5 Conclusions and Research Perspectives

This paper presented the first results of the SIGSNA research project. Aim of the SIGSNA project is to develop a comprehensive analysis of social interactions that take place in the Friendfeed SNS. In this paper we showed a first descriptive analysis of the whole network that pointed out the existence of a large variety of uses inside the SNS. In addition to that, we used a language identification system to make comparative analysis of SNS uses in different cultural contexts. As a preliminary investigation we have also performed a cluster analysis of the Italian Friendfeed network, with which we have identified the existence of two different clusters of users: weak users and highly dedicated users.

The SIGSNA project is characterized by the large database of entries that has been collected during the sampling period. Due to the large dimension of the database and to the high quality of the collected data the presented results have to be considered just as a first bite of the whole research that is still in progress. At the same time, our analysis is highlighting some computational limitations of traditional social data analysis tools, which cannot deal with the large amount of information produced by SNSs — in particular, traditional and text clustering algorithms implemented into widely used statistical tools could not be applied to the whole network, which presents hundreds of thousand users, millions of arcs, and millions of text posts. These limitations will drive the development of scalable techniques for the analysis of large and complex networks, which are necessary to deal with the size of current real social datasets.

References

1. Cosenza, V.: Osservatorio facebook (2009), `http://www.vincos.it/osservatorio-facebook` (retrieved on August 31, 2009)
2. boyd, d., Ellison, N.: Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication 13(1) (2007)
3. Siiback, A.: Online peer culture and interpretative reproduction on children's social networking profiles. In: COST conference proceedings (2009)
4. Hardey, M.: ICT and generations constantly connected social lives, the good the bad the challenging. In: COST conference proceedings (2009)
5. boyd, d.: Taken Out of Context: American Teen Sociality in Networked Publics. PhD thesis, University of California-Berkeley, School of Information (2008)
6. Celli, F.: Slide: Simple language identifier in perl. Technical report, University of Trento (2009)
7. Kaufman, L., Rousseeuw, P.: Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York (1990)

Predictability in an 'Unpredictable' Artificial Cultural Market

Paul Ormerod[1,*] and Kristin Glass[2]

[1] Volterra Consulting, London, UK
pormerod@volterra.co.uk
[2] New Mexico Institute of Mining and Technology, New Mexico, USA
kglass@icasa.nmt.edu

Abstract. In social, economic and cultural situations in which the decisions of individuals are influenced directly by the decisions of others, there is an inherently high level of ex ante unpredictability.

We examine the extent to which the existence of social influence may, paradoxically, increase the extent to which the choice which eventually emerges as the most popular (the 'winner') can be identified at a very early stage in the process. Once the process of choice has begun, only a very small number of decisions may be necessary to give a reasonable prospect of being able to identify the eventual 'winner'.

We illustrate this by an analysis of the music download experiments of Salganik et.al. (2006). We derive a practical rule for early identification of the eventual 'winner'. We validate the rule by applying it to similar data not used in the process of constructing the rule.

Keywords: social influence, unpredictability.

1 Introduction

Enormous resources are devoted to the task of predicting the outcome of social processes in domains such as economics, public policy, and popular culture. But these predictions are often woefully inaccurate. The two most striking characteristics of cultural markets, for example, are *inequality*, in that hit songs, books, and movies are many times more popular than average, and *unpredictability* (for example, Arthur 1989, Bentley et.al. 2007).

Consumer choice in such industries is governed not just by the set of incentives described by conventional consumer demand theory, but by the choices of others (Potts et.al. 2008), so that the payoff of an individual is an explicit function of the actions of others.

In Salganik et al. (2006), researchers constructed an online music market and examined the role social influence played in the songs which participants chose to download. The experiment revealed that increasing the extent to which participants were able to observe the selections of others led to an increase (decrease) in the popularity of

* Corresponding author.

S.-K. Chai, J.J. Salerno, and P.L. Mabry (Eds.): SBP 2010, LNCS 6007, pp. 354–359, 2010.
© Springer-Verlag Berlin Heidelberg 2010

the most (least) popular songs and a decrease in the predictability of song popularity based on quality. Experimental studies, such as those conducted in social psychology (Asch 1953) reach similar conclusions regarding the effects of social influence.

This paper examines the extent to which the existence of social influence may *increase* the extent to which winners can be identified at a very early stage in the process of consumer choices in a market. Section 2 describes the data, section 3 sets out some initial analysis, and section 4 derives a prediction rule.

