

@media print {#ghostery {display:none !important}}

f

i

®

s

t

m

✕

ñ

d

@

ψ

PEER-REVIEWED JOURNAL ON THE INTERNET

HOME

ABOUT

LOGIN

REGISTER

SEARCH

CURRENT

ARCHIVES

ANNOUNCEMENTS

SUBMISSIONS

Home > Volume 14, Number 10 - 5 October 2009 > Jernigan

f

i

®

s

t

m

✕

ñ

d

@

ψ

PEER-REVIEWED JOURNAL ON THE INTERNET

Gaydar: Facebook friendships expose sexual orientation

by Carter Jernigan and Behram F.T. Mistree

USER

Username

Password

☐ Remember me

Login

JOURNAL CONTENT

Search

All

Search

Browse

By Issue

By Author

By Title

Other Journals

FONT SIZE

CURRENT ISSUE

ATOM

1.0

RSS

2.0

RSS

1.0

ARTICLE TOOLS

Abstract

Print this article

Indexing metadata

How to cite item

Email this article (Login required)

Email the author (Login required)

ABOUT THE AUTHORS

Carter Jernigan

Behram F.T. Mistree

Abstract

Public information about one’s coworkers, friends, family, and acquaintances, as well as one’s associations with them, implicitly reveals private information. Social networking Web sites, e-mail, instant messaging, telephone, and VoIP are all technologies steeped in network data — data relating one person to another. Network data shifts the locus of information control away from individuals, as the individual’s traditional and absolute discretion is replaced by that of his social network. Our research demonstrates a method for accurately predicting the sexual orientation of Facebook users by analyzing friendship associations. After analyzing 4,080 Facebook profiles from the MIT network, we determined that the percentage of a given user’s friends who self-identify as gay male is strongly correlated with the sexual orientation of that user, and we developed a logistic regression classifier with strong predictive power. Although we studied Facebook friendship ties, network data is pervasive in the broader context of computer-mediated communication, raising significant privacy issues for communication technologies to which there are no neat solutions.

Contents

- Introduction
- Facebook overview
- Literature review
- Methodology
- Results
- Discussion
- Conclusion

Introduction

There is an old saying that “birds of a feather flock together.” The lesson is that people are self-segregating, such that the composition of your friends reflects on you. If the majority of your friends were male, one might predict you to be male. If many of your friends were a particular race, one might predict you to be that race. If many of your friends were gay, one might predict you to be gay. This predictive power works even though one knows nothing about you, as long as one knows something about your friends.

With the advent of computer-mediated communication, it has become disturbingly easy to log and track the web of human interactions. The phone company stores data on who calls whom, for example, and that data builds a social graph. Likewise, consider a social networking Web site such as Facebook where connections between users are clearly visible. Facebook allows each user to create a profile, and users can also designate other Facebook users as “friends.” Although Facebook users can control the content of their own profiles, they cannot control the content of their friends’ profiles. This suggests that significant personal information about a Facebook user can be determined by analyzing his network of friends. We were motivated to demonstrate this property of social networks in an effort to create awareness among users.

We tested the hypothesis that the sexual orientation of a given Facebook user can be determined based on the sexual orientations of that user’s friends and found that self-reporting gay male MIT Facebook users have almost an order of magnitude higher percentage of gay male friends than heterosexual male users. Thus, the ability to determine the sexual orientation of an arbitrary Facebook user represents a serious violation of privacy. Although one might immediately consider discrimination as the primary consequence of being “outed,” revealing one’s sexual orientation can be a deeply personal experience for some individuals. A study of lesbian, gay male, and questioning students at the University of Maryland showed that 92.1 percent of participants had disclosed their sexual orientation to their friends, while only 78.7 percent had disclosed to their parents (Lynch, 2005). Such disclosure patterns indicate that lesbian, gay male, and bisexual male or female (LGB) individuals exercise discretion during the “coming out” process. Although our work may seem immediately relevant only to Facebook and the gay community, the broader ramifications of our work emphasize how little control one has over one’s own privacy in online communities.

Facebook overview

Originally designed for college students as a social networking Web site, Facebook allows each user to create a profile, complete with information such as home address, mobile phone number, interests, religious views, and even the requisite data for online dating like sex, relationship status, and sexual orientation. In addition to creating individual profiles, Facebook users can also designate other users as

Window size: x  
Viewport size: x

"friends," send messages, and post pictures. Figure 1 presents a screenshot of a Facebook profile rendered in a Web browser. Notice the advertisement on the left. Is this advertisement the result of hetero-normative marketing? Is this advertisement appropriate, given the omission of sexual orientation, from the profile? On Facebook, sexual orientation is denoted by a field labeled "Interested In" — (men, women, men and women, blank), although in this case it was intentionally omitted by the owner of the profile.



Figure 1: Facebook profile screenshot of one of the authors taken November 2007. Notice the advertisement on the left.

The aggregation of any large set of personal information increases the potential for abuse. According to an Associated Press report, individuals commonly snoop through their employer's private customer data while on the job. For example, the IRS disciplined 219 employees in 2007 for browsing through confidential information on individual taxpayers (Foley, 2008). Most recently, the U.S. State Department fired employees for snooping into Barack Obama's passport files (Tumulty, 2008).

With such a backdrop, it is not surprising that Facebook has its own privacy issues. For example, employers analyze prospective employees' Facebook profiles during the hiring process (Fuller, 2006). Historically, many of the privacy concerns raised involve information explicitly disclosed by users to Facebook. Consider Kevin Colvin, an intern at Anglo Irish Bank, who missed work for a "family emergency." Colvin's boss subsequently came across a picture on Facebook that revealed Colvin was actually attending a Halloween party, and Colvin was fired as a result (Valleywag, 2007). In instances like this, one cannot help but blame Facebook users for compromising their own privacy. It is important to remember that Facebook collects data that users are providing. Users make a choice to register for a Facebook account, fill in their profiles, and upload photos. Later in this paper, we discuss what Facebook can do to educate users of the consequences of posting such information online.

With an active user base of over 250 million (Facebook, 2009), the amount of personal information collected by Facebook is staggering. Facebook users upload an average of 33 million pictures a day (Facebook, 2009) and post intimate details in their profiles. Industry analysts believe the ballast of Facebook's financial worth to be in its massive aggregation of information about its users and its advertising-driven business model (Guth, *et al.*, 2007). The larger Facebook's user base, the more consumers advertisers can reach with each advertisement; the more information Facebook collects about its users, the better advertisers can selectively target market segments. With so much data from so many users stored on its servers, how does Facebook protect its users' privacy?

Facebook's policies ward off some attacks against user privacy. For example, spammers and telemarketers are forbidden from harvesting contact information from Facebook profiles (Facebook, 2007). In the end, though, much of the burden of privacy is on the individual user. To help users protect themselves, Facebook employs explicit privacy controls in addition to the implicit controls dictated by its Terms of Use and Privacy Policy. Facebook is built around the construct of "networks," which generally map to schools, businesses, or geographic regions, and these networks provide a basic level of access control on Facebook. By default, profiles are only visible to other members of the same network, although friends can always view each other's profiles. For example, MIT students might join the MIT and Boston networks and could therefore view and be viewed by other members of those networks. This default setting permitted us to access a significant number of profiles in the MIT network because we are members of the MIT network.

Facebook friendships provide another level of access control, as friends can always view each other's profiles. Essentially, Facebook allows users to create a limited profile [1], which contains a user-specified subset of the fields in a user's full profile. Then specific people can be designated to see only the limited profile. For example, if Alyssa P. Hacker did not want Ben Bitdiddle to see her phone number, Alyssa could specifically hide her phone number in her limited profile and tell Facebook that Ben should be shown only this limited profile. Alternatively, Alyssa could add Ben to a blocked list that prevents Ben from seeing anything about Alyssa or even that Alyssa has a Facebook profile. It should be noted that Facebook users also have two additional options to secure their privacy: (a) leaving fields blank or (b) listing bogus information. For instance, if Alyssa P. Hacker is uncomfortable with sharing her mobile phone number in her profile, she can either (a) leave the mobile phone number field blank or (b) list a bogus mobile phone number.

