Modeling Polarization Dynamics in Online Communities

Hazhir Rahmandad, Virginia Tech, hazhir@vt.edu

Mohammad Mahdian, Yahoo! Research, Mahdian@alum.mit.edu

Abstract

The advent of Internet and other communication technologies has drastically increased the volume of communication in human societies. One might hope that increased communication will lead to a higher degree of mutual understanding and resolution of conflicts. An opposing, and somewhat counter-intuitive, point of view is that the reduction in the cost of communication would make it easier for people to interact with other like-minded individuals despite geographic distance, thereby polarizing the society. In this paper, we study some of the basic dynamics underlying this problem. We develop an agent-based model to capture the dynamics of an online community where agents post stories and read and vote on others' stories. We show that different combinations of parameters can lead to different macro-level behavioral modes in this model, and give anecdotal evidence from a large online community to support the predictions of the model. In particular, we identify four types of communities based on their dynamics: majority dominated, competitively polarized, converged, and diversified. We discuss the implications of each of these forms on the social welfare and the stability of the community.

1- Introduction and Motivation

Internet plays an increasing role in communication, knowledge creation, and information consumption, thus understanding its role in homogenizing or polarizing society becomes critical. Theoretical considerations provide two alternative views on this issue. One perspective suggests that in the digital age communication cost reductions break the barriers to interaction among different people, leading to a global village (McLuhan 1994) where location no more controls interaction (Toffler 1980; Negroponte 1995). Internet is expected to boost democracy (Westen 1998) as it reduces power asymmetry between government and citizens (Rheingold 1991) and allows for grassroots public dialogue (Grossman 1995) which is at the heart of deliberative democracy (Fung and Wright 2001). By reducing the costs of organizing, internet should also empower minorities and smaller or more distributed interest groups to mobilize around issues they care about and thus accelerate pluralism in the society (Bimber 1998). At the firm level, internet is expected to increase user participation in innovation, product development, and marketing through rapid consumer feedback, virtual prototyping, online user communities, viral marketing, virtual worlds, social networking, and other web-based technologies (Prandelli, Verona et al. 2006; Buyukozkan, Baykasoglu et al. 2007). Internet could also increase the efficiency of financial markets when online communities share timely information (Antweiler and Frank 2004; Das and Chen 2007).

A second view offers different hypotheses, highlighting the internet's polarizing impact (Van Alstyne and Brynjolfsson 2005). The basic argument holds that limitations in time and information processing capacity require people to choose among different information sources (Simon 1957). Internet allows one to more easily find sub-communities that share her point of view. This promotes in-group reinforcement of similar arguments since people prefer to interact with similar-minded others (Gilovich 1991; McPherson, Smith-Lovin et al. 2001). Given limited time and capacity for information consumption and communication, this mechanism can therefore reduce interactions among people holding different opinions. The prospects of internet balkanizing public discourse could be unsettling as people's tastes evolve in their social networks: interactions with different minded people induce empathy, reduce stereotypes, and increase tolerance (Gurin, Hurtado et al. 1994; Aronson and Patnoe 1997; Boisjoly, Duncan et al. 2006). On the other hand from mock jury experiments to the formation of fringe radical movements, when similar-minded people join and interact in a group opinion polarization is likely (Sunstein 2002). Polarization of public discourse can reduce the production of public goods and increase tension and conflict (Alesina, Baqir et al. 1999). Therefore internet's impact on democracies may be more nuanced than initially envisioned, and include negative effects (Levine 2002). At the organizational level, if online communities interact based on homophily principle (Chen, Gu et al. 2008), the internet-induced communication may also hurt organizational performance as contributing individuals reinforce each other's biases rather than bringing new insights, leading to herd behavior and increased risk (Delong, Shleifer et al. 1990).

