PACIS Proceedings

PACIS 2001 Proceedings

Association for Information Systems *Year* 2001

Unveiling the Traffic Dynamics of A Virtual Community

Eugenia Huang
National Chengchi University

Unveiling the Traffic Dynamics of A Virtual Community*

Eugenia Y. K. Huang
Department of Management Information Systems
National Chengchi University
eugenia@nccu.edu.tw

Abstract

The research results of a quasi-experiment on an award-winning website are reported. A virtual community was built on this website to study its performance under the influence of two treatments. The first treatment was an administrative modification – a change of membership policy that restricted nonmembers' access to a popular area of the website. The second treatment was the addition of an advanced feature that helped cultivate a virtual community, using combined push and pull capabilities. Data of four traffic performance metrics were recorded: pageview, visitor count, daily registration, and average duration. The data points preceding the treatment were analyzed using time-series ARIMA model to predict the data points subsequent to the treatment. In order to illustrate the effect of the treatment, the actual data prior to the treatment, the actual data following the treatment, and the predicted data are plotted on the same graph. This allows a clear visual inspection of how the treatment affects the data trend. The results show that for the first treatment, pageview, visitor count, and daily registration, all increased considerably**.** However, the levels of pageview and visitor count dropped back down very fast, while daily registration stayed high. Conversely, the seemingly positive treatment did not result in an increase of average duration. Unlike other performance metrics, average duration decreased. For the second treatment, the levels of pageview, visitor count, and average duration not only increased, but also stayed high. However, the treatment did not tend to change daily registration, as the level did not change much.

Keywords：Website, Traffic, Virtual Community, Quasi-Experiment, Time-Series

1. Introduction

1.1. Research Background and Purpose

*This research is partially supported by the National Science Council of the Republic of China under the project number NSC-89-2416-H-004-090.

When you open up a store, the golden rule is “Location! Location! Location!” When you open up a website, the location may be relatively irrelevant; but you’re constantly reminded of “Traffic! Traffic! Traffic!” That is why many dot-com companies strive for traffic before they intend to make profits. The recent dot-com stock value plunge not only indicates that making profit online is intricate but also that the niceties of building and retaining traffic are yet to be fully explored. Since the emergence of commercial uses of the Internet, web organizers have attempted many ways to increase Website traffic. A handful of them were phenomenal and became well-touted cases (Seybold, 1998), yet many others failed and disappeared from the Internet landscape. Those that succeeded may have approached their objectives systematically and learned some reasonably sure ways to build traffic (Watson, 1998). Nevertheless, no matter how famous or publicized the success stories may be, it is most likely that the web organizers will keep the traffic-building particulars confidential in order to sustain leading positions. As for the so-called best practices, their essence of operation is not actually revealed along with the legendary success stories.

Hagel’s (1997, 1998, 1999) publications on virtual communities affirmed them as an effective way to build and retain traffic. With many Web organizers believing the same, organizations and Web sites are adding virtual community capability to their Web services. By 1998, six of the twenty heaviest trafficked Web sites were primarily community based (Hanson, 2000). The formation of virtual communities on the Internet has been part of the Internet’s culture since its inception some three decades ago, when the Internet was serving in closed research groups. However, e-commerce marketers have just started to pay close attention to this potentially powerful tool in building up the critical mass needed for successful e-commerce. The explosive growth of the Internet and the emergence of the World Wide Web have enriched the way people communicate. WWW has become the synonym of Internet because most people do not know, or have forgotten, how to surf the Internet in other ways. Therefore a virtual community is almost always based on a Website, and both researchers and practitioners are trying to discover the ways to attract and retain members in a virtual community, and in turn, to harness the opportunities for e-commerce. Although virtual communities are generally believed to be a valuable avenue to building traffic, it’s dynamics are not well known. Thus, this paper attempts to answer questions concerning website traffic through close observations of the development of a virtual community.

