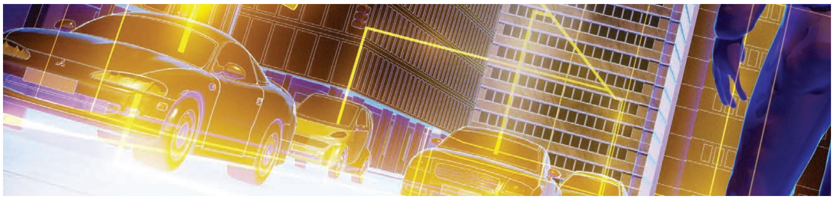
Data Protection in Healthcare Social Networks

Jingquan Li, Texas A&M University–San Antonio

// *The open philosophy of contemporary healthcare social networking sites can result in unauthorized use and disclosure of sensitive personal health data. A set of system requirements can help tackle these problems.* //



HEALTHCARE SOCIAL NETWORKING sites (HSNSs) have grown rapidly in recent years. HSNSs, like Sermo (www.sermo.com) and PatientsLikeMe (www.patientslikeme.com), provide healthcare professionals and consumers with tools and services to facilitate communication and information sharing within virtual communities to improve both healthcare and public health. Broadly, HSNSs are either healthcare professional- or consumer-oriented. HSNSs for professionals provide an online social environment for medical professionals and researchers to easily establish contact with each other and share clinical insights, observations, and medical knowledge. HSNSs for consumers let users share information and stories about their medical problems, treatments, and healthy living approaches, thereby supporting and inspiring each other.

However, the inherent openness of most HSNSs has raised serious privacy concerns. People are often willing to reveal many more private details within their online communities than they otherwise would. Unlike general social networks such as Facebook and Twitter, HSNSs maintain a vast repository of sensitive health-related data. Such repositories could be an attractive target to malicious insiders and outsiders; the openness philosophy, which many in the HSNS community cherish, could result in unauthorized commercial use and disclosure of health data.

Prior research about how best to protect data privacy has mainly concerned general social networks and narrowly focused on specific technologies such as privacy settings[1,2] and data anonymization.[3–5] However, the fundamental challenge to data protection in healthcare social networks is more system-related than technical. Because health information is highly sensitive, it's crucial for service providers, system architects, and researchers to develop private and secure HSNSs. The purpose of this article is to identify specific privacy and security problems on HSNSs and define a set of privacy-preserving system requirements.

Healthcare Social Networking's Impact on Healthcare

Looking at HSNSs as an evolving social networking paradigm could illuminate a way to new healthcare dimensions. First, social networking helps bring about effective behavioral change—normally the hardest element of any medical treatment. Most online patient communities are characterized by two functions: informational and empathetic support.[6] HSNSs let patients harness the wisdom of crowds in considering treatments for their respective diseases. Furthermore, patients' perceived empathy could be vital in patient recovery, specifically by enhancing patients' compliance with

0740-7459/14/$31.00 © 2014 IEEE

treatment protocols and the pace of healing. Thus, HSNSs have the potential to change patients' health behaviors and significantly enhance their lives by helping them change damaging behaviors and start positive behaviors.

Marketing

HSNSs could also boost online marketing and advertising in healthcare. HSNSs let individuals use online technologies to provide dynamic online content, such as opinions, insights, and experiences, and these sites can then sell the resulting information to pharmaceutical and medical device companies that are developing or selling relevant products and services to consumers. Marketers (in the health domain and elsewhere) have developed powerful data mining tools to exploit such information about individuals and groups and use it to create effective marketing campaigns.

Government

HSNSs also influence industries and governments that monitor users' conversations and use such feedback in decision making. Some HSNSs receive funds from healthcare institutions, pharmaceutical companies, financial institutions, or government agencies. Such stakeholders are interested in early indicators that could affect healthcare, such as symptoms of an impending epidemic. These entities can glean information directly from large communities through analyses of user-generated content that will help to both plan future investments and influence the healthcare and financial industries as well as regulatory bodies.

