The Privacy Jungle:
On the Market for Data Protection in Social Networks

Joseph Bonneau
Computer Laboratory
University of Cambridge
jcb82@cl.cam.ac.uk

Sören Preibusch
Computer Laboratory
University of Cambridge
sdp36@cl.cam.ac.uk

Abstract

We have conducted the first thorough analysis of the market for privacy practices and policies in online social networks. From an evaluation of 45 social networking sites using 260 criteria we find that many popular assumptions regarding privacy and social networking need to be revisited when considering the entire ecosystem instead of only a handful of well-known sites. Contrary to the common perception of an oligopolistic market, we find evidence of vigorous competition for new users. Despite observing many poor security practices, there is evidence that social network providers are making efforts to implement privacy enhancing technologies with substantial diversity in the amount of privacy control offered. However, privacy is rarely used as a selling point, even then only as auxiliary, non-decisive feature. Sites also failed to promote their existing privacy controls within the site. We similarly found great diversity in the length and content of formal privacy policies, but found an opposite promotional trend: though almost all policies are not accessible to ordinary users due to obfuscating legal jargon, they conspicuously vaunt the sites' privacy practices. We conclude that the market for privacy in social networks is dysfunctional in that there is significant variation in sites' privacy controls, data collection requirements, and legal privacy policies, but this is not effectively conveyed to users. Our empirical findings motivate us to introduce the novel model of a privacy communication game, where the economically rational choice for a site operator is to make privacy control available to evade criticism from privacy fundamentalists, while hiding the privacy control interface and privacy policy to maximise sign-up numbers and encourage data sharing from the pragmatic majority of users.

Contents

1 Introduction **4**

2 Related Work **5**

3 Survey Methodology **5**
3.1 Selection of Sites 5
3.1.1 General-Purpose Sites 5
3.1.2 Niche Sites 6
3.2 Evaluation Methodology 7
3.2.1 Data Collection 7
3.2.2 Data Provided During Signup 7
3.2.3 Technical Set-up 9

4 Data **9**
4.1 Market Dynamics 9
4.1.1 Network Size 9
4.1.2 Site Popularity: Traffic Data 9
4.1.3 Geographical Distribution: American Dominance 10
4.1.4 Site Evolution 10
4.1.5 Multilingualism 11
4.1.6 Competition 11
4.1.7 Business Model 11
4.2 Promotional Methods 11
4.2.1 Promotion of Social Interaction 12
4.2.2 Promotion via Network Effects 12
4.2.3 Promotion of Functionality 13
4.2.4 Promotion of Privacy 13
4.3 Presentation of Terms of Use and Privacy Policy 14
4.3.1 Privacy Policy Acknowledgement 14
4.3.2 Privacy Policy Review 15
4.4 Data Collected During Sign-up 15
4.4.1 Over-Collection of Demographic Data 15
4.4.2 Requirement of Real Names 16
4.4.3 Requirement of Email Addresses 16
4.5 Privacy Controls 17
4.5.1 Profile Visibility Options 17
4.5.2 Fine-Grained Controls 18
4.5.3 Permissive Defaults 18
4.5.4 User Interface Problems 19
4.6 Security Measures 20
4.6.1 Use of TLS Encryption and Authentication 20
4.6.2 Phishing Prevention 20
4.6.3 Online Safety Guidance & Abuse Reporting 21
4.7 Privacy Policies 21
4.7.1 Technical Accessibility 23
4.7.2 Length 23
4.7.3 Legal Issues 23
4.7.4 Data Claims 24

4.7.5 Availability of P3P Policies 24
4.7.6 Self-Promotion within Privacy Policies 24

5 Data Analysis 25
5.1 Privacy vs. Functionality 25
5.2 Privacy vs. Site Age 27
5.3 Privacy vs. Size 27
5.4 Privacy vs. Growth Rate 28
5.5 Privacy Promotion and Claims vs. Actual Privacy Practices 28

6 Economic Models 28
6.1 The Privacy Communication Game 29
6.1.1 Reducing Privacy Salience 29
6.1.2 Discouraging Privacy Fundamentalists 30
6.1.3 Reducing Privacy Criticism 30
6.1.4 Evolution of Communication 31
6.2 The Effects of Lock-in 32
6.3 Privacy as a Lemons Market 32
6.4 Privacy Negotiations 33

7 Limitations 34

8 Conclusions 34

References 36

A Imputed Metrics 42
A.1 Privacy Score 42
A.1.1 Data Collection Subscore 42
A.1.2 Privacy Control Subscore 42
A.1.3 Privacy Policy Subscore 42
A.2 Functionality Score 42

1 Introduction

In the past decade, social networking sites have become a mainstream cultural phenomenon [27]. Social networking has become one of the most popular activities on the web, with the top sites boasting hundreds of millions of users, and social networking sites representing 16 of the world's 100 most-visited sites [1]. Their popularity amongst the younger generation is even higher, with studies finding more than 80% of American university students active social network users [8, 48], commonly spending at least 30 minutes every day on social networks [51]. The ubiquity of social networking in youth culture has been likened to an addiction [30].

Social networks have also obtained a poor reputation for protecting users' privacy due to a continual flow of media stories discussing privacy problems [44]. Popular media angles include the disclosure of embarrassing personal information to employers [37, 19] and universities [76], blackmail using photos found online [68, 72], social scams [49, 39, 12], and user backlash against newly introduced features [75, 80].

Despite the focus of the English-speaking media on Facebook, MySpace, and occasionally Bebo, and the common perception of an oligopolistic market, there is a flourishing supply of social networking services, with dozens of large general-purpose sites competing alongside thousands of niche sites. In our study, at least 25 different services were found to be the most popular social network in at least one country [1].

There is also a common misconception that privacy violations occur routinely because the generation of (mostly younger) social networking users fundamentally do not care about privacy. This is contradicted by studies where most social network users do express an interest in privacy [8, 31, 26, 42]. Given the plethora of competing sites, the functional similarity of most social networks, and users' stated concern for privacy, market conditions appear prime for sites to compete on the basis of privacy. This was our overarching research question as we conducted—to the best of our knowledge—the largest and most comprehensive field study in the academic literature of the global social network market. Past studies have focused on studying users of social networks; our study is unique in that we compare the sites themselves. We have attempted to collect as large a sample of sites as possible, focusing on what is offered and promoted in the way of privacy rather than on user behaviour.

Our contribution is threefold. First, we report the results of a thorough analysis of the privacy supply in the social networking market (Section 4). Our data supports some common assumptions, such as a generally low quality of privacy policies, usability problems, and poor security practices. It also provides some surprises such as promotion of photo-sharing being far more common than game-playing, and a huge diversity of privacy controls available in different networks which is not effectively conveyed to users.

Second, we aggregate our data into overall privacy and functionality scores for each site, and use these to find which general factors may influence a site's privacy practices (Section 5). Again, we find interesting results, such as niche sites offering significantly less sophisticated privacy controls than general-purpose sites, positive correlations between privacy and the age, size, and popularity of a site. Privacy and functionality aren't strong correlated, but sites that promote on privacy are often found having less favourable privacy practices. We also find evidence that sites with better privacy are growing ahead of the market, while those that mention their privacy are falling behind.

Finally, we propose a novel economic model to explain the observed under-supply and under-promotion of privacy as a rational choice by the competing social networking providers. Our model assumes the existence of consumers with varying degrees of privacy concern. We conjecture that websites seek to maximise their desirability to both populations by not raising privacy concerns for the majority of users, while minimising criticism from the privacy-sensitive. We explore this, along with other economic explanations, in Section 6.

2 Related Work

Given the broad aims of our study, there is a large body of relevant prior research. Social networks have been an active research area in several academic disciplines in recent years. Sociologists have studied them from an ethnographic perspective, examining why they have become popular and what motivates individuals to participate [27, 26, 29, 30, 85]. Others have used surveys to examine users' attitudes towards social networks, in particular with regards to information sharing and disclosure [8, 48, 31, 42]. User studies have also been performed by automatically analysing crawled profiles [54, 8, 51, 48]. Computer scientists have performed more quantitative studies of social graph formation, using web crawlers to study the size and link structure of social graphs [62, 24, 53, 41].

Security and data protection in social networks has recently become an active research area recently. Many researchers have outlined the potential threats and risks associated with using social networking services [17, 74]. White-hat studies have identified many security flaws due to implementation errors in social networks [34, 21, 20]. Security researchers have also taken a constructionist approach. Several interfaces have been proposed for allowing users to more easily manage privacy [56, 66, 11, 77], a few of which we saw beginning to be deployed in sites we analysed. Some have proposed new architectures which can provide stronger privacy guarantees [36, 35, 23, 10, 71], while others have recommended implementing privacy-preserving front-ends for existing social networks [43, 59].

Privacy issues specifically arising from the graph of friendship links have been studied as well, and have identified graph privacy as a major issue. Social context was shown to make phishing attacks much more successful [46]. Several studies have indicated that knowledge of the social graph can enable accurate inference of private data [90, 55, 91]. It has been shown that it is impossible in practice to "anonymise" a social graph by removing names due to the amount of unique structure contained within it [15, 64, 38]. Social graph privacy has also been shown to be very fragile in that obtaining a relatively small amount of information enables many useful calculations to be made [28, 52, 22, 21, 63].

Privacy has also been extensively studied from an economics perspective. The problem has been formally modelled as an economics trade-off between disclosing information and gaining access to desirable online services [82, 70, 45]. Researchers have utilised surveys to gauge user attitudes about privacy, consistently showing a high stated concern for privacy [25, 78, 61]. Direct observational studies of users have often contradicted these studies though, showing that users often violate their stated privacy concerns [6, 79, 40, 69]. Economists have attempted to resolve this "privacy paradox" by proposing models which describe why users' long-term preferences for privacy are ignored or forgotten when interacting with a website [83, 57, 7, 9]. This has been shown specifically to occur in the case of social networks, where individuals with high self-reported privacy awareness revealed significant amounts of data on their profiles [8]. Other research has focused on privacy policies, usually finding them to be far too difficult for ordinary users to understand [61, 86]. Computer scientists have proposed technological solutions to improve users' ability to make privacy choices [73, 5, 33, 18, 16].

3 Survey Methodology

3.1 Selection of Sites

We selected 45 social networking sites for our survey, the complete list is provided in Table 1. Our goal was both to conduct an exhaustive survey of the major, general-purpose social networking sites, and include several representatives of other common social-networking niches for comparison.

3.1.1 General-Purpose Sites

Our operational definition of a *general-purpose* social networking service is one which anybody is free to join, people commonly present their real-world identity, and the primary use of the site is interacting with others via profile pages on the Web. This excludes sites whose primary purpose is sharing content (e.g. YouTube, Flickr), sites which enforce limited membership (invitation-only networks such as A Small World), or sites where few users reveal any real-world information about themselves (such as online poker websites). While some of these services contain all or almost all features of general-purpose social networking sites, they can be separated by their different patterns of typical use. For example, a web crawl revealed that average users of YouTube and Flickr make less than 10% as many connections as those using Orkut [62].

Our definition is mostly functional, and does not exclude sites which are mainly populated by specific demographics of users. Several of the sites we regarded as general-purpose target a specific demographic niche. For example, BlackPlanet is targeted to African Americans living in the United States, Eons is targeted at the older generation, and MyYearbook and Bahu are targeted specifically at teenagers. MocoSpace is a general-purpose social network on the web which additionally aims specifically to be usable via mobile devices. However, we still regard these sites as general-purpose as their feature set is similar to other general-purpose sites, they simply cater to a specific group of people with their marketing and graphic design.

An important omission from our study is sites which are not available in English. This includes several very large general-purpose sites, such as the Russian site VKontakte, the Japanese site Mixi, and the Spanish site Tuenti. This decision was necessary to ensure fair comparison between sites, particularly for privacy policies where word choice is critical. Our focus on the Web also excludes communication services such as Instant Messaging, online role-playing games such as World of Warcraft, and 3D virtual worlds such as SecondLife.

Within this definition, though, we were able to include 29 popular general-purpose sites from around the world, listed in full in Table 1. We enforced a minimum size of 500,000 users for general-purpose sites to keep the study tractable.

3.1.2 Niche Sites

In addition to general-purpose social networks, we examined 16 *niche* social networking services, also listed in full in Table 1. These sites either have a subset of general-purpose sites' functionality or are used in significantly different ways.

- **Business-networking sites** differ from general-purpose in that they specialise in maintaining professional contacts and searching for new jobs. Users typically share much less personal information, yet more professional information on these sites. They often implement specific features for specifying and managing business relationships and are frequently used for job-searching. We included LinkedIn, XING, and Viadeo, the most popular business-networking sites.

- **Media recommendation sites** specialise in allowing users to recommend and share films and music. While they have many features of general-purpose sites, users often interact with others based on similar tastes in music or movies, rather than real-world social connections. We included Last.fm, Imeem, Flickster, and Buzznet in this category.

- **Reunion sites** specialise in allowing people to search for old acquaintances from school or the military rather than actively maintaining profiles. They often aggregate contact information only and are designed to facilitate off-line connection rather than on-line interaction. We included Classmates.com and myLife (formerly Reunion.com) as representatives of this genre.

- **Activity-focused sites** centre around allowing users to perform a specific activity. Habbo and Gaia Online are two pre-eminent gaming-centric social networks. CouchSurfing is designed for students and youth to share accommodation while travelling.[1] Finally, we included the surging micro-blogging service Twitter in this category, though arguably it is in a niche by itself.

— **Privacy-specific sites** have specific privacy-enabling features. Experience Project is designed as a pseudonymous social network for users to share intimate stories with strangers who have had similar life experiences. Imbee is a fledgling social network aimed to be safe for younger children, with strong administrative oversight and parental controls. Kaioo is a non-profit social network designed to be community-owned and governed, for those uncomfortable trusting their social details to a private company. We included Imbee and Kaioo in our survey due to their unique privacy goals, though neither site has an established user base yet.

3.2 Evaluation Methodology

We conducted a standardised, scripted evaluation for each website. The evaluations were conducted in February 2009, and all data is accurate as of the time of evaluation. Due to the rapid evolution of social networking, several data points had already changed by the time of analysis, but we kept all values as a consistent snapshot of the time of collection, recorded alongside the data itself.

3.2.1 Data Collection

First, we collected general information about the site, such as its launch date, estimated user count and traffic ranks, country of operation, and ownership status (presented in Section 4.1). Next, we examined the publicly viewable sections of the webpage which are presented to non-members who visit the site (typically after receiving an invitation by friends who are already members of the site). These offer the most valuable insight into the marketing strategies used by social networks, since very few rely on traditional advertisements. We recorded the selling points used to encourage visitors to sign up (Section 4.2).

Next, we signed up for each site, recording the amount of personal information required in order to register an account (Section 4.4). We also recorded the means by which users are presented with the sites' Terms of Use and/or Privacy Policy during sign-up (Section 4.3). We then evaluated the extent of privacy controls available to users of the site, and the default values provided with a new account (Section 4.5). In addition to privacy controls, we recorded general security features like the use of encryption, the existence of help pages for controlling privacy, and the existence of infrastructure for reporting abuse (Section 4.6).

Finally, we evaluated the formal privacy policy provided by each site (Section 4.7). Evaluation criteria for the privacy policies included accessibility, length, collection and retention of user data, the role of third-party advertisers, and compliance with privacy laws.

In addition to the raw data points, we computed and included in our dataset aggregate metrics per site. In particular, we define scores for data collection, privacy control, privacy policies, and functionality, presented in Table 7. These are described in detail in Appendix A.

3.2.2 Data Provided During Signup

To ensure fair comparison, we supplied consistent data when asked to the fullest extent possible, and consistently withheld any information which was not mandatory. We signed up for an account with each site using the name "Upton Sinclair,"[2] a birth date of September 20, 1978, the Cambridge postcode `CB30DS`, and other standardised personal information consistent in all created accounts. We provided the same Yahoo! email account with a *ymail.com* suffix to each site. We only varied this information in

[1] Because its intended use is connecting strangers, CouchSurfing is notable for having a complicated reputation system built into the site to encourage safety.

