Privacy in Online Social Networks

Michael Beye[1], Arjan Jeckmans[2], Zekeriya Erkin[1], Pieter Hartel[2], Reginald Lagendijk[1], and Qiang Tang[2]

[1] Information Security and Privacy Lab, Faculty of EEMCS, Delft University of Technology
{m.r.t.beye; z.erkin; r.l.lagendijk}@tudelft.nl
[2] Distributed and Embedded Security, Faculty of EEMCS, University of Twente
{a.j.p.jeckmans; q.tang; pieter.hartel}@utwente.nl

Summary. Online Social Networks (OSNs) have become part of daily life for millions of users. Users building explicit networks that represent their social relationships and often share a wealth of personal information to their own benefit. The potential privacy risks of such behavior are often underestimated or ignored. The problem is exacerbated by lacking experience and awareness in users, as well as poorly designed tools for privacy-management on the part of the OSN. Furthermore, the centralized nature of OSNs makes users dependent and puts the Service Provider in a position of power. Because Service Providers are not by definition trusted or trustworthy, their practices need to be taken into account when considering privacy risks.

This chapter aims to provide insight into privacy in OSNs. First, a classification of different types of OSNs based on their nature and purpose is made. Next, different types of data contained in OSNs are distinguished. The associated privacy risks in relation to both users and Service Providers are identified, and finally relevant research areas for privacy-protecting techniques are discussed. Clear mappings are made to reflect typical relations that exist between OSN type, data type, particular privacy risks and privacy-preserving solutions.

1 Introduction

In recent years, Online Social Networks (OSNs) have attracted many millions of users worldwide. Even though Social Networks have always been an important part of daily life, now that more and more people are connected to the Internet, their online counterparts are fulfilling an increasingly important role. OSNs have also become a hot topic in areas of research ranging from sociology to computer science and mathematics.

Aside from allowing users to create a network to represent their social ties, many OSNs facilitate uploading of multimedia content, various ways of communication and sharing many aspects of daily life with friends. People can stay in touch with (physically remote) friends, easily share content and experiences and stay up to date in the comfort of their own home or when on the move.

However, benefits aside, potential threats to user privacy are often underestimated. For example, due to the public nature of many OSNs and the Internet itself, content can easily be disclosed to a wider audience than the user intended. Users often have trouble revoking or deleting information, and information about a user might even be posted by others without their consent. Privacy in OSNs is a complicated matter and is not always intuitive to users, especially because it is not always similar to how privacy works in real-life interactions.

Ideally, users should be able to trade some privacy for functionality, without their information becoming available beyond the scope they intend. For example, a user of a self-help OSN (e.g. www.patientslikeme.com) would like to meet people with the same medical condition, but doest not want everyone to know about his ailment. Even in less extreme cases, the importance of privacy is often underestimated.

In this chapter, we will observe the privacy risks OSN users face, what causes them and which techniques may help to minimize these risks. To this end, we first look at OSNs as they currently exist (Section 2), leading to a classification of OSNs based on their type and purpose, and a classification of data types in OSNs. We then map these to associated privacy risks in relation to both fellow users and Service Providers (Section 3), and finally give an overview of existing research into privacy-enhancing technologies (Section 4). Through tables, the relationships between these various aspects are mapped, providing a comprehensive overview. In Section 5 conclusions are drawn.

2 Classifying Online Social Networks

Let us begin by framing the concept of Online Social Networks, and observe how OSNs have become as widely used as they are today. This will help us understand the purpose of OSNs (which forms the basis for our classification), but also help to illustrate the needs of users, the environment they navigate, and potential threats as discussed in further sections.

2.1 Definition of an OSN

Boyd and Ellison's widely used definition [7] captures the key elements of any OSN:

Definition 1. *An OSN is a web-based service that allows individuals to:*

1. *construct a public or semi-public profile within the service,*
2. *articulate a list of other users with whom they share a connection,*
3. *view and traverse their list of connections and those made by others within the service.*

The terms to describe a connected user include, but are not limited to: 'friend' (www.facebook.com and www.myspace.com), 'professional' (www.linkedin.com), 'relative' (www.geni.com), 'follower' (twitter.com), 'subscriber' (www.youtube.com). Typically a connection is bidirectional (symmetric), but this is not always the case. For example, 'following' on Twitter or 'subscribing' on Youtube are one-way relationships.

2.2 The Rise of Online Social Networks

The first OSN to see the light of day was Sixdegrees in 1997 [14]. SixDegrees allowed users to create profiles, list and message their friends and traverse friends listings, thus fitting Boyd and Ellision's definition of an OSN. Even though there were millions of users, these did not have that many direct friends and SixDegrees did not offer much functionality besides messaging. The website finally shut down in 2000 [7].

During this period other websites started adding OSN features to their existing content, essentially becoming OSNs, with various degrees of success. In the years that followed, new OSNs started from scratch and began to offer functionality beyond simply listing and browsing friends. Ryze.com and later www.linkedin.com tailored to professionals looking to exchange business contacts, while www.friendster.com focused on dating and finding new friends. Friendster became widely used and experienced technical (performance and hardware) and social (fake profiles and friendship hoarding) difficulties due to its rapid growth. The technical issues and actions taken to combat the social difficulties eventually caused many users to seek out other OSNs. Despite this, Friendster is still popular, particularly in the Phillipines, Indonesia and Myanmar [44].

The popularity of Friendster encouraged the creation of other similar "social OSNs", like www.myspace.com and www.orkut.com. While Myspace has become popular among youth worldwide, Google's Orkut has attracted a predominantly Brazilian and Indian crowd [44]. Aside from these clearcut "social OSNs", a wide variety of niche OSNs have emerged, each catering to a particular interest. Adding the social structure of an OSN to existing services can often enrich them, making them more useful and attractive to users, or binding users to providers. Currently, OSNs form an integral part of the Internet.

As we have seen, not all OSNs are alike: they can serve different uses for disparate target audiences. A clear classification of OSNs can help us to understand what OSNs mean to their users and how they are used, which in turn will help us to structure our thoughts on privacy in OSNs.

2.3 Existing Classifications

It is remarkable that hardly any classifications for OSNs exist in scientific literature, even though OSNs are studied in many disciplines. However, some pseudo-scientific blogs and marketing resources offer relevant thoughts on the matter, a selection of which are summarized below.

Classifications by Topical Focus

Lovetoknow.com [17] classifies OSNs based on their topical focus:

- *Informational*. Seeking answers to everyday problems

- *Professional*. Helping you to advance within your career or industry
- *Educational*. Collaborate with other students or academic projects
- *Hobbies*. Conduct research on their favorite projects or topics of interest related to personal hobbies
- *Academic*. For important collaboration within the scientific community, over the Internet
- *News*. Those that publish “community content”

Such a topical point of view seems very relevant, although the categories of Informational, Educational and Academic seem to have some overlap.

Onlinebrandmanager.com [40] first classifies OSNs into four main areas:

- *Dating / friendship*
- *Alumni networks*
- *Career / business related*
- *Hobby / group networks*

They then state that these can be further split up into: *Book communities*, *Business Networking & Professionals*, *Family*, *Friends*, *Hobbies & Interests*, *Languages*, *Video Sharing*, *Photo Sharing*, *Audio Sharing*, *Mobile Communities*, *Shopping*, *Social Bookmarking*, *Students* and *Travel & Locals*. They note that these are broad categories, where a specific OSN may fit several categories. We remark that subcategories do not always seem logical extensions of the main categories, and their interrelation is not clearcut. Note however, that many categories are again topical, while some categories seem to focus on the *purpose* for which users visit the OSN.