Later sections of this paper question whether the privacy options discussed here provide adequate privacy safeguards or merely create an illusion of privacy. Specifically, we will analyze how seemingly innocuous friendship associations reveal intimate details about Facebook users.



## Literature review

"One of the most persistent empirical regularities that has been observed among social relationships is the tendency towards *equal status contact*." [2] In essence, people socialize with people who are like them in terms of categories like gender, sexual orientation, age, race, education, and religion.

### Sex segregation

Consider that children on the playground segregate themselves based on sex: six-year olds favor playing with same-sex playmates in a ratio of 11:1 [3]. Although adults no longer worry about "cooties," a

significant sex bias persists in adult friendships such that same-sex friendships are more frequent than cross-sex friendships for both men and women. One study showed that men had 65 percent male friends versus 35 percent female friends, while women had 70 percent female friends versus 30 percent male friends (Reeder, 2003). Another study revealed that this phenomenon persists in lesbian, gay, and bisexual individuals [4]. In short, males prefer males for friendship, and females prefer females for friendship.

Sexual orientation segregation

Lesbians and gay men draw the majority of their friends from the LGB community while bisexual women and men draw the majority of their friends from the heterosexual community, as discovered by Paz Galupo in a recent study of close friendships and LGB individuals [5]. The study allowed participants to report on up to eight close friends, which is reasonable because of the limited human capacity to maintain close friends. Supporting this notion of limited capacity, Malcolm Gladwell poses the following exercise in his book *The Tipping Point*: "Make a list of all the people you know whose death would leave you truly devastated." The average answer to the exercise is 12, and those 12 people make up your "sympathy group." [6] Galupo's participants listed an average of 4.13 close friendships [7], which is reasonable considering Gladwell's "sympathy group" of 12 also contains family. Of those 4.13 close friends reported by LGB participants in Galupo's study, lesbian and gay male participants reported an average of 1.72 heterosexual friends and an average of 2.28 lesbian and gay friends [8]. Galupo based her research on data from a larger study, and the heterosexual participants of that larger study reported an average of 4.26 close friends, of which 0.17 were reported to be lesbian, gay, bisexual, or "questioning." [9] Put another way, lesbians and gays draw 55 percent of their close friends from the LGB community, while heterosexuals draw only four percent of their close friends from the LGB community. These numbers are surprising, especially when viewed in the context of the frequency of homosexuality, as discussed below.

Frequency of homosexuality

Human sexuality is complex and defies binary classification, according to Alfred Kinsey:

"Males do not represent two discrete populations, heterosexual and homosexual. The world is not to be divided into sheep and goats. Not all things are black nor all things white. It is a fundamental of taxonomy that nature rarely deals with discrete categories. Only the human mind invents categories and tries to force facts into separated pigeon-holes. The living world is a continuum in each and every one of its aspects. The sooner we learn this concerning human sexual behavior the sooner we shall reach a sound understanding of the realities of sex." [10]

A number of studies have attempted to determine the frequency of homosexuality, although these studies have struggled with the complexity of sexual orientation. Is homosexuality defined by same-gender attraction, same-gender sexual activity, self-identification as LGB, or something else altogether? Studies have reported the frequency of homosexuality in several ways to capture the complexity of defining sexual orientation. For example, while Kinsey reported that 30 percent of the male population had incidental homosexual experiences or reactions at some point in their lives, only four percent of the male population was exclusively homosexual throughout their lives [11]. Because our research relies on public self-identification of same-gender interest in Facebook profiles as a sentinel value for LGB identity, Table 1 shows only frequency of homosexuality statistics with a high standard for determining homosexuality, such as same-gender sexual identity or primarily same-gender sexual activity. We assumed that Facebook users are unlikely to identify publicly a same-gender interest in their profiles because of a single homosexual experience, for example.

Table 1: Frequency of homosexuality. Note: These statistics are specifically gathered from the United States.		
Study (Year)	Men	Women
Kinsey, et al. (1948; 1953) primarily same-gender sex (includes bisexuals)	13.0%	7.0%
Laumann, et al. (1994) same-gender sexual identity (includes bisexuals)	2.8%	1.4%
Sell, et al. (1995) same-gender sex since age 15 (includes bisexuals)	6.2%	3.6%

When put in the context of Table 1, the fact that lesbians and gay males draw 55 percent of their close friends from the LGB community, while heterosexuals and bisexuals draw four percent, strongly suggests a correlation of LGB individuals' having more LGB friends. There are a number of possible explanations for such self-segregation phenomena. For example, people are already geographically separated, which therefore reduces opportunities for people to interact with others in different locales. Another theory is that individuals like to reinforce their self-identities, behaviors, and attitudes by associating with people similar to themselves. Regardless of their origins, homophily patterns are likely self-sustaining because family, friends, and other associates exert pressure to conform to their norms [12]. This phenomenon is known as "social proof." [13]

Taking a step back, if equal status contact is such a persistent empirical regularity in social relationships, how might such self-segregation manifest itself in online social relationships? As noted, people who identify as lesbian or gay have nearly an order of magnitude more close LGB friends than heterosexual individuals, so how does this fact then scale up and have any significance in a social networking system like Facebook, where a person may have hundreds of "friends?"

Facebook "friends" or real friends

In *The Tipping Point*, Gladwell presents the idea of a "social channel capacity." [14] He argues that our time and cognitive capacity to maintain relationships with people are limited, and the social channel capacity is an articulation of that limitation. For humans, the social channel capacity is about 150 people. Think of this group as "the number of people you would not feel embarrassed about joining uninvited for a drink if you ran into them in a bar." [15] On Facebook, many users have Facebook "friend" associations with well over 150 individuals. What kind of real-world relationships does the Facebook user have with his "friends" in excess of the channel capacity?

A phenomenon called the principle of locality heavily biases friendship formation in the real world based on proximity. Facebook builds on the principle of locality with the construct of "networks," which generally map to schools, businesses, and geographies. Users draw the majority of their Facebook friends from their "network," which implies that Facebook friends know each other in the real world. People use social networking Web sites like Facebook to augment their "off-line [social] network" with people they already know. According to Mikolaj Jan Piskorski at Harvard: "Interviews with [social networking Web site] users revealed significant interest in understanding their friends' social lives (and where respondents fit into it) — something that is hard to establish in the off-line world." [16] In other words, Facebook allows people to understand where they stand in a web of friendships, such that Facebook serves as a "coordinating system" for real-world social connections (Kochan, 2005).

Data mining

If Facebook really does map to real-world social relationships, what might we learn from analyzing those friendship connections? Such analysis falls under the term "data mining." Consider technology developed by the Air Force to detect insider threats by data mining e-mail messages (Okolica, et al., 2008). The

researchers built a social graph from a large corpus of e-mail messages, and detected individuals who may have been alienated or had a hidden agenda. If data mining can reveal hidden relationships, what might data mining reveal from a large corpus of data from a social networking site like Facebook?

Forming our hypothesis

Real-world self-segregation should carry over into online social networks. Because males have more male friends, and LGB individuals draw many of their friends from the LGB community, one would expect gay males on Facebook to have a higher proportion of gay male friends than heterosexual males. Because females have more female friends, and LGB individuals draw many of their friends from the LGB community, one would expect lesbians on Facebook to have more lesbian friends than heterosexual females.