[2] In honour of the pioneering investigatory journalist.

Site	Traffic Rank	Users (M)	Country	Category
Windows Live Spaces	4	120	USA	General-purpose
Facebook	5	175	USA	General-purpose
MySpace	7	250	USA	General-purpose
hi5	17	60	USA	General-purpose
SkyRock	43	13	France	General-purpose
Friendster	45	95	USA	General-purpose
NetLog	71	35	Belgium	General-purpose
Tagged	75	70	USA	General-purpose
Orkut	83	67	USA	General-purpose
LiveJournal	85	18	Russia	General-purpose
Bebo	119	40	USA	General-purpose
PerfSpot	124	20	USA	General-purpose
meinVZ	156	12	Germany	General-purpose
Multiply	161	12	USA	General-purpose
Badoo	168	19	UK	General-purpose
Sonico	183	33	Argentina	General-purpose
Ning	187	1	USA	General-purpose
CyWorld	315	20	South Korea	General-purpose
Xanga	346	40	USA	General-purpose
MyYearbook	406	15	USA	General-purpose
BlackPlanet	1021	18	USA	General-purpose
Plaxo	1486	20	USA	General-purpose
MocoSpace	2582	2	USA	General-purpose
Hyves	4166	8	Netherlands	General-purpose
Impulse	4782	1	Bulgaria	General-purpose
Yonja	5142	4	USA	General-purpose
Bahu	9977	1	France	General-purpose
Nexopia	12109	1	Canada	General-purpose
Eons	17872	1	USA	General-purpose
LinkedIn	149	35	USA	Business-networking
Imeem	186	30	USA	Media recommendation
Last.fm	317	21	USA	Media recommendation
Twitter	338	6	USA	Micro-blogging
Classmates.com	519	40	USA	Reunion
Gaia Online	628	7	USA	Gaming
MyLife	796	58	USA	Reunion
BuzzNet	954	10	USA	Media recommendation
Flixster	975	62	USA	Media recommendation
XING	1023	7	Germany	Business-networking
Viadeo	3280	7	France	Business-networking
Habbo	3349	124	Finland	Gaming
CouchSurfing	4326	1	USA	Travel
Experience Project	8878	2	USA	Privacy-specific
Kaioo	120679	n/a	Germany	Privacy-specific
Imbee	248170	n/a	USA	Privacy-specific

Table 1: Evaluated Social Networks, $N = 45$. User count in millions, rounded.

a few necessary cases, such as Bahu, which prohibits users over the age of 25, or for US-targeted sites which required US postal codes.

3.2.3 Technical Set-up

Recognising that websites may tailor interaction based on any observable data about the user, we were careful to keep the interaction conditions constant. All browsing was performed using IP addresses from the Cambridge Computer Laboratory's address space `128.232.*.*`. During sign-up and interaction with the studied websites, we used Mozilla Firefox v 3.0.6 running on OpenSUSE 11.1 Linux, configured to accept all cookies. We made use of the Screen Grab! v 0.95 plugin to capture images of web pages, as well as the CipherFox v 1.76 plugin to examine TLS connection details.

Examination of sites' Terms of Use and Privacy Policies was performed using a separate machine, running Internet Explorer 7.0 on Windows Vista. This was done to ensure that these documents would be presented as would be shown to a non-member of the site who is considering signing up.

4 Data

This section summarises our major observations from the data we collected. In addition to the figures presented in this section, we have made our entire dataset available online for public analysis.[3]

4.1 Market Dynamics

4.1.1 Network Size

The number of large social networks is impressive, though it is difficult to fairly assess their relative size. It is impossible to externally determine the number of members of a site, so we have relied on the sites' own claims, where available, and the most recent external estimates in other cases, giving us a rough estimates of network size in Table 1.

Member counts mean different things on different sites, for example, many users of Habbo control multiple accounts, inflating the reported number of users, while operating multiple accounts is uncommon and/or banned on other sites. Ning provides a particularly challenging case, as the service allows users to create "their own social network" from a template. Statistics are only published on the number of social networks created (500,000+) which surely underestimates the total number of users.

There are also problems due to large numbers of inactive or rarely-accessed accounts. Windows Live Spaces is particularly problematic because it automatically creates a profile page for every Hotmail user, leading to a huge number of reported users, despite many not actively maintaining their profile. This points to the larger problem of user account statistics including inactive or rarely-accessed accounts. Finally, we were unable to locate any reliable estimates for Imbee and Kaioo, both still too small report user numbers.

4.1.2 Site Popularity: Traffic Data

Due to the problems with network size, we feel that traffic data is a fairer indicator of a site's popularity, though this has complexities as well. We relied on the publicly available Alexa traffic rankings [1]. While these are commonly used as a general indicator of the amount of traffic a site is receiving, the algorithm to compute them is not publicly available so it is impossible to scientifically evaluate their accuracy.

Furthermore, because traffic rankings are produced at the second-level domain granularity, there are several difficulties for social networks which either share a domain with other services, or are spread across several domains. Windows Live Spaces again appears far more popular than it actually is, because

[3] `http://preibusch.de/publ/privacy_jungle/`

spaces.live.com shares its traffic rank with *search.live.com* and other more popular services. Collectively, the *live.com* domain has the #4 traffic rank, although the social networking service accounts for just 1.9% of this traffic. On the other hand, MeinVZ operates under both the *meinvz.net* and *studivz.net* domains, which rank 380 and 156, respectively. In these cases, we simply took the rank of the highest-ranking domain, since there is no way to combine opaque or sub-divide opaque rank data.

4.1.3 Geographical Distribution: American Dominance

With two thirds of our sites head-quartered in the USA, we were initially concerned that our study appeared heavily biased towards American-operated sites, especially given our decision to exclude non-English language sites. However, after analysing usage data we now believe that this mostly reflects the concentration of global web companies in the Silicon Valley area, as indeed most of the American-operated sites are based in the San Francisco Bay Area. We identified an interesting trend in that a number of large sites are based in the United States or at least nominally owned by American parent companies, despite being far more popular in foreign markets [1].

Orkut was famously designed for the US market by Google but has caught on primarily in Brazil and India, where it is now the most popular service. Hi5 is probably the best example, being founded in 2003 in San Francisco, and maintaining a traffic rank of just 96 in the USA, but being the most highly trafficked social networking site in countries as diverse as Honduras, Romania, Thailand and Angola. LiveJournal was founded and run in the USA for almost a decade despite being most popular in Russia, until finally being purchased by a Russian media conglomerate last year. Friendster is an interesting example: it was once the most popular service in the US market, but usage there has drastically fallen off [27], though it remains very popular in Asia, where it is the most popular service in Indonesia, Malaysia, and the Philippines. While these sites have caught on elsewhere despite being designed for the American market, Yonja was founded in the US by Turkish immigrants, and is almost exclusively visited by users from Turkey, though it is still operated in the US. Bebo has followed the opposite path to American ownership, starting in London and being recently purchased by US-based AOL, Inc., despite the majority if its users living in the UK and Ireland.

4.1.4 Site Evolution

Another interesting trend we observed by studying site histories is that many of the sites studied were not originally launched with the goal of becoming large social-networking services, but have evolved into them over the years. Facebook began as a service only for US university students, and MeinVZ similarly began as a directory service for German university students called StudiVZ. Both are now multi-lingual services open to the general public.

Other sites began with simple functionality, and gradually added social features to the point that they now look like general-purpose sites in many respects. LiveJournal, Xanga, and SkyRock (formerly SkyBlog) all began as blogging services, Classmates and MyLife both began with the goal of finding old classmates, and the media-sharing sites began only with anonymous media-ranking functionality. Similar to Zawinski's Law which predicts that all software expands until it can send mail, we propose a new law that all websites expand until users can add each other as friends.

The average age of the networks in our study is just 5.2 years, 5.07 for the general-purpose sites and 5.46 for the others. Impulse, Bahu, Kaioo, and Sonico were the only sites studied which launched within the past 2 years. Classmates, launched in 1995, is by far the oldest, with the next oldest being LiveJournal, CyWorld, and BlackPlanet, all launched in 1999. All of these sites had substantially different purposes when they launched.

4.1.5 Multilingualism

The degree of multilingualism in the sites surveyed was high, indicating that the online social networking paradigm is popular across many different cultures. The average site was offered in 9.1 languages, although the median was just 2, and the standard deviation was 11.1. There is a bimodal distribution between a number of sites offered in just 1 or a small handful of languages, and some very well internationalised sites. 7 sites (NetLog, hi5, Orkut, LiveJournal, Facebook, Windows Live Spaces, and PerfSpot) were offered in at least 25 languages. PerfSpot took the lead with an impressive 46 languages, including Cebuano, Estonian, and Tamil.

4.1.6 Competition

In addition to the variety of languages offered, we analysed country-specific traffic rankings provided by Alexa to approximate the national markets in which sites are competing for new users. As a rough heuristic, we considered two sites to be "competing" within one national market if their traffic ranks are within a factor of two of each other. Using this metric we found significant competition is occurring; every single site surveyed is competing to catch another social network in at least one market. In the English-speaking world, Facebook, MySpace, and Bebo are fighting in different orders as the top 3 in the UK, Ireland, Australia, and New Zealand, with Facebook and MySpace alone competing at the top in the USA and Canada.

There is a common market dynamic throughout Europe, with most countries having a home-grown service competing against a larger, international challenger (usually Facebook but also MySpace, Hi5, and others). Facebook is currently one spot in the rankings behind local competitors SkyRock and Bebo in France and Ireland, respectively, and has recently overtaken local competitors Hyves, meinVZ, and Impulse in the Netherlands, Germany, and Bulgaria, respectively. Even CyWorld, which has dominated the South Korean market for a decade, is now seeing competition from Friendster and Facebook which have slipped into the top 20 sites for the country.

4.1.7 Business Model

Most sites rely on advertisements for revenue, with only the non-profit sites CouchSurfing and Kaioo, and the children's site Imbee not displaying advertisements. We also observed that 7 of the 29 general-purpose sites (24%), but 10 of the other 16 (63%) offered paid premium memberships. These premium memberships typically allow more space for uploading pictures, more control of one's profile, and the removal of advertisements. Premium memberships were offered on all of the business-networking sites and reunion-focused sites, and seem to be a major revenue stream: XING, for instance, generates 80% of its revenue from the 8% of users who are premium members [89]. Many other sites offered the ability for users to buy each other virtual gifts, though these typically sell for only $1 or €1.

Overall, there is a lack of reliable data on the financial situation of social networks, with almost all of them still privately held and operating as start-ups reliant on outside financing. The global market for social networking advertisements is estimated to be US$2.66 billion in 2009 [88], but some market analysis has questioned the profitability of sites given the slow growth of advertising revenue and sites' large operating costs [47].

4.2 Promotional Methods

Most social networks rely on word-of-mouth promotion and there is very little external advertising. However, most sites promote themselves aggressively to non-members who visit in the hope of converting visitors into new users. We compared this promotional process across networks, grouping the most common promotional tactics used into several categories displayed in Figure 1.

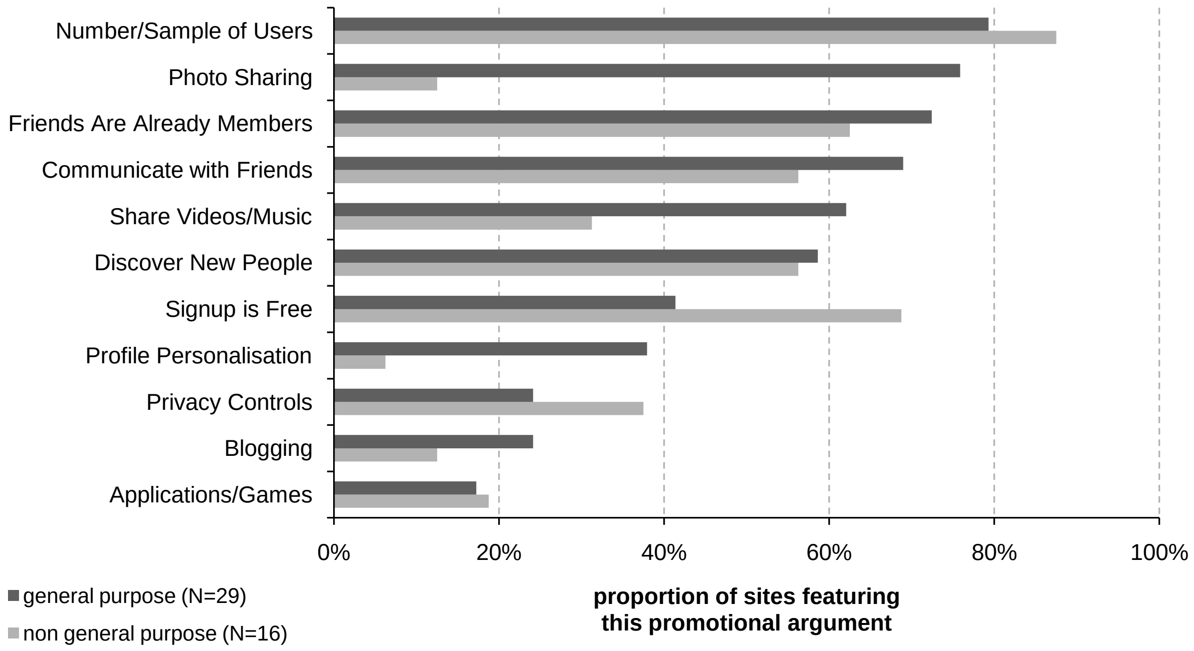


Figure 1: Promotional technique prevalence

4.2.1 Promotion of Social Interaction

Unsurprisingly, a very common marketing strategy is promotion of social interaction on the site. This was observed in sites promoting the ability to send messages using the site (20 / 69%), and extending the possibility of meeting new friends (17 / 59%). These approaches seem to loosely capitalise on the network effects of the site to indicate that one should join based on the ability to interact with other users already on the site.

4.2.2 Promotion via Network Effects

Capitalising on network effects was an explicit strategy for 23 general-purpose sites (79%) which showed a sample of user photos from the site and/or explicitly listed the number of user accounts. This was in fact the most common promotion observed in the sites studied. In addition to listing the total number of users, often as a live counter, many sites listed the number of users who were either currently logged in or came from the same area as the visitor in a further attempt to make the site appear actively used. 21 sites (72%) employed some variation of the argument that "Your friends are already using the site." Network effects were in fact even more commonplace in the niche sites surveyed, being used by all of the media-sharing sites, business-networking sites, gaming sites, and reunion sites.

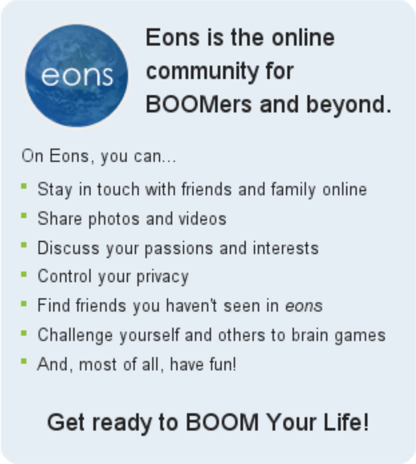
Of the sites showing sample user profiles, some designated user profiles or content as "featured" while others purported to be showing a random sample. In no case did we actually see a random sample that changed when visiting from a separate IP address, indicating that most "sample" users were in fact selected specifically to make the network seem appealing.[4]

The heavy use of network effects is no surprise. User surveys have usually found that the most common reason given for joining a site is because users felt that the majority of their friends were already using it [26, 42].

[4] We certainly noticed a preponderance of attractive and young individuals featured in these photos.

number of promotional arguments	promotion on privacy	
	no	yes
$\leq$ avg	23	2
$>$ avg	9	11
significance	$p = 0.0008$	

Table 2: Privacy as a promotional argument is found significantly more often when many other arguments are also deployed.



Eons is the online community for BOOMers and beyond.

On Eons, you can...

- Stay in touch with friends and family online
- Share photos and videos
- Discuss your passions and interests
- Control your privacy
- Find friends you haven't seen in *eons*
- Challenge yourself and others to brain games
- And, most of all, have fun!

