Paying attention to each other in visible work communities: Modeling bursty systems of multiple activity streams

Jamie F. Olson
Institute for Software Research
Carnegie Mellon University
Pittsburgh, PA 15214
jfolson@cs.cmu.edu

James Howison
Institute for Software Research
Carnegie Mellon University
Pittsburgh, PA 15214
jhowison@cs.cmu.edu

Kathleen Carley
Institute for Software Research
Carnegie Mellon University
Pittsburgh, PA 15214
kathleen.carley@cs.cmu.edu

ABSTRACT

Online work projects, from open source to wikipedia, have emerged as an important phenomenon. These communities offer exciting opportunities to investigate social processes because they leave traces of their activity over time. Unlike traditional work teams, the participants in these communities are widely dispersed and work without centralized management. The question arises, then, as to the extent to which these are in fact communities: is the group simply the sum of the individuals that make it up, or does the group function as a social unit? We explore this question in the temporal domain.

We argue that the rapid visibility of others' work afforded by the information systems used by these projects reaches out and attracts the attention of others who are peripherally aware of the group's online space, prompting them to begin or intensify their participation, binding separate individual streams of activity into a social entity.

Previous work has suggested that for certain types of bursty social behavior (e.g. email), the frequency of the behavior is not homogeneously distributed but rather can be divided into two generative mechanisms: active sessions and passive background participation. We extend this work for the case of multiple conditionally independent streams of behavior, where each stream is generated by these two generative mechanisms. Our model can characterized by a double-chain hidden markov model, allowing efficient inference using expectation-maximization. We apply this model to visible work communities by modeling each participant as a single stream of behavior, assessing transition probabilities between active sessions of different participants. This allows us to examine the extent to which the various members of the community are influenced by the active participation of others. Our results indicate that an active session by a participant at least triples the likelihood of another participant beginning an active session.

SIGKDD 2010 Washington, DC USA

1. INTRODUCTION

Novel information technologies have given rise to new organizational forms, such as open source and Wikipedia. Understanding these organizations is an important challenge both because they are increasingly important in the world and for the lessons they might provide about human behavior generally. In addition better understandings may help these projects as they seek to sustain their success.

They are interesting to study, in part, because their success is a puzzle. It is a puzzle because they work in conditions expected to be problematic for collaboration: they are high distributed, they have little control over their potential participants and they work in complex domains of work. These organizational forms are diverse, but share two distinguishing characteristics of particular interest for this paper: their participants are largely volunteers and, as such, have to be attracted to work on the projects. Secondly, their work is made public through information technologies.

Volunteer participation immediately raises the question of motivation, which has been extensively studied [13, 14, 8, 7, 23, 4, 10]. This work cites motivations like a) the product itself, b) learning by doing, c) intellectual stimulation, d) self-efficacy, e) building reputations and f) learning from others. This work speaks of motivations in a very general form and is almost always based on surveys or interviews [5]. The motivations share the characteristic that they are all very rational explanations of behavior and we believe that the possibility exists that these explanations are influenced by post-hoc justifications of the time spent on these projects. In any case they don't explain an interesting question: why do participants choose to participate at a particular time, and what might that tell us about the interesting success of these new organizational forms?

Four years of participant observation in an open source project, reported in [9], illuminated two as yet under-explored aspects of participation in these projects. The first, at risk of some obviousness, is that it occurs in and around one's real life context. Participation is only possible when one is awake and at one's networked computer. Even at one's computer, participation occurs in and around other activities, such as paid work, school work, socializing, web browsing, internet banking and so on. In addition, the material environment is quite noisy, with potential participants frequently shifting between applications and peripherally attending to multiple communication channels, such as open email clients, IRC chat sessions, Facebook notifications and so on. This dynamic environment means that the project is competing for attention with these other things: a potential participant's broad motivations must be converted into specific contribution actions undertaken at a specific time.

A potential participant's attention can be drawn to the project and its work in different ways. A first is that they may wish to accomplish something in particular and *intend* to focus their attention primarily on the project's work. A second is that their attention may be drawn by intersection with the project as an *incidental consequence* of other activities, such as when one is using a reference manager to write a paper and finds an annoyance, or when one is researching a city for a trip and reads a Wikipedia page.

These two mechanisms appear to relate to the temporal patterns of work in these communities. We observe (and provide evidence below) that there are two main such patterns of work: small numbers of events sporadically dispersed in time and short periods of high intensity, where many events come quickly, which we call active sessions. Sporadic participation seems likely to result from attention being drawn to the project, but not held for a substantial period of time, instead returning to other competing activities, such as firing off a quick bug-report while writing an academic paper. Active sessions, on the other hand might be generated by intentional, focused work, where participants are pursuing specific goals.

