

Getting to “Know” People on the Web 2.0

Yiwei Cao, Anna Glukhova, Ralf Klamma, Zinayida Petrushyna

(RWTH Aachen University, Ahornstr. 55, 52056 Aachen, Germany)

Databases & Information Systems

{cao|glukhova|klamma|petrushyna}@dbis.rwth-aachen.de)

Abstract: Web 2.0 platforms such as media sharing and social network sites (SNS) concern people in everyday life to a great extent. People are enabled to reach out to various media and up to now, it is nearly impossible to use digital identities ex ante or to recreate users' identities ex post across different platforms. In this paper, we explore important methodologies in Web 2.0 such as cross-media analysis and social pattern based analysis based on a survey in this area, aiming at cross-platform information diffusion across social network sites. Open issues are discussed to explore the challenges and solutions in this new research area.

Keywords: social network analysis, Web 2.0, cross-platform, cross-media, Web services, identity management, XML, Semantic Web, Multimedia

Categories: H.5.1, H.3.5, H.3.3.

1 Introduction

Nowadays, media information sharing takes place rather on virtual social networks than in real life. Moreover, knowing people on the Web 2.0 is different from our experiences in real life. There are mere technical issues like different email addresses, different computers in the network, different account information in social platforms, different authorization methods like password based accounts or public key infrastructures and so on. In addition, social and psychological issues like preferred anonymity result in creation of many faked online identities or faked avatars etc. Furthermore, there are also commercial, security, and privacy issues like personal account data locked in vendor databases as a company asset, identity data kept on secure servers not shared with other sites and data protection laws preventing companies from sharing account information, even if users explicitly allow it.

Nevertheless, online reputation of people depends on their social capital, i.e. the closest neighbours, including knowledge about their favourite topics, their roles in the network, their behaviour, their mutual trust relation etc. Solutions to the requirements may not be novel. But it is challenging to deal with the technical, the socio-psychological, the legal, and the commercial issues at the same time. Here we concentrate on some aspects of things already available to sift the field.

The mechanics of social networks consider their members as actors of a community. Information sharing depends on the media we use. If we want to share private information, we normally do not use a public blog but an email. If we want to spread a message immediately, we post it in a frequently-read forum or tell it to a blabbermouth. In short, information diffusion aims to find media and people relevant to a particular piece of information. People's relevance is assessed by their knowledge

about the topic, by their position in the network, by their behaviour, by their mutual trust relations, etc. Hence, we propose the cross-media and social pattern-based analysis with social network analysis [Degeenne and Forse 03] approach to realizing cross-platform information diffusion and to manage identities across social network sites in Web 2.0.

The rest of the paper is structured as follows. In Section 2 we give an overview of concrete solutions in cross-platform social networking. We introduce new methodologies and approaches related to research cross-platform social networking aspects in Section 3. The paper concludes with a short discussion.

2 What is already there in Cross-platform Social Networking?

Social network sites (SNS) emerge in recent years to enable users to create, search, share, and publish all kinds of information including not only diverse media content but also people's opinions and activities. Web 2.0 has become a kind of new culture with great social impacts. Virtual openness and reach out to communities feature Web 2.0 users. People can get to know people who they do not know in real life by reaching out to sharing information. Everybody is enabled online to share their ideas, videos, pictures, activities online. We compare people "know" people online and offline and find the comparison useful to understand identity issues well (cf. Table 1).

Activities	How to know people in reality?	How to know people the on Web 2.0?
Identity	name + social security id	User name + OpenID
Authentication	name + face	email address (user name) + password
Communication	handshaking, conversation, glance, seeing a film, meeting in a bar, visiting ...	Commenting, emailing, chatting, virtual discussions, emoticon
Knowing people in the networks	usually yes	yes or no (virtual friends)

Table 1: Differences on identifying people in real life and online

The prevalent SNS are shortly discussed as follows.

Facebook (<http://www.facebook.com>) as one of the most popular social network service has opened APIs also for third parties. Social network data is aggregated by news feeds to show recent activities within users' networks, such as who posts comments in whose profile, who tags whom in photos or who exchanges applications with whom.