Get ready to BOOM Your Life!

Figure 2: Weak privacy promotion in a long feature list (Eons)

4.2.3 Promotion of Functionality

Where general-purpose social networks choose to promote their actual functionality, the ability to share photos was by far the most common feature mentioned, advertised by 22 sites. Sharing videos and music was almost as common, mentioned by 18 sites. This suggests the interesting possibility that photo-sharing may be the real killer-application which is driving social-networking growth.[5] Every single general-purpose site we surveyed implements the ability to upload multiple photos, whereas only 5 of the 16 other sites implemented photo-sharing, making the ability to share and tag photos seem to be a requirement for a general-purpose social network. This difference is indeed highly significant at $p = 0.04$.

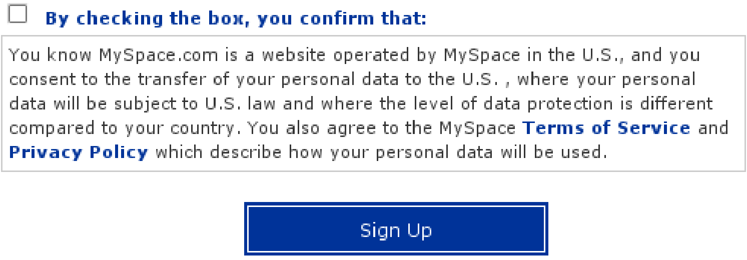
In contrast, the ability to install applications or play games was relatively rarely promoted given the huge amount of attention received by the Facebook development platform and the claim that is decisive factor in the site's popularity [41]. Facebook itself did not mention its application platform. 14 of the surveyed sites implement some platform for third-party applications which users can add to their profiles, but only 5 mention this promotionally, indicating this is not yet considered a major selling point. Other functionality, such as the ability to blog (promoted by 7 sites) and the ability to customise one's profile (11 sites) were similarly much less common in marketing than photo and media-sharing.

The fact that account sign-up is free was promoted by 21 sites, although all the general-purpose sites we surveyed offered free accounts. The freeness of the accounts may be relatively unsurprising today as consumers are conditioned to expect web services to be free. However, 7 of surveyed general-purpose sites do offer premium accounts, usually removing advertising and offering more storage for a monthly fee. 4 of these 7 optionally paid-for sites still promoted free sign-up, a higher percentage than sites without paid accounts. Similarly, there was an increase in promotion based on sign-up being free among the niche sites, despite a higher proportion of them offering paid memberships. This is possibly an indication that consumers are less likely to expect sites in the areas of music, gaming, and business to be free.

4.2.4 Promotion of Privacy

Finally, privacy was used as a selling point in 7 out of 29 general-purpose sites, and when it was mentioned it was typically in a vague and general fashion. 4 sites explicitly mentioned privacy: PerfSpot

[5] Indeed, Facebook hosts more user photos than any other website, with over 40 billion.



□ **By checking the box, you confirm that:**

You know MySpace.com is a website operated by MySpace in the U.S., and you consent to the transfer of your personal data to the U.S. , where your personal data will be subject to U.S. law and where the level of data protection is different compared to your country. You also agree to the MySpace **Terms of Service** and **Privacy Policy** which describe how your personal data will be used.

Figure 3: Terms of Use and Privacy Policy acknowledgement (MySpace)

claimed “unmatched privacy controls,” meinVZ offered “a wide range of settings that concern your privacy,” Eons mentioned the ability to “control your privacy,” and Sonico to “share photos, videos, and your interests privately.” 3 other sites made vague reference to access control: Windows Live Spaces stated that you decide “who sees your space, and who doesn’t,” Multiply claimed it was easy to share photos with “more people you want and fewer people you don’t,” and Hyves stated “you decide which information is available to whom.” Hyves also deserves commendation for promising “we’ll never sell your information,”—the only site we observed making such a guarantee. None of these promotions made any reference to or linked to the site’s privacy policy, no site attempted to use the contents of its privacy policy as a promotional tool. 2 of the 3 business-networking sites mentioned privacy, as did 2 privacy-specific sites, but just 2 of the 11 other niche sites mentioned privacy.

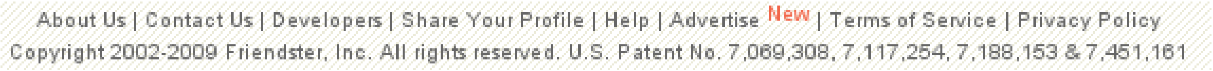
In addition to the relative rarity with which privacy was mentioned promotionally, we found strong evidence that it is not used as a primary argument, but as one of many items in a long feature list. For general-purpose sites, sites mentioning privacy used an average of 8.0 promotional categories, whereas sites not mentioning privacy used an average of 5.74. Privacy was never mentioned by a site which used fewer than 5 other promotional arguments. Fisher’s exact test reveals strong statistical significance in that privacy only emerges as a “yet another” argument (Table 2). The promotional page from Eons (Figure 2) provides a typical example of privacy being mentioned in a nondescript way, lost among other features.

4.3 Presentation of Terms of Use and Privacy Policy

We recorded the means in which users where presented with the site’s Terms of Use and Privacy Policy, as signing up is effectively a legal agreement governed by the these documents. Typically, there is a disclaimer placed near the submission button during signup which contains a reference to the Terms of Use, and sometimes the Privacy Policy as well. A particularly clearly-stated example from MySpace is shown in Figure 3. Unfortunately, most sites made scant mention of their privacy policies during sign up.

4.3.1 Privacy Policy Acknowledgement

Despite signing up being considered legal acceptance of the Privacy Policy in every site studied, only 5 of the 29 general-purpose sites required actively checking a box to indicate acknowledgement of the privacy policy, whereas 12 require checking a box to acknowledge the terms of service. 17 sites mentioned the privacy policy on the signup page, although only 11 of these placed the privacy policy reminder on the user’s critical path, with 3 placing it in the page’s margin and 3 placing it below the submission button. Results were even worse for the other sites surveyed, with 10 sites of 16 mentioning the privacy policy, but only 4 placing the reminder above the submission button.



About Us | Contact Us | Developers | Share Your Profile | Help | Advertise New | Terms of Service | Privacy Policy
Copyright 2002-2009 Friendster, Inc. All rights reserved. U.S. Patent No. 7,069,308, 7,117,254, 7,188,153 & 7,451,161

Figure 4: Privacy Policy link hidden in bloated page footer (Friendster)

4.3.2 Privacy Policy Review

In addition to not forcing users to actively acknowledge the privacy policy, very few sites encouraged users to read it. MeinVZ was a commendable exception, displaying a condensed version of the site's privacy policy on a separate page during the signup process. MySpace and Viadeo both displayed shorter extracted paragraphs from their privacy policies, and Imbee gave users a strong nudge to "Read our PRIVACY POLICY!" However, the remaining sites, including 27 of the 29 general-purpose sites, included essentially no pertinent information about the privacy policy on the user's path to creating an account.

Ten general-purpose sites made no reference to the privacy policy at all. Of the 17 general-purpose sites which did mention the privacy policy, 4 of them forgot to include a link to actually read it. 11 sites failed to mention the policy but provided a link in a standardised page footer, and 5 offered no link at all. On the sites linking to the privacy policy from a footer, typically it was grouped with many other links including help info, contact info, and information for advertisers. A glaring example is shown in Figure 4, as Friendster buried its privacy policy link along with 7 other links and a list of patents held. An additional 2 sites made the mistake of including links which did not open in a new window or tab, meaning that clicking on them would interrupt the signup process. Of the non-general-purpose sites, 4 failed to provide links to their privacy policies during signup, and 2 more included links not opening in a new window.

4.4 Data Collected During Sign-up

While signing up for each of the networks in our study, we recorded the amount of data which must be reported create a new account. We also recorded the amount of data which is requested but not required, though we consistently chose to withhold such data. We found remarkable variation between the general-purpose sites as to what data was collected, summarised in Figure 5.

4.4.1 Over-Collection of Demographic Data

In general, far more personal data is collected than is needed for a user to interact with a social networking service, particularly gender and birth date information. Gender was required by 20 sites and requested by 4 others. A full date of birth was required by 24 sites and requested by 2 others.[6]

These two pieces of data are both useful to personalise the site but should not be mandatory. We did observe several sites promoting reminders of friends' birthdays as a reason to use the site; the huge popularity of this feature could be a reason that this data is often required [30]. Similarly, the majority of the sites offer demographic search capabilities: 22 out of 29 sites allow finding fellow members based on location, gender, interests, and other data, instead of just name.

Photographs and information on employment and university affiliations are similarly unnecessary, but were not required except in the case of BlackPlanet, which requires a user's "Job Type." BlackPlanet was an outlier as it also requested a user's race, ancestry, income level, and sexual orientation during sign-up. Yonja went a step further in actually requiring users to report their sexual orientation.

[6] Six of the sites requesting a user's data of birth provided a check-box to hide the visibility of the date of birth on the form in which it was requested.

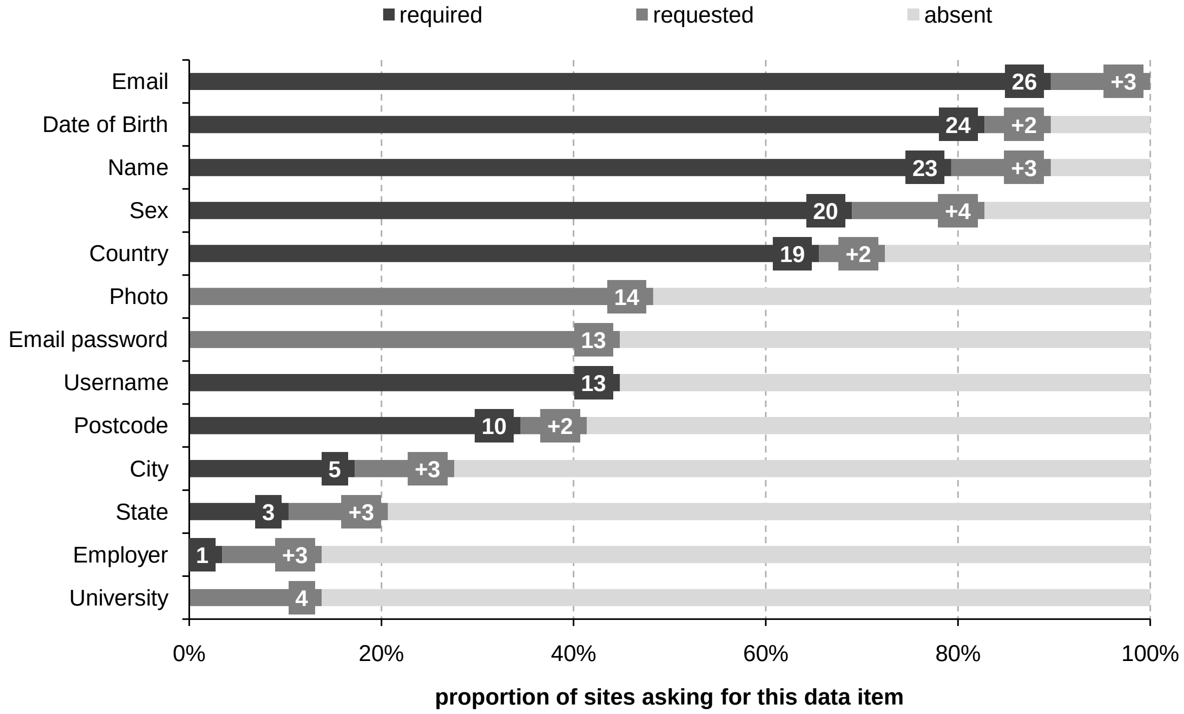


Figure 5: Visibility of profile data, general-purpose networks, $N = 29$

4.4.2 Requirement of Real Names

The widespread requirement of reporting names is similarly troubling, as 23 of the 29 sites require entering one's full name to join the site.[7] Only 3 sites were purely pseudonymous (Nexopia, Xanga, MocoSpace), with 3 other sites (LiveJournal, SkyRock, BlackPlanet) requesting a name but not requiring it. Of the sites which do not require a name, Xanga, LiveJournal and Skyrock all began as blogging services and have since transformed into social networking services, indicating that pseudonymity may be more desirable for blogging services than for general-purpose social networks.

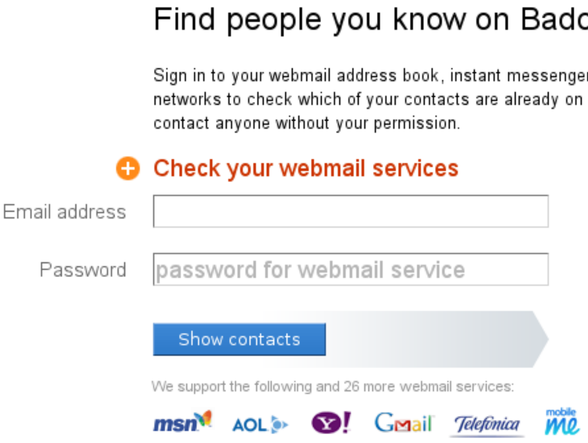
In addition to the 6 sites for which a name is optional and a pseudonym is the main identifier on the site, 7 more sites require a pseudonym or username for the site. This does not provide much privacy however as names are still displayed on all of these sites. From the non-general-purpose sites, the gaming websites, 2 media-sharing sites and ExperienceProject were strongly pseudonymous, not collecting names at all.

The utility of pseudonyms on social networks is controversial. One study reported that an excess of fake profiles was a contributing factor in Friendster losing popularity [27], while others found that many youth desire the ability to sign up under pseudonyms [85, 30].

4.4.3 Requirement of Email Addresses

It is also notable that every site required an email address to join, including the privacy-specific sites. Most of the general-purpose sites (26 out of 29) further require email verification in order to use the site, with only Hyves, meinVZ, and MyYearbook not verifying email addresses. Requiring a valid email address could be seen as an anti-spam technique, although 25 of the general-purpose sites already require

[7] Of course, this is never strongly verified, and there is anecdotal evidence of fake names commonly being provided [26].



Find people you know on Badoo

Sign in to your webmail address book, instant messenger, or search for people on your social networks to check which of your contacts are already on Badoo. We will not store your login info or contact anyone without your permission.

Check your webmail services

Email address

Password: password for webmail service

We support the following and 26 more webmail services:

Figure 6: Interface to import address book from external webmail account (Badoo)

their own CAPTCHA to sign up. Although it is easy to obtain free and disposable email addresses online, most users will enter their real email-address, making the insistence on email addresses a needless privacy violation since they are not necessary for interaction with a social networking site. [8]

Almost half of the sites requested the password to one's email address as well, in order to automatically retrieve a person's friends from their email provider. A typical interface is shown in Figure 6. In addition to this feature, 4 sites offer an "invite friends" feature which will send invitations to join the network to every email address listed in the user's webmail account. On top of generating spam, these features are poor user training for phishing, as they reinforce the habit of entering passwords into third-party websites.

4.5 Privacy Controls

After signing up, we examined the privacy controls offered by each site. While almost every site has invented its own unique terminology to describe access control, we were generally able to map these into categories which were common across sites. One limitation of our approach is that we did not verify the correct functioning of the privacy controls, which would require creating multiple test accounts with each site and examining the visibility of one profile from another under different settings.

4.5.1 Profile Visibility Options

The fundamental privacy setting is the visibility of one's profile page, which is the main display of one's personal information on the site. Options provided were profiles accessible to the public internet, profiles only accessible to site members, limitations by geographical or other sub-networks, and limits to friends only or friends of friends. The availability of these levels of control is displayed in Table 3. Only 3 sites provided no choice on profile visibility, with Skyrock making all profiles internet-public by default, and Yonja and Multiply making profiles viewable by all members.

