

# Stochastic Models of Large-Scale Human Behavior on the Web

Kristina Lerman<sup>1</sup> and Tad Hogg<sup>2</sup>

1. USC Information Sciences Institute

2. HP Labs

## Abstract

We describe stochastic models of user-contributory web sites, where users create, rate and share the content. These models describe how aggregate measures of activity arise from simple models of individual users. This approach provides a tractable, approximate method to understand user activity on the web site and how this activity depends on web site design choices, such as what information on other users' behaviors is shown to each user. We illustrate this approach in the context of user-created content on the news rating site, Digg.

## Introduction

The Web is becoming more complex and dynamic as sites allow users to contribute and personalize content. Such sites include Digg, Flickr and YouTube where users share and rate news stories, photos and videos, respectively. Additional examples of such web sites include Wikipedia and Bugzilla, enabling anyone to contribute to encyclopedia articles or help develop open source software. These social web sites also often allow users to form explicit links with other users whose contributions they find interesting and highlight the activity of a user's designated friends (Lerman and Jones 2007) to help users find relevant content.

Web sites often provide users with aggregate summaries of recent activity. For example, both Digg and Flickr have a front page that features 'hot' (popular or interesting) content. News organizations, such as The New York Times, allow users to subscribe to or embed RSS feeds of their most popular (e.g., emailed) stories in the users' own pages. Feedback between individual and collective actions can lead to nonlinear amplification of even small signals. For example, the 'Digg effect' refers to the phenomenon where a 'hot' story on the social news aggregator Digg brings down servers hosting the story that are not equipped to deal with heavy traffic that a popular story on Digg generates.

Aggregate contributions of many users determine the structure and usefulness of user-participatory web sites. Understanding this emergent behavior will enable, for example, predicting which newly contributed content will likely become popular, identifying productive ways to change how information display on web sites, or changing incentives so as to improve the content.

Copyright © 2009, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

The behavior of an individual web user is certainly complex and unpredictable, governed by a myriad of social, economic, emotional and cognitive factors, and often subject to unpredictable environmental influences, such as the weather or the stock market. Nevertheless, the combined activities of many users often give remarkably robust aggregate behaviors (Wilkinson 2008).

In this paper, we present a stochastic processes-based framework for modeling regularities in aggregate behavior of web users. The models can be written down directly from the individual behavior descriptions, and quantified with empirical observations of a representative sample of users. This methodology is valid for behaviors that can be modeled as Markov processes (of varying complexity). Though at first glance this may appear overly restrictive, many online activities have a fairly limited set of actions for users and present information based on little or no historical context of particular individuals, allowing a few state variables to capture the main context involved in user actions. Like all formal approaches, the stochastic methodology has limitations, which we also discuss. We illustrate the approach by modeling online behavior on the social news aggregator Digg, which allows users to submit and rate news stories by voting for the stories they like. Digg promotes highly rated stories to the front page, in essence allowing it to emerge from the decisions made by its users. We show that stochastic models are able to explain some features of the collective user behavior.

## Stochastic Models

Rather than account for the inherent variability of individuals, stochastic models focus on the behavior of *average* quantities representing aggregate properties of the system. In the context of a participatory web site, such quantities include average rate at which users contribute new content and rate existing content. Such macroscopic descriptions often have a simple form and are analytically tractable. Stochastic models do not reproduce the results of a single observation — rather, they describe the 'typical' behavior. These models are analogous to the approach used in statistical physics, demographics and macroeconomics where the focus is on relations among aggregate quantities, such as volume and pressure of a gas, population of a country and immigration, or interest rates and employment.

We represent each user as a stochastic process with a

small number of states. This abstraction captures much of the individual user complexity by casting their decisions as inducing probabilistic transitions between states. This modeling framework applies to stochastic processes of varying complexity. In this paper, we focus on simple processes that obey the Markov property, namely, a user whose future state depends only on her present state and the input she receives. A Markov process can be succinctly captured by a diagram showing the possible states of the user and conditions for transition between those states.

