A Familiar Face(book): Profile Elements as Signals in an Online Social Network

Cliff Lampe, Nicole Ellison, Charles Steinfield
Michigan State University
409 Comm Arts Building
East Lansing, MI 48824
lampecli@msu.edu, nellison@msu.edu, steinfie@msu.edu

ABSTRACT
Using data from a popular online social network site, this paper explores the relationship between profile structure (namely, which fields are completed) and number of friends, giving designers insight into the importance of the profile and how it works to encourage connections and articulated relationships between users. We describe a theoretical framework that draws on aspects of signaling theory, common ground theory, and transaction costs theory to generate an understanding of why certain profile fields may be more predictive of friendship articulation on the site. Using a dataset consisting of 30,773 Facebook profiles, we determine which profile elements are most likely to predict friendship links and discuss the theoretical and design implications of our findings.

Author Keywords
Social network sites, profile elements, signaling theory

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
User profiles are an integral part of social network sites and can include a vast array of user-contributed content. However, little is known about the specific effects user profiles have on interactions in online communities. Intuitively, we believe that profiles can help create a sense of presence and must garner positive outcomes for their users given the time commitment they require to complete and keep updated; yet we do not know what types of included information matter. This gap in understanding motivates our basic research question: how do elements in a profile influence the outcomes of using an online social network?

Online communities have different goals, but a common and important enterprise is forming connections between users. This is especially true for online communities that focus on articulating social networks, such as Facebook, MySpace, Friendster and Orkut, where the number of friends a user lists may act as a simple proxy for their connectedness in the network. Connections between users in an online community may be important for facilitating other tasks of the group [17, 20], reducing misbehavior [8, 18], and building types of social capital [11, 19], among other potential benefits [31].

To examine the role of profile elements in the formation of online connections, we focus on Facebook.com, an online social network site. Facebook, as with similar sites like MySpace and Friendster, allows users to create in-depth profiles describing themselves, and then to establish explicit links with other users, who are described as “friends” by the system. Facebook is a particularly appropriate site to study as it has profile creation and network articulation as primary community tasks, meaning that there is a consistency of action across the different users that allows for variance to be more clearly articulated and examined. Also, although Facebook is now open to those without academic affiliations, at the time data were collected, Facebook communities generally corresponded to existing offline network membership, typically related to academic environments like universities.

This offline connection has several implications. First, it allows the establishment of a natural boundary around the network that assists when determining who is a member and who is not. Second, the connection to an offline network might increase the likelihood of profile use by offline contacts, as the chances that a relationship formed in the online environment will extend to an offline meeting. This means that profile information has more opportunities to be verified than in other online communities. Third, participation may be reinforced by offline connections, contributing to a take-up rate for a given population that may be higher than normal. As mentioned above, Facebook is divided into networks based on affiliation with a particular offline institution. Consequently, users from University A are not, by default, considered part of the network of University B. Our study focused on one network within Facebook, namely the network defined by membership in Michigan State University (MSU). Membership in a university network is defined by Facebook as having a valid email address assigned by that university.

Facebook is a site that allows for users to create profiles and articulate their social networks. This is done through “friend” requests, wherein one user asks another to approve the connection. If the relationship is approved, the person shows up on the users’ friends list, and vice versa; friendship links in Facebook are mutual. Friend links are one way in which individuals traverse through the network, using links to travel from one profile to another [3].

Other research has shown that Facebook relationships tend to start offline and are then articulated online [15]. Respondents to surveys of Facebook memnbers have shown that they list offline friends most often, and only rarely list people they have met online [15]. While the ties to offline networks have previously meant that profiles are tied to unique individuals, that is changing as the site matures and users play with their capabilities. Some fake profiles, for the school mascot for example, are becoming more common.

In this paper, we report on an empirical study of the MSU Facebook community. Specifically, using data collected from all accessible MSU Facebook members via an automated script, we explore how profile elements relate to the number of friendship connections among users.

Literature Review

In this section, we explore prior work that establishes the basis for our primary proposition: That the amount and type of information included in user profiles should affect the number of articulated relationships in the online community.

Individuals form impressions of others in order to decide whether to pursue or continue a relationship [22]. In initial impression formation, individuals form impressions very quickly -- in as little as three minutes in face-to-face settings [23]. In order to achieve relational and other goals, individuals attempt to manage these impressions, strategically emphasizing some characteristics while de-emphasizing others [13]. These same self-presentational behaviors exist online. However, online self-presentation is more malleable and subject to self-censorship than face-to-face self-presentation due to the asynchronous nature of computer-mediated communication (CMC) and the fact that CMC emphasizes verbal and linguistic cues over less controllable nonverbal communication cues [26].