1.2. Website Traffic Performance Metrics

This research concerns the traffic performance of a virtual community website. With the advances of information technologies, the visitors always leave "traces" that can help web organizers assess the performance. In particular, it is possible to extract the website activity data from the web server's log files. As the first attempt of quasi-experiment on Website traffic, this paper focuses on four metrics extracted from the virtual community performance metrics proposed by Adler(199), Winett (1988), and Shafer (1999): pageview, visitor count, daily registration, and average duration. Pageview is the total number of pages viewed by the visitors in a day; visitor count is the number of unique visitors in a day; daily registration is the number of new registered members in a day; average duration is the average time visitors spend in the website for one visit. Average duration is the same as Adler's length/session.

2. Research Design

There exist principles, beliefs, and viewpoints about ways of tackling the problem of developing a successful virtual community. Yet, quantified accounts of the developmental process of virtual communities are scarce. Measures claimed to be effective in nurturing a virtual community are rarely supported by quantified data. The research results reported in this paper are the lessons learned through the experience gained in developing a virtual community. In a nine-month time frame, the growth of this virtual community was recorded and studied using a quasi- experimental research design. A quasi-experiment uses a setting that is far more realistic than in a laboratory experiment. An Interrupted Time-Series Design Model was chosen to facilitate data collection and analysis. The time that a treatment was applied was used as a demarcation point. The data collected before the treatment were used as the "comparison group" data, and the data collected after the treatment were used as the "experiment group" data. Participants were gathered naturally on the Internet without a sampling procedure. The data were taken at regular intervals to satisfy the required data property in time series analysis. Another property required is that relationships exist among data points, i.e. the data are related in time. The development of a virtual community is a process of time. Therefore it offers the ground to assume that the data points are related in time. The choice of a quasi-experimental design over a laboratory experimental design is obvious: a laboratory experimental design has too many controls that would render the study on virtual community dynamics invalid.

The treatments were applied after the process stabilized. The type of virtual community chosen was interest-oriented, i.e. members are attracted to the community and stay loyal due to some common interests. Out of the four types of virtual communities: Community of Transaction, Community of Interest, Community of

Fantasy, and Community of Relationship, Community of Interest is believed to create flow easier and faster than other types (Armstrong, 1996). This experimental website intended to attract young people who were interested in the newest trends in the areas of gourmet food, fashion, and fitness. This web theme was chosen after many brainstorming rounds. The possible ideas for a web theme were confined to the basic rationale that the Website developers are college students and that the students should not be asked to grow an unfamiliar interest, because growing a new interest and gaining in-depth knowledge in the interest area would take quite a long time. Considering this rationale, having a group of college students who are already interested in fashions develop a Website focusing on young people's fashion interests seems to be a sensible choice, because this choice of interest is quite general among young people. One limitation of this quasi-experimental design is that it cannot test the effect of the treatments combined in a different order, because the participants would have very likely been sensitized through the first round of experiment. However, in regard to a virtual community, the idea of a quantitative study based on a quasi-experimental design is unprecedented and of vital importance in the Internet and e-commerce development.

The domain name of this website was *fit.nccu.edu.tw*. Besides extensive editorial efforts in providing fresh contents in four categories every week, it also had many features that facilitated community interactivity. The four categories of contents were women's fitness, men's fitness, gourmet food, and street candid pictures.

The following push and pull communication tools were implemented to facilitate community interactivity:

Push tools:

(1) E-mails between Web masters and users, e-mails among members, with automatic reply capability.

(2) Weekly electronic newsletter sent out every Monday. Highlights of new contents comprise the weekly newsletter.

Pull tools:

(3) Four bulletin boards corresponding to four content categories with hot topic ranking shown on the first page of the Website.

(4) Chat rooms with scheduled hosts, and completed with private chat and automatic user sponsoring capability. System administrators can send system announcements to all the visitors in the chat room.

(5) Instant messaging with dedicated chat room capability: This is a combination of push and pull tools. This capability is introduced later as the second treatment.

The above communication facilities were visible to visitors. In the background, there were numerous functionalities that were only available to system administrators for

maintenance, enhancement, and data collection purposes. The variables that were measured and reported in this paper include the basic web metrics: pageview, visitor count, daily registration, and average duration.

Since this is a "life" study, only the treatments that were believed to have positive influence on the growth of virtual communities were considered. This is because it is more meaningful to identify positive influences than negative influences. In the mean time, if negative influences are applied and they actually pull down the virtual community performance significantly, then the quasi-experiment cannot go on, due to lack of data.

