Social Networks, Gender and Friending: An Analysis of MySpace Member Profiles[1]

Mike Thelwall

School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, UK. E-mail: m.thelwall@wlv.ac.uk
Tel: +44 1902 321470 Fax: +44 1902 321478

In 2007, the social networking web site MySpace apparently overthrew Google as the most visited web site for U.S. web users. If this heralds a new era of widespread online social networking, then it is important to investigate user behaviour and attributes. Although there has been some research into social networking already, basic demographic data is essential to set previous results in a wider context and to give insights to researchers, marketers and developers. In this article the demographics of MySpace members are explored through data extracted from two samples of 15,043 and 7,627 member profiles. The median declared age of users was surprisingly high at 21, with a small majority of females. The analysis confirmed some previously reported findings and conjectures about social networking, for example that female members tend to be more interested in friendship and males more interested in dating. In addition, there was some evidence of three different friending dynamics: oriented towards close friends, acquaintances, or strangers. Perhaps unsurprisingly, female and younger members had more friends than others, and females were more likely to maintain private profiles, but males and females both seemed to prefer female friends, with this tendency more marked in females for their closest friend. The typical MySpace user is apparently female, 21, single, with a public profile, interested in online friendship and logging on weekly to engage with a mixed list of mainly female 'friends' who are predominantly acquaintances.

Introduction

In many Web 2.0 sites, pairs of members can register as 'friends', which typically gives them privileged or easier access to each other's resources or communication channels (e.g., instant messaging, email, blogging, photographs, commenting). In some sites the primary purpose of friendship is as a shortcut to finding relevant resources by starting with those of people with shared interests. In social networking sites like MySpace and Facebook, however, the primary purpose appears to be to socialise in the sense of engaging in communication for its own sake (boyd, 2008; Pew Research Center for the People & the Press, 2007). This may include exchanging messages with no real information content (known as 'phatic communion' (Malinowski, 1923), e.g., "hope you are fine") as well as discussing topics of shared interest and planning face-to-face meetings. For teenagers, it seems that their MySpace activity can be deeply embedded into their lives, for example influencing their self-esteem (Valkenburg, Peter, & Schouten, 2006). A consequence of this is that social networking is not only an interesting new form of online communication but it is also one that is important in its own right.

Whilst offline friendship has been extensively studied, social networking friendship, as a relatively recent phenomenon, is much less understood. This is a significant omission because of the importance of MySpace, which apparently eclipsed Google as the most visited web site by U.S. users at the end of 2006 (Prescott, 2007), and the key role of friendship in these sites. The research that has been completed so far (reviewed below) has given many valuable insights into how and why social networking is used; particularly in terms of teenagers. A weakness of most previous research, however, is that it has been qualitative - or mixed methods (Tashakkori & Teddlie, 1998) but reported in a predominantly qualitative fashion (e.g., boyd, 2007) - and hence, it forms conjectures rather than presenting testable evidence. This is not a criticism: qualitative and mixed method approaches are *more* valuable than quantitative research for investigating the real meaning of new culture-related

phenomena like social networking friendship. Nevertheless, quantitative methods are needed to test and confirm the insights of qualitative research, to provide method triangulation, and to produce new findings. Quantitative data is also essential to give a broad overview of the demographics of social networking for site designers, researchers and advertisers. Although some useful statistics are published by commercial market research organisations, the underlying data and its origins are typically guarded and so claims cannot be tested. For example, one press release claimed that the majority of MySpace visitors were over 35 in August 2006 (http://www.comscore.com/press/release.asp?press=1019, accessed July 24, 2007), which does not seem credible and cannot be directly checked. In this case the results were based upon monitoring the activity of 2 million consenting clients and so it seems possible that the results overestimate the ages of MySpace visitors due to children using their parents' computers.

This article presents a quantitative exploration of friendship in social networks to identify factors relating to three key properties: friendship circle size, age and gender. This narrow focus allows a detailed statistical exploration into a basic yet important aspect of the social networking friendship. MySpace is used for the analysis because this was, at the time of writing (June 2007), the most popular U.S. social networking web site.

Online and Web 2.0 friendship: Friending vs. befriending

Social networking sites are essentially web servers that allow Internet users to register, to create a personal profile and then use this profile to communicate with selected others. A social networking site member will be able to add a picture and biographic information to their profile home page. They will also be able to find and connect with other members by agreeing to become friends. Some or all of each member's friends will be listed on their profile page, along with their photographs. Friends have special privileges, such as the ability to message each other and write comments on each other's profile page. Most general social networking sites also offer other facilities, such as a blog, online photograph albums and video hosting. In contrast, some social networking sites are oriented towards more specialist services, such as news reading (Digg), photo sharing (Flickr) and bookmark sharing (del.icio.us) – for these, friendship and related activities can primarily be a "collaborative filtering" (Konstan et al., 1997) aid to information retrieval (Golder & Huberman, 2006; Lerman, 2006). For example, in Flickr, friends' pictures (friends are called 'contacts' in Flickr) may be more relevant because they include shared acquaintances, a shared hobby (e.g., birds), or have a similar artistic taste.

In this article, the term *friending* is used to connote social network friendship connections. The first mass social networking site, Friendster (which started in 2003), introduced the ability for members to register each other as 'friends'. Although this function was probably originally intended to reflect real world friendships, its use in practice was very different (Donath & boyd, 2004). Overall, individual users and groups of users probably negotiate the meaning of 'friend' in any social network site, and 'acquaintance' is probably a more accurate general description. MySpace allows differentiation amongst friends, through a small top friends select list. Although the concept of friending in social networks is apparently the glue that keeps them together, it is a complex construct and one that can cause conflict. For instance, decisions about who are in the top friend set can be traumatic and 'defriending' someone by dropping them from a friend list can result – deliberately or accidentally – in upset feelings (boyd, 2006).

Note that offline friendship should not be reified: describing someone as a friend online is merely a summary statement of personal feelings "using culturally approved terminology" (Duck, 1992, p. 33) and its meaning is dependent upon culture and time. A reasonable generalisation, however, might be that friends tend to care about each other, have some things in common and expect this to continue for at least a moderate period of time.

LiveJournal is an example of an unusual social network site because of its orientation towards blog-like journals. A LiveJournal user's friends are probably the people with journals considered to be worth reading (Fono & Raynes-Goldie, 2005). This is partly a consequence

of LiveSpace friending not needing to be reciprocal, like Flickr 'contacts', but unlike most other spaces (e.g., MySpace, Facebook, Live Spaces). As with other spaces, however, LiveJournal users have and sometimes use the ability to create semi-private content that is only available to other users designated as friends. In this sense friendship is sometimes intrinsically related to trust (Fono & Raynes-Goldie, 2005). Fono and Raynes-Goldie (2005) also mention several other connotations of friendship.

- Courtesy: Sometimes friends may be added (or friendship reciprocated in LiveJournal or Flickr) out of a sense of politeness to avoid giving offence.
- Declaration: Designating someone as a friend may be partly a public declaration of friendship.
- Nothing: Some users may see friendship as signifying nothing.
- Offline facilitator: Friendships may be convenient for communication to coordinate offline activities.
- Online community: Friendships may be genuine purely online friend-like relationships.

An additional meaning for friend, especially in MySpace is 'fan'. MySpace encourages bands to join via music.myspace.com. Registering as a friend of a musician or band in MySpace typically gives free access to some online music and news bulletins as well as pictures, videos and general information. Although MySpace friending is an equal, reciprocal relationship, in practice the fan relationship is unequal but is not differentiated from 'normal' friendship in any way. MySpace, unlike Friendster and Facebook, has made a deliberate attempt to cultivate fan relationships (boyd, 2006).