These same processes of impression formation and management take place in online settings, albeit slightly differently due to the affordances and constraints of CMC. In online environments, traditional identity cues, such as accent and style of dress, are not available. Early research assumed this forced online interactants to operate in a vacuum of identity cues, with attendant negative consequences for interpersonal relationship and community formation [6, 20]. However, subsequent work developed a more optimistic assessment (for review, see [30]), noting that CMC groups just needed more communication time than face-to-face groups, in order to compensate for CMC’s slower rate of exchange [27]. Walther’s Social Information Processing theory posits that online users compensate for the lack of traditional cues in online environments by looking towards other kinds of cues, such as spelling ability [24, 29, 28]. Evidence for SIP has been generated in online contexts such as MUDs [25] and online dating [11].

Online interactants seeking to form impressions of their communication partners must assess not only the content of the identity claims made by others, but also the veracity of these claims. As Donath [9] writes, “In order for a signal to have its intended effect, the receiver must both understand and believe it.” Although deception in offline environments is common [7], the ability to selectively self-present online [26] means that some kinds of misrepresentation (e.g., “gender-bending”) are more easily accomplished via CMC. In some online environments such as online dating, misrepresentation is a significant concern [11]. Users in online environments rely on a variety of cues to make determinations about one another; however, all these cues are not deemed equally credible. For instance, Goffman [13] notes that identity cues can be intentionally given or unintentionally given off, and that we are more likely to privilege those cues that are perceived to be unintentional as opposed to strategically constructed. This ability to engage in deceptive self-presentation online is compounded when interactants do not share a social network and therefore have less access to “information triangles” such as mutual friends who might confirm or deny information [2].

Theoretical motivations

We draw upon three theories to help explain how profile construction might affect participation in online communities. Signaling theory addresses the *type* of information that can be placed in profiles, suggesting that profile elements act as signals that may prove something about the identity of the user. These signals can be manipulated by senders to communicate personal qualities, or interpreted by receivers to make judgments about the characteristics of other users. We use common ground theory to explain the motivation of filling out profiles, which is to establish common frames of reference that enhance mutual understanding. Transaction cost theory bridges the two former theories and suggests that certain profile elements may facilitate the production of shared referents, which usually involves costly negotiations between participants, and makes it easier for interactants to engage in other forms of communication (such as email).

Signaling and the verifiability of profile entries

Signaling theory addresses a basic question: what keeps signals reliable in contexts where deception can be beneficial [9]? Donath argues that a signaling system must evolve so that it is beneficial for participants to produce reliable signals, but costly to produce deceptive ones. Building on contemporary signaling theory, she distinguishes between many different kinds of signals, including those that reliably indicate possession of some quality simply through observation of the signal, which are termed assessment signals, and those that only indicate a quality through social convention, which are termed conventional signals. As she notes, lifting a heavy weight is an assessment signal that reliably indicates that a person is strong. Wearing a Golds® Gym T-shirt is a conventional signal that suggests that the person works out and therefore is likely to be in shape, but is easy enough to acquire that the wearer might actually be weak. However, as easily seen from these examples, conventional signals are cheaper to produce and more prone to deception. Donath points out that online contexts generally only support conventional signals – people construct their profiles with words or images that are easily manipulated – and therefore the question of how and why signals in online profiles are reliable is relevant.

Signaling theory provides clues as to why Facebook profiles might be more "honest" than profiles found in other online fora. Donath and boyd [10] argue that a shared social network can provide explicit or implicit verification of identity claims. Therefore, the structure of Facebook should encourage more truthful profiles, or misrepresentations that are playful or ironic as opposed to being intentionally deceitful. There are different types of elements that can be signaled in a Facebook profile - some more verifiable, and hence more easily tied to the ability to police their honesty, than others. Such elements might include those that place a member in a particular physical locale (e.g. their high school and home town, their location on campus), a particular community or group on the campus (e.g. their major and their classes) or other verifiable personal and physical attributes (e.g. their relationship status and gender). Assuming that people will be more likely to articulate a relationship with (or, in Facebook lexicon, to “friend”) those who include these verifiable identity cues (such as high school) in their profile, we thus expect that the more these elements are included, the more “successful” one will be in the Facebook context, using size of friendship network as a proxy for success. Hence, we propose that inclusion of these elements will have an impact on number of friends, and their presence may enhance the signaling value of other aspects of the profile that are less verifiable (e.g. interests).

Establishing common ground in online communities

Common ground theory tells us why the inclusion of more information in the profile in general should lead to greater numbers of friends. Including location information (hometown, etc.) establishes common ground, and interests reveal personality aspects that can help people make decisions about declaring friendship links.

