friends, for place recommendations in LBSNs.

As shown in Fig. 2, users and places are two essential types of entities in LBSNs, where users (denoted as u_1, u_2, u_3, u_4) are interconnected via social links to form a user social network and places (denoted as $l_1, l_2, \ldots, l_6$) are connected with users via their "check-in" activities, which generally reflects the users' preferences on places. Also illustrated in the figure, the places (geocoded by $\langle$longitude, latitude$\rangle$) are constrained geographically. Obviously the previous user check-in activities are very useful for recommending places to users. With the check-ins available in LBSNs, the conventional *collaborative filtering (CF)* techniques can be adopted by treating places as the "items" (which is the general term used in CF algorithms to refer to things to be recommended). The basic idea is that users' preferences to places can be deduced from other users who exhibit similar visiting behaviors to places in previous check-in activities. Thus, user-based or item-based collaborative filtering techniques may be applicable (Adomavicius and Tuzhilin 2005). Additionally, the social relationships/network of users, which are handily available in the LBSNs, can also be explored to enhance performance of place recommendations. Especially, in recent years, several studies have shown that social friends tend to share common interests and thus explored the social friendship in the process of collaborative filtering for making recommendations (Konstas et al. 2009; Ye et al. 2012).

While the ideas above aim to explore the essential information available in LBSNs, i.e., the user-location interactivities and user-user social links, the *geographical influence* naturally existing in the activities of users and their geographical proximities cannot be ignored. According to Tobler's first law of geography, "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970). Thus, a user intuitively tends to visit nearby



Location-Based Social Networks, Fig. 2 Social links between friends and check-in activities between users and places in an LBSN

places. Due to the geographical and social nature of the LBSNs, the following influences, including (i) the geographical influence between users and places, (ii) the geographical influences among places and (iii) the social influence among users, can be factored in for places recommendations in LBSNs.

The approach adopted in Ye et al. (2011c) is to develop an integrated collaborative recommendation framework that discover places of users' interests by considering (a) user preferences, (b) social influence, and (c) geographical influence. The users' implicit preferences of places are derived from commonly visited places (recorded in their check-in activities), i.e., two users who have checked into a lot of common places are considered as similar users. As a result, a preferred place l of a user u can be recommended to another user u' who exhibits very similar check-in behavior as u. The social influence between users are captured and compared in two forms based on the number of common friends and the ratio of common friends, respectively. Moreover, the geographical influence of places on a user is explored by investigating the "geographical clustering phenomenon" of user check-in activities in LBSNs, and the geographical influence is learned as a power-law probabilistic model and incorporated via Bayesian theory. Finally, a unified location recommendation framework is developed to fuse user preference to places along with the social influence among users and the geographical influence among places. Empirical study upon two large-scale real datasets, collected from Foursquare and Whrrl (which is an LBSN acquired by Groupon (http://www.groupon.com) in 2011), shows that the proposed framework is more effective than the other state-of-the-art place recommendation techniques.

Semantic Annotation of Locations

Meaningful descriptions and tags of places provide valuable clues/information for location-based services, e.g., location search and place recommendation. As tags of places in LBSNs are mostly generated by volunteered users, there are noises existing in the UGGSD in LBSNs. For example, a research study shows that about 30 % of places do not have any tag or text description in the dataset collected from Foursquare and Whrrl (Ye et al. 2011b). Thus, there is a need to annotate the places with no tags in LBSNs. In 2005, Liao et al. propose a relational Markov network to label places with the activities that have occurred in these places. In Liao et al. (2007), the authors hierarchically construct conditional random fields to extract and label places simultaneously. However, only four types of names (i.e., work, home, friend, parking) are included in the label sets. In 2011, Lian et al. treat the location-naming problem as the location-based search problem. They propose a local search framework to integrate different kinds of popularity factors and personal preferences and apply a learning to rank technique to realize location-naming services. In 2010, Lin et al. model the naming preference when sharing locations with others to study the factors that influence location-naming preference, including a place's entropy, social group, recipient's familiarity to sharing places, and sharer's comfortable level of sharing locations. In 2011a, Cheng et al. study location-based traffic patterns revealed through location-sharing services and find that these traffic patterns can identify semantically related locations to help predicting the semantic category of uncategorized locations. The work in Ye et al. (2011b) aims to address the problem of semantic annotation for places in location-based social networks such as Foursquare and Whrrl, by treating it as a multi-label classification problem. The authors investigate the check-in patterns corresponding to each individual place and behavioral regularity of each individual user to derive features for classification. Besides, the study in Ye et al. (2011a), investigating the temporal dimension of feature types in location-based social network, could be used to determine tags for places.

Research also explore geo-tagged photographs in Flickr to support location-based services (Jaffe et al. 2006; Rattenbury and Naaman 2009). In 2009, Rattenbury et al. develop methods for extracting place semantics according to the tags that are assigned to photos taken in near that place. In 2006, Jaffe et al. propose a framework for automatically selecting a summary set of photos from a large collection of geo-referenced photographs. The summary algorithm is based on spatial patterns in photo sets, as well as textual-topical patterns and cues of user/photographer identities. The algorithm can be expanded to incorporate social, temporal, and other factors. In 2007, Ahern et al. investigate how to analyze the tags associated with the geo-referenced Flickr photos to generate aggregated knowledge in the form of "representative tags" for arbitrary areas in the world. They use these tags to develop a visualization tool, World Explorer, for exploring the content of the data via a map interface which displays the derived tags and the original photo items. In 2009, Crandall et al. investigate how to organize a large collection of geo-tagged photos. Their approach adopts content analysis (based on text tags and image data) and structural analysis (based on geospatial data). In 2010, Sizov et al. propose a multimodal Bayesian model which combines text features (e.g., tags as a prominent example of short, unstructured text labels) with spatial knowledge (e.g., geo-tags and coordinates of images and videos), aiming to construct better algorithms for content management, retrieval, and sharing. In 2011, Yin et al. propose to discover and compare geographical topics from GPS-associated document such as Flickr photos. They develop a latent geographical topic analysis approach (LGTA) that combines location and text, where this approach not only finds regions of interests but also provides effective comparisons of the topics across different locations.