Social networking friendships have perhaps been explored most systematically, at least from a quantitative perspective, in terms of Facebook. Facebook is a general social networking site that supports text based communication between friends as well as picture sharing. In its early days it was exclusively for college students and still has this flavour although it is no longer restricted. For instance it seems that there is a class divide, at least in the U.S., between education-oriented Facebook users and predominantly non college-educated MySpace users (boyd, 2007). Despite the lack of geographic boundaries of the internet, most Facebook communication takes place between students at the same college (Golder, Wilkinson, & Huberman, 2007), and its primary role is often seen as cementing friendships that initially began offline (Ellison, Steinfield, & Lampe, 2006). Although distant friendships do exist, they don't seem to generate as much messaging traffic. In temporal terms, students seemed to fit Facebook use into their working pattern, engaging in social networking in parallel with studying, presumably due to the convenience of both tending to require an internet-connected computer (Golder et al., 2007). Some commercial research into general social networking supports the idea of social networking not being a separate activity to some extent by showing that many users maintain profiles in multiple sites and switch from one to the other to check each one consecutively (Prescott, 2007).

A U.S.-based survey of teen internet users included some questions about social networking friendship (Pew Research Center for the People & the Press, 2007). It found some gender differences, with boys claiming to use it more frequently for flirting than girls – although there are only 17% flirters overall, and the gender discrepancy may reflect differing perceptions of flirting unless a lot of gay or non-reciprocal flirting occurs. Both genders used social networking to plan with friends (91%) and about half of users tried to make new friends (boys more than girls). Young users seem to communicate very informally in MySpace, with strong swear words occurring in a majority of youth profiles (Hinduja & Patchin, in press; Thelwall, in press) and a significant minority containing evidence of illegal activities such as underage alcohol use (18% of under 18s) (Hinduja & Patchin, in press).

Objectives

The objective of this research is to identify personal factors that associate with different age ranges, friendship circle sizes and genders in MySpace. Age ranges are important since previous research and much media interest has focussed on teenage users and has discussed

how young users create their own social norms within social networking. Friendship circle size is also significant because of the hypothesised difference between offline friendship and the looser concept of friendship in social networks. Finally, previous research has identified small yet apparently significant gender differences in social networks and so this is an essential issue to explore.

Data

The raw data for this article are three samples of MySpace public user profiles. Each MySpace user has a personal identification number, and these numbers are apparently given out in sequence.

We identified approximately the last ID issued by MySpace on July 3, 2007 and for the first collection selected every 10,227th ID starting at 1,939 (a random starting point) to give a large total sample size (20,064). The profile page associated with each of these user IDs was then downloaded via the URL http://profile.myspace.com/index.cfm?fuseaction=user.viewprofile&friendid= followed by the user ID. These profiles were downloaded at a rate of 8,000 per day over three days to avoid overloading the MySpace server. This is the 'all members collection', and encompasses MySpace users from a wide variety of joining dates.

The second collection is the 'July 3, 2006 members collection' and consists solely of users who joined on July 3, 2006, as identified by trial and error through MySpace ID ranges, selecting 10,000 IDs (90,306,349 to 90,316,348). These were downloaded over four hours on July 17, 2007 starting at 2.30am central U.S. time in an attempt to access the data when most users were asleep in order to (a) minimise impact upon the MySpace servers and (b) capture comparable data. The main data set for this article is the all members collection but the July 3, 2006 members collection is used for supplementary analyses when time of joining is important.

Each profile page in each collection was automatically scanned to extract the following information.

- If the page returned an error (account closure, retrieval error, or incorrect format data)
- If the account was a music profile – typically used by singers or bands
- If the profile was private or public
- Number of registered friends
- Date of last access
- Age
- Gender
- Location (country)
- Aim of user (e.g., here for: friendship, networking).
- Religion
- Status (e.g., married, single)
- Sexual orientation
- Ethnicity
- Attitude to having children

All profiles resulting in an error message or a music profile were removed from the analysis, leaving a total of 15,043 (all members collection) and 7,627 (July 3, 2006 members collection). The reason for removing music sites is that these are likely to be fundamentally different in purpose to personal sites.

From the date of last access we calculated the estimated number of days since the profile was last accessed by counting the days from the day of data collection, adjusting this to central U.S. time. This is unavoidably an estimate since it disregards accesses by users during the day of data collection but after the profile had been downloaded.

Note that for most users not all of the above information was present. In particular, the minimum age for MySpace is 14 and profiles for users under 16 are automatically set to "private". Other users can also set their profiles to private. Private profiles are only visible to

registered friends: others (including the program that downloaded the profiles for this research) see a public cut-down profile revealing: date of last access, location (optional), gender and age. Most information apart from age, date of last access and gender is optional and some users do not include it in their profiles. As a result the data contains many missing values: for example most (12,187 in the all members collection) users did not report their ethnicity or had a private profile. In the statistics below, data are automatically removed from any analysis when one or more of the reported values are missing (Table 2 is an exception).

The third data set, the '403 member collection', is a random sample of 403 public, valid, non-music profiles from the second data set for members with at least 2 friends.

On a technical note, the profiles were downloaded using the web crawler SocSciBot 4 (socscibot.wlv.ac.uk) and parsed into a plain text file format (tab delimited) using an extra program added to SocSciBot 4 (see the author for instructions to access this and the extracted MySpace data). The same end result could be obtained using other software such as the Perl module WWW-MySpace (search.cpan.org/dist/WWW-Myspace/) (e.g., Escher, 2007). The data was summarised and graphed in Excel.

Analysis

This analysis takes the form of presenting descriptive statistics as well as cross-tabulation of multiple categories to identify patterns. In the reporting below of cross-tabulation results, for convenience the term 'significant differences' is used to connote differences that are both (a) statistically significant, with a Chi-square p value less than 0.01 and (b) large enough to be interesting, typically meaning at least a 5% difference. Point (b) is necessary because the large sample size allows even relatively small differences to show up as highly statistically significant.

The MySpace signup process

It is useful to briefly discuss the signup process for MySpace with regard to the information available in user profiles. On the signup screen (on July 11, 2007 accessed from the UK) the (compulsory) information requested includes: First name; Last name; Country (from a drop-down box, initially – and presumably normally correctly – set to a location worked out by the server); Date of birth; and Gender. This gives the member a basic profile with an (optional) photograph and the system automatically calculates their age, date of last access, and number of registered friends. In order to add some of the remainder of the information in the bullet list above, the user has to log on to their profile, select the Edit Profile link, and then select the Basic Info tab and fill in a form (Occupation: free text entry, no default setting; Ethnicity: drop down box, no default setting; here for: multiple choice from Dating, Serious Relationships, Friends, Networking - no default setting). In addition to the basic information link, following the Background & Lifestyle link gives more choices (Marital Status: one choice from Swinger, In a Relationship, Single (default), Divorced, Married; Sexual Orientation: one choice from: Bi, Gay/Lesbian, Straight, Not Sure, No Answer (default); Religion: drop down box with one option as in Fig. 7, default = no answer; Children: drop down box with one option as in Fig. 9, default = no answer).

From this process it can be seen that more casual users may not get round to entering the optional information and other users may decide not to report it. The only current default setting that might affect statistical analyses is that of "Single" being the default marital status, but the use of predefined categories is another limitation in others (e.g., there is no Sikh religion option). Note also that new users automatically receive a friendship request from "Tom" who is ostensibly a person but represents a friendly MySpace help and information feature. Users may reject Tom as a friend, accept and view him as a real friend or accept him and view him as a help agent. Hence a person with 1 friend is likely to have just Tom as their friend so the difference in the data between 0 and 1 friend is probably not significant.