It is important to point out that limiting profile views to only members of the site provides very little privacy, since membership is free and easy to obtain for every site surveyed. This distinction is really only useful for privacy in that search engines will not crawl the pages from sites with members-only privacy settings. Sites probably choose to limit visibility to members only in order to force visitors to

[8] Email addresses are used as login names on most sites, but this could be easily changed.

visibility level	default	optional	unavailable
public Internet	41%	-	59%
all site users	48%	28%	24%
sub-networks only	7%	17%	76%
friends of friends	-	24%	76%
friends only	3%	79%	17%

Table 3: Visibility of profile data amongst general-purpose sites, $N = 29$: Most sites make profiles publicly visible by default.

feature	separate ACL	profile ACL	no ACL
profile commenting	62%	21%	17%
messaging	52%	28%	21%
photo viewing	52%	41%	7%

Table 4: Access controls for additional features, general-purpose networks, $N = 29$

sign up to be able to view people's profiles. Facebook takes an interesting hybrid strategy, showing a limited "public listing" of profiles to visitors and search engines, using this to encourage membership.

4.5.2 Fine-Grained Controls

Many sites offer more fine-grained control of profile visibility, with 13 general-purpose sites offering a line-item setting where individual data items may have different visibility, argued to be a crucial feature for privacy management [77]. An average of 10 different profile items were configurable, with Windows Live Spaces offering the most at 27. Of these, only Facebook and LinkedIn offered the useful "audience view" feature, allowing users to see how their profile looks to different groups of users [56].

We found 8 sites which implemented a version of role-based access control by giving users the ability to segregate their friends into abstract, self-defined groups and make access control decisions at the group level. Of these 8, only 2, PerfSpot and Plaxo, made friend grouping mandatory, as has been shown to greatly enhance users' ability to control their privacy [66]

Other common privacy controls regulate photo access and the ability to send messages and post public "comments" to other users on the site. Access control offerings for these features are shown in Table 4. Most sites offered the ability to restrict these features separately, only Skyrock and Badoo, which operate with all profiles being completely open, did not provide the ability to limit visibility of photos.

4.5.3 Permissive Defaults

The main problem observed, however, was not lack of options but the almost universality of open defaults. Estimates have varied in previous literature, and depend on the site in question, but between 80 and 99% of users are typically found to never change their privacy settings [8, 53, 51]. For more obscure privacy-violating features such as those described in Table 5, fewer than 1% of users are thought to opt-out [22, 21]. A significant number of users are not even aware that privacy controls exist in social networks, estimated in two different studies at 26% [48] and 30% [8].

As seen in Table 3, all but 3 of the general-purpose sites (90%) leave new profiles initialised to be completely visible to at least all other members of the site by default. Of these, Friendster's default limitation to a user's continent and Facebook's limitation to a user's sub-networks provide relatively little privacy gain, since anybody can create a profile in these networks. Only Bebo defaulted users to

feature	implemented	opt-out	% opt-out
user event stream	14	11	79%
online status visibility	25	22	88%
profile viewing notification	16	12	75%
profile searchability	29	27	93%

Table 5: Most general-purpose sites have privacy-invasive discoverability features enabled by default and require manual opt-out from the user, $N = 29$

a friends-only view among the general-purpose sites. Similarly poor default privacy was found in the niche sites, with only the child-specific site Imbee using friends-only privacy by default (and in fact as the only option).

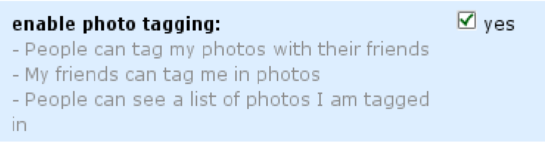
Often, default privacy settings left unnecessarily detailed data traces available to other users. The publication of a stream of user events, such as "Upton uploaded a new photo" or "Upton changed his relationship status," and of the user's online status can be aggregated into temporal usage patterns with serious privacy implications. A network user determined to monitor other users' behaviour may often benefit from demographic search capabilities to spot interesting surveillance targets, since most sites enable user discoverability beyond the name. Search was implemented on all of the sites; only two general-purpose sites (Eons and Badoo) forced users to manually opt-in for the profiles to be indexed. Finally, bilateral profile viewing notifications constitute a privacy dilemma in that enabling them per default unveils the casual stalker but constitutes a hurdle for inconspicuous network browsing. Table 5 summarises the proportion of sites requiring opt-out instead of opt-in for these privacy-invasive services.

4.5.4 User Interface Problems

In addition to the problem of permissive default settings, we observed many possible user interface problems which could limit the ability of users to effectively use the available privacy controls. This was reflected by a survey which found that 24% of Facebook users did not understand the implications of their own privacy settings [8]. There is also anecdotal evidence from the web that users are confused about privacy settings, such as a guide to configuring one's privacy settings for social networks which was downloaded over 500,000 times [65]

Many sites presented controls in an excessively complex way, although academic studies have found that providing users too much information and too many configuration options can harm usability [87]. Facebook had the most complex settings, with 61 options to select spread across 7 different privacy settings pages. LinkedIn also stood out with 52 settings and 18 pages. Windows Live Spaces suffered from particularly poor usability. For each of its 27 settings, the user most load a new page to examine the value of the setting, load a second page to edit the setting, and then click "SAVE" and submit a form to record the new setting. The average general-purpose site offered 19.2 privacy settings on 3.7 separate pages (median 16 / 2). Users also face an often overwhelming array of choices for controlling the amount of email received from the site, with an average of 13.0 email settings available, with only Nexopia, SkyRock, and Yonja not allowing users to control the amount of email received.

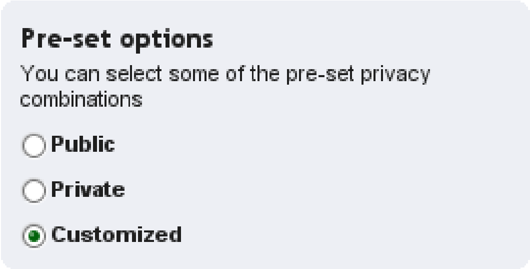
In addition to the complexity observed, we found many cases of confusing settings, ambiguous wording, and inconsistent use of terminology between sections of the same site's privacy settings. Orkut provides a telling example in Figure 7. The check box marked "enable photo tagging" actually relates only to the ability of others to tag photos, and also controls the ability to view a list of a user's tagged photos even if that user tagged the photos personally. The first sentence also includes a confusing dangling modifier; it is not clear if the phrase "with their friends" refers to who is being tagged or who is doing the tagging. Badoo provided another confusing example, offering the choice between making one's profile visible to "any users" or "only members." It is assumed that "any users" can include non-registered-members,



enable photo tagging: ☑ yes

- People can tag my photos with their friends
- My friends can tag me in photos
- People can see a list of photos I am tagged in

Figure 7: Coarse-grained privacy setting with potentially confusing wording and non-standard input controls ("☑ yes")(Orkut)



Pre-set options

You can select some of the pre-set privacy combinations

○ Public

○ Private

● Customized

Figure 8: Pre-selected combinations of privacy settings (Sonico)

though after selecting the "only members" setting it was displayed as "any members." Only 6 sites offered written help in managing privacy settings, exacerbating the problem of confusing terminology and labelling.

A very nice but rare feature was pre-set combinations of privacy settings which could be selected with one click. This was offered by Sonico, offering basic "Public," "Private," and "Custom" settings (Figure 8), and NetLog which offered "Meet new people" and "Keep in touch with my friends" settings, each with an additional "high privacy" option. MySpace also offered pre-set combinations of settings, but only to control who on the site is allowed to message a user.

4.6 Security Measures

4.6.1 Use of TLS Encryption and Authentication

We found an appallingly low adoption of the standard Transport Layer Security (TLS, formerly SSL) protocol. 17 of the 29 general-purpose sites failed to use TLS during log-in, and of the 12 using it, only 3 wrapped the entire log-in page in a TLS connection. The other 9 only used TLS for the HTTP `POST` action, which is undesirable because it prevents browsers' TLS indicators from being displayed, making users more susceptible to phishing. TLS adoption was slightly better in the other sites surveyed, with 6 of the 16 using TLS for the entire login page, and 2 for the `POST` action only.

A common error observed even among sites using TLS for login was forgetting to use TLS during the signup process, when passwords are also entered. 6 sites which used TLS during login did not use it at all during signup, with 2 sites making the opposite mistake. Both mistakes are a sign of careless implementation, as the sites clearly have the ability to deploy TLS but forget that there are two common situations where passwords are entered. Plaxo provided a particularly bizarre example of TLS inconsistency, using TLS to protect the requested email password for its "retrieve friends" feature but failing to protect the password entered as part of the signup data itself. Overall, 21 of the general-purpose sites and 9 other sites used no TLS during signup.

Disappointingly, only one website surveyed, the business-network XING, provided TLS for all interaction with the site. Curiously, despite this strong security practice, XING was not one of the sites which promoted itself on the basis of privacy. In fact, of the 13 sites which did promote themselves based on privacy, 7 employed no TLS whatsoever, and only 2 provided TLS for their complete log-in pages. [9]

4.6.2 Phishing Prevention

There was a glaring lack of attention paid to phishing in the sites surveyed. Not a single site used any anti-phishing mechanisms during login, such as personalised images displayed next to password prompts.

[9] We suggest that comprehensive TLS encryption might be used as a promotion technique for evading traffic logging schemes deployed in the European Union.

Only two websites surveyed (MySpace and BlackPlanet) made any mention of phishing in warning users only to enter their password at their site. Every single site sent us emails containing links requesting us to log-in to the site, easy for phishers to replicate fraudulently.

Coupled with the poor use of full-page TLS for log-in described in Section 4.6.1 and the common practice of requesting passwords for external email accounts described in Section 4.4, this represents an industry-wide disregard for the problem, though it has been made a point of government policy emphasis [17]. Academic research demonstrated years ago the power of "social phishing" using compromised account due to the social trust inherent in communication on social networks [46]. There is also empirical evidence that phishing is commonplace in large social networks [13, 17], and that phishers are now using stolen social network accounts to request money from unsuspecting online "friends" [39].

4.6.3 Online Safety Guidance & Abuse Reporting

Preventing abuse is another important challenge for social networks, as research has suggested cyber-bullying by peers is a significant threat [17], and the majority of young users report being harassed by another user to the extent that they blocked them [30]. Encouragingly, we observed widespread deployment of three mechanisms for preventing cyber-bullying: the ability to block access by specific users, the ability to report specific user profiles directly from the profile page, and web forms for reporting abuse. Every site implemented at least one of the three options, including at least one interface for reporting abuse, with the exception of Plaxo. However, in many cases the abuse reporting web form required clicking on several links to reach. Habbo made the bizarre choice to require completing a CAPTCHA before submitting an abuse report.[10] 10 general-purpose sites failed to implement the much more user friendly "Report User" ability on each profile page. Only one site, PerfSpot, provided a telephone hotline to speak with a representative.

11 general-purpose sites also provide help pages for maintaining online safety, with 9 providing specific help pages for parents. More sites could easily provide such pages, since many of the pages had very little unique content and mostly contained links to the plethora of non-profit online safety pages available on the web [2]. Only 6 general-purpose sites provided help pages for managing privacy. Again, there was a lack of correlation with sites promoting their privacy and providing privacy settings help, with only 1 site, Multiply, doing both.

4.7 Privacy Policies

Besides being a legally binding contract between the social network operator and its users, the privacy policy is the only primary source that a prospective user can rely on to give informed consent for data collection, as is required in the EU. Therefore, it is critical that sites post documents which are accessible both technically and linguistically. The results of our inspection of the privacy policies are summarised in Table 6. Two sites, SkyRock and Impulse, failed to provide a privacy policy separate from their Terms of Use. We analysed SkyRock's Terms of Use section on data protection practices since it was clearly labelled "Protection of Users' Privacy and Personal Data". We were unable to count Impulse's one-line statement on users' privacy[11] as an actual privacy policy. For completeness, we still report the analysis results of this statement as "Impulse (T&C)" in Table 6.

The quality of a privacy policy is not to be confused with the quality of data protection standards the site implements. Rather, as an enabler for informed consent, a policy should give a good account of the practices regardless of whether these are beneficial or detrimental for a user. As such, a site that honestly and clearly states horrific data collection, usage, and sharing has a better policy than a site with nebulously-phrased text that does not mention data sharing with third parties.

[10] In fact, Habbo utilised a more difficult CAPTCHA for reporting abuse than for signing up.

[11] Impulse's complete statement on privacy: "guarantees not tot [sic] share users' personal information with third parties (except for the cases provided by the law) and not to use it for any other purposes except those of the site;"