Next we introduce the approach taken in Ye et al. (2011b) to automatically annotate the places with *semantic tags*, which is motivated by observing about 30 % of all places in Whrrl and Foursquare lacking any meaningful textual descriptions or tags. Similar to Ye et al. (2011c), this research tries to explore the user behavior exhibited in check-ins captured in LBSNs, which has been logically represented in Fig. 2. As shown, the user-place relationship/check-ins are captured as a bipartite graph. However, different from the Foursquare data used in Ye et al. (2011c), the check-in information also contains the "time" in addition to "who" and "where." Let U and P denote the sets of all users and places in the system, respectively. Users and places are connected through a set of check-in activities $C = \{\langle u, p, h\rangle | u \in U \wedge p \in P \wedge h \in H\}$, where H is a set of time stamps. Each check-in $c \in C$ describes that "a user u has checked in a place p at time h." Additionally, a place p_i may be annotated by users with a set of semantic tags $T_i \subseteq T$, where T presents the universal space of tags in an LBSN.

There are plenty of semantic tags, e.g., restaurant, shopping, nightlife, and hotel, and a place may be associated with multiple tags. For instance, a place associated with a tag *restaurant* may also be tagged with *bar*. Hence, the task of annotating places in LBSNs with semantic tags may be addressed as a *multi-label classification* problem, which can be addressed by learning a binary supporting vector machine (SVM) for each tag in the tag space. Thus, the key issue becomes what features can be used to learn the SVMs discriminatively. Selecting the right features is critical as they have a significant impact on the effectiveness of the trained SVMs. In this work, the user behaviors and unique features of places captured in the check-in activities are explored.

Several observations are made in this work: (i) human behaviors are not completely random, e.g., people usually visit restaurants for lunch at around noon, and (ii) some users exhibit patterns in their activities, e.g., some people may visit similar places (i.e., those that have the same categories of tags) at the same time. By leveraging these observations, features of places are extracted in two complementary forms: (1) explicit patterns (EP) at individual places and (2) implicit relatedness (IR) among similar places. The EP features, corresponding to a given place, are obtained from all check-ins at the place via a statistical analysis on the total number of unique visitors, the total number of check-ins, the maximum number of check-ins by a single visitor, the distribution of check-in time in a week, and the distribution of check-in time in a day (24 h scale). On the other hand, the IR features aim to capture the relatedness among places by exploiting the regularity of user check-in activities to similar places.

Extracting the EP features is quite straightforward. Thus, the technical contribution made in this work is how to derive the IR features of a place without tags from its "related" places. To facilitate extraction of IR features, a *network of related places* (NRP) that captures the relatedness among places is built by exploring regularities of user check-ins to similar places. To build an NRP, two graphs that capture the user-place and time-place relationships from the user check-in activities are derived and the Random Walk and Restart technique Tong et al. (2006) is run on these graphs to estimate the relatedness between pairs of places. In the obtained NRP, place pairs with high relatedness values imply high similarity in the tag space and thus are linked. Moreover, the probability for a specific tag being labeled to a place (called *label probability*) is determined from its linked (i.e., similar) places as the IR feature. Finally, both the EP and IR features are used to train the binary SVMs.

Through an experimental study using a real dataset collected from Whrrl, it has been shown that using all EP and IR features achieves the better performance than using EP or IR features alone. More importantly, both of the EP and IR features contribute significantly in unique aspects to the classification of different semantic tags. For example, the EP features are effective in labeling tags such as *restaurant* and *nightlife* because most people exhibit similar behaviors in visiting restaurants and nightlife places. On the other hand, the IR features are excellent for tagging places related to *shopping* as some people exhibit strong patterns in such activities.

Future Directions

As pointed out in this article, UGGSD in LBSNs represent a wealth of rich information resources which are valuable for scientific research and innovative applications. There are many potential applications that can be realized not only by exploiting the relationships between user and user, place and place, and user and place. While some pioneering research have been reported in recent years, they mostly focus on a small number of simple data types and thus have only scratched the surface of the field. More research studies are needed to incorporate the heterogeneous and dynamic content associated with users, places, and their interactions in LBSNs. For example, how to deal with the dynamic evolving social network among users and the frequent updates of visit to places is a challenging issue.

Acknowledgments

The authors thank the editors, professors Hady Wirawan Lauw and Eepeng Lim, and anonymous reviewers for their constructive comments to improve the quality of this article.

Cross-References

- ▶ Detection of Spatiotemporal Outlier Events in Social Networks
- ▶ Futures of Social Networks: Where Are Trends Heading?
- ▶ Location-Based Status Updates and Camera Phone Apps in Social Networks
- ▶ Mobile- and Context-Aware Applications of Social Networks
- ▶ Modeling and Analysis of Spatiotemporal Social Networks
- ▶ Modeling Social Preferences Based on Social Interactions
- ▶ Recommender Systems Using Social Network Analysis: Challenges and Future Trends
- ▶ Recommender Systems: Models and Techniques
- ▶ Social Influence Analysis
- ▶ Social Web Search
- ▶ Spatiotemporal Footprints in Social Networks
- ▶ Spatiotemporal Personalized Recommendation of Social Media Content

References

Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans Knowl Data Eng 17(6):734–749

Ahern S, Naaman M, Nair R, Yang JH-I (2007) World explorer: visualizing aggregate data from unstructured text in geo-referenced collections. In: Proceedings of the 7th ACM/IEEE-CS joint conference on digital libraries, Vancouver, pp 1–10

Backstrom L, Sun E, Marlow C (2010) Find me if you can: improving geographical prediction with social and spatial proximity. In: Proceedings of the 19th international conference on World Wide Web, Raleigh, pp 61–70

Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. Nature 439:462–465

Cheng Z, Caverlee J, Lee K (2010) You are where you tweet: a content-based approach to geo-locating twitter users. In: Proceedings of the 19th ACM international conference on information and knowledge management, Toronto, pp 759–768

Cheng Z, Caverlee J, Kamath KY, Lee K (2011a) Toward traffic-driven location-based web search. In: Proceedings of the 20th ACM international conference on information and knowledge management, Glasgow, pp 805–814

Cheng Z, Caverlee J, Lee K, Sui DZ (2011b) Exploring millions of footprints in location sharing services. In: AAAI international conference on weblogs and social media (ICWSM), Barcelona

Cho E, Myers SA, Leskovec J (2011) Friendship and mobility: user movement in location-based social networks. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining, San Diego, pp 1082–1090

Chow C-Y, Bao J, Mokbel MF (2010) Towards location-based social networking services. In: Proceedings of the 2nd ACM SIGSPATIAL international workshop on location based social networks, San Jose, pp 31–38

