IPRO 305 Final Report

Fall 2010

User-Generated Map Content

Faculty Advisors: Prof. James Burstein, Prof. Christopher Lam

Sponsor Advisers: Mr. Wei Lee, Mr. David Ko

Table of Contents

Executive Summary

The focus of IPRO 305 was to aid in NAVTEQ's user-generated map content vision. User generated map content is a type of data that cannot be created, realistically, by any one organization. Instead, local communities and organizations need to participate to create the map data. This project determined what is required for such "community groups" to participate and help NAVTEQ provide better map content to customers.

Our first objective was to identify and organize community groups. Each community group was identified by its specific opinions about what data they would like to see on a map.

Our second objective was to determine an incentive structure that would motivate community groups to provide the map data. Each group was given general information on an incentive and the time frame they had in order to earn the incentive. As the semester progressed, map data gathering was tracked and at the end of the experiment, the map data was processed using an analysis methodology developed by the IPRO 305 team.

In the beginning, five community groups were identified. However, given time constraints and economic issues only three of these groups were pursued. Once the community groups were identified and created, IPRO 305 team members were assigned to each community group as a liaison for the project.

As a liaison, IPRO 305 team member's purpose was to offer community groups training and guidance throughout the project. Originally, each group was to be formed with eight to ten group members. Each group member was to receive a Nokia Smartphone loaded with a mobile data collection application. However, due to unanticipated economic

issues, this could not be realized. Consequently, other means for users to enter data were created.

Owing to a small sample of people, the results of the experiment do not bear much statistical worth, although the IPRO 305 team did manage to get the community groups to gather data. However, the major aim of this semester was to design the map-data gathering experiment and conduct experiment trial run. In this regard, the success has been phenomenal. Each community member took an exit survey at the end of the data collection phase of the experiment. The main purpose of the survey was to get feedback on the experiment conducted by the IPRO 305. The results of these surveys clearly showed the willingness of community members to collect data so long as they had a easy to use and working data collection tool.

The action strategy of our experiment constantly evolved throughout the semester often requiring us to modify our initial plans. What was learned in this semester will prove valuable for future semesters' success. From the issues with the data collection tools to the size and amount of community groups, we managed to gather data that was unexpected from us. The user feedback from this semester will be a valuable resource to future semesters when deciding how to proceed with the comprehensive experiment. The data collection tool was not entirely bug free and stable before handing it to the users. This gave us an opportunity to devise our own plan of action and recommend alternative methods of data collection. We have already identified a handful of community groups to pursue next semester.

Purpose and Objectives

IPRO 305 was asked to investigate the various incentives that would motivate community groups to enter data into maps so that the maps become more informative, up-to-date and useful to the community in general. Our sponsor NAVTEQ, a subsidiary of Nokia, specializes in producing accurate maps which it sells to various firms such as Garmin and MapQuest. NAVTEQ recently started investigating the prospect of crowd-sourcing as a means of enhancing their pre-existing maps.

NAVTEQ has always valued accuracy of data content on their maps. For data collection, their team would travel and scan across the entire US for details of map locations to generate exhaustive and accurate maps. Unfortunately, this is an expensive and time consuming process. By the time the survey of the entire US is done, a lot of new changes take place at many locations rendering the information on NAVTEQ's maps obsolete. To address this issue, NAVTEQ decided to adopt the paradigm of crowdsourcing, i.e. map makers use crowd as a source of information: common users enter data into the maps to make them more accurate and more useful.

In furtherance to this goal, NAVTEQ is eager to know what incentives, monetary or otherwise, would best motivate various users to generate map data. This is where IPRO 305's role comes into play. We are tasked with helping NAVTEQ in following ways:

## *Formation of dedicated volunteer Community Groups*

We would form various dedicated community groups that would generate map data for NAVTEQ. These community groups would be a well-knit team, with their members sharing certain common passions. Thus the data they enter into maps would reflect the

Points of Interest (POI) of people who have similar interests. This will give NAVTEQ an edge over its competitors, because NAVTEQ can provide user specific map data to people indicating their Points of Interest. NAVTEQ does not have a direct way to liaise with such communities. We, in the long run, would help NAVTEQ create such community groups.