Clark [4] describes community membership as an important characteristic to consider when trying to assess the amount of shared understanding that already exists between participants. Membership in different communities will involve generic and particular knowledge that can be used as common frames of reference to build common ground.

This type of common ground is important in interactions because it facilitates understanding and fosters cooperation between participants, especially in cases where group members are not well-known to each other or are interacting through information and communications technology [5, 16].

Profile elements provide clues about fellow participants in an online community that may act like the subtle interviewing done in face-to-face environments to find shared communities of membership or common ground [10]. Participants of Facebook may see someone is in their concentration of study, which indicates that the person will understand a range of references that a person in another major may not. If one sees another profile listing a shared hometown, they might expect the person to understand references to landmarks and events in that area.

Reducing the cost of connection

The signals that can be included in user profiles help reduce the cost of finding the common referents that lead to increased understanding between participants. In the economic literature, transaction costs are costs incurred in the process of economic exchange, and can explain why it is that certain types of markets fail [32]. They include costs of search (e.g. for products, trading partners, price discovery, etc.), negotiation (e.g. legal and contracting costs), settlement (e.g. fees for currency exchange and costs to transport products), and monitoring (e.g. product inspection and other means to verify that terms are being met). If transaction costs are too high, exchange may not occur. In this case theorists suggest that firms internalize production, substituting generalized labor contracts – often termed hierarchy – for specific market contracts.

Electronic transactions over the Internet – or e-commerce – are often considered to have reduced transaction costs, enabling markets to function where hierarchy might have prevailed otherwise [1]. Key to this is the effect on search costs. The Internet makes it easy to locate alternative suppliers of a desired good or service, especially when few vendors are present in a local market.

Just as in economic exchanges, search costs can easily influence the likelihood that a connection will be made in an online social network. Profiles make it easier to find others who have matching interests or other desired attributes. Indeed, matchmaking sites that help users find romantic partners can easily be analyzed from a transaction costs theory perspective, using such concepts from the theory as search costs, opportunism, and asymmetric information.

In the case of an online social network site like Facebook, the ability to search profiles reduces the costs associated with finding former high school classmates, people in the same concentration or major, or people in the same dormitory, even without a potential partner's name. The more profile elements there are, the more refined, and hence accurate, a search can be. Thus, profiles reduce search costs, and therefore reduce the costs of making connections. Hence, from a simple transaction costs theory point of view, we should expect that the more profile elements a person has, the more likely others will find him or her, and thus make connections.

Transaction cost theory might also explain why some types of profile elements may be more important than others, mainly when the connection is being made with a new acquaintance. In this case, elements that are more like the "assessment" signals noted above (which are potentially verifiable), might be considered as helping protect against opportunistic behavior. Profile elements that are less reliable are more difficult to monitor. Hence, a transaction cost theory perspective would suggest that these verifiable elements would have more of an impact on the likelihood of making a connection than other elements. Hence, we would expect that both verifiable elements and contact elements would have a stronger effect on the number of friends than profile elements that only provide preference information.

Research Questions

This review highlights a number of ways in which profile entries may be associated with online social network users' articulated friendship networks. We next present the methods used to enable us to explore several research questions suggested by this review, including:

1. What are the relationships between the various types of profile entries and the number of friends a user has on their social network site?
2. Are some types of profile entries more strongly associated with the number of friends listed?

METHODS

For an initial analysis of the use of profiles in online communities, we used automated scripts to measure the use of fields in Facebook profiles and associate those with the number of links between participants. This section will describe that process, and discuss how we code profile elements into different types of signals.

Data collection

As mentioned above, Facebook is divided into networks defined by membership in offline institutions, and this study focuses on the network defined by membership in Michigan State University. It is not well known how networks within Facebook may differ from each other, but the MSU Facebook network seems broadly comparable to descriptions of Facebook networks at other academic institutions [14, 21].

Data were collected from the MSU Facebook site using automated scripts that downloaded information in profile fields and saved that information in offsite databases. The data reported here were collected between April 1 and April 13, 2006, which was the amount of time necessary to collect information from every user on the site. Facebook users have the ability to restrict viewing of their profile to their friends, in which case the script would not be able to read the profile either. This occurred in 7,634 cases, or 19% of the profiles, leaving us with 30,773 usable profiles. This is a problem for analysis in that these users are consequently excluded, and may have characteristics that make them different from the population of users who do not elevate their privacy settings. There are several potential issues with this data collection method besides the loss of individuals with high privacy settings. For example, this data collection method is fragile to changes made by the Facebook design team. Over its brief history, Facebook has added and removed fields, changed data entry points and altered interface options. This means that not every user profile on Facebook is going to have the same data. Similarly, many of the profile fields on Facebook are optional which means that for some users fields will be empty.

