Photobase - A Research Platform to Investigate Peer Production and Collaborative Sensing Systems

Brian J. Taylor

May 27, 2008

Abstract

Collaborative peer production systems like Wikipedia and Flickr succeed (and fail) based upon the voluntary participation of the users of those systems. Individual user participation varies and is influenced by any number of different variables. Research in peer production systems has thus far analyzed them in a non-experimental fashion and has been unable to address why certain behaviors occur within a collaborative content generation system. A parallel to peer production systems as seen from a networking perspective is participatory collaborative sensing. Experiments on participatory sensing is concerned more with how to build and utilize such systems and misses looking at the data collection process itself. The work presented in this report introduces Photobase, a small peer production/collaborative sensing research platform that enables investigation into what influences and effects individual user participation in these systems. A preliminary experiment demonstrates that a two-fold improvement in participation is achievable when users view collaboration as a competition. We also demonstrate that users tend to participate up front and then slack off and when participating, users stay within their defined familiar boundaries.

1 Introduction

A peer production system is defined as any system where a community of individuals comes together with the intent to generate and share content amongst members of that community. Peer production systems are commonly ascribed as belonging to the recent trend of the internet dubbed as Web 2.0 by O'Reilly media [15] but it could be argued they've been around at least since Usenet.

A well known example of a peer production system is Wikipedia, a collaboration whose goal is the creation of a free and complete encyclopedia. Another example is Flickr, a photo sharing site where users upload photos, discuss and tag anyone's photos, favorite photos and join different subgroups based on subject matter. Other examples include blogs, community news sites and even the open source software community.

One quality that makes these systems interesting is that in general, the only incentive to participate is intrinsic to each individual. Some users get a sense of enjoyment from sharing, others may contribute to watch their work being used, and some contribute as a way to give back to the community through a sense of obligation from using the community in general. Even with a lack of any kind of financial or other direct incentive, there are many examples of peer production systems that can produce incredible amounts of high-quality content. Wikipedia contains 2 million articles in the English language alone and Flickr contains more than 2 billion images.

A parallel to peer production systems as seen from a networking perspective is participatory collaborative sensing networks. Collaborative sensing systems are networks of individual sensors working together to collect local information that is then processed and transformed into global information. The individuals in the sensing system could be coordinated or uncoordinated. Autonomous agents designed to gather data and interact with one another fit within this definition. So do individuals with sophisiticated mobile cell phones since it is common for the phones to come equipped with GPS receivers, wi-fi, bluetooth, radio communications, accelerometers, cameras, microphones, and touch screens.

1.1 Research on Peer Production and Participatory Sensing

Users in a participatory system actively generate content in the form of sensor data, and share it with the other agents of the system. Users in a peer production system produce creative or information content used by other users of that system. The distinction between sensor data and creative or informational content is trivial. In both systems agents generation data that other agents can consume or enhance. Since these two systems mirror each other, many of the research questions posed for peer production systems have a parallel in participatory sensing and vice versa.

Much of the research in peer production systems has thus far analyzed them in a non-experimental fashion and has been unable to address why certain behaviors occur within a collaborative content generation system. Many of the research papers are of a descriptive rather than explainable nature. There is much more research in collaborative sensing. Some of it is from the viewpoint of how to build and utilize such systems, though [10] appears to be pursuing the same opportunities we are.

In order to research these kinds of systems, we need a research platform that is small, controllable, repeatable and has something of value to offer to its participants. The platform also needs to be able to incorporate the application of treatments and control groups into the study to evaluate the effect of the treatment on the variables under investigation. Treatments and control groups would allow us to discern the relationship between causes and effects in the system. Establishing how aspects of these systems influence user participation would be of great value to our understanding and improvement of them.

To address the lack of experimentation, we present a research platform called Photobase. Photobase has been designed to enable investigation into what influences individual user participation in these systems. While preliminary, Photobase shows much promise. The results in section 5 show a two-fold improvement in the participation levels of individuals simply by presenting information to them in a competitive-like fashion. Observations on user behavior in Photobase suggest users perform some initial collaboration then slack off and that they tend to stay to areas they are already familiar.

The next section describes the Photobase implementation and its strengths for conducting research. Section 3 and 4 discuss the simulation and experimental design used to conduct a preliminary experiment. The results in Section 5 demonstrate our observations about participation in general and suggest competition encourages higher levels of content generation. Related work is found in section 6 with a discussion on future work in section 7.

2 Photobase

Photobase is a small photo-sharing peer production/participatory collaborative sensing system. It is an experimental system where users work together to create a comprehensive photo database. Part of the software for Photobase operates on mobile phones and part consists of the experiment website. Photobase was specifically designed to enable research in peer production/collaborative sensing systems. It permits experimental control over what users see and what capabilities they have on the phone and on the website.

Nokia graciously loaned to the project 22 N95 mobile cell phones. Participants are given one of the N95s with the Photobase software installed. With the phone they can take a photo which is tagged with the current GPS coordinates, date and time, and the userid. Photos are sent to the Photobase server which processes and stores the information. Through a webpage interface, participants can view the photos and their current rating. Ratings are from one (poor) to five (excellent) stars and photos are rated only through the website.

2.1 Photobase Design

Approximately three months went into the design and development of Photobase. The components of Photobase on the N95 were developed in Python. The website was developed using HTML, PHP, Javascript, and AJAX. The core Photobase DB on the webserver used the mySQL database management system. Photobase was designed as to appear like an easy-to-use Web 2.0 application and to make it simple for a user to take, rate, and share photos.

2.1.1 The Nokia N95 and Python Interpreter

The N95-1 is a higher-end Nokia smartphone that comes equipped with an integrated GPS, runs Symbian OS v9.2 with the 3rd Edition Symbian 60 (S60) user interface, and has 100MB of internal memory. The N95 supports WAP 2.0 protocols (HTTP and SSL) that run on TCP/IP protocols as well as support for GPRS and EGPRS data services. A version of the Python interpreter has been developed for the S60 OS called PyS60. PyS60 has much of what one would expect from Python plus it allows for mobile phone development with access to much of the phones core functionality.

The N95 contains two cameras, a high-resolution camera on the back with a resolution up to 5 megapixels and a low-resolution camera on the front. Photobase only uses the camera on the back and limits the photo quality to 65, 536 colors with a resolution of 1024x768 (less than a megapixel). This keeps the filesize to approximately 200KB which is acceptable for transfers over the data connection. Higher quality images are possible, but would unnecessarily slow down the data rate of Photobase.

2.1.2 Map Mode

When Photobase starts, it begins in Map Mode (see Figure 1). The Map Mode presents a bounded map of the UMass campus that includes most of the main campus and many of the dorms. Map Mode is a re-implementation of the basic Google Maps functionality using Google Map satellite tiles. This was done because no web browser tested on the N95 was capable of displaying Google Maps well or was capable of displaying the additional photo information required for the project. Google is developing its own map implementation as a stand-alone for smartphones, but this application currently has limited functionality and did not meet the research needs. Users can pan through the map and can zoom in and zoom out. Map Mode also displays the current signal strength of the wireless connectivity, the current battery power level, and the status of the GPS lock.