Both these mechanisms, however, are driven by individual causes, meaning that participants incidental participation and active sessions would not respond to other participants. The project would, in effect, be an accretion of individual work, simply a place for individuals to work, rather than a social entity or a collective project.

Careful consideration of the attentional environment of potential participants suggests a third attentional mechanism: that of visible work. By this we mean that the work of other participants becomes visible to a potential participant, drawing their attention towards the project. If the changes were only visible if the potential participant were to specifically visit the project pages (pull), then this would not be an attention drawing mechanism on its own. However many systems have channels which push visible activity by others directly to participants (push), such as email lists, IRC channels, RSS feeds and watchlists, which inform potential participants when changes are made to project venues. Thus work by others "reaches out" quickly to potential participants and has the ability to attract their attention to the project.

We argue that this socio-technical characteristic of visible work systems means that participants are likely to synchronize their attention to the project. Not only does such visible work attract attention but it also signals that other participants are awake, online and paying attention to the project (This point has been made in the "social presence" literature, particularly in the context of distance learning, e.g. [11, 20, 19]). This seems likely to create a particularly likely time for a participant to turn their primary attention to the project, taking them out of their incidental work pattern and into an active session. Once multiple participants are actively working, they may continue to prompt each other to continue to work, through obvious mechanisms like directly talking, asking and answering questions, solving problems which would otherwise have blocked work and seen a participant turn their attention to other non-project work. Other, less obvious, mechanisms might also operate, such as being annoyed by another participant's work (see Figure 1) and working to correct it, or working to demonstrate to another who has answered a question that their effort was worthwhile, thereby demonstrating your competence. In short, the mechanism of visible work, might operate to move a project from a collection of individuals towards a social entity, where the individual's work patterns are affected by and relate to those of others.

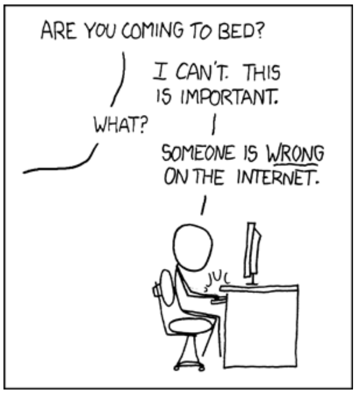


Figure 1: "Duty Calls" XKCD provides a humorous look at social attention extracting more time than potential participants might otherwise have contributed. Reprinted with permission, see http://xkcd.com/about/

Such a mechanism might help to explain the interesting success of these new organizational forms, who seem able to attract participants more effectively than offline volunteer work, which has historically struggled. Prior to the information technologies that make visible work possible, voluntary work tended to require participants to be sufficiently motivated that they would plan and set aside specific time to volunteer. Since the volunteer needed to be physically present this tended to require long periods of contiguous time. Visible work systems mean not only that small periods of "free" time can be harnessed, but that projects can actively attract time away from planned activities, bringing participants to contribute more than they might rationally have planned for. Lest this sound too sinister, it is worth noting that many participants report finding work in these communities a welcome relief from other activities. In the words of a participant in an open source project to participant who had just asked a question,

> *No problem. I was grappling with authentication and welcomed the distraction ;-)*[1]

Motivated by this theory, this paper builds a model to analyze temporal data from Wikipedia a leading example of a visible work community. The overall intention of the model is to explore the proposition that participant's temporal work patterns are responsive to one another. We first examine the issues in modeling human behavior in the temporal domain, identifying an existing model as a base and extending it to match our domain more closely. Our extensions are described in detail. We then turn to our dataset,

[1] http://tech.groups.yahoo.com/group/jena-dev/message/42760

providing relevant contextual background on Wikipedia and descriptive statistics on temporal patterns overall and for individual participants. We then present the results of the model and discuss their interpretation and limitations in the context of the theory above. Finally we conclude and discuss appropriate future work, both for improving our exploration of this theory and alternative applications of the new model introduced in this paper.

2. MODELING TEMPORAL BEHAVIOR

A large body of work suggests that many human behaviors are heavy-tailed and bursty in the temporal domain [2, 21, 17, 1, 6, 18, 22, 15]. They disagree, however, in their explanation of these properties. Some propose a priority-queue where individuals choose high priority tasks over low priority tasks [2, 21], similar to preferential attachment in network evolution [3]. Although these models reproduce several of the aforementioned heavy-tail and burstiness of human temporal behavior, they are inconsistent with several important properties of real world human behavior, notably circadian rhythyms and infrequent "sessions" of high activity [17, 1]. Recently, certain nonhomogeneous Poisson processes were shown to be able to replicate the same heavy-tail and burstiness [17]. The nature of these cascading Poisson processes allow researchers to include mechanisms like "session" and circadian rhythm directly in the model.