Flickr (<http://www.flickr.com>) is a photo sharing platform that includes an explicit social network. This is exposed by the Flickr API exploring contact lists data and tagging activities.

Del.icio.us (<http://del.icio.us>) is a social bookmarking system which allows users saving, tagging, and describing links. In addition, an explicit social network can be established by adding interesting people into the list to see their bookmarks.

YouTube (<http://www.youtube.com>) is a video sharing web application allowing users collaborating around video files applying tagging and ranking. Users can manage a contact list to create an explicit social network.

The other prevalent SNS are presented as a summary in Table 2. The summary is done according to our observations of the presented data in SNS. Statistical analysis refers to media popularity, freshness. Semantic analysis is an extraction of data, community actors or media according to the interests, tags, and location etc. Structural analysis is performed on social patterns.

SNS	Shared information							Communication					Analysis				
	Photos	Videos	Music	Bookmarks	Stories	Contacts	Cross-media Activities	Activities	Ranking	Comments	Tagging	Forums	Friend's list	Emails	Statistical	Semantic	Structural
Del.icio.us				●		○	○	○			●		○		●		
YouTube		●				○	○	○	●	●	○	●	○		●	○	
Blogger	○	○	○		●	○	○	○		●	●						
Last.fm		○	●			○	○	○	○	○	●	○	○		●	○	
Facebook	●		○		●	○	○	●	○	●	●	●	●	●	●	○	●
LinkedIn						○	○	○	○				○	○			○
Twitter						○	○	●									
Mixx	○	○	○		○	○	○	○	○	●	●		○		●	○	●
Flickr	●	●				○	○	○		●	●		●	●	○		
FriendFeed	○	○	○	○	○	○	○	●		●	●		○				
○ - Though the feature is supported, it is not the focus of SNS; the support is poorly implemented, compared to the others																	
● - the feature is supported and SNS make an accent on it as one of the main features																	

Table 2: A summary to the comparison of Social Network Sites (SNS)

Content and systems interoperability is a main issue of information system nowadays. Web 2.0 technologies like RSS facilitate information sharing across different Web 2.0 platforms. In addition, cross-platforms have been realized by synchronization technologies between MSN Messenger and Windows Media Player, between Outlook/iCal with Google Calendar, and among Last.fm and Windows Media Player and iTunes. For example, socialization is performed by informing people what music their buddies are listening to. From viewpoint of social network platforms a single sign-in over different platforms is crucial for interoperability concept.

The cross-platform aspect is novel and emergent for social network information aggregation and diffusion. This aspect is well identified and covered in [Guy et al. 08]. SNS have a wide coverage of applications with diverse content created by users ranging from simple user contacts to multimedia. When a user is a novice to a social network site, he needs to register, login, create his/her user profile, and get his/her social network via email address books. Through recent booming acquisition of start-

up social network sites by the market leading companies, users are supported to single-login to some extent. However, the interoperability of social network information across SNS is still limited.

The Friendfeed Approach to Social Aggregation

Above all, *Friendfeed* (www.friendfeed.com) allows users aggregating 41 different SNS services including Flickr, Del.icio.us and YouTube. What happens within supported SNS can be well traced, so that users are able to monitor their buddies' activities correspondingly. Besides, Socialthing!, Spokeo, SecondBrain and Iminta are all the similar platforms like Friendfeed to serve as social aggregators for cross-platform information diffusion.

In conclusion, the Friendfeed approach makes a first successful step towards cross-platform information diffusion on Web 2.0. However, the information is passed within the friend network limited in SNS. Users could have partly different networks on different SNS, which are not taken into account totally. No SNS like MySpace or Facebook etc. which consists of many different and non-interoperable for each other services such as photo sharing, blogging, instant messaging and so on) can be aggregated. Only "homogenous" SNS services, that are built according to the same model and use standard services, can.

3 Conceptual Cross-Platform Information Diffusion

According to the data presented in the last section, we propose four main aspects of cross platform information diffusion: social network analysis, cross-media, social pattern, and identity management. The existing approach is social aggregation of diverse SNS media including services and content created by user communities. We observe media and communities as a whole. Not only media is dealt with in a cross-platform manner, but also social networks need to be analysed across platforms. Hence, social pattern analysis as well as social network analysis needs to be employed to draw a social graph [Hinchcliffe 08b] on it (cf. Figure 1). Identity management is the key to fulfilling cross-platform information diffusion on cross-platform networks.