In this paper, we contribute to this discussion by proposing an agent-based model for the dynamics of interactions in an online community, and identifying different modes of behavior by simulating this model. Our model captures two types of entities: users, who each have an opinion in a one-dimensional opinion space (for example, in the domain of politics, this can correspond to the spectrum from the extreme leftist to the extreme rightist), and a level of motivation which determines how active they are in the community. The other type of entity capture stories that are posted by the users in the system. Each story corresponds to a point in the opinion space, and attracts a number of readers and a number of positive votes, based on the number of users who happen to read the story and agree with it. There are a number of feedback loops in the system: reading a story influences the opinion of an individual; seeing many stories that express an opinion similar to one's opinion increases the level of motivation; and receiving more votes increases the prominence of a story and its readership. These feedback loops lead to non-trivial behavior in the community, which we will study by simulating the model.

This simple schema captures how many online communities operate at a high-level. For example, Digg is an online community that operates almost exactly according to our model. Even communities like facebook, flickr (where stories correspond to pictures that agents post), and twitter operate very similarly to our model. Our research was particularly motivated by one Digg-like community named Balatarin. Balatarin is a Persian-language Digg-like community that has been very effective in shaping the online dialog in the domain of Iranian politics during the recent years. In this system users can post links (often news items) of their interest from different sources, and other users can come and vote on those items they like. Items with large numbers of vote (over a given threshold, which has changed over time) become hot, and appear on the first page (where as any new item goes into the new item page). Balatarin has existed for about four and half years, during which during which approximately 26,500 registered users have posted more than 1.2 Million links and have submitted more than 31 Million votes on these posts. This community attracts over 700 thousands unique monthly visitors who visit near 30 million pages.

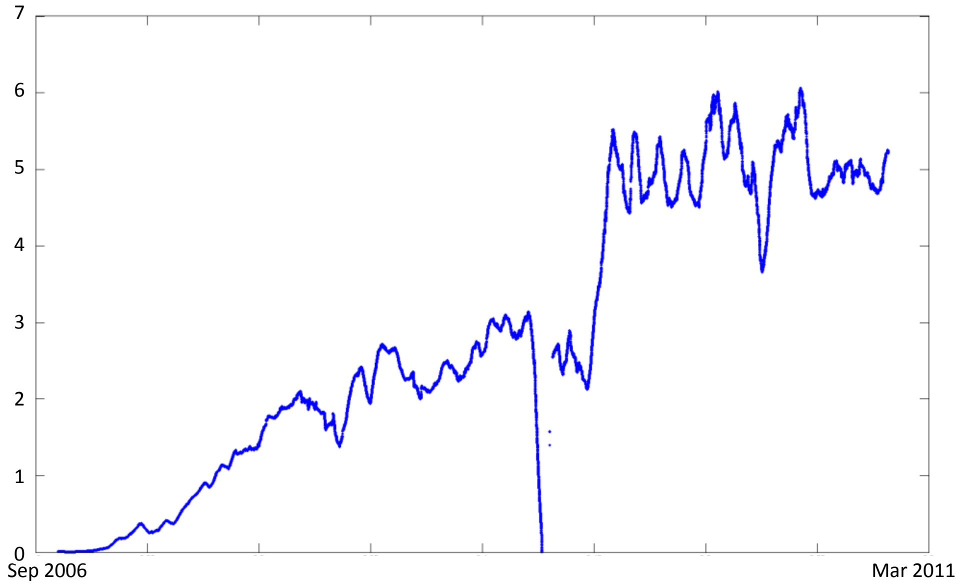

Figure 1- Number of stories posted on the Balatarin system over the 4.5 years since its inception. Y axis shows the number of stories (in 100,000s) in a two-week sliding window over time.

Figure 1 shows the plot of the number of votes as a function of time (using a sliding window with the length of two weeks). As it can be seen from the figure, the activity on the website has been increasing on average, with some noise corresponding to external political events. The large spike in the number of votes corresponds to the Iranian presidential election in 2009 and the events that followed, and the drop to zero corresponds to a time period when the website was temporarily hacked.