With the representation of users based on a small set of relevant states, the same set of states for all users, and transitions depending only on the state and not the individual user, the system as a whole is described simply by the *number* of users in each state at a given time. That is, the system configuration is defined by the occupation vector:  $\vec{n} = (n_1, n_2, \dots)$  where  $n_k$  is the number of users in state  $k$ . For example, in the context of Digg,  $n_k$  could be the number of users who have voted for story  $k$ .

The occupation vector changes as people use the web site, e.g., to view, post and rate content. In principle, one could follow the history of the system, giving a sequence of occupation vectors. However, to investigate typical behavior rather than the details of, say, how users rate a particular story on Digg, we consider a collection of histories of similar content (as determined through a few characteristic properties). This grouping allows the model to generalize from simply describing what has already been observed to predict behavior of similar content that may arise in the future.

The next step in developing the stochastic model summarizes the variation within the collection of histories with a probabilistic description. That is, we characterize the possible occupation vectors by the probability,  $P(\vec{n}, t)$ , the system is in configuration  $\vec{n}$  at time  $t$ . The evolution of  $P(\vec{n}, t)$ , governed by the Stochastic Master Equation (Kampen 1992), is almost always too complex to be analytically tractable. Fortunately we can simplify the problem by working with the average occupation number, whose evolution is given by the Rate Equation

$$\frac{d\langle n_k \rangle}{dt} = \sum_j w_{jk}(\langle \vec{n} \rangle) \langle n_j \rangle - \langle n_k \rangle \sum_j w_{kj}(\langle \vec{n} \rangle) \quad (1)$$

where  $\langle n_k \rangle$  denotes the average number of users in state  $k$  at time  $t$ , i.e.,  $\sum_{\vec{n}} n_k P(\vec{n}, t)$  and  $w_{jk}(\langle \vec{n} \rangle)$  is the transition rate from configuration  $j$  to configuration  $k$  when the occupation vector is  $\langle \vec{n} \rangle$ .

Using the average of the occupation vector in the transition rates is a common simplifying technique for stochastic models. A sufficient condition for the accuracy of this approximation is that variations around the average are relatively small. In many stochastic models of large numbers of components, variations are indeed small due to many independent interactions among the components. More elaborate versions of the stochastic approach give improved approximations when variations are not small, particularly due to correlated interactions (Oppen and Saad 2001). User behavior on the web often involves distributions with long tails, whose typical behaviors differ significantly from the average (Wilkinson 2008). In this case we have no guarantee

that the averaged approximation is adequate. Instead we must test its accuracy for particular aggregate behaviors by comparing model predictions with observations of actual behavior, as we report below.

In the Rate Equation, occupation number  $n_k$  increases due to users' transitions from other states to state  $k$ , and decreases due to transitions from the state  $k$  to other states. The equations can be easily written down from the user state diagram. Each state corresponds to a dynamic variable in the mathematical model — the average number of users in that state — and it is coupled to other variables via transitions between states. Every transition must be accounted for by a term in the equation, with transition rates specified by the details of the interactions between users.

In summary, this stochastic modeling approach to typical aggregate behavior requires specifying the aggregate states of interest for describing the system and how individual user behaviors create transitions among these states. The modeling approach is best suited to cases where the users' decisions are mainly determined by a few characteristics of the user and the information they have about the system. These system states and transitions give the rate equations. Solutions to these equations then give estimates of how aggregate behavior varies in time and depends on the characteristics of the users involved.

The descriptions of aggregate behavior by the Rate Equations is universal, meaning the same mathematical description can be applied to a variety of systems governed by the same abstract principles. This approach was used successfully to study the behavior of several distributed robot systems (Lerman et al. 2001; Lerman and Galstyan 2002; Martinoli, Easton, and Agassounon 2004; Galstyan, Hogg, and Lerman 2005). Stochastic models have also been applied to group behavior in social science, with model parameters estimated from social surveys (Robins et al. 2007).