Social Network Analysis Methodologies

Social networks indicate a special kind of networks where social relationships are viewed as edges in graphs with values in a social context. The understanding of relationships between members, dependencies between networks and technologies, as well as semantic examination of networks and members identities study is complex tasks of social networks exploration. Within social networks, actors and their actions are viewed as interdependent rather than independent and autonomous units [Wasserman and Faust 94; Vega-Redondo 07]. Recently, there are much research and many available systems for finding relevant people [Lin et al. 08; Guy et al. 08]. In order to distinguish nodes within a network, SNA focuses on the study of in-degrees and out-degrees of a node, i.e. centrality measures. Nodes with the high degree centrality are the most visible in the network, adjacent to the members, and contribute to the identity of the network [Wasserman and Faust 94]. The betweenness centrality identifies how many paths are going through members, how many times they are

bridges of information. The Edge Betweenness Centrality (EBC) [Freeman et al. 91] is useful to identify network subgroups and to recognize social patterns.

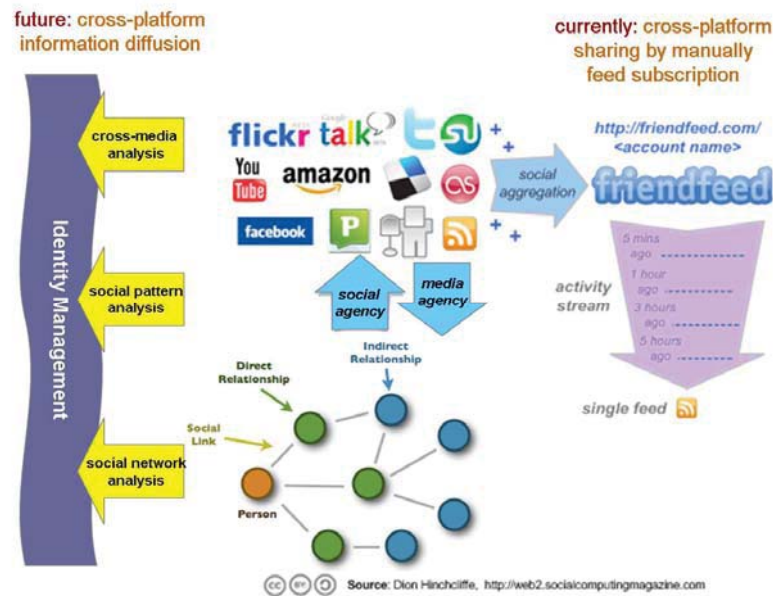


Figure 1: Cross platform for social network sites (adapted from [Hinchcliffe 08a; Hinchcliffe 08b])

Cross Media Analysis

We deal with social network data from different sources which are of a private (e.g. private SMS, private chat), a semi-public (e.g. mailing lists, group chats, forums, protected wikis) or a public nature (e.g. blogs, public wikis, public image or video sharing) [Klamma et al. 07]. Digital media make information easy to be copied and aggregated with little additional workload [Hogan 07]. The requirements on the combination of data from the media create the need of cross-media analysis application in the examinations of social networks.

Applying this we need network models that focus on individuals and consider the network structural environment as opportunities for or constraints on individual actions. Network models conceptualize social, economic, and political structures as lasting patterns of relations among actors. Among others, Actor Network Theory (ANT) [Latour 91] is appropriate for observing networks formed by humans and media, as it does not distinguish between human and non-human actor. ANT model reveals relations between technologies for social networks and identifies one prototype for all technologies and members of social networks in order to ease the comparison of influences of technologies and to alleviate the merge of social networks [Klamma et al. 06]. Furthermore, semantic analysis can be additionally applied besides cross-media approaches, with which social network data can be traced

back to the term of “semantic nets” as well as semantic networks based on graph theories [Shastri 91].