2 The Data

The Salganik et al. experiment created an artificial 'music market' in which participants downloaded previously unknown songs either with or without knowledge of previous participants' choices. Increasing the strength of social influence increased both inequality and unpredictability of success.

We examined data for 18 experimental worlds, in each of which the same 48 songs were available for downloading. The detailed description of the available data for each of these worlds is available at http://opr.princeton.edu/archive/. In 16 of the worlds a social signal is present. In 8 of these worlds, the person making the choice of whether or not to download was given information on the previous number of downloads carried out by other people, with the songs sorted into popularity at that time. We denote these experiments as being 'strong positive externality process' or strong PEP for short.

Table 1. Various information on the distributions of the final outcomes of the experiments

mean/median	max	N	max/N	experiment
1.3	57	659	8.65	11
1.77	154	1021	15.08	12
1.24	81	834	9.71	21
2.02	158	968	16.32	22
1.39	65	733	8.87	31
1.96	114	892	12.78	32
1.13	66	871	7.58	41
1.7	165	1103	14.96	42
1.21	68	755	9.01	51
1.85	161	1109	14.52	52
1.16	61	944	6.46	61
1.96	135	941	14.35	62
1.17	69	1013	6.81	71
1.77	154	1149	13.4	72
1.14	44	819	5.37	81
2.27	179	926	19.33	82
1.09	77	1571	4.9	91
1	79	2193	3.6	92

In a further 8 worlds, the same information was provided, but it was not sorted into rank order. We denote these experiments as being 'weak positive externality process' or weak PEP for short. Finally, in two of the worlds there is no social signal at all, designated 'no PEP'.

Table 1 sets out information on the final outcomes in each of the experiments.

***Notes*:** Mean/median is the mean number of downloads across the 48 songs at the end of the experiment divided by the median. 'Max' is the number of downloads of the 'winner', the most frequently down loaded song, and N is the total number of downloads. Max/N is 'max' as a percentage of N. The final column is simply the identification tags we assigned to each experiment

3 A Heuristic Prediction Rule

Our aim in this paper is heuristic. Specifically, we examine whether a rule can be discovered which will enable *ex ante* the top ranked song at the end of each experiment to be identified. In other words, we are not trying to predict the exact number of downloads (or market share) at the end of each experiment, but to see if the 'winner' of each experiment (i.e. the top ranked song at the end) can be identified *ex ante*.

A key characteristic of processes of agent choice or selection in which the decisions of others are taken directly into account is that the final outcome of any such process will typically exhibit considerable right-skew (for example, Simon 1955, Bentley et.al. 2009).

We examine the data on a step-by-step basis and to see at what point the outcome could be regarded as exhibiting a right-skew distribution. For a Gaussian distribution, the theoretical mean of the data is equal to the theoretical median. Denote by MM the ratio of the mean to the median. In an empirical setting, MM may deviate from 1 even if the data are Gaussian especially in a small sample. But the deviation is extremely unlikely to be more than 1.1, even with a sample size as small as 20 (calculations available from the authors).

In contrast, in right-skew distributions, the theoretical MM is distinctly larger than 1. For an exponential distribution, with rate parameter λ, the mean is $1/\lambda$ and the median is $\log(2)/\lambda$. So the MM theoretically is $1/\log(2)$, or around 1.44. For a lognormal, where μ is the mean of the natural log of the variable and σ is the standard deviation, the theoretical median is $\exp(\mu)$ and the theoretical mean $\exp(\mu + \sigma^2/2)$, so again MM > 1. And for the power law, empirical estimates of MM will in general give a value > 1 even if the population mean does not exist.

4 Results

We therefore calculated the mean/median value at each step of each experiment (though in the very early stages this ratio does not exist given than the median number of download is zero). We averaged this across the 8 'strong' and 8 'weak' positive externality experiments and across 2 experiments with no such externality.