Our work is heavily influenced by the model proposed by Malmgren and colleagues [17], which they subsequently simplified [16]. They propose a Markov mixture of Poisson processes cast as a double-chain hidden Markov model. Specifically, they use a mixture of two Poisson processes, represented by the hidden states in their double-chain hidden Markov model. When in the active state, events are generated by a homogeneous Poisson process with rate ρ_a. In the passive state, events are generated by a nonhomogeneous Poisson process with rate $\rho_p(t)$ that depends on the current time. The passive state is intended to represent a simple version of circadian rhythyms and is defined by two square pulse distributions p_d, p_w and a rate parameter ρ_0.

$$\rho_p(t) = \rho_{p0} W p_d(t|\tau_{d0}, \tau_{d1}, \epsilon_d) p_w(t|\tau_{w0}, \tau_{w1}, \epsilon_w) \quad (1)$$

Where the square pulse distribution with period τ is defined as

$$p(t|\vec{\tau}, \epsilon) = \begin{cases} w & (t \text{ modulo } \tau) \in [\tau_0, \tau_1) \\ \epsilon w & \text{otherwise} \end{cases} \quad (2)$$

$$w = (\epsilon\tau + (1-\epsilon)(\tau_1 - \tau_0))^{-1} \quad (3)$$

such that probability density between τ_0 and τ_1 is elevated relative to rest ($\epsilon < 1$). In order to represent circadian rhythyms, p_d represents the activity during the hours of the day with $\tau_d = 24$ and p_w represents elevated activity during some portion of the week($\tau_w = 7$).

The EM algorithm is used to jointly estimate the hidden states and the parameters, $\theta = \{\varepsilon, \rho_a, \rho_{p0}, \tau_{p0}, \tau_{p1}, \epsilon_d, \tau_{w0}, \tau_{w1}, \epsilon_w\}$. The complexity of these nonhomogeneous Poisson processes means that direct update formulas for the M-step are not available. However, because the likelihoods are convex with respect to the parameters, Powell's method can be used to obtain maximum likelihood estimates for the parameters.

3. MODEL

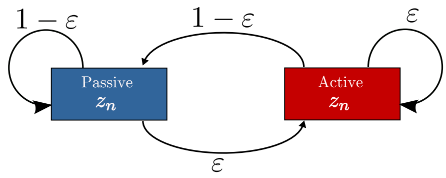


Figure 2: A single activity stream is characterized by two states: Active and Passive

We extend the work of [16] to model multiple streams of activity. The goal is to model interactions between these streams while preserving the computational properties of the DCHMM. We assume that these K streams are not independent but are conditionally independent given their state information. We use the two-state DCHMM but repeat that chain for each activity stream such that each stream is conditionally independent of all other streams given its current state.

$$T_k \perp T_j | Z_k \quad (4)$$

For each stream k we also repeat the distributions over emission probabilities, with $Z_k = 1$ indicating that stream k is in an active session and $Z_k = 0$ indicating that the stream is latent.

$$T_k | Z_k = 1 \sim \text{Poisson}(\rho_{ka}) \quad (5)$$

$$T_k | Z_k = 0 \sim \text{Poisson}(\rho_{kp}(t)) \quad (6)$$

Again, the rate, $\rho_{kp}(t)$ for each latent nonhomogeneous Poisson process is governed by the square pulse distributions.

$$p_k(t|\vec{\tau}_k, \epsilon_k) = \begin{cases} w_k & (t \text{ modulo } \tau) \in [\tau_{k0}, \tau_{k1}) \\ \epsilon w_k & \text{otherwise} \end{cases} \quad (7)$$

$$w_k = (\epsilon_k\tau + (1-\epsilon_k)(\tau_{k1} - \tau_{k0}))^{-1} \quad (8)$$

Because Poisson processes are memoryless (equation 10), we can easily construct likelihoods for each observation in the K streams. For an observation, o_n occuring at time t_n in activity stream s_n, the likelihood of the system being in each state is shown in equation 11.