Social Pattern Based Analysis

Social patterns include similar behaviour of members over their social networks. The following social patterns are useful for cross-platform information diffusion. The members are *brokers*, when they transmit knowledge between sub-groups in networks [Kilduff and Tsai 06]. The brokers are usually those, who span *structural holes* that are the relationships between two non-redundant neighbours of the network [Burt 92]. The size of a member social capital, i.e. the closest neighbours of the member in a network, increases, if the number of structural holes he spans grows [Granovetter 73]. A broker is a bottleneck in the network and enjoys his/her power. He/she controls information flows between groups. Brokers are useful to spread information to non-relevant people. The members are *hubs*, when these have a relatively high number of relations with *authorities*. The members are the *authorities*, if many of the others have relations with them [Kleinberg 99]. Usually, there are many *hubs* and few *authorities* in the network. Such distribution over nodes indicates that the network is scale-free and obeys the power law [Barabási 03].

Identity Management

Microsoft Live ID and Google accounts first brought identity into sight. Concept of identities is of great importance in digital social networks [Neuenschwander et al. 05] and one of examples of its implementation is the standard of OpenID. The standard has been developed from a lightweight HTTP based URL authentication framework to an open community-driven platform [Recordon and Reed 06]. Started with 9 million users on LiveJournal.com for bloggers, more and more social network sites support OpenID. The most significant benefit is to enable single-login, so that users do not need to register with another set of username and password to a new service. Users' wish whether to use OpenID is specially mentioned in [Powell and Recordon 07], which leads to trust issues potentially. Users might question how safe it would be to sign in each social network sites to access personal mail box with the same user name and password. So far, OpenID is an efficient method to identify users across platforms. What will happen, if one's OpenID is once stolen? A personal identity repository can be established, if email addresses traced. However, these emails are not allowed to be used for further purpose. In short, security, privacy and trust are particularly critical.

4 Discussion and Conclusion

In this paper we collected and analyzed some of the literature related to new approaches of analyzing and introducing social data across different platforms. We have argued that identification of right people and right media is essential to many new tasks in the social web like the effective and efficient diffusion of information. Because of our view that analysis is as important as operational support for cross-platform systems, we argued to identify people by a combination of pattern-based social network analysis giving us more insights in the social agency of people and cross-media analysis giving us the opportunity to gain knowledge out of the

embedding of agency in media. The other components that have to be taken into consideration in the future for the paper topic are the media types we are using for any task in SNS. These affect our behaviour and our position in a society. Technologies and their influences have to be included in the analysis of social network data as their competent members. Furthermore, the examination of Part-Of-Speech taggers in social network content and its emotions resolves ambiguity about the sense of content [Lakoff and Johnson 80; Pennebaker et al. 07]. Through all presented in the paper approaches, it is possible to reflect the social network data changes dynamically and to keep social network data processing up-to-date. Based on the research presented here, it is possible to develop new mechanisms for social network data collections and for developing new services for cross-platform information diffusion services.