Community-generated map data is something that is not yet included in NAVTEQ's maps. Therefore, the overall purpose of this IPRO has been to locate community groups and persuade them to generate the data desired by NAVTEQ. The community groups should be from different areas of the metropolitan Chicago area in order to have a well-rounded representation. Also, the community groups should involve different age groups in order to understand the differences (and possibly similarities) of the data points entered by each age group. Also, different age groups may enter data more or less frequently and at different time intervals. For example, a younger age group, ages 18-25, may enter data more frequently from Friday to Sunday.

There are many different types of community groups NAVTEQ may be interested in. For example, one type of community group NAVTEQ may be interested in would possibly be a biking group. The biking group would consist of both males and females aged from 25-40 and would mostly enter their data during the early mornings on the weekends. The biking group would be capable of providing a variety of different types of data for NAVTEQ. This may consist of different bike trails around the area, the conditions of the bike trails, different biking routes that may be used to reach the same location, and bike parking locations, etc.

The main advantage of utilizing community groups is that they would be able to provide the most up-to-date data. Currently, NAVTEQ operates several vehicle-based mapping instruments that need to be driven through different areas in order to obtain the most recent and accurate map data. However, the use of such mapping instruments is rather expensive and may be reduced if community groups are implemented in data collection. Community groups would be able to provide more accurate and recent data simply because they are also the consumers of the map data. The members of the group typically visit many of the data points that they provide because it is part of the passion that they all share. The biking group would be able to indicate whether the conditions of a biking trails have changed or if a bike shop has posted a sale much faster than the current method that NAVTEQ practices.
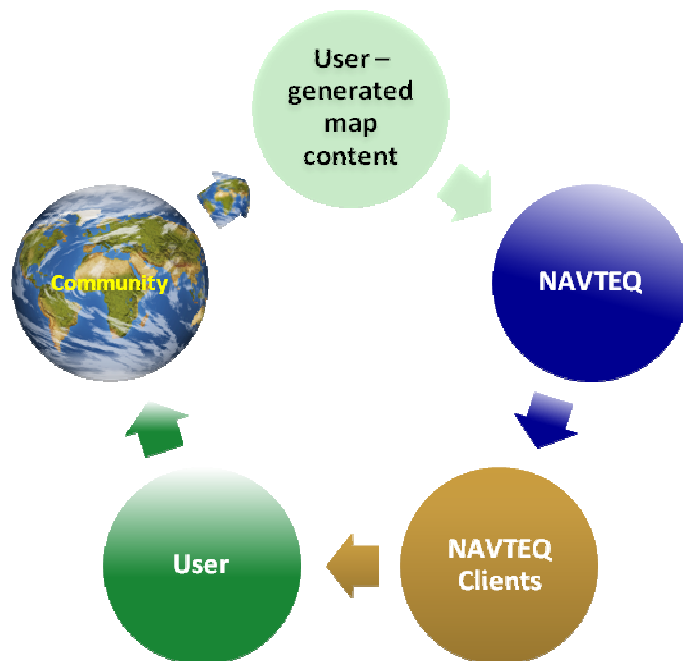
As shown in the figure above, the community enhances the maps by user-generated map content, which NAVTEQ provides to its clients, who in turn incorporate the same in their

end-products which are used by the users who are also part of one or more communities. This completes a cycle wherein, the map content generated by community is ultimately used by the community at large.

Investigation into the Efficacy of Various Incentives

We would investigate into what incentives would best impel a particular community to enter data into the maps. Thus for example, some community would respond well to a $20 'Best Buy' gift card while others might respond better to $20 rebate on Flight Tickets, while yet others would be happy to work just for the self-fulfillment. This study will help NAVTEQ invest its resources more wisely in crowdsourcing than its competitors.