Crandall DJ, Backstrom L, Huttenlocher DP, Kleinberg JM (2009) Mapping the world's photos. In: Proceedings of the 18th international conference on World Wide Web, Raleigh, pp 761–770

Crandall DJ, Backstrom L, Cosley D, Suri S, Huttenlocher D, Kleinberg J (2010) Inferring social ties from geographic coincidences. Proc Natl Acad Sci 107(52):22436–22441

Eagle N, Pentland AS, Lazer D (2009) Inferring friendship network structure by using mobile phone data. Proc Natl Acad Sci 106(36):15274–15278

Goldman J, Shilton K, Burke J, Estrin D, Hansen M, Ramanathan N, Reddy S, Samanta V, Srivastava M, West R (2009) The Wilson foundation report – participatory sensing. http://wilsoncenter.org/sites/default/files/participatory_sensing.pdf

Goodchild MF (2007) Citizens as sensors: the world of volunteered geography. Geo J 69(4):211–221

Hecht B, Hong L, Suh B, Chi EH (2011) Tweets from Justin Bieber's Heart: the dynamics of the location field in user profiles. In: Proceedings of the 2011 annual conference on human factors in computing systems, Vancouver, pp 237–246

Jaffe A, Naaman M, Tassa T, Davis M (2006) Generating summaries and visualization for large collections of geo-referenced photographs. In: Multimedia information retrieval, Santa Barbara, pp 89–98

Konstas I, Stathopoulos V, Jose JM (2009) On social networks and collaborative recommendation. In: Proceedings of the 32nd international ACM SIGIR conference on research and development in information retrieval, Boston, pp 195–202

Lian D, Xie X (2011) Learning location naming from user check-in histories. In: Proceedings of the 19th SIGSPATIAL international conference on advances in geographic information systems, Chicago, pp 112–121

Liao L, Fox D, Kautz HA (2005) Location-based activity recognition using relational markov networks. In: Proceedings of the 19th international joint conference on artificial intelligence, Edinburgh, pp 773–778

Liao L, Fox D, Kautz HA (2007) Extracting places and activities from GPS traces using hierarchical conditional random fields. Int J Robot Res 26(1):119–134

Lin J, Xiang G, Hong JI, Sadeh NM (2010) Modeling people's place naming preferences in location sharing.

In: Proceedings of the 12th ACM international conference on ubiquitous computing, Copenhagen, pp 75–84
McGee J, Caverlee J, Cheng Z (2011) A geographic study of tie strength in social media. In: Proceedings of the 20th ACM international conference on information and knowledge management, Glasgow, pp 2333–2336
NASA Science – Citizen Scientists (2012) http://science.nasa.gov/citizen-scientists/. Accessed 18 Nov 2012
Onnela JP, Saramäki J (2007) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci 104:7332
Rattenbury T, Naaman M (2009) Methods for extracting place semantics from flickr tags. ACM Trans Web 3(1):1
Scellato S (2011) Beyond the social web: the geo-social revolution. ACM SIGWEB Newsl 2011: 5:1–5:5
Sizov S (2010) Geofolk: latent spatial semantics in web 2.0 social media. In: Proceedings of the third ACM international conference on web search and data mining, New York, pp 281–290
Tobler WR (1970) A computer movie simulating urban growth in the detroit region. Econ Geogr 46:234–240
Tong H, Faloutsos C, Pan J-Y (2006) Fast random walk with restart and its applications. In: Proceedings of the sixth international conference on data mining, Hong Kong, pp 613–622,
Wang D, Pedreschi D, Song C, Giannotti F, Barabási A-L (2011) Human mobility, social ties, and link prediction. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining, San Diego, pp 1100–1108
Want R, Hopper A, Falcão V, Gibbons J (1992) The active badge location system. ACM Trans Inf Syst 10(1):91–102
Yang D-N, Shen C-Y, Lee W-C, Chen M-S (2012) On socio-spatial group query for location-based social networks. In: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining, KDD'12, Beijing. ACM, New York, pp 949–957
Ye M, Yin P, Lee W-C (2010) Location recommendation for location-based social networks. In: Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems, GIS'10, San Jose. ACM, New York, pp 458–461
Ye M, Janowicz K, Mülligann C, Lee W-C (2011a) What you are is when you are: the temporal dimension of feature types in location-based social networks. In: Proceedings of the 19th SIGSPATIAL international conference on advances in geographic information systems, Chicago, pp 102–111
Ye M, Shou D, Lee W-C, Yin P, Janowicz K (2011b) On the semantic annotation of places in location-based social networks. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining, KDD'11, San Diego. ACM, New York, pp 520–528
Ye M, Yin P, Lee W-C, Lee DL (2011c) Exploiting geographical influence for collaborative point-of-interest recommendation. In: Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval, Beijing, pp 325–334
Ye M, Liu X, Lee W-C (2012) Exploring social influence for recommendation: a generative model approach. In: Proceedings of the 35th international ACM SIGIR conference on research and development in information retrieval, SIGIR'12, Portland. ACM, New York, pp 671–680
Yin Z, Cao L, Han J, Zhai C, Huang TS (2011) Geographical topic discovery and comparison. In: Proceedings of the 20th international conference on World Wide Web, Hyderabad, pp 247–256
Zheng Y (2011) Location based social networks: users. In: Computing with spatial trajectories. Springer, New York, pp 243–276
Zheng Y, Zhang L, Ma Z, Xie X, Ma W-Y (2011) Recommending friends and locations based on individual location history. ACM Trans Web 5(1):5

Location-Based Status Updates and Camera Phone Apps in Social Networks

Larissa Hjorth
School of Media & Communication, RMIT University, Melbourne, VIC, Australia

Synonyms

Geomedia; Geotagging; GPS; Locative media

Glossary

Apps Mobile phone applications

Geotagging Using a location-based service to associate geographic location with other kinds of data such as a photograph or text message

GPS (Global Positioning System) Technologies that calculate the position of a ground-based device through satellite communication

LBS (Location-Based Service) Applications deploy the position of a device and its user, in the provision of a value-added service

LBSN (Location-Based Social Networking Applications) Examples include Facebook Home

SMS (Short Messaging System) The technical term for mobile phone texting that is characterized by a word limit of 160 English characters

Definition

Replacing the navigation of place from paper to digital maps such as Google Maps has transformed how we experience and conceptualize online and offline relationships across corporate, governmental, familial, and personal realms.

