ISBN: 978-972-8924-65-2 © 2008 IADIS

PREDICTING GROUP SIZE IN SECOND LIFE

Mitul Shukla, Nik Bessis, Marc Conrad, Gordon Clapworthy, Yong Yue
University of Bedfordshire, Department of Computing and Information Systems, Institute for Research in Applicable Computing, Park Square, Luton, LU1 3JU, UK

ABSTRACT

Groups play a central role in structuring the various activities performed by avatars populating the social platform Second Life. This is of particular interest when those groups are used as a means to market a certain product or service. Analysis of group activity or membership within a virtual community could be used to inform a marketing strategy. Although the work presented here is not in itself a marketing strategy, we feel that an investigation into the identification and suitable use of mathematical models evaluating whether there exists a correlation between time and the size of a group would be of value. Such a mathematical model could assist as a tool allowing the measurement of the success or failure of such a marketing campaign. The appropriate means to collect suitable input data is also discussed as is how that data would inform the mathematical model.

KEYWORDS

Virtual Environments, size correlation, mathematical model.

1. INTRODUCTION

Second Life is receiving more and more academic interest, one factor for this, as Castranova points out about Virtual Environments is, "that large games should be thought of as, in effect, social science research tools on the scale of the supercolliders used by physicists" (Castronova 2006). However, Second Life is distinct from Massive Multiplayer Online Role Playing Games (MMORPGs) such as World of Warcraft and Everquest, although both environments do allow for computer mediated shared experiences where users can interact with one another. Within Second Life the user, represented in-world by an avatar, determines what they want to do without an explicit storyline, plot or the necessity to achieve pre-determined goals. A large body of work related to how groups form and evolve in similar internet applications, such as social networking sites, is available, see Backstrom et al. 2006 and Mislove et al. 2007 for evaluations and analysis.

The basic premise of the research is to find out if there is a correlation of group size and time in the Second Life virtual environment and if this correlation can be mathematically modelled. The result of such an investigation could be valuable:

- For group owners to identify/determine which in-world age demographic to target advertising of group membership towards, based on present group membership.
- To understand the time frame involved to reach a particular group size in order to determine what actions are required to reach a target group size.

Assume that a company wishes to launch a viral marketing campaign in Second Life and therefore a group is created and fifty avatars have been successfully convinced to join that group. The question is, whether this is successful or not? (This example is far from fiction and can be seen for instance in the recent Warner Brother's promotion of the TV show 'Gossip Girl' that facilitates Second Life groups for this purpose). To address this issue we need to identify a parameterised step function $b_P(t)$ that maps elapsed time to the expected group size for a given time t. This would then serve as a baseline function that equipped with an appropriate measure μ would allow the quantification of success. For instance if $f(t)$ denotes the actual number of group members at time t the area enclosed between the two functions, namely $\mu(s) := \int_{[0..s]} f(t) - b_P(t)\, dt$ would provide a measurement for the success of the campaign where negative values denote a failure.

2. GROUPS IN SECOND LIFE

An avatar in Second Life can join a group either by invitation, private application or public application. Groups can be joined in Second Life either for free or by paying a joining fee.

Groups in Second Life share characteristics with Communities of Practice (CoP), Communities of Interest (CoI) and Virtual Communities (VC). The term 'Communities of Practice' was coined by Lave and Wenger (1991) and essentially relates to "groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly" (Lave and Wegner, 1991). Bosua and Scheepers point out that when, "…CoPs are geographically dispersed and cannot communicate face-to-face, some infrastructure is needed to promote collaboration, communication and participation of its members" in (Grudin 1994). They go on to say, "…we believe that Web-based technologies may be successfully harnessed as a medium for tacit knowledge exchanges between members of CoPs and also between CoPs within and across organisational boundaries." (op. cit.). Appropriate mechanisms for this are to be found in Second Life.

Involvement in a Community of Practice in Second Life should bring with it at least the reward of shared knowledge and learning. The result of the research proposed here could have value to members or organisers of any group that falls under the categorisation of a Community of Practice for determining group sustainability as explained previously.

There are similarities between groups in Second Life and those found in online social networking sites such as Orkut (www.orkut.com) or Facebook (www.facebook.com). The study and analysis of social networks is not a new phenomenon and there is much literature available in this subject area. The subsequent analysis of online social network sites has grown from the previous body of work and is in itself now sizable. Typical domains of interest and investigation are the 'small world' effect, diffusion and clustering with regards to groups and group membership. Graph theory and network theory tend to dominate this area of research as the most applicable tools. We would recommend the reader to the works of Backstrom et al. 2006 and Mislove et al. 2007 for a grounding in this field of investigation.

3. RESEARCH STRATEGY

3.1 Data Collection

In this section we describe the appropriate and available methods of data collection relevant to this research, as well as pertinent constraints. A number of 'in-world' data collection methods are possible to capture the size of a group at a given point in time or an avatar's membership of various groups. Descriptions of data collection methods and their benefits and drawbacks follows.