3. Treatments and Grouping of Data

The F.I.T.[1] virtual community used an escalated membership strategy to alleviate the veteran vs. newbie problem as described by Hanson (2000). Veterans are the expert users, and the newbies are the novice users. Newbies desire expert advice, while veterans may view newbies as an annoyance. Hanson stated, "This communication value asymmetry can kill an online discussion group." Escalated membership can help alleviate this problem and keep the discussions focused. In the F.I.T. virtual community, regular membership required limited information, while formal membership required more detailed personal information. To encourage visitors to become formal members, it was stated clearly that only formal members were allowed to participate in prize drawings, which were offered regularly. Within the nine months that F.I.T. was in operation, there were two treatments. The first was applied after about 40 days of the initial operating date. During the forty days, pageviews of each category were collected to determine which category was the most popular. Then a change of membership policy was applied to allow the access to the most popular category to formal members only. It turned out that the most popular category was the street candid picture section. The second treatment was the addition of a new subsection that was based on instant messaging capability. The new subsection allowed formal members to cultivate a stronger bond through posting personal profiles, pictures, and daily journals. If a member identified another member as having common interests and this other member was also online, then he or she could send messages instantly to the member or invite the person to chat. At the same time, the total number of visitors online and the login names of the members who were currently online were shown on a separate window. A simple click on a formal

[1] This Website received the following recognitions: 1) Excellence Award – R.O.C. Third College IS Project Competition, 2) Excellent Website Recommendation in the category of fashions – link through "285" Website – a Business Weekly Conglomerate sponsored Website.

member's login name in this window would connect the visitor to the new subsection showing the profile of the clicked member. However if the visitor was not a formal member, further browsing in the subsection was not permitted. In attempting to browse the subsection, the visitor would encounter a friendly invitation to register as a formal member.

The timeline in Figure 1 shows the dates when F.I.T. started and stopped operating, when the treatments were applied, and when service was interrupted. The reasons for service interruptions were network or power outages[2].

We extracted data from two time periods. The first period was from April 13 to June 3; the second period was from October 11 to December 13. The first treatment was applied during the first time period, the second, during the second time period. It was desirable for other data to be useful as well, however data from the initial growing period may not be reliable enough to conduct analyses, because the time series may not be stable yet. Also, unfortunately, from March 30 to April 8, the service was not entirely ideal. After the service resumed, we had to wait again for a few days for visitors to return. The time span from early June to mid September was very quiet as far as the Web traffic was concerned. Summer vacation had a surprising impact on all four basic measures. Though the service and content updates continued as before, the visitor counts, for example, fell to about fifty percent. The influence stretched out to at least ten days before and after summer vacation started and ended. Table 1 shows the grouping of data.

4. Data Collection and Analyses

In this paper, an initial analysis of *F.I.T.* experiment is reported. The data used to conduct time series analysis include pageview, visitor count, daily registration, and average duration. The state-of-the-art ARIMA (Auto Regressive Integrated Moving Average) model was drawn on in time series analysis. For time-series analysis, there is no definite requirement for data set length. A large sample assumption is not required, as usually required by many other analysis methods, which are derived based on the assumption of a normal distribution. The important guideline in time-series prediction is that the predicted data set length is not to be longer than the original data set. A series of the daily-recorded data were first divided into two groups according to the day the treatment was applied. Then the first group of data, i.e. the data collected before the treatment, was processed to obtain its ACF (Autocorrelation

[2] A painful lesson was learned that Tai Power, TAnet, and even the campus network are not reliable enough to support a round-the-clock service. Tai Power is the sole supplier of Taiwan's electricity. TAnet is the ISP of all educational organizations in Taiwan.

Function) and PACF (Partial Autocorrelation Function), in order to determine the parameters in ARIMA model. If it appeared that more than one set of parameters was suitable, then the residuals were processed to determine an optimal model. Once the parameters of ARIMA model were determined, the selected model was used to predict the data points into the future. The original, first group data were thus extended to the "future". However, we have the actual "future" data, which were the second group of data. The data in the first group and the second group, along with the predicted data were then plotted on the same graph. This allows a clear visual inspection; one can visually compare the actual and the predicted performance after the treatment. The predicted data represent the possible trend as if the treatment were never there. Thus, the comparison shows the effect of the treatment.