Why focus on sexual orientation? Firstly, from a technical perspective sex and sexual orientation data is easy to access. Facebook profiles have fields for "Sex" — (male, female, blank) and "Interested In" — (men, women, men and women, blank). Note that Facebook does not adequately support the complexity of human sex and sexual orientation, and therefore certain subjects, such as transgender identities, cannot be addressed by our study. Regardless, the "Sex" and "Interested In" fields are easily read and tallied by computer. The other fields in a Facebook profile are editable text fields that maintain no invariants to restrict the contents of the field. For example, the "Religion" text field is complex because there are many different religions around the world, many of which can be represented with multiple names like "Catholic" versus "Roman Catholic."

Secondly, gender, race, and even sexual orientation may be considered "master status" categories, characteristics one cannot help but notice when observing a subject. According to one study: "perception of sexual orientation rests, at least in part, on the perception of the body's shape in motion." [17] By observing the gait of another person, an observer can determine the sexual orientation of that person with accuracy above chance [18]. The ability to detect such characteristics without physically observing a subject introduces a new threat to privacy.

If we could indeed find a strong correlation between friendships and sexual orientation, it would represent a significant privacy risk because network data — data that relates one user to another — is not generally considered sensitive information and is afforded little protection under the Fourth Amendment in the U.S. For example, although a warrant is required to obtain a wiretap, a warrant is not required to log telephone numbers dialed (*Smith v. Maryland*, 1979).



Methodology

Sexual orientation is impossible to manipulate as an experimental variable, so designing an experiment to determine whether a causal relationship exists between sexual orientation and friendship is no easy task. We conducted a correlational study using archival data recorded by Facebook. Such an experimental design has a major shortcoming: it cannot determine causation, although it can reveal correlation.

Subjects

Our study targeted MIT students who used Facebook and listed MIT as their primary Facebook network. Individuals were not included in our study for the following reasons only: (a) Not having a Facebook account; (b) Having Facebook privacy settings that made their profiles inaccessible; (c) Having a primary network other than the following: MIT, MIT '07, MIT '08, MIT '09, MIT '10, MIT '11, MIT Grad Student '07, MIT Grad Student '08, MIT Grad Student '09, MIT Grad Student '10, MIT Grad Student '11, MIT Grad Student '12, MIT Grad Student '13, MIT Grad Student '14, MIT Alum '07; (d) or having fewer than 12 Facebook friends, as described in the [Detecting "abandons"](#) section.

Our subjects were "recruited" by an automated spider called Arachne, which is discussed in the following section. After filtering by sub-network and detecting abandons, our dataset comprised Facebook profiles of 6,077 students associated with MIT. Of these, 4,080 disclosed their sex and these are broken down in [Table 2](#). Our subjects were 42 percent male, 25 percent female, and 32 percent unreported. For comparison, [Table 3](#) contains statistics on the entire MIT student population during fall 2007 when we collected our data.

Table 2: Subject demographics.		
Note: Sex and sexual orientation are self-reported.		
Self-reported sexual orientation	Male	Female
Heterosexual	1,544	762
Bisexual	21	35
Homosexual	33	12
Not reported	947	726

Table 3: 2007 MIT student enrollment.		
Source: MIT, 2007a; 2007b.		
Class	Male	Female
Undergraduate	2,315	1,857
Graduate	4,226	1,822

Procedure

Arachne

The sheer volume of data we wished to gather required automated collection of profile and friend information from Facebook. We needed a "spider," a program that automatically and methodically browses the World Wide Web. Google and other major search engines use "spider" or "crawler" technology. Our spider, called Arachne, (a) signed into Facebook, (b) received cookies from Facebook, and (c) downloaded Web pages with profile and friend information for each member of the MIT network. Because we were downloading thousands of pages from Facebook, Arachne ran continuously from 24–29 October, 31 October–5 November, and 7–12 November 2007. This data collection was performed without Facebook's permission.

Consider Facebook to be a large social graph. Each user is a vertex of that graph and friendship between two users is an edge. Assuming that the MIT network is a connected graph, it should be possible to traverse the entire graph from any starting point. Essentially, Arachne performed a breadth-first search on the graph, starting with Facebook id 700146, the id of one of the authors, as the root of the search tree.

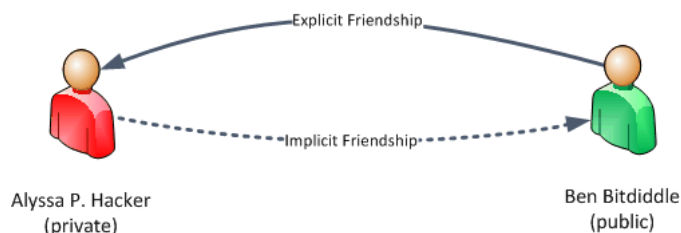
It should be noted that Arachne was conditioned to search for URLs embedded in Facebook friendship lists related to messaging. In order for Arachne to have catalogued a Facebook user, that user must have: (a) been friends with id 700146 because friends can always view a friend's profile, or (b) permitted messaging from non-friends. See [Figure 2](#) for an example of a Facebook profile that Arachne could not catalogue. A cursory analysis suggested that such anti-messaging privacy settings screened no more than two percent of profiles.



**Figure 2:** Example of Facebook friends who do not allow messaging.

#### *Implicit friendships*

Implicit friendships provide the basis for detecting the sexual orientation of individuals who make their Facebook profiles and friendship associations private. Suppose that Alyssa P. Hacker and Ben Bitdiddle are Facebook friends. Alyssa's profile is private, while Ben's profile is public. As a third party, one cannot see that Alyssa is friends with Ben, although one can see that Ben is friends with Alyssa. [Figure 3](#) visualizes this implicit friendship.



**Figure 3:** Example of implicit and explicit friendship associations. Although Alyssa's profile is private, her friendship with Ben can be discovered simply by viewing Ben's public Facebook profile.

#### *Detecting "abandons"*

A small percentage of Internet users "abandon" using a given Internet service, briefly trying the service but never returning (Jupiter Research, 2006). Although many Facebook users have hundreds of friends and 45 percent of users return to the site daily (Facebook, 2009), a small fraction of Facebook users have only a few friends and have likely abandoned their profiles. To prevent users with few friends from skewing the results of our study, we implemented an abandonment detection algorithm.

Upon creating a new account, a new Facebook user has zero friends. Adding friends is a multistep process, and adding a large number of friends is time-consuming. The possibility of abandonment declines as the number of friends, and therefore the time investment, increases. Robert Cialdini in his book *Influence: Science and practice* presents the notion of "commitment and consistency":

"A study done by a pair of Canadian psychologists (Knox and Inkster, 1968) uncovered something fascinating about people at the racetrack: just after placing bets they are much more confident of their horses' chances of winning than they are immediately before laying down the bets. Of course, nothing about the horse's chances actually shifts; it's the same horse, on the same track, in the same field; but in the minds of those bettors, its prospects improve significantly once that ticket is purchased." [19]

Once a person commits to something, even if the commitment is small, he is more likely to be consistent and stick to that commitment. This is the "foot in the door" technique. The time investment of adding friends on Facebook is an act of commitment, which reduces the likelihood that a user abandons his account.

What might be the commitment threshold for Facebook users? Upon first logging on, a new Facebook user might call to mind the friends in his "sympathy group" and during that first session might add those 12 friends and family members, spending about fifteen minutes doing so. And so he is committed. We elected to use this 12-friend threshold to detect abandonment, because friend data are usually available, even for private profiles, due to implicit friendships. Our abandonment algorithm removed all profiles with less than 12 friends in the MIT network from our directed graph model. Because the edge counts of the remaining nodes were reduced by this removal, the process was iteratively repeated until the number of nodes in the graph remained unchanged.

#### *Statistical methods and classifier design*

For our analysis, we partitioned our data into six sex orientation groups:

1. heterosexual females
2. heterosexual males

3. bisexual females
4. bisexual males
5. homosexual females
6. homosexual males

We analyzed the friends of users who self-reported in each of these sex orientation groups by finding the percentage of friends that fell into each sex orientation group. We then created a simple logistic regression model that was only dependent on one parameter — the percentage of a person's Facebook friends that identified as gay male.