At the heart of this argument is the concept of separation of scales, which holds that the details of microscopic (user-level) interactions are only relevant for computing the values of parameters of the macroscopic model. This principle applies broadly to naturally evolved systems, as found in biology and economics, and designed technological artifacts (Courtois 1985; Simon and Ando 1961; Simon 1996). From the perspective of large-scale group behaviors, this decomposition often arises from processing, sensory and communication limitations of the individuals and the restricted range of actions. In effect, these limits mean users can only pay attention to a relatively small number of variables (Hogg and Huberman 1987).

### Example: Digg case study

As an example of the stochastic modeling approach, we examine aggregate behavior on Digg, a social news aggregator that relies on users to submit and rate stories. When a user submits a story, it goes to the new stories queue. There are a few new submissions every minute and they are displayed in reverse chronological order of their submission time, 15 stories to a page. A user votes on a story by "digging" it. Sufficiently popular stories are promoted to the front page. Although the exact promotion mechanism is kept secret and

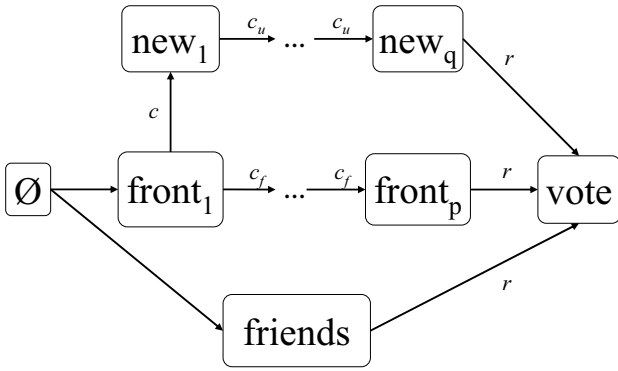


Figure 1: State diagram of user behavior with respect to a single story. A user starts in the  $\emptyset$  state at the left, may then find the story through one of the three available interfaces and then may choose to vote on it.

changes occasionally, it appears to take into account the number of votes the story receives and how rapidly. Digg’s popularity is fueled in large part by the emergent front page, which is formed through the actions of many users.

Digg allows users to track friends’ activities (stories they recently submitted or voted for) through the *Friends Interface*. The friend relationship is asymmetric. When user  $A$  lists user  $B$  as a *friend*,  $A$  is able to watch the activities of  $B$  but not vice versa. We call  $A$  the *fan* of  $B$ .

We identified stories that were submitted to Digg over the course of approximately one day in May 2006 and followed them over several days. Of the 2858 stories submitted by 1570 distinct users over this time period, only 98 stories by 60 users made it to the front page. For this example, we restrict attention to modeling the behavior of stories eventually promoted to the front page, though the model can also describe the dynamics of the majority of stories that never reach the front page. While in new stories queue, a story accrues votes at a slow rate. Once it is promoted to the front page, it accumulates votes at a much faster rate. As the story ages, accumulation of new votes slows down (Wu and Huberman 2007), and the story’s rating saturates at some value, depending on how interesting it is to the Digg community (see Fig. 3(a)). In this section we illustrate the stochastic approach by constructing a model describing how the number of votes received by stories changes in time. Our goal is to produce a model that can explain — and predict — the patterns of collective voting on Digg and how this relates to the ways Digg enables users to discover new content.

## Behavioral Model

Consider the behavior of a Digg user. When the user visits Digg, she can choose to browse its *front* pages to see the recently promoted stories, *new* stories pages for the recently submitted stories, or use the *friends* interface to see the stories her friends have recently submitted or voted for. She can select one of the stories to read, and depending on whether she considers it interesting, *vote* for it. Alternatively, after perusing Digg’s pages, she may choose to leave it. Figure 1

shows the model of user’s behavior. The user’s environment, the stories she is seeing, is itself changing in time, depending on users’ actions. A newly submitted story is visible on the new stories pages for 24 hours after the submission, but also to the submitter’s fans through the Friends interface. If the story accumulates enough votes, it is promoted to the front page, and becomes visible there. With each vote, a new story also becomes visible to the voter’s fans through the “dugg upcoming” part of the Friends interface, which shows the newly submitted stories that user’s friends voted for.