The time-series analysis package in SPSS (SPSS, 1994; Box, 1994) was used to process the data. Figures 2 and 3 are the ACF and PACT plots of the first group data in the first period (P1G1). Since its ACF is decreasing gradually, and its PACF is either decreasing gradually or rapidly, there are two candidate models, i.e. ARIMA(1,0,0) and ARIMA(1,0,1). Consequently, the ACF and PACF of the residuals of these two models need to be further examined to determine an optimal model. The resulting AIC's (Akaike Information Criteria) are as in Table 2.

Therefore the optimal model is ARIMA(1,0,0), for it has a smaller AIC. The ARIMA(1,0,0) model was then used to predict the data that coincide with the time period of P1G2. This series of predicted data, and the data in P1G1 and P1G2 are plotted in Figure 4. It illustrates that the treatment caused an abrupt increase in the pageview for a few days. Then the pageview decreases to about the same level as what was expected in the absence of the treatment, as can be seen from the time series prediction curve.

Following the same procedure, the performances in visitor count, daily registration, and the average duration are plotted in Figures 5, 6, 7. In Figure 5, the visitor counts behave the same as the pageview in Figure 4, i.e. a sharp increase followed by a sharp drop, in only a few days, to the same level as what was expected in the absence of the treatment. As for the number of daily registration in Figure 6, the rise on the treatment day was very dramatic. The change of membership policy was very effective; the daily registration increased ten-fold, and it did not drop as fast as the pageview or visitor count. Several weeks after the treatment, the daily registration was still considerably higher than what was expected from time series prediction. Interestingly and surprisingly, the average duration did not increase, as can be seen in Figure 7. In fact, the average duration became considerably shorter due to the treatment.

The results of the quasi-experiment in the second period are shown in Figures 8, 9, 10,

and 11. Evidently, the pageview, visitor count, and average duration went up considerably and stayed at higher levels. Only daily registration did not seem to be affected by the treatment.

Out of the eight figures of time-series prediction comparisons, the figures for daily registration, i.e. Figures 6 and 10 do not seem to have good time-series prediction results. The prediction curves in these two figures appear to be too "flat." However, further inspection on the sample data reveals one common trait: the curves of the original data contain many up-down zigzags, i.e. the processes are not stationary. Therefore, it is not unreasonable that the time-series predictions give rise to flat curves that are in the vicinity of the average level of the zigzags. Usually, when a process is not stationary, the sample data should be transformed until a stationary process is obtained. However various transformation methods were tried and none of them produced stationary processes. Perhaps the fast variations of the sample data caused the processes to deviate from a time-series. Further analysis of daily registrations using ANOVA and the nonparametric Kruskal-Wallis method was conducted. It was confirmed that the daily registrations before and after the treatment were indeed statistically significantly different. For the first treatment, the Kruskal-Wallis method was used because the sample data sets did not pass the Levene test. For the second treatment, ANOVA is used because the sample data sets did pass the Levene test. The reason that the data sets for other metrics were not subject to ANOVA or Kriskal-Wallis tests is that data from stationary processes do not satisfy the assumption of independence.

Table 3 summarizes the effects of the first and the second treatments on the four Web performance metrics.

5. Research Limitation

The strength of a quasi-experiment research design lies in its realistic setting, which is essential for studying website traffic performance. However, a quasi-experiment design offers no effective means to control external variables. Thus a general condition for the above analysis to be valid is that nothing else, which can affect the results, happens during the experiment, except the treatments. This is the main concern of a quasi-experiment design, and this type of problem is called the history effect. The explanation or exclusion of the history effect relies closely on the recollections on what happens. In this research, a supporting instrument was carefully devised to allow inspection on the possibility of history influence. During the process of this experiment, a journal was kept online to record any events that occur each day, even if they were remotely associated with the website activities. Every web master

was responsible for filling in the journal entries, and a designated person is in charge of their correctness and completeness. When analyzing the data, outliers were carefully examined and checked against the journal to spot potential "history" factors.