Independent Variables

In this analysis, the independent variables being studied are the various fields of the Facebook profile. We measured whether available fields had been populated by each user, and in some cases how much information had been added to the field. We divide the profile elements into four categories: control variables, referents (e.g. location) variables, preference variables, and contact variables. Tables at the end of each section summarize these variables and describe their categories.

Control variables

In our analyses we tracked two variables that have been shown in social network literature to affect network characteristics: sex of the member and length of membership in the network. Of the MSU Facebook users, 93.8% listed their sex with 53% listed as females. The age of account is a field that was automatically displayed at the time of data collection. To these characteristics, we added a third descriptive variable likely to have an effect on participation, namely institutional status. As Facebook is largely oriented to academic populations, one would expect different participation from sub-populations within those institutions. To wit, being an undergraduate is likely to create a different Facebook experience than being faculty. Finally, we added to the set of control variables the "Last Updated" field, which is an automatically displayed profile element indicating when the user last made any change to

Field	Description	%
Sex	Gender of the user.	93.8%
Status	Type of institutional member	100.0%
Member Since	Auto field listing account creation date	100.0%
Last Updated	Auto field listing last profile update time	100.0%

Table 1: Proportion of XSU Facebook community by institutional status.

their profile. Since we did not have measures of user activity like page views or number of times the profile is updated, the last updated field acts as a very rough measure of account activity. Accounts that have not been updated in months signal something different than accounts updated within the past week. Table 1 displays the profile fields included in these measures and the percentage of users populating these fields. At the time of data collection, these fields were typically auto-filled, so have high participation rates.

Referents Index

The first index of variables was based on fields related to common points of reference among users. Listing shared past referents allows the profile creator and user to find common ground, and share narratives. Users who list the same high school, even if they didn't attend it together, could make references to shared traditions, teachers or physical locations. Table 2 lists the profile fields included in this referents index, and the percentage of users who populate those fields.

Field	Description	%
Hometown	Town of residence before joining MSU	83.3%
High School	School attended before college	87.1%
Residence	On-campus housing information	45.1%
Concentration	Major field of study	89.5%

Table 2: Fields included in the referents index, and percentage of users populating that field.

Interests Index

A second index was created based on the use of profile fields that express personal preferences and self-descriptive information. Favorite movies and music, activities and an open-ended "About Me" field are interpreted as conventional signals, by which the user is crafting an image of themselves for other users. These signals are easy to produce, though harder to verify. Still, these profile elements may be used to judge similarity between the profile consumer and the profile producer. For example, shared movie favorites might indicate similarity between participants, with anticipation of agreement along other axes. Table 3 displays the profile fields included in this index, and the percentage of users who populate that field.

Field	Description	%
About Me	Open-ended field	59.8%
Interests	Items separated by commas become linked.	77.7%
Favorite Music	Same as above	78.2%
Favorite Movies	Same as above	80.1%
Favorite TV Shows	Same as above	46.5%
Favorite Books	Same as above	66.9%
Favorite Quotes	Content not linked.	73.8%
Political Views	Drop down list	60.9%

Table 3: Fields included in the interests index, and percentage of users populating that field.

Contact index

A third index contained profile fields whose presence seemed to indicate a willingness to share off-site connections with others. Users populated a variety of fields

Field	Description	%
Relationship Status	Drop down list of relationship states	78.5%
Looking for	Check button list of types of relationships sought	50.8%
Website	Allows for multiple web addresses to be input	29.1%
Address	Current address – non campus	13.5%
Birthday	Drop downs to select day/month/year	83.8%
AIM	AOL Instant Messenger screen name	67.8%
Email	Default is the user's school email, which is also used to join the site	92.3%

Table 4: Fields included in the contact index, and percentage of users populating that field.

that would allow the user to be contacted via other media, ranging from email to specific offline location information. Birthday is included, since birthday notices are broadcast to a user's friend list. Table 4 shows the profile fields included in the contact index, and the percentage of users who populate these fields.

Dependent Variables

Our primary dependent measure is the total number of "friends" a user has on Facebook. Facebook allows for two types of "friendship" links: with users at the same institution or with users at other institutions (now called "networks"). Our script tallied this information using the number of friends listed in each user's profile. Table 5 reports on the mean and median number of both friend links, as well as the standard deviation of the mean. In addition to friendship links, we report the average ratio of friend links from the same school over all friendship links. This gives a rough measure of the proportion of the user network that exists at the same institution.