site	PP present	PP new window	PP requires JavaScript	P3P full	P3P compact	PP dated	PP word count	PP mobileOK	PP zoomable	PP durable Url	PP printable	PP savable	PP textually structured	PP contains operator email address	PP contains operator postal address	PP contains seal	PP external dispute resolution	PP Safe Harbor participant	PP specifies national laws	PP specifies data locations	PP specifies data retention period	PP IP address collected	PP browser data collected	PP external data collected	PP shares with third parties	PP data anonymised for third parties	PP shares with search engines	PP shares with law enforcement	PP has third-party advertisers	PP user can delete data	PP user notified of changes	PP changes take effect delayed	TC minimum age
Badoo	y	n	n	y	n	n	1713	100	y	y	y	y	y	n	y	n	n	n	n	n	n	y	y	y	y	y		y		y	y		18
Bahu	y	n	n	n	n	n	266	91	y	y	y	y	y	y	n	n	n	n	n	n	n				n					y	n	n	13
Bebo	y	n	n	n	n	y	2842	42	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y		y	y	y	y	n	13
BlackPlanet	y	y	n	n	n	y	2888	100	y	y	y	y	y	y	y	n	n	n	y	n	n	y	y	y	y	n		y	y	u	y	n	14
BuzzNet	y	n	n	n	n	y	1781	86	n	y	y	n	y	y	y	n	n	n	n	y	n	y	y		y	y	y	y	y	y	n	n	13
Classmates.com	y	y	y	y	y	y	4934	48	y	y	y	y	y	y	y	y	y	n	y	y	n	y	y	y	y	n		y	y	u	y	y	13
CouchSurfing	y	n	n	n	n	n	1211	53	y	y	y	y	y	n	n	n	n	n	n	n	n	y		y	y	n				y	y	n	18
CyWorld	y	y	n	n	n	y	1870	54	y	y	y	n	y	y	n	n	n	n	n	y	n	y	y		y	y		y	y	u	n	y	13
Eons	y	y	n	n	y	y	1814	80	y	y	y	y	y	n	y	n	n	n	n	n	n	y	y	y	y	y		y	y		y	n	13
Experience Project	y	n	n	n	n	y	1502	91	y	y	y	y	y	y	n	n	n	n	n	n	n	y	y		y	n		y		y	y	n	13
Facebook	y	y	n	n	u	y	3765	56	y	y	y	y	y	n	y	y	y	n	y	y	n	y	y	y	y	n	y	y	y	y	y	y	
Flixster	y	n	n	y	n	y	2254	65	y	y	y	y	y	n	y	n	n	n	n	y	n	y	y		y	n		y	y	y	y	n	13
Friendster	y	y	y	n	n	y	1973	83	y	y	y	y	y	y	y	n	n	y	y	y	n	y	y		y	y		y	y		n	n	
Gaia Online	y	n	n	n	n	y	2249	63	y	y	y	y	y	y	y	n	n	n	n	y	n	y	y		y	y		y	y	u	y	n	13
Habbo	y	y	n	n	y	y	4166	47	n	y	y	y	y	y	y	n	n	n	y	y	n	y	y	y	y	n		y	y	y	n	n	13
hi5	y	n	n	n	n	y	2193	87	y	y	y	y	y	y	y	n	n	n	n	y	n	y	y		y	n		y	y	y	y	n	
Hyves	y	y	n	u	n	y	1706	38	y	y	y	y	y	n	n	n	n	n	y	y	n	y	y		y	n		y			n	n	0
Imbee	y	n	n	n	n	y	2240		y	y	y	y	y	y	y	n	n	n	y	n	n	y						y		y	n	n	
Imeem	y	n	n	n	n	y	1887	46	y	y	y	y	y	y	y	n	n	n	y	y	n	y	y		y	y		y	y	y	y	n	13
Impulse	n			n	n																												13
Kaioo	y	n	y	n	n	n	1418	57	y	n	y	n	y	y	n	n	n	n	y	n	n				n			y	n	u	n	n	14
Last.fm	y	y	n	n	n	y	4374	37	y	y	y	y	y	y	y	n	n	n	y	y	n	y	y	n	y	n		y	y		n	n	13
LinkedIn	y	y	n	n	y	y	4957	58	y	y	y	y	y	y	y	y	y	y	y	y	n	y	y			y	y	y	y	u	y	n	18
LiveJournal	y	n	n	n	n	y	2655	69	y	y	y	y	y	y	n	n	n	n	y	n	n	y	y	y	y	n		y	y	u	y		13
meinVZ	y	y	n	n	n	y	8455	52	y	y	y	y	y	y	y	n	n	n	y	y	y	y	y		n			y	y	y			18
MocoSpace	y	n	n	n	n	n	1344	71	y	y	y	y	y	y	n	n	n	n	n	n	n	y	y		y	y		y	y	y	y	n	13
Multiply	y	n	n	n	n	n	2142	65	y	y	y	y	y	n	n	n	n	n	y	n	n	y	y		y	y		y	y	y	n	n	13
MyLife	y	n	n	y	y	n	4083	57	y	y	y	y	y	n	y	y	y	n	n	n	n	y	y	y	y	n	y	y	y	y	y	n	13
MySpace	y	y	n	n	n	y	2738	45	y	y	y	n	y	y	y	n	n	n	n	y	n	y	y		y	n		y	y		n	n	13
MyYearbook	y	n	n	n	n	y	955	57	y	y	y	y	y	n	y	n	n	n	n	n	n	y	y	y	y	n		y	y			n	13
NetLog	y	y	n	n	n	n	311	94	y	y	y	y	n	y	y	n	n	n	y	y	n							y		u			
Nexopia	y	n	n	n	n	y	2752	39	n	y	y	y	y	y	y	n	n	n	n	y	n	y	y		y	y		y	n	u	y	n	13
Ning	y	n	n	n	u	y	4135	79	y	y	y	y	y	y	y	n	n	n	y	y	n	y	y	y	n			y	n	u	n	n	13
Orkut	y	n	n	u	n	y	3073	80	y	y	y	y	y	n	y	n	n	y	y	y	n	y	y	y	y	y		y	y	y	y	n	
PerfSpot	y	n	n	n	n	y	2108	20	y	y	y	y	y	y	y	n	n	n	y	n	n	y	y		y	n		y	y	n	n	n	13
Plaxo	y	n	n	n	n	y	4271	51	y	y	y	y	y	y	y	y	y	y	n	n	y	y	y		y	y		y		y	y	y	13
SkyRock	n	n	n	n	y	n	641	69	y	y	n	y	y	n	y	n	n	n	y	y	n	y	y					y		y			
Sonico	y	n	n	n	n	n	923	73	y	y	y	y	y	y	n	n	n	n	n	n	n				n					u			
Tagged	y	y	y	n	n	y	2799	68	y	y	y	y	y	y	y	n	n	n	y	y	n	y	y		y	n		y	y	u	n	n	13
Twitter	y	n	n	n	n	y	1535	67	y	y	y	y	y	y	n	n	n	n	y	y	n	y	y		y	n		y		y	y	n	13
Viadeo	y	n	n	y	y	n	2848	18	n	y	y	y	y	n	y	n	n	n	y	n	n	y	y		n					y	n	n	18
Windows Live Spaces	y	y	n	n	u	y	4361		y	y	y	y	y	n	y	y	n	y	y	y	n	y	y	y	y	y		y	y		y	n	
Xanga	y	n	n	n	n	n	4948	55	y	y	y	y	y	n	n	n	n	n	n	y	n	y	y		y	y	y	y	y	u	y	n	13
XING	y	n	n	n	n	y	3237	58	y	y	y	y	y	n	n	n	n	n	y	n	n	y	y		n		y				y	n	17
Yonja	y	n	n	n	n	n	1557	56	y	y	y	y	y	n	n	n	n	n	n	n	n	y	y		n			y	y	u	y	y	18
Impulse (T&C)	u	n	n	n	n	n	32	67	y	y	y	y	y	n	n	n	n	n	n	n	n				n			y			n	n	13

Table 6: Privacy Policy evaluation results. Fields are left blank where an evaluation criterion was inapplicable or if the site did not specify the information required for evaluation. Cells marked ‘u’ indicate implementation, but with errors, for the P3P policies and only partial data erasability for the criterion “PP user can delete data”.

4.7.1 Technical Accessibility

It is critical for privacy policies to be accessible to a variety of web browsers and devices to avoid disenfranchising any users. As social networks grow, adherence to good accessibility principles is increasingly important to enable use from mobile devices and by elderly and disabled individuals who may have special accessibility needs [81].

Despite this, we noticed numerous accessibility problems. 15 sites opened their privacy policies in a new window, which can be blocked by pop-up blocking browsers or unsupported by mobile devices. 4 sites required JavaScript to display the privacy policy, which is incompatible with older browsers or some mobile devices. 4 sites deployed privacy policies which did not allow zooming, 4 sites deployed policies which could not be saved, and 1 site (SkyRock) had a policy which could not be printed. These errors were not committed by the same few sites, 21 sites made at least one such accessibility error.

We also verified accessibility for mobile devices using the W3C mobileOK Checker [84], which checks a Web page against a defined set of recommended guidelines derived from best practices for the mobile Web and issues scores between 0 and 100. This is a rigorous test which is also a good indicator of accessibility in general. Only 2 sites, Badoo and BlackPlanet, received a perfect score. Even MocoSpace, targeted specifically at mobile devices, had numerous problems and received a score of just 71.

4.7.2 Length

Given the diversity of written privacy policies and the lack of a standardised vocabulary, we recorded the textual length only in place of a subjective measure of readability.[12] Only 10% of users claim in surveys to have read the privacy policy of their social-networking site [48], and examinations of server logs indicate the actual rate may be far less than 1% [86].

Privacy policies in general were too long to be expected to be read by most users, although the length varied greatly. The mean length was 2,633 words, with a median of 2,245, and a very large standard deviation of 1,546 words. The three shortest policies were all translated from originally French-language sites, the 266, 311, and 641 word policies of Bahu, NetLog, and SkyRock, respectively. The longest policy was the 8,455 word epic from meinVZ, nearly 3,500 words longer than the next longest, that of LinkedIn. There were 12 policies longer than 3,000 words, which are all far too long to provide usable privacy information.

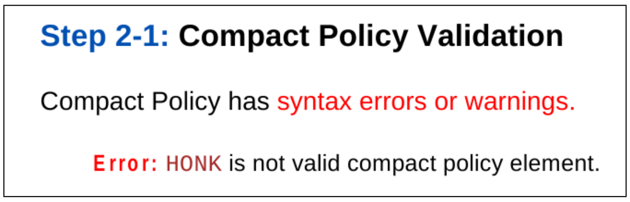
4.7.3 Legal Issues

Due to the nature of privacy policies as legal contracts, it is critical for them to provide some basic contractual information. Nevertheless, 13 sites failed to provide a date on their privacy policies, 15 sites didn't list a physical contact address, 17 sites didn't provide an official email address, and 7 sites provided no contact info at all.

21 sites reserved the right to change the terms without notice, making them of questionable contractual value. Only 5 sites guaranteed a minimum notice period before changes could take effect.

Finally, there were problems with specifying legal jurisdiction, especially pressing given the noted discrepancy between the geographic location of operators' headquarters and their targeted regional markets (Section 4.1.3). 20 sites did not specify which nation's data protection laws they followed, and 20 sites did not specify in which nations data would be stored and processed in. Only 17 sites specified both, which would be required information in the case of a dispute. The EU Safe Harbour Agreement, designed to enable compliance with the EU Data Protection Directive for foreign companies with EU customers, was only acknowledged by 6 sites, despite the prevalence of this geographic pattern. 6 sites specifically named an external party to arbitrate disputes arising from the privacy policy.

[12] Subjectively, we generally found readability to be poor.



Step 2-1: Compact Policy Validation

Compact Policy has syntax errors or warnings.

Error: HONK is not valid compact policy element.

Figure 9: P3P compact policy file validation errors (Facebook)

4.7.4 Data Claims

Regarding the actual claims that were made in the policies, there was significant variation, but a pattern emerged of few meaningful rights being assigned to users and operators reserving many data collection and sharing rights for themselves. In addition to user-uploaded profile data, 40 sites specifically reserved the right to record IP addresses and/or browser data. No sites promised not to collect such data, the other 5 sites left the issue unspecified. 14 sites also reserved the collect user data from external sources. Most sites were unclear on this point, with only Last.fm promising not to do so. Few data retention guarantees were made, with only Bebo, meinVZ, and Plaxo providing specific limits on how long they could retain user data. 21 sites did explicitly grant users the right to have their data deleted upon request, as is legally required in the EU, with 24 sites either providing an incomplete guarantee or leaving the point in question.

Operators also often reserved many rights to share user data. 32 explicitly reserved the right to share with third parties, while only 8 promised not to. Of the 32, 17 promised to anonymise user data (although academic research has proven this is impossible for realistic social graph data [15, 64]). 39 sites indicated they would share data with law enforcement when required to do so, with 6 failing to mention this.

4.7.5 Availability of P3P Policies

We evaluated the adoption of policies conforming to the W3C's P3P format [4], designed to enable users to quickly determine if a site's privacy practices are acceptable given the user's privacy preferences [73]. P3P has been argued to be a critical element in enabling privacy protection in the future [11], and has been shown to strongly influence user decision-making when its display is mandatory [40].

We saw low adoption of P3P among sites surveyed, with only 7 sites implementing a full P3P policy, 5 of which parsed correctly. Badoo and Hyves were the only general-purpose sites with correctly implemented policies. 10 sites implemented a compact policy, 7 correctly, including just SkyRock and Eons among the general-purpose sites. The lack of P3P adoption and the existence of incorrectly written policies indicates a negative attitude toward the P3P project by some site operators. As shown in Figure 9, Facebook's P3P compact policy provided a vivid example, consisting of an incorrect policy element name "HONK." This seems crafted specifically to mock users with P3P-displaying browsers.[13]

4.7.6 Self-Promotion within Privacy Policies

Despite the poor observed quality of privacy policies, an interesting trend was that many sites included promotional claims about their privacy practices within the privacy policies themselves. Some typical examples are shown in Figure 10. We recorded the use of such reassuring but legally meaningless phrases in privacy policies. Typically, these are written in simple English and make strong claims that privacy is an important consideration within the site. Overall, we observed this tactic in 34 of the 45 sites studied, with 21 of 29 general-purpose sites and 13 of 16 other sites making such claims. We also observed 7 sites displaying a graphical privacy seal next to their privacy policy, despite none of them using the

[13] Indeed, this is a vulgar word in German, making it particularly insulting to a substantial portion of Facebook's users.

> *At Badoo your privacy is of paramount importance. As the custodians of your personal information, we have developed this policy to ensure that your privacy is always protected while you are using the Badoo network.* —Badoo

> *Hyves consists of a network of friends. We deal with your information as you would expect from friends. So Hyves takes your privacy very seriously and will deal with your information with due care.* —Hyves

> *We have a pretty simple privacy policy. We are reasonably sure this won't annoy anyone.* —Last.fm

> *It is Buzznet's policy to respect the privacy of Members.* —Buzznet

Figure 10: Examples of self-promotion within privacy policies.

seal on their main signup page to convey the quality of the privacy policy, as the seals are intended. In Section 5.5, we report the lack of correlation between these privacy claims and good privacy practices.

5 Data Analysis

Viewing our data as a whole, we wish to infer which factors are correlated with good privacy practices in social networking sites. This is a complicated question because it is difficult to exactly answer what constitutes "good practice." For example, an increase in privacy controls available may be seen as good to a certain point, but usability problems may arise from overly complex privacy setting pages [87].

Despite these difficulties, we defined and computed a synthetic privacy score according to the formulae explained in Appendix A. We will use this privacy score to make broad inferences about a site's privacy practices. This privacy score included three subscores summarising a site's data collection practices (see Appendix A.1.1), privacy control interface (see Appendix A.1.2), and privacy policy (see Appendix A.1.3). We deducted points for unnecessary data collection, awarded points for privacy-enabling features and also for accessibility and usability of the privacy policy.

To compare the privacy practices of a site with the site's overall functionality (see Appendix A.2), we defined an additional functionality score which awarded points for the number of non-privacy features implemented by a site. This score awarded points for providing features such as photo uploading and tagging, profile commenting, event streaming, and support for third-party applications.

The privacy and functionality scores for each site are shown in Table 7. Examining the overall privacy score, we found Bebo, LinkedIn, and GaiaOnline to have the overall best privacy practices, while Badoo, CouchSurfing, and myLife scored the lowest. Using our functionality score, we found Facebook, MySpace, and Windows Live Spaces to be the most feature-rich sites, while Twitter implemented the fewest features.

5.1 Privacy vs. Functionality

We found only a non-significant positive relationship between the functionality score and privacy score (Figure 11, left). However, there is a pronounced relationship between a site's general functionality and its privacy-specific functionality. A correlation between the functionality score and the privacy control score yields a positive regression coefficient of $r = 0.50$ at $p = 0.0003, N = 45$ as determined by a t-test. Sites that provide more functionality in general also offer more advanced features and support for configuring data sharing. Yet, this is in fact an inherited effect since general-purpose sites, which provide

Site	1 – Data Collection Score	Privacy Control Score	Privacy Policy Score	Privacy Score	Function-ality Score
Badoo	.33	.07	.33	**.23**	.40
Bahu	.24	.22	.43	**.35**	.50
Bebo	.62	.44	.57	**.70**	.60
BlackPlanet	.29	.26	.54	**.46**	.50
BuzzNet	.29	.22	.43	**.37**	.60
Classmates.com	.33	.22	.63	**.51**	.30
CouchSurfing	.14	.30	.26	**.26**	.30
CyWorld	.14	.47	.50	**.51**	.50
Eons	.24	.36	.48	**.46**	.50
Experience Project	.81	.19	.30	**.44**	.30
Facebook	.10	.61	.41	**.53**	.90
Flixster	.33	.26	.48	**.44**	.40
Friendster	.29	.30	.48	**.44**	.60
Gaia Online	.81	.44	.46	**.69**	.30
Habbo	.81	.37	.48	**.66**	.50
hi5	.43	.32	.43	**.48**	.70
Hyves	.29	.41	.41	**.47**	.70
Imbee	.05	.37	.57	**.46**	.30
Imeem	.71	.15	.57	**.55**	.50
Impulse	.43	.34	.13	**.30**	.30
Kaioo	.57	.15	.46	**.43**	.20
Last.fm	1.00	.22	.48	**.64**	.40
LinkedIn	.52	.39	.67	**.70**	.50
LiveJournal	.48	.60	.37	**.62**	.50
meinVZ	.38	.41	.65	**.65**	.40
MocoSpace	.52	.30	.43	**.49**	.30
Multiply	.05	.36	.39	**.34**	.40
MyLife	.29	.07	.43	**.28**	.30
MySpace	.29	.41	.43	**.48**	.80
MyYearbook	.24	.44	.17	**.33**	.70
NetLog	.52	.30	.35	**.44**	.60
Nexopia	.33	.22	.46	**.40**	.30
Ning	.52	.41	.48	**.59**	.70
Orkut	.43	.35	.46	**.51**	.70
PerfSpot	.19	.63	.48	**.61**	.60
Plaxo	.29	.44	.57	**.58**	.40
SkyRock	.38	.11	.39	**.31**	.40
Sonico	.00	.33	.37	**.30**	.30
Tagged	.24	.22	.35	**.30**	.60
Twitter	.81	.26	.30	**.49**	.10
Viadeo	.43	.15	.50	**.41**	.20
Windows Live Spaces	.33	.47	.50	**.58**	.80
Xanga	.76	.48	.37	**.65**	.50
XING	.24	.37	.57	**.52**	.30
Yonja	.57	.33	.37	**.49**	.40

Table 7: Privacy and Functionality Scores. See Appendix A for the calculation details). In this table, the Data Collection Score is inverted and normalised to span $[0, 1]$.

functionality	privacy score	
score	$\leq$ avg	$>$ avg
$\leq$ avg	13	9
$>$ avg	9	14
significance	$p = 0.24$	

	privacy score		priv. control score	
category	$\leq$ avg	$>$ avg	$\leq$ avg	$>$ avg
gen. purpose	14	15	10	19
niche	8	8	11	5
significance	$p = 1.00$		$p = 0.03$	

Figure 11: There is a positive, yet not significant relationship between functionality and privacy as revealed by Fisher's exact test, 2-tailed on the contingency tables between a site's functionality score and its privacy score (left) (data z-transformed and dichotomised by above / below average partition). General-purpose and niche sites cannot be differentiated based on their privacy practices (middle), but general-purpose sites offer more complete privacy settings and better support for configuring them (right). $N = 45$.