In order to stimulate the community groups in participating in the data collection process, NAVTEQ designed a cell phone application titled "King of the Road". The purpose of developing the application is to provide a medium for all of the community groups to enter the map data. NAVTEQ provided 25 Nokia N97 Mini cell phones preloaded with the application to be used by various community group members. With this mobile device, a community group member would be able to enter map data which would then be uploaded to NAVTEQ map database. In this mobile application, each member receives 10 points per map data entered.

Although the concept of community groups collecting the map data is easy to visualize, there must be some sort of a driving force that compels the groups for the map data collection. Therefore, the IPRO team discussed several incentives that may persuade the community groups into participating in data collection:

- **No incentive:** The community group may participate in the map data collection for no gain at all. The group may be willing to participate for a variety of reasons, such as enjoying the competitive nature of 'King of the Road' mobile application or helping the local community.

- **Community-based incentive:** The group participates in collecting data with hopes of gaining something that would benefit their community. The prize may be related to the passion that the community group shares. For example, the biking group may compete in winning more bike racks throughout their community.

- **Monetary, individual incentive:** The community group would compete amongst themselves in order to win a single grand prize. The individual member with the greatest number of data entered would win a certain amount of monetary award.

These incentives were suggested by the IPRO 305 team as the most effective driving force for the community groups towards map data collection. All three types of incentives were incorporated into a statistical model that was developed by the team (Refer to the

APPENDIX 1: Sample Experiment Results). The statistical model used all three types in order to determine which incentive proved the most effective.

In conclusion, the overall purpose of this IPRO was to locate and motivate different community groups in generating map data. NAVTEQ hopes to apply the data gathered by these community groups into their maps in order to have the most accurate and varied information in their maps. A variety of community groups should be selected involving different age groups, locations, and passions. A cell phone application was developed in order to allow these community groups to enter their data. In order to motivate the community groups, different types of incentives were formed by the IPRO to drive the

community groups into collecting the data. These incentives were included in a statistical model to evaluate the effectiveness of each incentive.

Organization and Approach

Approach

In order to carry out the research, our IPRO team initially had planned to form 5 community groups of 8 to 10 members each. Our modus operandi was to organize a contest amongst the community groups and offer them various prizes as incentives and statistically determine what incentive impelled them the most.

In accordance with this plan, each of the community groups would undergo three stages:

1) Data Collection sans any incentives: The community group would be asked to enter data but would not know that there is a provision of any incentive, monitory or otherwise. This would serve as a control group experiment, which can be compared with various other experiments with different types of incentives to numerically analyze efficacy of various incentives.

2) Data Collection with collective incentive: The community group would compete with other community groups in data collection. Whichever team earns higher points will be given a group prize. The prize would be decided prior to the contest based upon inputs from the community groups.

3) Data Collection with individual incentive: The members within a community group would compete against each other. The winner would get an individual prize.

In furtherance to this end, our Sponsors had planned to provide the community group members with Nokia N97 Smartphones with GPS and Data Plan. The phones had an

application game namely "King of The Road". Whenever an individual would visit a place and would like to mark it in as a Point of Interest (POI), one would simply need to open the application "King of the Road", which would automatically mark the location on map and the individual would just need to enter name and specifics of the location. The game would also keep track of the user's score.

It was later realized that, due to financial constraints, our sponsors would be unable to provide us with enough number of phones. Also, there was a delay in acquisition of phones. IPRO 305 team therefore used this time to restructure the plans. As per the new plan, we limited the number of community groups to 3. After restructuring our community groups we initiated the data collection process and started working on inputs given by the groups.

Based upon inputs from the users, it was realized that the application "King of The Road" had many bugs and was not suitable for map data collection phase. We collaborated with the sponsors' software team to rectify the bugs, but it was found that the problems were more intricate than that the initial findings. After due deliberation, our team recommended that we use web based interface for data collection and validation. Upon this recommendation, our sponsors made available web-based map data collection tool to the community groups; where moderators from our IPRO team could track and verify the data entered by the community groups. However one community group was requested to use 'King of the Road' application to help find out its impact on the data collection phase.