Classifications by Topical Specificity

In contrast, Hudsonhorizons.com [27] uses topical *specificity* to divide OSNs into two groups:

- *Broad-range*. “Some social networking websites, such as Facebook, fall into the ‘general’ category; they accommodate folks of **all interests and backgrounds**. On this type of social networking websites, members can often include a list of their interests, and then locate members with similar interests by searching for keywords and key phrases. The main purpose of general social networking sites is to serve as a social platform where people can reunite with old friends, stay connected with current ones, and even make new acquaintances.
- *Niche*. “Other social networking sites have **tight, niche focuses**, and cater to specific groups of people. Social networking sites can revolve around sports, dating, culture, hobbies, ethnicity, education, romance, entrepreneurship and more.”

Note that the topic in question, or the goal behind it (dating vs. talking about hobbies vs. learning) does not play a role in their classification.

The following quote is from Enid Burns on Clickz.com [10], regarding advertising through OSNs:

> Many of these sites target communities defined by their affinity to a vertical industry, business model, or interactivity type, unlike Myspace and Youtube, which are designed to appeal to the mass population.

Again, the broad distinction made here seems to be on topical specificity.

Liz Gannes on Gigaom.com [22] also devotes a blog entry to classifying OSNs. The following three terms form the core of her argument:

- *Blank slates*. Gannes names Myspace and Bebo as typical examples. These seem to be what others might call "broad-range" or "general" OSNs.
- *Target audiences*. Targeted to a specific niche; Gannes compares them to ad verticals.
- *Existing interests / existing communities*. She names last.fm as an example, where OSN functionality is well-integrated into an existing activity (listening to music). This category seems to center around integrating OSN functionality into an already established, successful community.

Gannes also mentions "social tools": sites that have a certain goal in mind, such as LinkedIn. She describes the difference between a social network and a social tool as "a place to hang out for X kind of people" vs. "a place to get X done." One of her readers states that OSNs seem to have two main purposes: "communication" (networking with pre-existing group) and ”self-expression" (social network is just a feature). Another reader proposes "Community around shared services" (e.g. Del.icio.us) as a separate category. We note that this echoes concepts from previously discussed sources: topical specificity (general social OSNs vs. niche OSNs), and the *purpose* of an OSN playing a central role.

Classifications by Other Criteria

Bernard Lunn on Readwriteweb.com [34] also divides OSNs into two types, namely "Open networks" and "Gated communities". This distinction is centered around trust – in some communities (they name OSNs for relatives, doctors or models) trust may be more important than in others, and users will wish to interact in a gated community, that shields them from the outside world. Lunn notes that this does not directly relate to the size of the OSN or its significance to society, although the concept of gated communities seems related to both Hudsonhorizon's *niche OSNs* and restricted membership OSNs.

Dave Emmett, on his blog [19], looks at the effect an OSN has on users' personal networks:

- *Tightening*. Deepens existing relationships. Examples include Facebook, Dopplr and Friendfeed.
- *Broadening*. Adds new connections. Examples include Twitter, Brightkite, Flickr and Youtube

One of his readers comments that this seems related to the concepts of “bridging and bonding” in social sciences theory. The main difference here lies in the audience that a user intends to reach.

Dominique Cardon on internetactu.net [11] discusses the visibility and interactions of users in OSNs. He discussion on the following categories is freely translated from French:

- *The Screen*. People only meet through criterion search, and are otherwise invisible. Users are matched online and test their compatibility in the real world.
- *The Clear-obscure*. People share information on their daily private lives, but mostly to a select audience. These settings are about strengthening pre-existing relationships or explore friends-of-friends.
- *The Lighthouse*. People portray their identity, preferences or content with the general audience. Uses its high visibility to expand beyond real-life friends and find a larger audience.
- *The Post-it*. Users show their presences and availability through contextual clues, but to a restricted circle. The real and virtual worlds are highly interwoven.
- *The Magic Lantern*. Users employ personalized avatars as pseudonyms to decouple their online and offline identities. Interactions are mostly virtual and rarely extend into the real world.

Some remarks are made to relate network size, structure, homogeneity and growth to visibility. Cardon continues to discuss navigation methods, like “criterion referenced search engine”, “friend network navigation” and “user-activity driven search”. It is clear that the *goal* of users in an OSN plays an important role throughout Cardon’s views.

Finally, Wikipedia [15] offers the following thoughts on OSN classification:

> [...]Although online community services are sometimes considered as a social network service in a broader sense, social network service usually means an individual-centered service whereas online community services are group-centered. Social networking sites allow users to share ideas, activities, events, and interests within their individual networks.
> The main types of social networking services are those which contain category places (such as former school-year or classmates), means to connect with friends (usually within self-description pages) and a recommendation system linked to trust. Popular methods now combine many of these[...]

One can imagine yet other ways to distinguish between OSNs:

- *Source of revenue*. The OSN Service Provider can earn his revenue through direct or indirect means (subscriptions or micro-payments versus advertisements or data sales).
- *Membership type*. This can be open, select or invitation only.
- *Wideness of user base*. Does the OSN attract a worldwide, national or regional audience, or does it target a specific demographic or subculture.

2.4 New Classification of OSNs Based on Purpose

Recall that our classification is intended to structure our thoughts on privacy in OSNs. We feel that the *purpose that an OSN fulfills to its user base* is the main factor to determine the functionalities it offers, which in turn dictates what sort of data exists in the network, and how users can interact – this data and user interaction are what privacy is all about. Therefore, our approach may come to resemble those of [40] and [17] most, although we feel that none of the above classifications provide completeness, non-overlapping categories or a true focus on purpose.

For each category, some illustrative examples will be provided. We make our first broad distinction between OSNs that focus on *connections* and those that focus on *content*.

Connection OSNs

Connection OSNs focus on the social connections and interactions between users, by providing users with a social contact list, channels for interaction or matching services. Their general purpose is usually to connect users to new or existing friends and acquaintances, or to provide an easy way to maintain such relationships.

Dating. Dating sites are websites that aim to help users to find the love of their life – many dating sites incorporate OSN aspects these days. Each user has login credentials and usually a profile to attract potential lovers. Connections are typically in the form of love interests, but friendship links are also common; user groups may also exist. Traversing the OSN is often based on browsing, searching or recommendation generation, rather than through navigating existing connections. Messages exchanged between users are often kept private to these users, although in some cases comment sections viewable by others are offered. Behavioral information can be kept by the OSN to provide better recommendations. Example dating sites are www.match.com, www.paiq.nl and www.plentyoffish.com.

Business. These OSNs aim to provide professionals with useful business contacts. Searching for profiles does not always require signing up. Profiles display users' capabilities and work field as well as a means to contact them. This is usually done through the OSN via personal messaging. Users can also add other users to their network of connections so that other professionals can see who the user is working or has contact with. An example of this class is www.linkedin.com, which requires a subscription for premium services.

Enforcing real-life relationships. These OSNs are not aimed at finding new friends, but (re)connecting users with existing friends or acquaintances. Examples include family-oriented OSNs, college or ex-classmate focused networks, such as www.mylife.com, www.odnoklassniki.ru and www.plaxo.com.

Socializing. Fitting the more traditional view of social networks, what others might call a "blank slate" or "broad-range network". Here users can connect with current friends and find new ones. Most types of information can be found in OSNs of this class; often a lot of this information is (semi-)public. The revenue for the OSN Service Provider often comes from advertisements and selling information about the OSN, but can sometimes be combined with a subscription for additional functionalities (as with www.hyves.nl for example). In order to attract new users and bind them, this type of OSN usually has a lot of additional functionalities such as social and competitive games. For a user the value of a social OSN is often largely determined by the number of friends that use that particular OSN. Some well known examples of this class are www.facebook.com, www.orkut.com and www.myspace.com.

