Predictive Analysis for Social Processes

Kristin Glass, Richard Colbaugh, and Dennis Engi

Sandia National Laboratories
P.O. Box 5800, Albuquerque, NM 87185

{klglass | rcolbau | dengi} @sandia.gov

Abstract

This paper presents a new approach to predictive analysis for social processes. A key element of the proposed methodology is proper characterization of the interplay between the *intrinsic* aspects of a social process (e.g., the persuasiveness of an argument) and the *social dynamics* which is its realization (e.g., the way the argument propagates through a segment of society). We show that this interplay can be modeled within a novel multi-scale framework that is sociologically and mathematically sensible, expressive, illuminating, and amenable to formal analysis. We then develop a scientifically rigorous, computationally tractable approach to predictive analysis. Among other capabilities, this analytic approach enables assessment of process predictability, identification of measurables which have predictive power, discovery of reliable early indicators for events of interest, and scalable, robust prediction. The potential of the proposed approach is illustrated through a case study involving early warning analysis for mobilization/protest events.

Introduction

Enormous resources are devoted to the task of predicting the outcome of social processes, in domains such as public policy, economics, popular culture, and national security, but these predictions are often woefully inaccurate. Consider, for instance, the case of cultural markets. Perhaps the two most striking characteristics of these markets are their simultaneous *inequality*, in that hit songs, books, and movies are many times more popular than average, and *unpredictability*, so that well-informed experts routinely fail to identify these hits beforehand. Examination of other domains in which the events of interest are outcomes of social processes reveals a similar pattern – market crashes, regime collapses, fads and fashions, and social movements involve significant segments of society but are rarely anticipated.

It is tempting to conclude that the problem is one of insufficient information. Clearly winners are qualitatively different from losers or they wouldn't be so dominant, the conventional wisdom goes, so in order to make good predictions we should collect more data and identify these crucial differences. Recent research in the social and behavioral sciences calls into question this conventional wisdom and, indeed, indicates that there may be fundamental limits to what can be predicted about social systems. Consider social processes in which individuals pay attention to what others do. Recent empirical studies offer evidence that the *intrinsic* characteristics of the outcomes of such processes, such as the quality of the various options in a social choice situation, often do not possess much predictive power. For example, the study reported in (Walls 2005) finds that the (intrinsic) attributes normally considered to be important when assessing the likelihood of movie box office success, such as the director, actors, genre, and critic reviews, are not statistically significantly related to box office revenues. Similar results can hold in adoption of innovations (e.g., Arthur 1989), diffusion of sociopolitical behaviors (e.g., Hedstrom et al. 2000), sales in online markets (e.g., Leskovec et al. 2006, Colbaugh et al. 2008), and trading in financial markets (e.g., Shiller 2000). The available experimental evidence, although much more limited, also supports this conclusion (e.g., Salganik et al. 2006).

This research provides compelling evidence that, for many important social processes, it is not possible to obtain useful predictions using standard methods which focus on the intrinsic characteristics of potential process outcomes. We propose that accurate prediction, if it is possible at all, requires careful consideration of the subtle interplay between the intrinsics of a social process and the underlying *social dynamics* which is its realization. This paper presents a new approach to predictive analysis which exploits this idea. We propose a novel multi-scale modeling framework for social processes that captures the interplay between a process's intrinsic features and its dynamics in a sociologically-grounded way and which yields models that are amenable to formal analysis. We then formulate predictive analysis questions in terms of social system reachability, and present a mathematically rigorous, computationally tractable method for deciding reachability and, consequently, for answering prediction questions. The analytic methodology enables assessment of process predictability, identification of measurables with predictive power, discovery of early indicators for events of interest, and scalable, robust prediction. The potential of the proposed approach is illustrated with a social movement warning analysis case study.

Predictive Analysis

Problem formulation

We formulate prediction problems as questions about the expected dynamics of a system of interest. Given a social process, a set of candidate process measurables, and the behavior about which predictions are to be made, we represent the process as a stochastic dynamical system evolving on some state space, relate the observables to process characteristics (e.g., parameters) and/or states, and encode the behavior of interest in terms of (the satisfaction of) reachability conditions.