Increasingly, location-based services (LBSs) have become an integral part of social network sites (SNS) like Facebook as part of mobile, social, and locative media convergence of the smartphone uptake. With the rise of location-based social networking (LBSN) and camera phone mobile apps like Instagram, geotagging is becoming the default, rather than choice, for many users of mobile applications. Many smartphone applications come with built-in LBS that are active unless the user explicitly turns it off. Social mobile media like Facebook Home are a key example of the ways in which LBS is becoming a norm in the practices of information sharing whereby users can become mere nodes in the corporate commercialization circuits. This shift in LBS and LBSN becoming increasingly mainstream has major implications upon various levels, especially around privacy and surveillance (Perusco and Michael 2007). This entry explores this phenomenon by focusing upon LBSN camera phone practices. By focusing on research coming from media, Internet, and cultural studies, this entry reflects upon the changing nature of people's visual representations and experiences of place in an increasingly location-aware media environment.

Introduction

Replacing the navigation of place from paper to digital, geotagged maps have transformed how we experience and conceptualize online and offline relationships across corporate, governmental, familial, and personal realms. From providing convenience at a fingertip to helping criminal investigations, using locative media has become an essential part of everyday life for individuals, families, communities, businesses, and government agencies. For some, the locative media can help in new forms of surveillance; for others, this phenomenon represents a new epoch in rethinking privacy.

As an important part of mobile social media practices – epitomized by the smartphone phenomenon – camera phone mobile apps, with built-in geotagging features, have become the norm. So how is this overlay between locative media and camera phones creating new types of representations and experiences of place and its relationship to privacy? (Perusco and Michael 2007; Abbas 2011).

Existing research has shown that the growth of locative media is impacting upon cultural practice and policy and on our experiences of place, time, and mobility in ways that are dynamic and uneven across cultural, generational, and temporal realms (Abbas 2011). Given the recent and uneven movement of locative media from new media to everyday practice, there are numerous gaps in the research that require analysis if we are to know contemporary notions of privacy and surveillance in an age of Big Data (Andrejevic 2007). Such a phenomenon requires us to not only reconceptualize how place, privacy, and sociality are being overlaid but also how to study such practices as they increasingly become messy across mobile, social, and locative media spaces.

In this entry we explore the growing phenomenon of LBS, often called "locative media" or "geomedia" (Lapenta 2011a, b). The rise of locative media parallels the move to accessing the online increasingly through mobile media as heralded by the adoption of smartphones unevenly throughout the world. Locative media have significant implications for the ways localized notions and enactments of privacy and surveillance are developing across micro- and macro- and individual and collective levels. Valued at 2.9 billion USD, location-based services are predicted to almost triple to 8.3 billion by 2014 (Fitchard 2012). Although there has been discussion about corporate and governmental surveillance in an age of Big Data (e.g., Farman 2010; Andrejevic 2007; Cincotta and Ashford 2011), the growth in new forms of social surveillance in families is creating an additional – and to date under-researched – layer of everyday practices that are both shaped by, and shaping, locative media (Schofield Clark 2012; Sengupta 2012).

Key Points

The convergence between social, locative, and mobile media impacts how we understand cartographies as not just geographic but determined by social relationships. The use (and non-use) of locative media reflects broader connections between identity and place and raises important issues about privacy (Fusco et al. 2010; Cincotta and Ashford 2011). In order to understand these complex shifts in how place is narrated through social media vis-à-vis locative media, this entry explores the changing role of camera phone practices vis-à-vis locative media. What we see is that once locative media becomes an almost default setting in much sharing of camera phone practice on social media like Facebook, the motivations and types of images shared changes. With locative media like geotagging allowing users to know the physical, geographic setting, camera phone practices become more about sharing more emotional, social, and abstract qualities of a place. Moreover, with the rapid rise of photo apps that are location based, the ability to share almost immediately via multiple social network sites (SNSs) provides users with numerous options that can both blur different communities and distinguish. Camera phones progressively play an important role in our understandings and visualizations of place, especially as they become entangled in locative media practices.

Locative, Social, and Mobile Media at a Crossroad

While Hyunjin waited for Soohyun in a café in Shinchon (Seoul, South Korea), she toyed with her iPhone. Having downloaded some of the numerous photo applications (apps), she began to experiment. Finally, she was happy with the Instagram app. It featured various types of lenses that transformed her ordinary coffee into a polaroid picture from yesteryear. She liked using Instagram because she could take and share pictures and her location almost immediately through various forms of social media such as Twitter and Facebook. She then quickly uploaded the image to Facebook, geotagged with her location, with the caption "Waiting." While the relevance of the caption might be lost on many of her friends, for Soohyun it served as a reminder that she was keeping her friend waiting as she dashed from the train station. A minute later, Soohyun was standing in front of Hyunjin holding her phone with Hyunjin's Facebook page open and her Instagram picture of her coffee. Soohyun then said "Not anymore!" Both girls laughed.

Toshi had never understood why people used the LBS game *Foursquare* until after the earthquake, tsunami, and nuclear disaster of March 2011 in Tokyo (called "3/11"). During 3/11, he tried in vain to make contact with his parents, finally making contact hours later and after much duress. Access to social network sites like Mixi was difficult with mobile phone (*keitai*) networks jamming. Moreover, with loss of electricity, Toshi found himself in an unfamiliar situation – no working *keitai*. So when some sort of normality was established after 3/11, Toshi began to constantly use *Foursquare* not to play the game but as a way to give his friends and family a way to always know where he was.

As a mother of two adult children, Penelope had begun to reacquaint herself with new media. The first thing she did was buy an iPhone. After a decade of using one of the first-generation "classic" Nokias, which was only capable of SMS (Short Messaging System) and voice calls, Penelope's world quickly expanded. Having never used Facebook prior to the iPhone purchase, Penelope was quick to adopt new social and even locative media practices. She loved catching up with old friends overseas that she hadn't seen for decades. It created a new world of possibilities and conversations. But her children weren't so keen. Once upon time, Facebook was their social media world. Now, people over 55 are the fastest growing users of Facebook. In particular, Penelope is part of a new generation of parents that are able to tacitly maintain a friendly eye on their children through the long arm of locative media. By checking in on Facebook, they are able to see where their children are and with whom. So Penelope's children and their friends, like many others, responded to their parent's online presence by changing their privacy settings and defining their online boundaries.