The contest was thus continued and a substantial amount of data was collected to make at least some preliminary recommendations. The details of analysis are discussed in the

section Analysis and Findings. The long term goal of this project for future semesters is to generate some actual data for NAVTEQ and statistically find out what incentives best drive community users to generate map data. This will help NAVTEQ invest in crowdsourcing more wisely.

Organization

Since fall 2010 was the formative semester of the IPRO 305, we allowed a flexible structure for the team in the beginning. As the semester progressed, our goals evolved and we devised a defined modus operandi, the roles of individual team members became clearer and specialized.

In the initial stages, when we were soliciting various community groups, our IPRO team, we would have a group of members available to meet the prospective community groups. This was done on a rotating basis. Later, when the community groups were singled out, we assigned two members each for every community group. These two members would interact with and address issues of their respective community groups. Later, we also realized we need one or two dedicated members for communication with our sponsors and resolve technical issues. Our IPRO team also constituted dedicated members responsible for IPRO deliverables and some members responsible for data analysis.

Therefore the organization of our team was dynamic in nature evolving with the stages our IPRO progressed.

IPRO Team Structure

Following were the Individual based roles assigned to some team members for specific tasks

Team Leader: Responsible for preparing agenda for IPRO meetings, running the session and supervising other team members.

Minute Taker: Responsible for taking minutes of all IPRO meetings.

Following was structure of Sub-teams:

Team Leader

Community Group Liaisons	IPRO Deliverables Team	Data Analysis Team
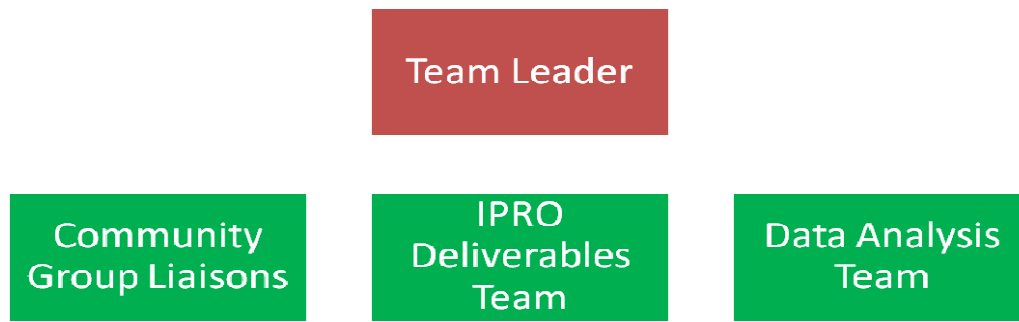
1. *IPRO Deliverables Team:* Responsible for ensuring IPRO department deliverables are ready on deadlines. It was responsible for posters, presentations and brochures required for IPRO day.

2. *Community Group Liaison Team:* Responsible for communicating with community groups, explaining them the experiment, addressing their issues and conducting the exercise of data collection.

3. *Data Analysis Team:* Responsible for statistical analysis model and the actual data analysis of the data generated by community users so as to give recommendations to NAVTEQ.

Following are the Community Groups which partook as volunteers in the IPRO 305:

Chicago Night Life: A group of young individuals interested in various restaurants, bars specials deals related to Chicago night life

Downtown Evanston: A group of professionals in Evanston interested in business locations, renters etc.

Northwestern University Football: A group of managers from Northwestern University Football interested in football field locations, various restaurants around and marking meeting places on maps.

There were two more groups that were ready to volunteer, but due to paucity of Nokia N97 phones, which were an indispensible part of our experiment initially, we couldn't accommodate them.

Analysis and Findings

This IPRO semester's goals included designing an experiment, conducting a trial run of that experiment, analyzing the experiments results, and providing guidance for following semesters to implement a full scale version of the experiment. This section describes the results of our attempts to complete each of these goals.

Experiment Design

The primary goal of this experiment is to determine how to best incentivize community groups to contribute map data to NAVTEQ. In this experiment, several community groups will be recruited to collect data. Each community group will be composed of 5-10 group members who share a common interest.