	Alexa rank		user count	
privacy score	$\leq$ med	$>$ med	$\leq$ med	$>$ med
$\leq$ avg	15	7	15	7
$>$ avg	7	16	8	15
significance	$p = 0.02$		$p = 0.04$	

	privacy score		policy length	
age	$\leq$ avg	$>$ avg	$\leq$ avg	$>$ avg
$\leq$ avg	16	10	17	8
$>$ avg	6	13	7	12
significance	$p = 0.07$		$p = 0.07$	

Figure 12: Larger and more popular sites as well as more mature sites have significantly better overall privacy protection and they feature longer privacy policies, as revealed by Fisher's exact test, 2-tailed on the contingency tables, data z-transformed. (Note that a lower rank means more popularity.) The privacy score increasing with age cannot be attributed to one single privacy subscore: there is no significant relationship between a site's age and its data collection, privacy policy or privacy control subscores. $N = 44$ for the privacy policy length, $N = 45$ otherwise.

better privacy controls (Figure 11, right) have a significantly higher functionality score than niche sites ($p = 0.01$).

5.2 Privacy vs. Site Age

We find a positive relationship between the age of a site (the time elapsed since it went online) and its privacy score. Sites that have been in existence for a longer time also have a significantly longer privacy policy in terms of word count, which can be explained by a (reactive or pro-active) privacy policy engineering process (Figure 12 right). The lack of (negative) relationship between functionality score and privacy score indicates that network operators fail to exploit their users' willingness to give up more privacy when they receive more benefits in return (discussed further in Section 6.4).

5.3 Privacy vs. Size

Similarly, the resource constraints of the social network operator give an economic explanation for our finding that P3P is implemented more often among larger sites (Figure 13). One can expect that bigger companies can more easily devote resources to deploying P3P policies. Unlike the mere presence of written privacy policies, the implementation of P3P policies is not mandated by law. As such, an operator who has invested in deploying a P3P policy has undertaken measures towards privacy enhancement beyond the required minimum. Similarly, more popular sites (by traffic rank and by user count) have an overall higher privacy score (Figure 12, left).

P3P deployed	user count $\leq$ average	 $>$ average	Alexa rank $\leq$ median	 $>$ median
yes	7	7	7	7
no	23	6	15	16
significance	$p = 0.08$		$p = 1.00$	

Figure 13: P3P policies are deployed more often on sites with above average user count ($N = 43$). However, there is no relationship between a site's popularity in terms of Alexa count and its P3P deployment ($N = 45$). p-values by a two-tailed Fisher's exact test.

5.4 Privacy vs. Growth Rate

Our sample provides evidence that privacy-enhanced sites have grown ahead of the market lately. The privacy score is positively associated with both the three-month change in traffic rank and the three-month change in page views. Similarly, the privacy control score is positively associated with the change in page views but negatively with the change in traffic rank, with only the latter relationship being significant, though ($p = 0.08$). It is possible that both phenomena may have a common cause such as the market concentrating on big sites with extensive configuration possibilities.

It is noteworthy that sites which promote on privacy are falling behind with respect to those sites which do not promote on privacy. Sites promoting on privacy have a weakly significant below-average traffic rank increase ($p = 0.10$). Implications of this are discussed further in Section 6.1.1.

5.5 Privacy Promotion and Claims vs. Actual Privacy Practices

A site's privacy claims do not necessarily indicate good privacy practices. We tested for a relationship between the privacy score and its constituent subscores with a site promoting on privacy and vaunting its data protection in the privacy policy. No significant relationship could be found between embellished claims in the privacy policy and actually good practices as captured by the privacy scores. On the contrary, sites that promoted privacy on their signup pages have a below-average privacy score ($p = 0.11$). Still, there is a weak positive relationship between the quality of a privacy policy and the existence of promotional arguments related to data protection ($p = 0.19$).

We conclude that sites mentioning good privacy practice during the signup phase actually have less favourable privacy practices, but they are well communicated in the privacy policy. These results can be interpreted as being similar to the adverse selection effect of privacy seals for general websites [32], or perhaps as the supply side analogy to the discrepancy between stated and actual privacy preferences on the demand side of the social networking market [8].

6 Economic Models

The diversity we found in the privacy practices across the sites indicates there are no universal rules for privacy in social networking. The market continues to be fluid and experimental, with some of the variation in privacy practices surely due to irrational decisions by site implementers. However, we have analysed the data and found it supports several compelling models for why poor privacy may be a rational choice for social network operators. In particular, we propose a novel model which explains our observed data, the privacy communication game. We will then compare this game-theoretic explanatory approach with other economic models traditionally applied to privacy design choices.

6.1 The Privacy Communication Game

We propose a novel model to explain the varying levels of privacy-related advertising within a single site, taking into account heterogeneous privacy preferences in the user population, and the temporal dynamics of privacy concerns. We call this model the *privacy communication game*.

In our model, different users have different privacy concerns and the social network's strategy can be seen as an attempt to optimise its interaction with each group. Previous research has provided evidence that Web users can be divided into three groups based on privacy concerns: the *marginally concerned*, the *pragmatic majority*, and the *privacy fundamentalists* [6], a taxonomy originally due to Westin. The predominant group of users, the *pragmatic majority* claims when asked to be interested in privacy but has been shown in previous studies to forget about privacy when given an attractive service [6] or monetary rewards such as discounts [79].

In parallel, it has also been shown that providing more assurance of privacy can actually make non-fundamentalists less comfortable than simply ignoring privacy [57]. However, privacy fundamentalists care deeply about privacy, and may actively investigate a site and complain to non-fundamentalists if they are dissatisfied with a site. A successful site will therefore play a game of minimising the concerns of the fundamentalists while simultaneously minimising the awareness of privacy for the non-fundamentalists.

Expressed slightly more formally, the action space for the social network operator in the privacy communication game is {communicate, hide}. There are two categories of user, namely {non-fundamentalist, fundamentalist}. All users must choose between {sign up, cancel}, while the fundamentalists will also choose between {complain, silence}. Non-fundamentalists are inclined towards "sign up" when seeing "hide"; fundamentalists are inclined towards "cancel" and "complain" when seeing "hide" and vice versa when seeing "communicate".

Because the operator is inclined towards opposite strategies for the two groups of users, it can improve its outcomes by filtering the two groups based on observed signals about users' privacy preferences and then discriminating its strategy based on the user's type. This is in some sense similar to the practice of price discrimination, as the site operator aims to serve both groups of customers in a dedicated way.

A more complex model would account for third parties such as journalists who can more strongly influence the public [60]. Eventually, only privacy negotiations with individualised communication strategies based on non-clonable signals will enable the service provider to choose individually optimal privacy strategies and to take the corresponding communication actions.

The following subsections derive the network operator's optimal strategy in the privacy communication game and relate it to our empirical evidence.

6.1.1 Reducing Privacy Salience

When facing non-fundamentalist users, the goal of the network operator is to encourage not just sign-up but also disclosure of information. Since social networks are more valuable to each user the more of their friends' data is available, operators may seek to create an environment where people feel free to disclose their data, which for non-fundamentalists is best achieved by making minimal reference to privacy.

Talking about privacy, even in the context of underlining the site's positive privacy features, may have negative consequences for the social networking operator because the mere mention of data protection raises concerns amongst the visitors. This phenomenon is known as *privacy salience*, or privacy-priming. Experiments have shown that providing strong privacy assurance can actually make people less likely to disclose personal information than if none were provided [57]. Similarly, a study on P3P browsers found that users exposed to explicit privacy information reported higher privacy concerns afterwards [40]. Many users taking part in a survey about privacy on social networks were found to have restricted their visibility settings after taking the survey [8].

Due to privacy salience effects, even promoting positive privacy practices might actually fan fears and drive customers away or reduce their willingness to reveal personal information. This would have a negative impact on the valuation of the network by its two most important customer groups: users and advertisers. Ceteris paribus, a user of the site will perceive a the network as less useful when the amount of social information for viewing is decreasing—for instance due to users not entering personal information due to privacy concerns. For advertisers, less complete profiles limit the ability for targeted advertising.

This may explain the behaviour of not promoting on privacy (Section 4.2.4) and minimising mention of a site's privacy policy during sign-up (Section 4.3). Social networks have another powerful tool to decrease privacy salience, which is to showcase other users who have posted photos and other personal behaviour, making this behaviour seem normal and safe (Section 4.2.2). This is corroborated by evidence from previous studies, which suggest that the majority of users can be enticed to enter more personal data by an animated character requesting it, or by framing the data input as situationally acceptable [57, 79].

Additionally, surfacing privacy concerns can be mitigated proactively by establishing trust with users without mentioning privacy. User studies have found that the quality and professionalism of a site is more effective in establishing trust then the contents of a privacy policy or the existence of privacy indicators [18]. This may explain survey results in the case of MySpace and Facebook, two sites mainly differing by screen design at first site, which find that Facebook is strongly trusted by its users [8, 48] more so than MySpace [31]. In our study, Facebook reached a privacy score of 0.53 compared to MySpace's 0.48, only coming out slightly ahead. The extra trust in Facebook may represent Facebook's cleaner and more consistent layout rather than its privacy practices.

6.1.2 Discouraging Privacy Fundamentalists

Fundamentalists make up a small portion of the market (estimated between 17% [6, 25] and 30% [79]), thus their participation may not be crucial for a social network's success, in particular because they are the least likely customers to begin with. Initial growth of a networking site will be created by less privacy-concerned early adopters. Individuals with strong privacy beliefs are significantly less likely to use social networks, as indicated by surveys [8], after they feel compelled to because their friends have already joined [26].

Most importantly, though, they may be less valuable or even have negative value as customers because of their privacy-conscious actions on a site. This has opportunity costs in that fundamentalists will upload less personal information, which is correlated both to having fewer friends on the site and using it less frequently [54, 48]. This makes these users less valuable for targeted advertising (we conjecture they are also likely to click on advertising links). There may also be indirect costs, however, such as the existence of fundamentalists with limited profiles or strict privacy settings raising the privacy salience of non-fundamentalists. Direct costs accrue naturally to the network operator from providing a typically free service.

The undesirability of privacy fundamentalists as social networking users may explain several trends we noticed where sites seem to avoid simple privacy practices that seem relatively cheap. For example, the poor deployment of TLS authentication and encryption (Section 4.6.1), the failure to implement P3P (Section 4.7.5), and the requirement of real names and gender (Section 4.4.2) are all likely to deter privacy fundamentalists, despite these being relatively small changes to make to the site. Similarly, there are often features which are not supported for users with strict privacy settings. Two Facebook users who both make their profiles unsearchable are given no support to become friends on the network [22]. These observations may reflect a rational choice to discourage privacy fundamentalists from joining.

6.1.3 Reducing Privacy Criticism

While fundamentalists make up a small enough population that the network may not wish them to join, they may exert power beyond their numbers by complaining to non-fundamentalists, ruining the network's attempt to minimise privacy salience. Indeed, even small, advantageously placed groups can influence the opinion in networks: fundamentalists may in fact be bloggers or journalists who wield a disproportionate influence over other users' opinions of the site [60]. Thus, the network it is strongly inclined to reduce their criticism. Another important class of privacy fundamentalists may be parents, who may not use the service themselves but are afraid of their children's online activities. It has been shown, for example, that people are consistently more afraid of privacy threats to their own children than they are to themselves [8].

As a result, while access to the legally-required privacy policies is veiled from non-fundamentalists, it is in the service provider's own interest to address privacy concerns to fundamentalists who may actually reach the documents and incorporate it into their decision to establish trust with the site [16]. We recall that, in addition to the decision whether to join or not to join, the fundamentalists potentially complain. This could explain the frequency with which operators vaunt their good privacy practices within their privacy policies, while not making such claims elsewhere on the site (Section 4.7.6). A particularly interesting manifestation of this strategy are (paid) privacy seals that are embedded in the privacy policy but not posted on the main pages of the site.

Similarly, social networking sites frequently make strong claims about their privacy practices when confronted with critical media attention due to major scandals. For example, in February an American teenager was caught soliciting naked pictures of under-age male Facebook users for blackmail purposes. In a press release responding the story, Facebook's first sentence was "Facebook seeks to provide a safe and trusted environment by offering users industry-leading tools that control access to their information..." [50]. This can be seen as another consequence of the privacy communication game, as Facebook realises it needs to strongly promote its privacy practises to concerned users reading news articles.

This quote also points to the deployment of overly-complicated privacy settings with open defaults as a rational strategy for reducing privacy complaints while still minimising salience. We frequently observed open-by default settings (Section 4.5.3), which is a good choice because most users will not adjust their privacy settings [53, 48, 8, 22, 35, 21]. We also observed many cases of privacy controls which we considered too numerous or confusing to be practical (Section 4.5.4). Deploying such settings may be optimal because it will prevent non-fundamentalists from managing their privacy, while still giving fundamentalists the control they desire given sufficient effort to understand the interface.

6.1.4 Evolution of Communication

Finally, we propose within our privacy discrimination model that a site's optimal strategy may evolve over time as its user base changes. It can be expected that non-fundamentalists will dominate the early adopters of a social network. This has been found by ethnographic studies, as more privacy-concerned individuals report that they only join a social network when they feel compelled to do so after many of their friends have joined [26, 42]. Similarly, privacy fundamentalists, particularly journalists, may be less inclined to complain about newer sites with lower membership, focusing on major players instead. Individual users also reported that their privacy concerns increased over time when using a network [26], suggesting that the user base may inherently drift towards privacy fundamentalism as time passes.

Speculatively, an optimal strategy for a network may, therefore, be to begin with no privacy controls to minimise privacy salience and encourage growth, while slowly implementing privacy features as it ages and the user base complains, or mass media criticises unfavourable data protection mechanisms. This may explain the common perception of social networks as following a "functionality first" paradigm, which Facebook's CEO acknowledged by stating that "growth is primary" in the industry[92]. We found evidence for this in the strong correlation of improved privacy practices in older networks in our survey (Figure 12).

6.2 The Effects of Lock-in

Lock-in is an entrenched feature of the market for social networks, with users facing high-switching costs to create accounts on competitive networks. In addition the cost of learning a new interface, users have been found to invest significant amounts of time in building up their profiles, which is lost if the user changes networks [48, 26]. Previously, it has been argued that lock-in is an endemic problem in security applications which harms the quality of products on the market [58]. The same model may apply to social networking accounts, as lacking data portability or data extraction prevention make it impossible for a user to move his data out and to a new network if it hypothetically offered better privacy.

This theory is supported by our survey, which found very little evidence of portability of profiles between sites. No site which we studied offered any interface for exporting one's profile data, friendship links, or photos in a simple way.

We also found strong evidence that sites attempt to erode their competitors' lock-in advantages by offering to automatically retrieve friends from a user's email inbox, making it easier to get a new account started (Section 4.4.3). Smaller social-networking sites could potentially request a user's old account from a competitive site to retrieve profile information, but this is against most sites' terms of use and has already led to at least two lawsuits: Facebook sued startup Power.com in January for allowing users to enter their Facebook login details and then fetching their account data, after similarly suing to shut down Google's FriendConnect service in May 2008 [14]. This demonstrates that sites are aware of the lock-in they possess and are actively fighting to maintain it.