Research

HSNSs have an enormous potential to accelerate medical research and science and be a conduit for healthcare providers to achieve better patient outcomes. Advances in digital medicine such as patient-driven research have already started to move medical systems' focus to patients. HSNSs globally attract a large number of users with widely varying demographics. Users not only search and collect data but also generate and share data with each other; pharmaceutical and equipment companies, insurers, universities, and research labs have purchased data from HSNSs—which has in turn created revenue that feeds back into the sites. User-generated content helps researchers better understand the real-world medical value of therapies, drugs, and medical devices so they can improve them and speed up the development of new solutions for patients. Thus, the biggest gains that HSNSs will provide will come in the shape of new, better treatments for difficult diseases.

Advances in digital medicine such as patient-driven research have started to move medical systems' focus to patients.

Privacy

HSNSs, although offering new opportunities to improve healthcare, also present a major issue of data privacy. The data shared via HSNSs often contains sensitive personal information about chronic diseases, mental health, psychiatric care, sexual behavior, fertility, abortion, sexually transmitted diseases, HIV, substance abuse, physical abuse, genetic predispositions, and so on. The misuse of such personal health information could cause embarrassment, humiliation, discrimination, economic hardship, and even identity theft. The EU, through directives such as the 1995 Directive 95/46/EC on the protection of personal data, requires 27 member states to enact legislation to ensure that citizens have a right to privacy. In the US, however, HSNSs aren't currently covered under the Health Insurance Portability and Accountability Act of 1996 (HIPAA) privacy rule or the enhanced Health Information Technology for Economic and Clinical Health (HITECH) Act.[7] Thus, HSNSs could take advantage of powerful commercial interests at the expense of users' personal privacy.

Case Studies

Two case studies show the types of privacy problems that might be inherent to the current generation of HSNSs.

Sermo

Sermo (www.sermo.com) is a physician-only community in which users can collaborate on difficult cases, exchange observations, and get help with patient care and practice management. While a surgical resident at Harvard's Beth Israel Deaconess, Daniel Palestrant noticed that doctors learned a lot from informally chatting with colleagues but they had no means of sharing, comparing, or evaluating those insights with members of the wider medical community. Moreover, many physicians working in rural areas found it difficult to schedule and attend conferences or courses to keep themselves up to date and learn from other health professionals. In 2006, Palestrant launched Sermo (named after the Latin word for *speech*) with the vision of forging a platform of physician communication and collaboration. Today, Sermo is the largest online physician community.

> Palestrant launched Sermo with the vision of forging a platform of physician communication and collaboration.

Sermo doesn't make money from selling advertising. The idea behind Sermo's business model is to create content that is valuable to outside parties, including hedge funds, Wall Street firms, and big pharmaceutical companies, to allow them access to the community and charge them for this access. User-generated content is commercially valuable to the community to learn the usefulness of drugs or medical devices and thereby gain exclusive insights into products. A chief concern from a privacy perspective is how these information-sharing partnerships can be used against an individual. According to the site's privacy policy, Sermo is not responsible for the confidentiality, security, or distribution of users' own personal information by third parties outside the scope of its contract with them. Furthermore, the site fails to be fully transparent about its practices related to the sharing of users' profiles with third parties.

Although Sermo promises physicians an exclusive, closed community where their conversations are accessible only to professional peers, its sheer size has made it easy for people to have access to confidential conversations for various purposes. For example, a physician-turned-hospital administrator looking for negative information about a trouble-making internist or a malpractice attorney using a brother-in-law's login ID to troll for potential cases could obtain access to Sermo. In fact, Sermo doesn't rely on invitations, but rather on four pieces of personal information for verification: a physician's name, the medical school he or she attended, the date of graduation, and the Drug Enforcement Agency (DEA) number. These pieces of information can be easily obtained from public databases. Thus, a malpractice lawyer, patient, insurer, journalist, medical consultant, or anyone else, with the help of public databases, can check and see what's going on behind the physician community's supposedly closed doors. Under its privacy policy, Sermo states,

