Sociapedia: Online Collaboration Over a Social Network
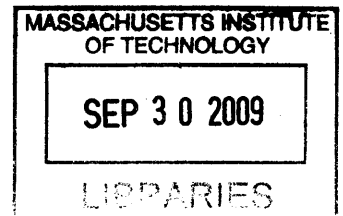
by

Mihir Kedia

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2009

Author ....
Department of Electrical Engineering and Computer Science
August 7, 2009

Certified by ....
David R. Karger
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by ....
Terry P. Orlando
Chairman, Department Committee on Graduate Students

Sociapedia: Online Collaboration
Over a Social Network

by

Mihir Kedia

Abstract

This paper introduces Sociapedia, a collaborative social network application that encourages friends to contribute content about each other. Sociapedia is designed upon the same principles that allowed ordinary web users to construct the comprehensive encyclopedia Wikipedia, and is built on top of the largest social network currently in existence. Several lessons from the evolution of Sociapedia are discussed.

In order to learn how people would use a collaborative social network like Sociapedia, a field study was conducted over the period of one week. We established that Sociapedia generates accurate information that is not otherwise present, and characterize the nature and structure of this information. We also compare methods for spreading applications in Facebook. Finally, we analyze Sociapedia's growth and offer several suggestions for increasing its reach throughout Facebook.

Thesis Supervisor: David R. Karger

Title: Professor of Electrical Engineering and Computer Science

Contents

1

Introduction

The Sociapedia project was started in order to create a *people encyclopedia*: a repository of information about regular people. While we can easily go to the library and look up information about famous figures throughout history, it is much harder to find out more about the people we interact with every day.

Online websites exist that are devoted to this goal. These websites are called *social networks*, and they create a platform over which hundreds of millions of people can create profiles and interact with each other. Users can create profiles on these sites to tell others more about themselves: their hobbies, favorite TV shows, and more. This is interesting information, but is not necessarily complete. We would like to find a way to collect information beyond that which users write about themselves.

Our approach is to get this information from friends and colleagues of the user. We want to create a site where any of your friends can come and help write articles about you. There are already some existing projects that try to collect information about users through their friends, but we believe that our encyclopedic approach can collect additional information that these methods do not.

This approach is similar in style to the one that built the Wikipedia website, and our first step was to take a closer look at this enormous encyclopedia. Wikipedia is an encyclopedia that contains over three million comprehensive articles on a huge variety of topics (Johnson). What makes Wikipedia extraordinary, however, is its atypical authorship. While most encyclopedias are the result of painstaking years of labor by a select team of experts, Wikipedia has always allowed anyone to write content and edit articles. Wikipedia has 10 million registered contributors, and over 300 million individual contributions or edits have occurred since its creation in 2001 (Wikipedia).

How can we harness the same forces that built Wikipedia to create our people encyclopedia? One thought would be to have Wikipedia include articles about regular people, but this would not be suitable. Wikipedia has over 300 million unique visitors per month (Morgan), so even a tiny fraction of that user base is capable of generating enormous amounts of content. What we need, however, is for a much larger proportion of people to write about their friends and colleagues. In essence, Wikipedia lacks the social structure needed to allow users to discover and contribute information about their friends.

The social networking sites from before, though, provide exactly the social structure we need. If we could augment a social network with a component that allowed users to write content about their friends, we could use it to collect the additional information that we believe existing locations are missing.

We built Sociapedia as a custom application that runs on top of Facebook: one of the largest social networks currently in existence. Sociapedia is based directly on Wikipedia, allowing users to create and edit articles about their friends. Designing Sociapedia was not a trivial task, and we devote a portion of this thesis to sharing the thinking and rationale behind the features and interaction model we developed.

There were several key areas we wanted to understand, and we ran a one-week field study in search of answers. Our primary aim was to find whether or not this model could work: can we get useful information about ordinary people from others in their social circle? The existence of this thesis is proof of our success, and we are able to show that Sociapedia generates accurate social information that is not found elsewhere. We also categorize the type of information present on Sociapedia, showing exactly what types of information users are inclined to share.

Another aim we had was to understand the potential of Sociapedia. Could it grow to encompass all of Facebook? We first studied recruitment techniques, showing which of them were the most effective. We also looked as the users who chose to contribute information: who did they write about and why were some of them more prolific than others? Finally, we tackled the growth question directly. While our conclusion is unfortunately negative, we share several ideas for modifying Sociapedia to be able to expand across its entire social network.

2

Background

A reasonable amount of domain knowledge is necessary in order to fully understand Sociapedia. This section gives an overview of two essential concepts: wikis and social networks. More specific knowledge is introduced as needed later.

Wikis

Many readers are probably familiar with the online site Wikipedia. This website is a comprehensive encyclopedia with an accuracy that approaches that of the famous Encyclopedia Britannica (Giles). What is fascinating about Wikipedia, however, is that its articles were not written by a selected team of experts; rather, any visitor to the site is allowed to make edits and contribute to the articles.

You can see an example, a Wikipedia article about Madagascar, in Figure 1. Note how a natural organization has built up around the information, with information cleanly separated into Culture, Economy, Politics, and more. Although not contained in the screenshot, an ontology has also been built around the articles, placing them in various categories (Madagascar is part of *East Africa, Island Countries*, and *States and Territories Established in 1960*, among others.)
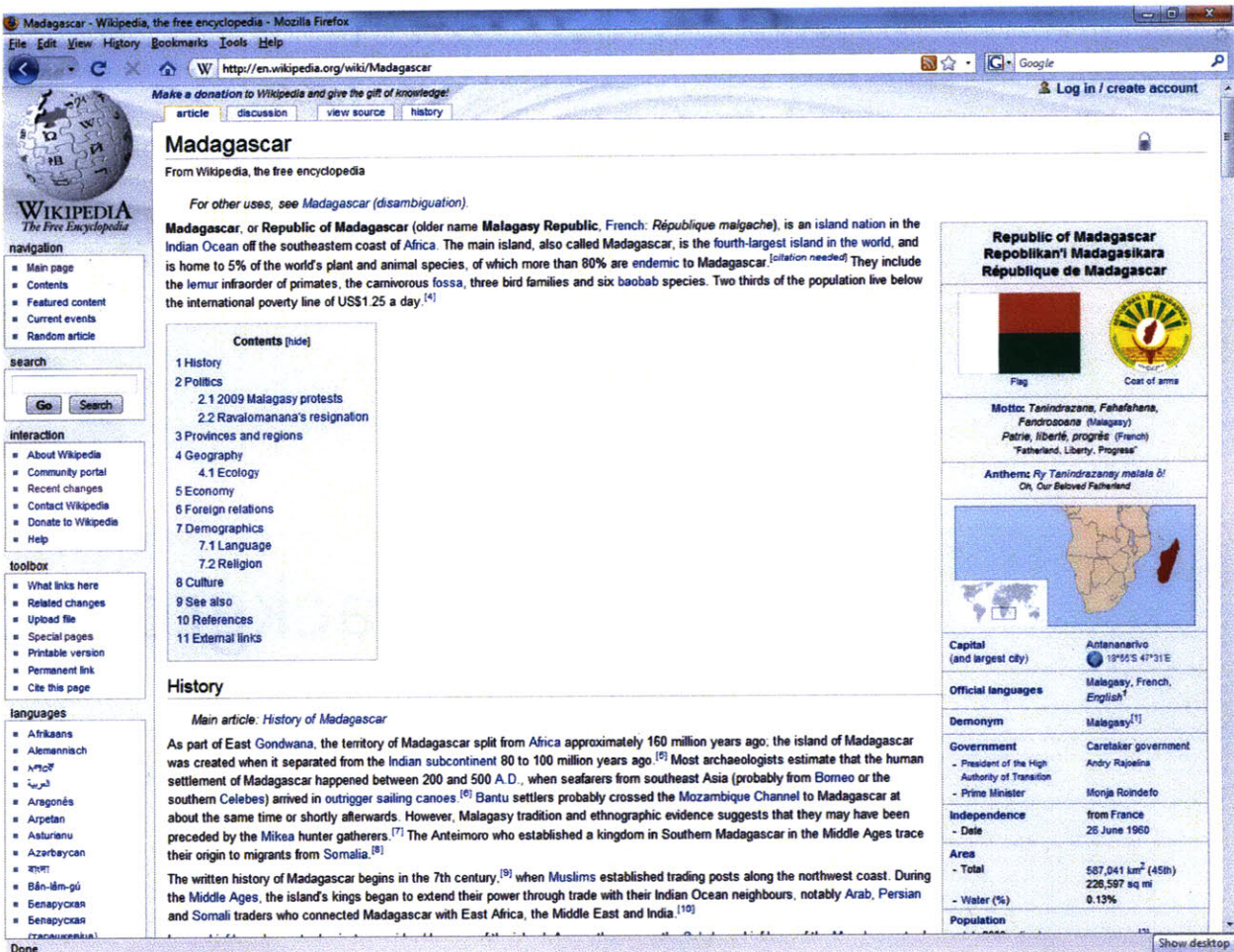
Figure 1: A Wikipedia article about Madagascar. Information is separated into different sections.

Wikipedia is an example of a *wiki*: a website built around the idea of crafting content with a collaborative community. Rather than hiring a team of experts to meticulously author text, wikis rely on the consumers of that text – the visitors to the website – to help with its creation.

A screenshot of the editing interface is located in Figure 2. Wikis are designed to make editing a very simple task, encouraging their visitors to contribute to the site. They use a simple markup language to allow editors to stylize information – typing "'some text'", for example, makes the text bold.

As mentioned earlier, wikis allow anyone to write and edit content. This includes reverting changes made by other people. One can imagine the problems that might arise from this. Therefore, all wikis *track* the changes made to their content. Thus, a visitor can view the history of an entire page and see who contributed which content. This also makes contributors responsible for what they choose to add, incentivizing them to contribute accurate content.