The OpenSocial project [3] has been started to promote interoperability between sites. Seven of the sites surveyed implement OpenSocial applications, yet only Ning made any mention of this fact and none of the sites implement the project's goal of allowing users to take their profile data between sites. It is telling that sites have embraced OpenSocial to prevent application developers from being locked-in to one site's platform and ensure a large number of applications are available, but have avoided using it to actually allow users to more freely move between sites.

Thus, most users are locked into their current social network, meaning sites are primarily competing for the sign-up of new users. This is particularly problematic for privacy advocates. First, most new users have little data uploaded and thus their privacy is less of a concern, making data protection less of a selling point for a new account. Second, it can be difficult to assess the full spectrum of privacy controls before a significant amount of data is uploaded, thus it is even more difficult for users to asses privacy controls when considering joining. Sociological evidence may support this, as teenagers are infatuated with sharing when they first join a network, before eventually growing frustrated with the "drama" generated by people viewing their social networking pages [26]. Thus, lock-in may explain a lack of motivation for promoting privacy practices or building privacy controls, as users may be significantly locked-in to the network by the time they are concerned about privacy.

The lock-in model may be complementary to the privacy communication game model. Whilst the lock-in model captures the temporal dynamics of privacy preferences of the social network usage life-cycle and thereby explains why offering few privacy controls do not present a hurdle for joining the network, unlike the privacy communication game, lock-in effects do not account for heterogeneous privacy preferences among the user population and cannot fully explain the existing privacy practices.

6.3 Privacy as a Lemons Market

The market for privacy in social networks also fits the model of a lemons market well, as has been shown to occur in general for privacy and websites [83]. Because users have so much trouble assessing a site's privacy, sites have less incentive to provide good functionality and the market is dominated by "lemons." As argued previously, the obfuscated language employed by privacy policies deliberately deprives consumers of adequate information about what privacy is truly being offered by a website, preventing sites from needing to compete on privacy. This is consistent with our findings for social-networking privacy policies, which suffered from many usability problems (Section 4.7).

It is made stronger by our findings that privacy is not mentioned promotionally (Section 4.2.4), P3P—a potential remedy against information asymmetry—is rarely enabled (Section 4.7.5), and privacy controls are excessively numerous and confusing (Section 4.5.4). Moreover, we found that promotional privacy claims were inversely correlated with superior privacy practices (Section 5.5), meaning users are in fact receiving misinformation. For these reasons, it is difficult for end users to adequately assess the privacy functionality of a social networking site. Indeed, in our evaluations it typically took around one hour just to collect rote data on privacy features offered by each site.

This model integrates with a privacy communications game well. The inability of non-fundamentalist users to distinguish between good and bad privacy further lessens the incentive for sites to promote their privacy, when doing so may raise privacy salience and have adverse effects.

6.4 Privacy Negotiations

The paradigm of *privacy negotiations* views a user's choice to use a social-networking service as a privacy trade-off, weighing the functional benefits they get from a social networking site against the privacy they have to give up in order to qualify for these benefits [82, 70]. A similar optimisation can be made to determine if and when to reveal specific data items once signed up or to determine if and when to delete information or leave the platform. There is some evidence that users may rationalise their social network use this way, some survey respondents stated that they consider giving up some personal information to be the price of a useful, free service [30].

Whilst such a utility maximisation problem can easily be stated formally, the subjective valuations associated with benefits as well as with privacy costs make a computational solution unrealistic—not withstanding systematic psychological distortions in privacy-related decision-making. In particular, the valuations need to be formed for non-monetary benefits and costs, under limited information, over expected values with small probabilities, and subject to uncontrollable externalities. Even if a user possessed all required information, the cognitive burden and finite resource one is ready to spend would make her use simple heuristics.

Regarding the economics of privacy on social networks, this resort to heuristics has two major implications. First, network operators who assume fully rational behaviour of their users may see their expectations over the users' actions unfulfilled. Optimisation procedures over privacy designs that assume a *homo economicus* are unlikely to yield successful results in practice. Second, operators may gainfully exploit the users' inability to make fully informed choices. When heuristics are used as decision rules, these can be tricked. An example is hiding bad privacy practices in the fine-print and equipping a privacy policy with a seal instead (Section 4.7.6).

In the decision heuristics, users will contrast perceived advantages with perceived disadvantages. The higher the user perceives the functional benefit, the more she is willing to provide information. The entirety of the promotional arguments a site uses to induce sign-up can be interpreted as increasing the perceived benefits in a privacy negotiations settings.

Our data suggest that social network operators do not yet strategically exploit the tradeoff between functionality and data protection as two alternative sources for a user's utility as they compete for users. We found no evidence that sites with more funcionality are able to offer less privacy, our data instead showed a weak trend in the opposite direction (Section 5.1). Nor did we observe any evidence that price discrimination with different (privacy, functionality)-bundles is implemented within individual sites. It could be argued that in the context of social networks site functionality is less important than network effects, which grow with the number of relevant peers, i.e. the number of potential contacts. However, sites more attractive by popularity or user count also exhibit a higher privacy score (Figure 12). These trends lead us to generally reject a privacy negotiations paradigm. Still, this observation does not preclude that some users may consciously perform a cost-benefit analysis before joining a site.

7 Limitations

In light of the scale and the scope of this study, some limitations should be kept in mind that apply to all phases of our study. First, the selection of sites and criteria to assess them might be improved. We have given account of our sampling criteria in Section 3. They are aimed at defining a tractable sample for which an exhaustive evaluation could be performed. While considerable effort has been made to identify all sites that fall into the operational definition, it might be possible that some sites were missed. The sample size is particularly sensitive to cut-off levels defined on user count. Due to the scarcity of resources, expanding our sample would have forced us to compromise on the depth of analysis. The authors have adopted the point of view that—at a later point in time—the sample could be expanded more efficiently in breadth than in depth, henceforth our selection of 45 sites evaluated at approximately 260 criteria each. It might be possible we missed an important evaluation criterion or metadata entry. Our choices were driven by our analysis needs, lessons from past field studies, our expectations regarding discriminatory metrics, and eagerness for conciseness and completeness. We did not attempt to evaluate some more qualitative elements, such as the usability of privacy controls or the readability of privacy policies, relying on very rough indicators like word count and number of settings instead.

Second, the evaluation process needed to be done manually which introduces inevitable human error. Fine-grained evaluation criteria with little room for interpretation and operational definitions including tool support for evaluation (for instance in determining the privacy policy word count) are intended to keep this error small. The evaluation apparatus, as described in Section 3 was kept constant as much as possible. The evaluation tasks were split among the authors on a per criteria basis rather than on a per site basis.

Third, the scores we define, the privacy score and its constituting subscores for data collection, the privacy policy, and the privacy controls, as well as the functionality score, can be debated. We consider the definitions sound by intuition and we provide statistical backup for the definitions (Cronbach's α). Other scores may be defined at the reader's discretion. In being explicit on the calculation formula (Appendix A), we enable the reader to assess the suitability of each of these scores for her or his own analyses. Equally, in making the dataset publicly available, we provide the necessary data to define any alternative score.

Fourth, the authors acknowledge that durability of the data is limited given the high mutability of the market in social networking. Even so, the value of the dataset does not only originate in being the most comprehensive snapshot. It can also be used as an historical data point in longitudinal analyses.

Fifth, our analyses and the economic models we advance as explanations for the empirical evidence might be scrutinised. By making our dataset publicly available, we encourage the community to challenge our interpretations and conclusions.

8 Conclusions

Online social networking has become an indispensable activity, and research must keep up with the phenomenon. With the mass adoption of social networking sites over the last eighteen months, a scholarly review of privacy practices "in the wild" was overdue. Given our data, we have serious concerns about the current state of affairs.

In particular, we have found strong evidence that the social networking market is failing to provide users with adequate privacy control. The market is still in an early stage of aggressive competition for users that may eventually yield to a more static and consolidated supply. Our results suggest that the naive application of utility maximisation theory fails to capture all the intricacies of the market for privacy in social networking. Experimental economics has long suggested that users' privacy-related decision-making is systematically distorted from full rationality and subject to limited information. We have found compelling evidence that a major problem is the lack of accessible information for users, encouraged by sites' strong incentives to limit privacy salience as part of the privacy communication game: the data suggests that sites may have evolved specifically to communicate differently to users with different levels of privacy concern.

Assuming that better privacy awareness and protection would be beneficial for users, regulation may be necessary in order for a privacy market to function properly. Reducing information asymmetry is an important first step, through standardised "privacy nutrition labels" [67] which can communicate privacy practices in a non-textual format to help users make more informed privacy choices. Increasing privacy salience is of critical importance. This could be achieved by requiring sites to provide clear, Web-integrated interfaces for users to see exactly what personal data of theirs is held, and exactly which parties have access to it. User access to data is a core principle of the EU Data Protection Directive, but we argue it must be far easier and more integrated into the user experience to be effective. Finally, reducing lock-in effects through mandated data portability may be necessary to increase consumer choice in social networks. In this area, regulation seems most promising and may pay off in the short run.

We also think that much more research is necessary on the dynamics of privacy in social networks. Our results hint at many promising areas for further inquiry. The privacy salience phenomenon and its role in social networking in particular needs further analysis. We are planning a user experiment to study privacy-related decisions on social networks, focusing on the role of communication and privacy-functionality trade-offs each user has to solve. Research is also needed on methods to make privacy information more understandable, and better user interfaces for configuring social network access controls. We hope that our study, along with our published dataset, will be an important starting point.

Acknowledgements

The authors would like to thank Alastair Beresford, Jonathan Anderson, and Ross Anderson for their support and feedback.

References

[1] Alexa: The Web Information Company, Mar 2009.

[2] OnGuard Online. *www.onguardonline.gov/*, 2009.

[3] OpenSocial Project. *www.opensocial.org*, March 2009.

[4] Platform for Privacy Preferences (P3P) Project. *http://www.w3.org/P3P/*, March 2009.

[5] Mark S. Ackerman. Privacy in Pervasive Environments: Next Generation Labeling Protocols. *Personal Ubiquitous Comput.*, 8(6):430–439, 2004.

[6] Mark S. Ackerman, Lorrie Faith Cranor, and Joseph Reagle. Privacy in e-Commerce: Examining User Scenarios and Privacy Preferences. In *EC '99: Proceedings of the 1st ACM conference on Electronic commerce*, pages 1–8, New York, NY, USA, 1999. ACM.

[7] Alessandro Acquisti. Privacy in Electronic Commerce and the Economics of Immediate Gratification. In *EC '04: Proceedings of the 5th ACM conference on Electronic commerce*, pages 21–29, New York, NY, USA, 2004. ACM.

[8] Alessandro Acquisti and Ralph Gross. Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. In *Privacy Enhancing Technologies – LNCS 4258*, pages 36–58. Springer Berlin / Heildelberg, 2006.

[9] Alessandro Acquisti and Jens Grossklags. Privacy and Rationality in Individual Decision Making. *IEEE Security and Privacy*, 3(1):26–33, 2005.

[10] Jonathan Anderson, Claudia Diaz, Joseph Bonneau, and Frank Stajano. Privacy Preserving Social Networking Over Untrusted Networks. *Second ACM SIGCOMM Workshop on Online Social Networks*, 2009.

[11] Annie I. Antón, Elisa Bertino, Ninghui Li, and Ting Yu. A Roadmap for Comprehensive Online Privacy Policy Management. *Commun. ACM*, 50(7):109–116, 2007.

[12] Michael Arrington. Elaborate Facebook Worm Spreading. *TechCrunch*, Aug 2008.

[13] Michael Arrington. Phishing For Facebook. *TechCrunch*, Jan 2008.

[14] Michael Arrington. Facebook Defends Its Turf, Sues Power.com. *TechCrunch*, 2 Jan 2009. eMarketer.

[15] 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.

[16] Gaurav Bansal, Fatemeh Zahedi, and David Gefen. The Moderating Influence of Privacy Concern on the Efficacy of Privacy Assurance Mechanisms fo Building Trust: A Multiple Context Investigation. In *ICIS 2008: International Conference on Information Systems*, 2008.

[17] David Barroso, Richard Barle, Patrice Chazerand, Melissa de Zwart, Jeroen Doumen, Slawomir Gorniak, Mateusz Kaźmierczak, Markku Kaskenmaa, Daniel Benavente López, Adam Martin, Ingo Naumann, Ren Reynolds, Janice Richardson, Christian Rossow, Anna Rywczyoska, and Michael Thumann. Security and Privacy in Massively-Multiplayer Online Games and Social and Corporate Virtual Worlds. Technical report, ENISA - European Network and Information Security Agency, 2008.

[18] France Belanger, Janine S. Hiller, and Wanda J. Smith. Trustworthiness in Electronic Commerce: the Role of Privacy, Security, and Site Attributes. *The Journal of Strategic Information Systems*, 11(3-4):245 – 270, 2002.

[19] Rosemary Bennett. Plea to ban Employers Trawling Facebook. *The Times*, 25 Mar 2008. The Times.

[20] Bonneau, Joseph. New Facebook Photo Hacks. *http://www.lightbluetouchpaper.org/2009/02/11/new-facebook-photo-hacks/*, 2009.

[21] Bonneau, Joseph and Anderson, Jonathan and Danezis, George. Prying Data out of a Social Network. In *ASONAM 2009 : Advances in Social Networks Analysis and Mining*, 2009.

[22] Bonneau, Joseph and Anderson, Jonathan and Stajano, Frank and Anderson, Ross. Eight Friends Are Enough: Social Graph Approximation via Public Listings. In *SNS '09: Proceeding of the 2nd ACM Workshop on Social Network Systems*, 2009.

[23] Sonja Buchegger and Anwitaman Datta. A Case for P2P Infrastructure for Social Networks - Opportunities and Challenges. In *Proceedings of WONS 2009, The Sixth International Conference on Wireless On-demand Network Systems and Services*, Snowbird, Utah, USA, February 2-4, 2009.

[24] Duen Horng Chau, Shashank Pandit, Samuel Wang, and Christos Faloutsos. Parallel Crawling for Online Social Networks. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 1283–1284, 2007.

[25] Cranor, Lorrie F., Joseph Reagle, and Mark S. Ackerman. Beyond Concern: Understanding Net Users' Attitudes about Online Privacy. Technical Report TR 99.4.3, AT&T Labs, 1999.

[26] danah boyd. Why Youth (Heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life. *Youth, Identity, and Digital Media*, pages 119–142, 2008.

[27] danah boyd and Nicole Ellison. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 2007.

[28] George Danezis and Bettina Wittneben. The Economics of Mass Surveillance and the Questionable Value of Anonymous Communications. *WEIS: Workshop on the Economics of Information Security*, 2006.

[29] Donath, J. and boyd, d. Public Displays of Connection. *BT Technology Journal*, 22(4):71–82, 2004.

[30] Catherine Dwyer. Digital Relationships in the "MySpace" Generation: Results From a Qualitative Study. In *HICSS '07: Proceedings of the 40th Annual Hawaii International Conference on System Sciences*, page 19, Washington, DC, USA, 2007. IEEE Computer Society.

[31] Catherine Dwyer, Starr Roxanne Hiltz, and Katia Passerini. Trust and Privacy Concern within Social Networking Sites: A Comparison of Facebook and MySpace. In *Proceedings of the Thirteenth Americas Conference on Information Systems*, 2007.

[32] Benjamin Edelman. Adverse Selection in Online "Trust" Certifications. *WEIS: Workshop on the Economics of Information Security*, 2006.

[33] Serge Egelman, Janice Tsai, Lorrie Faith Cranor, and Alessandro Acquisti. Timing is Everything?: the Effects of Timing and Placement of Online Privacy Indicators. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 319–328, New York, NY, USA, 2009. ACM.

[34] Adrienne Felt. Defacing Facebook: A Security Case Study. *www.cs.virginia.edu/felt/fbook/facebook-xss.pdf*, 2007.

[35] Adrienne Felt and David Evans. Privacy Protection for Social Networking Platforms. *Workshop on Web 2.0 Security and Privacy*, 2008.

[36] Adrienne Felt, Pieter Hooimeijer, David Evans, and Westley Weimer. Talking to Strangers without Taking their Candy: Isolating Proxied Content. In *SocialNets '08: Proceedings of the 1st workshop on Social network systems*, pages 25–30, New York, NY, USA, 2008. ACM.