As an illustrative example, consider the problem of predicting ultimate market share in a cultural market (e.g., music or films) in which "buzz" about products propagates through various social networks. If, in a market containing two products with indistinguishable "intrinsic appeal", it is possible for one of the products to achieve a dominant market share, we might view the market to be unpredictable. Conversely, a predictable market would be one in which market shares of indistinguishable products evolve similarly and market shares of superior products are typically larger than those of inferior ones. Prediction, of course, then involves estimating the ultimate market share of a product of interest, perhaps based on measures of appeal. In our formulation, market share dominance by product A is associated with a region of market share state space, and the condition that A eventually achieves such dominance and simultaneously possesses an appeal that is indistinguishable from product B is easily written as the satisfaction of a state space reachability condition.

To formulate prediction questions in terms of reachability, the behavior about which predictions are to be made is used to define system state space subsets of interest (SSI). Candidate measurables allow identification of indistinguishable starting sets (ISS), that is, those sets of initial states and system parameters which cannot be resolved with the available data. This setup permits the four predictive analysis tasks of interest – predictability assessment, identification of useful measurables, warning analysis, and prediction – to be performed in a systematic manner. Predictability assessment involves determining which SSI can be reached from ISS and deciding if these reachability properties are compatible with the prediction goals. More specifically, if moving between state-parameter pairs within an ISS leads to unacceptably large variation in the probability of reaching an SSI then the process is *initial state (IS) unpredictable* (and is IS predictable otherwise). Alternatively, if the probabilities of reaching two qualitatively distinct SSI are both high then the process is *eventual state (ES) unpredictable* (and is ES predictable otherwise). This analysis leads naturally to a way to identify those measurables with the most predictive power: these are the components of the vectors that comprise the ISS for which IS or ES predictability is most sensitive. If the system's reachability properties are incompatible with the prediction goals – if, for instance, "hit" and "flop" in a cultural market are both reachable from a single ISS – then the given prediction question should be refined in some way. Possible refinements include relaxing the level of detail to be predicted (by redefining the SSI) or introducing additional measurables to resolve the ISS.

If and when a predictable situation is obtained, the problems of discovering reliable early indicators for events of interest and forming robust, useful predictions can be addressed. These problems are also naturally studied within the reachability framework. Warning analysis involves identifying *indicator sets*, that is, process state space subsets with the property that observing a trajectory entering an indicator set implies that the event of interest is likely to occur. Prediction entails estimating the probability that the process will evolve to an SSI and quantifying the uncertainty associated with this estimate.

Social process models

In many social situations, individuals are influenced by observations of (or expectations about) the behavior of others, for instance seeking to obtain the benefits of coordinated actions, infer otherwise inaccessible information, or manage complexity in decision-making. Processes in which observing a certain behavior *increases* an agent's probability of adopting that behavior are often referred to as positive externality processes (PEP). PEP have been widely studied in the social and behavioral sciences and, more recently, by the computer science and informatics communities. One hallmark of PEP is their unpredictability: phenomena from fads and fashions to financial market bubbles and crashes appear resistant to predictive analysis (although there is no shortage of *ex post* explanations for their occurrence!). These considerations are important for national security applications as well. For example, there is increasing recognition that collective dynamics are central to numerous security-relevant social movements, including revolutions, political and religious radicalization, and cultural/ethnic conflicts.

A key step in understanding PEP dynamics, and the impact of these dynamics on predictive analysis, is the formulation of appropriate social dynamics models. Toward this end, we propose a class of multi-scale models for social processes which capture the function and structure of the underlying systems, including the role played by social networks, through the use of three modeling scales:

- a *micro-scale*, for modeling the behavior of individuals;
- a *meso-scale*, which represents the collective dynamics within "social contexts";
- a *macro-scale*, which characterizes the interaction between the social contexts.

The micro-scale model enables the interplay between process intrinsics and social influence to be characterized and quantified; observe that this interaction is particularly important for PEP dynamics. The meso-scale level captures the dynamics of individuals interacting *within* social contexts, that is, within localized settings (e.g., defined by workplace environment or family structure) in which interaction between individuals is (approximately) "fully mixed" (e.g., Hedstrom et al. 2000). Finally, social interaction *between* contexts is modeled at the macro-scale level. Distinguishing between the way individuals interact within and across social contexts in this way provides an analytically tractable means of capturing the important social network structures present in most social processes. The basic modeling framework is depicted in Figure 1.