Like Wikipedia, Sociapedia is a wiki (the name similarity intentionally signals this.) Its implementation and appearance is very similar to that of Wikipedia, but it is slightly more restrictive of who can see and edit content. This will be discussed further in the Design section.
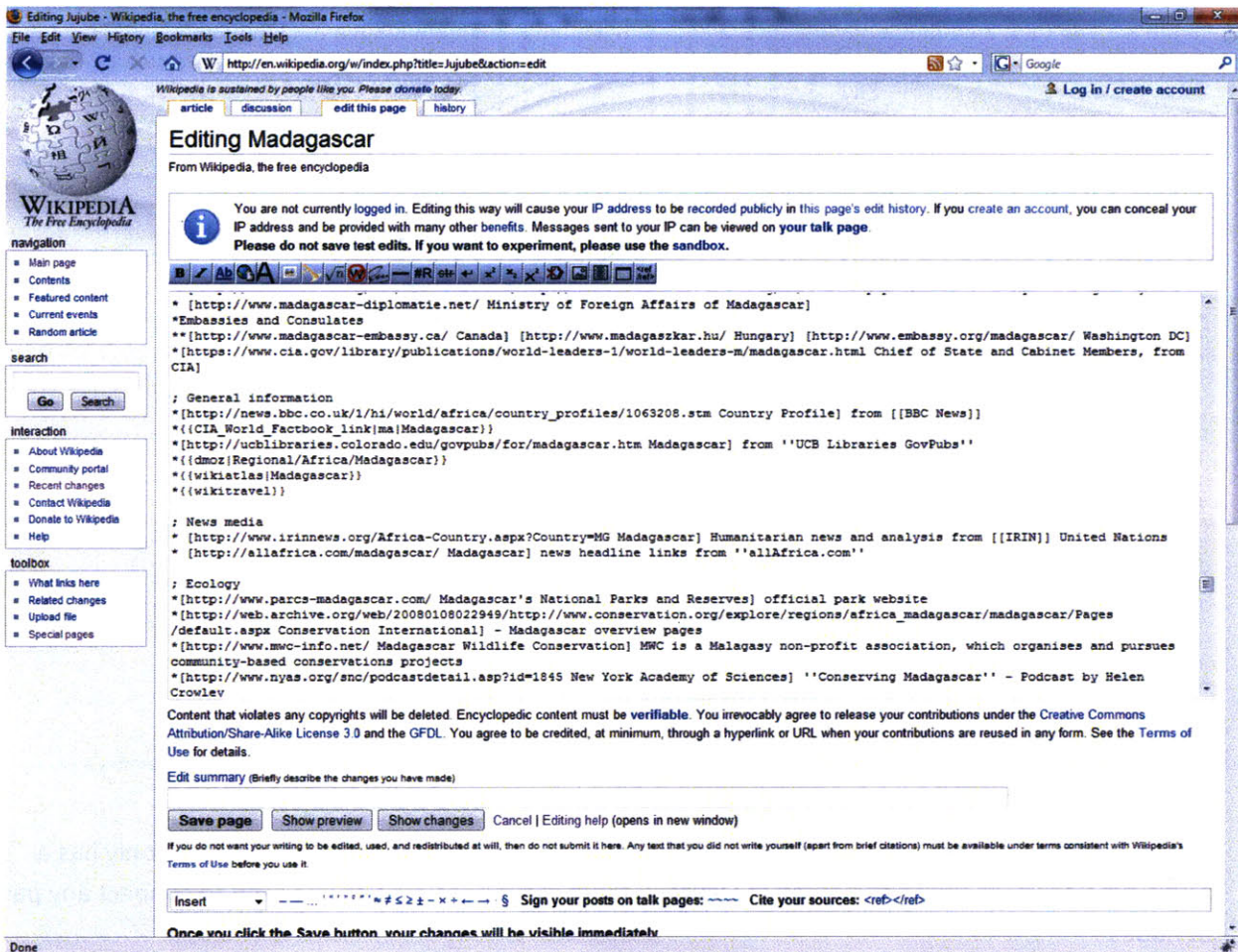
Figure 2: The interface for editing a Wikipedia article. The buttons above the textbox automatically insert the correct syntax for modifying text.

Social Networks

Scientifically, a social network consists of a collection of people and the set of ties between them. We tend to visualize networks using a graph, as shown in Figure 3. Each circle in the graph is called a *node*: they represent the people in the social network. Each line (which is called an *edge*) represents the relationship between the two nodes it connects.

While social networks have been studied in the sciences for decades (Wasserman and Faust), they have recently come to have a different meaning to the general public: online websites that allow you to interact with friends. Some well known examples of these sites would be MySpace and Facebook. They allow you to create profiles for yourself, denote friendships with other people, and interact with them in many different ways. For this thesis, we focus on Facebook: one of the largest social networks in existence and a reasonably canonical implementation of the social networking concept.

Facebook

Facebook is currently the largest social network on the internet, with over 200 million active users (Facebook). Each user of the site each has their own account and can make requests to "friend" other

11
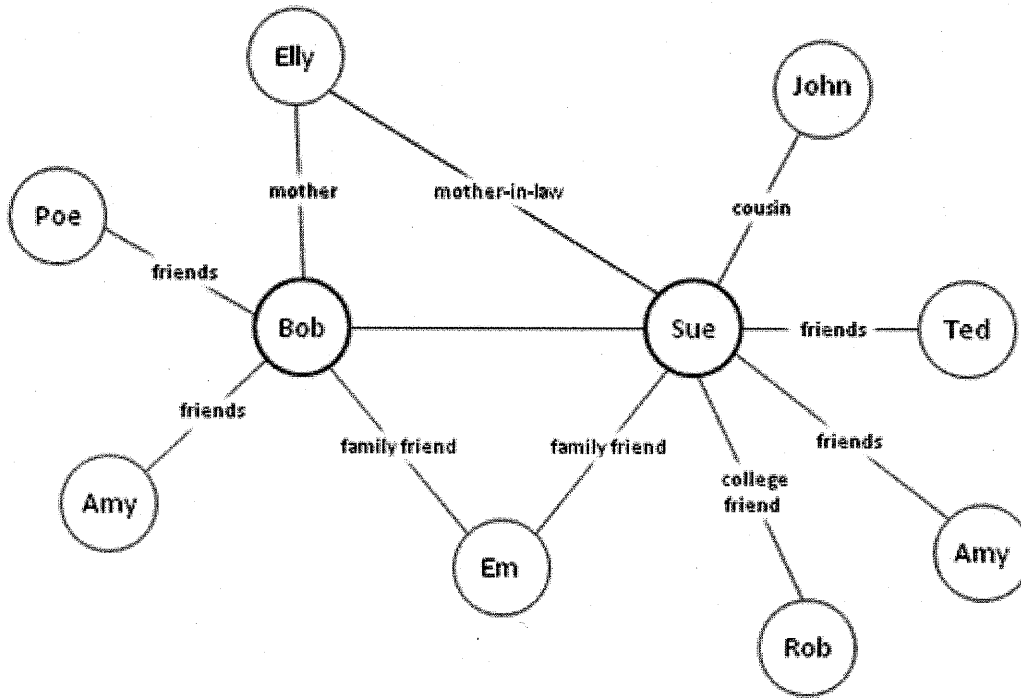
Figure 3: An example of a social network. Each circle represents a person and each edge between them represents a relation. Please don't tell Sue about Amy..

accounts. If the other user accepts, this creates a social tie as described above. Facebook only has a single type of tie, so unlike the graph in Figure 3, there is only one type of line used to connect any pair of nodes[1].

The goal Facebook was founded upon was to accurately reflect the real world's social network. This is not reflective of all online social networks: online massive multiplayer games, for example, purposefully break from this idea.

Facebook has evolved into a fairly complex platform, with numerous types of supported interactions. Since many of these features are used in the development of Sociapedia, we give a thorough overview of them here.

Profile Info

When a user joins Facebook, a *profile* page is created for them. An example is shown in Figure 4. The profile initially contains the w*all tab* (which contains the user's *wall*, a concept discussed in the next section) and the *info tab*. Facebook asks new users to write down personal information about themselves, and this information is used to populate the info tab. Some typical information that people record includes:

- Their current location and workplace (or college, depending on age)

---

[1]Facebook used to support the ability to label a friend relationship through "friend details", where you would indicate how you knew that friend. This feature was dropped due to lack of usage, though.

Figure 4: An example of a profile. Note how each profile consists of multiple tabs
– the *Info* tab contains personal information about the profile user.

- A list of received degrees and other educational history. Sometimes they also list their work history.
- Favorite TV shows, music, and movies
- Hobbies, interests, and activities

The info tab only consists of a small number of fixed sections (e.g. *interests* or *relationship status.*) Because of this, users can only contribute information that falls within its template. The advantage of this rigid structure is that profiles all across the site are homogeneous – whenever you encounter a new person on Facebook, you can instantly process and understand their personal information. There are other popular social networks, like MySpace, that have taken a different approach: they allow users to fully customize their profile even to the point of adding background music.

Facebook users are confident that profiles portray them in an accurate way (Lampe, Ellison and Steinfield).

13
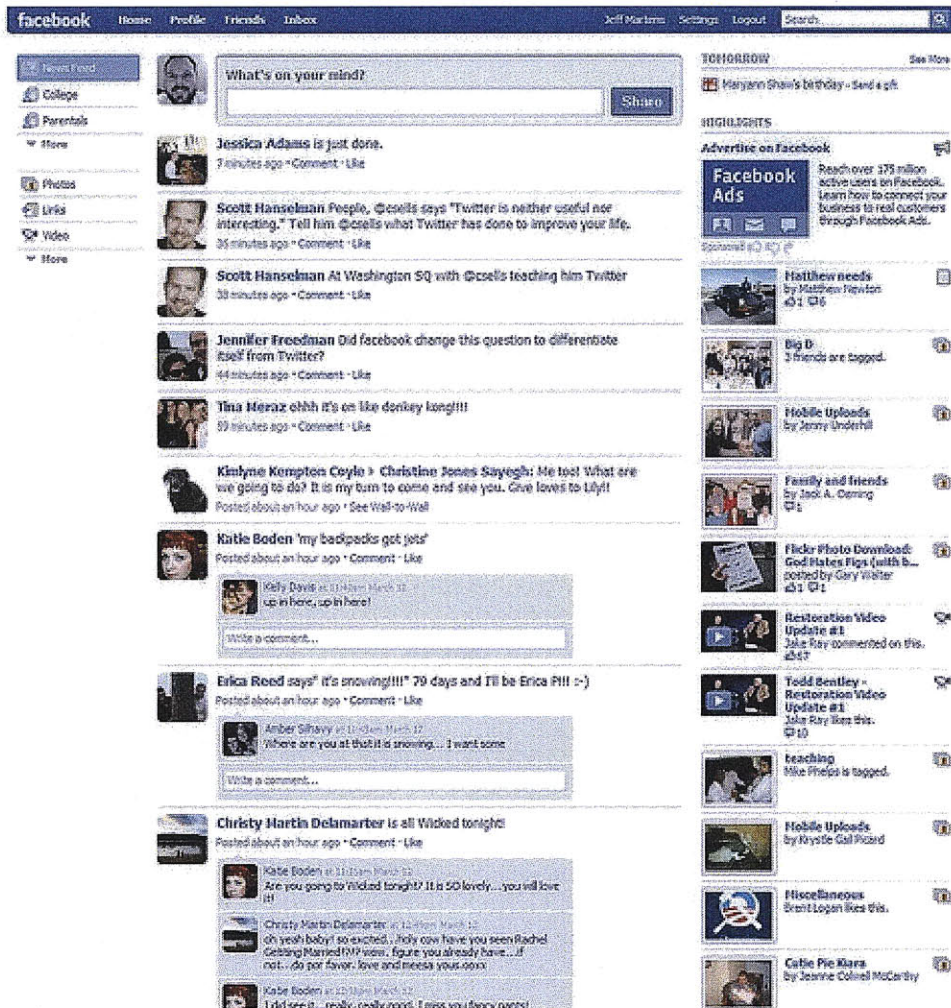
Figure 5: The news feed. The central column is a stream of recent friend activity.