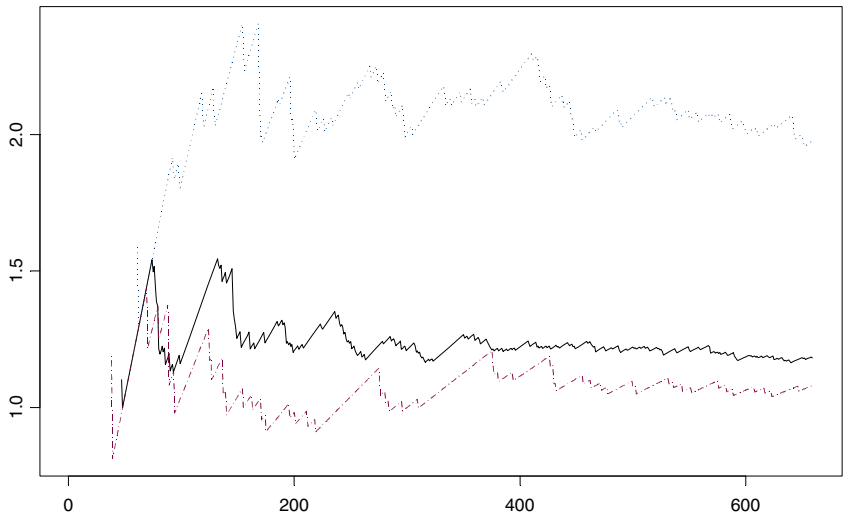

Fig. 1. Dotted line at top is average at step k of the mean/median of the 8 strong PEP experiments; solid line is average at step k of the mean/median of the 8 weak PEP experiments; broken line at bottom is average at step n of the mean/median for the 2 non-PEP experiments. For n close to zero, the median is zero. The data is plotted up to step 659, the length of the shortest experiment.

It is evident that at a fairly early stage in the process, the different types of experiment become differentiated using the mean/median criterion. The next question is therefore whether the empirical mean/median is a useful tool with which to make early identification of eventual 'winners' in the experiments.

As an initial exercise, we selected the first time in each experiment that the mean/median > 1.10, with the next step also being above 1.10. We compared the rankings at this step, τ say, with the rankings in the final step, N.

Specifically, we examined whether the eventual overall winner, the one with the most downloads at time N, can be identified in any way at time τ. Table 2 sets out information on this, along with the percentage of total steps in the experiment which corresponds to step τ.

Column 1 is identifies the experiment in the database we used, and the numbers have no significance as such. Column 2 shows the percentage of total steps in the experiment at which the mean/median > 1.10 for the first time. Column 3 shows the number of downloads of the market leaders at that time. Note that in general it is very small. Column 4 indicates by yes/no whether the winner at time N at the end of the experiment was also the unequivocal leader at time τ. Column 5 indicates by yes/no whether the winner at time N was one of a group of joint leaders at time τ, and column 6 shows the number of joint leaders at time τ.

In nine of the experiments, the eventual winner was either the unequivocal of the joint winner at time τ. Step τ as a percentage of the total number of steps (individual downloads) in the experiment varied between 3.06 and 8.05.

Table 2. Outcome of the use of the decision rule in identifying eventual winners

experiment	τ/N	maximum single download at step τ	winner at time N and winner at time τ	Joint winner at time N, winner at time τ	number of joint winners
11	8.04	2	no	yes	6
12	4.11	7	yes	n/a	n/a
21	5.4	9	no	no	n/a
22	4.75	3	no	yes	4
31	4.5	4	no	no	n/a
32	5.05	4	no	yes	3
41	3.1	2	no	no	n/a
42	3.63	4	no	no	n/a
51	7.02	4	no	yes	2
52	4.26	8	yes	n/a	n/a
61	3.5	4	no	no	n/a
62	5.74	4	no	yes	3
71	3.06	3	no	yes	2
72	5.31	13	yes	n/a	n/a
81	5.74	8	no	no	n/a
82	5.11	8	yes	n/a	n/a
91	2.42	3	no	no	n/a
92	1.41	3	no	no	n/a

In experiment 21, at step τ, where τ is 6.4 per cent of N, the eventual winner was placed joint second. In experiment 32, the eventual winner was third at step τ. The rule was less successful in the other experiments, but not completely without value.

Of the 8 experiments which exhibit strong positive externality processes, the winner at time N can always be identified very early, either unequivocally or as part of a small group, using the mean/median > 1.10 criterion. In addition, as mean/median evolves over time, it rapidly becomes apparent which experiments are strong positive externality processes.

So the simple statistic, the mean/median, appears to be a useful way of a) identifying at an early stage whether a process is governed in part by positive externalities in agent choice and b) identifying at an early stage in processes which do show evidence of positive externalities the choice which will eventually ‘win’ the process.

We checked the validity of the MM rule with 2 further data sets from Salganik which were not used in the process of generating the rule. These had older, more male, and more international participants that were recruited differently from those in the experiments used to develop the rule. So the two provide a useful test of the rule.