Photos that have been taken in Photobase are represented on the map as a marker. By panning to a marker and clicking on it, the photo taken at that location is displayed on the screen, along with the photo's current rating. Currently, photos cannot be rated on the phone interface; users must use the website to rate photos.

From Map Mode users have several options. They can ask the server to update their phone with the current set of photos. This is useful as they are taking photos and want to know when their photo has been uploaded. Other users might also be taking photos at the same time, and refreshing the markers allows users to see the most up to date set of photos. Even over a small

Figure 1: The Photobase Map Mode displaying photos on the UMass campus.

period of time, users can generate a large amount of photos, and the photo markers can obscure the map or make it difficult to tell photo locations apart. To alleviate this, users are given the ability to choose a subset of the photos they want to see. Photo views include all photos, the best, average, or worst rated photos, a set of 20 random photos, a set of the 20 nearest photos to their current GPS location, and the last 20 photos that were uploaded to the website. Users also have the option to switch from Map Mode to Camera Mode.

Photobase is always trying to acquire a GPS lock. Without a GPS lock, some of the Photobase functionality is unavailable. When there is no GPS lock users cannot center the map, cannot get the set of photos nearest to them and cannot enter Camera Mode. Since the GPS lock can be acquired and lost over time, the last known good GPS coordinates are always stored. This way, so long as there was at least one GPS lock, the phone has latitude and longitude coordinates to work with.

2.1.3 Camera Mode

The Camera Mode (Figure 2) allows users to take geotagged photos. When a user takes a photo, the photo is quickly displayed on the camera and then they are prompted as to if they want to save the photo or not. If they save the photo, the picture is saved to phone memory and Photobase will transmit the photo when it is convenient. Photos are tagged with the latitude and longitude coordinates, time and date, and userid. The server does some minor processing like creating thumbnails and adds the photo entry into the DB.

The camera settings are all set to default. This means the users cannot zoom, change the flash, change the photo quality, or set photo exposure. This was a design decision to make the Camera Mode easy to use and to eliminate photographic proficiencies any one user may have over another. With everyone using the same settings, photo taking is simply point and click. From Camera Mode, users can switch back to Map Mode.

2.1.4 Website

The Photobase website [1] requires user login to track who is using the site, what photos are available to them, and individual user ratings. Implementation of the website was done through a

[1]http://prisms.cs.umass.edu/ btaylor

Figure 2: The Photobase Camera Mode in action.

combination of HTML, PHP, Javascript, AJAX, and mySQL queries. The website was designed to be cross-browser compatible. Photos visible in a world are displayed on the left hand side of the browser frame. In the middle of the browser is a Google Maps API which contains markers at the locations of the photos. Users can click on the left hand photo or the marker to generate an info window in the map that scrolls to the location of the photo. A thumbnail of the image, a left rotation button, a right rotation button, and a 5-star rater that displays the current photo rating are shown in the info windows (see Figure 3).

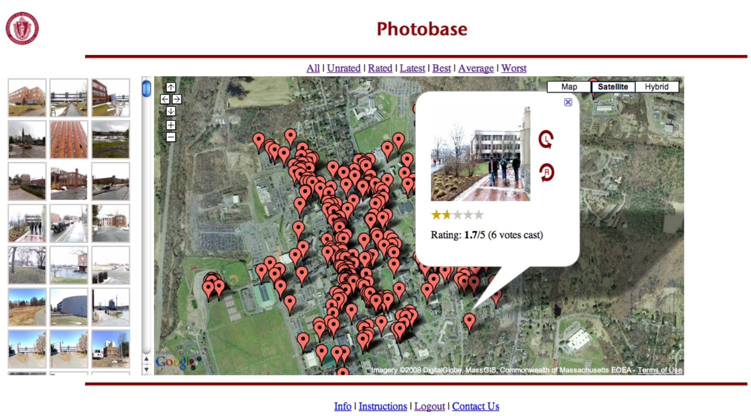


Figure 3: Example view of the Photobase Website.

By clicking on the thumbnail in an info window, the original full-sized image opens in a new popup window. If the user has not previously rated the photo then by moving the mouse over the stars, the stars light up indicating a value they can vote for (i.e., moving from left to right the first star lights, then the second and so on). If the user has already rated, then the star rater displays the current rating both in text and by coloring in the stars in gold. While user ratings are in whole numbers, the average rating can be fractional. The Camera Mode is implemented so that the camera is held in a landscape orientation. It is still possible for a user to hold the camera in portrait orientation, but the photos will be subsequently rotated. To correct for this, users can click on the thumbnail and use one of two rotation buttons to rotate the image in place. Any user may rotate any image, and users were asked to refrain from misusing the rotations.

As photos come into a world, it is quite likely that the markers will begin to obscure the map view. Users can filter the photos they view by choosing a subset of the available photos. Available options include the set of photos they have rated, the set of photos they have not rated, the last 20 photos to arrive, the best, worst, and average rated photos, and all photos in the world.

2.1.5 Data Recording

Data is recorded throughout Photobase, both on the phone and on the website. Data that is collected on the phone includes information on when Photobase is run (and by which user), how the Map Mode is used (which photos are being viewed), and on photos being taken (previously discussed in the Camera Mode section). On the website data is collected on who logs in and logs out, which photos a user views, how the user rates a photo, and if the user rotates a photo or not. Some data was recorded into a mySQL database while some was recorded to an experiment logfile.

2.2 Photobase as a Research Platform

While Photobase itself is an easy-to-use collaborative environment, its the control of what the users see and how they use the system that makes it interesting as a research platform. When the mobile phones query the webserver to obtain the set of current photos, we can choose which photos the user sees. When they take a photo and it is uploaded to the server, where that photo is stored and who has access to it can be assigned. Likewise, when they log on to the website, the photos they can see, the photos they can rate, and if they see additional website functionality like a leaderboard are also controlled.

2.2.1 Worlds

The Photobase research platform utilizes user views known as worlds. The idea of worlds comes from [14] and [11] where users were placed into one of N different worlds to study their behavior. (For further information, consult Section 6). The use of worlds is an excellent way to study how a treatment on a world affects it. In Photobase, individual participants can be placed into any world. Users, photos, photo ratings are all self-contained to the world. The only photos a user can see in the world are those either taken by them or someone else in the same world. Users may also only rate photos available to them in their world. By using worlds, we can obtain very strong causal relationships.

2.2.2 Causal Inference

Causal inference allows us to identify dependent cause-effect realtionships between variables [8]. Causal inference and causal modeling differ from traditional statistical modeling which relies on demonstrating correlation between variables. In causality, the goal is to eliminate all other possible contributions to an effect through the use of independence relations. If variable A is shown to be statistically related to variable B, we may only be able to predict values of B based on values of A. But if a causal link is established that says variable A causes variable B, then we can manipulate A to obtain a desired effect in B.