[37] A. Finder. For Some, Online Persona Undermines a Resume. *The New York Times*, Jun 2006.

[38] Frankowski, Dan and Cosley, Dan and Sen, Shilad and Terveen, Loren and Riedl, John. You Are What You say: Privacy Risks Of Public Mentions. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 565–572, New York, NY, USA, 2006. ACM.

[39] Dan Frommer. What a Nigerian Facebook Scam Looks Like. *The Business Insider*, Jan 2009.

[40] Julia Gideon, Lorrie Cranor, Serge Egelman, and Alessandro Acquisti. Power Strips, Prophylactics, and Privacy, Oh My! In *SOUPS '06: Proceedings of the second symposium on Usable privacy and security*, pages 133–144, New York, NY, USA, 2006. ACM.

[41] Minas Gjoka, Michael Sirivianos, Athina Markopoulou, and Xiaowei Yang. Poking Facebook: Characterization of OSN Applications. In *WOSP '08: Proceedings of the first workshop on Online social networks*, pages 31–36, New York, NY, USA, 2008. ACM.

[42] Tabreez Govani and Harriet Pashley. Student awareness of the privacy implications when using facebook. *http://lorrie.cranor.org/courses/fa05/tubzhlp.pdf*, 2005.

[43] Saikat Guha, Kevin Tang, and Paul Francis. NOYB: Privacy in Online Social Networks. In *Workshop on Online Social Networks – WOSN 2008*, pages 49 – 54, 2008.

[44] Seda Gürses, Ramzi Rizk, and Oliver Günther. Privacy Design in Online Social Networks: Learning from Privacy Breaches and Community Feedback. In *ICIS 2008: Proceedings Twenty Ninth International Conference on Information Systems*. ACM, 2008.

[45] Il-Horn Hann and Kai-Lung Hui and Tom S. Lee and I. P. L. Png. Online Information Privacy: Measuring the Cost-Benefit Trade-off. *23rd International Conference on Information Systems*, 2002.

[46] Tom Jagatic, Nathaniel Johnson, Markus Jakobsoon, and Filippo Menczer. Social Phishing. *Communications of the ACM*, 50(10):94, 2007.

[47] Jessi Hempel. Is Facebook Losing Its Glow? *Fortune Magazine*, April 2009.

[48] Harvey Jones and Jose Hiram Soltren. Facebook: Threats to privacy. *http://web.mit.edu/jsoltren/www/facebook.pdf*, 2005.

[49] K.C. Jones. Facebook Admits Sexual Assault Suspect Used Site. *Information Week*, 6 Feb 2009.

[50] Kincaid, Jason. Wakeup Call: Facebook Isn't a Safe Haven. *TechCrunch*, Feb 2009.

[51] Ethan Kolek and Daniel Saunders. Online Disclosure: An Empirical Examination of Undergraduate Facebook Profiles. *National Association of Student Personnel Administrators journal*, 2008.

[52] Aleksandra Korolova, Rajeev Motwani, Shubha U. Nabar, and Ying Xu. Link Privacy in Social Networks. In *CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management*, pages 289–298, 2008.

[53] Balachander Krishnamurthy and Craig E. Wills. Characterizing Privacy in Online Social Networks. In *WOSN: Workshop on Online Social Networks*, pages 37 – 42, 2008.

[54] Cliff A.C. Lampe, Nicole Ellison, and Charles Steinfield. A Familiar Face(book): Profile Elements as Signals in an Online Social Network. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 435–444, New York, NY, USA, 2007. ACM.

[55] Jack Lindamood and Murat Kantarcioglu. Inferring Private Information Using Social Network Data. *WOSN: Workshop on Online Social Networks*, 2008.

[56] Heather Richter Lipford, Andrew Besmer, and Jason Watson. Understanding Privacy Settings in Facebook with an Audience View. In *1st Conference on Usability, Psychology, and Security*. USENIX Association, 2008.

[57] George Loewenstein. Keynote Speech: Searching for Privacy in all the Wrong Places: A Behavioral Economics Perspective on Individual Concern for Privacy. *WEIS 07: The Seventh Workshop on the Economics of Information Security*, June 2007.

[58] Tom Lookabaugh and Douglas Sicker. Security and Lock-in. *WEIS '03: Proceedings of the Third Workshop on the Economics of Information Security*, 11 April 2003.

[59] Matthew M. Lucas and Nikita Borisov. FlyByNight: Mitigating the Privacy Risks of Social Networking. In *WPES 08 - Workshop on Privacy in the Electronic Society*, page 1, 2008.

[60] M.E. McCombs and D.L. Shaw. The Agenda-Setting Function Of Mass Media. *Public Opinion Quarterly*, 36(2):176–187, 1972.

[61] George Milne and Mary Culnan. Information Privacy: Measuring Individuals' Concerns About Organizational Practices. *Journal of Interactive Marketing*, 18(3), 2004.

[62] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and Analysis of Online Social Networks. In *IMC '07: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, pages 29–42, 2007.

[63] Shishir Nagaraja. The Economics of Covert Community Detection and Hiding. *WEIS: Workshop on the Economics of Information Security*, 2008.

[64] Arvind Narayanan and Vitaly Shmatikov. De-anonymizing Social Networks. *30th IEEE Symposium on Security & Privacy*, 2009.

[65] Nick O'Neill. 10 Privacy Settings Every Facebook User Should Know. http://www.allfacebook.com/2009/02/facebook-privacy, Feb 2009.

[66] Anthony Onwuasoanya, Maxim Skornyakov, and Jonathan Post. Enhancing Privacy on Social Networks by Segregating Different Social Spheres. *Rutgers Governor's School of Engineering and TechnologyResearch journal*, 2008.

[67] Patrick Gage Kelley and Joanna Bresee and Lorrie Faith Cranor and Robert W. Reeder. A "Nutrition Label" for Privacy. *Symposium On Usable Privacy and Security (SOUPS) 2009*, 2009.

[68] Ed Pilkington. Blackmail claim stirs fears over Facebook. *The Guardian*, 16 July 2007. The Guardian.

[69] J. C. Poindexter, Julia B. Earp, and David L. Baumer. An Experimental Economics Approach Toward Quantifying Online Privacy Choices. *Information Systems Frontiers*, 8(5):363–374, 2006.

[70] Sören Preibusch. Implementing Privacy Negotiations in E-Commerce. *Lecture Notes in Computer Science*, 3841:604–615, 2006.

[71] Sören Preibusch and Alastair R. Beresford. Privacy-Preserving Friendship Relations for Mobile Social Networking. *W3C Workshop on the Future of Social Networking*, 2009.

[72] David Randall and Victoria Richards. Facebook Can Ruin Your Life. And So Can MySpace, Bebo... *The Independent*, 10 Feb 2008. The Independent.

[73] Joseph Reagle and Lorrie Faith Cranor. The Platform for Privacy Preferences. *Commun. ACM*, 42(2):48–55, 1999.

[74] David Rosenblum. What Anyone Can Know: The Privacy Risks of Social Networking Sites. *IEEE Security & Privacy Magazine*, 5(3):40, 2007.

[75] Tracy Samantha Schmidt. Inside the Backlash Against Facebook. *Time Magazine*, 6 Sep 2006.

[76] Jessica Shepherd and David Shariatmadari. Would-Be Students Checked on Facebook. *The Guardian*, 11 Jan 2008. The Guardian.

[77] Andrew Simpson. On the Need for User-Defined Fine-Grained Access Control Policies for Social Networking Applications. In *SOSOC '08: Proceedings of the workshop on Security in Opportunistic and SOCial networks*, pages 1–8, New York, NY, USA, 2008. ACM.

[78] H. Jeff Smith and Sandra J. Milberg. Information Privacy: Measuring Individuals' Concerns About Organizational Practices. *MIS Q.*, 20(2):167–196, 1996.

[79] Sarah Spiekermann, Jens Grossklags, and Bettina Berendt. E-privacy in 2nd Generation E-commerce: Privacy Preferences Versus Actual Behavior. In *EC '01: Proceedings of the 3rd ACM conference on Electronic Commerce*, pages 38–47, New York, NY, USA, 2001. ACM.

[80] Louise Story and Brad Stone. Facebook Retreats on Online Tracking. *The New York Times*, 30 Nov 2007.

[81] Henny Swan. Social networking across devices: opportunity and risk for the disabled and older community. *W3C Workshop on the Future of Social Networking*, 2009.

[82] Hal R. Varian. Economic Aspects of Personal Privacy. *Topics in Regulatory Economics and Policy*, 2002.

[83] Tony Vila, Rachel Greenstadt, and David Molnar. Why We Can't Be Bothered to Read Privacy Policies: Models of Privacy Economics as a Lemons Market. In *ICEC '03: Proceedings of the 5th International Conference on Electronic commerce*, pages 403–407, New York, NY, USA, 2003. ACM.

[84] W3C, Mobile Web Best Practices Working Group, Checker Task Force. W3C mobileOK Checker, 2009.

[85] E.J. Westlake. Friend Me if You Facebook: Generation Y and Performative Surveillance. *TDR: The Drama Review*, 52(4):21–40, 2008.

[86] T. Wham. Transcript of the FTC Workshop on Information Privacy: Measuring Individuals' Concerns about Organizational Practices. *http://www.ftc.gov/bcp/workshops/infomktplace/transcript.htm*, 2001.

[87] Alma Whitten and J. D. Tygar. Why Johnny can't encrypt: A usability evaluation of PGP 5.0. In *8th USENIX Security Symposium*, 1999.

[88] Debra Aho Williamson. Social Networking Ad Spending. *eMarketer*, 13 May 2008. eMarketer.

[89] XING AG. Press release: XING AG increases revenues by 80 percent and continues to grow profitably, 2009.

[90] Wanhong Xu, Xi Zhou, and Lei Li. Inferring Privacy Information via Social Relations. *International Conference on Data Engineering*, 2008.

[91] Elena Zheleva and Lise Getoor. To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles. *WWW: The International World Wide Web Conference*, 2009.

[92] Mark Zuckerberg and Holger Schmidt. Facebook CEO Mark Zuckerberg: Our Focus is Growth, Not Revenue. *Frankfurter Allgemeine Zeitung / FAZ.NET*, 8 Oct 2008.

A Imputed Metrics

We provide here a complete description of the synthetic scores we calculated for privacy and functionality, used in our analysis in Section 5. We stress that these scores do not represent a complete evaluation of the sites, but are only designed to allow analysis of trends within the data. The complete scores can be viewed in Table 7.

A.1 Privacy Score

The privacy score aggregates three subscores: one representing the amount of data collected, one representing the extend of privacy controls provided, and one representing the quality of the privacy policy provided. The overall privacy score, is the arithmetic mean of the three subscores 'data collection score' (DCS), 'privacy control score' (PCS), and 'privacy policy score' (PPS), according to the following formula. To ensure an equal weighting, we normalised each subscore based on the highest-observed and lowest-observed value for that subscore. The data collection score is reversed, since a higher DCS is less privacy-protecting.

$$\text{Privacy Score} = \frac{\frac{\text{PCS}-\min(\text{PCS})}{\max(\text{PCS})-\min(\text{PCS})} + \frac{\text{PPS}-\min(\text{PPS})}{\max(\text{PPS})-\min(\text{PPS})} + 1 - \frac{\text{DCS}-\min(\text{DCS})}{\max(\text{DCS})-\min(\text{DCS})}}{3}$$

A.1.1 Data Collection Subscore

The data collection score is calculated from the number of data items a site collects and the mandatoriness of each data field. The data collection score is calculated as 1pt for each required data item plus 0.5pt for each requested data item. Data items that are verified or validated are counted with +2pts. The use of such a score finds its motivation in a value of $\alpha = 0.51$ for Cronbach's alpha over the 34 evaluation criteria regarding data collection. From the definition, a higher data collection score is more privacy-invasive and a higher privacy score is more desirable from a data protection point of view. The recalibrated data collection score is mapped to the $[0, 1]$ interval based on the minimum and maximum DCS observed in our sample.

A.1.2 Privacy Control Subscore

The privacy control score represents the number of features implemented by a site which allow a user to control their privacy. In particular, it awards 1 point for most such features, as listed in Table 8. We chose not to include several features of dubious value, such as number of privacy settings available, since it is unclear at which point this lessens usability of the site. We included features for ensuring privacy and safety on the site.

A.1.3 Privacy Policy Subscore

We scored privacy policies based on their availability during the signup process, their accessibility, the information provided, and the existence of P3P version of the policies. We did not attempt subjective scoring of the quality of the privacy policies. The detailed criteria are available in Table 9.

A.2 Functionality Score

The functionality score represents the number of non-essential features implemented by a site. In particular, it awards 1 point for most features, and 2 points for allowing third party applications, considering this to be worth extra in enabling a large variety of extra functionality to be created. The criteria are listed in Table 10.

feature	**scoring**	**adjustments**
data security		
full TLS during login	+2.0pt	−1.0pt POST only
can logout from site	+1.0pt	
privacy configurability		
Friends-only visibility available	+1.0pt	+1.0pt if default setting
Network-only visibility available	+1.0pt	+1.0pt if default setting
Profile line-item ACL available	+1.0pt	+0.1pt each item up to a maximum of +3.0pts
Photo, commenting, messaging ACL	+1.0pt each	
List-based ACL	+1.0pt	
Audience view	+1.0pt	
Search view	+1.0pt	
Pre-set privacy combinations	+1.0pt	
Email settings	+1.0pt	
privacy support by site operator		
Privacy settings Help	+1.0pt	
Block User Button	+1.0pt	
Report User Button	+1.0pt	
Abuse Reporting	+1.0pt	+1.0pt telephone
Privacy Tips	+1.0pt	
Safety Tips	+1.0pt	
Parent Tips	+1.0pt	

Table 8: Privacy Control Score Criteria. Adjustments are applied to the original item scoring. The possible maximum score is 27.

feature	**scoring**	**adjustments**
accessibility		
PP acknowledgement checkbox	$+1.0$pt	
PP link presented during signup	$+1.0$pt	-0.5pt requires JavaScript and -0.25pt pop-up window
T&C link presented during signup	$+1.0$pt	-0.5pt requires JavaScript and -0.25pt pop-up window
number of links available to PP	$+0.5$pt per link	up to a maximum of $+2.0$pts
number of links available to T&C	$+0.5$pt per link	up to a maximum of $+2.0$pts
usage		
PP exists	$+1.0$pt	-1.0pt requires JavaScript
PP dated	$+1.0$pt	
mobileOK percentage score	$+0.00.. + 1.00$pt	
PP bookmarkable	$+1.0$pt	
PP printable	$+1.0$pt	
PP savable	$+1.0$pt	
PP textually structured	$+1.0$pt	
privacy statements		
Can delete personal information	$+1.0$pt	-0.5pt partial deletion only
National laws obeyed specified	$+1.0$pt	
Privacy seal	$+1.0$pt	
Collection of data from external sources	-1.0pt	
Sharing data with third parties	-1.0pt	-2.0pt not anonymised
contact details		
Contains operator email address	$+1.0$pt	
Contains operator postal address	$+1.0$pt	
privacy policy mutability		
Users notified of changes to PP	$+1.0$pt	
Delay period before changes effective	$+1.0$pt	
P3P deployment		
Full P3P policy	$+1.0$pt	-0.5pt if not correct
Compact P3P policy	$+1.0$pt	-0.5pt if not correct

Table 9: Privacy Policy Subscore Criteria. Adjustments are applied to the original item scoring. The possible maximum score is 23. (T&C: Terms and Conditions)

feature	**scoring**
meta-functionality	
Third-party applications	+2.0pt
interactivity	
Profile commenting	+1.0pt
Sending messages to other users	+1.0pt
Photo uploading	+1.0pt
Photo tagging	+1.0pt
Event streaming	+1.0pt
connectivity	
Data export protocol	+1.0pt
OpenID	+1.0pt
Linkable “profile badges”	+1.0pt

Table 10: Functionality Score Criteria. The possible maximum score is 10.