Using the 'Search' function of the Second Life interface allows for Group searches to be executed. At regular intervals a predetermined choice of words could be entered into the search field and the results recorded. This process essentially records group size over time. However, this method of data collection does not give any information on current *active* membership. This is true as far as this method does not differentiate between users currently using Second Life and those who have not for a number of years.

An alternative would be for the researcher to go to a given location where the profiles of passers-by could be activated. Then, the 'date of birth' and the number of groups that are currently joined by that avatar could be recorded. This approach does raise some interesting questions about location, traffic flow and privacy. With regards to location, should the location be random or pre-selected? If the location is random, then can traffic flow to the location be guaranteed? If the location is pre-selected then can obvious bias be overcome?

Bias here is in the form of selecting a location that is 'easier' for the researcher in so far as traffic flow is good. Bias of location could also occur as only certain types of user would be attracted to visiting that location, for example certain types of product being on sale in that area or a general meeting place for a given social/geographical/(etc) grouping. Traffic flow or the number of available avatars to get data from would need to be substantial enough for the data set to be valid within a realistic time frame. The time zone that a user is in would also need to be taken into account. Again bias could arise in the selection of the data set if

only users from the same time zone as the researcher are chosen. Therefore the data set would have to be collected in a way as to take into account the different peak and off peak traffic from differing time zones.

One possible approach is to use one of the 'tele-hub' locations where new avatars are sent to after completing their induction training. This type of location is useful as traffic flow is usually high with new avatars arriving regularly throughout any given 24 hour period. Profiles of avatars could be selected and their name recorded with only those being 'born' that day being used for the data set. Then on a regular basis the profiles of given avatars could be searched for and the number of groups joined recorded. What can not be determined by any of the methods mentioned is the amount of time spent in-world by a given avatar.

By using the group listings provided from the group searches described earlier, another strand of investigation is possible. From the group search results of each term an arbitrary group can be chosen. The details for each of these chosen groups can be individually viewed and the membership of that group can be determined and recorded. Once the membership of a group has been recorded then the affiliation of each member to other groups can also be viewed, see Figure 1. This method would give an insight into whether there is a correlation between group members and the number of groups joined.

Figure 1. Group membership information available from a profile.

The screen capture shown in Figure 1 shows on the left hand side an Avatar's profile, from which affiliation to other groups can be ascertained. While on the right a group profile can be seen, this has been activated by double clicking its name in the Avatar profile. The group profile in turn shows the current member listing; it is the use of this member listing, which would be used in the first instance to identify individual members. A combination of the methods outlined here could enable a cube analysis approach with regards to Recency, Frequency and Abandonment. The Recency element would only be possible if 'Last Login' data was collected, see Figure 1. As such, the Recency element would provide data about the last time an Avatar was active in world, the Frequency element would provide data with regards to the number of groups an Avatar has joined, while the Abandonment element would provide data to identify which groups are of high or low value to an Avatar in terms of churn rate.

The use of a software agent to collect data in a similar manner to the approaches mentioned is also possible. However, there exists the question of whether using this type of agent for data mining purposes is in contradiction to paragraph 4.2 of the Second Life Terms of Service which states "You agree to use Second

Life as provided, without unauthorized software or other means of access or use" (Linden Lab, 2008) Finally, a general concern exists about the ethical issues related to the privacy of an avatar in Second Life.

3.2 Possible Mathematical Approaches

A mathematical approach to building a model has the advantage of accuracy based on a given data set and the potential for replication or validity checking by other parties. The following is an overview of various mathematical approaches that could be employed to analyse group size within the Second Life environment as a method to predict if a group will grow, decline or remain static based on individual participation.

Set Theory is essentially the study of collections of objects or elements and the interrelation between two or more collections of the elements contained there in. Here we would model a Second Life group as a set of avatars. Relationships between groups can be interpreted as different types of intersections (empty, subset, non-empty). Conceptually each group, at a given point in time, would be an element of the powerset $P(A)$ where A denotes the set of all avatars.

First Order Logic is a deductive form of logic that assumes the world contains objects, relations and functions. First Order Logic uses predicates and quantification as opposed to the declarative statements of propositional logic. A predicate enables the attribution of a property to an object in a given statement. Quantification here allows the formal treatment of terms such as "all", "every" and "some". This would allow the modelling of statements such that all members of group A and group B also join group C.

Graph Theory typically uses the abstract concepts of nodes being interlinked by edges and an incidence function to represent the relationship between nodes in a topographical manner. Graph theory can be applied across a diverse range areas, for example, the associations of individuals on a social network site, the structure of molecules or modelling a transport network. Here we would interpret the nodes of a graph as the various avatars in Second Life which are connected to each other whenever they share a common group. We may also envisage the labelling of edges according to the number of groups shared.

The concept of Markov Processes, comes from the realm of stochastic processes and conceptualises that a future state can be determined by the present state but is independent from past states. A discrete time Markov process is also known as a Markov chain. A Markov chain describes a chain of processes or events wherein the probability of moving from one to the next is dependant upon the current state. Applied to Second Life that would mean that transitions are triggered by the decisions of avatars to join or leave groups.