	Mean	Std. Dev	Median
Same School	95	86	75
Other School	84	78	68
Ratio	0.53	0.18	0.53

Table 5: Friendship links for MSU Facebook members

Analysis approach

We explored the research questions by illustrating how the median number of friends differs based on different values in individual profile fields. We further used a multiple regression approach to see which clusters of profile elements were most strongly associated with the total number of friends. The main independent variables were the indices representing the theoretically derived groups of profile fields as described earlier – assessment, conventional, and contact. We used the log of the number of friends as a dependent measure in the regression analysis, because the friend distribution was highly skewed with a long tail created by smaller numbers of users with extremely large numbers of friends. We further standardized all of the non-nominal variables in order to deal with unequal scale ranges. We report the statistical significance of the results, even though our sample more closely resembles a census of the MSU Facebook community, and with this sample size nearly all differences would be "significant." However, we can conceptualize this group of users as a sample of the entire Facebook community, and so use the relative differences in estimate strengths to guide interpretation.

RESULTS

Profile use

Facebook users participate widely in the fields that allow them to present themselves to other users. On average, users complete 59% of the fields available to them, and in some fields display a significant amount of information. Tables 1-4 show whether individual profile fields are used, but does not show anything about how much information is added when the fields are open-ended. The categories "About Me," "Interests," and "Favorites" are the most open fields available to Facebook users, with users able to articulate many preferences that shape the public persona they are trying to present to others. Table 6 shows the average and median number of items in each of these fields, except in the "About Me" field for which the average and median number of characters is reported. We report the median value as it is the measure of central tendency more resistant to the presence of strong outliers, which is the case for this dataset.

Profile field	Average items	Median items
Interests	6.5	5
Favorite Book	2.5	1
Favorite Movies	6.5	5
Favorite Music	8.8	3
Favorite TV Show	2.1	0
About Me*	157	36

Table 6: Median items in open-ended profile fields
*** number of characters in the field**

"Friends" links

As with other online social networks, Facebook has labeled declared relationships between members as "friends" links. MSU Facebook members are actively participating in the creation of friendship links. The number of friends at MSU vs. those at different institutions is relatively stable across users, with a strong correlation between the number of different types of friends (Pearson r=.694). In the analyses below, we label friends also in the MSU Facebook network as "same institution" and those in different networks as "other institution."

User characteristics are related to number of friends

As one might expect, the median number friends a user has will vary on demographic dimensions. Status in the institution, as shown in Table 7, is a particularly important factor associated with median number of friends both at the same institution and at different institutions.

Undergraduate members have more friends than any other group in the network. Females have slightly more friends than do males. The older the account, the more friends of each type a user will have as well, with a Pearson correlation coefficient of r=0.37 between the age of the account and friends at the same institution, and a Pearson correlation coefficient of r=0.24 between the age of the account and friends at other institutions.

	Same institution	Other institutions
Status		
Undergrad	87	83
Alumnus/Alumna	45	33
Grad Student	27	19
Faculty	29	8
Staff	14	7
Sex		
Female	80	76
Male	73	63

Table 7: User characteristics associated with differences in number of friendship links

Using profile fields is related to number of friends

The act of populating profile fields is strongly associated with the number of friendship links a user will have. In Table 2 we showed which profile fields were available to users and what percentage of those users populate that field with information. In each case where that field is populated, users show a higher median number of friends. T-tests of those differences show significance on every dimension of $p<.001$, though the size of the dataset limits the usefulness of significance testing. Of more importance are the practical differences when information is present vs. when it is not.

Profile Field	Same institution		Other institutions	
	yes	*no*	*Yes*	*no*
High School	92	35	89	21
Favorite Music	83	37	75	33
AIM	100	50	93	52
Birthday	80	26	73	21
About Me	88	56	78	52

Table 8: Five profile elements associated with the largest difference in number of friends.

Table 8 reports the fields for which the five largest differences in number of friends at the same institution is evident between those who entered information and those who did not. One explanation for the importance of the High School, AIM and Birthday fields is that they help support the maintenance of pre-existing social networks that are being reified on Facebook. For example, a user might keep track of high school friends using the site. Favorite Music and About Me are slightly different in that they act as cues about the identity of the Facebook member, so may just as likely be targeted to users not in a pre-existing social network with the user.

The amount of information in profiles is weakly associated with number of friends

It's not just whether the information is present that is related to the number of friendship links, but how much information is presented. Table 9 reports the Pearson correlation coefficients between the number of items in open-ended profile fields and the number of friendship links held by the user. "Favorites" is a sum of all the entries in the five "favorites" fields.