> *It is important to remember that whenever you voluntarily disclose information in a forum, through e-mail or elsewhere, that information can be collected and used by others. If your personal information is accessible to the public, you may receive unsolicited messages from other parties in return. Ultimately, you are solely responsible for maintaining the secrecy of your personal information.*

PatientsLikeMe

PatientsLikeMe (www.patientslikeme.com) is an online community that lets users share information and real-world experiences that can improve the lives of patients with life-changing diseases. PatientsLikeMe was founded in 2004 by three engineering graduates from MIT—brothers Jamie and Ben Heywood and their friend Jeff Cole. The website grew from the efforts of the Heywood family to care for another brother, Stephen, who was diagnosed with amyotrophic lateral sclerosis (ALS). The idea was that if patients with ALS could share details about their symptoms and treatments, then better treatment plans and options could be identified by the collective wisdom on this site. PatientsLikeMe has grown from a single network around ALS to one that addresses more than 1,200 health conditions.

Unlike exclusive communities like Sermo, PatientsLikeMe is a public, open community. On PatientsLikeMe, users from around the world write very personal things in forums, often using their real names. Although the website offers privacy settings, patients can still volunteer their personally identifiable information. Thousands of users have access to sensitive health information, and some could disclose negative things—intentionally or unintentionally—about others. Furthermore, malicious individuals or entities can create fake accounts to obtain information from innocent users. Therefore, it's difficult to prevent unauthorized use and disclosure of personal information.

Users could also fall victim to the site's business model. Like Sermo, PatientsLikeMe doesn't sell advertising. Rather, the company earns its revenue through data-sharing partnerships with, for example, pharmaceutical and medical device companies. PatientsLikeMe's own research teams have also utilized the site's user-generated content. Although information sharing is limited to deidentified data, it poses new threats to patient privacy because sophisticated data mining and deanonymization techniques could reidentify people, even from such anonymous information.[4,8] One concern is that targeted marketing could influence patients to seek drugs they don't need or spend more money on branded drugs rather than generics. Moreover, employers or health insurers could gain access to a patient's profiles, leading to potential penalties against the patient. Finally, the site doesn't meaningfully disclose how the data is collected and used, although it gives its partners many patients' digital profiles that can include conditions, concerns, fears, and behaviors.

Privacy Risks

Both open and closed communities have the potential to introduce a new range of privacy risks. For instance, the service provider has the ability to observe and accumulate information that users share online and maintain that data in perpetuity. Users are increasingly sharing intimate details online with little concern for who might be viewing their information. Some users are even willing to share personally identifiable information, such as real names and photos, together with medical records with the site. Almost any information users post online can spread so quickly that they can't feasibly withdraw the information if they change their minds about sharing it. For example, a user might post a psychiatrist's notes about his or her serious mental illness to get feedback on the condition or treatment. Then, another user might save that information and either communicate it to other people or post it on the Web. Sooner or later, it could reach the inbox of someone the user really doesn't want to see it, such as an employer or insurer. Furthermore, users typically have no control over retention periods for their shared data, which could live forever on the Web.

Another concern is that an HSNS environment's flexibility and complexity can create problems for users trying to assess risks associated with sharing personal data. Social network formation is often accomplished with such ease that the networks are often more expansive than one might expect, leading users to misjudge their actual exposure.[9] This architectural fact has privacy implications that many users surely fail to consider, not the least of which is that profiles and messages available within a gated community can be copied, intentionally or often unintentionally, and made available on the Web.[10] For example, data scrapers can copy messages from the community via fake accounts. Another related issue is that many HSNSs allow for complex data-sharing partnerships and usage scenarios—many of them use it as their core business model. Therefore, health data can be exposed to various recipients who might not have the users' best interests in mind.

Health data can be exposed to various recipients who might not have the users' best interests in mind.