The Wall

The *wall* allows users to share information about their daily lives. Once people join Facebook, they can write posts about their real-life activities, sharing them with friends who now live in distant areas. Facebook allows users to make different types of posts:

- **A status message.** This is a short textual message that usually indicates what the user is currently up to.
- **Photo albums.** Users can post photos from various trips and events they go to
- **Note.** These are longer blocks of text that can contain any content.

There are additional types of posts, but knowledge of the above are all that is relevant to this thesis.

When posting a photo or note, the user can also *tag* it with the names of other Facebook users. Tagging is actually a mechanism that exists across many websites: it refers to the association of a word or phrase with content. The purpose or meaning of this association varies greatly depending on the content type – blog posts tend to contain keywords that categorize the article, whereas photos are usually tagged with

the names of the people contained within them. All tags are non-hierarchical in nature, having no structural relation with each other.

Facebook does not allow users to tag content with anything other than the names of other Facebook users. The tag usually indicates that the user is related to the item (e.g. contained within the photo or a subject of the note.) Tagging is a common and familiar activity on Facebook.

The final important detail about the wall is that it is chronological in nature. Unlike the info tab, which serves static information and can be changed at will, users add content to the top of the wall, pushing older content lower on the page.

News Feed

When people post content, it shows up on their wall: therefore, one way of finding out how what a friend has recently been up to is to check their wall. This does not scale well, though – some people have over a thousand friend relations. Instead, Facebook constructs a stream of recent friend activity on their homepage called the *news feed*. An example is shown in Figure 5. It is hard to overstate the importance of this feed: it is the main window of users into their friends' activities.

Even the feed would not be able to keep up with the content from a thousand friends, though. Instead, Facebook creates the feed by algorithmically sampling every post made by the user's Facebook friends, trying to select the content they believe users to be most interested in. Thus, there is no guarantee that a post made by a user will show up on their friend's feeds. This makes life especially hard for researchers, as it is impossible to analyze exactly how effective posts from the news feed are in spreading applications.

Interactions

Facebook is more than just a directory service of your friends – it is also a means of interaction. For the above posted items, any friend is allowed to write a *comment* about that item. This comment is public: anyone who can see the item can also see its associated comments.

Whenever someone has a comment posted about their item, a *notification* is sent to them. Notifications are a messaging system implemented by Facebook – whenever a new notification is triggered, the user is informed on every Facebook page they visit until they check its contents.

Tags are another form of interaction, although one that occurs at the time an item is posted. Users receive a notification whenever they are tagged on an item.

Applications

The final Facebook component we need to cover is the *platform*. The platform is an infrastructure that allows outside developers to create new methods of interaction within Facebook. Third-party companies develop custom *applications* that users add to their account, and these applications allow users to interact with each other in a variety of different ways. Some examples include:

- **Lexulous**, which is a game very similar to scrabble (Agarwalla and Agarwalla).

- **stat.us**, which allows users to post status messages for other people (MIT Haystack).
- **Sociapedia**, which is the subject of this thesis.

Applications can post to a user's wall just like other Facebook content. This requires the user's explicit permission for every post, though. Applications can also send notifications, which do not require permission.

Applications usually live on a separate page within Facebook. Users can go to these pages as easily as they can visit the profiles of friends. In addition, some applications choose to augment the profile by displaying additional information about the user. This information is displayed in separate tabs (one for each application) and can contain almost any content.

Applications usually spread via *invitations*. Invitations are messages sent by applications users to recruit their friends into joining the application. Facebook takes some measures to prevent applications from indiscriminately spamming users. At present, 63 percent of all application invitations are ignored in Facebook[2] – a fact that played an important role in our design of Sociapedia.

---

[2] Facebook Developer application statistics, 5/30/09

3

Related Work

There are few other research systems similar to Sociapedia. Nevertheless, we start by covering a research application that is in the same space as Sociapedia. We also cover various non-research Facebook applications that allow friends to contribute content about each other.

Collabio

Collabio is a social tagging Facebook game developed at Microsoft Research (Bernstein, Tan and Smith, Collabio). People have fun while playing the game, and as a side-effect they generate accurate information about their friends.

You can see an example of Collabio in Figure 6. Users are initially presented with an obfuscated cloud of words describing the subject (tags) previously contributed by other users. As the user tags their friend, they either add additional tags to the cloud or reveal matching tags that have already been contributed. Tags that have multiple contributors gain prominence, since those are likely to be the most accurate tags. Users also receive more points for guessing those tags.

Greg Smith
Stanford Alumnus/Alumna
Microsoft

Tag Greg to reveal each hidden item. One point for each tag, another point for each other friend who used the same tag to describe Greg!

| | Tag! |

Greg's friends have tagged him with:

... . ajax .....-.. .... band ... be ....
..............-..... c# ....
cruise ... . dev .... dogs ....
....∘.......... ....
....
.... hacker ....
....-...... .... lsjumb ....-..... ....
microsoft .... mscs msr ....
..... poker ....
.... smoky .... stanford ....
.... vegas vibe ....-........ ....
.... wii ....

Choose someone else:

| Start typing a friend's name | Go |

People who know Greg best:

| Amy Karlson | You | Raman Sarin |
| 96 points | 85 points | 83 points |

| Ed Cutrell | Patrick Baudisch | Desney Tan |
| 81 points | 81 points | 78 points |

My Score: 85 points	
microsoft	12 points ✕
poker	11 points ✕
stanford	11 points ✕
vibe	9 points ✕
msr	8 points ✕

Figure 6: An example of a Collabio screen. The visible tags are the ones that have been guessed by the user. Each guessed tag contributes to the user's score, with more popular tags contributing more points.

This system incentivizes people to contribute accurate tags. You know little about tags contributed by others except for their presumed accuracy, so the best way to achieve a high score is to contribute as many accurate tags as possible.

Collabio was developed as a research project, and Microsoft Research carefully analyzed the application and resulting data. They proved that contributed tags were accurate and that they contained information that was unavailable elsewhere (Bernstein, Tan and Smith, Collabio: A Game for Annotating People within Social Networks).

Clearly, Collabio and Sociapedia collect different types of information. While Sociapedia asks users to write flowing text about a person, Collabio limits itself to single-word tags. We look at whether Sociapedia generates information that Collabio does not in our Field Study section.

A more important difference between the two systems, though, is that Collabio does not allow friends of the application user to see that user's tags unless they too use Collabio. In addition, it does not allow

18

a Collabio user to see tags that they themselves did not generate. Thus, all active Collabio users are contributors: you cannot find out more about your friends without either contributing new information or corroborating existing information. On the other hand, it is possible for a Sociapedia user to make extensive use of the application without contributing any content about their friends.

Facebook Applications

There are a number of Facebook applications that attempt to get data about a user from their friends. Unlike Collabio, these were written purely for entertainment and no research about them is publically available.

- Like Collabio, iDescribe is a social tagging application. Users are given a collection of words and have to select five that they believe describe their friend well. Choices tend to be qualities like "honest" or "dependable", so the resulting tags are far more constrained compared to Collabio. We also saw little overlap between the information people chose to contribute in Sociapedia and qualities like these. Tags like "honest" or "dependable" are tricky because they are closer to opinion than fact: it is hard to objectively define them, much less show that they are meaningful descriptors of a particular person.
  iDescribe is fairly popular, with around 400,000 active monthly users at the time of this thesis (Yazeed).
- Define Me is similar in nature to iDescribe, but allows users to write any tag about their friend. While anyone can see what tags a user has been associated with, the contributor for each tag is anonymous.
  Define Me currently has about 20,000 active monthly users (Define Me).
- Compare People is an application where you can compare two friends with various questions, like "who is better at dancing?" Votes are anonymous, and you can see how you rank in various areas compared to all of your friends (e.g. of all people being compared, who is the best dancer?)
  Compare People is extremely popular, with 2 million active users at the time this thesis was written (Chainn, Inc.).
- Testimonials is an application that allows Facebook users to write testimonials about other Facebook users. Users are allowed to delete any testimonials they wrote, and can mark others for deletion: if enough votes are gathered, the testimonial is deleted. As we discuss in the Field Study section, these testimonials do not usually contain the type of information that Sociapedia generates.
  Testimonials currently has about 70,000 monthly active users (Chainn, Inc.).

25 Random Things About You Meme

There is one final social interaction on Facebook that falls in the same space as Sociapedia. Over the last year or so, a common meme – the 25 Random Things About You note – has swept Facebook. A user is drawn into this meme when they get a notification that they have been tagged in a friend's note. Upon looking at the note, they discover that the friend has written 25 personal facts about themselves and

have tagged 25 friends they want to know more about. The note usually includes official rules similar to the following:

> *Rules: Once you've been tagged, you are supposed to write a note with 25 random things, facts, habits, or goals about you. At the end, choose 25 people to be tagged. You have to tag the person who tagged you. If I tagged you, it is because I want to know more about you.*

Like other memes, there are a number of variants of the note also spreading throughout Facebook. In some notes, a set of predefined questions is given for the user to fill out. These questions retain the random nature of the original note, e.g. "what cereal brand do you tend to eat for breakfast?" The tagging rules tend to remain constant, however.

The information contained within these notes is not information that can be expressed in the profile info tab or in single-word tags. Rather, people choose to share random tidbits about themselves:

> *If I don't fall asleep in the first 20 minutes of lying down, I won't fall asleep in the next 2 hours.*

As we will cover later in the Field Study section, Sociapedia also captures information that is similarly not represented in these ways. Unlike this meme, however, Sociapedia gets this information from friends of the user.

4

Design

An example of a Sociapedia page can be found in Figure 7.

Notice how similar Sociapedia is to Wikipedia. This is no accident – Sociapedia was carefully designed to mimic both the look-and-feel and the behavior of Wikipedia. Wikipedia is one of the most popular sites on the Internet[3], so using a familiar look-and-feel will make it easier for people to use Sociapedia.

Another way of looking at this is the fact that wikis are complex concepts: users would just be confused if presented with an unfamiliar wiki website. It is important to exploit the familiar – Wikipedia – to prevent people from being driven away by the perceived complexity of our Facebook application.

---

[3] At the time this thesis was written, it was the seventh-most popular website on the Internet (Alexa, Inc.).

Figure 7: An example of a Sociapedia page. Note the similarity in design and appearance to Wikipedia

This design choice affects how people view Sociapedia. Wikipedia is an encyclopedia – an information reference on a variety of topics. Sociapedia attempts to capture the same feelings of authority and importance in Sociapedia – this notion that the information contained here is accurate and part of a larger ground truth. There may be people who are driven away by this image, but our intuition was that

there is a large swath of people who would find this Wikipedia-styling to be appealing, and that they could form the core of our user base.