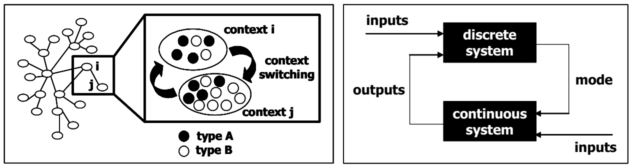


Figure 1: Multi-scale model for social process. The cartoon at left illustrates the basic model structure, in which individuals (black and white nodes) interact within social contexts (ellipses encircling nodes) via fully mixed dynamics and between contexts according to the network topology characterizing context interdependencies. The block diagram at right depicts an S-HDS encoding of the model.

We build and analyze multi-scale social dynamics models using a hybrid systems formalism (Colbaugh et al. 2008). A *stochastic hybrid dynamical system* (S-HDS) is a feedback interconnection of a continuous-time, continuous state-dependent Markov chain $\{Q, \Lambda(h(x))\}$ and a collection of stochastic differential equations indexed by the Markov chain state q:

$$\Sigma_{\text{S-HDS}} \qquad \begin{array}{c} \{Q, \Lambda(h(x))\}, \\ dx = f_q(x, p)dt + G_q(x, p)dw, \\ k = h(x), \end{array}$$

where $q \in Q$ (with $|Q|$ finite), $x \in X \subseteq \Re^n$ is the state of the continuous system, $p \in P \subseteq \Re^p$ is the system parameter vector, $\Lambda(x)$ is the matrix of (x-dependent) Markov chain transition rates, $\{f_q\}$ and $\{G_q\}$are sets of vector and matrix fields characterizing the continuous system dynamics, w is an m-valued Weiner "disturbance" process, and h defines a partition of the continuous state space into subsets labeled with index k. It is worth mentioning that much is known about how to construct vector fields $\{f_q\}$ and matrices $\{G_q\}$ to model PEP dynamics (e.g., Colbaugh et al. 2008).

An extensive discussion of general deterministic and stochastic hybrid systems can be found in (Bemporad et al. 2007). Briefly, S-HDS are feedback interconnections of continuous dynamics, such as the dynamics of individuals exchanging ideas within a social context, and discrete dynamics, capturing for instance the switching behavior encountered when an individual from one context moves to another and introduces an idea which is novel in the latter context (see Figure 1). As we show next, an advantage of representing multi-scale social dynamics using an S-HDS framework is that the resulting models are amenable to formal analysis.

Predictive analysis via reachability

Recall that we formulate predictive analysis in terms of reachability of social processes. More specifically, the preceding discussion shows that we need to develop a rigorous, tractable methodology for assessing reachability of S-HDS social process models. Particularly desirable are methods which can be applied in the presence of both stochastic dynamics and parametric uncertainty, as these are ubiquitous in social systems. An approach to reachability assessment that possesses these characteristics can be developed using the concept of "altitude functions" (Colbaugh et al. 2008).

The basic idea is illustrated in Figure 2. Consider a system with dynamics that evolve on a continuous state space, depicted in Figure 2 according to the vector field (arrows) shown in the figure. Assume it is of interest to decide whether states in the ISS can evolve to the SSI. Such reachability questions are usually approached by computing the reach set associated with the ISS region and determining if it intersects the SSI region, and this analysis is quite difficult for generic systems. Suppose, however, that it is possible to find an "altitude function" A(x) which has a level curve, say the A(x) = 0 surface, that separates the ISS and SSI regions and on which the vector field points toward the partition containing the ISS region (see Figure 2). In this case we can conclude that the SSI region is not reachable from the ISS region and, moreover, this conclusion is reached without computing system trajectories. Note also that this reachability assessment is carried out for entire sets of initial conditions (and system parameter values); these features of the process greatly enhance the practical utility of the analytic approach.