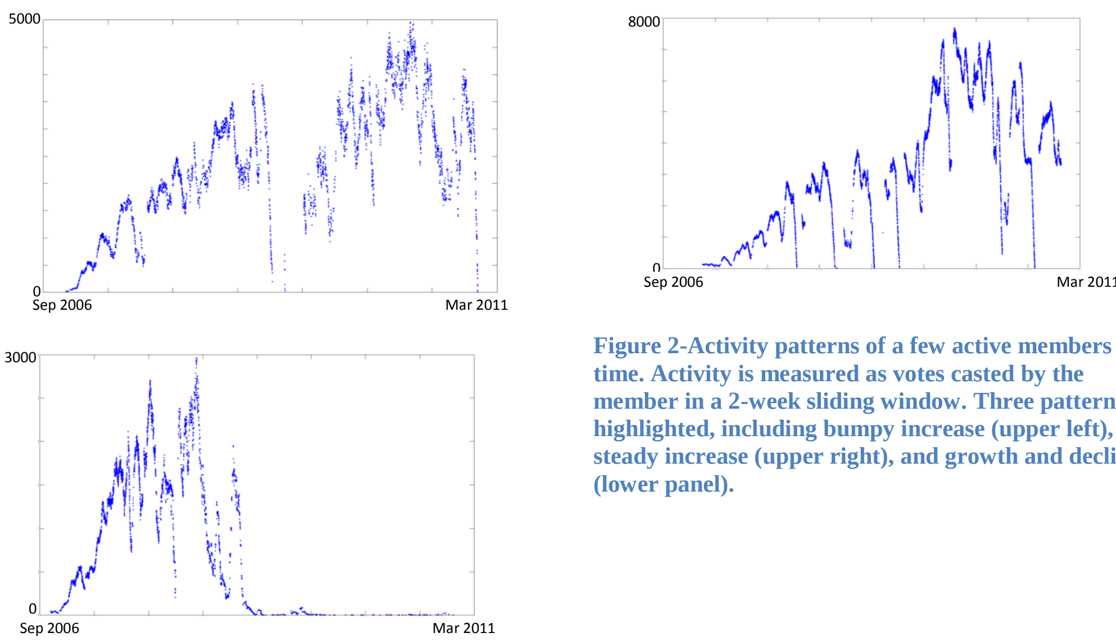

Figure 2-Activity patterns of a few active members over time. Activity is measured as votes casted by the member in a 2-week sliding window. Three patterns are highlighted, including bumpy increase (upper left), steady increase (upper right), and growth and decline (lower panel).

We also plot the number of votes submitted by a few active users, as a demonstration that the activity levels of different users have followed different trends. For example, Figure 2 shows two active users whose activity has been increasing, one with a significant decline in between (upper left panel) and one steadily (upper right panel), whereas another user (lower panel) has practically stopped contributing to the community after a couple of years even though s/he has been a major contributor earlier on. Later in the paper, we will see how the dynamics of the community can lead to such increase and decreases in the activity level of its participants. While the Balatarin case is an empirical motivation for this study, the model is not at this stage calibrated to this case and only captures some of the key feedback mechanisms we have observed in this online community.

2- Modeling and Analysis

In this section we discuss a simple dynamic model that captures some of the core features of online communities in general, and Balatarin in particular, and we use this model to explore the assumptions, boundary conditions, and parameter settings which underlie the dynamics of motivation and opinion in online communities.

2.1) Model Specification

In our modeling we focus on the dynamics of a fixed-size online community. While the dynamics of users joining and leaving a community are interesting and relevant to the problem at hand, we find the narrower model boundary useful for zooming on a few core features of problem at hand, namely the creation, consumption, and evaluation of content and the dynamics of motivation and opinion in online communities. We use an agent-based modeling framework with individuals and stories as the two main entities in our modeling work. The agent-based modeling framework allows us to identify the key feedbacks in the simulated community from following the more tangible mechanisms of interaction among distinct entities (Individuals and stories).