One way to identify causal relationships is through experimentation where a variable is set to a specific value. Setting the variable is called applying a treatment. In medical studies a treatment may be a new drug; in social sciences it may the use of a special program or activity. The experiments also require a control group which receives no treatment. Control groups may receive placebos or simply not have the treatment applied at all. By observing the differences in a variable under study between the two groups, we can draw conclusions as to the relationship between the treatment and that variable.

In our worlds, participants are placed by themselves with simulated users without their knowledge. Each world contains the same set of simulated users who have a pre-defined behavior. By doing this, each world has the same initial condition and, absent user activity, the same expectation of a final state. The combination of a treatment (or absence of a treatment) along with the users provide the differences between the worlds. This allows us to draw strong conclusions as to the effect of a treatment since every world is identical up to the user and the treatment. World views are not found in other peer production systems, making Photobase unique and in a prime position to conduct research.

2.2.3 Possible Research Questions

There are several unanswered research questions for peer production/participatory sensing systems including how the level of participation in these systems change over time, how users recognize and deal with poor quality information, if such systems work best in an uncoordinated or coordinated fashion, and if behavior in these systems is predictable. Photobase was designed to investigate these questions.

Influencing User Participation

Since peer production and participatory systems succeed (and fail) based upon the quality of the content and the amount of participation from their users, one may wonder what are the right ingredients to make a system a success. Identifying the specific causes will be difficult given that individual user participation varies and is influenced by all sorts of different variables. Users may participate because they enjoy doing so, out of a sense of obligation to the community, for glory and ego, or because contribution exists within the framework of a game or competition.

We can define user participation from the two different perspectives. If participation is defined from the perspective of participatory sensing, we might look at the number of photos, the geographical area covered by the photos, and ways to influence or maximize photo coverage of campus. From a peer production perspective, we might be interested in predicting future behavior based on past behavior and determining how participation levels change over time (i.e., do users of these systems participate a lot initially then slack off over time?). We may even want to know if users mimic each other, which if true, may add reinforcement of a phenomenon to a database but at the cost of minimizing the number of independent data points.

Larger peer production systems with tens of thousands or more members can certainly be analyzed to determine user behavior over time. But very little can practically be done to experiment with these systems to determine the exact mechanism that causes the behavior. Consider a large collaboration system where users could interact through message boards. If one tried to conduct an experiment and presented a different user interface to some small subset of users, chances are likely that these users would communicate the differences to the mass of other users, invalidating the experimental results. Even should an experiment be conducted, how can that system maintain a consistency for each user given that those users could access the website across different computers and different internet connections. Not all peer production systems require users to log in in order to use them, and it is very probable that some users would notice that the system behavior varies depending on how they access it. This too could invalidate the experiment.

Photobase as it is designed can be used to investigate influences on user participation with a greater ease than that of studying large, pre-existing systems. We can start with any number of worlds, initialized however we want, and comprised of any number of users. Participants know there is research being done, but they do not know the specific details of it and we should be able to expect them to behave naturally. Photobase requires user accounts, so we can control the user experience no matter how they access the website. Finally, since we're interested in the research and not actually building a hugely popular website, we can eliminate features that could potentially allow users to know if they were in a treatment or control group.

Handling Poor Quality Information

An important question for peer production systems is how they handle poor quality information, where poor quality could be one of vandalism, inaccurate or missing information. To study how users react to poor information, we might try injecting poor information into the system and studying their behavior. Do participants recognize poor information and if so, do they try to correct for it? How long does it take for poor quality information to be found? Does it require a consensus before this information is corrected or removed or do users take individual responsibility and act accordingly?

For most peer production systems, users would be very upset to learn that poor information was deliberately injected into the system. People have come to rely, at least for initial knowledge, on Wikipedia. If articles were tampered with to study user perception, it may damage Wikipedia's popularity, thereby influencing user behavior. We could not then answer the questions on poor information because we would not know if user reaction was due to the poor information in front of them or because they now call into question Wikipedia's reputation.

A possible solution, though not a preferable one, would be to seed new articles into Wikipedia and observe how long it takes for anyone to notice they are inaccurate and correct them. But how can we define what makes a good injection of "poor information"? If no article previously existed for what we enter, could it then be that it wasn't important enough to be of use to anyone, so the likelihood of someone catching the poor information is small. Or maybe the article we create is not well known, so people accept the incorrect content. We may not be able to distinguish between these two cases.

Photobase does not have these problems, mainly because it is not a well-established or large collaboration. Small size is an advantage. Since Photobase isn't well-established, users have limited prior expectations. Each world can be allowed to generate content such that it is generally consistent between all worlds. Then specific worlds can be selected for injection of poor information. Examples include uploading deliberately bad photos, placing photos in the wrong geospatial coordinates, inserting blank photos, photos with improper rotations, or removing photos of a certain area to create missing data. How the treatment worlds behave then sheds light on how users perceive and deal with poor quality information in general.

Social Network Structure Influence

If Photobsae were to be extended to incorporate social network capabilities, it could be used to study how social network structure influences the level of participation and the kinds of user participation. We might address questions such as "do users coordinate to maximize the UMass campus coverage?" and "does your social network influence your photo ratings?" It might be that social networks handle poor information differently. Could it be that users who can communicate identify poor information faster or might there be a negative reinforcement because users assume the information is appropriate since no one else has said anything.

With pre-existing systems, there is generally one type of social network. There may be subgroups or cliques within the network structure, but those communities probably follow a small world structure. Modifying existing social structure would probably be met with resistance. Again, this is where the world views in Photobase are an advantage. Each world could be assigned a specific type of social network structure (small world, random, power law, etc.). Since the users would be known ahead of time in the world, there positions could be initialized within the social structure enabling investigation into the above questions.

User Roles

Analysis of peer production systems has identified that users can be classified into different classes. For example, users might be classified as one of heavy content creator, minor editor, or vandal. The latter two classifications are so common they are given the informal names of "gnomes" and "trolls." Getoor identifies that using link prediction to predict user classification in Wikipedia is a part of the grand challenges of knowledge discovery [9]. Answering this question in Photobase might be easier than answering it in WIkipedia simply because of the sheer size of Wikipedia.

3 Simulation

A simulation was implemented to facilitate the notion of independent worlds. Individual participants were placed into worlds comprised of ten users. The human subjects were told that they were being placed with nine other real users, but in fact were placed into a world with nine simulated users. Each simulated user executed instructions over time making it appear as if they were active, uploading new photos throughout a day, and rating those photos. The simulation also made it appear as if the simulated users were rating the real user's photos.

To ensure consistency between the worlds, each world's simulated users had to have a consistent behavior. This meant that each simulated user had to upload photos and rate photos exactly the same and at the same times. To create the set of photos simulated users would upload, photos were collected during initial system tests. To create the ratings that simulated users would apply to those photos, a set of independent raters was used. The final and perhaps most important part was to generate ratings on real users photos. Again, the independent raters was the solution.