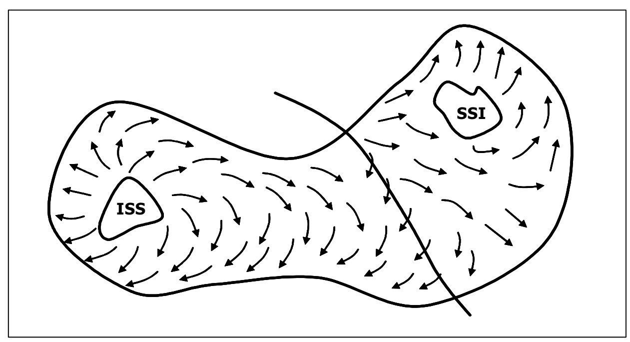


Figure 2: Cartoon of altitude function-based reachability analysis.

To make this basic idea more quantitative, consider the S-HDS model $\Sigma_{\text{S-HDS}}$. Denote by X, X_0, X_u, and P the sets of feasible states, initial states, undesirable states, and parameter values, respectively, and assume that X and P are bounded. We are interested in the following reachability problem: given system $\Sigma_{\text{S-HDS}}$ and the sets X, X_0, X_u, the *stochastic safety problem* involves computing an upper bound $\gamma \in [0,1]$ on the probability that any trajectory of $\Sigma_{\text{S-HDS}}$ will reach X_u; thus γ satisfies

$$\text{Prob}\{\underline{x}(t) \in X_u \text{ for some t}\} \leq \gamma$$

for all $x_0 \in X_0$, where $\underline{x}$ is the "stopped" process associated with x_0 and $\Sigma_{\text{S-HDS}}$ (roughly, $\underline{x}(t)$ is the trajectory of $\Sigma_{\text{S-HDS}}$ which starts at x_0 and is stopped if it encounters the boundary of X (Kushner 1967)).

Observe that this problem is central to our approach to predictive analysis. For instance, assessing IS predictability involves determining whether moving between state-parameter pairs within an ISS leads to acceptable variations in the probability of reaching an SSI, and this determination has at its core the stochastic safety problem.

We adopt an analysis methodology which is analogous to the one underlying Lyapunov function-based stability analysis (Sontag 1998): we seek a scalar function of the system state that permits the probability upper bound γ to be deduced *without* computing system trajectories. In order to derive such results we require an additional concept. The *infinitesimal generator B* for the process x(t) is given by $BA(x_0) = \lim_{t\to 0} (E[A(x(t)) \mid x(0) = x_0] - A(x_0)) / t$, where A(x) is any differentiable scalar function (Kushner 1967). The infinitesimal generator B is the stochastic analog of the Lie derivative. For example, the generator for $\Sigma_{S\text{-}HDS}$ can be written $(\partial A_q/\partial x)f_q + (1/2)\ \mathrm{tr}[G_q^T(\partial^2 A_q/\partial x^2)G_q] + \Sigma_{q'\in Q}\lambda_{qq'}B_{q'}$ (where Markov chain transition rates $\lambda_{qq'}(x)$ are defined so $\lambda_{qq'}(x) \geq 0$ if $q \neq q'$ and $\Sigma_{q'}\lambda_{qq'}(x) = 0\ \forall q$).

We are now in a position to state our main result for the stochastic safety problem:

Theorem 1: γ is an upper bound on the probability of trajectories of $\Sigma_{S\text{-}HDS}$ reaching X_u from X_0 while remaining in $Q \times X$ if there exists a family of differentiable functions $\{A_q(x)\}_{q\in Q}$ such that

- $A_q(x) \leq \gamma\ \forall x\in X_0,\ \forall q\in Q$;
- $A_q(x) \geq 1\ \forall x\in X_u,\ \forall q\in Q$;
- $A_q(x) \geq 0\ \forall x\in X,\ \forall q\in Q$;
- $(\partial A_q/\partial x)f_q+(1/2)\mathrm{tr}[G_q^T(\partial^2 A_q/\partial x^2)G_q]+\Sigma_{q'\in Q}\ \lambda_{qq'}B_{q'} \leq 0$ $\forall x\in X,\ \forall q\in Q,\ \forall p\in P$.