Results

Within the all members collection there were 4,090 members with private profiles (27%). This set includes 100% of the 14-15 age group and 18% of the remainder – matching Facebook research suggesting that privacy is often not seen as important (Acquisti & Gross, 2006), although it can be in some countries (Snyder, Carpenter, & Slauson, 2006). There were almost 50% more females (2,415) than males (1,616) with private profiles (a few did not declare a gender).

Figure 1 shows an approximate power law (see Barabási, 2002) in the number of days since the site was last accessed. The oldest recorded date was October 8, 2003, less than two months after the public launch of MySpace on August 5, 2003 (http://freemyspace.com/?q=node/13). The deviation from a pure power law straight line at the top left of the graph can be explained by errors in the dates due to crawling over a 24 hour period. Also the absence of the normal full "fat tail" at the bottom right of the graph can be explained by curtailment at the date that the site started. The left-hand side of the graph (recent accesses) is difficult to interpret because it is likely to include new users as well as old users who check their accounts frequently. Nevertheless, the right hand side of the graph reflects a significant number of users that had not logged on for a long time (half of the users had not checked their account for at least two and a half months), because they had given up MySpace, had a pattern of infrequent checking, or had switched to another MySpace account. Excluding users with 0 or 1 friends on the grounds that they are probably mainly inactive users, the median length of time since the last logon was approximately one week (8 days).

A cross-check was performed for associations between access times and other variables in the data. The only significant association was that females appear to check MySpace more frequently: a small majority of females checked in the previous 1-5 days.

It is useful to group the access dates for further analysis. Based upon a natural grouping of points into approximately straight line components of the graph, the following ranges were selected: 3-5 days and 6-90 days. Two further groups make up the remainder: 1-2 days and 91+ days.
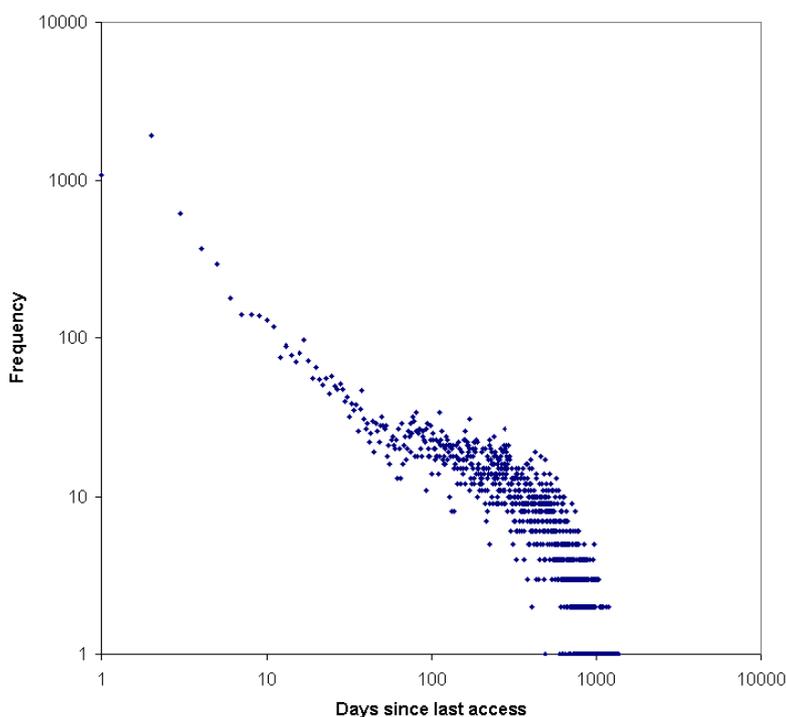

Figure 1. Days since last access for MySpace users (all members collection, note: log-log scale).

Figure 2 reports the same user access data as Figure 1 but arranged by MySpace ID. It is clear from this graph that the IDs are given out in ascending order (and hence MySpace has had over 200 million 'members' - although some people have multiple memberships). In addition, the graph also shows that many members only ever access the site once, when they sign up, and the majority of the remaining users have accessed the site within the past few weeks.
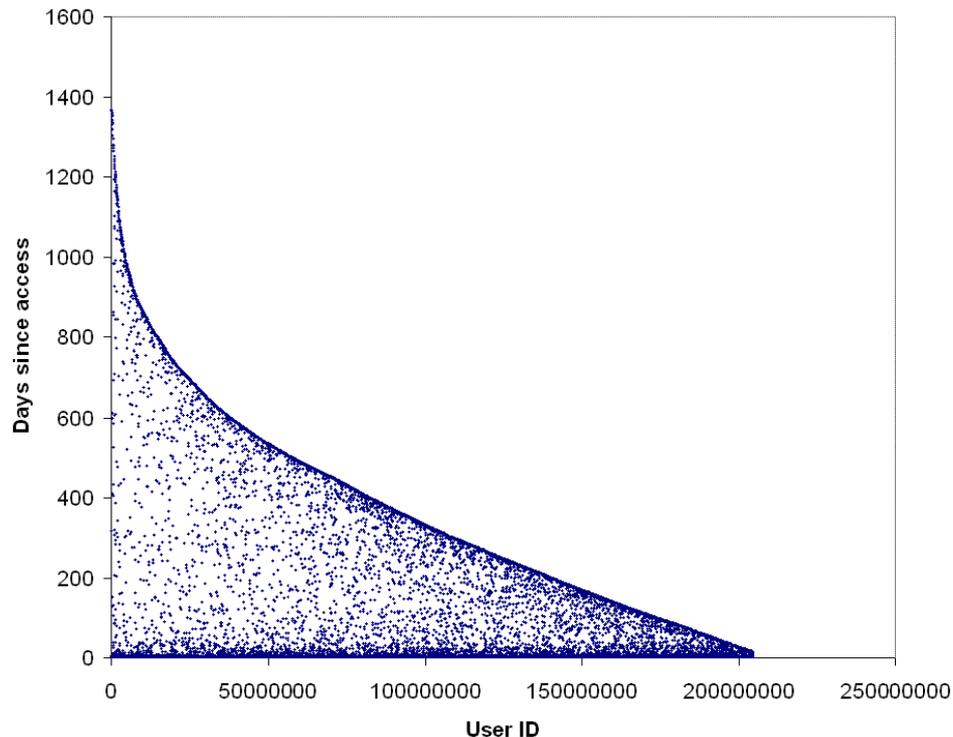

Figure 2. Days since last access against MySpace user ID (all members collection).