At an aggregate level, we focus on how the number of votes a story receives changes over time. Specifically, let  $N_{\text{vote}}(t)$  be the number of votes the story has received by time  $t$  after it was submitted to Digg. Using Fig. 1 as a modeling blueprint, the rate equation that governs how  $N_{\text{vote}}(t)$  changes during a time interval  $\Delta t$  is:

$$\Delta N_{\text{vote}}(t) = r(\nu_{\text{front}}(t) + \nu_{\text{new}}(t) + \nu_{\text{friends}}(t))\Delta t \quad (2)$$

where  $r$  measures how interesting the story is, i.e., the probability it will receive a vote once seen by the user, and  $\nu_{\text{front}}$ ,  $\nu_{\text{new}}$  and  $\nu_{\text{friends}}$  are the rates at which users find the story via one of the front or new pages, and through the friends interface, respectively. In terms of the general rate equation (Eq. 1), the occupancy vector  $\vec{n}$  describing the aggregate user behavior has the following components: the number of users who see a story via one of the front pages, one of the new pages, through the friends pages, and numbers of users who vote for a story,  $N_{\text{vote}}$ . Since we are interested in the number of users who reach the vote state, we do not need a separate equation for each state in Fig. 1, but simply let  $\nu_{\text{front}}$ , etc. represent the number of users who discover a particular story through the front, etc. page during some time interval. In addition to  $r$ , the parameters users’ state transitions are  $c$ , the rate at which users choose to browse new pages,  $c_f$ , the rate they move from one front page to the next, and  $c_u$ , the rate they move from one new page to the next.

The choice of  $\Delta t$  in Eq. 2 is somewhat arbitrary. For a discrete model, it could correspond to the rate at which the web site is updated or data collected to compare observations with the model. We could also choose to take the limit  $\Delta t \rightarrow 0$  to give a continuous-time model.

## Model parameters

Before we can solve Eq. 2, we must determine its parameters, which depend on the details of user behaviors, the sociological and psychological factors that are not readily available for measurement. Instead, we estimate them using data collected from Digg, as described below.

**Front page** The visibility of a story on the front page decreases as newly promoted stories push it farther down the list. The front pages stories are split into groups of 15, with the first front page displaying the 15 most recently promoted stories, page 2 the next 15 stories, and so on. While we do not have data about Digg visitors’ behavior, specifically, how many proceed to page 2, 3 and so on, generally when presented with lists over multiple pages on a web site, successively smaller fractions of those users visit later pages in

the list (Huberman et al. 1998). While this framework can incorporate any distribution of how users visit sequences of web pages and how they view stories presented in a list on individual pages, we consider a simple model that holds that users view all 15 stories presented on a page and some fraction  $c_f$  of users who view the current front page proceed to the next front page. Thus, if  $\nu$  users visit Digg within a unit time interval, say one hour, the rate users visit the second page is  $c_f\nu$ , and  $c_f^{p-1}\nu$  users per unit time see page  $p$  stories. This model, shown in Fig. 1, captures the decreasing likelihood with which users visit subsequent pages.

**New stories queue** A similar model describes how the number of users who see the story on the new stories queue changes, as the story is superseded with newer submissions. If a fraction  $c$  of Digg visitors proceed to the new stories pages, and of these, a fraction  $c_u$  proceed to the next new page, then the rate users see second page stories is  $cc_u\nu$ , and  $cc_u^{q-1}\nu$  is the rate for page  $q$  stories.