6. Conclusions

The above results give rise to the opportunities of some intriguing observations:

(1) Devising a clever strategy like our first treatment certainly accomplishes the direct purpose, i.e. successfully acquired large number of members in a very short time, especially in the first few days. A few days after the treatment, the effect was not as significant, though the curves did manage to stay at a somewhat higher level. As for the pageview and visitor count, the exciting moment was just like a shooting star – disappeared in a blink! The most unanticipated result was the decrease of the average duration that visitors stay in the Website. It was unanticipated, because the intuition suggests that a mechanism should either have a positive or negative influence across all performance metrics; the degree of influence may vary, depending on different metrics, but the qualitative influence should consistently be classified as either positive or negative. Thus, at first, it was thought that the decreasing of average duration was totally unexpected. However, there is a perfect explanation for this decrease: many new members are in lack of "stickiness" that is so desirable in virtual community. To be more precise, it is highly possible that they were "forced" to become members; otherwise, they will not be able to continue viewing new candid pictures. It does not take much time to view all the new candid pictures. Then, what happen after they are done viewing the candid pictures? We are hesitant to make this conjecture – perhaps they just leave, due to a lack of interest in other contents or services. Web organizers really need to ponder this finding, and ask themselves if it is worthwhile to attract many members who are not sticky and are not likely to become loyal. That is to say, it is wise to determine the value of visitors before committing resources, because it takes up one share of resource to serve one visitor, regardless of whether the visitor is valuable.

(2) Compared with the first treatment, instilling the second treatment required a lot more work beforehand. The first treatment is only a change in policy, while the second treatment takes considerable design and development efforts. In our case, toil and labor pays off: pageview, visitor count, and average duration, all increase considerably, and daily registration stays at about the same level. Though the immediate effects are not as striking as that of the first treatment, the curves all demonstrate long-lasting effects. It is very tempting to infer from this result that

deep investments on the push and pull tools, or the so-called content trees and communication rings of virtual communities will effectively help the virtual community grow steadily. The improvements on pageview, visitor count, and average duration we saw may not seem very significant. Yet, the results reported here are based on the measurements on all visitors. If only the data of the members are shown, the effect should be a lot more significant. Since the number of registrations is not affected by the second treatment, the positive effects from the second treatment are mainly the contributions of "old" members. Indeed, the second treatment is an offer of new functionality that benefits members only. If this group of "old" members, who deserve to be called loyal members, can be pinpointed, they will become the most valuable assets of the virtual community.

References

Adler, R. P., "Doing Research in Online Communities," Apr. 1999, available at http://www.digiplaces.com.

Box, George E. P., Jenkins, Gwilym M., and Reinsel, Gregory C., *Time Series Analysis – Forecasting and Control*, Prentice Hall, 3rd ed., 1994.

Armstrong, G. A. and Hagel, J. III, "The Real Value of On-Line Community," *Harvard Business Review*, May-June 1996, pp. 134-141.

Hagel, J. and Armstrong G**.** A**.** *Net Gain: Expanding Markets through Virtual Communities,* Harvard Business School Press, 1997.

Hagel, J. "Collaborate and Conquer," *Information Week,* September 14, 1998, pp. 275-277.

Hagel, J. and Singer, M., *Net Worth: Shaping Markets When Customers Make The Rules,* Harvard Business School Press, 1999.

Hanson, W., *Principles of Internet Marketing,* South-Western College Publishing, 2000.

Seybold , Patricia B. and Marshak, Ronni T., *Customers.com*, Random House, 1998.

SPSS Trends 6.1, SPSS Inc., 1994.

Watson, Richard T., Akselsen, Sigmund, and Pitt, Leyland F., "Attractors: Building Mountains in the Flat Landscape of the World Wide Web," *California Management Review*, Vol. 40, No. 2, Winter 1998, pp. 36-56.

Winett, B., "Tracking Your Visitors," April 1998, available at http://hotwired.lycos.com /webmonkey/98/16.