We have n=50 individuals who post, read, and vote for stories over a 100-period time horizon. Each individual i has an opinion (O_i) and a motivation (M_i) (both between [0-100]) which are dynamically changing[1]. We restrict the dynamics of opinion to a one-dimensional space. While in reality people typically discuss many different types of topics (e.g. politics, sport, art, economics, social issues) and have distinct opinions on each set, a one-dimensional opinion space will help us distinguish key dynamics more clearly. For example in the context of Balatarin the values on the opinion space could represent the continuum between pro-and anti- current government beliefs. The number of votes casted for story j up to this point in time is captured in the variable V_j which is the main state variable for a story. The following formulations are used to operationalize the model:

- ***Posting Stories:*** Individual i posts stories with rate R_i. That is, the posting of stories is a discrete event, which happens with intervals exponentially distributed with average time between posting of $1/R_i$. These stories have an embedded opinion of $S_j=O_i+\varepsilon_j$. Here the error ε term is drawn from a

[1] Dynamic variables are identified by upper case letters while parameters are set to be lower case.

normal distribution with standard deviation of v and is truncated to ensure S remains between 0 and 100.

- ***Reading Stories:*** Individual i reads stories with rate E_i. The individual picks randomly from the pool of currently posted stories. The reading probability for each story is proportional to V_j^y, where y is a parameter that describes the structure of story retrieval in the online community. For example, if stories with higher votes are shown on the first page (which is what happens in Balatarin, Digg, and many other environments with collaborative filtering), then those stories have a higher chance of being read and thus a higher value of y is a better descriptor of the system. A pure random selection of stories is described by y=0.
- ***Voting for Stories:*** After reading a story the individual may vote for the story with a probability that depends on how far the individual's personal opinion is from the story. We assume that individuals do not vote for stories that are further than a threshold (t) from their own opinion ($t>|O_i-S_j|$) and the probability of voting linearly increases as the story gets closer to the individual's opinion, up-to a maximum probability of x. Formally:

  $$prob(voting) = Max(0, x.\frac{t-|Oi-Sj|}{t}) \qquad (1)$$

  The threshold parameter t is important in our analysis and represents the openness of population members to hearing and appreciating opinions far from their own.
- ***Impact of Reading on Opinion:*** A basic finding of social influence theory is that people's opinion could be influenced by the opinions they are exposed to (Cialdini 2009). Therefore reading stories could have an impact on the opinion of the reader. However, if the story is not seen as credible, it does not impact the reader's opinion. We capture these mechanisms by assuming that individuals opinion is updated towards the read stories if those stories are closer to the current opinion than the threshold t, otherwise the opinion is not influenced by reading the story. Formally:

  $$O_{i,a} = (1-s).O_{i,b} + s.S_j \text{ if } |Oi-Sj| > t \qquad (2)$$

  Here the parameter s represents the sensitivity of individual to influence from stories read and subscripts a and b represent variable values after and before the transaction of interest (reading stories in this case). We assume initially individuals are uniformly distributed on the opinion space.
- **Natural Change in Motivation:** Individual's motivation (M) is assumed to include, among other determinants, natural rates of accumulation and depreciation that are independent of other dynamics in the model. These could represent factors dependent on the social and natural environment, availability of time, and personal characteristics of the individual. We assume that each period a fixed amount (a) of motivation is replenished, and a fraction of motivation is depreciated with motivation half-life of d. In this and all other equations regulating motivation dynamics, first order controls are in place to make sure motivation remains between 0 and 100.

  $$M_i(t+1) = M_i(t) + a - M_i(t)/d \qquad (3)$$

  We assume all individuals start with the same initial motivation of a.d, which is the steady state motivation level if only natural change factors existed and is parameterized to remain between 0 and 100.
- **Impact of Reading on Motivation:** A common observation in our study of Balatarin is that users enjoy reading stories that align with their opinion, but get demotivated when stories contrary to their opinion are dominant. We capture both these effects by assuming motivation to increase/decrease if stories are closer/further than t to the opinion held by the reader. The strength of this feedback is moderated by a parameter (c) and the shape of the function by parameter w:

$$M_{i,a} = M_{i,b} + \left(1 - M_{i,b}/100\right).Max\left(0, M_{i,b}/100\right).c.\left(\frac{t - |O_i - S_j|}{Max(100 - t, t)}\right)^w \qquad (4)$$