A final consequence of our choice is that we hope for users to believe that Sociapedia is not just another random application. There are an enormous number of applications already on Facebook, few of which widely known. We feel that our application's deliberate link to Wikipedia gives it an extra degree of importance even over an identically-functioning application that does not try to make this link.

Interface

You can see a shot of our interface in Figure 7. Sociapedia uses the same background, styling, and tab structure that appears in Wikipedia. Users who wish to edit the page or view the page's history find these buttons in the exact same location. Also, almost no changes have been made to the editing syntax. This means that a user can do anything they would typically do in Wikipedia in the same way here.

A search box is provided to allow users to look up friends who may be on Sociapedia. The search interface comes from Facebook rather than Wikipedia, as the Facebook search tool is far more effective at searching for users. Unlike Wikipedia, the search box cannot do full-text searches of Sociapedia entries – it is limited to searching for articles about friends. Adding the capability for full-text searches would certainly be worthwhile (but difficult), and we discuss this in the Future Work section.

The page in Figure 7 also contains a number of additional components that are not present in Wikipedia. They were developed as the design of the site evolved from a plain wiki into something more socially aware. To better understand this, we start by talking about our initial design for Sociapedia.

Early Design

The initial design of Sociapedia was very simple. Users would join the application and get a page automatically created for them, which they would post about to their wall. Any friend could then follow a link from that post to edit the page. The contents of this page could also be displayed in a profile tab, allowing friends who had not added the application to view the page.

This is perhaps the most literal interpretation of the "wiki on top of Facebook" application concept. Users joining the application would become part of a smaller wiki network living on top of Facebook, whereas those who did not would have no interaction or awareness of it. The fact that it followed this interpretation so faithfully, though, created major problems.

First, there was no natural way for this application to spread. People would have to notice that friends had added the application and then choose to add it themselves. The only other tool that Facebook provides is an *application invitation*, which a large percentage of Facebook users view as spam.

Also, this would only allow us to gather content about users who choose to add the application. Even the most popular applications are sparse when compared to the entirety of the Facebook user base, and this model would rely on a small number of Sociapedia users all being eager enough to contribute content about other Sociapedia users within their social circle. What we want is for those eager contributors to help create pages for everyone.

In essence, the problem was that the application was not taking advantage of the social network beneath it. The current design fixes the above problems by more tightly integrating Sociapedia into Facebook. The key is to modify the way in which pages are created for users.

Current Behavior

In the current model, users of the application are allowed to make pages for any of their friends on Facebook. If the subject is not yet a part of the application, they are sent a notification with a link to their page. Currently, users are required to join the application even to see their own article – they can do this by quickly clicking on a link before being forwarded to their Sociapedia page. Note that this means pages can exist for people who are not application users themselves.

This gives the application a nice way to spread – users can create pages for other people, who join the application and create pages for their friends in turn. This also allows anyone to have a page, rather than just the most enthusiastic application users.

In addition to this change, we added the additional components contained in Figure 7 to further help integrate Sociapedia into Facebook. We discuss each of them after talking about the current content access policy.

Content Access Policy

Facebook (and social networks in general) has a more restrictive content access model than traditional wikis. While wikis typically allow anyone, anonymous or not, to view and edit content, Facebook only allows people you have explicitly created a friend relation with and a few others[4] to view your data.

Since Sociapedia is a wiki, our first instinct was to allow anyone to read or edit any article. The problem, though, is that since it lives on top of Facebook, it is subject to the same privacy assumptions that people make about their social network data. The data that people choose to share about others might be data that the subject is not comfortable having anybody know.

We cannot be too restrictive, however. Wikis exist because a smaller number of people are willing to craft content that is consumed by a larger populace. If access is too restrictive or cumbersome, nobody might use the application.

In the current design of Sociapedia, only friends of a user are allowed to view or edit their page. It is unlikely that anybody who is not a Facebook friend of the user would be able to significantly contribute to the article anyway, so this policy gives enough freedom to editors while still respecting Facebook privacy concerns.

---

[4] The actual policy is somewhat complicated for historical reasons, and differs from person to person based on their privacy settings. A good approximation would be people who you are friends with and people living in the same locale as you.

Discussion Page

Like Wikipedia, Sociapedia includes an additional discussion page for each person. The discussion page is a place for contributors to talk amongst themselves about the content that should form the article itself. Users can ask questions to other friends of the article subject or debate content that should go into the article.

We included this mainly because it came with the software platform we modified to create Sociapedia. While it is identical to its counterpart in Wikipedia, we were not sure that social networking denizens would discover and use it. We talk about these pages more in the Field Study section.

Tagging Friends

Sociapedia implements tagging in much the same way that Facebook photos and notes do. If a user is looking at another person's Sociapedia page, they can tag that page with any mutual friend. We only allow mutual friends to be tagged because we want to make sure that you can only tag articles with people who are allowed to look at that page. Like Facebook tags, notifications are sent to users who are tagged on pages.

We expected users to tag a page with the users mentioned in the article text or to tag users they wanted to come and contribute more content to the page. We also believed that it could serve as a more natural alternative to an application invitation. Rather than sending an invitation to someone you wished to join the application, simply tag them on a page you think they could help with. We discuss the effectiveness of tagging as an invitation in the Field Study section.

Links Box

While the links section in the screenshot only contains a total of three links, it went through many iterations before that configuration was finally chosen. It is important to keep the number of links small, so that users notice and understand each one. On the other hand, there are many compelling features that could be added, and choosing among them is not easy. We selected the above three after careful thought.

- **Add Sociapedia to your Profile.** This shows the user a simple graphic that contains instructions on how to add the Sociapedia tab to their profile. Unfortunately, Facebook does not allow applications to do this automatically for the user.
- **See all Friends with Pages.** This feature was born after we started testing Sociapedia and realized that the search box alone was insufficient. While searching was great for going to a specific page, there was no way for a user to browse Sociapedia and look through the pages they could visit. This link takes users to a page containing a list of every friend on Sociapedia with an article. We would like to further develop this, and discuss creating an additional sidebar component in the Future Work section.
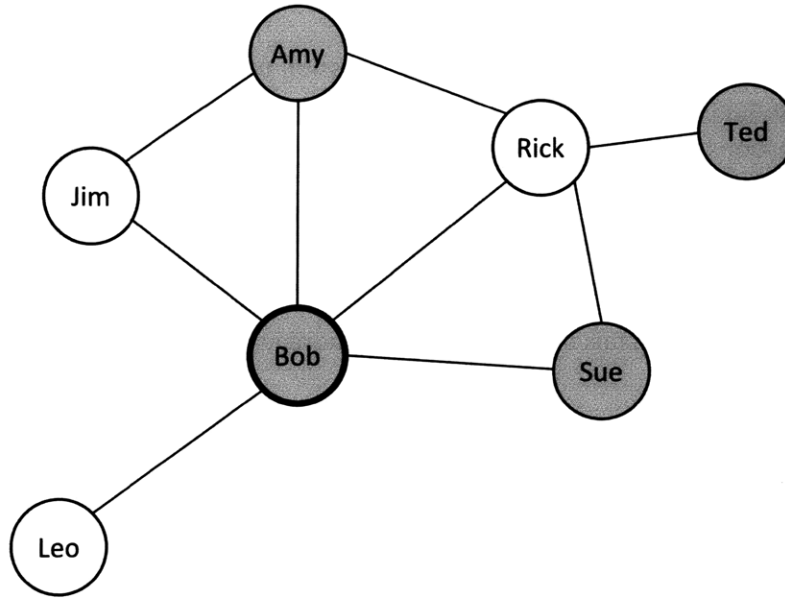
Figure 8: An example to show how the suggestion algorithm works. In this graph, shaded circles are already Sociapedia users and the user receiving suggestions is Bob.

- **FAQ.** The link name is a common internet abbreviation for Frequently Asked Questions. This page is where users can go for help about Sociapedia. While this is certainly important, we would like to find a different location for this link and use the valuable links box real estate for an additional social feature. Just as Windows application users know to look at the top-right menu or to hit F1 when they need help with something, so too might there be a natural place to put this link within the Sociapedia interface. We have yet to discover the right location, however.

Suggestions

The last Sociapedia component is the suggestions box. When a Sociapedia user visits a page, this box displays the names of several of their friends who do not yet have pages, suggesting that the user create those pages themselves. Initially, two different friends were displayed, but after early feedback from the field study, this was increased to five names.

Selecting which friends to display is an interesting challenge. We wanted to create an algorithm that would select close ties of the user: these are the people we thought users would most likely create pages for. It is possible to predict the tie strength between two people using only data from Facebook (Gilbert and Karahalios). Unfortunately, engineering constraints prevent us from doing this.

In the end, we implemented the following algorithm which we believe best approximates the goal above within our engineering constraints:

> *For each friend of the user, count how many of their friends are also Sociapedia users. Suggest the people who have lots of friends who are Sociapedia users and who do not have pages themselves.*

26

This has the effect of "filling in the gaps", so to speak. As an example, look at the network in Figure 8. In this graph, shaded nodes are already Sociapedia users and we are trying to generate recommendations for Bob. The algorithm we use would most highly recommend Rick, as he has 3 friends who are also Sociapedia users. It would also recommend Jim, who has 2, over Leo with just 1.

News Feed Posts

The last two sections in Design talk about Facebook features that Sociapedia uses in its interaction model. Whenever a user edits a page, they have the option of creating a news feed post about that edit. The following is an example of what these posts would contain.

> ### Ted edited <u>Jim Smith's Sociapedia page</u>.
> *.. and is an avid footballer and watcher of football. Jim hosts an awesome superbowl party every year, making his own queso and buying plenty of beer. The 75" tv doesn't hurt either...*

We made two design choices when designing news feed posts that we believe increased their effectiveness. The first is that our news feed posts involve two people: an editor and an article subject. Posts like this are uncommon on Facebook: a large portion of the feed consists of content posted by friends, like photos or notes. We believed that content involving the interaction between two friends was more compelling.

Data from our earlier prototypes supported this. Our initial design only involved one person: posts were published by the article subject, rather than the author of the change. Thus, rather than saying "Sue edited Bob's Sociapedia page", they would say "Help Bob fill out his Sociapedia page." These messages were almost completely ignored until we redesigned them to include both participants.

The second choice is that we show valuable social content within the news feed post. Whenever an edit was posted to the news feed, we included the contents of that edit. This differs from the way many of the applications in the Related Works section operate: they only present the fact that your friends are using their application rather than shedding light on how they are interacting. By showing valuable social content within the news feed post itself, we believe we create a greater incentive for users to join the application and see the rest of the available content.

Notifications

Sociapedia uses notifications in two ways. The most obvious one is to notify somebody whenever their page is edited. This is common sense – users should be informed whenever information on their page is updated in case somebody contributed something they would rather not have made public.

The other time we send notifications is when anybody edits a page that was previously edited by the user. In essence, we assume that any page edited by the user is one they are interested in and send them a notification when further edits occur to it.