Figure 3 gives the clearest indication of usage patterns, using the July 3, 2006 members collection. Over a quarter of users (28.5%) only logged on to their MySpace account on the day that they signed up, and a third (33.3%) did not log on again later than a week after signing up. A total of 18.9% accessed it in the early hours (assuming they were U.S. based) of the day of the survey or the day immediately before: most of these presumably check daily. Almost a third (31.4%) had checked in the previous week, and 40.0% in the previous month. In simple terms, it seems that about a third of members give up immediately, a third become regular users and the rest log on occasionally.
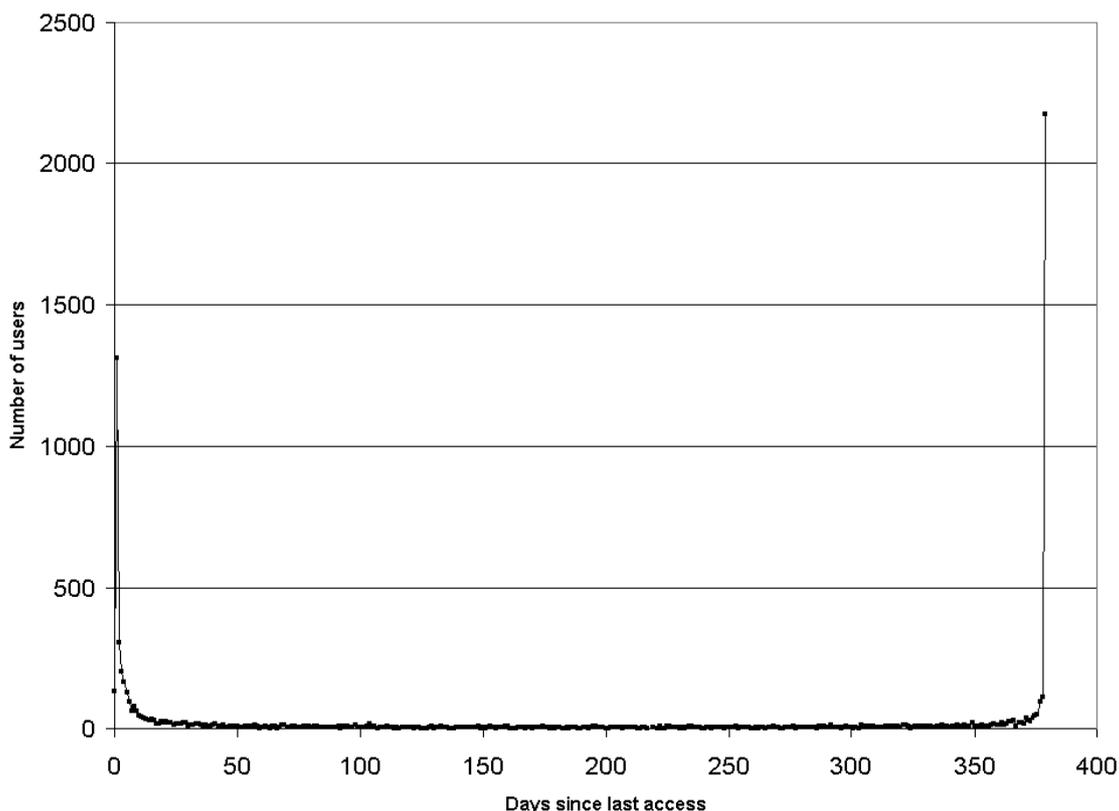
Figure 3. Days since last access (July 3, 2006 members collection).

Figure 4 shows the expected power law distribution of numbers of friends: most users have only a few friends but some have many. The median number of friends is 1, and the median number for users who have at least 2 friends is 27. The power law is not perfect, though, and any deviations from a perfect power law may suggest that different types of friending mechanisms are co-occurring. There appear to be four different patterns at work, consistent with previous research (boyd, 2006; Fono & Raynes-Goldie, 2005).

- *No friends*: 0-1 friends. There are more users with one friend than expected for the power law. As a result, this probably reflects people who are trying out the system and adopt Tom as their friend as part of this (Tom is the sole friend of about 99% of users with 1 friend).
- *Close friends*: 2-9 friends: There is an approximately straight line for users with 2-9 friends. This might represent users with friends that are all or mainly offline friends, i.e., interpreting friendship the same online as offline.
- *Acquaintances*: 10-90 friends: There is an approximately straight line covering users with 10-90 friends and this line has a different slope to the 2-9 friends slope. This suggests users following a different friending mechanism and would be consistent with those approximately equating MySpace friendship with acquaintanceship, and perhaps including a significant number of bands. This probably also includes some close friends.
- *Strangers*: 90+ friends: The tail of the graph seems to have a different average slope to the other parts. This could reflect members competitively trying to get as many registered friendships as possible. This probably also includes some close friends and acquaintances.

Note that the median of 27 friends for members with at least 2 friends falls in the acquaintanceship category above, which accounts for 40% of these users. Hence this could reasonably be viewed as the modal or "normal" interpretation of social network friending.

In terms of gender, there was a fairly even split between male and female users in the 0-1 and 2-9 categories but proportionally more males had 2-9 friends and more females had

more friends than this. About 9% more females than males had more than 9 friends based upon the average for both males and females – this small difference is still statistically significant because of the large numbers of data points involved (a Chi-square test for the whole table is significant with $p=0.000$).

Note that for the purposes of this discussion the friendship patterns have been identified visually. Although there are mathematical techniques for fitting lines to power law distributions (Rousseau & Rousseau, 2000), these are controversial and are an unnecessary complication that does not add to the discussion. Note also that in some of the analyses below, members with 0-1 friends are removed under the assumption that the vast majority are not active members, but members with private profiles that do not reveal friend counts are included in the relevant analysis (Figure 6), even though some of these will have 0-1 friends.
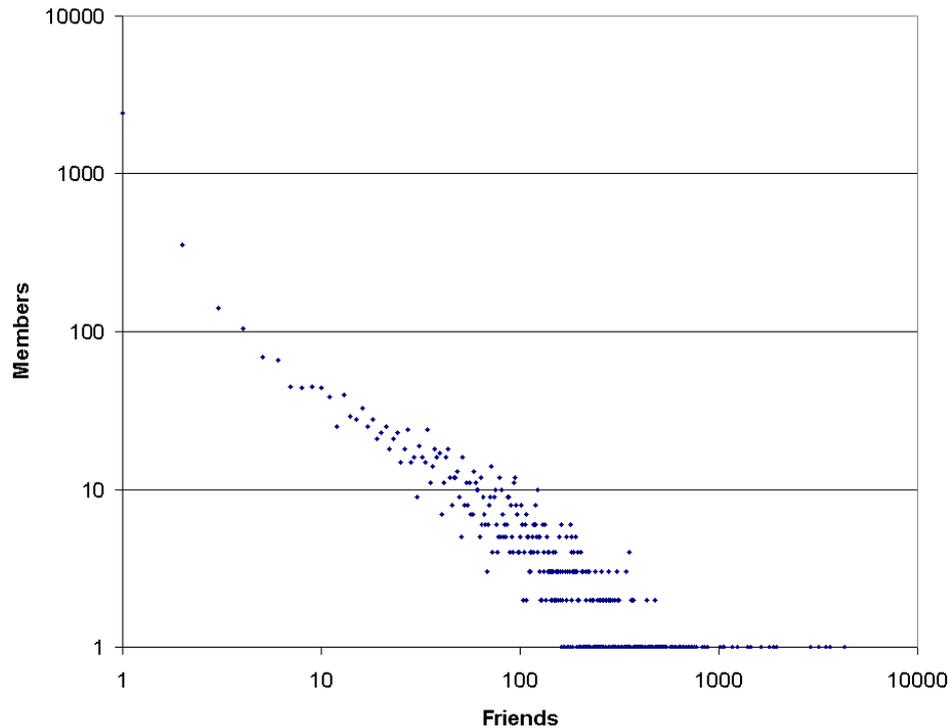

Figure 4. Friends scatter graph for users over 15 (July 3, 2006 members collection, note log-log scale and that the 199 users with 0 friends are not shown).