Profile field(s)	Same institution	Other Institutions
About Me	r=.13	r=.13
Interests	r=.21	r=.19
Favorites	r=.14	r=.15

Table 9: Pearson correlations between number of items in profile and number of friends

While not exceptionally strong correlations, these relationships between amount of self-descriptive content and number of friends have positive directionality. The direction of influence cannot be determined, but there is an association between how many items a person lists in their open-ended profile fields and the number of friendship links they have. There are many possible explanations for this, including that people with many friends have increased social pressure to add information to their profiles, or that active users both add information to profile fields and seek out people to list as "friends."

Multivariate analyses: factors predicting number of friends

The previous univariate analyses indicate that individual profile elements are associated with differences in numbers of friends among Facebook users. In order to see which sets of profile entries have the strongest association, this section describes a regression analysis of profile element use and how much their use affects numbers of friends.

Table 10 shows the results of an ordinary least squares regression (OLS) analyzing the effect these index variables, as well as select control variables, have on total number of friends listed by the user. Note that the adjusted R-squared value of 0.401 indicates this model explains around 40% of the variance in number of friends.

The indices were entered into the model in a stepwise fashion, with each index increasing the explanatory power of the model. The dependent measure for this model is the log of the number of friends, used to reduce the effects of a skewed distribution. Variables that included the amount of information in About Me, Interests and Favorites were also added, but did not have an effect on the model, and so are excluded in Table 10. This is significant in that it doesn't appear to matter how much information is included in profile fields, just that some information is included.

Term	Estimate	t Ratio	Prob>|t|
Intercept	-0.319	-14.71	<.001
Female*	0.042	10.02	<.001
Alumni	-0.045	-1.96	0.050
Faculty	0.027	0.40	0.688
Graduate	-0.201	-7.82	<.001
Staff	-0.230	-4.11	<.001
Undergraduate	0.449	20.28	<.001
Member since	-0.342	-78.62	0.000
Recency of last update	0.174	40.20	0.000
Referents Index	0.193	35.32	<.001
Interests Index	0.079	16.77	<.001
Contact Index	0.103	22.98	<.001
Adj. R-squared = .401 F=2061.10, P<.0001 N=30,773			

Table 10: Regression results predicting number of friends from provided profile information
* **dummy coded, and so compares females to males**

The results demonstrate that after controlling for gender, status at MSU, length of time as a Facebook member, and our proxy measure of user activity, profile field use has a slight positive association with the number of friends a user has on Facebook. As we would expect, undergraduates have more friends on the system than others, as do those who have been members for a longer period of time. The recency with which users had updated their profiles contributed to the explanatory power of the regression.

Regarding the three indices, the referents index had the largest coefficient, followed by the contact index and then the interests index. The fields in the referents index are harder to falsify than information in the interests field. In this way, the referents may be acting more similarly to assessment signals as described by Donath [9]. Referent profile fields may also be acting as mechanisms to ease search costs, or reduce transaction costs, in enabling latent ties from offline networks into the Facebook social network. That is to say, a user may not be searching for people who like Citizen Kane to add to their friends list, but they may be seeking out users who went to their high school.

DISCUSSION

We found that populating profile fields on Facebook is positively related to the number of friends a user will have listed. The amount of information posted in open-ended fields does not affect the number of friends when added to the indices of the presence of information in the profile fields.

Even after controlling for gender, time on the system, user status in the community, and the recency of updating, significant variation in the number of friends is associated with the supply of other types of profile information. Information that helps share common referents – same high school or home town, same major, same classes – should be important. From a common ground point of view, it provides an immediate establishment of common referents that can foster interaction. From a transaction cost point of view, it reduces the costs of search for potentially relevant contacts. Ellison *et al* [10] suggest that Facebook may be helping to make connections out of "latent" ties by supplying this information. This co-referent information seems to be different than Facebook activity in general. If it were simply the amount of information that was the determining factor, we would expect to see all of the profile elements to be related to number of friends roughly equally. This doesn't happen, which speaks to the importance of co-referent profile fields.

Preference information also contributed to the explained variance in number of friends, but showed the weakest association of our three indices. These conventional signals may be used more in a playful and ironic sense, and may not necessarily be relied upon as strongly to gauge similarity for finding new connections.

Theoretical Implications

This paper draws upon three theoretical threads to create a framework that elucidates the relationship between self-presentation, as measured by fields in a user profile, and relationship articulation as measured by "friend lists," using data gathered from a popular online social network site, Facebook. To date, this relationship has not been explored in a way that draws upon behavioral data and theory. Signaling theory can help explain why Facebook differs from other online contexts, due to the shared social network that can help warrant identity claims [9]. Common ground theory is used to explain how users can find areas of shared connection and share narratives. Finally, transaction cost theory sheds light on how these sites may make it easier for individuals to find others they want to connect with, as it makes visible commonalities and allows participants almost effortless ways to locate and communicate with others.