This may seem a bit odd, but it parallels the way Facebook handles this situation. Whenever a user comments on an item or photo, Facebook emails them every time another person comments on it too. Applications are limited to using notifications instead of email, but the principle is the same.

5

Engineering

This section talks about the engineering behind Sociapedia and may be more technical than the rest of this thesis.

We also include some engineering tips we discovered from writing Sociapedia. This may be useful to those readers who wish to write their own Facebook applications.

Background: Third-Party Applications in Facebook

Explaining the entire development toolkit for Facebook applications would be well beyond the scope of this thesis. Nevertheless, we present a brief overview in order to allow the reader to more fully understand Sociapedia's implementation.

Facebook applications are externally hosted websites that Facebook loads and renders as needed. Websites can communicate with Facebook through a REST API, allowing them to retrieve social information about the user. Facebook also provides a PHP 5 library that has additional features, like user session management and commonly-used functions.

Applications can also communicate with Facebook on the browser side by using a Javascript API library. This library can be used in conjunction with the PHP one, although they have no interaction with each other. Javascript API calls are made using AJAX and the results can be returned in HTML format, allowing for immediate rendering.

Finally, Facebook supports a set of XML tags called FBML. Since every application page goes through Facebook before being rendered to the user, it is possible for Facebook to manipulate the page before sending it. FBML tags tell Facebook to render certain data about a user in place of the tag (e.g. their full name, their birthday.) This reduces latency, as the application does not have to query Facebook for the necessary data before creating the page to show the user. In addition, this allows the applications to render data without running into privacy concerns, as the user's personal data is never sent to the application.

These technologies were all used in the implementation of Sociapedia.

Implementation

Sociapedia is implemented on top of Mediawiki – the same web platform that powers Wikipedia. This is a natural explanation for the similarity in look-and-feel we strived for in the Design section. The application is hosted on a virtual Linux machine running Apache and uses a local MySQL database to store information in the backend.

A heavily customized open-source Facebook Connect plugin is used to integrate Mediawiki into Facebook. Accounts are automatically created for users logged onto Facebook, making the sign-on process automatic and transparent. Some Facebook API calls are made to collect user information (most notably, their list of friends), but most work is done using generous amounts of FBML. This offloads a lot of the work from the backend, resulting in both lower latency and simpler engineering.

Some technical steps were necessary in order to achieve an acceptable level of performance. APC, a PHP bytecode cache, was installed to prevent PHP from recompiling the Mediawiki source code for every request. As mentioned above, Facebook API calls are made as rarely as possible and cached. Finally, we switched from using an external, dedicated database machine to a local MySQL installation. The reason for this is that Mediawiki makes numerous, low-cost database queries for every page request, and the latency penalty for communicating with an external machine for every database query was enormous.

Integration

Articles use Facebook IDs as titles, preventing any collision problems (a Facebook ID is a unique unsigned 64-bit integer assigned to each user.)

Facebook recently allowed users to claim a unique username that would serve to identify them to the world. Internet users are used to the idea of dealing with usernames from applications like instant messaging, so this could have been a better solution. Unfortunately, third-party applications were unable to learn or access these usernames while the application was being engineered.

The only data that the application requires from Facebook is the user's list of friends. From that list, we can create a list of every friend with an existing page, as well as calculating the best suggestions for new pages to create. This list is updated every twelve hours at present (although we may tweak this for future releases.)

Tips for Facebook Application Development

Logging: Logging is essential to web-based research projects. A large collection of data can be analyzed with ease, shedding light on the behavior and motivations of users. Since applications are rendered through Facebook, however, the default Apache logs are not very useful: all activity will appear to come from a small collection of IP addresses. Therefore, we had to implement our own custom logging system to collect data.

Latency: Facebook can be unforgiving with latency. Nothing drives away users like error messages, so it is important that applications reduce request latency as much as possible. However, this is made much more difficult by the fact that the latency for Facebook API calls can be quite high at times. Therefore, optimizing the application performance, reducing the number of API calls, and caching all data received from Facebook is essential for third-party application developers.

6

Field Study

In order to understand how people use Sociapedia in the wild, we conducted a one-week user study. We gathered ten people and gave them the following instructions:

> *We expect that you will spend 20 minutes per day using the application. This means being an active user of the application: writing and editing pages, tagging friends on pages, and so on.*
>
> *Sociapedia will ask you to post some of your editing activity to your news feed. We are requiring that you accept to share this activity on your news feed at least once every two days."*

These instructions are purposefully general – we only ask them to actively use the application rather than requiring them to write a specific number of pages or tag a certain number of people. This ensures that research findings are as bias-free as possible.

We did give them one specific directive – the requirement that they post to their news feed at least once every two days. As it turns out, most respondents posted more often than we required. This suggests that the natural behavior for Sociapedia users was not affected by this requirement.

All ten of our users were affiliated with the MIT CSAIL research lab. As a result, there was some overlap in their social networks. This allowed us to observe a wide variety of editing patterns and interactions over the week, with some pages being edited by multiple users in our study.

Of the ten users in our study, nine completed the entire one-week study. The tenth was not an active contributor to the application (although they were active in viewing pages.)

Note: We have included some quotations from Sociapedia pages below in italics. In order to protect user privacy, we have changed the names to another with the same gender.

Statistics

At the end of the study, Sociapedia contained 116 pages (75 of which were created during the study). 47.8% of those pages contained over 100 words and 73.3% contained over 40 words[5].

A majority of the articles – 53.1% – were authored by a single friend. 35.6% of the articles had multiple authors, and the remaining 11.3% were entirely self-written. Note that these figures are across all page sizes: if we limit ourselves to the large articles (over 100 words), then 56.4% of them had more than one author.

By the end of the study, Sociapedia contained 243 user accounts. 58 user accounts performed editing actions, and 34 accounts created at least one page. Within the week of the study, a total of 122 accounts were created. Of these accounts, 112 of them were friends of a study participant. 32 user accounts performed editing actions during the week we conducted the study (this includes our 9 active study participants.)

Article Content: What Did People Write

The easiest place to start is by looking at the content that users chose to contribute. Here are our key questions:

- ◆ What categories of information are generated in Sociapedia?
- ◆ Does Sociapedia provide correct information that cannot be found on the Facebook profile?
- ◆ How is this information written down? Do users structure and organize the information?
- ◆ How were discussion pages used? How did those pages affect article content?

Categorization Scheme

We went through each article and created a set of categories to describe the information contained within the articles. We describe each of the nine categories below. Note that unless otherwise stated, all quotations within this subsection are examples rather than quotes from Sociapedia articles.

---

[5] We picked 40 words because that a typical length for two sentences (Oxford University Press).

- **Professional Information (or College Information).** For adults, this contains information typically found on a resume. This could be about their current or prior jobs, as well as earned degrees.

  For students, this is information about their current college. If the subject is a graduate student, this could also include their current research interests; if the subject is an undergraduate, this could also include information about their choice of major or the courses they are currently taking.
- **Early Life.** For the purposes of this thesis, we consider any information about their high school years or prior to be part of this category. Information about their birthplace or where they grew up would also be included here.
- **Clubs and Organizations.** What clubs and organizations are they part of (sororities, student government, etc.)? This would not include something like "takes Jazz lessons every week", but would include "part of the MIT Jazz Ensemble."
- **Hobbies and Activities.** What does the subject do on a regular basis? This includes hobbies like tennis, skiing, or violin as well as "works out at the Z-center" or "always watches Lost with the roommates."
- **Likes/Dislikes.** What does the subject like? Radiohead? Chocolate? Muenster Cheese? One thing to watch out for is that a sentence like "John likes to go skiing every weekend" is an activity, not a like. This category also includes information like favorite TV shows, music, or movies.
- **Anecdotes.** These tales cover a specific incident in the user's past. They can be mundane, "The night before flying out to UIST 2008, Jeb and his skateboard parted ways unexpectedly and his pinky finger suffered the consequences", or even a single sentence, "Emily re-read the entire sixth [Harry Potter] book in one night."
- **Personality and Character.** Do people find the subject funny? Sociable? Always smiling? Also, what do they think about the person's character: are they honest to a fault or sometimes deceptive?

  We need to be careful here: this does not include comments like "Sue is awesome!" In this context, awesome is not a meaningful word about Sue's personality.
- **Relationships.** Information about the subject's relationships with others. The other person does not have to be a Facebook user.
- **Other.** There is a wealth of information that cannot be easily classified. An example from a Sociapedia article: "Amy even has a Facebook fan page dedicated to her. And with this her fan base has since grown internationally."

We developed these categories organically: reading through each article, we looked for the most popular types of information contributed by users. Each of the above categories is contained within at least ten pages. In addition, we looked at all remaining facts that fell into the *Other* category: there were no other categories that could contain more than two or three of them.
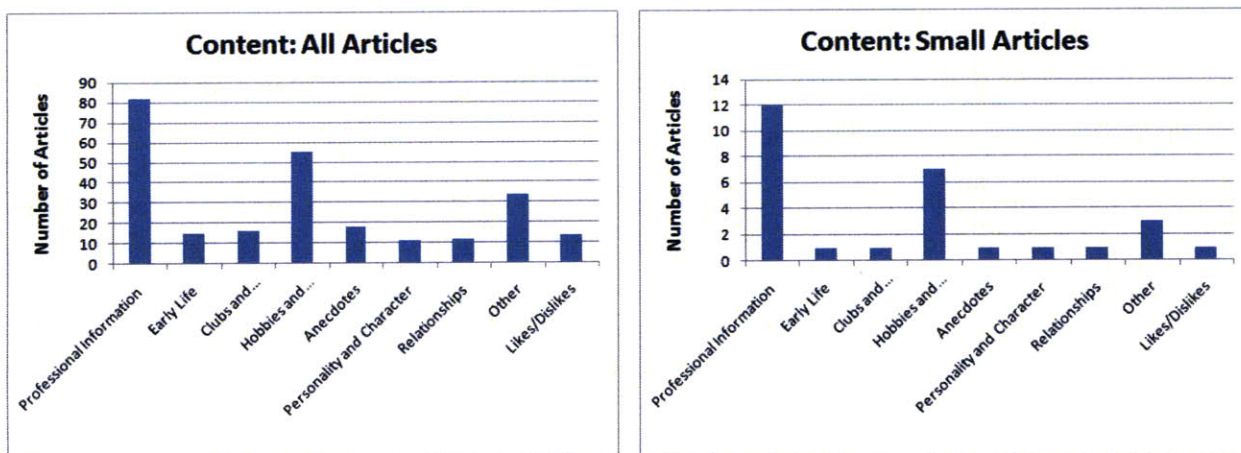
Figure 9: The left chart shows how many articles contained information in each listed category. The right chart shows the same, but only includes articles below 40 words in length.

We also tried to avoid overlap as much as possible when designing these categories. The only categories that have some potential for overlap are *Clubs and Organizations* and *Hobbies and Activities*, but we found no examples of overlap between them (or any other pairs of categories) during the classification process.