The accumulated health data can be used for many different purposes. HSNSs can release users' health data to numerous organizations including marketers, employers, pharmaceutical companies, insurance companies, government agencies, and others. Some HSNSs automatically share health data with third-party applications. Innovative data mining and consumer-profiling technologies can link data produced from different sources to produce useful personal digital profiles. Individual pieces of information don't communicate much about a person. Put together, however, little pieces of information such as demographic data, medical conditions, medication, children's names, pets, employers, updates about life-changing events, lists of hobbies, photos, friendship links, and group memberships create a bigger picture that could communicate a great deal of identifying information. Such digital profiles can be immensely valuable to companies looking to market products or, in the case of some insurers or employers, deny a policy or job. These digital profiles, maintained without direct government oversight, would also be an attractive target for hackers and identity thieves.

Another issue is the scale of security risk. Users face a variety of internal and external security risks in casually posting personal information online.[11] HSNSs create a perfect social and ecological environment for viruses, phishing and malware attacks, and social engineering attacks, mixing the exploitation of human vulnerability, easy and direct access to hundreds of millions of people, and an unprecedented amount of personal data.[12] Although encrypted transmission will improve security and authentication and access control will reduce unauthorized access, one hack into the system, error by a site operator, or misuse by an authorized user can compromise many digital profiles.

> User-generated content is a lucrative commodity and can be used for various commercial and medical purposes.

Although these risks aren't exclusive to HSNSs, the nature of HSNSs make their users much more vulnerable to such risks than those of "fun" social networks, such as Facebook, for several reasons:

- Health information is sensitive, and there is no way to withdraw health information about a person once the information is exposed and the resulting damage is done to that person.
- Patients with life-changing diseases are highly motivated to share more sensitive details in hopes that some exchange will help them or their peers find a better treatment.
- A person sharing health information can reveal health issues affecting family members, local communities, or local groups.

If an HSNS doesn't assure users that their information is private and secure, users might have fears about risks of sharing data and consequently stop swapping content; in that case, the site would no longer be able to provide data for medical research, public health, and other good causes.

Privacy-Preserving HSNS System Requirements

So far, most research and practice for data protection have focused on specific technologies and techniques. Analysis of HSNS privacy concerns show that even if technology innovations in access control, cryptography, and network security were deployed, and even if users were aware and competent in the use of sophisticated privacy settings, personal health data would still be exposed. Such exposure would include potential privacy violations by authorized users, including the omniscient service provider and its many business partners, but also privacy violations by authorized third-party applications, which could have the hidden purpose of capturing medical data and other personal, social, and professional data. In fact, Internet users are becoming increasingly concerned about online privacy and correctly believe that they're far more exposed today than they were a generation earlier.[13] From the user perspective, a willingness to provide personal information depends heavily on the level of anonymity of the information, its possible recipients, its possible uses or disclosures,[1] and its storage and transmission security. From the provider perspective, user-generated content is a lucrative commodity and can be used for various commercial and medical purposes. The problem with HSNSs isn't just that the privacy technology doesn't exist but that system requirements aren't formulated for the strong protection of personal data or aren't implemented properly. Therefore, the most fundamental challenge to HSNS information privacy is "more system-related than technical."[14] However, several key system requirements can better ensure privacy.

Protect Data by Default

An HSNS should make personal profiles fully private by default without requiring a user's action—an opt-in, maximum privacy approach. In other words, it should define a floor level of privacy protection, even if users don't change their privacy settings. Furthermore, it should collect only the minimum amount of person-specific data to accomplish its intended business purposes. For example, it might adopt an architecture in which the user is anonymous from the perspective of not just other users but even the platform itself.

Employ Privacy-Preserving Data Sharing

Privacy-preserving mechanisms should establish accessibility or electronic exchange of health data. Several methods exist for an HSNS to ensure data-sharing privacy. First, it can provide built-in mechanisms that afford users fine-grained control over how their personal information is shared. In addition to covering personally identifiable information, the mechanisms can let users set privacy policies for information about their medical conditions and other sensitive information. To reduce the burden on users, the site can provide intuitive user interfaces for changing privacy preferences. Second, the information exposed must be made anonymous via sophisticated anonymization techniques such as eliminating personal identifiers,[15] cluster-based and graph modification approaches,[3] and network data anonymization.[4,5]

A decrease in anonymity could make deidentified data reidentifiable owing to the range of publicly available background information readily available on the Web. Hence, the technique must prove that it requires considerable time, efforts, cost, and skill to reidentify anonymized data.[16] Third, an HSNS could provide means by which users can visualize their current exposure on the site. Users often underestimate the scope of information sharing. Graphic displays of the social network could help users appreciate the potential risks of disclosure.