Figure 5 presents data concerning the number of users with different groups of friends, but breaks down each category by the number of days since the last logon. It is clear from the graph that people with 0 or 1 friends are unlikely to be active users of the system. It seems that 0-1 friends represents a member not really using the system, other than perhaps to create a profile "to see what all the fuss is about". If this is true then just over half of the members don't really engage with MySpace. Even those who accessed the system within 6-90 days may have joined recently but have already given up or never intended to become active users.
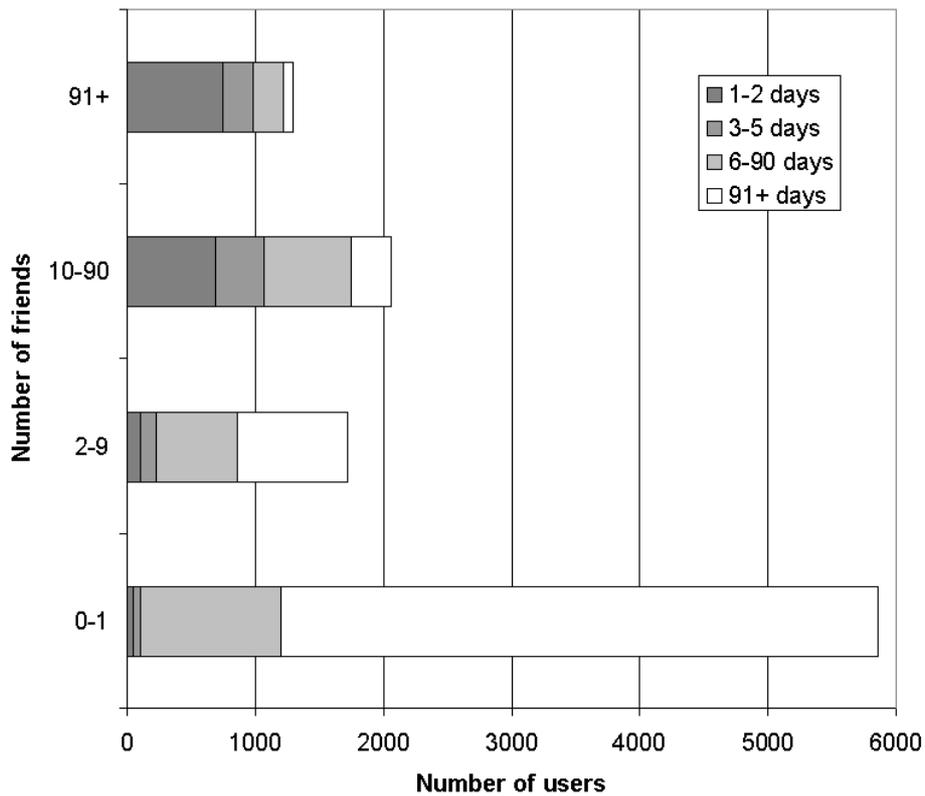
Figure 5. Number of friends and time since last logon of MySpace users over 15 (all members collection).

Basic user statistics

Figure 6 reports the gender distribution of age ranges, excluding users with 0 or 1 friends (but including users with private profiles that do not reveal how many friends they have). There is a significant youth bias to MySpace, although the median reported age is 21 – the median age of active users is slightly lower because the data ages annually with its members – for example the average age of users who logged on in the previous two months is 20. There are clearly significantly more female teen users but more males than females in their late twenties. Overall there are more females (7,658 - or 4,876 excluding those reporting 0 or 1 friends) than males (7,278 or 4,219 respectively). The surprisingly high number of centenarians could be due to children signing up their oldest living relatives but certainly reflects joke ages for at least some: several claimed to want children "someday"; and there are examples in the data like "Kevin" (apparently 102) who appears to be about 13 from his picture and has a typical teen action-packed MySpace profile.
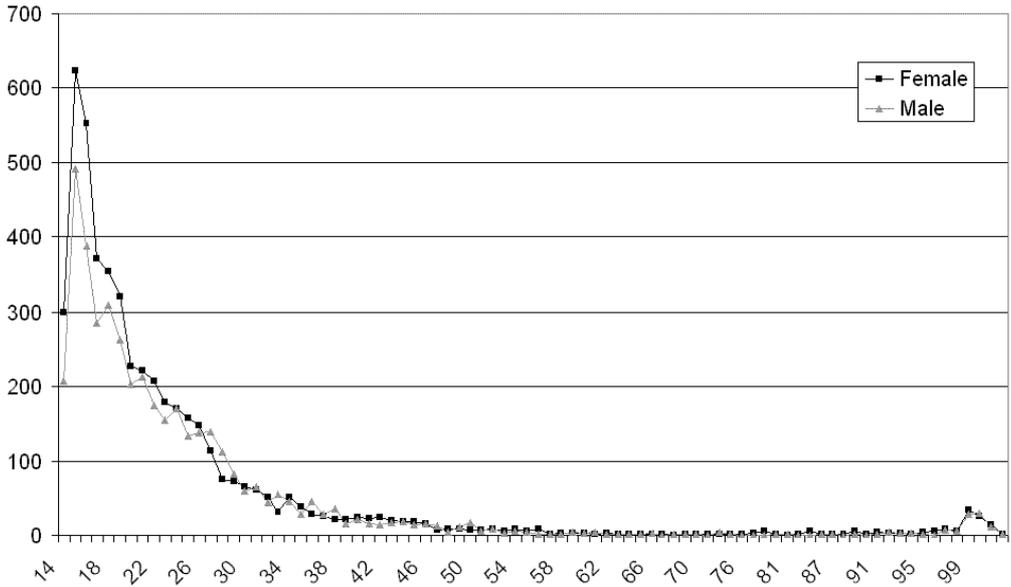
Figure 6. Ages and genders of MySpace users (all members collection, excluding those reporting 0-1 friends).

Figure 7 reports the age distribution of MySpace users' friendship circle sizes, using the categories introduced in the discussion relating to Figure 4. The results for users over 40 are excluded as unreliable due to the smaller numbers involved (under 100 in most cases – see Figure 6), and friend numbers are unavailable for 14 and 15 year-olds. The pattern is clearly for younger users to have more MySpace friends. Perhaps surprisingly, though, the very youngest group had the most friends. Presumably these friends were built by the 16 year olds when they were 14 and 15 and their profiles were private.
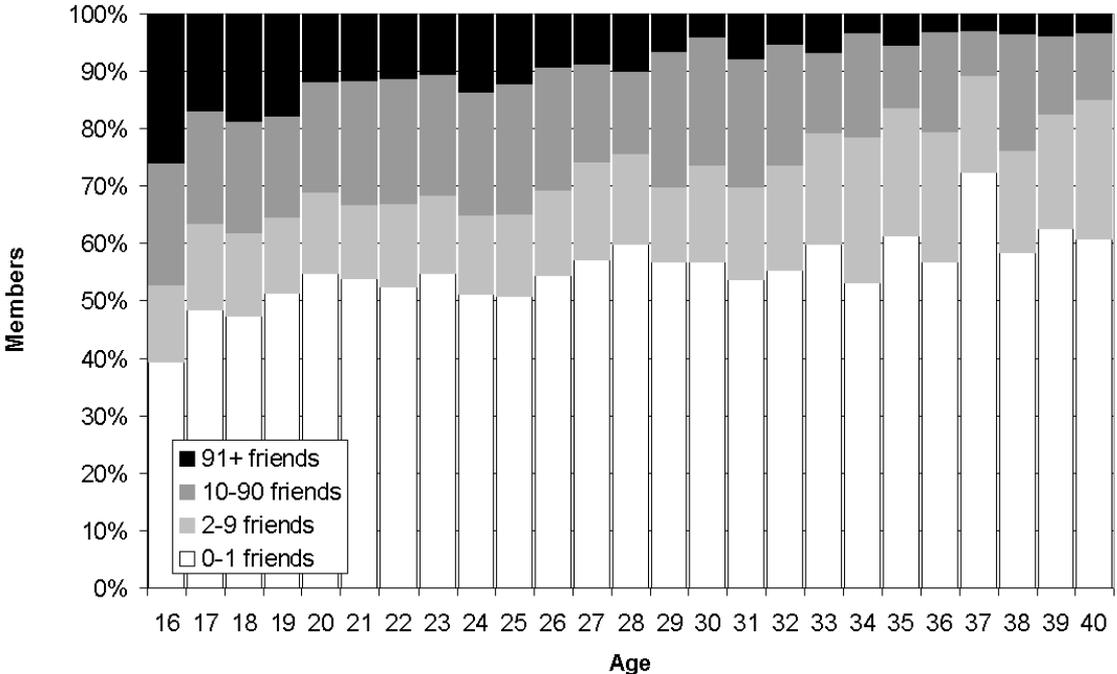

Figure 7. Ages and friendship counts of MySpace users 16-40 (all members collection).