Chat / Instant Messaging. Some (webcam-)chat websites (e.g. www.stickam.com) contain OSN features (friends list and profile). Some sources consider Instant Messaging (IM) services an OSN, if they allow users to store an explicit “address book” of friends. Popular IM clients include Windows Live Messenger (formerly MSN Messenger), AOL Instant Messenger (AIM), ICQ, Skype and Yahoo! Messenger.

Content OSNs

Content OSNs focus more on content provided by, or linked to by users. This content can be multi-media or information like knowledge, advice or news. The social interactions with other users usually revolve around and are driven by a search for information or the exchanging of said media.

Content sharing. Sharing of user-generated content can happen within a selected group, such as friends or family, or a far wider audience. Content that is shared is usually multi-media; this is often of potential interest to a wide audience, and even for selected audiences, e-mailing such content is cumbersome and often impossible due to size of the data. Uploading content generally requires users to sign up and log in; sometimes viewing content also requires logging in, or viewing is restricted through the use of hard-to-guess obfuscated URLs. Sometimes messages or tags can be added to the shared content, and especially in more open systems content tagging and recommendation may be an integral part of the system. User profiles, if any, are usually brief. Examples are Picasa (picasaweb.google.com), photobucket.com and www.youtube.com.

Resource recommendation. In some OSNs users do not focus on uploading content, but on recommending existing (usually professional, external) content or resources. Book-review sites like weread.com and URL-tagging communities like delicious.com are prime examples where external items are discovered, added to the system as links and finally tagged or rated. No actual content is created or uploaded.

Advice sharing. Offering a place for people to share their experience or expertise in a certain area with others, or to seek help and advice can be a focus for some OSNs. For example mothers-to-be (www.babycenter.com), medical patients (www.patientslikeme.com) or students (www.teachstreet.com) can help one another. Other examples include www.advogato.org for software developers, the now discontinued Cake Financial [13] and sciencestage.com.

Hobbies / Entertainment. Many OSNs focus on audiences that have similar interests and hobbies. Such OSNs may involve multi-media uploads, recommendation or advice sharing elements, but the main distinguishing feature is their homogeneous audience. This means that the topic of the OSN mainly determines its character and appeal for users. Examples are www.athlinks.com for athletes, www.care2.com for those interested in health and green living, or OSNs tied to gaming communities like Xbox Live (www.xbox.com/en-us/live/) or www.playfire.com. Entertainment OSNs might make money through advertisements or direct sales targeted to their user base's niche, or through subscriptions.

"News" sharing. Some OSNs focus on world news or gossip, but a multitude of (micro-)blogging OSNs provide a stage mainly for sharing "personal news", opinions and experiences. Examples are www.nowpublic.com, www.blogster.com, twitter.com, www.buurtlink.nl and www.gossipreport.com.

2.5 Data in OSNs

Now that we have an idea of the wide variety of OSNs and their purpose, let us take a look at the data that these systems can contain. From Boyd and Ellison's definition, we can already deduce that the following user-related data must exist in an OSN:

Profiles. A profile is tied to a user and is their representation to the outside world. Usually this is a self description, or the description of an alter-ego (pseudonym, avatar). This may typically include a short biography, a picture and attributes like age, gender, location and the like.

Connections. A connection exists between two users and can be of several types, like friend, colleague, fan, etc. A collection of connections can be represented by a graph.

Login credentials. Most OSNs require the user to login to make use of the service. A user account ties a profile to the user behind it, and to sign in the user needs certain login credentials. Such credentials can also be found in traditional websites, and this chapter will not pay special attention to the security issues surrounding them.

Depending on the goal of an OSN and the additional services it offers, other forms of information related to users can be involved:

Messages. We view messages in the broadest sense of the word. Any piece of data exchanged between a user and another user or a group of users is a message; these may contain text or multi-media. Messages form the basis for additional OSN functionalities. Interaction between users has been recognized as a rich source of information on the underlying social network, even more so than friendship graphs [51]. Note that in some cases a message can be instantaneous and short-lived, as in an Instant Messaging setting. In other cases messages may be stored for an indefinite time and be read long after being sent; think of blog posts or messages left on a user's "Wall" on Facebook. Note that in some cases the Service Provider stores messages, in others fellow users do (as with most

Instant Messaging applications). The sender of these messages often has little control of how long the messages are stored.

Multi-media. Actual content that can be attached to messages, but may also be uploaded to private or public data-spaces (e.g. Picasa photo album, a blog, Facebook “Wall”) or be attached to a profile. Examples are contents of blog entries (text, photos, video), or the photos, video, music and voice recordings that can be connected to a Myspace or Stickam profile.

Groups. A group is a collection of users, who usually share some common attributes, resources or privileges, for example: similar preferences or backgrounds, a collaborative document, or access to a common virtual space.

Tags. We define tags in the broad sense, as in collaborative filtering systems: descriptive keywords (meta-data) that are attached to content by users (either the uploader or other users). In Facebook terminology, ‘tagging’ refers to the specific case where a user identifies the people depicted in a photo by tags the photo with their names, thus explicitly linking these people to the picture.

Preferences / Ratings / Interests. Many OSNs provide their users with some type of matching or recommendation functionality for either content or peers. In order to provide relevant recommendations, information on a user’s attributes or preferences is required. Often, users are asked to explicitly express their preferences or rate items. The resulting information may be publicly visible (interests on a profile page, ratings for an item shows along with who provided them) or restricted to the Service Provider only. Sometimes, the Service Provider will derive (supplementary) information on users’ preferences and attributes from their behavioral information.

Behavioral information. By this we mean browsing history, profile settings, and any actions undertaken by the user while performing tasks within the OSN. Benevenuto et al. note that this type of meta-data is particularly rich [5]. Information such as preferences, friendships or even attributes such as physical location or demographic data can be inferred from it. Behavioral data is also found in traditional websites, although behavior there is not related to navigating a social network.

As said, not all OSNs involve information from all of the above categories. Which information is contained in a particular OSN mostly depends on its media-richness, the functionality it offers to users, and its business model. Some information is only available to the Service Provider (i.e. the OSNs software or operators), while other information is also available to (a subset of) the OSN users, or even the public at large.

Furthermore, some information is consciously supplied by users through the OSN’s graphical user interface, while other information is implicitly derived by the Service Provider by observing user behavior.

2.6 Summary

People use OSNs for a variety of purposes. In any case, to get the desired functionality (e.g. meeting with friends, attracting an audience, getting advice or recommendations), they will need to provide some personal information to the OSN. The type of user data in question depends on the functionality of the OSN, and its media-richness. Table 1 gives an impression of which data types may typically be expected in different types of OSNs. In tables, ● represents a likely match, • represents a possible match and · an unlikely match.

Table 1. Data types typically found in different types of OSNs.