During the course of this experiment, community groups will be offered various incentives intended to motivate data collection. Data contributed by community groups will be recorded for the purposes of data analysis. Entry and exit questionnaires will be administered to allow for additional quantitative and qualitative analysis.

To address the question of incentives, our team divided potential incentives into three categories.

1. **Individual Incentive:** Incentives given to an individual group member based on that individual's performance in collecting data

2. **Group Incentive:** Incentives given to the entire group based on the group's performance in collecting data

3. **No Incentive:** The group is not informed of any potential for receiving incentives

In order to compare the effectiveness of different incentives, community groups will be offered multiple incentives over the course of this experiment. Each community group will participate in two test periods. During each test period, a community group will be informed of a single incentive which they may receive based on their map data collection performance.

This experiment will last for 6 weeks which will be divided into 4 phases as detailed below.

Phase	Title	Length	Description
Phase 1	Introduction	1 Week	Community members complete the entry questionnaire and are introduced to a data collection tool. The community groups are then given one week to familiarize themselves with the data collection tool and report any problems that occur.
Phase 2	Test Period 1	2 Weeks	At the beginning of this phase, each community group is informed of an incentive for collecting data during the next two weeks.
Phase 3	Test Period 2	2 Weeks	As Phase 2, but with different incentives.
Phase 4	Finalization	1 Week	Community members are asked to complete an exit questionnaire and participate in a feedback meeting, if possible. Prizes are distributed and any materials distributed as part of this experiment are collected by the IPRO team.

One complicating factor is that users may become more accustomed or become less interested to data collection as the experiment progresses. Such factors can bias results.

To control for this bias, the incentives offered in each test period will be staggered as seen below.

Group	No Incentive	Individual Incentive	Community Incentive
A	Test Period 1		Test Period 2
B	Test Period 2		Test Period 1
C		Test Period 1	Test Period 2
D		Test Period 2	Test Period 1

To analyze the results of this experiment, pooled t-tests will be used to evaluate the effectiveness of community-based incentives compared to lack of incentives based on the results from community groups A and B. Pooled t-tests will also be used to evaluate the effectiveness individual incentives compared to community-based incentives based on the results from community groups C and D. Information from the entry and exit questionnaires will be used to search for statistically significant correlations between survey responses (age, use of social media, etc.) and data collection.

Trial Experiment

One of this semester's goals, for IPRO 305, is to conduct a trial run of the data collection experiment, and identify potential difficulties. The experiment was implemented based on the experimental design presented above. Some adjustments were made to deal with present constraints. All adjustments made are noted in this section.

Several difficulties were encountered during this experiment. The data collection method desired by NAVTEQ was the smart-phone application "King of the Road" available only on the Nokia N97 smart-phone. NAVTEQ initially intended to supply the IPRO 305 team with 40 Nokia N97 phones to conduct the study. However, NAVTEQ was only

able to supply 25 phones during this semester. As a result the size of the trial experiment was reduced to three groups consisting of 10, 8, and 5 members respectively.

A second difficulty arose when group members attempted to use the "King of the Road" application. The application did not function as intended and achieved only minimal functionality by conclusion of the experiment. To circumvent this problem, two groups were allowed to collect data using an online submission form provided by NAVTEQ. A third group continued to use "King of the Road", and as a result they were unable to collect data. Further, difficulties with the data collection method caused several group members to drop out of the experiment, potentially biasing the results.

As mentioned earlier, three community groups participated in this experiment. For this experiment, the following incentives were offered

1. **Individual Incentive:** The group was informed that at the end of the test period, the member who contributed the most data will receive a $50 gift card to Amazon.com

2. **Group Incentive:** The group was informed that at the end of the test period, each member will receive a $25 gift card to a local restaurant if the group collected enough data to pass an unspecified threshold

Due to time constraints for the trial experiment, the experiment's schedule was adjusted. All groups were allowed an introduction period and finalization period as detailed in the experimental design. However, groups test periods were reorganized as follows.