3.1 Step 1 - Simulating Photo Uploads

As Photobase was being developed, it needed to go through several iterations of the mobile phone user interface and website testing. The feedback and user experiences was valuable to improving the software but it also offered the opportunity to collect photos that could be used by the simulated users. As with the actual experiment, data was recorded for each photo when it came into the system, which included the GPS coordinates, date and time the photo was taken. A collection of more than 300 photos was gathered during the interface tests and from these 125 were chosen to have good enough quality to serve as the photos for the simulated users.

Since the duration of an experiment was decided to be three days, the test photos were chosen randomly and assigned to one of three days, day 0, day 1, or day 2. Day 0 corresponded to the first day of the experiment, while days 1 and 2 corresponded to the second and third days. Then for each photo, a random simulated user was chosen to act as the uploader for the photo. Simulated users were not chose uniformly, instead, they had the following chance percentages:

By choosing non-uniformly between the simulated users it enhances the appearance that they are real; some uploaded more often, others less often. Finally, the set of simulated upload instructions was constructed by ordering the photos according to their day and then original upload time. The

Table 1: Percentage chance a simulated user would upload a photo.

Position	Percentage Chance
0	10
1	5
2	5
3	20
4	10
5	10
6	15
7	10
8	15

instruction included the original latitude and longitude where the photo was taken.

3.2 Step 2 - Simulating Photo Ratings

Since simulated users were uploading photos, a way was needed to have these simulated users also rate photos, both their own as well as the real users submissions. Using a random system where a simulated user would randomly rate a photo was not acceptable. Random ratings provide no way of tying a measurement of quality (subjective or objective) to the photos. Other methods were considered such as using normal distributions for each photo, with the mean of the distribution reflecting a photos average rating as accumulated during the interface tests. But this only accounts for photos that have already been rated (perhaps by using the ratings collected in the user interface tests) and lacks the ability to create rating distributions for real users as they submit new photos.

Another possibility was to use a mixture model which used a combination of ratings from the user interface testing as well as from real users as the experiments were underway. This approach was rejected for one simple reason: it introduced a dependence between worlds. The rating behavior of one user could affect how the mixture model generates ratings for photos in other user's worlds. As an example, if one user rated every photo as a 5, this may cause simulated photo ratings to be near 5 and thus sway other users to rate the same photos equally high.

Instead, the chosen approach was to ask for a set of independent volunteers whose sole function was to rate photos. These raters act as independent quality assessment, judging the photos from the same goal we gave to the worlds. As they had no connection to the real users, this approach maintained world independence and as the raters worked throughout the entire experiment, their judgments were consistent for all worlds.

The raters were gathered from the UMass campus. It was desired to obtain a representative sample of the campus, so raters were invited from campus staff, undergraduates, graduates, and faculty. People who were not a part of the UMass campus were also invited, including one who was familiar with the campus and one who had never seen the campus before. In total there were eight volunteers consisting of three staff, one undergraduate, one graduate, two non-campus affiliated individuals, and one computer science department faculty member.

Since it would make no sense to have photos receive ratings before they were uploaded, the ratings needed to occur after a photo was uploaded.

For three days the simulation uploaded photos according to the upload instructions alone. Every time a rating was applied to a photo by one of the raters, we recorded which rater it was, which photo, what the rating was, and the day and time the rating occurred. Not every rater was faithful, but most were. By the end of the third day, five raters had rated most of the photos, and this was deemed enough to construct the simulated ratings. The lack of complete photo ratings mirrors how ratings would occur in a real system. Users may contribute periodically and some users may not contribute at all.

As was done with the upload instructions, the photo ratings were ordered according to the day and time they were applied. Ratings on the first day were set to day 0, on the second day to day 1, and third day to day 2.

3.3 Step 3 - Merging the Simulated Uploads with the Simulated Ratings

The set of simulated upload commands was merged with the set of simulated ratings to create a single set of instructions, ordered by their day and time Timing consistency was checked to ensure no photo was rated before it was uploaded. Dependency checks ensured that all ratings were on photos that had been uploaded. Each instruction was numbered with the first instruction being numbered 0.

Since there were nine simulated users, a number from 0 to 8 was used to identify the simulated user. However, each world contained a copy of this set of simulated users. To make the simulated users seem unique, each one was assigned a random name in each world. A table in the DB recorded the world id, the simulated users number, and their random username in that world. As an example, the simulated user with number 0 might be known as user367 in world 1 and as user923 in world 2. When the simulation performs an activity for position 0 in worlds 1 and 2, it simply does a lookup in the DB for each world to identify the actual simulated user. The real user in world 1 sees user367 upload photo A on day 0 at 3:20pm while the real user in world 2 sees user923 upload photo A on day 0 at 3:20pm. The different random simulated usernames help to minimize the chance that two participants might discuss the experiment and piece together how it worked.

```
169, 1, 10:27:38, 08, RATING, EXP12063845766.jpg, 1
170, 1, 10:27:59, 08, RATING, EXP12061228880.jpg, 1
171, 1, 11:02:54, 08, UPLOAD, EXP12060253969.jpg, 42.395833931248, -72.533135508279
172, 1, 11:03:50, 05, UPLOAD, EXP12063714824.jpg, 42.395474012462, -72.532271837303
173, 1, 11:10:25, 08, RATING, EXP11998858898.jpg, 2
174, 1, 11:10:31, 08, RATING, EXP12024817925.jpg, 2
```

Figure 4: Example of simulation instructions.

Figure 4 is an example of the upload and rating instructions. The first six fields in order are the instruction counter, the day and time the instruction is to be executed, the simulated user executing the instruction, the instruction type, and the specific photo that is the object of the instruction. For upload instructions, the latitude and longitude are also provided. For rating instructions the last field is the rating itself.

3.4 Step 4 - Running the Simulation

For each world, the simulation maintained a world info file that recorded the last completed instruction and the start date of the world. By having worlds with different start dates, we could accept volunteers as they became available, rather than gathering a group and waiting until we had a sufficient number to begin. With this approach, the day designation of 0, 1, or 2 used in the simulated instruction provided a day difference from the start of the world. Assume world 1 starts on March 30 and world 2 starts on March 31 and that the first upload instruction occurs on day 0 at 11:15am. Then the simulation would perform the upload for world 1 on March 30 at 11:15am and increment the world 1 last instruction to 1. For world 2, the first instruction would be executed on March 31 at 11:15am and its world info file would be updated to reflect the last instruction was 1. By maintaining the last completed instruction, the simulation is recoverable from stopped executions.

3.5 Continued Participation from the Independent Raters

To emulate the simulated users rating the real users photos, the independent raters continued rating photos during all of the experiments. The website was modified to allow the raters to see photos across all worlds. To prevent them from seeing duplicates of photos that were being simulated in multiple worlds, the SQL queries returned results grouped by photo filename. Raters were given no indication to whom the real photos belonged to nor which world they came from. For the real users, they had no indication who was rating their photos. All they could see was that their photos were being rated and the number of votes each photo had.