Our synthesis of signaling theory, common ground theory, and transaction cost theory suggested that in an online social network that is constructed around an existing offline community, certain types of profile information might be particularly important. One theoretical contribution of this approach is that it highlights the underlying shared thrust of each of these theories – namely, that they are concerned with how to reduce the costs associated with locating and evaluating communication partners. Common ground theory speaks to making interactions more efficient through assumptions about shared referents. Transaction costs approaches examine the ways in which interactions, such as finding “friends,” can be made more efficient through these systems.

Future Work
This paper is an initial step in a continuum of work related to how profiles influence interaction in online spaces. We have shown here a non-causal link between profile use and friendship links, but several questions remain. In other research, we have seen the importance of offline connections to Facebook friends [15] and how Facebook participation affects the genesis of social capital among its users[12].

Work is in progress now to explicate these results with information gathered through surveys and interviews. Cognitive walkthrough style interviews with Facebook users will be used to determine how users decide whom to add to their friends list, and which elements of a profile they find to be the best signals in assessing other users, both for friend requests and in general.

Additionally, we plan to track Facebook participation over time, with periodic surveys of a panel of MSU Facebook users conducted to determine how uses are changing.

Design Implications
Online communities depend on many-to-many interactions between participants with heterogeneous goals, backgrounds and characteristics. The lack of traditional social cues available in online interactions can make it challenging to develop a sense of fellow participants in online communities, although users do become adept at reading online cues and interpreting them. These findings indicate several possible design implications for fostering these interactions.

Given that common referents was a strong indicator of number of friends, additional search features could be added to online communities to create these connections. Allowing people to search by shared city, institution or job type may create a sense of connection that facilitates interactions. Another implication involves the choice of fields included in user profiles. While common referents were important in this case, other communities may find that personal preference fields matter more than in the Facebook case. The design guideline derived from these data is to include fields that can be used to highlight similarities between users.

Limitations
This study looks at one particular online community, Facebook, and sub-section of that site devoted to members of Michigan State University. In addition, the study looks at profile use at a particular moment in time, and since data collection Facebook has changed many of the fields used in profiles. While generalizing these results to all online communities should be discouraged, we do feel that the basic insight that profile elements can act as signals to facilitate social browsing stands as a general contribution.

Facebook is tied to offline interactions in a way few other online communities are. It could be that offline friends create social pressure to use certain profile fields, which would also be translated as online social ties. This means that the causal direction could well be in either direction – something our data is not able to sort out.

The data we collected represent information users contribute to the Facebook site, which has the benefit of being behavioral but the deficit of not indicating attitudes or motivations. This work cannot address either how users perceive profile elements, or how they feel their profile elements will be received by others. Additionally, many of the profile elements included in the regression model are recorded in the binary through our data collection. While univariate analysis seems to indicate the amount of information in profile fields isn’t important, the content of the fields may very well be. Both of these limitations need to be redressed in future research, particularly through interviews with users.

CONCLUSION
User profiles in online communities can play a role in the functioning of the site. This paper describes how users of Facebook populate fields, and that different profile elements have different consequences for the number of friends listed on the site. Profile fields that help users share common referents are more highly associated with numbers of friends than fields that express personal likes and dislikes.

Online communities may be able to use profiles strategically to foster their goals. Profile development, such as happens on Facebook, allows users to present themselves to their fellow community members, and get a sense of those with whom they are interacting. Further study in user profiles can help us understand both how we make sense of one another online, but also how to improve our impression formation and management processes in order to facilitate more productive interactions.

ACKNOWLEDGMENTS
We would like to thank Facebook for permission to use automatic scripts on the MSU site. We also thank Nathan Oostendorp for help in creating the scripts used to collect data.

REFERENCES

1. Bakos, J. (1999) Reducing Buyer Search Costs: Implications for Electronic Marketplaces. *Management Science, 43* (12). 1676-1692.
2. Berger, C.R. Planning and scheming: Strategies for initiating relationships. in Burnett, R., McGhee, P. and Clark, D. eds. *Accounting for relationships: Explanation, representaion and knowledge*, Methuen, New York, 1987, 158-174.