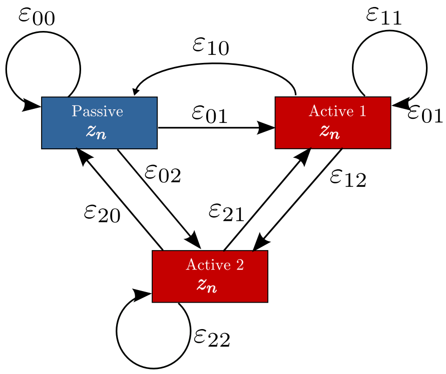
$$T \sim \text{Poisson}(\lambda) \quad (9)$$

$$P(T = t) = P(T < t_0)\, P(T = (t - t_0)) \qquad \forall t_0 < t \quad (10)$$

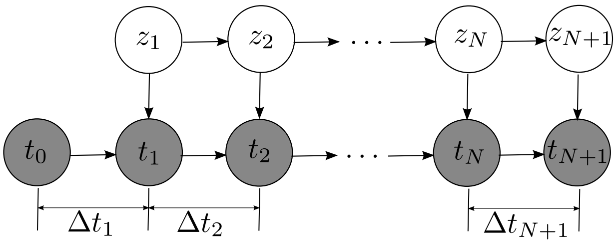
$$P(o_n | Z_n^* = k) = P(T_{s_n} = t_n | Z_n^*) \prod_{l \neq s_n} P(T_l > t_n) \quad (11)$$

In general, this leads to an explosion in the state space of the overall system, with 2^K states. This is compounded by the needed to learn transition probabilities between these states. We choose instead to limit configurations of the system to those where at most one of the K streams is active. This leads to $K+1$ states for the system Z^*, with the total number of parameters $|\vec{\theta}|$ growing with square of the number of activity streams. This limits the model to scenarios where only one activity stream can be active at each event. We discuss limitations deriving from this decision below.

$$Z_k = \begin{cases} 1 & Z^* = k \\ 0 & \text{else} \end{cases} \quad (12)$$



(a) State transitions



(b) Double-chain Hidden Markov Model

Figure 3: An example of the proposed model for two streams of events showing 3(a)the event-generating states and 3(b)the DCHMM that can represent this model.

4. DATA

We applied this model to a WikiProject in Wikipedia. A WikiProject is a group of Wikipedians who work to improve a section of Wikipedia. Projects include topics such as Music, Sports and Geographical regions. The Project identifies Articles which are considered "in scope" and tags them as associated with the WikiProject. We assume that participants are observing portions of these pages in their Watchlists, and discussions of work done for the WikiProject in the Project's collaboration pages. For this reason we argue that it is reasonable to believe that participants are aware of when others are actively working, making this an example of visible online work.

We chose to study a geographical WikiProject, Project Oregon, since that gives some confidence that participants are in the same time-zone. The Project Oregon "About Us" page states that it was founded in March 2005 and "experienced a lot of growth in late 2007 and 2008."

We accessed a March 12, 2008 dump of English Wikipedia and downloaded all revisions to Articles, and their associated Talk pages, marked as in scope for the Project. A datapoint consists of a user-id and a timestamp; we do not use data about which specific page was edited, since we wanted to capture the idea that participants could also be motivated to work on nearby pages, or indeed anywhere else in the Project's scope.

Overall the dataset consists of 354,793 revision events by 25,780 different users and 5622 articles. Because we seek to model interactions within a community, we limit ourselves to those we define as community members. In this case, we consider a participant to be a member of WikiProject Oregon if over the history of the project they engage in Talk activities and Edit activities at least 100 times. This reduced the dataset to 55,104 (15% of the total) revisions and 24 users (0.09%) (a typical skew in contribution is typical). We broke the dataset down by year, in part to minimize the impact of annual cycles (see above) and in part to assess change within the Project. Table 1 shows some aggregate information regarding the 3 periods studied.

Year	Total Revisions	Total Active Users
2006	12,126	17
2007	34,181	23
2008 (partial)	7,700	21

Table 1: Aggregate statistics for the years examined

The collected data supports the observations made above regarding temporal patterns in human behavior. Figre 4 shows the distribution of interevent times for some of the users in log-log scale. Although it is by no means scale-free, most users have heavy-tailed inter-event distributions, often going for long periods of time between edits. The inter-event distribution suggests, and qualitative examination of the user event streams (shown in Figure 5) confirms that there is substantial periodicity in the temporal patterns of edit behavior. This further motivates the cascading nonhomogeneous Poisson process used in the model.