← OSN types	Data types →	Profiles	Connections	Messages	Multi-media	Tags	Preferences/ratings	Groups	Behavioral information	Login credentials
Connection OSNs	Dating	●	•	●	•	·	●	•	●	●
	Business	●	●	●	•	·	•	●	●	●
	Enforcing real-life relationships	●	●	●	•	·	•	●	●	●
	Socializing	●	●	●	•	·	•	●	●	●
	Chat / instant messaging	•	●	●	•	·	•	●	●	●
Content OSNs	Content sharing	•	•	●	●	●	•	•	●	●
	Resource recommendation	●	·	•	●	●	●	•	●	●
	Advice sharing	●	•	●	•	•	•	●	●	●
	Hobbies / entertainment	•	•	●	•	•	•	●	●	●
	“News” sharing	•	•	●	●	•	•	•	●	●

3 Privacy Concerns in Online Social Networks

Because users need to reveal information to make use of the desired functionality of an OSN, there exists a tradeoff between functionality and user privacy. Making sure the OSN can provide the desired functionality is one thing, but when sharing a wealth of (personal) data, one should also consider what *undesired* results might occur. We have seen examples where data is potentially sensitive (e.g. medical or dating OSNs), and the open nature of online systems makes privacy a definite issue. In this section, we will look into the concept of privacy, its role in OSNs, and potential threats to users’ privacy.

3.1 Definitions Regarding Privacy

The word privacy has many subtly different meanings, each with their own definition. This ranges from “personal privacy” (which involves seclusion and bodily privacy) to “Information Privacy”, around which privacy on the Internet in general revolves.

Kang [29] uses the wording of the Information Infrastructure Task Force (IITF), as cited below:

> Information Privacy is "an individual's claim to control the terms under which personal information–information identifiable to the individual – is acquired, disclosed or used."

This concept of Information Privacy is strongly related to the notion of "Confidentiality", from the field of Information Security, but not to be used interchangeably. Confidentiality is concerned with the secrecy or disclosure of individual pieces of information, while Information Privacy also deals with the individual who is the subject of said information, the effects that disclosure have on this person and his or her control and consent.

When users collaborate in a Web2.0 setting, they generally share a lot of (personal) information. When users upload their data to an OSN, they usually have a *scope* in mind (as a quote from Palen and Dourish' below illustrates). Privacy involves keeping a piece of information in its intended scope. This scope is defined by the size of the audience (breadth), by extent of usage allowed (depth), and duration (lifetime). When a piece of information is moved beyond its intended scope in any of these dimensions (be it accidentally or maliciously), a privacy breach occurs. So, a breach may occur when information is shared with a party for whom it was not intended (disclosure), when information is abused for a different purpose than was intended, or when information is accessed after its intended lifetime. We also see this reflected in data protection laws, such as the Data Protection Act 1998 in the United Kingdom [43], where the use of *personal data* is not regulated in an all-or-nothing fashion, but limitations are imposed on the extent and duration of its use.

Palen and Dourish [42] identify three privacy boundaries with which individuals are struggling:

1. The disclosure boundary (managing the tension between private and public),
2. The identity boundary (managing self representation with specific audience, e.g. one will behave differently when at work than when among friends),
3. The temporal boundary (managing past actions with future expectations; user behavior may change over time).

Something to note at this stage, is that by no means all information that is uploaded to an OSN is considered personal data, and is thus not covered by laws regulating the use of personal data. Personal data is defined in [43] as:

> Personal data means data which relate to a living individual who can be identified –
> (a) from those data, or
> (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the *data controller*

The term Personally Identifiable Information (PII) is related (but not synonymous), and refers to “information that can be used to uniquely identify, contact, or locate a single person or can be used with other sources to uniquely identify a single individual”.

Particularly sensitive personaly information is often regulated by additional laws, such as the Health Insurance Portability and Accountability Act (HIPAA) for medical data, or *sensitive personal data* under Data Protection Act 1998, the definition of which includes:

> The racial or ethnic origin of the data subject, his political opinions, his religious beliefs or other beliefs of a similar nature, [...] his physical or mental health or condition, his sexual life, the commission or alleged commission by him of any offense [...]

However, the majority of data in an OSN does not clearly fall under these categories, and its storage, processing and use is not always strictly regulated.

3.2 Users and Privacy Management

Weiss [53] states that on the traditional Web, privacy is maintained by limiting data collection, hiding users’ identities and restricting access to authorized parties only, while the reality of OSNs is that data and identity become closely linked, and are often visible to large groups of people. It becomes harder for a user to monitor and control his personal information, as more of it becomes available online. Together, this makes managing information and privacy a lot more difficult.

Most OSNs offer their users privacy controls that are simple to use, but coarse, for example enabling users to set their entire profile as public, visible to friends only or private (visible only to the user). With growing demand from users and increased attention to privacy in the media, many OSNs (e.g. Facebook) have started offering their users more (apparent) control, like setting the visibility for individual items, or allowing users to organize their friends into categories. Another risk lies in the other extreme, when interfaces become overly complicated. If users do not understand the settings or find them too cumbersome, they may either set them incorrectly or ignore them and settle for sub-optimal privacy protection.

Gross and Acquisti [24] show in a case study that most users do not change the default privacy settings as provided by the OSN, while sharing a large amount of information on their profile. Tufecki [49] concludes in his case study that privacy-aware users are actually more reluctant to join social networks, but once they do join, they still disclose a lot of information. Another observation is that users’ privacy is regulated mostly through visibility, i.e. the privacy settings of the OSN, rather than through selective uploading. In general users are pre-occupied with the current visibility of their information and do not take into account future change and its implications. It seems that users implicitly trust OSN Service Providers to handle user data in a fair and conscientious way, and continue to do so in the future.

3.3 Service Providers and Trust

Besides difficulties in managing privacy towards other “users” (registered or not), there exists a completely different type of concern, originating from the user’s relationship with the OSN Service Provider. The main difference between users and the Service Provider is the type of information they can access. A user or outsider can generally only view public information. The Service Provider can generally view *all* data in the system, including private uploads, browsing behavior, IP addresses, etc. It is also the Service Provider who decides which data is stored, how long it is kept and how it is used or distributed. The user is also dependent on the Service Provider for tools to protect his privacy. Therefore, trust plays a big role in the relationship between a user and the Service Provider.

On a related note, the rules with regards to ownership and intellectual property of user-uploaded data can be deceptive. Some OSNs (for example Facebook) acquire a license to use such content through their terms-of-use policy. This license gives the OSN free reign to use or sell the data as it sees fit, without worrying about copyrights or other claims by the user. Facebook’s statement of rights and responsibilities [20] states:

> You own all of the content and information you post on Facebook, and you can control how it is shared through your privacy and application settings. In addition:
> For content that is covered by intellectual property rights, like photos and videos (“IP content”), you specifically give us the following permission, subject to your privacy and application settings: you grant us a non-exclusive, transferable, sub-licensable, royalty-free, worldwide license to use any IP content that you post on or in connection with Facebook (“IP License”). This IP License ends when you delete your IP content or your account unless your content has been shared with others, and they have not deleted it. [...]

This becomes unnerving once we realize that the interests of users and Service Provider can clash, especially in OSNs where the main source of revenue is not the users, but third party sales and targeted advertisement.

Finally, we note that many laws, including the Data Protection Act 1998 [43], focus on *“data controllers”*. There are no specific regulations for OSNs and they are currently treated as *“information services”* – online databases of information. The EU article 29 Data Protection Working Party [50] would like to see this changed, so that OSN Service Providers will be treated as “data controllers”, which will obligate them to adhere to laws for processing of user data. This should it easier to guarantee the trustworthiness of OSNs, by forcing them to be more privacy-friendly, ideally without hampering the services they offer to their users.

3.4 User-related Privacy Concerns

In many cases, privacy is breached by fellow OSN users, or unregistered visitors. This may be a deliberate act (snooping, hacking), or accidental (mismanagement of privacy settings by the user himself, lingering data), and can have serious consequences. Let us take a look at different privacy threats that involve disclosure to other users.