There was one final issue that came up when trying to classify information. This was *implication*: information could imply certain other facts that would fall into other categories. Knowing that someone is going to MIT, for example, tells you that they are not married and that they live in Cambridge. Should this implied information be considered part of the article?

We decided against it. We did not want to categorize articles based on the information a reader could glean from them, but from the facts that contributors explicitly chose to contribute. This is a somewhat subtle decision, and our hope is that it would not substantially affect our findings.

Content Breakdown

Each article was classified based on which of the above categories it contained information from. The thesis author's page, for example, contained *Professional Information, Activities, Anecdotes,* and *Other*.

The results are shown on the left chart in Figure 9. The three dominant categories are *Professional Information, Hobbies and Activities,* and *Other*. We discuss each of these below.

Professional Information. The popularity of this category can be traced directly to the large number of college students using the application. Most entries included information about their year and choice of major. Only three or four entries included information about classes taken, but 23 entries included information about current research interests.

Hobbies and Activities. When people list their hobbies on their Facebook profile, they usually respond with formal terms like "tennis" or "swimming" (this observation is based on our comparison in the Uniqueness section below.) Sociapedia can capture this information, but it also captures more informal

information. Some examples from different articles would be *Chinese yo-yo-ing, tinkerer,* or *regularly hits up the Z Center.*

Other. As we mentioned earlier, the contained facts are so varied that further categorization is impossible – we would need as many categories as facts! We mentioned a quote from this category above:

> *Amy even has a Facebook fan page dedicated to her. And with this her fan base has since grown internationally.*

Some other examples:

> *sometimes during his graduate years he'd go out and buy a loaf of bread as a snack.*

> *Stupid rules are an irresistible target for him, like one MIT bulletin board that said "Do not post anything here"*

Even if they cannot be easily classified, facts from this category are still interesting and valuable.

Short Articles

We mentioned in the statistics that 73.3% of articles contained over 40 words. The remaining articles are short: perhaps one or two sentences at most. What is the nature of these articles; did they still contain useful information, or were they just random sentences like "I have a new page!"?

We looked through our categorization of these short articles. A large percentage, 85.8%, contained factual information from one of the above categories. In addition, Figure 9 shows that the distribution of information was almost identical to the longer articles. Even short Sociapedia articles are valuable, a point we further reinforce in the next section.

Uniqueness

Due to privacy limitations, we are only able to access the Facebook profile of a small number of Sociapedia users. We randomly picked 12 users whose profiles we had access to and compared their Sociapedia page to their profile. These users had varying page lengths and categories of information.

10 out of 12 of these users had articles which contained information not present in their profile. The 11th article was ambiguous: it seemed to contain additional information but this was wrapped within humorous content and may not have been clear to a reader. We include the ambiguous quote from that article here:

> *Recently, Ashley went missing for two weeks. She claimed to have been staying with family visiting from Italy, yet she can provide no witnesses who saw her with any supposed "mom" or "brother" during this time. Therefore, it is reasonable to assume that Ashley has been plotting something evil. Friends of Ashley, be warned. According to Law and Order SVU, killers are most likely to attack those closest to them.*

We also compared the information present in Sociapedia articles to those given in Collabio tags. Excluding the Sociapedia application authors, there were 22 people with both Sociapedia articles and Collabio tags. In our comparison, we only included Collabio users that had tags from more than one person, but we included users regardless of the length of their Sociapedia articles (there were several users with articles that were ten-to-fifteen words.)

We found that 93.8% of users had Sociapedia articles that contained information not present in Collabio tags. What is most remarkable is that some of the Sociapedia articles were incredibly short. An article that was only 10 words in length still contained information not present in over 63 distinct Collabio tags!

Accuracy

We asked several users about how accurate their pages were, and got some interesting responses. The reason that some Sociapedia pages contain incorrect facts is not out of ignorance or malice, but humor:

> *Yes, my girlfriend implied I was gay. This is her way of teasing me, and I figured it would cause more trouble to correct it, so it is still there.*

These inside jokes can be fairly bold. On one person's page, the first paragraph contained this:

> *After a vicious attack by drop bears and hoop snakes coordinated by the sinister bunyip mastermind Quin-tan killed his family at the tender age of 3, Bob declared all of bunyip kind to be his mortal enemies and has dedicated his life to destroying them.*

When we were analyzing accuracy, we hypothesized that this type of joke information inhibited contributions. Our reasoning was that people would have no idea how to edit or contribute to the page short of deleting the fictitious content, which they would be reluctant to do. We went through the articles and looked at their edit history to substantiate this, but found no proof that this was correct. Indeed, we found examples where users contributed meaningful, correct information even after an entire fictitious article was written about the subject: they wove factual information directly into the fictitious content.

Joke information was pervasive. It was not limited to one social network or demographic, springing up in disparate pages across Sociapedia. However, joke information aside, the people we talked with all attested to the accuracy of the information contained within their pages. We also note that the edit histories contained very few delete actions: since article subjects are notified of every contribution, they would have easily been able to delete or fix any incorrect information. This fact alone suggests either that the information is accurate or that users were willing to tolerate inaccurate information on their pages; combined with the feedback from users, however, we have strong reasons to believe that contributed information was accurate.

Structure

We can also look at how well organized the articles were. As mentioned in the background section, Wikipedia articles have a clear organization, with content spread out over separate sections. Does this also occur in Sociapedia articles, which have far fewer editors?

Out of the articles with over 100 words, 41.8% of the articles had section headings. This ratio drops to 28.4% if we look at every article, but the former percentage is a better representation: we only want to look at articles that are large enough to contain many types of information.

Despite this separation, there is little commonality in the headings across articles. This makes sense, since there is no style guide or ability to decide on a standard. Page headings were arbitrarily created based on the topic that the contributor wished to write about, which is very different from the consistent page headings that Wikipedia enforces across all articles of a certain type. The most popular headings, *Research*, *Random Facts*, and *Other* were simply topics that multiple people shared, and even these popular headings only occurred in 4 pages each. 80.2% of all topic headings occurred only once.

Discussion Pages
The discussion pages were almost completely unused during the field study. The ones that were created tended to include meaningless content like, "Hello World!"

An explanation for this could be that the interface was poor: people did not notice the discussion page or did not use it because it was too cumbersome. Although the discussion pages had a social purpose, they used an interface from Wikipedia rather than Facebook.

However, it is still unclear whether discussion pages would be used even if the interface was changed. There was no evidence of contention within the set of articles created within the study, which is the primary use of discussion pages on Wikipedia (Kittur, Suh and Pendleton). This is a question that would have to be addressed in a future study, however.

Recruitment: How Does the Application Spread
There are three different ways that new people can be brought into Sociapedia:

- **Having a page created for them**. You get a notification when someone creates a page for you.
- **Seeing a post on the news feed**. Every time someone edits a page, they have the option of making a post to their feed. Since your news feed is an algorithmic sample of actions by your Facebook friends, you have a chance to see their posts in your feed.
- **Being tagged in a page**. If you're tagged in someone's page, a notification is sent to you.

These methods differ in scope and nature. While news feed posts command larger audiences, the personal user-to-user nature of a page creation or tagging might change the user's behavior in interacting with the application.

There are three questions we can ask about recruitment:

1. **How many people joined via each of the above methods?**

We analyzed the users who were active during the study, determining how each of them joined the application. The results are contained in the left chart of Figure 10. A surprisingly large percentage of users – 65.9% – joined the application because they saw a Sociapedia post about a friend's page in their news feed.

This is especially surprising because there is evidence that the current news feed is not friendly to applications. Facebook's news feed has gone through many redesigns, and the most recent one caused a severe drop in application traffic (Lynch). Our field study clearly shows that the news feed remains a potent tool for spreading applications, however.

2. How effective was each method? That is, how many people solicited by that method ended up joining the application?

Unfortunately, we are not able to determine how effective news feed posts are. Since news feeds are algorithmic samplings of content posted to friends' walls, we have no idea how many people saw each post.

Astonishingly, every single person who had a page created for them during the study joined the application. However, we have to be careful here: Sociapedia requires you to join in order to view your page. Thus, it is possible that users who had no interest in the application joined anyway just to look at what was written about them.

A better way to measure the effectiveness of creating pages for other people is to ask how many of those users went on to look at a friend's page. If they started exploring Sociapedia beyond their own page, we can say that they showed an interest in the application equivalent to those people who clicked on a news feed post to look at a friend's page.

It turns out that 75 percent of users who joined after having a page created for them went on to explore the application after looking at their own page. Thus, creating pages for other users is a highly effective recruitment tool.

The third recruitment type is tagging. Out of all of the prospective users who got tagged in a Sociapedia page, 25 percent of them went on to join the application. This is a lower rate than Facebook application invitations. Users who are tagged in a page are notified in the same way as users who have a page created for them, so we can conclude that tagging is an exceptionally poor recruitment tool.

New Users   **New Editors**
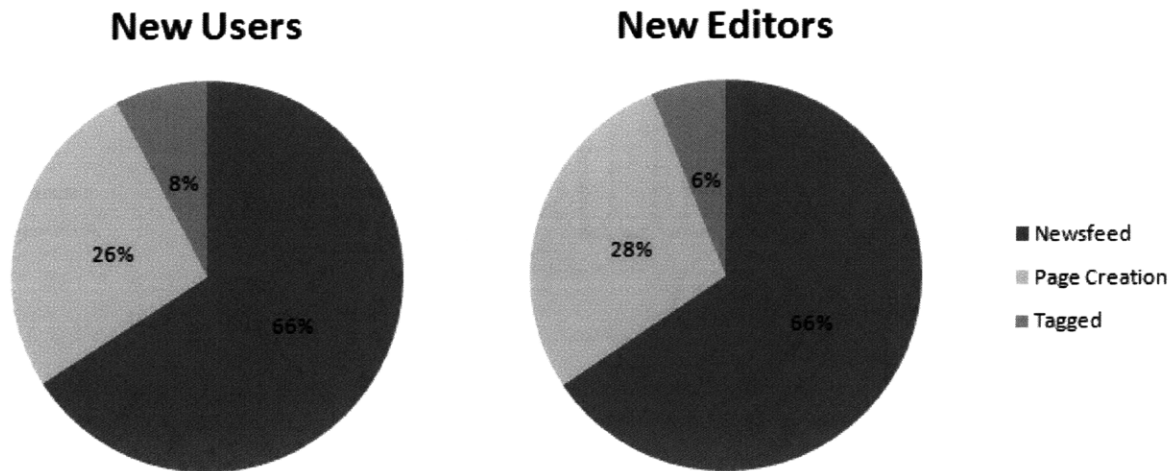
Figure 10: This chart shows the proportion of people who joined via each recruitment method. The left chart is over all new users during the week of the study, while the right chart only includes people who edited or created an article during that period.