Game Theory is concerned with strategic interactions between rational players to determine an optimal or strategic policy dependant on that players preference. Game Theory, then, allows for the analysis of a given players moves or strategies and attempts to make predictions based on those facts. Game Theory lends itself to the analysis and prediction of circumstances in many disparate fields of a competitive nature from economics, political science and business management through to military strategy. In the context of Second Life groups, Game Theory could be a means to interpret the joining or leaving of groups as strategic decisions in order to achieve a certain goal.

Table 1. Intentions and Assumptions.

Intentions:	• To investigate the existence of a correlation between group size and time in the Second Life Virtual Environment. • That the time aspect of the investigation can be expressed either in terms of a given period or in relation to the amount of time lived in world by a new avatar.
Assumptions:	• The investigation is not focussed on the specific nature of the psychological, social or economic factors that may be determinants on an avatar joining a group. • That avatars may or may not be members of multiple groups at any given time.

Table 1 shows the approach criteria of intentions and assumptions to this research, which needs to be explicit before a critical evaluation of mathematical approaches under the constraints of the available data can be carried out.

With these intentions and assumptions borne in mind, a relevant evaluation of mathematical approaches to building a model can be carried out.

The use of Game Theory and Graph Theory do not seem appropriate for use in the context of modelling here. Although Game Theory does lend itself to predictive capability it is fundamentally based on a 'game'

being played by competing players irrespective of the complexity of such a game (de Mesquita 1997). Likewise, the use of Graph Theory as a modelling tool under the circumstances here would be inappropriate as no relationship between one group and another is being investigated, that would fall more in the realm of the investigation of social networks (Backstrom 2006).

The use of Sets and Set Theory would appear to be entirely appropriate for this type of research. One description on the benefits of using Set Theory is that: "Set Theory with its emphasis on consistency and independence proofs, provides a gauge for measuring the consistency strength of various mathematical statements." (Jech 2002). Set Theory derivatives also exist such as Fuzzy Set Theory (Ulieru 2002) and Rough Set Theory (Wu 2004) which have been used in a wide variety of research disciplines such as "expert systems, machine learning, image processing, pattern recognition, knowledge discovery and control systems" (Wu 2004) as well as in looking at Virtual Organisations (Ulieru 2002). First Order Logic with its use of predicates and quantifiers also seems an entirely feasible addition to the 'toolkit' required to build an appropriate model to predict the size of a group within Second Life. In addition, the combination of First Order Logic and Set Theory is an extremely valuable modelling tool (Boyer 1987). In a predictive manner the Markov Process could be used in this context to focus on the size evolution of group membership.

4. CONCLUSION

In order to model and predict a group size within Second Life, issues pertaining to the nature of Second Life and the groups available in that environment have been considered. It is important to note, that our proposed research does not claim to serve as a marketing strategy but rather as a tool to facilitate marketing strategy. A treatment of the issues related to the appropriate methods of data collection available has also been made. Finally, we have appreciated various mathematical approaches and identified Markov Processes underpinned by a Set Theoretic model of the static aspect of Second Life groups as the most suitable approach.

REFERENCES

Backstrom, L. et al, 2006, Group formation in large social networks: membership, growth, and evolution. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining.* Philadelphia, United States of America, pp. 44-54.

Boyer, R. et al., 1987. Set theory in first-order logic: clauses for Gödel's axioms. *Journal of Automated Reasoning*, Volume 2 , Issue 3, pp. 287-327.

Castronova, E. (2006). On the Research Value of Large Games: Natural Experiments in Norrath and Camelot. CESifo Working Paper Series No. 1621 Available at SSRN: http://ssrn.com/abstract=87557

de Mesquita, B., 1997. A decision making model: Its structure and form. *International Interactions,* Volume 23, Issue 3, pp. 235-266.

Grudin, 1994 in *Symmetry of Ignorance, Social Creativity, and Meta-Design.* Fischer, G. 1999. *Proceedings of the 3rd Conference on Creativity & Cognition* (Loughborough, United Kingdom, October 11 - 13, 1999).

Jech, T. 2002 *Set Theory*. http://plato.stanford.edu/entries/set-theory/. Accessed 28/11/2007

Lave, J. & Wenger, E., 1991. *Situated Learning: Legitimate Peripheral Participation*, Cambridge University Press, Cambridge, UK.

Linden Lab, 2008 *Second Life Terms of Service.* http://secondlife.com/corporate/tos.php. Accessed 29/4/08.

Mislove, A. et al, 2007, Measurement and Analysis of Online Social Networks. *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement.* San Diego, California, United States of America, pp. 29-42.

Ulieru, M. 2002. *A Fuzzy Mathematics Approach To Modeling Emergent Holonic Structures.* Invited Chapter in Geometry, Continua and Microstructures, Academic Press, pp. 241-255.

Wu, C. et al., 2004. The rough set theory and applications. *Engineering Computations: International journal for computer-aided engineering and software.* Volume 21, Number 5, pp. 488-511.