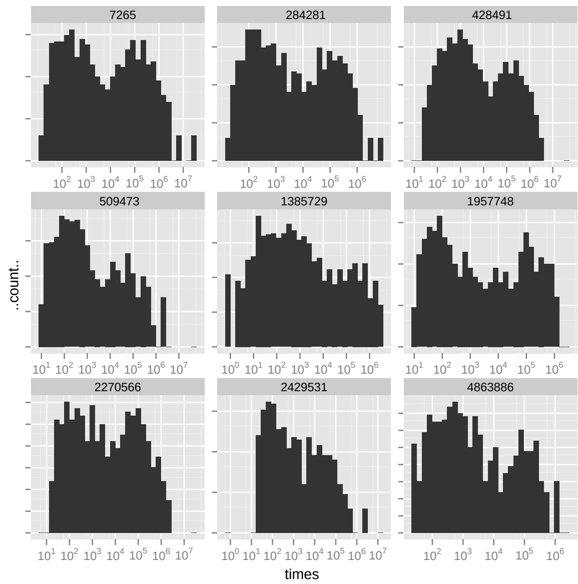


Figure 4: Distribution(Log-Log scaled) of inter-event times for some representative users

5. RESULTS

We apply our model to the WikiProject Oregon data described previously. The outputs of the model are parameters associated with the temporal distribution of revisions and a transition matrix, showing the likelihood estimates for transitions between the state of the system.

Figures 5-8 shows the parameters learned for temporal distribution of revisions across the three years. Most of the

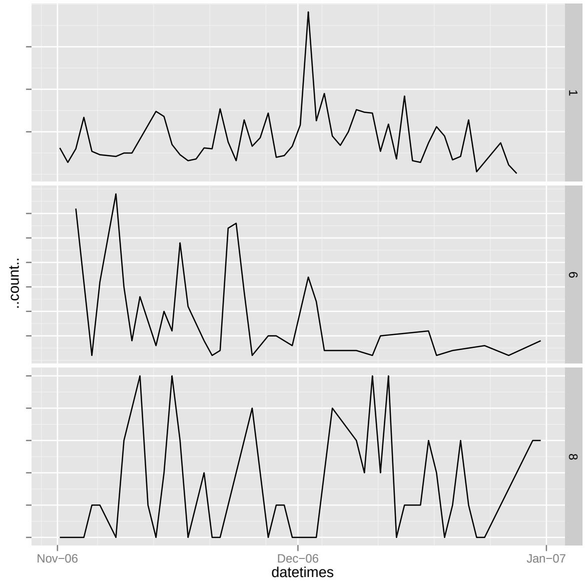


Figure 5: Temporal periodicity in selected users during November and December of 2006

revision parameters are quite stable across the three years. This is consistent with the results of Malmgren [16] which found consistent temporal patterns in emails. Worth noting is that the user population generally begins their WikiProject "wiki-week" on Thursday and works over the weekend.

The transition matrix shows transition probabilities between the states of the model; there are $K + 1$ states, one indicating an active session for each participant and one indicating that no participant is currently in an active state (system passivity). For each event (i.e. a revision) the model estimates the probability that the system is in each state.

Two types of transitions are are particularly relevant to the social theory motivating this study: the first is the probability of a transition to system passivity, given an active session (by any participant), informally written as P(Active|No Active). The second is the probability of a transition to an active session of any participant, given an active session by a different participant, informally written as P(Active|Active). A t-test comparing these transition probabilities, (means shown in Table 2), indicates that in all years P(Active|Active) is significantly greater than P(Active|No Active).

We can also construct a network based on the state transition matrix. Using the probability of any user spontaneously becoming active we dichotomize the state transition network, removing all transitions that have a probability less than twice the baseline P(Active|No Active). Using the 2007 parameters, this results in the sparse network shown in Figure 5.