Group	Individual Incentive	Community Incentive
DE		11/01/2010 - 11/14/2010
NU	11/01/2010 - 11/07/2010	11/08/2010 - 11/14/2010
NL	11/05/2010 - 11/14/2010	10/26/2010 - 11/04/2010

Data Analysis for Trial Experiment

After resolving complications with the data collection method, the group sizes were 8, 5, and 8 for the DE, NU, and NL groups respectively. Further, issues with the "King of the Road" application prevented the NL group from collecting data for the duration of the experiment. Among the DE and NU groups, data submission rates were highly variable. One member submitted 92 data points. Four members did not submit any data. Refer to Appendix 2 for further details.

The statistical methods described in the experimental design were implemented. Due to the small sample size, and varying conditions throughout the experiment, no statistically significant results were found with regards to the efficacy of various incentives.

However, for future semesters, a rigorous statistical analysis has been planned, the details of which are annexed in Appendix 1.

Conclusions and Recommendations

As explained in the above sections, during the fall 2010 semester, IPRO 305 established a framework for user-generated map content experiments and conducted a small-scale trial experiment. The team created a detailed experiment guide covering activities such as enrolling community groups, conducting test phases, and statistical data analysis. For the trial experiment, the IPRO 305 team was successful in enlisting three different community groups to participate in map data collection. Working with the corporate sponsor, NAVTEQ, necessary back-end tools were established to support the community-based data collection experiment. Throughout the semester, the IPRO 305 team was an effective conduit between the community groups and NAVTEQ. Although the trial experiment did not yield statistically significant results, NAVTEQ gained valuable feedback on their data collection tools from the community groups and the IPRO 305 team.

Going forward, we have constructed a set of recommendations for the future semesters. We would like to make following recommendation to future IPRO 305 teams:

- Verify that the data collection tools chosen for the experiment are bug-free

- Select data collection tools based on functionality and scalability

- Ensure that the data collection tools are user-friendly and sufficient training is provided to the community groups before the data collection phase

- Determine incentives for each community group based on the community group profiles

- During data analysis, take into account data collection tool problems encountered by the community groups during the experiment

- Determine data collection period for community groups based on their interest and the anticipated data volume. This could help target more community groups in one semester with shorter data collection periods.

The overall goal of our corporate sponsor NAVTEQ is to assess the crowd-sourced map data generation market. This will help NAVTEQ in determining how to use their resources effectively when they begin to use crowd-sourcing to generate their map data. In this semester, IPRO 305 team took the first step towards assessing the potential of the crowd-sourced data collection market. We believe that this ground work will help the IPRO 305 teams of the future semesters, to achieve higher success.

APPENDIX 1: Sample Experiment Results

The following is a sample thought experiment, wherein the data was fabricated by us and not obtained in an experiment. The purpose of this experiment is to illustrate the types of statistical analysis outlined in the Data Analysis section. Four community groups participated in this study. For this report, they are designated Groups A, B, C, and D. The following test schedule was used.

	N	Control (no incentive)	Individual Incentive	Community Incentive
	N	N=15	N=16	N=31
Group A	10	Test Period 1		Test Period 2
Group B	5	Test Period 1		Test Period 2
Group C	8		Test Period 2	Test Period 1
Group D	8		Test Period 1	Test Period 2

Groups C and D were given time-reversed roles in this experiment. The goal was to test how user input might be affected by the timing of each experimental period. In our results we saw a minor uptick in response rate during the second period of study, but the results did not show a statistically significant
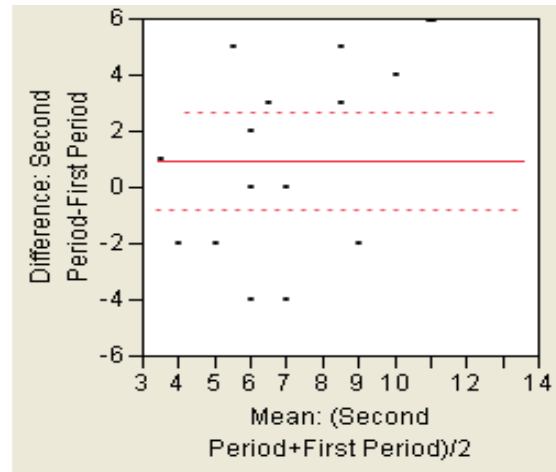

increase (alpha = 0.87). During the second period, members reported an average of 0.94 more POI's than in the first period. It should be noted that neither of these groups had a control period.