A logistic regression model produces a risk score,  $[0, 1]$ , where 1 implies the input has the attribute the model tests for, and 0 implies the input does not. This risk score is particularly useful for binary prediction; one can designate a specific threshold for the risk score at or above which the result is "yes" and below which the result is "no." There is a fundamental tradeoff between specificity, a measure of the model's susceptibility to false positives, and sensitivity, the probability the model correctly predicts a characteristic (Signorovitch, 2007).

To build our logistic regression model, we inputted all subjects that self-reported as gay male and all subjects that self-reported as bisexual or heterosexual male. Then we designated a risk score threshold to provide binary classification indicating a gay male user or a bisexual/heterosexual male user. To decide on a threshold and evaluate the tradeoff between sensitivity and specificity, we plotted a Receiver Operating Curve (ROC). The ROC curve plots sensitivity as the dependent variable versus one minus specificity as the independent variable and is evaluated by the Area under the Curve (AUC), which measures how accurately subjects are ranked by their risk score. An AUC of 0.7 or greater is satisfactory for clinical work (Signorovitch, 2007).



## Results

### Overview

For each user who self-reported in one of the sex orientation groups mentioned in the [Statistical methods](#) section, we found the percentage of friends that fell into each sex orientation group. We then averaged these percentages over the entire sex orientation group. [Table 4](#) provides a summary of these statistics. [Figure 4](#) shows a subset of the data in Table 4, highlighting the percentage of LGB friends per sex orientation group and revealing that a gay male has, on average, a much higher percentage of gay male friends than the other groups. As can be seen from these data, heterosexual males have 0.7 percent gay male friends on average ( $SD = 0.01$ ), while self-identified gay males have 4.6 percent gay male friends on average ( $SD = 0.05$ ).

Table 4: Percentage friends per sex orientation group.						
Sex orientation group	Percentage friends per group					
	Heterosexual	Bisexual	Homosexual			
<b>Heterosexual</b>						
Female	19.0%	22.4%	0.7%	0.5%	0.4%	0.8%
Male	13.9%	28.3%	0.5%	0.4%	0.3%	0.7%
<b>Bisexual</b>						
Female	15.5%	20.7%	1.4%	1.1%	0.3%	1.2%
Male	12.6%	22.3%	0.8%	0.6%	0.3%	1.9%
<b>Homosexual</b>						
Female	18.0%	23.6%	0.9%	0.7%	0.2%	0.8%
Male	13.1%	21.4%	1.1%	1.1%	0.4%	4.6%

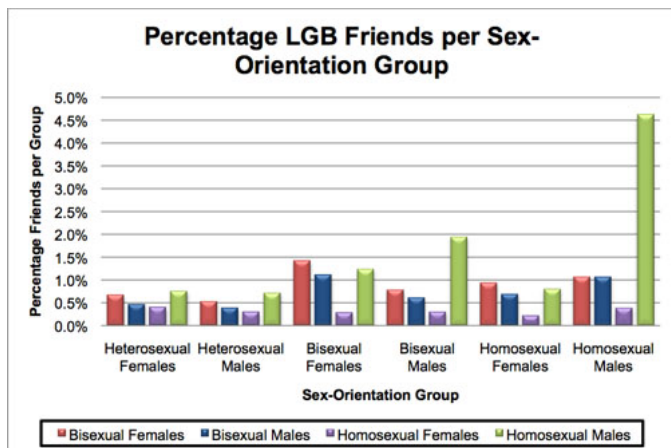


Figure 4: Percentage of LGB friends per sex orientation group.

### Implicit friendships

Recall that friendship associations of Facebook users with private profiles can be determined implicitly. A strong correlation exists between explicit and implicit friendship associations on Facebook. Across subjects with public profiles, the average number of explicit friends was 112.7, while the average number of implicit friends was 96.3. The Pearson coefficient between implicit and explicit friends of public profiles from Table 2 is  $r = 0.99$ , with  $n = 5,302$ . Implicit and explicit friends are *highly* correlated.

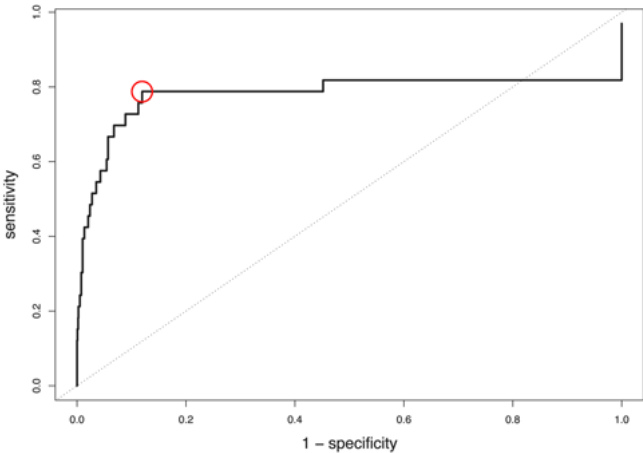
### Classifier

[Figure 5](#) shows the ROC curve of a logistic regression model generated from data on all MIT Facebook users that self-reported as gay male and all MIT Facebook users that self-reported as bisexual male or



heterosexual male. The presented ROC has a stellar AUC value of 0.83, much better than the clinical standard of 0.7.

The circle at the elbow of the ROC curve in Figure 5 indicates the sensitivity–specificity pairing we chose from which to derive our threshold value for our logistic regression. The threshold value that we arrived at was 1.89 percent: if more than 1.89 percent of a male Facebook user’s friends self-identified as gay, our logistic regression model classified that Facebook user as gay. The threshold value corresponds to a sensitivity of 0.78 and a specificity of 0.83, meaning that the threshold derived from this pairing correctly classified gay males from the training set 78 percent of the time and correctly classified non-gay males 83 percent of the time.



**Figure 5:** ROC curve of a logistic regression model generated from male MIT Facebook users that self-reported as gay versus users that self-reported as bisexual or heterosexual. The red circle at coordinates (0.78, 0.17) indicates the specificity (1 – 0.17 = 0.83) and sensitivity (.78) thresholds based on the results from the training set. A perfect correlation would appear as an L-shape with the elbow in the upper left corner at (0, 1). A completely random result would be appear as a diagonal line from (0, 0) to (1, 1).

The ROC classifier makes 17 percent Type 1 errors and 22 percent Type 2 errors. It should be noted that the costs of misclassification might be asymmetrical. For example, the error of misclassifying a Facebook user as a homosexual male might be perceived as more serious in certain contexts than misclassifying a user as a heterosexual male.

Validation

While the results shown in Figure 4 are interesting, the purpose of this research was to test whether one can *predict* information that a user keeps private. To test this hypothesis, we created a validation dataset of subjects who we knew to be gay male, as a privilege of our real-world acquaintances with them. [Table 5](#) reports the results of analyzing the friends of these subjects.