Figure 2(a) shows how the page number of a story on the new and front pages changes in time for three randomly chosen stories from the May data set. This behavior can be fit by lines  $q, p = k_{u,f}t$  with slopes

$$\begin{aligned} k_u &= 3.60 \text{ pages/hr} \\ k_f &= 0.18 \text{ pages/hr} \end{aligned} \quad (3)$$

on the new stories and front page respectively. Since each page holds 15 stories, these rates are  $1/15^{\text{th}}$  the submission and promotion rates respectively.

We use a simple threshold to model how a story is promoted to the front page. When the number of votes a story receives exceeds a promotion threshold  $h$ , the story is visible on the front page. Before that many votes, the story is visible on the new stories pages. This threshold model approximates Digg’s promotion algorithm as of May 2006, since in our data set we did not see any front page stories with fewer than 44 votes, nor did we see any upcoming stories with more than 42 votes. For evaluating the model, we take  $h = 40$ . Moreover, Digg imposes a recency requirement for front page stories, specifically Digg removes new stories after 24 hours.

**Friends interface** The Friends interface allows the user to see the stories her friends have (i) submitted, (ii) voted for, and (iii) commented on in the preceding 48 hours. Although people can use all these features, we only consider the first two and their effect on upcoming stories. Votes for front page stories mostly come from users seeing them directly on the front page so additional votes via the friends interface are a minor contribution. These uses of the friends interface closely approximate the functionality offered by other social media sites: e.g., Flickr allows users to see the latest images his friends uploaded, as well as the images a friend liked (marked as favorite).

Let  $S$  be the number of fans the story’s submitter has, i.e., users who are watching the submitter’s activity. As with other aspects of the model, we can use any distribution for the times fans visit Digg. A simple model takes these users

to visit Digg daily, and since they are likely to be geographically distributed across all time zones, the rate fans discover the story is<sup>1</sup>  $aS\Theta(1 - at)$  at time  $t$  since the story was submitted, where  $a = 1/24$  per hour. The step function accounts for the fact that the pool of fans is finite. As fans read the story, the number of potential voters gets smaller.

As the story receives votes, fans of those voters can see the story through the “newly submitted stories my friends dugg” part of the Friends interface. We use  $S_v$  to denote the combined number of fans of the previous  $N_{\text{vote}}$  voters. The number of users who see the story through this part of the Friends interface is  $aS_v\Theta(h - N_{\text{vote}})$ , where the step function accounts for the fact that only stories on the new pages are shown this way, and  $h$  is the vote threshold for promotion to the front page.

Figure 2(b) shows the average size of  $S_v$ , the combined number of fans of the first  $N_{\text{vote}}$  users to vote on the story. Although  $S_v$  is highly variable from story to story, its average value has consistent growth, approximately

$$S_v = 112.0 \log(N_{\text{vote}}) + 47.0 \quad (4)$$

## Dynamical model

In summary, the rates in Eq. 2 are:

$$\begin{aligned} \nu_{\text{front}} &= c_f^{p(t)-1} \nu \Theta(N_{\text{vote}}(t) - h) \\ \nu_{\text{new}} &= cc_u^{q(t)-1} \nu \Theta(h - N_{\text{vote}}(t)) \Theta(24\text{hr} - t) \\ \nu_{\text{friends}} &= a(S\Theta(1 - at) + S_v\Theta(h - N_{\text{vote}}(t))) \end{aligned}$$

where  $t$  is time since the story’s submission. The first step function in  $\nu_{\text{front}}$  and  $\nu_{\text{new}}$  indicates that when a story has fewer votes than required for promotion, it is visible in the upcoming stories pages; and when  $N_{\text{vote}}(t) > h$ , the story is visible on the front page. The second step function in  $\nu_{\text{new}}$  accounts for a story staying in the upcoming queue for 24 hours. The story’s current page number on the upcoming page  $q$  and the front page  $p$  change in time according to:

$$p(t) = (k_f(t - T_h) + 1)\Theta(t - T_h) \quad (5)$$

$$q(t) = k_u t + 1 \quad (6)$$

with  $k_u$  and  $k_f$  given by Eq. 3, and  $T_h$  is the time the story is promoted to the front page, before which  $p(t) = 0$ .