Groups A and B were both given a control period with no incentive during Test Period 1, and then a community-based incentive during Test Period 2. Members of these groups reported an average of 3 more POI's when given a community-based incentive than a when given no incentive.

Linear Fit
Linear Fit
Community-based = 3.0060729 + 1.0232794*Control
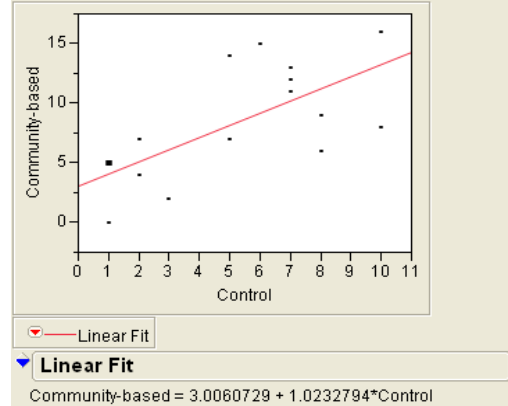
This trend was highly linear, as seen in above figure (alpha = 0.94). Group A reported an average improvement of 2.7, while Group B reported an average improvement of 3.6. Individually, Group A showed a statistically significant improvement (alpha = 0.98), but due to Group B's small size (n=5), they did not show a statistically significant improvement (alpha = 0.94). Taken together, we can say with 99% confidence that community incentives improved scores. Even if we assume that users naturally report 0.94 additional POI's in the second period and adjust our results accordingly, this result remains statistically significant (alpha = 0.95).

Groups C and D were each given one test period with individual incentives and one test
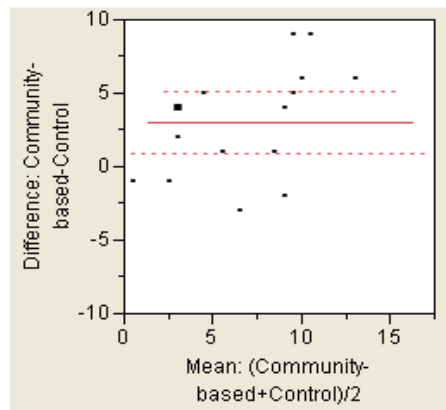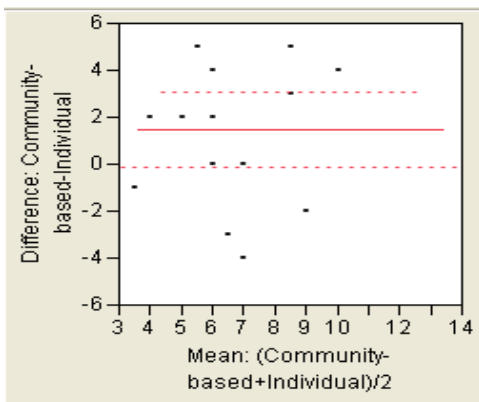
period with community-based incentives. Their test periods were staggered to control for changes in user behavior over the course of the study. On average, users reported 1.4

more POIs when offered community-based incentives than they did when offered individual-based incentives. Taken together, this result shows a statistically significant improvement when using community-based incentives over individual-based incentives (alpha = 0.96).

Since each group was offered community-based incentives during at least one test period, we will compare performance during the community-based incentive period among all group members to look for trends in user behavior. During this period, the group members (n=31) contributed an average of 8 POIs, with a 95% confidence interval of between 6.6 and 9.5 POI's for this mean. At the beginning and at the conclusion of the study, group members answered a questionnaire.