These three vignettes are but a few examples of the many ways LBS and LBSN are being deployed by young and old in their everyday media practice. These stories are part of a broader 3-year ethnographic study into changing mobile media practices and their relationship to intimacy and a sense of place (Hjorth and Arnold 2013). With the rapid rise of smartphones globally, convergence between locative, mobile, and social media is increasingly becoming an integral part of everyday media practice. Across numerous technical platforms, personal and cultural contexts, and through a wide variety of social media, people young and old are using social, locative, and mobile media to rehearse earlier forms of ritual and, at the same time, create new forms of intimacy and different contexts for the expression of intimacy. While locations like Seoul and Tokyo have long been centers of innovation in the invention and popularization of mobile media, the relationship between personal, social, locative, and mobile media is quotidian and, for most part, tacit in its familiarity. In other locations like Australia or the USA, convergence in the form of smartphones is nascent. Mobile social media are a global phenomenon, but they are also local at every point (Hjorth and Arnold 2013).

Devices such as the iPhone have become synonymous with this media evolution. Through this growth, we have witnessed a shift from the device being analyzed as communication medium to being understood as a networked media tool in which social media, games, locative media, and various forms of everyday creativity can be found. In the short few decades in which mobile phones have been readily available, this technology has changed from being a mere extension of the landline to being a sophisticated and convergent online mobile media portal. For many, mobile media are the key device and context for online and social media, with locations like Japan having more than a decade of mainstream mobile Internet. In China, three-quarters of its 485 million online users (318 million) access the Internet via mobile media. In locations like Singapore and Melbourne (Australia), the rapid adoption of smartphones has seen a rise in cross-generational social media usage – much to the disgust of some younger users who fear their parents won't understand much of the tacit etiquette. While mobile media is a global phenomenon, it is also reflective of the local at every level. The growth in mobile media has served to amplify the various complex dimensions of locations, rather than eroding the importance of place.

All over the world, Global Positioning System (GPS) has afforded LBS such as geotagging and Google Maps to become a pervasive part of everyday life through social media accessed via mobile platforms and devices such as smartphones, Android devices, tablets, and portable gaming devices. Moving beyond printed maps, mobile digital devices now frame and mediate our ability to traverse, experience, share, and conceptualize place. This shift appears to have a range of consequences for our relationships to place, intimacy, privacy, time, and presence. Locative media shapes, and is shaped by, a variety of factors such as culture, age, and temporal differences. The more locative media becomes a key part of media practice, the more the way in which we conceptualize and understand the complexity of place shifts (Wilken and Goggin 2012). Locative media not only transforms how we experience mobility (technological, geographic, and psychological) but also how this shapes the fabric of place as what Doreen Massey calls "stories so far" (Massey 2005).

Visualizing Place: A Case Study of Location-Based Services and Camera Phone Practices

While camera phone genres such as self-portraiture have blossomed on a global scale, vernacular visualities that reflect a localized notion of place, sociality, and identity making practices (Hjorth 2007; Lee 2005, 2009) are also flourishing. Smartphone apps like Hipstamatic and Instagram have made taking and sharing photographs easier and more interesting (Palmer 2012; Chesher 2012). With LBSN like Facebook and, LBS games like *Foursquare* and *Jiepang*, we see a further overlaying of place with the social and personal, whereby the electronic is superimposed onto the geographic in new ways. Specifically, by sharing an image and comment about a place through LBSs, users can create different ways to experience and record journeys and, in turn, create an impact upon how place is recorded, experienced, and thus remembered. This is especially the case with the overlaying of ambient images within moving narratives of place as afforded by LBSs. An example might be someone uploading a geotagged camera phone image onto Facebook Home whereby the information of place is recorded and shared in a variety of ways. This practice, in turn, impacts upon their experience of place as something that is mediated through networked media. While place and intimacy has always been mediated by language, memories, and gestures, it is the way in which it is being mediated that is transforming how we think about and practice place.

The rapid uptake of smartphones has enabled new forms of distribution and has provided an overabundance of apps, filters, and lenses to help users create "unique" and artistic camera phone images. Although the iPhone has been quick to capitalize on this phenomenon through applications such as Hipstamatic, other operating systems like Android have also had their share of this expanding market. So too social media such as microblogs and LBSs have acknowledged the growing power of camera phone photography, not only by affording easy uploading and sharing of the vernacular (Abbas 2011) but also by providing filters and lenses in order to further enhance the "professional" and "artistic" dimensions of the photographic experience (Mørk Petersen 2009).

Consider, for example, how social media applications for smartphones no longer ask you to go through multiple steps to attach images to a post. Formerly, if you wanted to post an image, you'd take the image using the phone's camera application that would store the image in the camera's library, which you would then access by attaching the image to a post. Sometimes you would even need to upload the image to an online image repository so it could be linked to the SNS. Now, many social media apps provide a photo button integrated into the app that allows you to take a picture and post it immediately, and social media companies provide their own image hosting servers that operate almost invisibly to the user. One way to understand the impact of LBS on camera phone practices is through a shift from *networked* to *emplaced* visuality. But firstly, let us explore the locative, social, and mobile media intersection and its impact upon camera phone visualities.

At a Crossroads: Smartphones, LBS, and Camera Phone Visualities

For Chesher (2012), the rise of smartphones like the Samsung Galaxy and the iPhone – with their attendant software applications like Instagram, Google Goggles, and Hipstamatic – has created new ways in which to think about camera phone practices and their engagement with both image and information. For Chesher, the iPhone universe of reference disrupts the genealogy of mass amateur photography. Chesher argues that up until camera phones, the Kodak moment dominated. This was then replaced by the Nokia moment and then further colonized by the iPhone through the plethora of camera phone apps available. Applications like Instagram which allow users to take, edit, and share photos partake in what could be called a second generation of camera phone and photo-sharing social media. With sites that allow the display of local Do-It-Yourself (DIY) creativity such as Flickr being the precursor, Instagram heralds a new generation of visuality in which the cult of the amateur is further commercialized. Launched in October 2010, Instagram quickly grew to boasting over 150 million uploaded images in less than 6 months. With these new applications, often working in collaboration with social and locative media, camera phone images have been given new contexts.