Prohibit Privacy-Invasive Uses or Disclosures

An HSNS must carefully weigh the risks and benefits of uses or disclosures of health data, giving special attention to legal constraints and ethical considerations. HSNS administrators might find it necessary to prohibit certain privacy-invasive uses or disclosures of health data regardless of user consent.[7] We know that users are concerned about employers, insurers, or credit card companies having access to their medical records. Therefore, health data must not be exposed for purposes of discrimination in the context of employment, insurance, or credit. Moreover, if an HSNS doesn't prevent inappropriate use or disclosure of personal information, the Federal Trade Commission (FTC) can bring action against the company for engaging in a "fraudulent, deceptive, and unfair business practices" (http://ftc.gov/opa/2011/11/privacysettlement.shtm). More importantly, privacy protection makes good business sense because it increases users' trust in the system and reduces costs for data breaches. Greater exchange of information would result, benefitting both the site and users.

Monitor and Audit Data Usage

Users should have clear ways to verify adherence to data protection. An HSNS should undergo privacy and security audits to track who views a user's profile. So far, transparency hasn't been a strong point of today's HSNS. Information mash-ups and the combination of multiple services create unexpected information flow through back channels, impeding users' ability to get a clear view of the way their data is propagating. There have been allegations and supporting evidence that some social networks don't follow the privacy policies to which they claim to adhere (http://ftc.gov/opa/2011/11/privacysettlement.shtm). An HSNS can't assure users of their privacy

TABLE 1

System requirements for data protection in healthcare social networks.

System requirements	Recommendations
Protect data by default	Make profiles private by default Collect the minimum amount of person-specific data
Employ privacy-preserving data sharing	Provide user controls over data sharing Share only deidentified data within and beyond the network Revere visibility and transparency
Prohibit privacy-invasive uses or disclosures	Prohibit inappropriate uses or disclosures of health data Comply with fair information practices
Monitor and audit data usage	Audit and log uses and disclosures Provide audit trails upon request
Provide system-wide security	Use system-wide preventive and detective security measures such as risk analysis, SSL logins, encryption, security audits and monitoring, and breach notification

unless patients can request an audit trail detailing when their data was accessed, by whom, and for what purpose.

Provide System-Wide Security

An HSNS must provide a secure place to keep health data. Identity thieves are increasingly using sophisticated methods and techniques via the Internet to collect personal data. Authenticating users and encrypting communication between users and the site provide enhanced security. But identity thieves can still create fraudulent accounts to steal health data. Owing to the sensitivity of health data, it's important for HSNSs to adopt encryption technology for the transmission and storage of personal data. Nevertheless, encryption won't stop insiders from tweeting or posting within the site. Furthermore, social networking creates an ideal virtual community for an attacker to hit a large number of targets quickly and effectively. There's no sure way to protect the community against social engineering attacks because the vulnerability stems from the natural human tendency to trust other people. One step HSNSs can take is to develop system-wide preventive and detective security requirements for protecting health data from both insiders and outsiders (see Table 1).

The requirements that I've described in this article by no means cover all privacy controls, but they form a basis for how to start thinking about the privacy of online health data. These requirements can also be extended to other virtual environments. Some research has recommended data protection through privacy settings for social networking in education;[4,5] other research has used sophisticated anonymization techniques for social networking in business.[3–5,8,11] More importantly, an HSNS needs to implement and enforce requirements as well as issue privacy policies based on these requirements. However, the complexity of the virtual environment—apparent by the rich user interactions, application interactions and mash-ups, and sophisticated user-generated content that it offers—poses unique challenges to implementing and enforcing the requirements.