We solve Eq. 2 subject to initial condition  $N_{\text{vote}}(0) = 1$ , because a newly submitted story starts with a single vote, from the submitter. We take  $\Delta t$  to be one minute. The solutions of Eq. 2 show how the number of votes received by a story changes in time for different values of parameters  $c$ ,  $c_u$ ,  $c_f$ ,  $r$  and  $S$ . Of these, only the last two parameters — the story’s interestingness  $r$  and number of fans the submitter has  $S$  — change from one story to another. Therefore, we fix values of the first three parameters  $c = 0.3$ ,  $c_u = 0.3$  and  $c_f = 0.3$  and study the effect of various  $r$  and  $S$  values. We also fix the rate people visit Digg at  $\nu = 10$  users per minute. The actual visiting rate may be vastly different, but we can always adjust the other parameters accordingly.

Figure 3(a) shows the evolution of the number of votes received by six real stories from our data set.  $S$  is the number of fans the submitter has at submission time. Figure 3(b)

<sup>1</sup> $\Theta(x)$  is a step function: 1 when  $x \geq 0$  and 0 when  $x < 0$ .

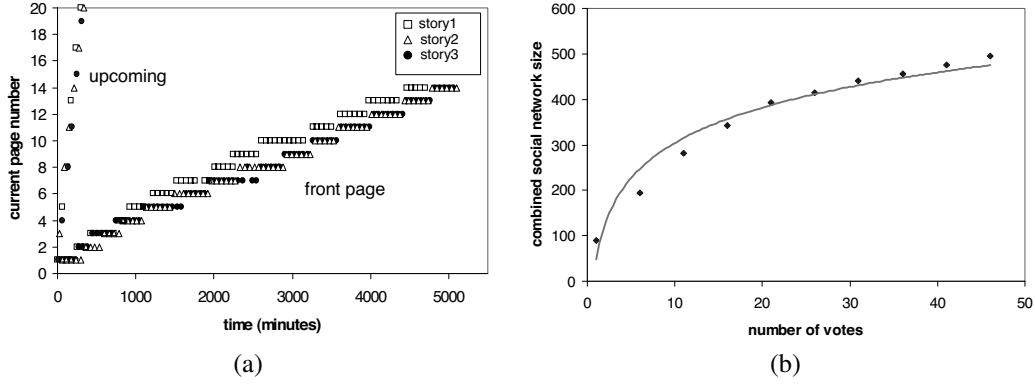


Figure 2: (a) Current page number of a story on the new stories and the front page vs. time for three different stories. Time is measured from when the story first appeared on each page, i.e., time it was submitted or promoted, for the new and front page points, respectively. (b) Growth of the number of distinct users who can see the story through the friends interface of the first 46 users to vote on a story. The points are average values for 195 stories, including those shown in (a), and the curve is from Eq. 4.