For Daniel Palmer (2012), iPhone photography is distinctive in three ways. First, it creates an experience between touch and the image in what Palmer calls an "embodied visual intimacy." While "touch has long been an important, but neglected, dimension in the history of photography . . . the iPhone, held in the palm of the hand, reintroduces a visual intimacy to screen culture that is missing from the larger monitor screen" (Palmer 2012). Second, the proliferation of photo apps for the iPhone has meant that there are countless ways for taking, editing, and sharing photos. No longer do camera phone images have to look like the poorer cousin to the professional camera. Third, and most important to our discussion here, is the role of GPS capability with the iPhone automatically "tagging photographs with their location, allowing images to be browsed and arranged geographically" (Palmer 2012).

As Palmer identifies, the placing of photos through tagging creates a different way of archiving and contextualizing the image. With the growth in LBS games like *Foursquare Jiepang* (China) and *flags* (South Korea) that allow users to "check in" and upload pictures, networked visuality and the attendant empowerment/control binary are further complicated. For some, LBSs provide new forms of überveillance (Cincotta and Ashford 2011), privacy invasion (Abbas 2011), and stalking (Gazzard 2011). For others, they highlight how the local continues to play a pivotal role in informing notions like privacy and sociality. LBSs allow already "friendly" surveillance, such as that present in parent/child relationships, to flourish in new ways (Sengupta 2012; Schofield Clark 2012). In these new geospatial visualities, the motivations and types of genres are changing to reflect the ways in which locative media impact upon how localities are shared and experienced.

LBS Camera Phone Images: Networked Versus Emplaced

When visuality becomes part of a networked culture, its meanings, contexts, and content change. Camera phones, as an extension of the networked nature of mobile media, are clearly defined by this dynamic. Although initial studies into camera phone visuality discussed it as part of networked media (Rubinstein and Sluis 2008; Ito and Okabe 2003; Ito and Okabe 2005, 2006; Villi 2013), this second generation of visuality – one that is characterized by locative media – is about new types of place-making exercises (Pink and Hjorth 2012). These exercises are emotional and electronic, geographic, and social – highlighting the complexity of ever-evolving notions of place. In each location, camera phone images are overlaid onto specific places in a way that reflects existing social and cultural intimate relations as well as being demonstrative of new types of what Pink calls "emplaced" visuality, in which locative media emplace images within the entanglement of movement (Perusco and Michael 2007).

First-generation "networked" visuality, when combined with LBSs in the obvious case of the smartphone, becomes "emplaced" visuality, that is, visuality overlaid by a moving, geospatial sociality. Incorporating movement in the theorization of visuality is important given the ways in which camera phone practices give way to an accelerated taking, editing, and sharing of a "moment" that is then contextualized through its place in the moving geographic and social maps of LBSs and social media. Whereas first-generation camera phone sharing was defined by the network (Ito and Okabe 2006; Villi 2013; Abbas 2011), second generation, characterized by LBSs like geotagging, becomes focused upon emplacement through movement (Perusco and Michael 2007). In other words, LBS camera phone culture is about reinforcing the process of the node rather than the product of the network. Images increasingly became about creating a sense of movement through an ambience of place. These images are "multisensorial" (Perusco and Michael 2007) in that they evoke more than the visual; they overlay information (such as location) with emotion.

For Pink, the combination of locative media with the photographic image requires a new paradigm that engages with the multisensoriality of images. It might at first seem odd to talk about images as being multisensorial because surely images are visual and so draw upon only one sense: vision. Pink draws on Tim Ingold's (2008) critique of the anthropology of the senses and of network theory and argues that by exploring the visual in terms of multisensoriality, one can reprioritize the importance of movement and place. For Pink, locative media provide new ways in which to frame images with the "continuities of everyday movement, perceiving and meaning making" (Perusco and Michael 2007).

By contrasting "photographs as mapped points in a network" with "photographs being outcomes of and inspirations within continuous lines that interweave their way through an environment – that is, in movement and as part of a configuration of place" (Perusco and Michael 2007), Pink argues that we must start to conceive of images as produced and consumed in movement. Here, we can think about how images are being transformed in the light of various turns – emotional, mobility, and sensory. Indeed, of all the areas to be impacted and affected, camera phones – especially with their haptic (touch) screen interface and engagement, along with their locative media possibilities – can be seen as indicative of Pink's (Perusco and Michael 2007) call for a multisensorial conceptualization of images. As Pink notes, the particular way in which text, image, and GPS are overlaid creates a multisensorial depiction of a location.

This shift can be viewed as the movement from a camera phone visuality that is networked to camera phone images which are "emplaced" (Perusco and Michael 2007). An image that is socially networked, tagged, and GPS located is "emplaced" in a number of ways. First, it is emplaced as one of the many images captured by a particular member of our intimate or social public and is contextualized by our relation to that person. Second, the copresence of many images arranged by time or place on one site places each image in the context of others to constitute a narrative and thus another context. Third, the GPS coordinates place the image in geographic space and invite the viewer to *recall* the place in question as well as *view* the image captured in the place in question, thus overlaying another context. Finally, the social distribution of the images creates a social public for those images, thus overlaying another context, and the image tags entered by the public overlay yet another context. As Mikki Villi notes in "Visual Mobile Communication on the Internet: Publishing and Messaging Camera Phone Photographs Patterns," "much of the traffic in photographs now circulates through digital networks and is facilitated by new platforms" (Villi 2013). As Lapenta notes, "what is really changing has little to do with the increasing numbers of images taken every day and more to do with the increasingly differentiated forms of photographic image production, aggregation and distribution" (Lapenta 2011b).

As Pink and Hjorth argue in "Emplaced Cartographies: Reconceptualising camera phone practices in an age of locative media" (Pink and Hjorth 2012), the rise of high-quality camera phones, along with the growth in distribution services via social and locative media, is heralding new forms of visuality. These new types of "copresent" visuality overlay and interweave online and offline cartographies in different ways – maps that require a revision of ethnography. They argue this shift can be framed in terms of a movement from networked visuality to emplaced visuality and sociality. These emergent ways in which online and offline cartographies are becoming overlaid and entangled – as well as the experiential environments associated with this – demand alternative ways for theorizing visuality and socialities of copresence. This requires that we go beyond earlier studies of camera phones (e.g., Ito and Okabe 2003; Ito and Okabe 2005, 2006) which focused on the three "s" – sharing, storing, and saving – in informing the *context* of what was predominantly "banal" everyday *content* (Koskinen 2007). Moreover, these earlier studies forwarded the idea of "ambient copresence" (Ito and Okabe 2006) in which burgeoning community platforms such as Flickr and social media like Facebook, camera phone practices embodied not only "networked visuality" but also emergent forms of user creativity (Mørk Petersen 2009). Departing from these ideas, we develop the notion of emplaced visuality, that is, a particular overlaying of various narratives that locate emotional, social, and mnemonic dimensions of place and presence beyond just the geographic or networked.