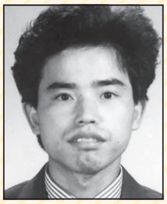
HSNSs have great potential to serve personal and professional information and communication needs, improve patient care, and enhance medical research and public health. The continuing growth of HSNSs relies on participation of more people and sharing of more accurate and credible content online. Ample opportunity exists for future research in this area. Although I've discussed fundamental requirements for HSNSs in this article, detailed requirement implementation and enforcement, business practices, site operator education, and user education are also necessary components of the privacy solution.

References

1. R. Gross and A. Acquisti, "Information Revelation and Privacy in Online Social Networks," *Proc. ACM Workshop Privacy in the Electronic Society*, ACM, 2005, pp. 71–80.
2. K. Lewis, J. Kaufman, and N. Christakis, "The Taste for Privacy: An Analysis of College Student Privacy Settings in an Online Social Network," *J. Computer-Medicated Comm.*, vol. 14, no. 1, 2008, pp. 79–100.
3. B.C.M. Fung et al., "Privacy-Preserving Data Publishing: A Survey of Recent Developments," *ACM Computing Surveys*, vol. 42, no. 14, 2010, pp. 1–53.

4. J.M. Kleinberg, "Challenges in Mining Social Network Data: Processes, Privacy, and Paradoxes," *Proc. 13th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, ACM, 2007, pp. 4–5.
5. E. Zheleva and L. Getoor, "To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles," *Proc. 18th Int'l Conf. World Wide Web*, ACM, 2009, pp. 531–540.
6. P. Nambisan, "Information Seeking and Social Support in Online Health Communities: Impact on Patients' Perceived Empathy," *J. Am. Medical Informatics Assoc.*, vol. 18, no. 3, 2011, pp. 298–304.
7. J. Li, "Privacy Policies for Health Social Networking Sites," *J. Am. Medical Informatics Assoc.*, vol. 20, no. 4, 2013, pp. 704–707.
8. A. Narayanan and V. Shmatikov, "De-anonymizing Social Networks," *Proc. 30th IEEE Symp. Security and Privacy*, IEEE CS, 2009, pp. 173–187.
9. J. Williams, "Social Networking Applications in Health Care: Threats to the Privacy and Security of Health Information," *Proc. 2nd Workshop Software Eng. in Health Care*, ACM, 2010, pp. 39–49.
10. N. Mooradian, "The Importance of Privacy Revisited," *Ethics Information Technology*, vol. 11, no. 3, 2009, pp. 163–174.
11. D. Rosenblum, "What Anyone Can Know: The Privacy Risks of Social Networking Sites," *IEEE Security & Privacy*, vol. 5, no. 3, 2007, pp. 40–49.
12. M. Brandel, "Baited and Duped on Facebook," *Computerworld*, vol. 43, no. 31, 2009, pp. 28–33.
13. D.J. Weitzner, "Personal Privacy without Computational Obscurity: Rethinking Privacy Protection Strategies for Open Information Networks," *Proc. 23rd Ann. Computer Security Applications Conf.*, IEEE CS, 2007, pp. 173–175.
14. D.J. Weitzner, "Beyond Secrecy: New Privacy Protection Strategies for Open Information Spaces," *IEEE Internet Computing*, vol. 11, no. 3, 2007, pp. 96–97.
15. B. Krishnamurthy and C.E. Wills, "On the Leakage of Personally Identifiable Information Via Online Social Networks," *Proc. 2nd ACM Workshop Online Social Networks*, ACM, 2009, pp. 7–12.
16. K. El Emam, "Risk-Based De-Identification of Health Data," *IEEE Security & Privacy*, vol. 8, no. 3, 2010, pp. 64–67.

ABOUT THE AUTHOR



JINGQUAN LI is an associate professor at Texas A&M University–San Antonio. His research interests include information security and privacy and data mining. Li received a PhD in information systems from the University of Illinois at Urbana–Champaign. He's a member of the IEEE Computer Society. Contact him at jli@tamusa.tamus.edu.