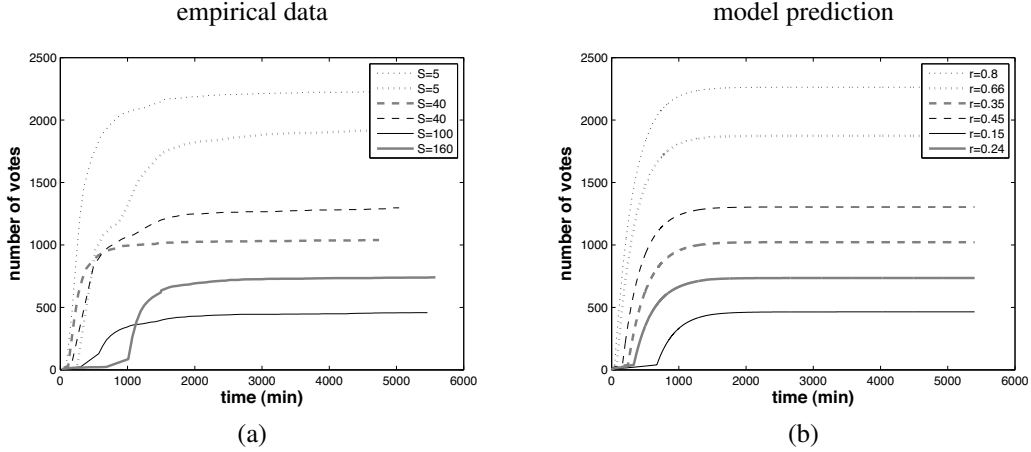


Figure 3: (a) Evolution of the number of votes received by six stories.  $S$  gives the number of submitter's fans. (b) Predictions of the model for the same values of  $S$  and values of  $r$  chosen so as the votes saturate at their observed values.

shows solutions of Eq. 2 for the same values of  $S$  and different values of  $r$  chosen to lead the predicted votes to saturate at their observed values. Overall there is qualitative agreement between the data and the model, indicating that the basic features of the Digg user interface we considered are enough to explain the patterns of collective voting. The more interesting stories (with higher  $r$  values) get promoted to the front page (inflection point in the curve) faster and receive more votes than less interesting stories. As discussed in (Lerman 2007), a story submitted by a poorly connected user (small  $S$ ) has to be very interesting (high  $r$ ) in order to be promoted to the front page, and vice versa, a less interesting story submitted by a highly connected user is able to get to the front page. The only significant difference between the data and the model is visible in the lower two lines. In the data, a story posted by the user with  $S = 100$  is promoted before the story posted by the user with  $S = 160$ , but saturates at smaller value of votes than the latter story. In

the model, the story with bigger  $r$  is promoted first and gets more votes. The disagreement is not too surprising, given the number of approximations in the model. Another effect not in the model is that a story could have a different  $r$  to user's fans than to the general Digg audience. The model can be extended to include inhomogeneous  $r$ .

## Discussion

We described a general approach to relating simple models of user choices to aggregate properties of systems involving large numbers of users. Modeling user-participatory web sites is one application of this approach, as we illustrated with how stories accumulate votes on the Digg web site. This example illustrates how the stochastic approach, based on user state diagrams such as Fig. 1, relates models of individual user behavior to aggregate behavior of the site. Observations allow estimating the quantitative rate parameters appearing in the model. Comparing solutions to the model

with observations can also help identify approaches to improving the model, e.g., by including heterogeneous preferences among users. The user state diagram is determined by the actions and information the web site makes available to users. Whether this approach results in a tractable model depends on the kinds of questions one is interested in and how much user behavior depends on details of user history or on the specific choices of other users rather than just a few aggregate measures provided by the web site.

The connection between user state transitions and aggregate behavior allows investigation of how changes to the web site may change aggregate behaviors. Such hypothetical uses of the modeling approach can suggest improvements to the web site. For example, Digg's promotion algorithm could take into account the number of fans a submitter has, making it more difficult for highly connected users to get uninteresting stories promoted to the front page.

This framework is particularly relevant when information on specific users is limited, as is their range of actions (e.g., posting stories and voting on them in Digg). The framework is less well suited to describing complicated history-dependent actions (e.g., individual users who remember how others treated them in the past as when forming reputations in an e-commerce context). Moreover, while the model can suggest how changes to underlying parameters or user behaviors will affect overall observations, the model provides correlations rather than causal connections between users and observed behavior. In general, there could be other effects, not included in the available observations of the users, that significantly affect behavior and therefore may limit inference from the model of changes that may achieve some more desired behavior (e.g., users spending more time at a web site). Nevertheless, the relations seen with stochastic models can suggest ways to improve the behavior which could be tested, either directly through experimental manipulation of the web site (Salganik, Dodds, and Watts 2006) or through smaller-scale experiments (Kagel and Roth 1995).