If a rater had rated a test photo (now a simulated photo) during the initial ratings collection, they would not be able to re-rate that photo. However, if they had not previously rated the test/simulated photo, then it was available for them to rate. There are two special cases. First, is when a rater rates one of the test/simulated photos after it has been uploaded. This implies the same unrated photo may be present in multiple worlds. The website catches this and when the photo is rated, applies the rating to all copies of this photo. The second happens when a rater rates a test/simulated photo prior to it being uploaded into the worlds. When this happens, a rating is recorded for the original photo, but is never applied to the copies. This would appear as a missing rating, but is better than having a photo appear as a fresh new photo in a world with ratings already attached to it.

Given that the raters were active during the experiment, one might question why simulated rating instructions were even used. The strongest argument is that given the ratings applied to the photos, the simulation can modify when those ratings occur. This would allow the experiment to investigate how the timing of ratings can influence participants. Even though this aspect of the Photobase design was not used, it is available for later experiments. Another argument is that having the raters first go through the set of test/simulated photos actually acted as a calibration period for the raters. This gave them a chance to become familiar with the website and comfortable with using their own judgments on the photos. It would only have been confusing for them to be presented with the same set of 125 photos again once the experiments had started.

4 Experimental Design

Volunteers were solicited through a computer science undergraduate email list, through announcement on the computer science lounge display, through word-of-mouth, and through announcement in an undergraduate operating systems course (CMPSCI 377). There was no monetary incentive offered. 18 people expressed an interest in participation; of those 13 actually took a phone for an experiment session.

Since this experiment involves human subjects, prior approval for the experiment was obtained from the university Independent Review Board (IRB). It was determined that the risk to human subjects in this research was minimal. There was a small risk that users might do something dangerous while taking photos, but they were strongly encouraged to stay safe and not to trespass or do anything else illegal for the experiment. User privacy was protected as much as possible.

When picking up a phone, a copy of the user's student ID, their local phone number, and email address was collected. However, at no time was this information linked to the data collected; it's sole purpose was to identify who had experiment equipment and who had returned it. Upon safe return, their personal information was returned to them. Instead we gave to each user a generic username, and all data collected in the experiment was linked to this userid. Every tester, rater, and user in this study were given a consent form and asked to sign it before they could participate.

4.1 Worlds

As mentioned in Section 2.2.1, the experimental design used world views to study the effect of a treatment on participation. It was decided that the world duration, the time a user had to participate in a world, would be set to three days. This decision was partly due to time limitations as there was only a limited amount of time to perform the experiments. It was also observed early on during initial system tests that participants tend to perform a lot of work initially and then slack off. Individual participants were placed into worlds comprised of themselves with nine simulated users as previously explained. The simulation makes it appear as if the simulated users were actively uploading and rating photos.

Because worlds are treated as self-contained environments, the individuals in the world can be given different goals to drive the subject matter of the photos as well as how the photos are judged. For the initial experiment, a single goal was chosen across all worlds that would motivate the participants and focus them on a consistent subject matter.

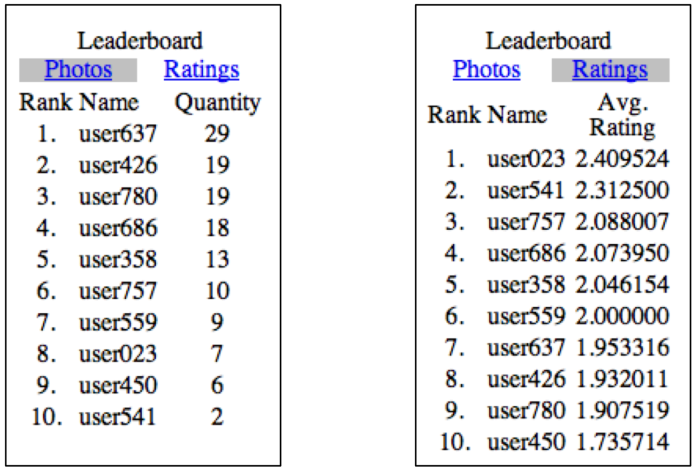
A blog dedicated to discussion of school financing recently posted the contents of an older article that identified the 20 ugliest campuses in North America [13]. The UMass campus came in second, beaten only by Drexel University. The claim was made that the buildings on campus were old, faded, lacked originality, and general made the campus look like an industrial park. Depending on a person's temperament, these can be taken as scathing attacks on the pride of the campus. This review in conjunction with some of the possible Photobase applications mentioned prior led to a clear choice for the goals of the participants: take photos that make the case that the UMass campus is more attractive than it has been given credit for. With this goal, the entire campus was available as subject matter with participants encouraged to take photos of buildings, landscapes, animal life, and general life on campus. They were cautioned to avoid photos which had nothing to do with the goal including photos off-campus and of friends hanging out.

4.2 Applying a Treatment

In addition to the general observations of how users would behave in this small peer production system, we applied a treatment to determine if it would have any effect on their participation. For this experiment, we defined as the treatment the availability of a leaderboard. The leaderboard displays two pieces of information for a world; one is the ranking of users by the number of photos they have uploaded, the other is the ranking of users by the average rating for their uploaded photos. When computing the average ranking, only those photos that have received a rating have been included in the calculation.

Worlds were divided into two categories: control and treatment. Users were randomly selected to be in one or the other. If the user is placed into a treatment world then they can see the leaderboard from the beginning of the experiment. If they are placed into a control group, then the leaderboard is absent as well as any indication, even in the website code, that a leaderboard exists.

The leaderboard introduces the appearance of a competition. As the simulated users upload



Leaderboard
Photos | Ratings

Rank	Name	Quantity
1.	user637	29
2.	user426	19
3.	user780	19
4.	user686	18
5.	user358	13
6.	user757	10
7.	user559	9
8.	user023	7
9.	user450	6
10.	user541	2

Leaderboard
Photos | Ratings

Rank	Name	Avg. Rating
1.	user023	2.409524
2.	user541	2.312500
3.	user757	2.088007
4.	user686	2.073950
5.	user358	2.046154
6.	user559	2.000000
7.	user637	1.953316
8.	user426	1.932011
9.	user780	1.907519
10.	user450	1.735714

Ranking by Photo Quantity   Ranking by Photo Rating

Figure 5: Example of the two views of the Leaderboard

photos, their positions will also change making it appear as if they are active. As the simulated ratings are applied, their rankings will also change on the board.

4.3 Initial Hypotheses

We had two hypotheses based on the leaderboard.

H_1: Users in worlds with a leaderboard will have a higher average number of photos than those without a leaderboard.
H_2: Users in worlds with a leaderboard will have a higher average photo rating than those without a leaderboard.

H_1 theorizes that by seeing their position on the leaderboard, users will want to avoid finishing in the bottom, and so will be more motivated to take photos to improve their ranking. H_2 theorizes that users will want to take better quality photos that prove the campus is attractive.

4.4 Experiments

Experiments were split between two waves with each wave consisting of two sessions. Each session lasted for three days. It was anticipated there would be 20 volunteers. Prior to handing out phones, worlds were randomly split into one of control or treatment. As participants arrived for phone pick up they were assigned into the worlds in sequential order, given a phone, and a generic website login. The phones internal identifier, the IMEI, was linked to each generic userid. When phones interact with the server, they send their IMEI with each communication. This way we could record which user was performing what action.