3. **Did the behavior of the user change based on how they were introduced to the application?**

We answer the third question by computing how many *editors* joined Sociapedia via each of the above methods (an editor is someone who has made at least one edit to another person's page.) A significant difference in the proportions of editors versus users would strongly imply that the way people join the application influences their future behavior.

The right chart in Figure 10 shows the recruitment method for every user who joined during the study period and later went on to become an editor. Comparing the two charts, one can see that the proportions are nearly identical. This is strong evidence for the fact that the way a user joined Sociapedia does not significantly influence their future behavior on the site.

Editors: How is Information Created?

Sociapedia relies on editors in order to spread. Whether new users join to see the content on their page or because they saw a news feed post about a mutual friend, almost every new user can be correlated with a single triggering edit action.

The most important question we can ask is how many people became editors. This will allow us to learn about how Sociapedia will spread: are there enough editors to allow Sociapedia to spread throughout all of Facebook? In addition, we want to understand the motivations behind people who choose to become editors.

Spread

We have to be careful when attempting to analyze the proportion of users who become editors: this metric is skewed by the presence of study participants. If we include them, we can say that 28.6% of all users during the study become editors; exclude them and the ratio drops to a conservative 21.6%. We
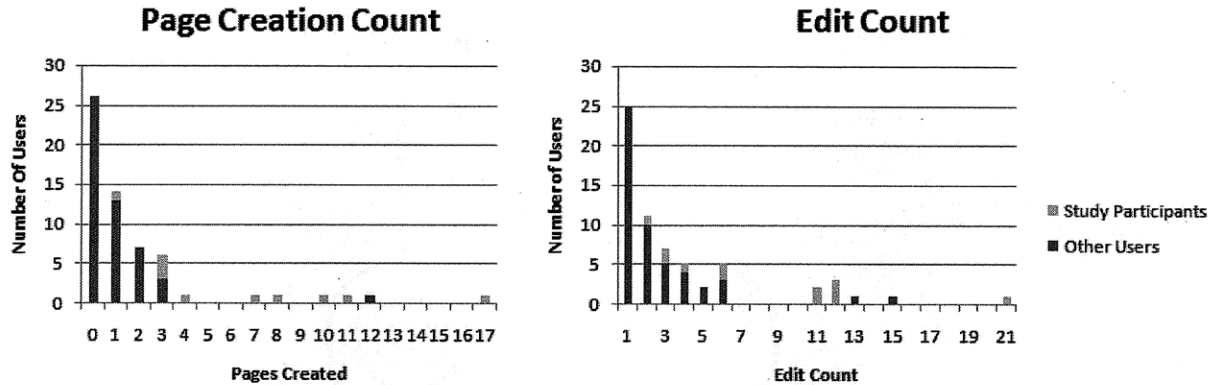
Page Creation Count

Edit Count

Figure 11: The left chart is a histogram of pages created by editors (so it only includes users who have made at least one page edit.) The right is a histogram of the number of edits made by each editor. Study participants (and the application authors) are highlighted in a different shade.

can also look at this proportion over the entire lifetime of Sociapedia: 23.9% of all users have become editors since Sociapedia was launched.

To determine whether or not Sociapedia can spread throughout Facebook, we use a mathematical model known as a *branching process*. This model works as follows.

- We start with a single process *p*.
- At every timestep, every existing process replicates itself a random number of times (according to some fixed probability distribution).
- Each process dies after replication

An important question we can ask about branching processes is whether they eventually become extinct: will there be some timestamp after which no more processes will remain? If μ is the expected number of processes that a single process will create, then extinction will definitely occur if μ is less than one (Grimmett and Stirzake).

In Sociapedia, each editor is a process that replicates itself by creating pages for other users. We want to know whether Sociapedia will eventually stop spreading completely, i.e. the process will go extinct. Combining the equation above with our statistics, we can say that unless editors create an average of at least four pages each (4.2 pages, to be precise), Sociapedia will eventually stop spreading.

Unfortunately, that is exactly the situation we are in. The left chart in Figure 11 shows the number of editors that have created a specific quantity of pages. Some key points:

- If we include the study users, editors created an average of 1.91 pages. Exclude them and this drops to 0.96 pages.
- With one exception, nobody who was not part of the study created more than three pages.

We can also look at how many page edits each user committed. News feed posts are triggered by page edits, and we saw above that they were the most effective recruitment tool. The right chart is a

histogram that shows the number of page edits made: editors made an average of 3.73 page edits, with non-study editors making an average of 2.59 edits.

Why were some users more active than others? Was there something that distinguished prolific editors? We talk about this a bit in the Behavior section below, but even after interviewing a few of our editors and looking at the data, this was not a question we were able to clearly answer. Regardless, we introduce some ideas to increase the productivity of editors in the Future Work section.

Motivations

We do not have a good way of predicting the people who chose to become editors. The one common metric we have over all Facebook users is their friend count: this gives us some indication about how active they have been on Facebook. However, there is no correlation between the average number of friends one has on Facebook and the probability of them becoming an editor.

We asked some editors directly about the pages they chose to edit. With one caveat, people usually edited the pages of their close ties. Some quotes from the editors:

> It's easier to write pages for closer friends because you know them better.

> I had much more emotion in my [close tie] posts and they were more fun to write, because I knew more about them personally

The caveat is that some people felt more pressure when writing about close friends. An example:

> Writing for closer friends was a bit harder because I felt under more pressure to do a good job of it

In looking at the responses, users who chose non-close ties to edit did so because it was "easier." This is our major social motivator (or demotivator, in this case): fear of how a close friend might react to the content the editor chooses to share.

Reciprocation

Past sociological research has shown that reciprocity is a major social motivator (Cialdini). Sociapedia has a simple notion of reciprocity – contributing to the page of the person who brought you into Sociapedia. Since this person may be the only person you know who already has a page, this is a very natural action we would expect users to take.

Reciprocity also plays a role in other applications in this space. When Collabio users were asked about the reasons they tagged someone, 82% of survey respondents cited that the friend had tagged them first. In addition, 82% of Collabio users who joined after being tagged reciprocated by tagging at least one of the friends who had tagged them (Bernstein, Tan and Smith, Collabio: A Game for Annotating People within Social Networks).

We hypothesized that people were much more naturally inclined to edit the page of the person who created their page. This would imply that the best way to get people to contribute to your page is to edit or contribute to theirs.

Here is how we analyze reciprocation. We take the set of users whose pages were created by another person and who edited at least one other page themselves. Out of these people, how many of them edited the page of the person who created their page? If reciprocity is a major factor, we would expect this number to be quite high.

As it turns out, it was surprisingly low. Only 42.9% of the users in the above group edited their creator's page, meaning that almost three-fifths of the users edited and created pages without touching the person who brought them into Sociapedia. Since Collabio's reciprocity rate is almost twice as large, this suggests that reciprocity is not nearly as powerful a motivator for Sociapedia as it is for other applications in this space.

Behavior

From the first section in this paper, we have an idea of what kinds of content were contributed about users. We also want to look at it by editor: who were the people who contributed more? Did they provide certain types of content, or was it dependent on the pages of the people they were editing?

We were not able to classify every single edit made to the pages due to time constraints (although we did look at the effects of the *initial* edit, which we discuss below. In specifically looking at the edits of users who made between 3 and 6 edits, however, we noticed something interesting. For the most part, there were no immediately discernable patterns: people's edits were heavily dependent on the identity of the page they were editing. There was one interesting common pattern among a portion of these more prolific editors, though: the *friend comment*.

Friend comments are signed remarks written in the first person at the end of an article subject's page. An example is shown below.

> *John Smith*
>
> *Tim knows how to passively dry dishes. His whiteboard is perfect for conveying your innermost whatever happens to be on your mind at the time thoughts. Also he can binary solo.*

While these comments include personal opinions the author has about the subject, they almost always also contain factual information. Much of the information naturally falls into the *other* category: interesting trivia about the subject that is unlikely to be found elsewhere. Quotes containing information from the other two popular categories (*professional information* and *hobbies and activities*) were also present during the study, though.

One question to consider is whether we could get the same information by completely scrapping articles; instead, just allow users to write comments about each other. This would be very similar to the Testimonials application from the Related Work section. Unfortunately, without the backdrop of Sociapedia, users usually only write personal opinions about the subject:

once a friend .. always a friend!!most helpful and most caring friend!! real understanding n
patient....and GOOGLE anytime and anywhere!!!!!!!

How did this editing pattern arise? It originated from one of our study participants and spread throughout those friends of hers that used the application. Intuitively, this editing pattern encourages a greater quantity of edits: it makes it easier to write content for any of your friends. However, we cannot statistically say that this editing pattern causes editors to write more; it is entirely possible that this was simply due to demographics.

We discuss a way to integrate friend comments into the editing interface in the Future Work section. We have to be careful: while it is clear that comments on Sociapedia contribute factual information, we do not want users to stop editing the actual articles and only write comments.

Initial Creation

How does the initial page content influence the outcome of the page? In particular, do certain types of information lead to longer articles or a greater number of contributors? To answer this, we looked at the information contained within the *first* edit made to any page (the information contributed by the page creator.) We then ran a statistical analysis with this against both the length of the article and the number of editors.

Our analysis found that there was no significant correlation: in other words, the information contained within the information from the page's creation did not predict the final length of the article or the number of people who would edit that article.

7

Conclusions

Information

The most valuable conclusion of our work is that Sociapedia generates information not easily available on Facebook. We have compared Sociapedia to both Facebook profiles and to another application in this space (Collabio), and have shown that even small Sociapedia articles contain unique social information.

We have also been able to characterize the nature of this information. Since many of our pages are about students, college information is very prevalent. Hobbies and other activities is another popular category, with articles focusing more on informal activities like "regularly hits up the Z Center" or "watching the new Colbert Report" – a contrast to the formal data from the Facebook profile section. Finally, many articles contain random tidbits that defy categorization yet are still factual and interesting.

Sociapedia does not have the organizational properties of Wikipedia, with only 41% of our larger articles possessing structure. This came as no surprise to us. Wikipedia has an entire infrastructure devoted to organizing articles: users can tag articles that do not meet the standards set forth in a style guide, and other users can then clean up the affected articles without regard to content. Since Sociapedia articles

can only be edited by a few people, we do not have a set of contributors who solely devote themselves to tidying up large numbers of articles

Application Recruitment

There are various methods to recruit new users into Sociapedia. The way new users use Sociapedia does not change based on how they are recruited, so the only thing we need to compare between methods is how productive they are.

Although we do not know what percentage of people that saw news feed posts went on to join the application (nor even how many users saw the posts), we do know that they were a highly productive recruitment tool. A majority of our users (almost two-thirds, in fact) joined after seeing a post on the Facebook news feed.

Every user who had a page created for them proceeded to join the application, and three-fourths of them went on to further explore the application. While this method was by far the most effective, its limited reach in comparison to news feed posts about page edits made it much less productive.

Finally, tagging was an exceptionally poor recruitment tool. Only a quarter of users who were tagged in pages went on to join the application, a percentage even lower than the 37% of users who join Facebook applications due to invitations.