3. boyd, d. (2006) Friends, Friendsters, and Top 8: Writing community into being on social network sites. *First Monday*, *11* (12).
4. Clark, H.H. *Arenas of Language Use*. University of Chicago Press, Chicago, IL, 1992.
5. Clark, H.H. and Brennan, S.E. Grounding in communication. in Resnick, L.B., Levine, J.M. and Teasley, S.D. eds. *Perspectives on socially shared cognition*, APA Press, Washington, 1991.
6. Culnan, M.J. and Markus, M.L. Information technologies. in Jablin, F.M., Putnam, L.L., Roberts, K.H. and Porter, L.W. eds. *Handbook of organizational communication: An interdisciplinary perspective*, Sage Publications, Thousand Oaks, 1987, 420-443.
7. DePaulo, B., Kashy, D., Kirkendol, S., Wyer, M. and Epstein, J. (1996) Lying in everyday life. *Journal of Personality and Social Psychology*, *70* (5). 979-995.
8. Donath, J.S. Identity and Deception in the Virtual Community. in Kollock, P. and Smith, M. eds. *Communities in Cyberspace*, Routledge, London, 1998.
9. Donath, J.S. *Signals, Truth and Design.* MIT Press, Cambridge, MA, forthcoming.
10. Donath, J.S. and boyd, d. (2004) Public displays of connection. *BT Technology Journal*, *22* (4). 71.
11. Ellison, N., Heino, R. and Gibbs, J. (2006) Managing Impressions Online: Self-Presentation Processes in the Online Dating Environment. *Journal of Computer-Mediated Communication*, *11* (2).
12. Ellison, N., Lampe, C. and Steinfield, C., Spatially Bounded Online Social Networks and Social Capital: The Role of Facebook. in *International Communication Association*, (Dresden, 2006).
13. Goffman, E. *The Presentation of Self in Everyday Life*. Doubleday, New York, NY, 1959.
14. Gross, R. and Acquisti, A., Information Revelation and Privacy in Online Social Networks. in *Workshop on Privacy in the Electronic Society*, (Alexandria, VA, 2005), ACM Press.
15. Lampe, C., Ellison, N. and Steinfield, C., A Face(book) in the Crowd: Social Searching vs. Social Browsing. in *ACM Special Interest Group on Computer-Supported Cooperative Work*, (Banff, Canada, 2006), ACM Press.
16. Olson, G.M. and Olson, J.S. Distance matters. in Carroll, J. ed. *HCI in the New Millennium*, Addison-Weasley, New York, 2001.
17. Preece, J. and Maloney-Krichmar. Online Communities. in Jacko, J. and Sears, A. eds. *Handbook of Human-Computer Interaction*, Lawrence Erlbaum Associates Inc., Mahwah, NJ, 2003, 596-620.
18. Reid, E. Hierarchy and power: Social control in cyberspace. in Smith, M.A. and Kollock, P. eds. *Communities in Cyberspace*, Routledge, London, 1999, 107-133.
19. Resnick, P. Beyond Bowling Together: SocioTechnical Capital. in Carroll, J. ed. *HCI in the New Millenium*, Addison-Wesley, 2001.
20. Sproull, L. and Kiesler, S. *Connections: New ways of working in the networked organization*. MIT Press, Cambridge, MA, 1991.
21. Stutzman, F., An Evaluation of Identity-Sharing Behavior in Social Network Communities. in *iDMAa and IMS Code Conference*, (Oxford, OH, 2005).
22. Sunnafrank, M. (1986) Predicted outcome value during initial interactions: A reformulation of uncertainty reduction theory. *Human Communication Research*, *13*. 3-33.
23. Sunnafrank, M. and Ramirez, A. (2004) At first sight: Persistant relational effects of get-acquainted conversations. *Journal of Social and Personal Relationships*, *21* (3). 361-379.
24. Tidwell, L.C. and Walther, J.B. (2002) Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. *Human Communication Research*, *28* (317-348).
25. Utz, S. (2000) Social Information Processing in MUDs:The Development of Friendships in Virtual Worlds. *Journal of Online Behavior*, *1* (1).
26. Walther, J.B. (1996) Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interactions. *Communication Research*, *23*. 1-43.
27. Walther, J.B. (1993) Impression development in computer-mediated interaction. *Western Journal of Communication*, *57*. 381-398.
28. Walther, J.B. (1992) Interpersonal effects in computer-mediated communication: A relational perspective. *Communication Research*, *19*. 52-90.
29. Walther, J.B., Anderson, J.F. and Park, D.W. (1994) Interpersonal Effects in Computer-Mediated Interaction. *Communication Research*, *21* (4). 460-487.
30. Walther, J.B. and Parks, M.R. Cues filtered out, cues filtered in: Computer-mediated communication and relationships. in Knapp, M.L. and Daly, J.A. eds. *Handbook of interpersonal communication*, Sage Publications, Thousand Oaks, CA, 2002, 529-563.
31. Wellman, B. (2001) Computer Networks as Social Networks. *Science*, *293* (14). 2031-2034.
32. Williamson, O.E. *The Economic Institutions of Capitalism.* The Free Press, New York, 1985.