Here the terms $\left(100 - M_{i,b}/100\right)$ and $Max\left(0, M_{i,b}/100\right)$ are first order controls that keep the motivation in the feasible range (0-100). The power function is normalized by Max(100-t,t) so that it remains under 1 in all cases; as a result parameter c determines the maximum change in motivation possible due to a reading. Also note that the sign of $(t - |O_i - S_j|)$ expression is conserved for the calculations of the power function so both negative and positive impacts are possible.

- **Impact of Posting on Motivation:** Cognitive dissonance theory, one of the tenants of social psychology, holds that through changing their opinions, motivations, justifications, etc, individuals try to reduce the dissonance caused by holding conflicting ideas (Festinger 1957). One of the empirical manifestations of cognitive dissonance theory is *effort justification* in which individuals reduce their cognitive dissonance from spending time on useless activities by increasing their internal interest in those activities (Aronson and Mills 1959). In our setting, we hypothesize such effect could indeed increase individual's interest in the online community when they spend time posting stories there. We capture this potential impact through a fractional (*h*) increase in motivation as a result of each posting:

$$M_{i,a} = M_{i,b} + h * (100 - M_{i,b})/100 \qquad (5)$$

- **Impact of Motivation on Posting and Reading:** Finally, the reading and posting of stories depends on the motivation of individual. Higher/lower motivation increases/reduces the reading and posting rates following the equations below. Parameters p and k determine the sensitivity of posting/reading to motivation:

$$R_i = r * (M_i/100)^p \qquad (6)$$

$$E_i = e * (M_i/100)^k \qquad (7)$$

- **Removal of Stories**: Stories do not stay in the system forever; rather, they are removed after an exponentially distributed life time with average life of *f*.

Together the mechanisms above create a few potentially relevant feedback mechanisms. First, stories posted depend on the current opinions of members, but also influence the change in the opinions of individuals. Therefore a feedback loop exists between the opinions and stories. Second, the stories that are consistent with the majority view expressed provide additional reinforcement for those views, which can further increase the motivations of individuals holding on those opinions. The story reading and voting process includes a potentially reinforcing loop, in which stories that are voted more, are more likely to be read, and thus to receive even more votes. Another reinforcing feedback emerges from the motivation equations, in which those more motivated to post stories receive additional reinforcement through the effort justification mechanism, leading to them having even higher motivation. Several balancing loops are embedded to contain motivation and opinion in their feasible range.

Some of these feedback loops are somewhat different from the feedback mechanisms commonly discussed in the system dynamics literature in the sense that they cross different levels of aggregation. Some feedback mechanisms (e.g. motivation reinforcement) are active at the level of a single individual, while others go through multiple individuals and stories, before they come back to feed back into those entities, creating a multi-level dynamic that requires use to follow the distributions of individuals in the opinion and motivation spaces to be able to capture the key dynamics.

2.2) Metrics of Performance

Simulating the model discussed above one can follow the dynamics of motivation (M) and opinion (O) for all the individuals in the model. We can also look at the dynamics of votes (V) for the stories posted in the system. These three are the main state variables in the system from which everything else can be deduced. However, they each represent a distribution for the individuals/stories active in the system. For example, the graphs in Figure 3 show the timeline of changes in the motivation (a) and opinions (b) for a single simulation run for a population of n=50 individuals. Also shown is the histogram of number of votes for different stories at the end of the simulation (c). The individual metrics are color coded based on initial opinions of individuals: those with higher values are given a lighter color, which distinguishes them throughout the simulation.

In this simulation (parameters reported in Table 1) starting from the same motivation level and uniformly distributed opinions, individuals grow to fall into three groups. First, those at the low and high levels of initial opinion end up adjusting their opinions slightly, but then losing interest (reduced motivation) and ending up contributing very little to the system in terms of posting, reading, or voting. The members of the other group, those with the middle-ground opinions, converge to each other in the opinion space, become very motivated, and continue contributing stories to the system. The distribution of votes per story shows a power law in which most stories receive very few votes and a few receive a very large number of votes.