In one of the data sets, the eventual winner was also the winner at time τ, when $\tau/N = 6.17$. In the other, the eventual winner was ranked second at time τ, and the eventual second was the winner at time τ. In this case, $\tau/N = 5.64$.

5 Conclusion

In markets where social influence is important in determining whether or not an agent decides to adopt a particular mode of behavior or buy a particular product or brand, a large literature shows that successful *ex ante* prediction of the eventual winner is either very difficult or impossible.

However, the existence of social influence means that it is often possible to identify the eventual winner at a very early stage of the process of choice by participants in the market. We illustrate this with an analysis of the artificial cultural market created by Salganik et.al (op.cit.). We derive a rule for early identification of the eventual winner, which we verify by using it successfully on two further experiments which were not part of the data sets used to create the rule.

Acknowledgement

We are grateful for comments from three anonymous referees for the 2010 International Conference on Social Computing, Behavioral Modeling, & Prediction.

References

1. Arthur, W.B.: Competing technologies, increasing returns, and lock-in by historical events. Economic Journal 99, 116–131 (1989)
2. Asch, S.E.: Effects of group pressure upon the modification and distortion of Judgements. In: Guetzkow, H. (ed.) Groups, leadership and men: Research in human relations. Russell and Russell, New York (1953)
3. Bentley, R.A., Lipo, C.P., Hahn, M.W., Herzog, H.A.: Regular rates of popular culture change reflect random copying. Evolution and Human Behavior 28, 151–158 (2007)
4. Bentley, R.A., Ormerod, P., Batty, M.: An evolutionary model of long tailed distributions in the social sciences (2009), http://arxiv.org/abs/0903.2533
5. Potts, J., Cunningham, S., Hartley, J., Ormerod, P.: Social network markets: a new definition of the creative industries. Journal of Cultural Economics 32, 167–185 (2008)
6. Salganik, M.J., Dodds, P.S., Watts, D.J.: Experimental study of inequality and unpredictability in an artificial cultural market. Science 311, 854–856 (2006)
7. Simon, H.A.: On a class of skew distribution functions. Biometrika 42, 425–440 (1955)

Improving an Agent-Based Model by Using Interdisciplinary Approaches for Analyzing Structural Change in Agriculture

Franziska Appel, Arlette Ostermeyer, Alfons Balmann, and Karin Larsen

Theodor-Lieser-Straße 2, 06120 Halle (Saale), Germany
appel@iamo.de

Abstract. Structural change in the German dairy sector seems to be lagged behind. Heterogeneous farm structures, a low efficiency and profitability are persistent although farms operate under similar market and policy conditions. This raises the questions whether these structures are path dependent and how they can eventually be overcome. To answer these questions we use the agent-based model AgriPoliS. The aim of our project is to improve assumptions in AgriPoliS by using it as an experimental laboratory. In a second part AgriPoliS will be used in stakeholder workshops to define scenarios for the dairy sector and communicate and discuss results to practitioners and decision makers.

1 Introduction and Motivation

Farms in the different parts of Germany operate under relatively similar market conditions and policy environments. In spite of that, a huge regional heterogeneity in terms of farm sizes and specialization is noticeable. This project focuses on the German dairy sector. One reason is that structural change in this sector seems to be particularly lagged behind. The very most dairy farms either operate with inferior techniques or apply them in a less economical way. Also the regional heterogeneity of farm structures is particularly large with, e.g., many small farms in the southern parts of Germany (e.g. Bavaria) and a relatively low number of large dairy farms in the north-eastern parts of Germany (e.g. Saxony-Anhalt). A second reason is that the dairy sector is particularly affected by the ongoing liberalization of the European Union's Common Agricultural Policy (EU CAP). Accordingly, dairy farmers, their representatives and politicians are highly concerned about the future of this sector. Lately the extension of the milk quota which should stepwise lead to a complete abolishment in 2015, causes a fall of the EU milk prices. Because of that, many dairy farmers are threatened in their existences and counteract the CAP reform via strong protests (Deutsche Welle 2009).

An important aim of this research is to analyze structural change in agriculture. Relevant questions are; what are the determinants of structural heterogeneity, is structural change path dependent (cf. David 1985, Arthur 1989, Balmann 1995) and can such path dependences be overcome? In this regard the question arises how fast the structure of the dairy sector is able to adjust to the new EU CAP and how farms

S.-K. Chai, J.J. Salerno, and P.L. Mabry (Eds.): SBP 2010, LNCS 6007, pp. 360–366, 2010.
© Springer-Verlag Berlin Heidelberg 2010