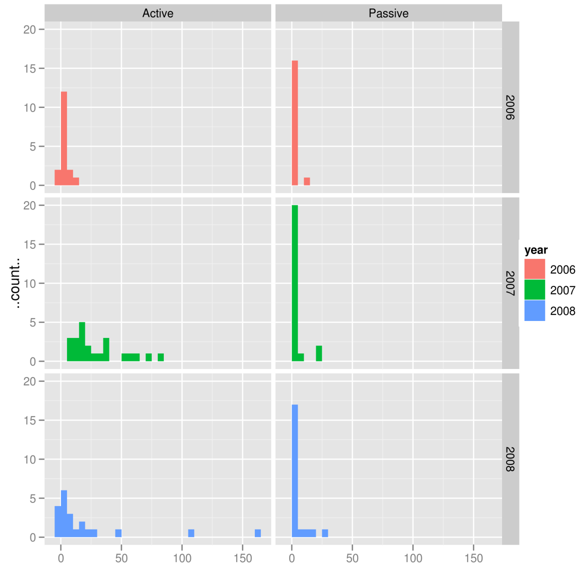
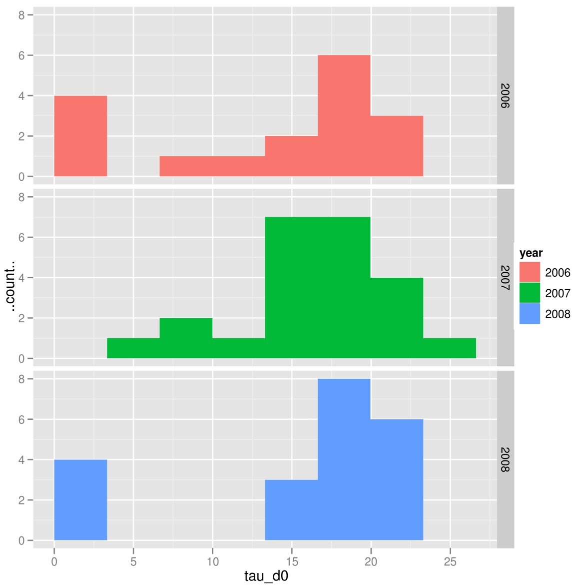
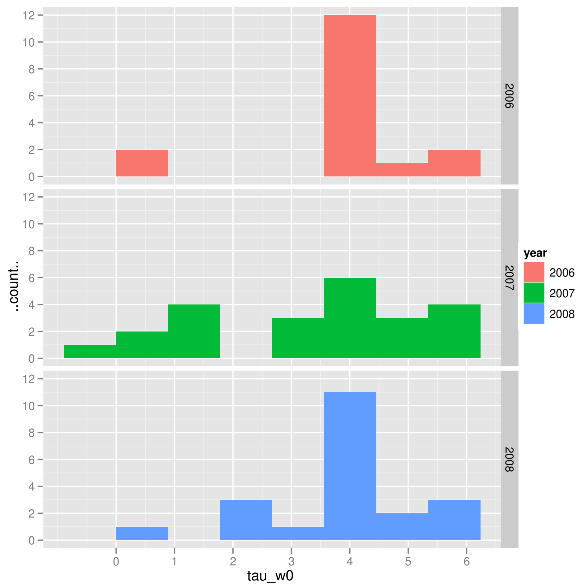


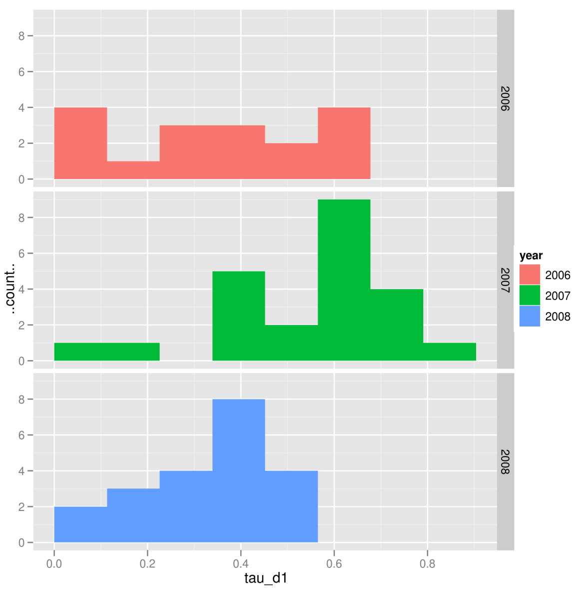
Figure 6: Active(ρ_a) and Passve(ρ_p) rate per hour



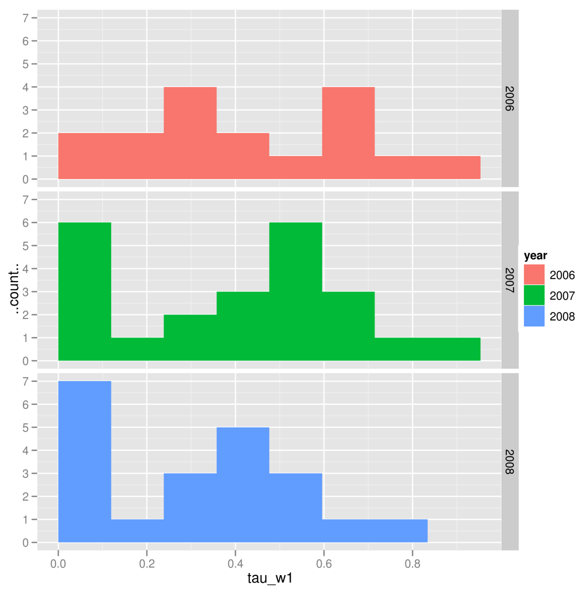
(a) Begin Day(τ_{d0})