Stranger Views Private Information. Users can falsely assume some information to be kept restricted to a certain audience, when in reality it is not. This can be due to design flaws on the part of the OSN Service Provider (e.g. private photos, videos and blogs being easily “hacked” on Myspace [28]), or a lack of understanding or attention to the privacy controls of the user himself. However, even Internet security experts can make mistakes with disclosing information [54]. When a stranger views such private information, user control over who views the information is lost and conflict occurs with the *disclosure boundary*. The above can apply to profiles, connections to fellow users, messages, multi-media, tags, group memberships etc. Rosenblum [46] shows that information in OSNs is far more accessible to a widespread audience than perceived by its owners, and can even end up in the mainstream media [26].

Unable to Hide Information from a Specific Friend or Group. Sometimes one would like to hide some information from a specific friend or a group of friends. Perhaps a user would not like to let a friend know that a surprise party is being planned or hide the pictures of a night out from his parents or employer. In real life, we easily manage the different social contexts that we belong to, but in OSN’s the lines that separate them tend to blur [31]. Not many OSNs provide the option to create groups of friends for different social contexts or hide information on a fine-grained level. This problem is related to Palen and Dourish’ *identity boundary* as users do not have the control to act differently towards one user or group of users, than towards others.

Other Users Posting Information About You. Even if a user is careful in controlling what information she posts to an OSN, she has no control over what other users post in the same OSN. Often, messages contain information about multiple OSN users, or even non-users. This problem is related to the *disclosure boundary*, because information is made more public than intended, in this case by others. It can occur when another user posts information about you which you do not want to be uploaded to the OSN, or when information disclosed privately to another user is made available to a larger audience. This can even be a deliberate act [45].

3.5 Provider-related Privacy Concerns

A completely different type of privacy threat involves the relationship between the user and OSN Service Provider, and in particular the trust that the user puts in the provider. This goes beyond user control of information, because the Service Provider usually designed or configured the systems underlying the OSN. Thus, he has full access to any user-related data, including browsing behavior and message logs. The associated threats are detailed below.

Data Retention Issues. When posting information to an OSN it is often impossible or very difficult to remove that information, for several reasons. On one hand, the Service Provider may intentionally prevent or hinder removal of data. Facebook for example does not provide users with the means to delete their profile, and has actively blocked third-party software that attempts to remedy this [35]. This is because the capital of an OSN often lies in the number of users, and data sales are sometimes part of the revenue. Facebook would like to store content forever [52]. Secondly, information (especially in a social context) tends to be replicated. People may spread information or multi-media and even store it locally and re-upload it at a later time. Finally, information that is apparently erased may still reside elsewhere on the OSN, for example in backups, to be found by others. Similarly, a resource may be disabled or seemingly deleted, but references to it (thumbnails, messages on friends' pages etc.) can remain visible to the outside world. An example of this is given by Bonneau [6] who tracked the availability of deleted photos. These are all violations of the *temporal boundary* as information is available longer than intended.

OSN Employee Browsing Private Information. The OSN Service Provider has full access to the data and its employees might take advantage of this. This is in conflict with the implicit trust that the OSN asks of its users. All information supplied to the OSN is at risk in this issue, up to and including behavioral information. Interviews suggest that some Facebook employees are able to view user information and it is left to the company to police this [39].

Selling of Data. The wealth of information that is stored on the OSN, is likely to be of value to third parties and may be sold by the OSN. User preferences, behavior and friendship connections are all potentially interesting for marketing purposes and research into social dynamics. Data sales can easily be in conflict with the implicit trust the user has in the OSN. Depending on the OSN's business model, it may be in the interest of the OSN to have its users share as much information as possible, to obtain a license or ownership wherever possible, to store it indefinitely and get maximal profits from sales. One example of an OSN that provides user data to third parties is PatientsLikeMe. To quote their website:

> PatientsLikeMe offers pharmaceutical companies the opportunity to reach and learn from thousands of patients with chronic diseases. Follow their symptoms, treatments, outcomes and attitudes. Evaluate real-world safety and efficacy data, and conduct targeted clinical trial recruitment. These are just a few examples of how our portfolio of services drives value at each stage of the drug development process.

Even though data is often anonymized before being sold to protect user privacy, re-identification is a remaining threat that is often overlooked or ignored. Backstrom et al. [3] show how users can be re-identified by looking for unusual points within the friendship graph of the actual OSN, and locating these in the anonymized dataset.

Targeted Marketing. Multiple pieces of information in the OSN can be combined to provide a high value profile of the user. This high value profile can then be used or exploited to present targeted marketing to the user. This again is a conflict of the implicit trust in the OSN, because information is used for a different purpose than intended by the user. An example of a company which uses OSN data for targeted marketing is TrustFuse [38].

> "All of this information could come in handy for Rapleaf's third business, TrustFuse, which sells data (but not e-mail addresses) to marketers so they can better target customers, according to TrustFuse's Web site"

3.6 Summary

Because OSNs contain massive amounts of useful and interesting data about large numbers of users, they form an interesting target for third parties, both private and commercial. Either through browsing/spidering, hacking attacks or legitimate data-sales, this information could end up in the wrong hands. The fact that the users are not always the source of revenue for an OSN (in the case of advertisement revenue and data sales) can lead to conflicting interests for users and Service Providers. Given the diverse and often extensive information available on OSNs, and the fact that threats may come from other users or even the Service Provider itself, the threats are numerous. Table 2 attempts to give a comprehensive overview. In this table concern is high (●), medium (•), or low (·).

Table 2. Privacy concerns for user data in OSNs.

← Privacy concerns	Data types →	Profiles	Connections	Messages	Multi-media	Tags	Preferences/ratings	Groups	Behavioral information	Login credentials
User related	Stranger views private info	●	●	●	●	•	●	●	•	·
	Unable to hide info from specific friend / group	●	●	●	●	•	·	·	·	·
	Other users posting information about you	·	·	●	●	●	·	·	·	·
Provider related	Data retention issues	●	●	●	●	●	●	●	●	·
	OSN employee browsing private info	●	●	●	●	●	●	●	●	•
	Selling of data	●	●	●	●	·	●	·	●	·
	Targeted marketing	●	·	·	·	·	●	·	●	·

4 Existing Research into Privacy-Protecting Technologies

We have seen that there is a wide variety of privacy issues that play a role in OSNs. Because the type of access differs greatly between users and Service Providers, the two main categories of threats require their own specific defense mechanisms. Despite the fact that prevention is no simple matter, research is being conducted in many areas to alleviate some of the aforementioned threats. To protect user data from fellow users, awareness and proper tools for managing and enforcing access policies play a leading role [2, 12, 31]. This does not work towards solving issues that involve untrusted Service Providers. Obscuring and hiding sensitive data from the providers [1, 25, 48], or removing them from the picture entirely [8, 9, 47] are the general approaches here, as we will see. We now proceed to a topical literature overview of research on mitigating privacy issues and tailoring to the privacy needs of users.

4.1 Anonymization

As pointed out in Sections 2 and 3.5, sales of information pertaining to the OSN is often a major source of revenue for the Service Provider. If this were to be done without any further consideration for privacy, users might take offense and leave the network (thus hurting revenue), or take justified legal action. Through *anonymization*, OSN Service Providers may try to remove the privacy issues associated with data sales, by *obscuring the link between users and data sold*.