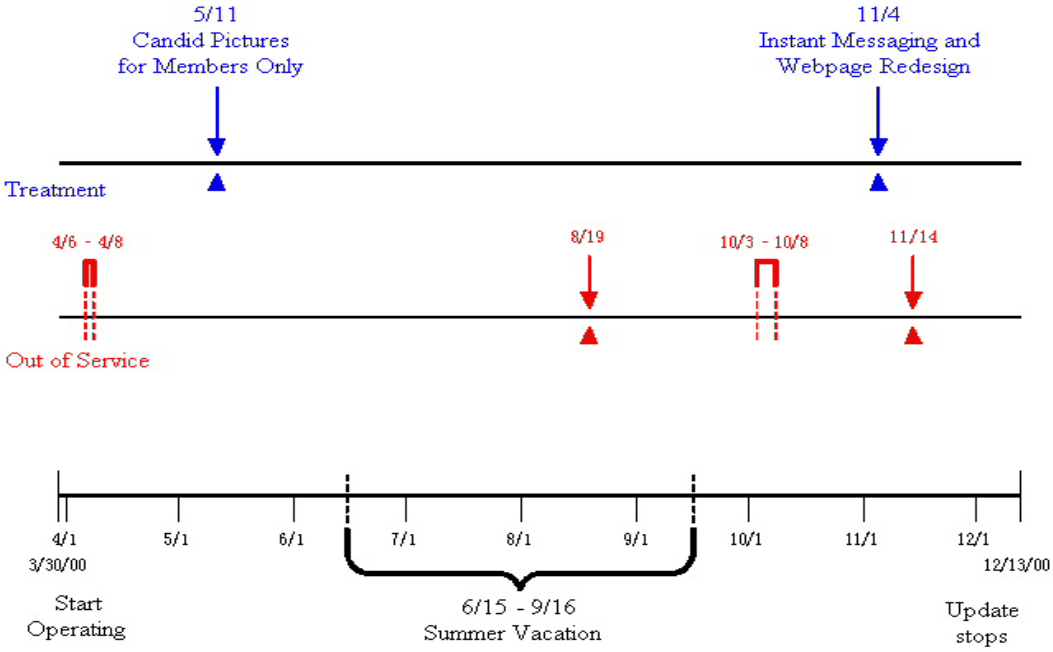

Figure 1. Timeline

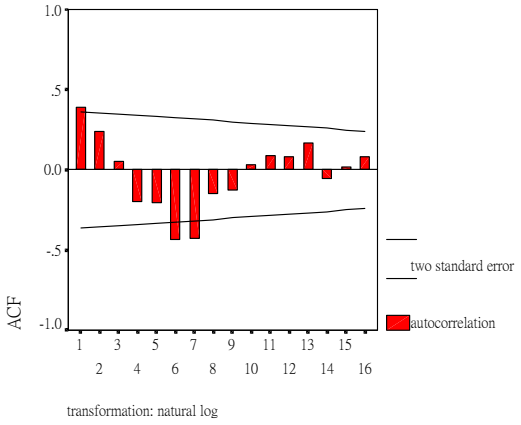

Figure 2. ACF of pageview.

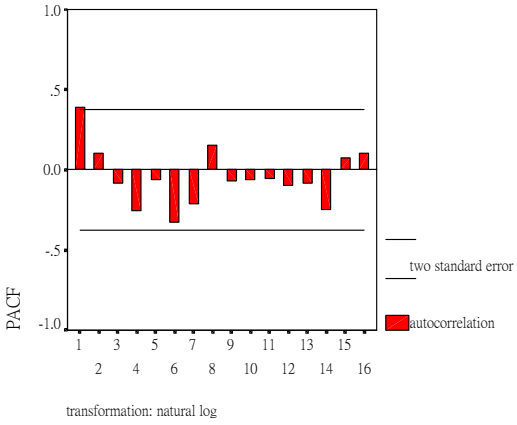

Figure 3. PACF of pageview.

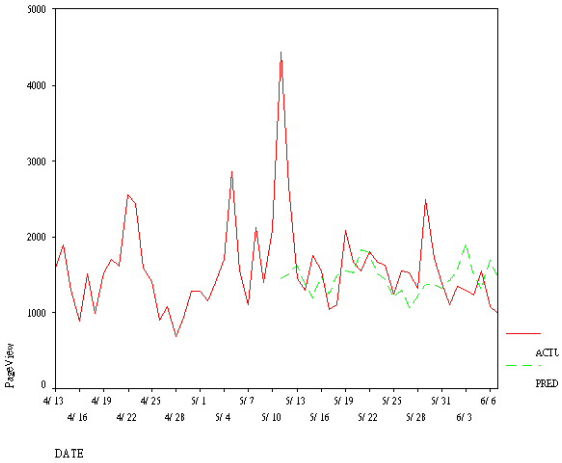

Figure 4. Pageview – first period.