(a) Begin Week(τ_{w0})



(b) Day Length(τ_{d1})



(b) Week Length(τ_{w1})

Figure 7: Elevated activity "wiki-day"

Figure 8: Elevated activity "wiki-week"

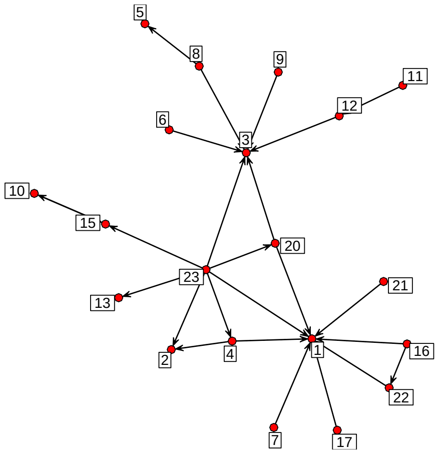


Figure 9: Potential influence network for 2007

Year	P(Active|Active)	P(Active|No Active)
2006	0.0195	0.004
2007	0.0385	0.012
2008	0.0271	0.002

Table 2: The conditional probabilities learned for each year

6. DISCUSSION

The central question of this paper is the extent to which participants' temporal work patterns are responsive to one another. Table 2 and the associated hypothesis test give confidence that transitions between active sessions of different participants are significantly more likely than transitions to an active session when no one is active. These results also indicate a substantial effect size, suggesting that an active session at least triples the likelihood of another user beginning an active session. We interpret this as evidence that visible activity by WikiProject members increases the probability that other members will begin an active session of work. The magnitude of this effect appears to be stronger in 2007 and 2008 than in 2006; this matches the statement on the WikiProject homepage that they became more active and organized beginning in 2007.

There is also evidence that some participants are more likely to active sessions in others. Figure 5 can be interpreted as an implied influence network. Node 23 has an outdegree of 7, meaning that their visible work may capture the attention of others and motivate their transition into active participation. Nodes 1 and 3 have high indegree meaning that their active sessions tend to follow those of the most others, suggesting that they attend to the visible work of a large number of WikiProject Oregon participants.

7. LIMITATIONS

There are two key limitations of this study, one relating to the model, one relating to our interpretation of the results.

The model constrains the system such that only one activity stream can be active at each event, restricting its ability to model simultaneous bursty work. In situations where multiple activity streams are in fact in simultaneous sessions, this model will instead identify numerous transitions back and forth between the active sessions. This will be reflected in increased stream transitions probabilities for overlapping periods. This means that the model cannot distinguish between multiple simultaneous activity sessions and a sequence of non-overlapping activity sessions. We do not believe that this threatens the overall result of the paper, since both patterns are signs of responsiveness between participants.

Furthermore, the Markov assumption in this limited state space means that transitions are now effectively conditioned on the single activity stream that was in an active session (or the latent state) rather than the set of previous session information. For relatively dense streams of activity, this may be problematic. If stream A undergoes an active session which leads to an active session in B, but a third irrelevant stream, C, is active in the time between when A stops and when B begins, the relationship between A and B will not be correctly inferred. Given the asynchronous capabilities of the information system this means we are unable to capture all sources of responsiveness, just those that are immediate and direct.

We have interpreted the social synchronization of participants indicated by the transition probabilities to indicate the operation of the attention effect theorized in the introduction to this paper; that is endogenously to the social system. However it is possible that attention is drawn to the project exogenously, through events occurring in the world and being reported on in the news or blogs. Such events are known to produces bursts of activity in traditional media [12] and so seem particularly likely to affect articles about entities currently in the news. This effect is particularly seen in biographies of living persons, where Wikipedia has most often had to institute temporary locks on article editing, due to a flood of participants attracted by the currency of the articles topic. There is no reason to believe that Project Oregon's scope would be particularly susceptible to this effect, but it would be great to extend the modeling framework to identify effects from periodic newsworthy events.

8. CONCLUSIONS AND FUTURE WORK

This paper argues that one of the sources of the surprising success for novel socio-technical organizational forms like open source software development and Wikipedia, is their ability to successfully compete for potential participant's moment-to-moment attention, and one of the reasons that they are able to do so is the social pull resulting from the visibility of other's work. This mechanism might explain why the experience of work in these projects is social, and the projects described as social communities, rather than simply aggregations of individual work.