User attributes and beliefs

Figure 8 shows the dominance of Christianity over other forms of religion, although given the U.S. base of MySpace users, the number of Protestants is surprisingly low. Perhaps the Christian–other category was used by Protestants who regarded themselves as part of a

particular Protestant church rather than as a generic Protestant. There did not seem to be a significant association between age and religion or gender and religion amongst MySpace users.
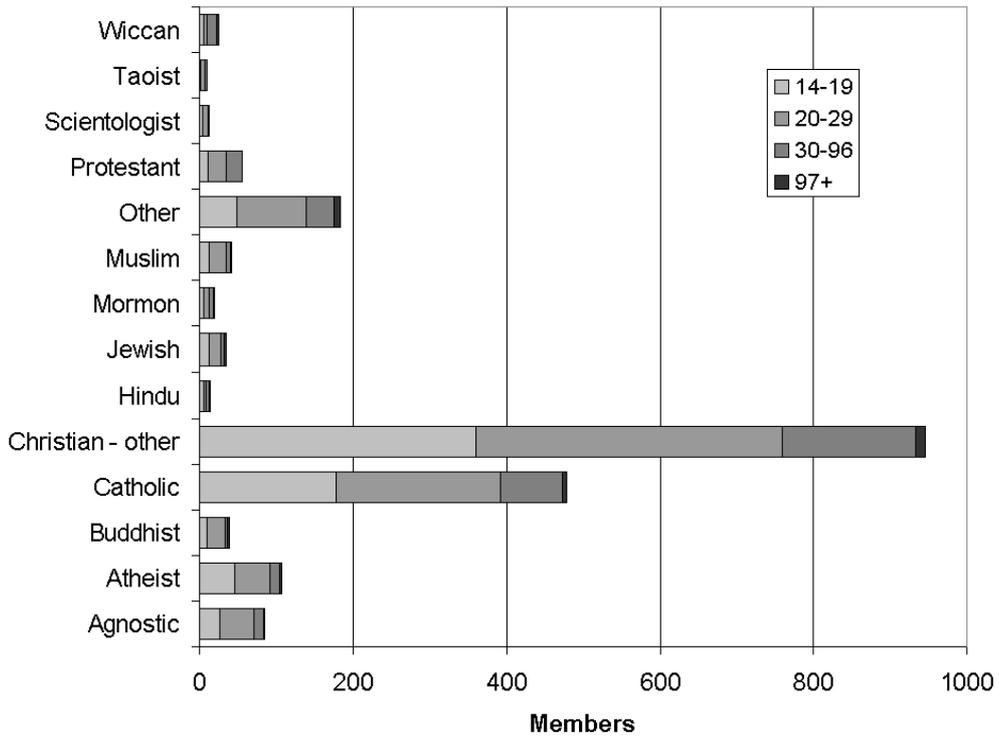
Figure 8. Religions of MySpace users over 15 broken down by age (all members collection, excluding those reporting 0-1 friends; 59% unknown).

Figure 9 is the most surprising graph in the data set. It apparently shows that there is a large difference in the desire to have children according to the number of friends, even if the 0-1 friends column is ignored. A related clear but unsurprising association (not shown) is that younger users were the more likely not to want children. It seems likely that "I don't want kids" used to be a default in MySpace, with many casual users never changing it.
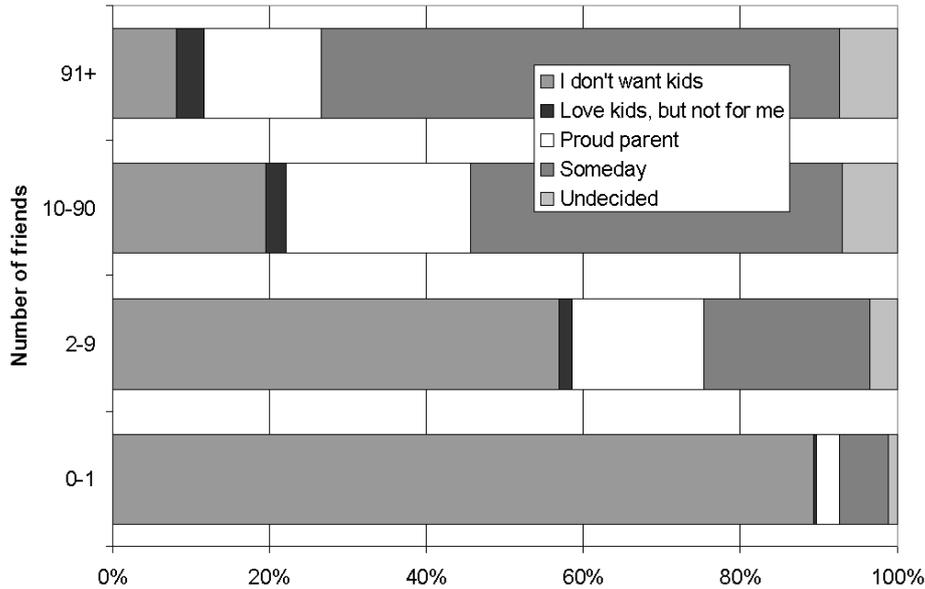
Figure 9. Child orientation of MySpace users over 15 by number of friends (all members collection; 41% unknown).

Figure 10 shows that most users are using MySpace for friendship although some also use it for dating, networking and serious relationships. Women are more likely to be interested primarily in friendship and men are more likely to be interested in dating, conforming to previous findings for teens (Pew Research Center for the People & the Press, 2007). Another pattern in the data (not shown) is that younger users were less likely to network.
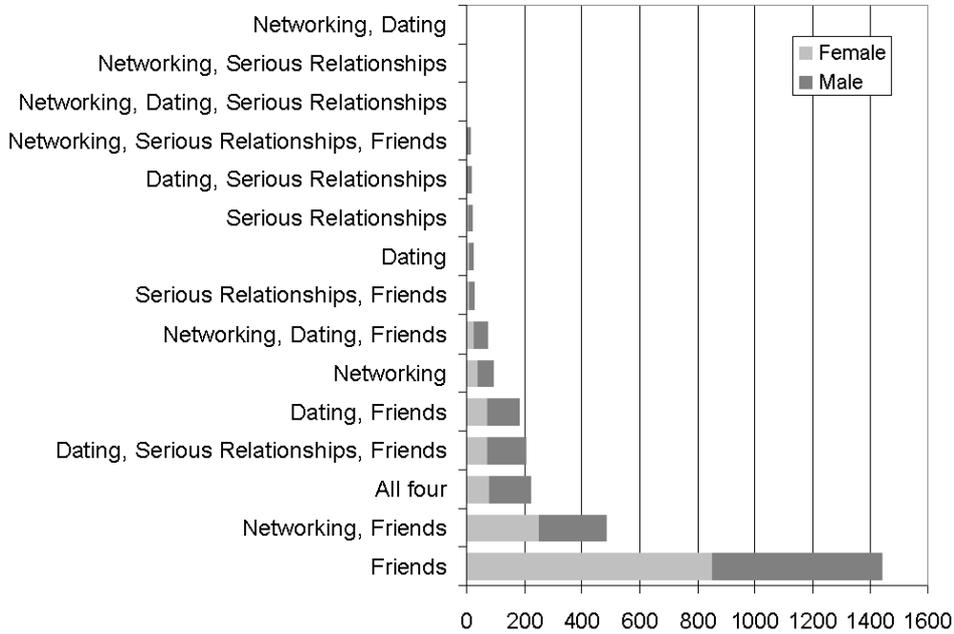
Figure 10. Objectives of MySpace users over 15 by gender (all members collection, excluding those reporting 0-1 friends; 88% unknown).