For Rubinstein and Sluis, the increasing "reliance on tagging for organization and retrieval of images is an indication of the importance of textuality for online photographic procedures… Tagging provides a substantially different way of viewing and interacting with personal photography" (Rubinstein and Sluis 2008). According to Rubinstein and Sluis, these networked images get transformed as part of metadata in which the original context is lost.

Rubinstein and Sluis's study of the networked image operates as a precursor to the second generation of camera phone practices in which GPS creates new levels for recording the "original" context with a geographic coordinate. This capacity to geo-locate or tag camera phone images invites us to conceptualize the production and consumption of images through movement and place. Therefore, rather than viewing networked images as part of acontextualized metadata as suggested by Rubinstein and Sluis, Pink and Hjorth argue that locative media provides new ways in which to frame images with the "continuities of everyday movement, perceiving and meaning making" (Pink and Hjorth 2012). By contrasting "photographs as mapped points in a network" with "photographs being outcomes of and inspirations within continuous lines that interweave their way through an environment – that is, in movement and as part of a configuration of place" (Perusco and Michael 2007), locative media camera phone practice can be conceived of images as produced and consumed in movement *Emplaced.*

In "The Digital Wayfarer" (Pink and Hjorth 2013), Pink and Hjorth further extend upon this discussion by draw on phenomenological anthropology to advance a framework for understanding contemporary and emergent experiences and practices of camera phone photography. They propose understanding camera phone photography as part of a continually shifting and emerging ecology of place where humans, the material, and digital are increasingly entangled. This, they argue, signifies a need to depart from the dominant "network" paradigms in visual/media culture and Internet studies towards a focus on "emplacement" whereby people, images, and technologies are always situated, in movement, and part of and constitutive of place. They argue that the traces made as camera phones are used, as mobile media weaving through material/digital environments with their users, thus become forms of visuality that are emplaced digitally and materially.

Pink and Hjorth note that new camera phone photography, overlaid by locative media, can be understood in terms of Ingold's wayfarer in two ways: to consider how wayfaring can encompass the online/offline places of everyday life and to consider the making and posting of images as a characteristic of the particular type of digital wayfaring we are interested in. This, they argue that "camera phone wayfaring" – as a byproduct of locative, social, and mobile media overlays – helps to understand how the routes and trajectories of life as lived become parts of configurations of place through which camera phone photos are made, posted, and viewed. But on the other side of this creative play with copresent visuality is the impact LBS has on how we understand information sharing and its impact upon notions of privacy and surveillance (Abbas 2011).

Conclusion

In this entry, we have explored the burgeoning role of LBS with social and mobile media convergence. Through the overlaying of social and LBS media, we are seeing new ways in which places can be mapped and visualized. By exploring the example of LBS camera phone practices, we can see how emerging visualities are shaping, and being shaped by, different notions of place.

The growth of locative media is having significant impact upon cultural practice, place making, and relationships in ways that are shifting, ongoing, and emergent (Abbas 2011; Cincotta and Ashford 2011; de Souza e Silva and Frith 2012; de Souza e Silva and Hjorth 2009; de Souza e Silva and Sutko 2010; Farman 2010; Fusco et al. 2010; Gordon and de Souza e Silva 2011). The first generation of locative media games provided insight into ways in which to rethink privacy, play, and place. However, second-generation LBS, harnessed by the smartphone revolution, has brought with it a flood of experiences that we are still trying to assess. This phenomenon has both positive and negative impacts upon localized notions of privacy and surveillance across both micro- and macro- and individual and collective levels (Fusco et al. 2010; Gazzard 2011). While much analysis has been conducted into experimental forms of locative media/art (de Souza e Silva and Frith 2012; de Souza e Silva and Hjorth 2009; de Souza e Silva and Sutko 2010; Lee 2009), the increased ubiquity of locative media through devices such as the smartphone will undoubtedly transform the way in which place, privacy, and mobility are articulated (Abbas 2011; Cincotta and Ashford 2011; Hjorth and Gu 2012; Lapenta 2011b; Perusco and Michael 2007).

Although second-generation LBS games like *Foursquare* and *Jiepang* are in their infancy, they represent an area of growing diversity and complexity within mobile media and communication. In particular, LBSs are changing how we visualize intimate cartographies through shifting camera phone practices. Whereas first generation of camera phone practices noted gendered differences (Hjorth 2007; Lee 2005), through LBSs, these gendered differences in visualities take on new dimensions – particularly in terms of its potential "stalker" elements (Cincotta and Ashford 2011; Gazzard 2011).

As LBS becomes progressively more integrated into social mobile media experiences, gendered, generational, and other differences will become magnified in the types of pictures not only taken and shared but also not shared. For example, in preliminary studies conducted with younger women, the interest in self-portraiture has diminished with the rise of locative media (Hjorth 2013). So too, young people "coming out" (as homosexuals) may have concerns about their privacy especially if living in conservative locations or cultures. Increasingly, we will see how different camera phone practices, and the choice to share to and with whom, will reflect a users relationship with LBS and privacy. Moreover, studies across the generations and cultural contexts need to be conducted if we are to fully realize the complex way in which LBSs are reflecting differing norms around privacy, community, and place (Hjorth 2013).

The mainstreaming of LBSs through smartphones is demonstrating the diverse ways in which privacy is understood across cultural, social, temporal, and generational contexts. Future studies in locative media will need to reflect upon these issues.