One pertinent question is whether they own a smart-phone. Overall, owning a smart phone did not have a statistically significant impact on how many POI's a group member contributed. Unfortunately, due to the small size of this study, no statistically significant results can be obtained by restricting our focus to only smart-phone owners.
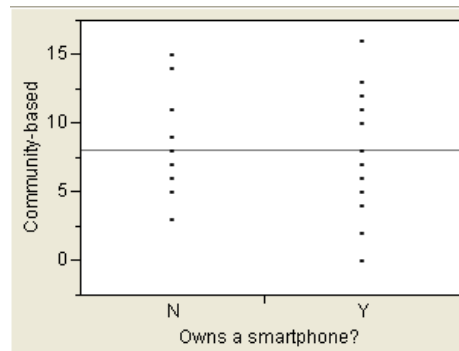
Group members were also asked to rate their enthusiasm about participating in the study on a scale of 1 to 5, with 5 being the most enthusiastic. All responses were in the range 3 to 5, and the response given had no statistically significant relation to the number of POI's contributed by group members.
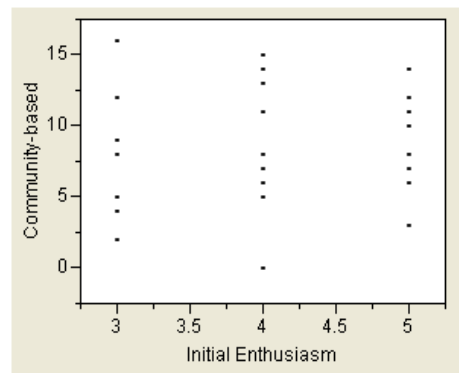
Group members were also asked to report their age. Age of the group member does seem to have an effect on the number of POI's contributed. The data appears to indicate decreasing activity as age increases, with a minimum near age 32 before when the trend reverses. However, this trend was not statistically significant (alpha = 0.94) due to the inclusion of only 10 group members over age 30. However, if we consider only group members under age 30, we do find a statistically significant trend of higher participation among younger group members (alpha = 0.98).
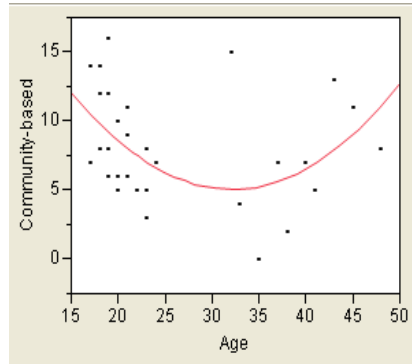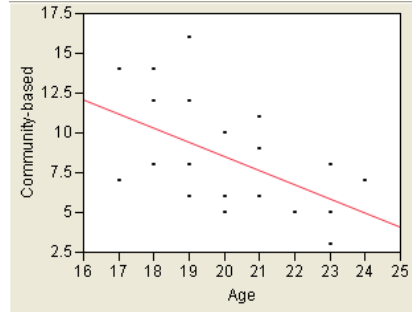
APPENDIX 2: Data Collected During the Experiment

DE = Downtown Evanston; NU = Northwestern University Football; CNL= Chicago Night Life.

Each row represents a member of the mentioned group and column represents a two week phase in the experiment. The numeric entries are the number of POI data entered.

Group	Introduction Phase	Community Incentive	Individual Incentive	Total Data Collected
DE	0	1	N/A	1
DE	0	92	N/A	92
DE	3	21	N/A	24
DE	0	8	N/A	8
DE	7	13	N/A	20
DE	0	5	N/A	5
DE	0	0	N/A	0
DE	0	0	N/A	0
NU	2	0	0	2
NU	3	0	1	4
NU	0	0	0	0
NU	0	0	0	0
NU	0	0	0	0
NL	15	0	4	19
NL	1	0	0	1
NL	0	0	0	0
NL	0	0	0	0
NL	0	0	0	0
NL	0	0	0	0
NL	0	0	0	0
NL	0	0	0	0