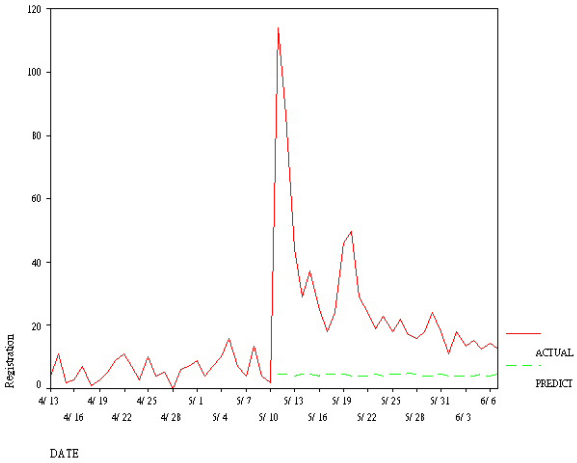

Figure 6. Daily registration – first period.

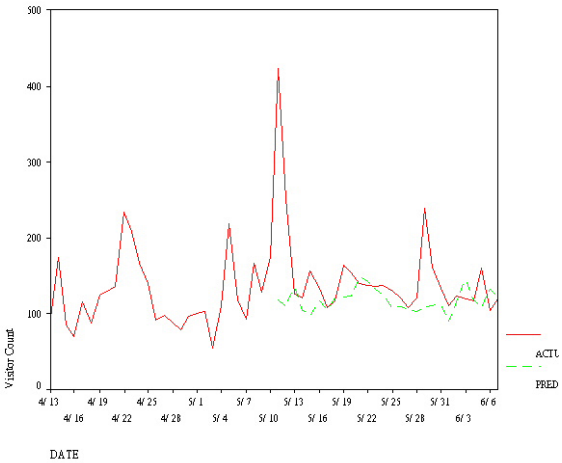

Figure 5. Visitor count – first period.

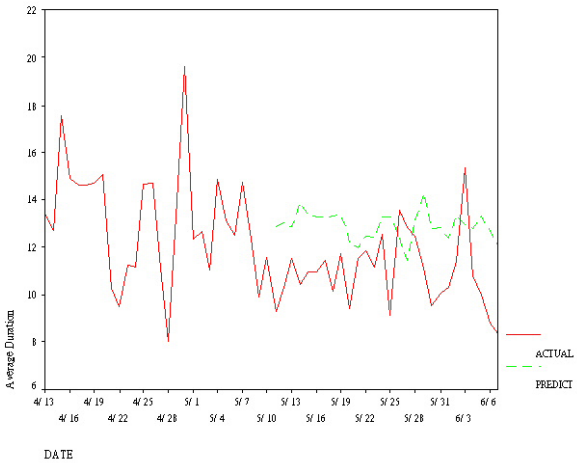

Figure 7. Average duration – first period

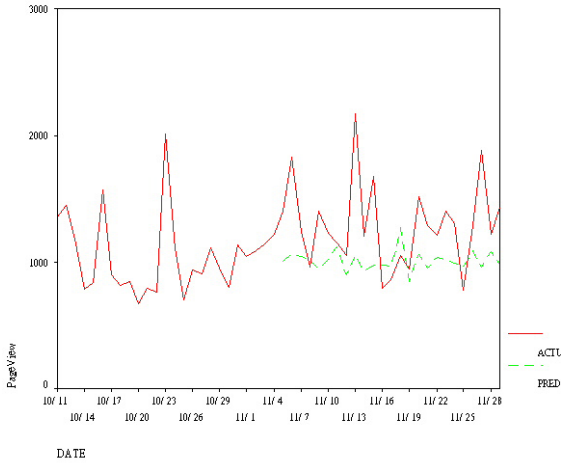

Figure 8. Pageview – second period.