Table 5: Percentage of gay friends of gay male subjects known to the authors <i>a priori</i> .					
Note: The percentage column indicates the ratio of gay male friends to all friends in the MIT network.					
Name	Profile privacy setting	Reported sex	Reported orientation	Percentage gay friends	Classified as gay
A	Private	Unknown	Unknown	13.21%	True
B	Not private	Male	Gay	10.18%	True
C	Not private	Male	Unknown	8.09%	True
D	Not private	Unknown	Unknown	7.10%	True
E	Not private	Male	Gay	5.53%	True
F	Not private	Male	Gay	5.26%	True
G	Not private	Male	Gay	5.10%	True
H	Private	Unknown	Unknown	4.56%	True
I	Private	Unknown	Unknown	4.19%	True
J	Not private	Male	Gay	3.96%	True
K	Private	Unknown	Unknown	3.80%	True
L	Not private	Male	Gay	3.70%	True
M	Not private	Male	Unknown	3.68%	True
N	Not private	Male	Unknown	3.30%	True
O	Not private	Male	Gay	3.27%	True
P	Private	Unknown	Unknown	2.86%	True
Q	Not private	Male	Gay	2.78%	True
R	Private	Unknown	Unknown	2.65%	True

S	Not private	Male	Gay	2.53%	True
T	Not private	Male	Gay	2.16%	True
Average heterosexual male				0.72%	False

## Discussion

Our results show strong predictive power for the sexual orientation of male MIT Facebook users. Table 5 shows statistics for members of our validation dataset, composed exclusively of gay males known to the authors *a priori*; all had between 3.0 and 18.3 times higher of a percentage of gay male friends than the average male user that reported as heterosexual. Such numbers are remarkably high and indeed were high enough for our logistic regression classifier to categorize these individuals correctly. We were only able to obtain sensitivity and specificity numbers from those that self-identified a sex and sexual orientation in their Facebook profile. This metric measures our model's ability to predict subjects that self-identify as gay male, rather than an ability to generally predict those on Facebook that are gay male.

Notice the members of our validation dataset whose profiles were private; the data is quite sobering. Without any information about a Facebook user beyond a list of his friends, one can accurately predict his sexual orientation.

As to why our hypothesis does not hold for lesbians, two possibilities are: our comparatively small sample size or even fundamental differences in the structure of lesbian communities. This is a topic for future research.

### Threats to validity

Although it is impossible to arrive at a causal inference due to our study's design, we have shown strong correlation. Despite such a strong correlation, it is difficult to eliminate all threats to the validity of our study.

#### Selection

While the validation dataset provides a startling example of how this research violates the privacy of individuals who try to make their Facebook profiles private, it was derived from individuals known to the authors to be gay males *a priori*. We consider this to be the single greatest source of selection bias in our study.

We limited our subjects to current MIT students, which introduced selection bias into our study. For one, we inherited the selection process of the MIT admissions committee. Secondly, MIT students may not reflect Facebook users as a whole, and there is self-selection bias in that not all MIT students join Facebook, although Facebook does maintain 85 percent market share of four-year U.S. universities (Facebook, 2009). There is additional self-selection bias in that some users decided to change their default privacy settings.

We detected MIT Facebook users based on their primary network affiliation. Joining the MIT network on Facebook requires an @MIT.edu or @alum.MIT.edu e-mail address. As long as MIT restricts such e-mail addresses to those affiliated with MIT, the MIT portion of the network on Facebook is somewhat difficult to forge. We then limited our subjects to active MIT students at the time of our study based on class year. It should be noted that Facebook users within the MIT network can switch between the different sub-networks without any verification. Hypothetically, a MIT alumnus from 1979 could have changed his network to MIT '11 and would have been included in our study. In practice, we believe such behavior is minimal because self-identifying with the wrong sub-network reduces Facebook's utility; the sub-network aids in finding friends within a class year on Facebook.

#### False reporting

Because Facebook makes no explicit request for accuracy when a user creates a profile and users exercise control over their profile content, our study might be vulnerable to false reporting. There has been some research analyzing the veracity of Facebook profile information. One such study concluded that there were "generally strong patterns of convergence" between the data that Facebook users self-report and the perceptions of those users by their friends (Gosling, *et al.*, 2007). Another perspective uses signaling theory: "A public display of connections [friendships] can be viewed as a signal of the reliability of one's identity claims." [20] By making friendship associations public, there is increased accountability at the cost of privacy. These studies suggest that the general accuracy of information on Facebook user profiles is high.

One method of verifying reporting accuracy is to compare our collected data against known trends. When comparing the homosexuality rates of our subjects to those discussed in the [Frequency of homosexuality](#) section, we observe a much lower incidence. Such a statistic does not disqualify our analysis. It may be the case that there is a lower incidence of homosexuality in the MIT student population compared to those populations examined in the cited studies. Alternatively, fewer users may choose to report their sexual orientation in a public forum such as Facebook.

#### Maturation

Is the gay male friendship correlation we observed based on real-world social groups that quickly assimilate new members of the MIT community? People who identify as active members of the LGB community might take part in LGB events at MIT immediately after arriving to campus. The friendship patterns of some individuals support this notion. [Figure 6](#) shows the associations between self-identified gay male subjects in our study, where the ovals represent individual Facebook users and the lines represent Facebook friendships. Consider p1 and p13, which have a degree around fifteen, which is a noticeably higher degree than the other vertices in the graph. Could these "hubs" be the assimilators of the MIT gay community?



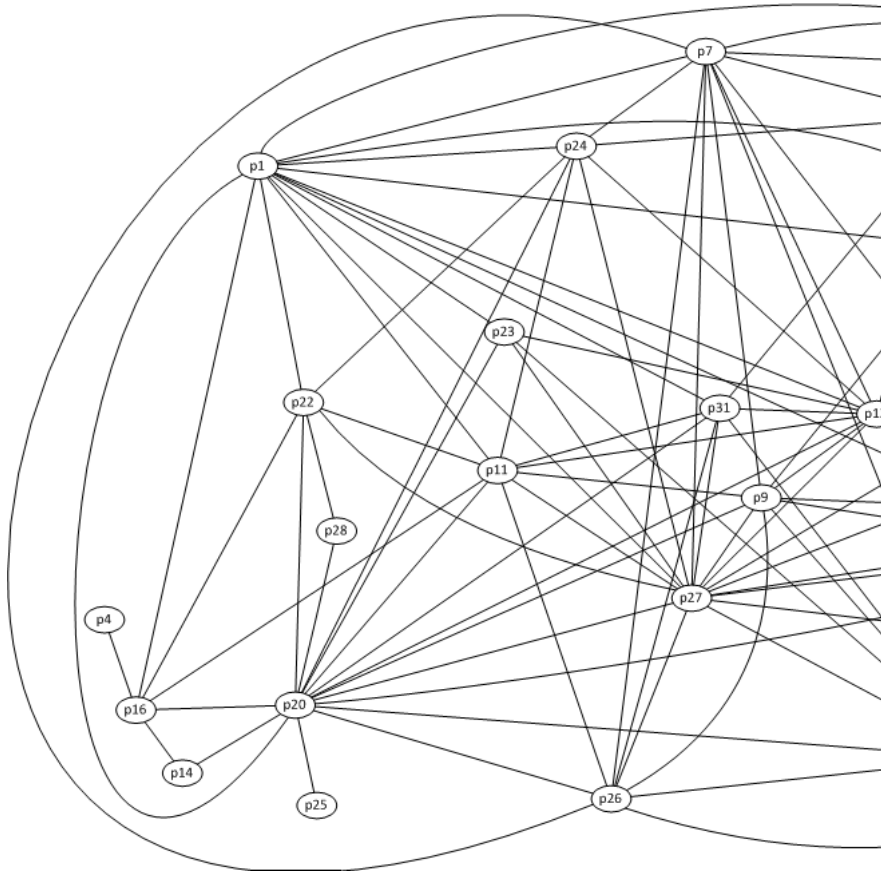


Figure 6: Social graph revealing Facebook friendship associations between self-identified g

The transition to a gay or lesbian sexual orientation may involve different stages. "For some individuals, bisexuality represents a developmental stepping stone on the way to forming a gay or lesbian sexual orientation." [21] This may explain the disproportionate number of bisexual individuals in our data. As discussed previously in the [Frequency of Homosexuality](#) section, one would expect only 0.8 percent of the male population to report as bisexual and between 2.8 percent and 13 percent of the male population to report as gay. Our data on the other hand show male bisexuality being quite close in frequency to male homosexuality with 21 males reporting as bisexual and 33 males reporting as gay.