Each participant was informed that the experiments would start the following morning and that they could use the time up until then to practice taking photos and using the website. During their practice they would be assigned to a world consisting of solely themselves and that when the experiments started, they would be placed into a world with nine other users. In actuality, they were placed into a new world consisting of nine simulated users. During the practice, no user was given the treatment.

On the morning of the experiment, all practice photos and practice ratings were removed from the database. Simulated users were generated for each world. The DB was updated to reflect which users were receiving the treatment. A world info file was created for each world and the simulation script was activated. Users were prompted once when their experiment began, and once more on the second day to take photos and use the website.

The simulation script ran for the full three days. At the end of an experiment session, users were sent an email informing them that the experiment was complete. Their website and phone access was not immediately removed. Since both waves of experiments ran into the weekend, users were not asked to return the phones until the Monday after their experiments began. All data that came into the DB was time stamped so removing photos and ratings after the end of the session was easy. This extra data, even though not part of the official experiment, was still recorded in case it provided any interesting observations we hadn't anticipated.

5 Discussion of Results

18 people expressed an interest in participation; of those 13 actually took a phone for an experiment session. These volunteers were split into 7 control worlds and 6 treatment worlds. As is the case with peer production systems in general, the set of users consists of active users and lurkers. The active users took photos and used the website to rate photos in their world. The lurkers took few, if any, photos, but mainly participated through rating photos on the website.

Obviously, with the limited number of participants, we cannot argue that our observations lead to any valid statistical conclusions. In order to approach any kind of meaningful results, we would need to have a minimum of 30-40 worlds split evenly between control and treatment. Still, what we observed is interesting and can direct future experiments and improved hypotheses.

5.1 Influence of the Leaderboard

Our first hypothesis described how we would expect user behavior to be influenced by the leaderboard. Table 2 show the activity for users in the control board while Table 3 shows the activity for users in the treatment.

Table 2: The levels of activity for users in the control group.

User	Num. Photos	Photo Ranking	Avg. Rating	Rating Ranking
user035	0	10	-	10
user070	20	2	2.44	1
user082	0	10	-	10
user091	6	9	3.16	1
user204	13	5	2.21	3
user211	1	10	3.0	1
user226	13	5	3.13	1
Averages (all/active):	7.6 / 10.6	6.2 / 7.2	2.79	1.4 / 3.8

The average results are presented for two cases: when only the quantities from participants who took or rated a photo are included and when all participant activity is included. The analysis

Table 3: The levels of activity for users in the treatment group.

User	Num. Photos	Photo Ranking	Avg. Rating	Rating Ranking
user023	7	8	2.41	1
user044	0	10	-	10
user058	0	10	-	10
user066	1	10	2.0	7
user100	39	1	2.41	2
user117	40	1	2.8	1
Averages (all/active):	14.5 / 21.75	5 / 6.7	2.4	2.75 / 5

will only consider the latter situation, but being able to see the results when excluding the inactive participants is also enlightening.

First, note that users in the treatment group produced twice the number of photos. This is actually true whether you include the inactive users or not. This would suggest that users who saw collaboration as a competition were likely to participate more than those who did not. But these results are for a small sample size. To determine if the difference between the means of the two groups is significant, a permutation test was run. We ran 10000 permutations where we selected a set of control and treatment values from the original pooled values. This yields a significance of the original result as 0.18. This is suggestive but not a statistically valid result. We would need additional experiments to improve that number. At this time we cannot accept hypothesis H_1.

The second hypothesis, H_2 cannot be accepted either. Users in the control group produce higher average rated photos than those in the treatment group. A possibility for this might be that users, in general, have a low probability of taking photos seen as excellent by the raters. As users take more photos to improve their ranking, they succomb to the judgments of the raters which lowers their average photo rating. Two of the three users who took the lowest quantity of photos are in the control group and yet they have some of the highest ratings.

5.2 Rating and Website Usage

It is possible that how users utilize the website and make ratings is influenced by the leaderboard. The next two tables show some of the website and photo rating statistics. The % Self View identifies the percentage of the time that a user spends looking at their own photos computed from their total number of photo views. The # Website Refreshes represents the total number of times the user refreshed the photo markers. The Avg. Self Rating is the average rating the user gave to their own photos. The Avg. Rating of User by Independent Raters is the average rating given to the user's photos by the set of independent raters. Finally, the Avg. Rate of Simulated Users by User is the average rating given to photos uploaded by the simulated users into that users world.

It is clear that users in general rate their own photos higher than they do photos from other people (as seen by the average rating users apply to photos from the simulated users.) It's also clear that users rate their photos higher than the independent raters rate the user's photos. The leaderboard appears to have no effect on how frequent users view their own photos. In both groups, the average percent of the time users viewed their own photos was around 24%.

Table 4: The website activity for users in the control group.

User	% Self View	#Website Refreshes	Avg. Self Rating	Avg. Rating of User by Independent Raters	Avg. Rating of Simulated Users by User
user035	0	26	None	0	2.37
user070	24.2	7	4	2.41	2.23
user082	0	3	None	0.0	None
user091	12.5	8	4.5	3.10	2.10
user204	34.3	64	3.84	2.88	1.47
user211	0	13	None	3	None
user226	0	None	None	2.16	None
Average:	23.6	20.1	4.11	2.71	2.04

Table 5: The website activity for users in the treatment group.

User	% Self View	#Website Refreshes	Avg. Self Rating	Avg. Rating of User by Independent Raters	Avg. Rating of Simulated Users by User
user023	0	25	None	2.44	2.10
user044	0	4	None	0	1.8
user058	0	2	None	0	None
user066	3.64	75	5.0	1.5	2.85
user100	46.56	54	3.26	2.30	2.77
user117	0	1	None	2.86	None
Average:	25.1	23	4.13	2.27	2.38

5.3 User participation level over time

We noticed that during the initial user tests, users would take photos on the first day they had the phones, then seem to stop producing photos. To see how users behaved in the actual experiments, we plot their photo activity over time in Figures 6 and 7 for the control and treatment groups respectively. The x axis reflects the hours in the experiment while the y axis is the cumulative number of photos they have taken.

As can be seen, most users take photos in a small window of time and then never take another photo. There are three notable exceptions to this observation, user070 in the control group and user100 and user117 from the treatment group. We might be able to argue that the behavior of user100 and user117 is due to the leaderboard. Over time their ranking by photo quantity would change as the simulated users would climb higher in the rankings. Someone who was motivated to "win" would be inclined to take additional photos as their position changed, but they may be inclined to only take enough to raise their ranking. Again, our limited number of observations is not enough to draw any strong conclusions, but they do suggest new hypotheses for the future.

5.4 Which locations are popular to photo?

There are two ways we can analyze the locations of the photos: where were photos taken and what subjects were common in the photos.