The other member attributes checked produced results that provide useful background demographic data. A large majority (77%) of users were from the U.S., although – surprisingly - the UK was in second place (5%), followed by Canada (2%), Australia, (2%), the Philippines (1%) and Mexico (1%). Over 90% reported a geographic location. Unsurprisingly, however, a large majority of the 26% who gave a sexual orientation chose

straight (93%). The relationship status of users (Table 1) seems suspect because there are very few divorced members. Presumably divorced people tend to be older than average and hence less likely to be MySpace users, but they might also declare themselves as single (or accept MySpace's default characterisation of single).

Table 1. Relationship status of MySpace users over 15, where given (all members collection, excluding those reporting 0-1 friends; 27% unknown).

Status	Frequency	Percent
Divorced	81	2%
In a Relationship	965	19%
Married	573	11%
Single	3344	66%
Swinger	97	2%

The ethnic origins reported in Table 2 are perhaps most noteworthy for the number of people who did not report this attribute. Over half of the members (2,596) with public profiles and at least 2 friends did not report an ethnicity. Perhaps it is often not seen as significant in cyberspace. Alternatively, however, ethnicity is obvious from many photographs or it may be seen as distasteful or uncool to report.

Table 2. Ethnic origin of MySpace users over 15 (all members collection; 51% unknown).

Ethnicity	Frequency	Percentage
Asian	140	6%
Black / African desc	297	12%
East Indian	12	0%
Latino / Hispanic	337	14%
Middle Eastern	11	0%
Native American	20	1%
Other	169	7%
Pacific Islander	32	1%
White / Caucasian	1459	59%

Friends' genders

The third sample of data, the 403 member collection, was used to investigate the influence of gender on friendship: whether males had more male friends than female and vice versa, either in the Top 8 friends list or the first page of the full list of friends. Each profile page was visited and the members' "Top 8" friends categorised into four groups: male friends, female friends, bands, and others. Friends were classified as bands if they were pop groups or used the MySpace music site, or 'others' if they were identified in any other way as a professional person with a fan base (e.g., actors, comedians). The remainder were classified as male or female, with the exception of the help agent 'Tom' who was ignored. The same process was repeated on the first full page of up to 40 friends reached by clicking the "all friends" link – these seem to be the member's first 40 friends. The limitation to the first page was a practical choice as some members had hundreds of friends.

A test for gender differences in the gender of Top 8 friends was positive (Mann-Whitney, $p=0.026$): females had a slightly higher proportion of female friends (60.5%) than did males (53.5%). In contrast, males had slightly more female page 1 friends (58.1%) than did females (54.1%), although the difference was not significant (Mann-Whitney, $p=0.196$). It seems that males and females do not seem to have greatly differing friend gender preferences for their closest friends (i.e., approximately 7%), which is surprising given recognised gender differences in offline relationships (Aukett, Ritchie, & Mill, 1988; Caldwell & Peplau, 1982; Elkins & Peterson, 1993; Reeder, 2003; Shrum, Cheek, & Hunter, 1988). Males possibly have more females as more distant friends, however, although the difference was not large enough to be statistically significant. In addition, there was a difference in those chosen to be friends,

with all members tending to choose more female friends in their Top 8 list (Wilcoxon Signed Ranks Test, p=0.000) and tending to have more female friends in their first page of friends (Wilcoxon Signed Ranks Test, p=0.000). In summary, both males and females share the same preference to have a majority of female friends, although females have a significantly greater preference for close friends to be female.

Limitations

This research has a number of limitations that are worth reviewing before the conclusions.

First, and most importantly, some of the data is misleading, and probably the age data is the most unreliable. Almost all of the data analysed is self-reported and presumably some of it is deliberately or accidentally incorrect. For example, members may lie out their age and probably there are many users under 14 who declare an older age in order to have a profile. A systematic study of profiles of youth apparently aged under 18 has found evidence of age inflation in 8.3% (Hinduja & Patchin, in press). For instance, in the sample used in the current study, one 21 year old female's profile declared an interest in "hot boys 11-14". This seems likely to be the work of a girl aged about 11, although it could be a joke comment or even the work of an inept paedophile. Moreover, some users may give a false date of birth (although this is not reported in the user's profile) as potentially useful information for identity theft (e.g., see http://news.bbc.co.uk/2/hi/uk_news/6910826.stm, accessed July 24, 2007). Perhaps most substantially, members may use MySpace to create a new online identity that is unrelated to their offline identity, perhaps including gender switching. This would be consistent with findings about deception for other online communication media (e.g., Utz, 2005). The pattern in Figure 6 – dominated by young users - gives some credence to the self-reported ages, although it seems likely that many of those aged 14-15 may be younger. The data may also have been affected by design decisions in MySpace, such as the default settings for the answers to some of the questions. Some friends will also be music sites, a qualitatively different type of friend to other members. Hence, although the member-entered data must be interpreted cautiously in respect of both the potential differences between members and their projected identity representations, and the potential biasing factors of the MySpace interface, it seems unlikely that the data is dominated by deception, especially because of the close connection between offline and online friendship.

Second, the 'all members' data set is heterogeneous in terms of the length of time that the users in the sample had been members. Users presumably add friends over time and so newer users with few friends may eventually have more friends, although in some of the analyses above they are grouped with older users with few friends. Membership duration may also affect all the data to some extent, for example if older users are more likely to get round to completing their profile information or to adopt copying practices with regard to reporting information (e.g., Perkel, 2006).

Conclusions

In terms of reported age, although the modal age range for MySpace is teen, MySpace is apparently not a predominantly teen site: it has a reported median age of 21 and slightly younger active users. Its youngest members are most likely to be female and there is a small female majority overall. Teen users have more friends than average and are the most likely to have a very large number (e.g., over 90). Amongst the other investigated variables, a lack of desire for children and more frequent logging on also associate with youth, as does an absence of interest in "networking". Whilst this data could be interpreted as referring mainly to the projected identity representations of the MySpace members, it seems reasonable that this predominantly matches with offline identities.

There is a huge variety of friendship circle sizes: a few members have tens of thousands of friends but the majority have 0 or 1. The sizes approximately follow a power-law distribution with break points consistent with three interpretations of friends: close friends; acquaintances; and strangers. This corroborates previous research that has pointed to different conceptualisations of online friendship by users (boyd, 2006; Fono & Raynes-

Goldie, 2005), but is somewhat at odds with Facebook research that emphasises its role in building relationships started offline (Ellison et al., 2006). Perhaps social networks allow stronger bonds to build with casual acquaintances – changing the nature of weak friendship – or, perhaps members build an extended network of acquaintances but concentrate most of their social networking activity on their best friends.