Proof: The proof is given in (Colbaugh et al. 2008). ■

The preceding theoretical result is of significant practical interest only if it is possible to efficiently compute altitude functions A(x). Toward that end, observe that the result presented in Theorem 1 specifies *convex* conditions which must be satisfied by the associated altitude functions; thus the search for altitude functions can be formulated as a convex programming problem (Parrilo 2000). Moreover, if the system of interest admits a polynomial description (i.e., the system vector fields are polynomials and system sets are semialgebraic) and if we restrict our search to polynomial altitude functions, then the search can be very efficiently carried out using sum of squares (SOS) optimization (SOSTOOLS 2007).

SOS optimization is a convex relaxation framework based on SOS decomposition of the relevant polynomials and semidefinite programming. SOS relaxation involves replacing the (non-)negative and (non-)positive conditions to be satisfied by the altitude functions with SOS conditions. As a simple example of the basic idea, consider the following relaxation of the first condition given in Theorem 1:

$$A_q(x) \leq \gamma\ \forall x\in X_0,\ \forall q\in Q \quad \rightarrow \quad \gamma - A_q(x) - \lambda_{0q}^T(x)\ g_{0q}(x) \text{ is SOS}$$

where the entries of the vector functions λ_{0q} are SOS and $g_{0q} \geq 0$ (entry-wise) when $x\in X_0$. More complex conditions on the A_q can be relaxed analogously. The relaxed SOS conditions are obviously sufficient and typically are not overly conservative. Software for solving SOS programs is available as a third-party Matlab toolbox (SOSTOOLS 2007). Importantly, the approach is tractable: for fixed polynomial degrees, the computational complexity of the associated SOS program grows polynomially in the cardinality of the discrete state set and the dimensions of the continuous state and parameter spaces.

Observe that altitude function-based reachability analysis provides the foundations for a rigorous, computationally tractable predictive analysis methodology. For example, Theorem 1 defines a procedure for determining (an upper bound for) the probability of a social process reaching the SSI from an ISS, and SOS programming (via SOSTOOLS) provides a tractable means of implementing this procedure. This capability directly enables assessment of both IS and ES predictability, as each involves deciding ISS to SSI reachability. The methodology also gives an efficient method for identifying measurables with predictive power and indicators useful for warning analysis. More specifically, repeated application of Theorem 1 allows exploration of the space of candidate measurables/early indicators to find those that are actually predictive for the behavior of interest. Finally, prediction entails estimating the probability that a given process will evolve to each of a set of SSI, conditioned on the observed measurables.

Sample theoretical result

We now apply the proposed predictive analysis framework to the early warning problem for social movements. Social movements are large, informal groupings of individuals or organizations focused on a particular issue, for instance of political, social, economic, or religious significance. Example issues which have spawned such movements include racial and gender equality, temperance, labor rights, political ideology, economic philosophy, religious fundamentalism, and environmental concerns. Study of these and other social movements has demonstrated that PEP dynamics are an important driver in their emergence and growth. Consider the problem of distinguishing successful social movements, that is, movements which attract significant following, from unsuccessful ones early in their lifecycle. A crucial step in the proposed methodology is identifying those measurables which are most useful for warning. This function is expected to be particularly important here, as there are myriad measurables that *may* have predictive power, and identifying which (if any) actually do is both challenging and critical for reliable warning analysis.

To enable development of a warning analysis capability for social movements, we first collect a family of models from the social movement theory (SMT) literature and formulate these within our multi-scale framework. More specifically, we derive an S-HDS representation for social movements which is of the form $\Sigma_{S\text{-}HDS}$, and find that this formulation enables simultaneous analysis of an entire collection of relevant SMT models (Colbaugh et al. 2008). Next, we quantify movement success by defining a subset X_u of the social system state space that corresponds to a level of movement membership consistent with movement goals, and we seek measurables which allow early identification of those movements that are likely to evolve to X_u. That is, we compute provably-correct upper bounds for the probability that any model in the collection will reach X_u from X_0. Because this computation does not require forward simulation and can be conducted for *sets* of initial states and parameter values, we can efficiently explore the way various measurables affect these probability bounds and identify those for which the probability of reaching X_u exhibits sensitive dependence. These measurables can then be tested for their utility in distinguishing successful and unsuccessful movements.