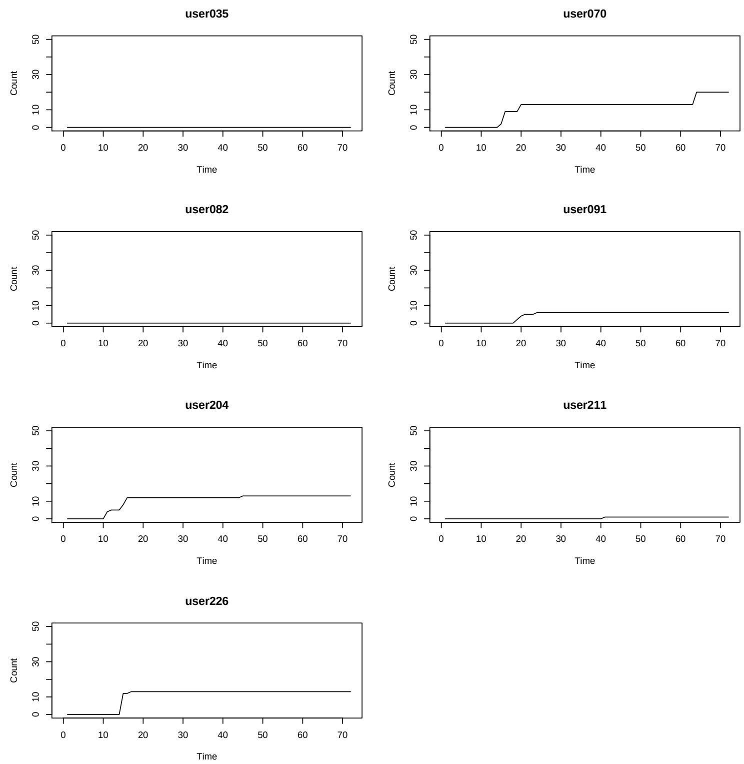


Figure 6: Photo quantity over time (Control Group)

The participants appear to have stayed to a core area of campus, not venturing to the outer areas that may not have been accessible by foot. Figure 8 shows the layout of photos. The major areas which were ignored or underrepresented are shown with white boxes and a numeric label. The locations are explained in the table below.

Very little can probably be inferred as to if the participants viewed these areas as attractive. Users were allowed to take photos using whatever means of travel they could, but it appears that most of the photos were taken on foot, suggesting that the areas which were photographed were those which were reachable by walking. The Hadley Equestrian Farm is on the outskirts of campus and would probably not be accessed by walking. The same can be said for the Waste Management and University Admissions. These are generally inaccessible and less frequently travelled, so they were probably ignored for their location.

The lack of a large number of photos in the Southwest Residential area could be due to the sample of the UMass student body population in our study. We may have just lacked students who lived there and so had no reason to visit those dormatories. That there were no photos in the athletic part of campus is suprising, especially since the Mullins Center is a relatively new and modern looking building housing both basketball and hockey. The large size of the athletic fields is also another feature which may make the UMass campus special, but these did not show up in the photos. The Studio Arts Building is a recently completed building, just finished in late 2007, and having some interesting archiecture such as an angled roof and a large glass front. Yet for being a recent campus addition, it was never photographed.

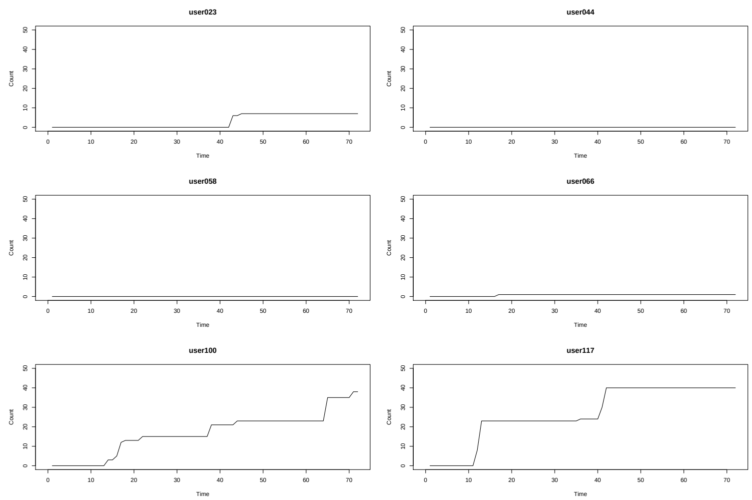


Figure 7: Photo quantity over time (Treatment Group)

Figure 8: Layout of Photo Locations on the UMass Campus

The other observation was that certain buildings, areas, or things would show up in almost every world. Since the users were unaware of each other's photos, the only photos they saw were the ones being uploaded by the simulated users. The tallest building, the W.E.B. Du Bois library is also located in the center of campus. It is not surprising then that it showed up in six of the nine worlds where users took photos. Other commonly photographed subjects included dormatories (in six worlds), one of the most recognizable buildings on campus, the Old Chapel (in four worlds), the Campus Center (in four worlds), and finally ducks (in three worlds). Of note is the Campus Center which is often derided as being the ugliest building on campus and the ducks whose recent return from winter to campus was only days before the beginning of the experiments.

By analyzing the contents of Photobase, we may be able to build predictive models as to the commonly traveled areas of campus as well as the areas the users of the system deem most attractive. Such information would benefit planning and development committees.

Table 6: Regions of Campus underrepresented in Photobase

Label	Location
1.	University Hadley Equestrian Farm
2.	Athletics Fields, McGuirk Football Stadium, Mullins Center, and other sporting complexes
3.	Southwest Residential Area
4.	International Programs Office, Research Administration, brand new Studio Arts Building
5.	Tilson Farm, Waste Management, University Admissions

6 Related Work

The idea of worlds comes from [14] and [11] where users were placed into one of N different worlds to study how music popularity might be influenced by existing popular opinion. The authors were researching the "cumulative advantage" effect where popular musical artists only become more popular over time and less popular artists never rise to fame. Their premise was that predicting which artist would become popular was generally a random process heavily influenced by users social interactions.

In order to study the cumulative advantage effect, users were presented with bands they had never heard of and were given the chance to listen, rate, and if they chose to, download music from the bands. Users were placed into one of two groups; group one only saw the names of songs and bands while group two had the names as well as how many times each song was downloaded. Group two was further split into eight worlds and each world started out exactly the same; all bands had zero downloads. By separating users into these worlds and having the worlds start off with the same initial conditions, the worlds were independent of each other save for the users. Results showed that among the eight worlds, no one band was consistently popular and when a band was popular, it was really much more popular than the alternatives.

6.1 Research in Peer Production Systems

There are several papers on the phenomenon of Wikipedia. [2] conducted a survey of Wikipedia users to better understand how their participation transformed over time. In general as user participation becomes more central and frequent, the individuals adopted new goals and took on more editorial roles. Their perception of and contribution to Wikipedia changes. It isn't clear how these observations apply to peer production systems in general (or even Wikipedia specifically) as the conclusions were drawn from informal evidence. [6] is another example of research on Wikipedia which analyzed the trust of the content being generated.