Basic anonymization involves removing any identifying (or identifiable) information from the data, while preserving other structures of interest in the data. As said however, different re-identification attacks [3] can be mounted to fill in the missing info from the data sold, e.g. by exploiting the topology of the network. Techniques for more thorough anonymization have been proposed, for example mixing of attributes, or modifying the graph structure in such a way that its properties stay mainly intact, while making re-identification hard or impossible [55]. Zhou et al. [56] give a good overview of this field and its problems. Recently the field of anonymization is shifting towards differential privacy [18], which aims to make users in released data computationally indistinguishable from most of the other users in that data set.

Anonymization techniques are usually simple to implement, and need to be performed only once on a given snapshot of the OSN before sales to third parties. The drawback is that it is hard to *formally prove* the security of these methods, as is done in classical cryptography. This mainly stems from the fact that information can only be *partially* removed or obfuscated, while other parts *must be kept intact* for the dataset to remain interesting for after-sale use. Because OSNs are such complex systems, it is nearly impossible to predict which pieces of data can be combined into identifiable information, or which external hints may become available for attackers to exploit. This is what makes it hard to definitively prevent (partial) recovery of the private information that was obscured through anonymization.

4.2 Decentralization

Research on *decentralization* of OSNs revolves around the concept of untrusted OSN Service Providers, and tries to prevent privacy issues where the implicit trust in the OSN is abused. Decentralization can be applied to different degrees. Either some of the power is taken away from the Service Provider, or he is removed from the picture altogether. An example with slight decentralization would be to set up direct links between users when chatting. In this way the chat data never passes through the server. An extreme form of decentralization would remove the OSN altogether and have all traffic take place through peer-to-peer networks. Generally the more decentralized the solution the better the protection from aforementioned privacy issues.

Buchegger and Datta [8] list the possible advantages and challenges for shifting to a fully decentralized OSN. One of the major obstacles is that all users will be made responsible for availability (and integrity) of (one another's) information. Because users cannot be expected to remain online constantly, peers or a trusted proxy should keep data available on a user's behalf. Doing this securely (with untrusted peers), reliably and efficiently poses a big challenge, especially because another of the main challenges in this area of research lies in version control. Given the churn of users and the rate at which data is updated, designing a fully decentralized OSNs is no simple task. Because a decentralized structure works strongly towards taking power away from the OSN Service Provider, it is contrary to the business model of many existing OSNs. This means that these will not be likely to adopt such a structure, or aid its development.

Some creative proposals have been made with the aim to overcome these challenges. Tootoonchian et al. [48] propose to decouple social relationships from the user data. User data (which they call "social data") will still reside on the OSN's server, but the relationships (the actual graph) will be in the form of attestations. An attestation can prove to a Service Provider that two users have a social relationship. These attestation can then be used for access control, granting the user access to the proper resources on any such OSN without requiring the user to sign up for every social network.

Freedman and Nicolosi [21] propose a method for verifying social proximity (friend of a friend) and give the list of bridging friends to one of the parties. In this scheme one of the parties looks forward, while the other looks backwards. With both using a one-way transform, one party compares these relationships. In this directional scheme, the party that is the target of the friend relationship has to consent. This party also has to give up more of his privacy, he sends out a key and receives an attestation. Considering that this party is not the initiator of the relationship this is a skewed trade-off.

Mezzour et al. [37] propose another method which works for longer paths. This method works by using a token flooding phase in which tokens are spread throughout the social graph. The user whom first sent these tokens can use a look-up to discover the path between him and the other user. However, revoking any of the relationships in the flooded path would require another flooding phase.

4.3 Privacy Settings and Management

The research in this field is devoted to finding methods to either give the user more control over their privacy settings, or make it easier for the user to manage such settings. In doing so, research in this area hopes to mitigate such problems as unauthorized data access by other users and the inability of users to hide information from a specific friend or group.

Some propose forms of automated assistance to set defaults or adjust privacy settings. Baatarjav et al. [2] propose a system that selects privacy settings according to some basic profile information. This profile information is used to predict a set of expected user preferences, based on statistical data. For example, if most single elderly ladies adopt a certain set of settings, this will be the default for new users in this demographic.

A similar approach is suggested by Maximilien et al. [36], where a privacy score based on the sensitivity and visibility of profile items is computed. This privacy score can then be compared among peers, and the privacy settings of peers can be mimicked if needed.

Goecks et al. [23] have created an overview of the challenges and problems of configuring privacy settings based on this type of collaboration. Most notable is *information cascade*, which is a snowball effect that can lead to the adoption of a certain set of privacy settings by many users. Because this process can also increase the score of an unwanted configuration, this herding behavior can eventually lead all users to share the same unwanted setting. In an extension of their system, Goecks et al. add an “expert set” of advanced users that has higher priority over regular users.

A different suggestion comes from Banks and Wu [4], where an interaction history facilitates privacy settings between users, using trust as a currency. This proposal has not been worked out in detail.

Another interesting research topic is the development of solutions to make information disclosure and privacy settings more gradual,fine-grained and transparent. The central question here is how to design appropriate tools for such fine-grained control, without overburdening the users or the system.

Privacy awareness among users can be enhanced by showing the user the consequences of his actions. According to Lipford et al. [32] this can be done by showing the user their profile as seen by others. Onwuasonanya et al. [41] study the behavior of users when given the ability to define subgroups among their online friends. An existing system that combines both of these features (and other privacy tools) is Clique (clique.primelife.eu), part of the Primelife project. This experimental OSN allows its users to create multiple “faces” to use in different social contexts. Each face has its own collections of contacts (e.g. friends, colleagues and family) and each piece of information can be made visible to any combination of people [31]. Users can check if the desired result is achieved by viewing their profile from the perspective of other users.

This type of solution is often comparatively cheap to implement, and mainly depend on OSN Service Providers making the right design choices. However, they require user awareness and acceptance in order to reach their full potential. Also, data collection and retention are key to many OSNs' revenue, so acceptance by Service Providers may be an even bigger issue.

4.4 Encryption

Encryption can be used as a tool to provide confidentiality and as the basis for integrity. Depending on how encryption is applied this can mean protection from unauthorized users or the Service Provider. It is often used as a building block in other proposals, for example in decentralized systems or in privacy settings and management tools.

Lucas and Borisov [33] propose to encrypt certain parts of a user's profile using public key cryptography. Keys are derived from user-supplied passwords in order to maintain access flexibility. A potential problem with this approach is the resulting low entropy of the keys.

Guha et al. [25] argue that encrypted messages on a profile are easy to spot by the Service Provider, which could result in sanctions if the provider disapproves of encryption. Their aim is akin to *steganography*, in that the Service Provider should not even be aware that information hiding is being applied. Their approach uses substitution of "information atoms" (e.g. age, or name-and-gender) to hide information. Keys and related material are exchanged out of band. The number of channels that are used for this scheme is high. Also, users that do not employ this system have no way to distinguish users that are hiding their information from users that are not. This makes profiles meaningless to such users, and could lead to cases of mistaken identity.

The advantage of cryptographic approaches is that they can solve many issues, if used properly. Through cryptography, one can protect data from other users, as well as the OSN. In addition, the security of such techniques can often be proven or verified comparatively easily. However, key management is a big issue, especially if there is a high amount of churn in friendship relations and group membership. Also, cryptography often involves expensive operations, for the user, the Service Provider, or both. Especially in large OSNs, scalability can easily become an obstacle.

4.5 Awareness, Law and Regulations

Research in this mainly non-technical field aims to enhance user awareness of the privacy issues that exist within OSNs, as well as compliance of both users and Service Providers to established laws and social conducts. It can aid users in specifying and respecting privacy boundaries, and may alleviate issues like snooping and “other users posting information about you”.