Instrumentation

Because the data were collected over a period of about two weeks, we do not have an instant snapshot of the Facebook social network. The network continued to grow and became more connected over the course of spidering. This caused the implicit and explicit friends to be inconsistent. Let us suppose the spider downloads Ben Bitdiddle's profile and friends list. After downloading Ben's information, Ben and Alyssa P. Hacker become Facebook friends. When the spider subsequently downloads Alyssa's profile and friends list, it will see that Alyssa is friends with Ben but will not update Ben to be friends with Alyssa. This type of instrumentation error was corrected by taking the intersection of implicit and explicit friends. In other words, all friends = implicit friends  $\cap$  explicit friends. A few profiles were found to be outliers. For example, one user deactivated his Facebook account and reactivated it halfway through our data collection process. These events caused his implicit friendships to be nearly zero, while his explicit friendships were correct. The intersection of both sets of friends resolved that discrepancy.

We used one of the authors as the root of the breadth-first search of the MIT network. He had 219 Facebook friends, a good set of seed profiles. The success of Arachne was predicated on the social network's being well-connected. If some set of profiles were orphaned from the rest by maintaining its own subgraph, Arachne would have missed downloading those profiles. Considering that our average subject had about 100 friends, the network was likely well-connected. Using the Facebook account of one of the authors to perform the spidering also represents a source of bias. Because Facebook uses friendship associations as a method of access control, our dataset potentially contains privileged information on some of that author's Facebook friends whose profiles would otherwise have been inaccessible. The overall effect of this privileged information on our results is likely small.

History

National Coming Out Day was 11 October 2007, thirteen days before the beginning of our study. Although unlikely to have caused a huge effect, it is possible that this event increased the number of self-reporting LGB individuals on Facebook immediately prior to the beginning of our study.

Ethics and consequences

Our analysis demonstrates a method of classifying sexual orientation of individuals on Facebook, regardless of whether they chose to disclose that information. Facebook users who did not disclose their sexual orientation in their profiles would presumably consider the present research an invasion of privacy. Yet this research uses nothing more than information already publicly provided on Facebook; no interaction with subjects was required. Although we based our research solely on public information, only a limited subset of our results, which contain no personally identifiable information, is presented in this paper to maintain subject confidentiality.

[Table 6](#) shows a subset of users from our validation dataset (see [Table 5](#)) who specified Facebook privacy settings that prevented non-friends, such as Arachne, from viewing their profiles. Despite having private profiles, the classifier correctly distinguished these individuals as gay males. Such classifications are startling not only for their accuracy but also for their potential ramifications in the LGB community.

Table 6: Subjects presented in Table 5 with private profiles accurately classified as gay male.				
Profile	Reported	Reported	Percentage	Classified

Name	privacy setting	sex	orientation	gay friends	as gay
A	Private	Unknown	Unknown	13.21%	True
H	Private	Unknown	Unknown	4.56%	True
I	Private	Unknown	Unknown	4.19%	True
K	Private	Unknown	Unknown	3.80%	True
P	Private	Unknown	Unknown	2.86%	True
R	Private	Unknown	Unknown	2.65%	True

In a study of sexual orientation and suicidal behavior at the University of Washington, students who identified as heterosexual but reported same-sex attraction or behavior were six times as likely to have attempted suicide in the previous year as opposite-sex attracted heterosexuals [22]. Students who identified as LGB were only twice as likely as opposite-sex attracted heterosexuals. Victimization over one's same-sex attraction or behavior was the risk factor that increased suicidal behavior for all groups. But why would same-sex attracted heterosexuals be worse off than LGB identified individuals? One theory is that these same-sex attracted heterosexuals are still in the process of developing their sexual orientation identity. Although these students might eventually identify as LGB, they have not reached the point of coming out. During that explorative period of sexual orientation identity development, discrimination may have a stronger negative effect. Our own study revealed a disproportionate number of bisexual subjects at MIT compared to the literature on frequency of bisexuality. These subjects may be using a bisexual identity as a stepping stone to a gay identity and could be at a higher risk of suicide if research like ours were used for discrimination. On the other hand, research like ours could possibly be used by mental health professionals to identify high-risk individuals and reduce the incidence of suicide.

Although same-sex marriage is a political hot topic, 29 states have no laws to protect LGB employees as of September 2009 (Human Rights Campaign, 2009). Although the Employment Non-Discrimination Act of 2009 is currently moving through Congress to prohibit employer discrimination based on sexual orientation, the LGB community still faces significant challenges on the discrimination front. Gay men face a "glass ceiling" not unlike that historically experienced by women in the workforce. According to a recent study, "cohabitating gay men earn nine percent less than unmarried cohabiting heterosexual men," although this discrimination is concentrated in management and blue-collar jobs [23].

Throughout our analysis we have left aside the troubling question of whether the identities of an individual's cohorts should shackle him to a particular identity in the eyes of a network data miner. Although U.S. Supreme Court precedent supports the notion of "freedom of association" (*NAACP v. Button*, 1963), network data eats away at that freedom. Until now, we have made little mention of our reliance on inference. Although the relations that we described are highly predictive, the threshold that we determined from our ROC curve ensured both false positives and false negatives. However, such a question becomes even more disturbing when one realizes that such a model has a statistical chance of being wrong.

#### Suggestions for Facebook

Facebook exists in two spaces: technical and user. In the technical space, Facebook's source code explicitly governs what users can and cannot do. In the user space, Facebook promotes a code of conduct governing what users should and should not do.

##### Technical space

Facebook provides users a number of explicit privacy controls to restrict who has access to their profiles, which were discussed in the [Facebook overview](#) section. Unfortunately, these explicit controls create a mirage of privacy that fades upon closer inspection. Despite the fact that the users exemplified in Table 6 had strict privacy settings, we were still able to correctly classify their sexual orientation. There are gaps in Facebook's own security procedures, so what else can Facebook do from a technical perspective?

Steve Jobs called digital rights management a "cat-and-mouse game" in his open letter to the computer and music industries (Jobs, 2007). Computer security is likewise a "cat-and-mouse game." No system is ever fully secure: for each security improvement made, new attacks are devised. Even so, Jobs believes there is significant value in security systems that "keep honest users honest." Such barriers stop casual intrusions, although they still do not block the more sophisticated hackers.

After performing the majority of our research, we learned that Facebook has monitors to detect scripted access and that Facebook deactivates users for "misuse" of the service (Kelly, 2008). Although Arachne was never designed for stealth, it somehow did not trigger these monitors when downloading 18,000 Facebook profiles over a two-week period. A simple threshold of maximum profile views per user per day could "keep honest users honest," while simultaneously blocking a simple spider such as Arachne. A maximum profile views per day threshold may not be enough though.

A distributed spider could also overcome such a maximum profile views threshold using a few dozen accounts, each making a fraction of the requests that a single account would, although a distributed spidering scheme requires far more sophistication. Regardless, Facebook must do more to impede overt spidering of their site.

Further, Facebook could impede the screen scraping process. Facebook's XHTML is human readable, simplifying the task of crafting XPath queries that extract data from Facebook profiles. As an example, the query `"/div[@id='Gender-data']"` returns the sex from a Facebook profile [24]. Facebook could eliminate obvious attribute tags such as "Gender-data" or "birthday," with less meaningful alternatives. Facebook could also introduce randomization into the XHTML source without altering page layout, which would make extraction of information much more difficult with XPath.

##### User space

Facebook allows users to create profiles online that can contain intimately sensitive information; yet Facebook does not truly explain the consequences of posting this information online. In many respects, this paper may have exploited Facebook's cursory notification. Given the results of our research, the users with private profiles, who our classifier accurately designated as gay in Table 5, might consider even stricter privacy settings or leaving Facebook altogether.