The large number of users with no real friends suggests that many people sign up to MySpace, perhaps out of curiosity, and then do not become active users. This conjecture is supported by the typically long length of time since their last logon for these members. Ignoring those with 0 friends or 1 friend, the median number of friends of the remainder of users (aged 16+) is 27, and 40% of members fall in the 10-90 friends "acquaintanceship" category which suggests that acquaintanceship is the modal interpretation of friendship in MySpace by active users.

The results with regard to gender confirm that female users are more likely than males to be mainly interested in friendship, although this is true for a majority of both genders. In contrast, male users are more likely to be interested in dating and/or serious relationships, although this involves a minority and so the significance of this should not be exaggerated. Female users also tended to have more friends, and probably log on more frequently, although the differences were not large. Male and female members both seemed to preferring to have a majority of female friends, but males preferred to have a significantly greater proportion of female friends. A corollary of this is the importance of cross-gender friendships, an area that has attracted relatively little research (Gee, 2004; Monsour, 2001; Scudder & Bishop, 2001). Apart from these results, and a female majority, especially within younger members, there was a reasonably even gender split in the categories investigated.

In conclusion, the results above provide quantitative evidence to back up a range of previous claims about patterns of social networking, at least with regard to MySpace users. In addition, they support previous survey-based findings about social networking, providing evidence triangulation to make these findings more certain. If there is a typical MySpace user then she is probably 21, single, with a public profile, interested in online friendship and logging on weekly to engage with a list of mixed but majority female 'friends' that are predominantly acquaintances (including bands). This is a very different profile to those of the teen users frequently discussed in the contemporary press, and so it seems likely that public perceptions of MySpace are at variance with reality.

Acknowledgements

References

Acquisti, A., & Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the Facebook. *Lecture Notes in Computer Science, 4258*, 36-58.

Aukett, R., Ritchie, J., & Mill, K. (1988). Gender differences in friendship patterns. *Sex Roles, 19*(1-2), 57-66.

Barabási, A. L. (2002). *Linked: The new science of networks*. Cambridge, Massachusetts: Perseus Publishing.

boyd, d. (2006). Friends, Friendsters, and MySpace Top 8: Writing community into being on social network sites. *First Monday, 11*(2), Retrieved June 23, 2007 from: http://www.firstmonday.org/issues/issue2011_2012/boyd/index.html.

boyd, d. (2007). Viewing American class divisions through Facebook and MySpace. *Apophenia Blog Essay (June 24)*, Retrieved July 12, 2007 from: http://www.danah.org/papers/essays/ClassDivisions.html.

boyd, d. (2008). Why youth (heart) social network sites: The role of networked publics in teenage social life. In D. Buckingham (Ed.), *Youth, identity and digital media* (pp. 119–142), Cambridge, MA: MIT Press.

Caldwell, M. A., & Peplau, L. A. (1982). Sex differences in same-sex friendships. *Sex Roles, 6*(7), 721-732.

Donath, J., & boyd, d. (2004). Public displays of connection. *BT Technology Journal, 22*(4), 71-82.

Duck, S. (1992). *Human relationships*. London: Sage.

Escher, T. (2007). The geography of (online) social networks. *Web 2.0, York University*, Retrieved September 18, 2007 from: http://people.oii.ox.ac.uk/escher/wp-content/uploads/2007/09/Escher_York_presentation.pdf

Elkins, L. E., & Peterson, C. (1993). Gender differences in best friendships. *Sex Roles, 29*(7-8), 497-508.

Ellison, N., Steinfield, C., & Lampe, C. (2006). Spatially bounded online social networks and social capital: The role of Facebook. *Annual Conference of the International Communication Association*, Retrieved April 23, 2007 from: http://msu.edu/~nellison/Facebook_ICA_2006.pdf.

Fono, D., & Raynes-Goldie, K. (2005). Hyperfriendship and beyond: Friendship and social norms on Livejournal, Association of Internet Researchers (AOIR-6), Chicago.

Gee, L. (2004). *Friends: Why men and women are from the same planet*. New York: Bloomsbury Publishing.

Golder, S. A., & Huberman, B. A. (2006). The structure of collaborative tagging systems. *Journal of Information Science, 32*(2), 198-208.

Golder, S. A., Wilkinson, D., & Huberman, B. A. (2007). Rhythms of social interaction: Messaging within a massive online network, 3rd International Conference on Communities and Technologies (CT2007), East Lansing, MI.

Hinduja, S., & Patchin, J. W. (in press). Personal information of adolescents on the Internet: A quantitative content analysis of MySpace. *Journal of Adolescence*.

Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., & Riedl, J. (1997). GroupLens: Applying collaborative filtering to Usenet news. *Communications of the ACM, 40*(3), 77-87.

Lerman, K. (2006). Social networks and social information filtering on Digg. *ArXiv.org*, Retrieved April 23, 2007 from: http://arxiv.org/abs/cs.HC/0612046.

Malinowski, B. (1923). The problem of meaning in primitive languages. In C. K. Ogden & I. A. Richards (Eds.), *The Meaning of Meaning: Routledge & Keegan Paul* (pp. 296-346).

Monsour, M. (2001). *Women and men as friends: Relationships across the life span in the 21st century*. New York: Lawrence Erlbaum Associates Inc.

Perkel, D. (2006). Copy and paste literacy: Literacy practices in the production of a MySpace profile. *Informal Learning and Digital Media*, Retrieved April 23, 2007 from: http://www.dream.dk/uploads/files/perkel%2020Dan.pdf.

Pew Research Center for the People & the Press. (2007). *Social networking websites and teens: An overview*. Retrieved June 4, 2007, from http://www.pewinternet.org/PPF/r/198/report_display.asp

Prescott, L. (2007). Hitwise US consumer generated media report. Retrieved March 19, 2007 from: http://www.hitwise.com/.

Reeder, H. M. (2003). The effect of gender role orientation on same- and cross-sex friendship formation. *Sex Roles, 49*(3-4), 143-152.

Rousseau, B., & Rousseau, R. (2000). LOTKA: A program to fit a power law distribution to observed frequency data. *Cybermetrics, 4*(1), Retrieved July 5, 2007 from: http://www.cindoc.csic.es/cybermetrics/articles/v2004i2001p2004.html.

Scudder, J. R., & Bishop, A. H. (2001). *Beyond friendship and Eros: Unrecognized relationships between men and women*. New York: State University of New York Press.

Shrum, W., Cheek, N. H., & Hunter, S. (1988). Friendship in school: Gender and racial homophily. *Sociology of Education, 61*(4), 227-239.

Snyder, J., Carpenter, D., & Slauson, G. J. (2006). MySpace.com: A social networking site and social contract theory. *Proceedings of ISECON 2006*, Retrieved April 23, 2007 from: http://isedj.org/isecon/2006/3333/ISECON.2006.Snyder.pdf.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology*. London: Sage.

Thelwall, M. (in press). Fk yea I swear: Cursing and gender in a corpus of MySpace pages. *Corpora, 3*(1), Retrieved November 14, 2007 from: http://www.scit.wlv.ac.uk/~cm1993/papers/MySpaceSwearing_online.doc

Utz, S. (2005). Types of deception and underlying motivation - What people think. *Social Science Computer Review, 23*(1), 49-56.

Valkenburg, P. M., Peter, J., & Schouten, A. P. (2006). Friend networking sites and their relationship to adolescents' well-being and social self-esteem. *CyberPsychology & Behavior, 9*(5), 584-590.