Kang and Kagal [30] propose a social mechanism to prevent data abuse by showing on a profile what is acceptable to do with the data and what is not. This relies on proper social conduct and no further technical support is provided to enforce it.

Onwuasoanya et el. [41] want to make it a requirement for the user to group his friends and consequently be able to set different privacy settings for each group. The aim is to provide users a simple and intuitive way to manage their privacy settings, thus increasing user awareness and control.

The system that Goecks et al. [23] propose, uses social collaboration to make it easier for users to set their privacy settings and make them more aware if their choice is different from the norm.

The Platform for Privacy Preferences (P3P) [16] is an initiative that aims to provide websites with a standardized format in which they can define their privacy policy. Visitors of the website can then, through client-side “user agents” (e.g. plug-ins for their webbrowser or applets), easily check the details of a privacy policy and see what will happen to data they submit. This system can help to increase user awareness, but only for users that employ agents and if websites properly define their privacy policies and adhere to them.

These non-technical approaches lack the power to actively enforce the changes they propose. Policies and regulations are not mandatory, and awareness is something that needs time to be raised. Specific laws dealing with personal information as related to the Internet and OSNs form an important and much needed tool, but often take long to be developed. Also, laws are generally used to solve matters *after* things go wrong, whereas most technical solutions attempt to *prevent* violations.

4.6 Summary

Table 3 shows which research discipline can contribute to address which privacy concern: a ● indicates that the technique is helpful to address a particular concern, while a · states that the technique does not seem applicable. None of the disciplines mentioned in this section offer complete privacy for OSN users. Because the issue of privacy is multi-faceted, it will require a multi-faceted solution. Several techniques will likely need to be combined to develop proper technical solutions to tackle the various privacy issues. In addition, Service Providers should be encouraged or required to implement such solutions, and users need to be made aware of the benefits of using them.

5 Conclusion

We have seen that OSNs are used by millions for a wide variety of purposes. In our classification, we chose to make the broadest distinction between *content* OSNs and

Table 3. Privacy concerns and relevant defenses.

← Privacy concerns	Relevant defenses →	Anonymization	Decentralization	Privacy settings and management	Encryption	Awareness, law and regulations
User related	Stranger views private info	●	·	●	●	●
	Unable to hide info from specific friend / group	·	·	●	●	·
	Other users posting information about you	·	·	●	·	●
Provider related	Data retention issues	●	●	·	●	●
	OSN employee browsing private info	·	●	·	●	●
	Selling of data	●	●	·	●	●
	Targeted marketing	●	●	·	●	●

connection OSNs. Users generally either look to share media and information, or simply socialize. The purpose of an OSN and its media-richness dictate which types of user data reside in the network. For privacy in turn, this implies which data may be at risk, and in what ways.

Because OSNs are such complex systems, the privacy issues are myriad. The fact that trust in the OSNs Service Provider is not always justified further complicates matters. Users are expected to protect their privacy with tools that are designed by a party that does not by definition have the same goals with regards to privacy, and even if they are protected from other users, the Service Provider could abuse his position of power.

Many areas of research can help to protect user privacy, ranging from technical (e.g. system design and cryptography) to non-technical (e.g. sociology and law). However, privacy protection should ideally be built into the system, without harming its operation by overburdening users or Service Provider, or hampering the OSNs functionalities. Also, we must realize that “The Privacy Problem” – if one could even formulate it as a single problem – will not be solved by any single research discipline. Our conclusion is, that in order to develop a full solution to protect consumer privacy, the strengths of several research areas will need to be brought together. Only thus can users be educated and empowered, will OSN Service Providers be forced to comply, and are legal steps possible when prevention fails.

Acknowledgments
The research for this work was carried out within the Kindred Spirits project, part of the STW Sentinels research program.

References

1. Jonathan Anderson, Claudia Daz, Joseph Bonneau, and Frank Stajano. Privacy-enabling social networking over untrusted networks. In Jon Crowcroft and Balachander Krishnamurthy, editors, *WOSN*, pages 1–6. ACM, 2009.
2. Enkh-Amgalan Baatarjav, Ram Dantu, and Santi Phithakkitnukoon. Privacy management for facebook. In R. Sekar and Arun K. Pujari, editors, *International Conference on Information Systems Security*, volume 5352 of *Lecture Notes in Computer Science*, pages 273–286. Springer, 2008.
3. Lars Backstrom, Cynthia Dwork, and Jon Kleinberg. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 181–190, New York, NY, USA, 2007. ACM Press.
4. Lerone Banks and Shyhtsun Felix Wu. All friends are not created equal: An interaction intensity based approach to privacy in online social networks. In *IEEE International Conference on Computational Science and Engineering*, pages 970–974, 2009.
5. Fabrcio Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Virglio A. F. Almeida. Characterizing user behavior in online social networks. In Anja Feldmann and Laurent Mathy, editors, *Internet Measurement Conference*, pages 49–62. ACM, 2009.
6. Joseph Bonneau. Attack of the zombie photos. online, 2009. `http://www.lightbluetouchpaper.org/2009/05/20/attack-of-the-zombie-photos/`.
7. Danah Boyd and Nicole B. Ellison. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2007.
8. Sonja Buchegger and Anwitaman Datta. A case for p2p infrastructure for social networks - opportunities and challenges. In *WONS 2009, 6th International Conference on Wireless On-demand Network Systems and Services*, pages 161–168, Snowbird, Utah, USA, February 2009.
9. Sonja Buchegger, Doris Schiöberg, Le H. Vu, and Anwitaman Datta. Peerson: P2p social networking: early experiences and insights. In *SNS '09: Proceedings of the Second ACM EuroSys Workshop on Social Network Systems*, pages 46–52, New York, NY, USA, 2009. ACM.
10. Enid Burns. Marketing to social networking sites, targeted. online, 4 2007. `http://www.clickz.com/3625536`.
11. Dominique Cardon. Le design de la visibilit : un essai de typologie du web 2.0. online, 2 2008. `http://www.internetactu.net/2008/02/01/le-design-de-la-visibilite-un-essai-de-typologie-du-web-20/`.
12. Barbara Carminati, Elena Ferrari, and Andrea Perego. Private relationships in social networks. In *ICDE Workshops*, pages 163–171, 2007.
13. Anonymous contributor on Wikipedia.org. Cake financial. online, 4 2010. `http://en.wikipedia.org/wiki/Cake\_Financial`.
14. Anonymous contributor on Wikipedia.org. Sixdegrees.com. online, 4 2010. `http://en.wikipedia.org/wiki/SixDegrees.com/`.
15. Anonymous contributor on Wikipedia.org. Social network service. online, 4 2010. `http://en.wikipedia.org/wiki/Social\_network\_service`.