This study produced two main results. First, the degree to which movement-related activity shows *early diffusion* across multiple social contexts is a powerful distinguisher of successful and unsuccessful social movements. Indeed, this measurable has considerably more predictive power than the volume of such activity and also more power than various system intrinsics. Second, significant social movements occur with finite probability only if both 1.) the intra-context "infectivity" of the movement exceeds a certain threshold and 2.) the inter-context interactions associated with the movement take place with a frequency that is larger than another threshold. Note that this is reminiscent of, and significantly extends, well-known results for epidemic thresholds in disease propagation models.

Case Study: Muslim Mobilization

Consider the task of identifying reliable early indicators for successful social movements, for instance mobilization in response to perceived affronts to a segment of society. The theoretical result summarized above suggests that social network dynamics are critical to social movement success. Moreover, the results show that the features of these dynamics which may be useful early indicators of movement success are practically measurable in many applications. Diffusion across social contexts often can be inferred from analysis of public opinion and demographic data, as this measure requires only incomplete information regarding the relevant social networks. Alternatively, we show below that *online* (i.e., Web-based) social activity can sometimes serve as a proxy for these social dynamics.

In this case study we investigate whether diffusion across social contexts is a useful early indicator for successful Islamic mobilization and protest events. The study focuses on Muslim reaction to six recent incidents, each of which appeared at their outset to have the potential to trigger significant protest activities:

- publication of photographs and accounts of prisoner abuse at Abu Ghraib in Spring 2004;
- publication of cartoons depicting the prophet Mohammad in the Danish newspaper Jyllands-Posten in September 2005;
- distribution of the DVD "I was blind but now I can see" in Egypt in October 2005;
- the lecture given by Pope Benedict XVI in September 2006 in which he quoted controversial material concerning Islam;
- Salman Rushdie being knighted in June 2007;
- republication of the "Danish cartoons" in various newspapers in February 2008.

Recall that the first Danish cartoons event ultimately led to substantial Muslim mobilization, including massive protests and considerable violence, and that the Egypt DVD event also resulted in significant Muslim mobilization and violence. In contrast, Muslim outrage triggered by Abu Ghraib, the pope lecture, the Rushdie knighting, and the second Danish cartoons event all subsided quickly with essentially no violence. Therefore, taken together, these six events provide a useful setting for testing whether the extent of early diffusion across social contexts can be used to distinguish nascent Islamic mobilization events which become large and self-sustaining (and potentially violent) from those that quickly dissipate.

A central element in the proposed approach to early warning analysis is the measurement, and appropriate processing, of social dynamics associated with the process of interest. Indeed, the preceding results suggest that in many cases reliable warning analysis *requires* such data. In the present study we use online social activity as a proxy for "real world" diffusion of mobilization-relevant information. More specifically, we use blog discussions as our primary data set. The "blogosphere" is modeled as a graph composed of two types of vertices, the blogs themselves and the concepts which appear in them. Two blogs are linked if a post in one hyperlinks to a post in the other, and a blog is linked to a concept if the blog contains (a significant occurrence of) that concept. Among other things, this blog graph model enables the identification of blog communities, sets of blogs with intra-group edge densities that are significantly higher than expected (Newman 2003); these blog communities serve as one of our proxies for social contexts.

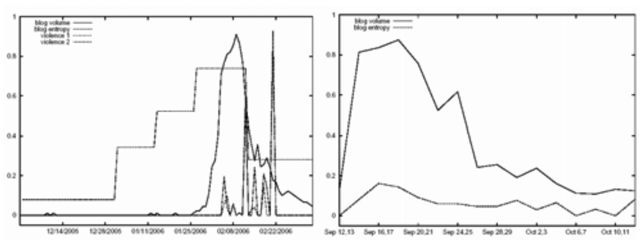
We propose the following procedure for warning analysis given a potential "triggering" event of interest.

1. Use key words/concepts associated with the triggering event to collect relevant blog posts and build the associated blog graph.
2. Identify the blog social contexts (e.g., graph community-based, language-based).
3. Construct the post volume time series for each social context. Compute the post/context entropy (PCE) time series associated with the post volume time series. We define the PCE for a given sampling interval t as $PCE(t) = -\Sigma_i f_i(t) \log(f_i(t))$, where $f_i(t)$ is the fraction of total relevant posts during interval t which occur in context i.
4. Construct a *synthetic* ensemble of PCE time series from (actual) post volume dynamics using a general S-HDS social diffusion model (see Predictive Analysis section).