Growth

At present, Sociapedia cannot completely spread across Facebook. Editors create far too few pages, making it likely that the application's spread would eventually stop completely. There are three approaches we can take in attacking this problem, all of which can be done simultaneously.

First, we can try to increase the number of page edits. This would increase the news feed posts, which is our best recruitment tool. Admittedly this is a bit of an oversimplification: one person making twenty news feed posts is far different than twenty people each making a single news feed post. Nevertheless, any increase in page edits can only be beneficial. We know that many prolific editors wrote comments about friends on their Sociapedia page, so making this easier is something we address in the Future Work section.

The second approach is trying to increase the number of pages created by editors. If editors were to create roughly four pages each, Sociapedia could spread throughout Facebook. Unfortunately, they only create an average of one page at present. We need to add an incentive to editors to create at least four pages. Many social applications use a *badge system*, and we discuss this further in the Future Work section.

Finally, we can try and increase the number of editors. Currently, almost a quarter of all users become editors. Increasing this by a mere eight percent would mean that editors would only have to create an average of three pages, although we get diminishing returns beyond that (in order to get down to an average of two pages, we would have to increase the number of users who become editors to fifty

percent.) We have two ideas that we outline in the Future Work section: making the edit button more visible and increasing the number of people who receive notifications for page edits.

8

Future Work

Good design is about iteration. Learning from one's mistakes and continually trying out new approaches is far more effective than any amount of planning or prior calculation. This section suggests a number of changes to be made for future iterations of Sociapedia. We divide suggestions into four categories

Suggestions to Increase Page Edits

Comments

We have to be very careful here. If we ask people to write comments, they may stop editing the original article altogether. We also do not know if comments contain the same variety of information as Sociapedia articles.

Nevertheless, they are definitely worth exploring further. One approach we could take would be to add a second edit box to the bottom of the editing interface where users could write a personal comment. This box would almost be like a separate page for each user: it would retain its content and could be later edited by that user. Article subjects would be allowed to delete (but not edit) comments. Sociapedia would create a new *Comments from Friends* section at the end of every page and append each user's comment with signature to this section.
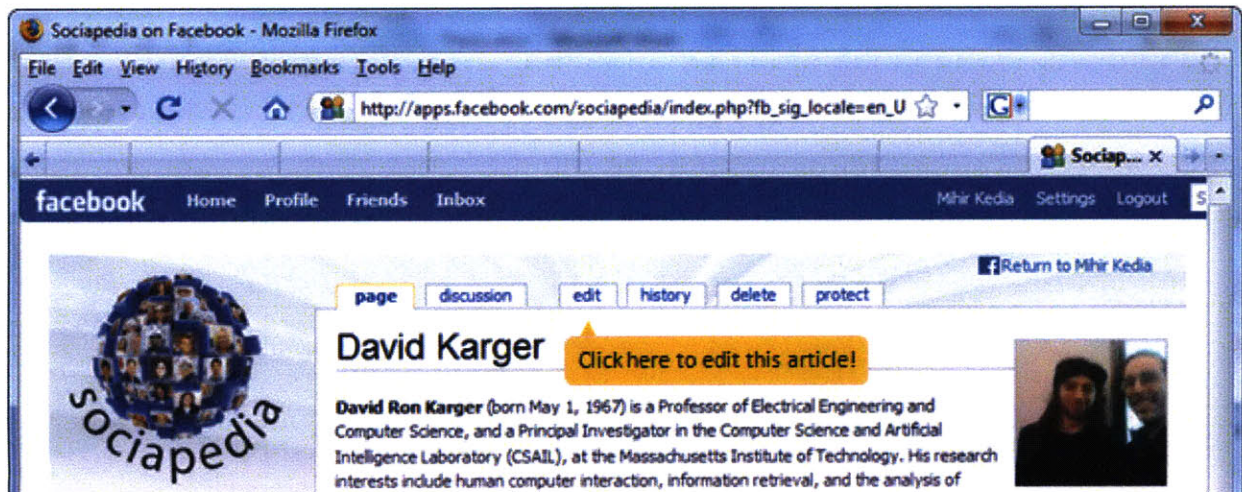
Figure 12: An example of how to make the edit button more visible while still maintaining the Wikipedia interface.

We also had a more unusual idea: Specifically ask users to write about the relationship between them and the article subject, and then add the content to *both* the subject's page and the user's page. Since we are asking for information about the relationship between the two users, there is nothing wrong or weird about placing it on the editor's page. In this model, both users could edit this content blurb at will.

This latter idea gives us two nice benefits. First, every single person is able to make a unique contribution. Rather than being intimidated by the task of editing the existing article and adding in additional information, users are able to write something that only they can contribute. Second, it increases the amount of content people can read on Sociapedia, since this information is visible to both the editor's circle of friends and the article subject's circle.

Suggestions to Increase the Number of Editors

Expanded Notifications

Currently, users only get notified if a page they previously edited is changed. This means that somebody who never edits a page will never get a notification (beyond edits to their own page.) Our idea here is to send notifications to users whenever any of their friend's pages gets edited. Our hope is that this could increase the number of people who become editors.

In order to prevent accusations of spamming, we would allow users to configure how often they received notifications. Users could choose from "any friend's page", "any page I've previously edited", "any page I'm tagged on" or "just my own page." The default setting would be "any friend's page", though.

Visible Edit Button

We also want to make the edit button more prominent. Figure 12 shows one possible way of doing this: a brightly colored bubble that appears on certain pages and points to the edit button.

We want to show this bubble on pages that the user is most likely to edit, which is a similar problem to choosing users for the Suggestions box. Thus, we can modify the existing suggestions algorithm to decide when to show the bubble. We would show the bubble until either they make at least two or three edits or they fail to click on the edit link a large number of times, implying that they are choosing not to edit rather than simply not knowing how to edit.

Suggestions to Increase the Pages Created

Badge System
In many websites, people who are active participants get rewarded with a badge that demonstrates how active they have been. A similar system might encourage editors to create more pages. In particular, we want to give a badge to any editor who creates at least four new pages: as we mentioned in the Conclusions section, this would allow Sociapedia to spread throughout Facebook.

This system can help with the other goals as well (e.g. giving badges for users who edit enough pages.) It is most important for increasing the number of created pages, however, because we want to get to a particular threshold for each editor.

Other

Full-Text Searching
As mentioned in the Design section, the search box only allows you to find the names of other Sociapedia users. It would be nice to expand this and allow users to search the text of their friend's pages. Unfortunately, this is not trivial. Almost every search engine library in existence today assumes that all queries should be run over an identical repository of documents. In Sociapedia, however, users each have a unique set of pages to search over: the set of their friends who use the application. We were unable to find any existing search libraries for general use that could handle this in a scalable fashion, which means we would have to engineer our own solution. Thus, while we would like to introduce full-text searching into Sociapedia, this is not a feature that can be added anytime soon.

Recently Changed Box
When we realized that a search box was insufficient for exploring Sociapedia, we added a page that would allow users to see a list of every friend they had on Sociapedia. We were unable to develop this further due to time constraints. One thing we would like to do in the future is to move a lot of this functionality into a new *Recently Changed* box.

This box would display the most recently edited articles within your Facebook friend circle, as well as the name of the editor and the age of the edit. Users would be able to click on a link at the bottom of the box to see a more extensive list (or to navigate to the currently existing page.)

We would like to add this because we believe that a browsing interface is an important component that belongs on every page, rather than simply linking to one within the Links box. Showing users the most recently edited pages allows them to easily find content they have not yet consumed.

Bibliography

Agarwalla, Jayant and Rajat Agarwalla. Lexulous. 1 July 2009
<http://www.facebook.com/apps/application.php?id=3052170175>.

Alexa, Inc. wikipedia.org - Traffic Details from Alexa. 24 August 2009. 24 August 2009
<http://www.alexa.com/siteinfo/wikipedia.org>.

Bernstein, M., D. Tan and G. Smith. Collabio. 18 August 2009
<www.facebook.com/apps/application.php?id=20574313186>.

Bernstein, M., et al. "Collabio: A Game for Annotating People within Social Networks." To appear in UIST
2009. ACM Press. (2009).

Chainn, Inc. Compare People. 20 August 2009
<http://www.facebook.com/apps/application.php?id=2433486906>.

—. Testimonials. 10 August 2009 <http://www.facebook.com/apps/application.php?id=2588290420>.

Cialdini, Robert B. Influence: The Psychology of Persuasion. Collins; Revised edition, 1998.

Define Me. 20 August 2009 <http://www.facebook.com/apps/application.php?id=2446888439>.

Facebook. Press Release: Facebook Receives Investment from Digital Sky Technologies. 26 May 2009. 7
August 2009 <http://www.facebook.com/press/releases.php?p=103711>.

Gilbert, Eric and Karrie Karahalios. "Predicting Tie Strength With Social Media." CHI. Boston: ACM, 2009.

Giles, Jim. "Internet encyclopaedias go head to head." Nature (2005/12/15): 900-901.

Grimmett, G. R. and D. R. Stirzake. Probability and Random Processes, 2nd ed. Oxford: Clarendon Press,
1992.

Johnson, Bobbie. English Wikipedia hits three million articles. 17 August 2009. 18 August 2009
<http://www.guardian.co.uk/technology/blog/2009/aug/17/wikipedia-three-million>.

Kittur, A., et al. "He says, she says: conflict and coordination in Wikipedia." Proceedings of the SIGCHI
Conference on Human Factors in Computing Systems. San Jose: ACM, 2007. 453-462.

—. "Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie." alt.chi
(2007).

Lampe, C. A., N. Ellison and C. Steinfield. "A Familiar Face(book): Profile Elements as Signals in an Online
Social Network." In Proc. CHI (2007): 435-444.

Lynch, C.G. "The Fragile Facebook Economy: Developers Struggle As Rules Change." 28 January 2009.
CIO. 1 July 2009

<http://www.cio.com/article/478272/The_Fragile_Facebook_Economy_Developers_Struggle_As_Rules_ Change>.

MIT Haystack. stat.us. 1 July 2009 <http://www.facebook.com/apps/application.php?id=40915749365>.

Oxford University Press. Sentence Length. 29 June 2009. 20 August 2009 <http://www.askoxford.com/betterwriting/plainenglish/sentencelength/>.

Wasserman, Stanley and Katherine Faust. Social Network Analysis. Cambridge University Press, 1994.

Wikipedia Contributors. Joe The Plumber. 17 August 2009. 18 August 2009 <http://en.wikipedia.org/wiki/Joe_the_Plumber>.

Wikipedia. Statistics. 18 August 2009. 18 August 2009 <http://en.wikipedia.org/wiki/Special:Statistics>.

Yazeed. 20 August 2009 <http://www.facebook.com/applications/iDescribe/5331531411>.

—. iDescribe. 20 August 2009 <http://www.facebook.com/applications/iDescribe/5331531411>.