16. Lorrie Cranor, Marc Langheinrich, Massimo Marchiori, Martin Presler-Marshall, and Joseph Reagle. The platform for privacy preferences 1.0 (p3p1.0) specification. online. http://www.w3.org/TR/P3P/.
17. Ryan Dube and Mary Beth P. Adomaitis. What types of social networks exist. online, 3 2009. http://socialnetworking.lovetoknow.com/What_Types_of_Social_Networks_Exist.
18. Cynthia Dwork. Differential privacy. In *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II*, pages 1–12, 2006.
19. Dave Emmett. Taxonomy of social networks. online, 6 2009. http://davemmett.wordpress.com/2009/06/15/taxonomy-of-social-networks/.
20. Facebook.com. Statement of rights and responsibilities. online. http://www.facebook.com/terms.php.
21. Michael J. Freedman and Antonio Nicolosi. Efficient private techniques for verifying social proximity. In *Proceedings of the 6th International Workshop on Peer-to-Peer Systems (IPTPS07)*, pages 1–7, Bellevue, WA, February 2007.
22. Liz Gannes. A taxonomy of social networks? online, 2 2007. http://gigaom.com/2007/02/09/social-network-taxonomy/.
23. Jeremy Goecks, W. Keith Edwards, and Elizabeth D. Mynatt. Challenges in supporting end-user privacy and security management with social navigation. In Lorrie Faith Cranor, editor, *Symposium on Usable Privacy and Security*, ACM International Conference Proceeding Series, pages 1–12. ACM, 2009.
24. Ralph Gross and Alessandro Acquisti. Information revelation and privacy in online social networks. In *WPES '05: Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80, New York, NY, USA, 2005. ACM.
25. Saikat Guha, Kevin Tang, and Paul Francis. Noyb: privacy in online social networks. In *Proceedings of the first workshop on Online Social Networks (WOSP)*, pages 49–54, New York, NY, USA, 2008. ACM.
26. Nelson Hernandez. President apologizes for questionable photos, 10 2007. http://www.washingtonpost.com/wp-dyn/content/article/2007/10/17/AR2007101702244.html.
27. Hudsonhorizons.com. Types of social networking websites. online, 2010. http://www.hudsonhorizons.com/Custom-Website-Solutions/Social-Networking/Types-of-Social-Networks.htm.
28. Arpit Jacob. How to hack myspace private profile picture and video. online, 4 2007. http://www.clazh.com/how-to-hack-myspace-private-profile-picture-and-video/.
29. Jerry Kang. Information privacy in cyberspace transactions. *Stanford Law Review*, 50(4):1193–1294, 1998.
30. Ted Kang and Lalana Kagal. Establishing social norms for privacy in social networks. unpublished, 2009.
31. Ronals Leenes. *Context Is Everything - Sociality and Privacy in Online Social Network Sites*, volume 320/2010, chapter 4, pages 48–65. Springer Boston, 2010.
32. Heather Richter Lipford, Andrew Besmer, and Jason Watson. Understanding privacy settings in facebook with an audience view. In *UPSEC'08: Proceedings of the 1st Conference on Usability, Psychology, and Security*, pages 1–8, Berkeley, CA, USA, 2008. USENIX Association.
33. Matthew M. Lucas and Nikita Borisov. Flybynight: mitigating the privacy risks of social networking. In *Proceedings of the 7th ACM workshop on Privacy in the electronic society (WPES)*, pages 1–8, New York, NY, USA, 2008. ACM.

34. Bernard Lunn. Social network types, motivations, and the future. online, 9 2007. http://www.readwriteweb.com/archives/social_network_types_motivations.php.
35. Paul MacNamara. Facebook blocks 'web 2.0 suicide machine'. online, 1 2010. http://www.networkworld.com/news/2010/010410-buzzblog-facebook-blocks-suicide-machine.html.
36. E. M. Maximilien, T. Grandison, Kun Liu, T. Sun, D. Richardson, and S. Guo. Enabling privacy as a fundamental construct for social networks. In *Proc. International Conference on Computational Science and Engineering CSE '09*, volume 4, pages 1015–1020, August 29–31, 2009.
37. Ghita Mezzour, Adrian Perrig, Virgil D. Gligor, and Panos Papadimitratos. Privacy-preserving relationship path discovery in social networks. In Juan A. Garay, Atsuko Miyaji, and Akira Otsuka, editors, *Cryptology and Network Security*, volume 5888 of *Lecture Notes in Computer Science*, pages 189–208. Springer, 2009.
38. Stefanie Olsen. At rapleaf, your personals are public. online, 8 2007. http://news.cnet.com/At-Rapleaf,-your-personals-are-public/2100-1038_3-6205716.html.
39. Nick O'Neill. ”anonymous” facebook employee interview: Fact vs fiction, 1 2010. http://www.allfacebook.com/2010/01/anonymous-facebook-employee-interview-fact-vs-fiction/.
40. Onlinebrandmanager.com. Types of online social networks. online. http://onlinebrandmanager.org/social-media/social-network-types/.
41. Anthony Onwuasoanya, Maxim Skornyakov, and Jonathan Post. Enhancing privacy on social networks by segregating different social spheres. *Rutgers Governor's School of Engineering and Technology Research Journal*, 3:1–10, 2008.
42. Leysia Palen and Paul Dourish. Unpacking ”privacy” for a networked world. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 129–136, New York, NY, USA, 2003. ACM.
43. UK Parliament. Data protection act 1998, 1998. http://www.legislation.gov.uk/ukpga/1998/29/contents.
44. Pingdom.com. Social network popularity around the world. online. http://royal.pingdom.com/2008/08/12/social-network-popularity-around-the-world/.
45. Warren Riddle. Cyberbullied teen sues ex-classmates, their parents, and facebook, 3 2009. http://www.switched.com/2009/03/04/cyberbullied-teen-sues-ex-classmates-their-parents-and-faceboo/.
46. D. Rosenblum. What anyone can know: The privacy risks of social networking sites. *IEEE Security & Privacy*, 5(3):40–49, May 2007.
47. Amre Shakimov, Alexander Varshavsky, Landon P. Cox, and Ramn Cceres. Privacy, cost, and availability tradeoffs in decentralized osns. In Jon Crowcroft and Balachander Krishnamurthy, editors, *WOSN*, pages 13–18. ACM, 2009.
48. Amin Tootoonchian, Stefan Saroiu, Yashar Ganjali, and Alec Wolman. Lockr: better privacy for social networks. In *CoNEXT '09: Proceedings of the 5th international conference on Emerging networking experiments and technologies*, pages 169–180, New York, NY, USA, December 2009. ACM.
49. Zeynep Tufekci. Can you see me now? audience and disclosure regulation in online social network sites. *Bulletin of Science Technology Society*, 28(1):20–36, February 2008.
50. Alex Turk. Opinion 5/2009 on online social networking. Technical Report 01189/09/EN WP 163, Article 29 Data Protection Working Party, 6 2009. http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2009/wp163_en.pdf.

51. Bimal Viswanath, Alan Mislove, Meeyoung Cha, and P. Krishna Gummadi. On the evolution of user interaction in facebook. In Jon Crowcroft and Balachander Krishnamurthy, editors, *Workshop on Online Social Networks*, pages 37–42. ACM, 2009.
52. Chris Walters. Facebook's new terms of service: ”we can do anything we want with your content. forever.”, 2 2009. http://consumerist.com/2009/02/facebooks-new-terms-of-service-we-can-do-anything-we-want-with-your-content-forever.html.
53. Stefan Weiss. The need for a paradigm shift in addressing privacy risks in social networking applications. In *The Future of Identity in the Information Society*, volume 262, pages 161–171. IFIP International Federation for Information Processing, 2008.
54. David M Williams. Online identity expert loses control of nsfw r-rated online pics, 3 2009. http://www.itwire.com/your-it-news/home-it/23975-online-identity-expert-loses-control-of-nsfw-r-rated-online-pics.html.
55. Xiaowei Ying and Xintao Wu. Randomizing social networks: a spectrum preserving approach. In *Proceedings of the SIAM International Conference on Data Mining*, pages 739–750. Society for Industrial and Applied Mathematics, 2008.
56. Bin Zhou, Jian Pei, and WoShun Luk. A brief survey on anonymization techniques for privacy preserving publishing of social network data. *Special Interest Group on Knowledge Discovery and Data Mining Explorations*, 10(2):12–22, 2008.