A practical challenge for using these models is identifying the relevant states for the users and estimating the transition rates among these states. To some extent, online activities simplify this problem through their limited range of actions and information provided to users. However, web sites can become more personalized over time, e.g., with collaborative filtering for recommendations based on history. This leads to more history-dependence in user behavior and the open question of whether the history-dependence can be summarized in simple additional state variables for the user – such as probability a recommendation is relevant being a function of number of visits the person has had to a site. If so, the model only requires a few additional state variables – in this case number of visits – to regain the Markov property. Alternatively, we can generalize the model to allow the transition to the next state to depend not just on the current state but also some fixed number of past states, as has been applied to dynamic task allocation (Lerman et al. 2006).

As web sites develop greater complexity and personalization, model-based design tools could help identify aggregate consequences of design choices of actions and information provided to users. More broadly, such models could also

complement economic or game theory analyses of the incentives for participation provided to the users.

## References

- Courtois, P. J. 1985. On time and space decomposition of complex structures. *Communications of the ACM* 28(6):590–603.
- Galstyan, A.; Hogg, T.; and Lerman, K. 2005. Modeling and mathematical analysis of swarms of microscopic robots. In Arabshahi, P., and Martinoli, A., eds., *Proc. of the IEEE Swarm Intelligence Symposium (SIS2005)*, 201–208.
- Hogg, T., and Huberman, B. A. 1987. Artificial intelligence and large scale computation: A physics perspective. *Physics Reports* 156:227–310.
- Huberman, B. A.; Pirolli, P. L. T.; Pitkow, J. E.; and Lukose, R. M. 1998. Strong regularities in World Wide Web surfing. *Science* 280:95–97.
- Kagel, J., and Roth, A. E., eds. 1995. *The Handbook of Experimental Economics*. Princeton Univ. Press.
- Kampen, N. G. V. 1992. *Stochastic Processes in Physics and Chemistry*. Amsterdam: Elsevier Science, revised and enlarged edition.
- Lerman, K., and Galstyan, A. 2002. Mathematical model of foraging in a group of robots: Effect of interference. *Autonomous Robots* 13(2):127–141.
- Lerman, K., and Jones, L. 2007. Social browsing on flickr. In *Proc. of International Conference on Weblogs and Social Media (ICWSM-07)*.
- Lerman, K.; Galstyan, A.; Martinoli, A.; and Ijspeert, A. J. 2001. A macroscopic analytical model of collaboration in distributed robotic systems. *Artificial Life* 7:375–393.
- Lerman, K.; Jones, C. V.; Galstyan, A.; and Mataric, M. J. 2006. Analysis of dynamic task allocation in multi-robot systems. *International Journal of Robotics Research* 25(3):225–242.
- Lerman, K. 2007. Social information processing in social news aggregation. arxiv.org preprint cs.cy/0703087.
- Martinoli, A.; Easton, K.; and Agassounon, W. 2004. Modeling of swarm robotic systems: A case study in collaborative distributed manipulation. *Int. Journal of Robotics Research* 23(4):415–436.
- Opper, M., and Saad, D., eds. 2001. *Advanced Mean Field Methods: Theory and Practice*. Cambridge, MA: MIT Press.
- Robins, G.; Pattison, P.; Kalish, Y.; and Lusher, D. 2007. An introduction to exponential random graph (p\*) models for social networks. *Social Networks* 29:173–191.
- Salganik, M. J.; Dodds, P. S.; and Watts, D. J. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311:854–856.
- Simon, H. A., and Ando, A. 1961. Aggregation of variables in dynamic systems. *Econometrica* 29:111–138.
- Simon, H. A. 1996. *The Sciences of the Artificial*. Cambridge, MA: MIT Press, 3rd edition.
- Wilkinson, D. M. 2008. Strong regularities in online peer production. In *Proc. of the 2008 ACM Conference on E-Commerce*, 302–309.
- Wu, F., and Huberman, B. A. 2007. Novelty and collective attention. *Proc. of the Natl. Acad. Sci.* 104:17599–17601.