[1] provides a lengthy discussion on peer production systems and makes a strong argument for how they will revolutionize the economy of information. It presents a nice history of peer production systems from open source software to Wikipedia to massive online role playing games. But the approach of the text is philosophical in nature and relies heavily on description and not investigation of peer production.

6.2 Research in Participatory Sensing Systems

Collaborative sensing can be sub-divided into two approaches: opportunistic and participatory sensing. In opportunistic sensing, the agents in the system may not be aware that sensory data is being collected. Opportunistic sensing is automated and can be done in the background. The sensors in this approach wait until a set of requirements is met before collecting data, i.e. when a device's power levels are not low or.

Participatory sensing is collaborative sensing that involves the active participation of the agents in the system. When thinking of people and their cell phones, the people must use and interact with their phones. They must be the ones making the decision to collect and transmit the data. Without active involvement by the agents in the system, data collection would not happen.

[3] introduces the concept of participatory sensing and identifies many different participatory sensing applications including a geotagging photodocumentary that uses an already-deployed network of mobile devices. The authors argue that participatory sensing campaigns could sprout from citizen concerns yielding bottom-up, grassroots sensing to combat concerns like traffic congestion and pollution. User roles are defined for the network-assisted goal-oriented data gathering which align well with roles found in peer production systems.

[5] evaluated the opportunistic vs. the participatory approach to collaborative sensing in an example of using cell phones to take photos to provide panoramic street-views of major cities. Their results indicated that so long as there were enough people, the participatory approach would outperform the opportunistic but that decreasing interest in a sensing project would lead to opportunistic approaches being the best. They argued that systems that prompted users to take photos would eventually annoy the users and they would be less inclined over time to participate. But this assumption did not take into account incentives or other factors which could be designed into the system to influence participation.

[10] looks to come closest to our work. As part of a participatory sensing project to collect general data from mobile phones, the authors plan to use incentives in the form of credits to encourage participation and higher quality data gathering. Their results have not been published but it looks to focus purely on incentivization.

6.3 Research with Mobile Phone Networks

Mobile phones continue to become increasingly complex with the types of sensors they carry. It is now common for cell phones to come standard with GPS receivers, wi-fi, bluetooth, radio communications, accelerometers, cameras, microphones, and touch screens. After-market add-ons include hardware that will charge your cell phone with solar power. University researchers have investigated adding carbon monoxide and radiation detectors to cell phones. And phones like the Nokia 5140 come with built-in thermometers and compasses.

There are many projects investigating how mobile phones can be used as sensing devices. Most notably Nokia has supported several research projects and even has a website dedicated to the topic called SensorPlanet [12]. The Reality Mining Project from MIT [7] makes use of Nokia mobile phones to investigate how social networks evolve over time. Project goals included predicting user behavior, modeling conversational context, and inferring social network structure [4].

7 Conclusion and Future Work

Photobase has been presented as a research platform capable of investigating peer production and participatory collaborative sensing systems in ways that no other research has considered. The design of the simulation, use of worlds to control the views of users, and control that we can maintain of the phones and website allow us to draw causal relationships between system features and user behavior.

Our early experiments show that when users see the collaborative environment as a competition, they are incentivized to produce more content. In our case, users who saw a leaderboard produced twice the number of photos as users who did not. The results also suggest that users contribute during a small period of time and then slack off, but that those who see the leaderboard might be more inclined to participate for longer periods. We also demonstrate that users will rate their own contributions higher than they rate the contributions of others. A final observation was that users contribute from their familiar areas. In the case of Photobase, photos were taken in commonly accessible areas and areas which were harder to get to were greatly underrepresented.

Future work will look at improving the experimental protocols and opening the pool of potential volunteers to the UMass campus in general. 13 participants was a small number and we may need to consider financial incentives to encourage participation but not so much as to bias our results. There are additional questions we are interested in such as how participants react to poor information and if coordination affects participation levels and campus coverage.

8 Acknowledgements

There are several people who helped to make this project as successful as it was. These include Rich Hankins from Nokia who made possible the loan of the Nokia N95 mobile phones and a grant from Nokia to help support the phone data plans. Rich also provided valuable assistance during the development of the Photobase software. I'd also like to give my appreciation for the independent raters and Photobase participants, who volunteered for nothing more than the good of science and the success of this research. My gratitude also to David Jensen and Mark Corner, my advisors on this project, whose guidance led me to doing better research.

References

[1] Benkler, Y. The wealth of networks: how social production transforms markets and freedom. 2006. Yale University Press.

[2] Bryant, S.L., A. Forte, A. Bruckman. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. Proceedings of the 2005 international ACM SIGGROUP conference on supporting group work. pp 1-10. 2005.

[3] Burke, J., D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, M.B. Srivastava. Participatory sensing. ACM SenSys Workshop on World-Sensor-Web. October 31, 2006.

[4] Eagle, N. and A. Pentland. Reality mining: sensing complex social systems. Personal and Ubiquitous Computing. 10(4): 255-268. 2006.

[5] Lane, N.D., S. Eisenman, M. Musolesi, E. Miluzzo, A. T. Campbell. Urban sensing systems: opportunistic or participatory? In Proc. of Ninth Workshop on Mobile Computing Systems and Application (HotMobile 2008), Silverado Resort, Napa Valley, CA, USA, Feb. 25-26 2008.

[6] Lih, A. Wikipedia as participatory journalism: reliable sources? metrics for evaluating collaborative media as a news resource. Nature. 2003.

[7] MIT Media Lab; Reality Mining. htp:// reality.media.mit.edu . Last Retrieved: April 10, 2008.

[8] Pearl, J. Causality: Models, Reasoning and Inference. Cambridge, 2000.

[9] Piatetsky-Shapiro, G., C. Djeraba, L. Getoor, R. Grossman, R. Feldman, M. Zaki. What are the grand challenges for data mining?: KDD-2006 panel report. ACM SIGKDD Explorations Newsletter. 8(2):70-77. 2006.

[10] Reddy, S. J. Burke, D. Estrin, M. Hansen, and M.B. Srivastava. A framework for data quality and feedback in participatory sensing. Conference on Embedded Network Sensor Systems. Proceedings of the 5th international conference on embedded network systems. pp. 417-418. November 2007.

[11] Salganik, M., P. Dodds, D. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. Science, 311: 854-856, 2006.

[12] Sensorplanet. http:// sensorplanet.org Last Retrieved: April 10, 2008.

[13] The 20 Ugliest College Campuses in the USA - A List of all Ugly Colleges. Last Retrieved April 8, 2008. http://www.thinkfinancial.com/blog/college-admissions/the-20-ugliest-college-campuses-in-the-usa-a-list-of-all-ugly-colleges. January 11, 2008.

[14] Watts, Duncan. Is Justin Timberlake a product of cumulative advantage? The New York Times, April 15, 2007.

[15] Web 2.0. Last Retrieved April 10, 2008 from Wikipedia. http://en.wikipedia.org/wiki/Web_2.0