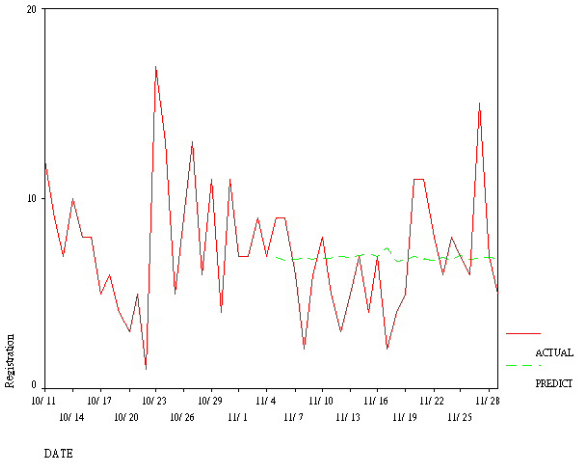

Figure 10. Daily registration – second period.

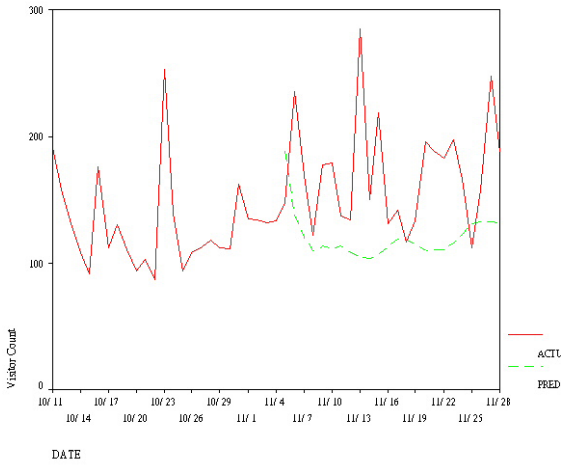

Figure 9. Visitor count -- second period

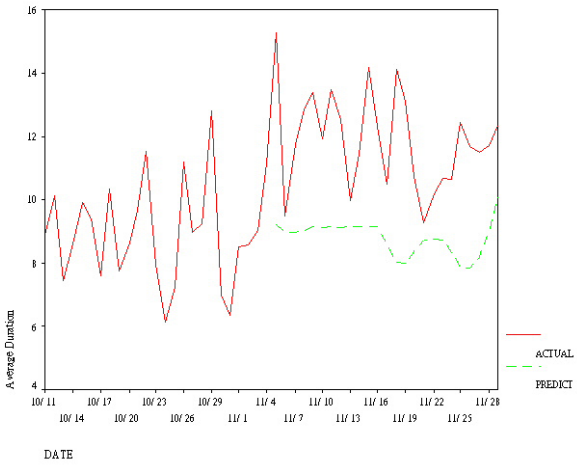

Figure 11. Average duration – second period.

First period: 4/13 to 6/3	Second period: 10/11 to 12/13
Treatment 1: 5/11 First group: 4/13 to 5/10 (**P1G1**) Second Group: 5/11 to 6/3 (**P1G2**)	Treatment 2: 11/4 First group 10/11 to 11/4 (**P2G1**) Second group 11/5 to 12/13 (**P2G2**)

P1G1 – Period 1, Group 1. **P2G1** – Period2, Group 1.

P1G2 – Period 1, Group 2. **P2G2** – Period 2, Group 2.

Table 1. Grouping of data.

Model	AIC
ARIMA（1,0,0）	18.508256
ARIMA（1,0,1）	20.516547

Table 2. AIC's of the residuals of ARIMA(1,0,0) and ARIMA(1,0,1) – first period, first group (P1G1).

Performance Metric	**First Treatment**	**Second Treatment**
Pageviews	Go up and drop rapidly (Fig. 4)	Go up and stay high (Fig. 8)
Visitor Counts	Go up and drop rapidly (Fig. 5)	Go up and stay high (Fig. 9)
Number of Registrations	Go up and stay high (Fig. 6)	No change (Fig. 10)
Average Durations	Go down (Fig. 7)	Go up and stay high (Fig. 11)

Table 3. The effects of treatments.