References

- [Barabási 03] Barabási, A.L.: “Linked: The New Science of Networks”, *J. Artificial Societies and Social Simulation*, 6(2), 2003.
- [Burt 92] Burt, R.: “Structural Holes: the Social Structure of Competition”, Harvard University Press, Massachusetts, 1992.
- [Degenne and Forse 03] Degenne, A., Forse, M.: “Introducing Social Networks, Sage Publications, 2003.
- [Freeman et al. 91] Freeman, L.C., Borgatti, S.P., White, D.R.: “Centrality in valued graphs: A measure of betweenness based on network flow”, *Social Network*, 13:141–154, 1991.
- [Guy et al. 08] Guy, I., Jacovi, M., Shahar, E., Meshulam, N., Soroka V., Farrell, S.: “Harvesting with SONAR- The Value of Aggregating Social Network Information”, *Proc. ACM Conference on Human Factors and Computer Resources, CHI’08*, Florence, Italy, 2008.
- [Granovetter 73] Granovetter, M.S.: “The Strength of the Weak Ties”, *American Journal of Sociology*, 78:1360–1380, 1973.
- [Hinchcliffe 08a] Hinchcliffe, D.: “Social Aggregators Emerge to Manage Digital Lifestyles”, *Dion Hinchcliffe’s Web 2.0 Blog*, March 19, 2008, <http://web2.socialcomputingmagazine.com/{12.04.2008}>.
- [Hinchcliffe 08b] Hinchcliffe, D.: “The Social Graphs: Issues and Strategies in 2008”, *Dion Hinchcliffe’s Web 2.0 Blog*, January 16, 2008, <http://web2.socialcomputingmagazine.com/{12.04.2008}>.
- [Hogan 07] Hogan, B.: “Using Information Networks to Study Social Behavior: An Appraisal”, *IEEE Data Eng. Bull.* 30(2): 6-14, 2007.
- [Kilduff and Tsai 06] Kilduff M., Tsai, W.: “Social Networks and Organizations”, SAGE Publications, London, Thousand Oaks, New Delhi, 2006.
- [Klamma et al. 06] Klamma, R., Spaniol, M., Cao, Y., Jarke, Y.: “Pattern-based cross media social network analysis for technology enhanced learning in Europe”, Nejd, W., (eds): *Innovative approaches for learning and knowledge sharing: Proceedings of First European Conference on Technology Enhanced Learning, EC-TEL 2006*, Crete, Greece, October 1-4, 2006, Berlin: Springer, 2006 (LNCS 4227), 242-256.
- [Klamma et al. 07] Klamma, R., Cao Y., Spaniol, M.: “Watching the Blogosphere: Knowledge Sharing in the Web 2.0”, Nicolov, N., Glance, N., Adar, E., Hurst, M., Liberman, M., Martin,

- J.H., Salvetti, F., (eds): International Conference on Weblogs and Social Media, Boulder, Colorado, USA, March 26-28, 2007, 105-112.
- [Kleinberg 99] Kleinberg, J.: "Authoritative Sources in a Hyperlinked Environment", J. ACM, 49:604–632, 1999.
- [Latour 91] Latour, B.: "Technology is Society Made Durable", Law, J., (eds): A Sociology of Monsters: Essays on Power, Technology and Domination, London: Routledge:103–31, 1991.
- [Lakoff and Johnson 80] Lakoff, G., Johnson, M.: "Metaphors We Live By", Chicago: University of Chicago Press, 1980.
- [Lin et al. 08] Lin, C.-Y., Ehrlich, K., Griffiths-Fisher, V., Desforges, C.: "SmallBlue: People Mining for Expertise Search MultiMedia", 15(1):78-84, January-March, 2008.
- [Neuenschwander et al. 05] Neuenschwander, M., Lewis J., Blum, D.: "Enterprise Identity Management: Moving from Theory to Practice", Burton Group, Identity and Privacy Strategies, In-Depth Research Overview, June, 2005.
- [Pennebaker et al. 07] Pennebaker, J.W., Chung, C.K. Ireland, M., Gonzales, A., Booth, R.J.: "The Development and Psychometric Properties of LIWC2007", LIWC Inc, Austin, Texas 78703 USA in conjunction with the LIWC2007 software program, <http://www.liwc.net/LIWC2007LanguageManual.pdf> {10.04.2008}.
- [Powell and Recordon 07] Powell A., Recordon, D.: "OpenID: Decentralised Single Sign-on for the Web", Ariadne issue 51, April 30, 2007, <http://www.ariadne.ac.uk/issue51/powell-recordon/> {10.04.2008}.
- [Recordon and Reed 06] Recordon D., Reed, D.: OpenID 2.0: "A platform for user-centric identity management", Proceedings of the second ACM workshop on Digital identity management, Alexandria, Virginia, USA, ACM Press, 11-16, 2006.
- [Shastri 91] Shastri, L.: "Why semantic networks?", Sowa, J.F. (eds): Principles of Semantic Networks, Morgan Kaufmann Publishers, inc., 1991.
- [Vega-Redondo 07] Vega-Redondo, F.: "Complex Social Networks", Cambridge University Press, 2007.
- [Wasserman and Faust 94] Wasserman S., Faust, K.: "Social Network Analysis", Cambridge University Press, 1994.