5. Perform motif detection: compare the actual PCE time series to the synthetic ensemble series to determine if the early diffusion of activity across contexts is “excessive”. Flag events with excessive early diffusion for further (e.g., manual) analysis.

This approach to early warning analysis was applied to the Islamic mobilization case study. If early diffusion of discussions across blog communities is an indicator that the associated Islamic mobilization event will be large, we would expect to observe such diffusion with the mobilization associated with the first Danish cartoons and Egypt DVD events and not with the other four events. Additionally, we would expect this early diffusion to be “excessive”, relative to the synthetic ensemble, for the first two events and not for the latter four.

As can be seen in Figure 3, this is precisely what we find. In the case of the first Danish cartoons event (Figure 3, top left), the entropy of diffusion of relevant discussions across blog communities (dashed curve) experiences a dramatic increase a few weeks before the corresponding increase in the volume of blog discussions (solid curve); this latter increase, in turn, occurs before any violence (dotted and dashed-dotted curves). In contrast, in the case of the pope event (Figure 3, top right), the entropy of diffusion of discussions across blog communities (dashed curve) is small relative to the cartoons event, and any increase in this measure lags discussion volume (solid curve). Similar curves are obtained for the other four



Time series motif analysis

Event	Motif
Danish cartoons 1:	1/1—1/26/2006.
Egypt DVD release:	10/2–10/9/2005.
Abu Ghraib story:	none.
Pope lecture:	none.
Rushdie knighting:	none.
Danish cartoons 2:	none

Figure 3: Sample results for Islamic mobilization case study. The time series plots at the top correspond to the first Danish cartoons event (left) and the pope event (right). In each plot, the solid curve is blog volume and the dashed curve is blog entropy; the Danish cartoon plot also shows two measures of violence (dotted and dashed-dotted curves). The table at the bottom summarizes the results of the motif analysis study.

events (but are not shown because of space limitations). More importantly, the proposed motif detection process also yields the expected result: motifs are found only for the Danish cartoons and Egypt DVD events (Figure 3, bottom), and these motifs precede significant blog volume and real world violence. This case study suggests that early diffusion of mobilization-related activity (blog discussions) across disparate social contexts (blog communities) may be a useful early indicator of successful mobilization events.

Acknowledgements

The research described in this paper was sponsored by the Department of Homeland Security and by Sandia National Laboratories through Sandia’s Laboratory Directed Research and Development Program.

References

Arthur, W. 1989. Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal* 99: 116-131.

Bemporad, A., Bicchi, A. and Buttazzo, G. eds. 2007. In *Proc. 10th Int. Conf. on Hybrid Systems: Computation and Control,* Pisa, Italy.

Colbaugh, R., Glass, K. and Willard, G. 2008. Analysis of complex networks using aggressive abstraction. SAND 2008-7327, Sandia National Laboratories.

Hedstrom, P., Sandell, R. and Stern, C. 2000. Mesolevel networks and the diffusion of social movements: The case of the Swedish Social Democratic Party. *American Journal of Sociology* 106: 145-172.

Kushner, H. 1967. *Stochastic Stability and Control*. New York: Academic Press.

Leskovec, J., Adamic, L. and Huberman, B. 2006. The dynamics of viral marketing. In *Proc. 7th ACM Conf. Electronic Commerce*, Ann Arbor, MI.

Newman, M. 2003. The structure and function of complex networks. *SIAM Review* 45:167-256.

Parrilo, P. 2000. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD dissertation, California Institute of Technology.

Salganik, M., Dodds, P. and Watts, D. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311: 854-856.

Shiller, R. 2000. *Irrational Exuberance.* New Jersey: Princeton University Press.

Sontag, E. 1998. *Mathematical Control Theory* Second Edition. New York: Springer.

http://www.cds.caltech.edu/sostools/. 2007.

Walls, W. 2005. Modeling movie success when ‘nobody knows anything’: Conditional stable-distribution analysis of film returns. *J. Cultural Economics* 29: 177-190.