Throughout the paper we have presented this effect as attracting more participation and suggested that is a reason for the surprising success of visible work communities. It is clear, however, that that having too many participants descend on a page at once is likely to lead to coordination problems. In Wikipedia social responsiveness has a negative side too, as participant's attention can be drawn by opposing perspectives, descending into edit wars. It may be that the mechanism of motivation and attracting investigated in this paper has an inflection point, above which its effects become negative, prompting system administrators to dampen its effects by instituting temporary edit locks, slowing down the stream of attention grabbing edits.

The model developed in this paper should have usefulness beyond the application presented in this paper. In particular it seems well suited to investigating situations where the system is constrained to be in a single state. An example of where this might prove useful is in modeling individual's attention flows throughout the day, perhaps making use of data collecting by agents on people's computers.

9. REFERENCES

[1] L. Amaral, D. B. Stouffer, and R. D. Malmgren. Log-normal statistics in e-mail communication patterns. pages 1–18, 2006.

[2] A.-L. Barabási. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(May):207–211, 2005.

[3] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509, 1999.

[4] P. O. Chin and D. Cooke. Satisfaction and coordination in virtual communities. In *Proceedings of the Tenth Americas Conference on Information Systems*, New York, New York, 2004.

[5] K. Crowston, K. Wei, J. Howison, and A. Wiggins. Free/libre open source software development: What we know and what we do not know. Technical report, Syracuse University FLOSS Project, 2008.

[6] F. Deschatres and D. Sornette. Dynamics of book sales: Endogenous versus exogenous shocks in complex networks. *Physical Review E*, 72(1):16112, 2005.

[7] R. A. Ghosh, G. Robles, and R. Glott. Free/libre and open source software: Survey and study floss. Technical report, International Institute of Infonomics, University of Maastricht: Netherlands, 2002.

[8] A. Hars and S. Ou. Working for free? Motivations of participating in FOSS projects. *International Journal of Electronic Commerce*, 6(3):25–39, 2002.

[9] J. Howison. *Alone Together: A socio-technical theory of motivation, coordination and collaboration technologies in organizing for free and open source software development.* PhD thesis, Syracuse University, School of Information Studies, 2009.

[10] W. Ke and P. Zhang. Participating in open source software projects: The role of empowerment. In *Proceedings of ICIS08 HCI Workshop*, 2008.

[11] B. Kehrwald. Understanding social presence in text-based online learning environments. *Distance Education*, 29(1):89–106, 2008.

[12] J. Kleinberg. Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, 7(4):373–397, 2003.

[13] K. Lakhani and E. von Hippel. How open source software works: "free" user-to-user assistance. *Research Policy*, 32:923–943, 2003.

[14] K. Lakhani and R. Wolf. Why hackers do what they do: Understanding motivation efforts in Free/F/OSS projects. Working Paper 4425-03, MIT Sloan School of Management, 2003.

[15] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins. Microscopic Evolution of Social Networks. In *Proceedings of the ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 2008.

[16] R. D. Malmgren, J. M. Hofman, and D. J. Watts. Characterizing Individual Communication Patterns. In *Proceedings of the ACM SIGKDD international conference on Knowledge discovery and data mining*, volume 15, pages 607–615. ACM, 2009.

[17] R. D. Malmgren, D. B. Stouffer, and A. E. Motter. A Poissonian explanation for heavy tails in e-mail communication. *Proceedings of the National Academy of Sciences*, 105(47):18153–18158, 2008.

[18] K. Naruse and M. Kubo. Lognormal Distribution of BBS Articles and its Social and Generative Mechanism. *Web Intelligence*, 2006.

[19] L. Rourke, T. Anderson, D. Garrison, and W. Archer. Assessing social presence in asynchronous text-based computer conferencing. *Journal of distance education*, 14(2):50–71, 1999.

[20] J. Short, E. Williams, and B. Christie. *The social psychology of telecommunications.* John Wiley Sons., New York, 1976.

[21] A. Vázquez, J. a. G. Oliveira, Z. Dezsö, K. I. Goh, I. Kondor, and A.-L. Barabási. Modeling bursts and heavy tails in human dynamics. *Physical Review E*, 73(036127):1–19, 2006.

[22] M. Vlachos, C. Meek, Z. Vagena, and D. Gunopulos. Identifying similarities, periodicities and bursts for online search queries. *International Conference on Management of Data*, 2004.

[23] Y. Ye and K. Kishida. Toward an understanding of the motivation of open source software developers. In *Proceedings of 2003 International Conference on Software Engineering (ICSE)*, Portland, Oregon, USA, May 3-10 2003.