Cross-References

▶ Collective Intelligence for Crowdsourcing and Community Q&A
▶ Facebook's Challenge to the Collection Limitation Principle
▶ Flickr and Twitter Data Analysis
▶ Game Theory and Social Networks
▶ Location-Based Social Networks
▶ Mobile- and Context-Aware Applications of Social Networks
▶ Mobile Communication Networks
▶ Privacy Preservation and Location-based Online Social Networking
▶ Social Networking Sites
▶ Spatiotemporal Personalized Recommendation of Social Media Content

References

Abbas R (2011) The social implications of location-based services: an observational study of users. J Locat Based Serv 5:3–4, 156–181

Andrejevic M (2007) iSpy: surveillance and power in the interactive era. The University Press of Kansas, Lawrence, KS

Chesher C (2012) Between image and information: the iPhone camera in the history of photography. In: Hjorth L, Burgess J, Richardson I (eds) Studying mobile media: cultural technologies, mobile communication, and the iPhone. Routledge, London/New York, pp 98–117

Cincotta K, Ashford K (2011) The new privacy predators. Women's Health. http://www.purehacking.com/sites/default/files/uploads/2011_11_00_Australian_Womens_Health_November.pdf. Consulted 10 Jan 2012

de Souza e Silva A, Hjorth L (2009) Playful urban spaces: a historical approach to mobile games. Simul Gaming 40(5):602–625 de Souza e Silva A, Sutko D (eds) (2010) Digital cityscapes. Peter Lang, Berlin

de Souza e Silva A, Frith J (2012) Mobile interfaces in public spaces: locational privacy, control, and urban sociability. Routledge, New York

Farman J (2010) Mapping the digital empire: Google Earth and the process of postmodern cartography. New Media Soc 12(6):69–888

Fitchard K (2012) By 2016, 70 M families will keep tabs on each other via GPS. In: GIGAOM. http://tinyurl.com/9ctsjyb. Accessed 10 Feb 2013

Fusco SJ, Michael K, Michael MG (2010) Exploring the social implications of location based social networking: an inquiry into the perceived positive and negative impacts of using LBSN between friends. In: Giaglis G (ed) 9th IEEE international conference on mobile business. IEEE, Athens. Available at http://works.bepress.com/kmichael/190. Accessed 10 Feb 2013

Gazzard A (2011) Location, location, location: collecting space and place in mobile media. Converg: Int J Res New Media Technol 17(4):405–417

Gordon E, de Souza e Silva A (2011) Net locality. Wiley, New York

Hjorth L (2007) Snapshots of almost contact. Continuum 21(2):227–238

Hjorth L (2013) The place of the emplaced mobile: a case study into gendered locative media practices. Mobile Media Commun J 1(1):110–115

Hjorth L, Arnold M (2013) Online@AsiaPacific: mobile, social and locative media in the Asia–Pacific. Routledge, New York

Hjorth L, Gu K (2012) The place of emplaced visualities: a case study of smartphone visuality and location-based social media in Shanghai, China. Contin: J Media Cult Stud 26(5):699–713

Ingold T (2008) Anthropology is not ethnography. Proc Br Acad 154:69–92

Ito M, Okabe D (2003) Camera phones changing the definition of picture-worthy. Jpn Media Rev. Retrieved from http://www.ojr.org/japan/wireless/1062208524.php. Accessed 10 Feb 2013

Ito M, Okabe D (2005) Intimate visual co-presence. Paper presented at Ubicomp, Takanawa Prince Hotel, Tokyo, Japan, 11–14 Sept 2005. Retrieved from http://www.itofisher.com/mito/. Accessed 10 Feb 2013

Ito M, Okabe D (2006) Everyday contexts of camera phone use: steps towards technosocial ethnographic frameworks. In: Höflich J, Hartmann M (eds) Mobile communication in everyday life: an ethnographic view. Frank & Timme, Berlin, pp 79–102

Koskinen I (2007) Managing banality in mobile multimedia. In: Pertierra R (ed) The social construction and usage of communication technologies: European and Asian experiences. Singapore University Press, Singapore, pp 48–60

Lapenta F (2011a) Geomedia: on location-based media, the changing status of collective image production and the emergence of social navigation systems. Vis Stud 26(1):14–24

Lapenta F (2011b) Locative media and the digital visualisation of space, place and information. Vis Stud 26(1):1–2

Lee D-H (2005) Women's creation of camera phone culture. Fibreculture J 6. http://www.fibreculture.org/journal/issue6/issue6_donghoo_print.html. Consulted 3 Feb 2006

Lee D-H (2009) Re-imaging urban space: mobility, connectivity, and a sense of place. In: Goggin G, Hjorth L (eds) Mobile technologies. Routledge, London/New York, pp 235–251

Massey D (2005) For space. Sage, London

Mørk Petersen S (2009) Common banality: the affective character of photo sharing, everyday life and produsage cultures. PhD Thesis, ITU Copenhagen

Palmer D (2012) iPhone photography: mediating visions of social space. In: Hjorth L, Burgess J, Richardson I (eds) Studying mobile media: cultural technologies, mobile communication, and the iPhone. Routledge, New York, pp 85–97 (88)

Perusco L, Michael K (2007) Control, trust, privacy, and security: evaluating location-based services. IEEE Technol Soc Mag 26(1):4–16. Copyright 2007 IEEE. http://ro.uow.edu.au/cgi/viewcontent.cgi?article=1521&context=infopapers. Accessed 9 Apr 2013

Pink S, Hjorth L (2012) Emplaced cartographies: reconceptualising camera phone practices in an age of locative media. Media Int Aust 145:145–156

Pink S, Hjorth L (2013) The digital wayfarer. In: Goggin G, Hjorth L (eds) The Routledge mobile media companion. Routledge, New York

Rubinstein D, Sluis K (2008) A life more photographic: mapping the networked image. Photographies 1(1):9–28

Schofield Clark L (2012) The parent app: understanding families in the digital age. Oxford University Press, New York

Sengupta S (2012) Big Brother? No, It's Parents. New York Times, June 25. http://tinyurl.com/aysm5ud. Accessed 30 March 2013

Villi M (2013) Visual mobile communication on the internet: publishing and messaging camera phone photographs patterns. In: Cumiskey K, Hjorth L (eds) Seamlessly mobile. Routledge, New York (forthcoming)

Wilken R, Goggin G (2012) Mobile technology and place. Routledge, London

Recommended Reading

de Souza e Silva A, Frith J (2012) Mobile interfaces in public spaces: locational privacy, control, and urban sociability. Routledge, New York

Farman J (2010) Mapping the digital empire: Google Earth and the process of postmodern cartography. New Media Soc 12(6):869–888

Gordon E, de Souza e Silva A (2011) Net locality. Wiley
Hinton S, Hjorth L (2013) Understanding social media. Sage, London
Hjorth L, Arnold M (2013) Online@AsiaPacific: mobile, social and locative media in the Asia–Pacific. Routledge, New York

Locative Media

► Location-Based Status Updates and Camera Phone Apps in Social Networks

Logic Programming

► RIF: The Rule Interchange Format

Logit Models

► Exponential Random Graph Models

Longitudinal Network Analysis

► Analysis and Visualization of Dynamic Networks