Currently, Facebook only describes the functional effect of specifying particular privacy settings. However, Facebook does not contextualize these controls by telling a user why he might want to keep particular information privileged, nor does it give examples of the potential consequences of revealing particular information. Such a lack of disclosure may give users the feeling that the provided controls are arbitrary inconveniences rather than protections against real threats. Consider that at the time this research was conducted, the average Facebook user could view only 0.15 percent of all Facebook profiles due to network boundaries (Kelly, 2008), although 0.15 percent of Facebook's then 68 million strong user base was still 102,000. Would you want your information visible to 102,000 people? Additional candor by Facebook may only crudely address the techniques employed by this paper; however, they would empower users to make more informed decisions.

Facebook has historically taken an opt-out approach when releasing new features on its Web site, although opt-in approach would greatly enhance user privacy. Consider Facebook's Beacon program, which launched in November 2007 and immediately began gathering additional private information about Facebook users' interactions with affiliate Web sites without the users' consent. In one instance, users were shocked to find holiday purchases from Overstock.com appearing on Facebook for all their friends to see, ruining the surprise of the gift (Holahan, 2007). An opt-in approach would have reduced Beacon's value to Facebook but simultaneously would ensure that user privacy is protected. An opt-in approach promotes the principle of least surprise by maintaining the status quo. Facebook subsequently shut down Beacon to settle a class-action lawsuit (Ortutay, 2009).

How would an opt-in approach have prevented our study? The default privacy settings on Facebook provide no protection to our data mining technique, because new Facebook profiles are visible to their entire network by default. Only 25 percent of Facebook users change the default privacy settings on their profiles according to Facebook's Chief Privacy Officer (Kelly, 2008). Over 75 percent of Facebook users are vulnerable to the methods presented in this paper! In traditional software, user customization is significantly higher. In a study of WordPerfect users, 92 percent of participants customized the software, making an average of 9.1 changes over a 28-day period [25]. Clearly, there is a disparity between the customization of Facebook privacy settings versus other computer software. Because Facebook's customization percentage is so low, Facebook should change its default privacy settings to stricter standards, requiring users opt-in to make themselves susceptible to our data mining techniques.

The problem of network data has no neat solutions. Perhaps the only concrete answer is to educate users. Increased candor by Facebook would empower its user base to make more informed decisions about what information to share. Anything less than complete disclosure by Facebook falls far short of this mark. As such, we encourage Facebook to support this and other research.




## Conclusion

The world is filled with companies and services trafficking in network data—data that relates one person to another. E-mail providers, telephone companies, instant messaging services, and social networking Web sites know with whom you communicate. Such information can reveal profoundly intimate details of a person's life. In particular, we accurately predicted a Facebook user's sexual orientation based on information about that user's Facebook friends. The privacy controls of Facebook, a multi-billion dollar corporation, offer anemic protection against such an analysis: our model built from relatively simple network data was mostly unimpeded by Facebook's privacy efforts. Future extensions of this work need not be limited to Facebook and could be applied to telephone call records or even e-mail transactions, as those communications rely on social connections. Who is to say that companies are not already doing the type of network analysis presented here behind closed doors?

Extensions of our work to other networks has profound ramifications. Network data shifts the locus of information control away from individuals. Each individual's traditional and absolute discretion is replaced by that of members of his social network. One's coworkers, friends, family, and acquaintances, as well as one's associations with them, implicitly contain private information. Although our research focuses on sexual orientation, there are many possible extensions. For example, in studies of friendships among college students, males are more likely to choose their friends based on common activities [26].

Unfortunately, few apparent solutions truly address the problems identified. The weaknesses described and exploited in this analysis are not a fault of a particular company or service but an inherent property of any service that tracks human interactions. With the increasing prevalence of such services, what recourse do individuals have? The individual truly concerned with his privacy might unplug from the world. Scott Adams, the creator of *Dilbert*, suggests an alternative:

"Let's say that someday technology will allow anybody to find out every possible thing about my life. I can compensate by being so uninteresting that nobody could survive the process of snooping on me without lapsing into a coma." [27] 

## About the authors

**Carter Jernigan** received his B.S. in Computer Science and Engineering from the Massachusetts Institute of Technology in 2008, with additional studies in psychology and management. Carter is currently employed as a software engineer, but his career in the computer industry began when he founded his own business, The Computer Coach of Charleston, LLC., at age twelve. Carter's other projects include *Locale*, a \$300,000 grand prize winner of the Android Developer Challenge. "My goal in life will always be to invent something cool and unusual," says Carter. In his spare time, Carter practices to become an expert mixologist and also started Eat Your Way Through Boston, a club to explore new restaurants around Boston.

**Behram F.T. Mistree** received his B.S. in Electrical Engineering and Computer Science in 2007 and his M.Eng. in Electrical Engineering and Computer Science in 2008 from the Massachusetts Institute of Technology. He is currently pursuing his Ph.D. in Electrical Engineering at Stanford University. Behram is interested in a range of subjects, but daydreams most frequently about novel sensor networks and systems. He hopes his current studies are preparing him to approach and solve fun interdisciplinary problems that help people. In his spare time, he reads some fantasy fiction, cooks bread, and thinks about technology.

Correspondence concerning this article should be addressed to gaydar [at] mit [dot] edu.

## Acknowledgements

Behram's work was indirectly supported by a National Science Foundation Graduate Research Fellowship under Fellowship no. 2007050798.

## Notes

<sup>1</sup> After we collected our data, Facebook added additional features allowing its users to create custom groupings of friends with differing levels of access to their profiles.

<sup>2</sup> Laumann, *et al.*, 1994, p. 16.

<sup>3</sup> Maccoby, 1998, p. 22.

<sup>4</sup> Galupo, 2007, p. 143.

<sup>5</sup> Galupo, 2007, p. 139.

<sup>6</sup> Gladwell, 2002, pp. 176, 177.

<sup>7</sup> Galupo, 2007, p. 143.

<sup>8</sup> Galupo, 2007, p. 145.

<sup>9</sup> Galupo, personal communication, 10 December 2007.

<sup>10</sup> Kinsey, *et al.* 1948, p. 639.

<sup>11</sup> Kinsey, *et al.* 1948, pp. 650–651.

12. Laumann, *et al.*, 1994, p. 16.
13. Cialdini, 2001, p. 100.
14. Gladwell, 2002, p. 177.
15. Gladwell, 2002, p. 179.
16. Piskorski, 2007, pp. 28–29.
17. Johnson, *et al.*, 2007, p. 332.
18. Johnson, *et al.*, 2007, p. 321.
19. Cialdini, 2001, p. 53.
20. Donath and boyd, 2004, p. 73.
21. Brannon, 2005, p. 289.
22. Murphy, 2007, p. 84.
23. Elmslie and Tebaldi, 2007, pp. 448, 451.
24. Note that Facebook is inconsistent by providing a field called "Sex" — (male, female, blank) in the Facebook profile presented to users which maps to a field called "Gender-data" in the actual XHTML source.
25. Page, *et al.*, 1996, p. 342.
26. Brannon, 2005, p. 221.
27. Adams, 1997, p. 208.

## References

- S. Adams, 1997. *The Dilbert future: Thriving on stupidity in the 21st century*. New York: HarperBusiness.
- L. Brannon, 2005. *Gender: Psychological perspectives*. Fourth edition. Boston: Pearson/Allyn and Bacon.
- R.B. Cialdini, 2001. *Influence: Science and practice*. Fourth edition. Boston: Allyn and Bacon.
- J. Donath and d. boyd, 2004. "Public displays of connection," *BT Technology Journal*, volume 22, number 4, pp. 71–82. <http://dx.doi.org/10.1023/B:BTJ.0000047585.06264.cc>
- B. Elmslie and E. Tebaldi, 2007. "Sexual orientation and labor market discrimination," *Journal of Labor Research*, volume 28, number 3, pp. 436–453. <http://dx.doi.org/10.1007/s12122-007-9006-1>
- Advocate, 2007. "ENDA, State by State" (14, 20 November).
- Facebook, 2009. "Facebook | Statistics," at <http://www.facebook.com/press/info.php?statistics>, accessed 26 July 2009.
- Facebook, 2007. Facebook | Terms of Use (15 November), at <http://www.facebook.com/terms.php>, accessed 2 April 2008.
- R.J. Foley, 2008. "Worker snooping on customer data common," *Associated Press* (28 February), at [http://news.yahoo.com/s/ap/20080223/ap\\_on\\_hi\\_te/snooping\\_workers](http://news.yahoo.com/s/ap/20080223/ap_on_hi_te/snooping_workers), accessed 4 March 2008.
- A. Fuller, 2006. "Employers snoop on Facebook," *Stanford Daily* (20 January), at <http://daily.stanford.edu/article/2006/1/20/employersSnoopOnFacebook>, accessed 3 April 2008.
- M.P. Galupo, 2007. "Friendship patterns of sexual minority individuals in adulthood," *Journal of Social and Personal Relationships*, volume 24, number 1, pp. 139–151. <http://dx.doi.org/10.1177/0265407506070480>
- M. Gladwell, 2002. *The tipping point: How little things can make a big difference*. Boston: Back Bay Books.
- S.D. Gosling, S. Gaddis, and S. Vazire, 2007. "Personality impressions based on Facebook profiles," *International Conference on Weblogs and Social Media* (27 March), at <http://www.icwsm.org/papers/3--Gosling-Gaddis-Vazire.pdf>, accessed 18 November 2007.
- R.A. Guth, V. Vara, and K.J. Delaney, 2007. "Microsoft bets on Facebook stake and Web ad boom," *Wall Street Journal* (25 October), p. B1.
- C. Holahan, 2007. "Facebook may revamp Beacon," *BusinessWeek* (28 November), at [http://www.businessweek.com/technology/content/nov2007/tc20071128\\_366355.htm](http://www.businessweek.com/technology/content/nov2007/tc20071128_366355.htm), accessed 12 December 2007.
- Human Rights Campaign, 2009. "Workplace," at [http://www.hrc.org/issues/workplace/workplace\\_laws.asp](http://www.hrc.org/issues/workplace/workplace_laws.asp), accessed 3 October 2009.
- S. Jobs, 2007. "Thoughts on music" (6 February), at <http://www.apple.com/hotnews/thoughtsonmusic/>, accessed 11 December 2007.
- K.L. Johnson, S. Gill, V. Reichman, and L.G. Tassinari, 2007. "Swagger, sway, and sexuality: Judging sexual orientation from body motion and morphology," *Journal of Personality and Social Psychology*, volume 93, number 3, pp. 321–334. <http://dx.doi.org/10.1037/0022-3514.93.3.321>
- Jupiter Research, 2006. "Retail Web site performance: Consumer reaction to a poor online shopping experience" (1 June), at [http://www.akamai.com/dl/reports/Site\\_Abandonment\\_Final\\_Report.pdf](http://www.akamai.com/dl/reports/Site_Abandonment_Final_Report.pdf), accessed 28 November 2007.
- C. Kelly, 2008. "Facebook," *Harvard Journal of Law & Technology Symposium*, Harvard Law School (13 March).
- A.C. Kinsey, W.B. Pomeroy, C.E. Martin, and P. Gebhard, 1953. *Sexual behavior in the human female*. Philadelphia: Saunders.
- A.C. Kinsey, W.B. Pomeroy, and C.E. Martin, 1948. *Sexual behavior in the human male*. Philadelphia: Saunders.
- T. Kochan, 2005. "15.668 People and organizations," *MIT OpenCourseWare*, at <http://ocw.mit.edu/OcwWeb/Sloan-School-of-Management/15-668Fall-2005/CourseHome/>, accessed 15 November 2007.
- E.O. Laumann, J.H. Gagnon, R.T. Michael, and S. Michaels, 1994. *The social organization of sexuality: Sexual practices in the United States*. Chicago: University of Chicago Press.
- J.W. Lynch, 2005. "The relationship of lesbian and gay identity development and involvement in lesbian, gay, bisexual, and transgender student organizations," at <https://drum.umd.edu/dspace/bitstream/1903/2667/1/umi-umd-2582.pdf>, accessed 11 December 2007.
- E.E. Maccoby, 1998. *The two sexes: Growing up apart, coming together*. Cambridge, Mass.: Belknap Press of Harvard University Press.

Massachusetts Institute of Technology (MIT), 2007a. "Number of students by course and year" (5 October), at <http://web.mit.edu/registrar/www/stats/yreportfinal.html>, accessed 8 December 2007.

Massachusetts Institute of Technology (MIT), 2007b. "Number of women students by course and year" (5 October), at <http://web.mit.edu/registrar/www/stats/womenfinal.html>, accessed 8 December 2007.

H.E. Murphy, 2007. "Suicide risk among gay, lesbian, and bisexual college youth," PhD dissertation, University of Washington, Seattle.

*NAACP v. Button*, 1963. 371 U.S. 415 (U.S. Supreme Court, 14 January); see [http://www.oyez.org/cases/1960-1969/1961/1961\\_5](http://www.oyez.org/cases/1960-1969/1961/1961_5), accessed 25 September 2009.

J.S. Okolica, J.L. Peterson, and R.F. Mills, 2008. "Using PLSI-U to detect insider threats by datamining e-mail," *International Journal of Security and Networks*, volume 3, number 2, pp. 114–121. <http://dx.doi.org/10.1504/IJSN.2008.017224>

B. Ortutay, 2009. "Facebook to end Beacon tracking tool in settlement," *Associated Press* (21 September), at <http://www.google.com/hostednews/ap/article/ALeqM5iHu1iUmbDqb2SuYZV2Zd3DSqYSbAD9AS1IO80>, accessed 30 September 2009.

S.R. Page, T.J. Johnsgard, U. Albert, and C.D. Allen, 1996. "User customization of a word processor," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver), pp. 340–346.

M.J. Piskorski, 2007. "I am not on the market, I am here with friends: Using online social networks to find a job or a spouse" (15 January), at <http://blog.hbs.edu/faculty/mpiskorski/>, accessed 25 September 2009.

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

R.L. Sell, J.A. Wells, and D. Wypij, 1995. "The prevalence of homosexual behavior and attraction in the United States, the United Kingdom and France: Results of national population-based samples," *Archives on Sexual Behavior*, volume 24, number 3, pp. 235–248. <http://dx.doi.org/10.1007/BF01541598>

J. Signorovitch, 2007. "HST.951/6.873 Materials" (17 October), at Stellar MIT Course Management System: 2007, accessed 8 December 2007.

*Smith v. Maryland*, 1979. 442 U.S. 735 (U.S. Supreme Court, 20 June); see <http://supreme.iustia.com/us/442/735/>, accessed 25 September 2009.

K. Tumulty, 2008. "Snooping into Obama's passport," *Time* (21 March), at <http://www.time.com/time/politics/article/0,8599,1724520,00.html?xid=rss-politics-cnn>, accessed 2 April 2008.

Valleywag, 2007. "Your privacy is an illusion: Bank intern busted by Facebook" (12 November), at <http://valleywag.com/tech/your-privacy-is-an-illusion/bank-intern-busted-by-facebook-321802.php>, accessed 5 December 2007.

---

## Editorial history

Paper received 26 July 2009; accepted 22 September 2009.

---

Copyright © 2009, *First Monday*.

Copyright © 2009, Carter Jernigan and Behram F.T. Mistree.

Gaydar: Facebook friendships expose sexual orientation  
by Carter Jernigan and Behram F.T. Mistree.  
*First Monday*, Volume 14, Number 10 - 5 October 2009  
<http://firstmonday.org/ojs/index.php/fm/article/view/2611/2302>



A Great Cities Initiative of the University of Illinois at Chicago [University Library](#).

© *First Monday*, 1995-2016.