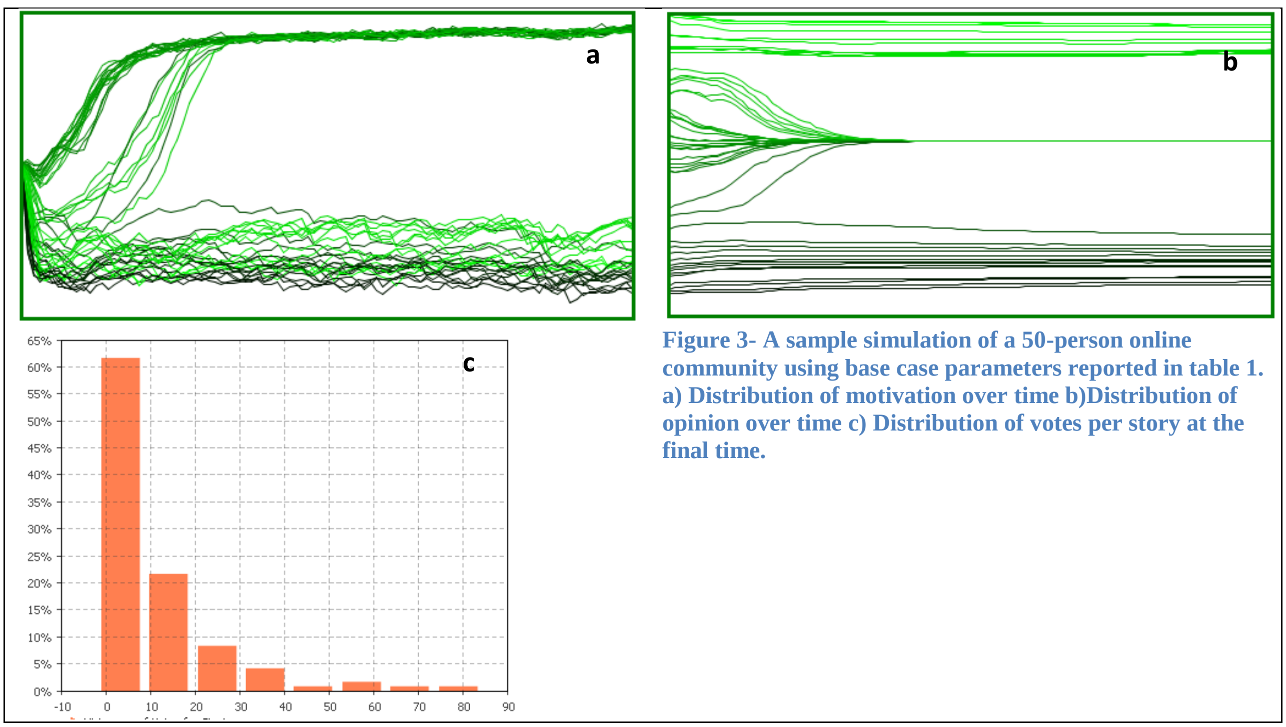

Figure 3- A sample simulation of a 50-person online community using base case parameters reported in table 1. a) Distribution of motivation over time b)Distribution of opinion over time c) Distribution of votes per story at the final time.

The model is stochastic and therefore different random numbers streams will lead to different realizations of the dynamics. While graphs like those in Figure 3 are useful for getting a detailed view into a single stochastic realization, they can not be scaled over many realizations to provide statistically reliable insights. For this reason we need to specify aggregate metrics to summarize the key characteristics of these distributions across individuals. Given that we know the initial distributions of the variables, metrics calculated at the final time are likely to be most informative. Moreover, the metrics should inform the distinction between polarized and homogenized communities, which is the main phenomenon of interest in our study. While mean and standard deviations are useful metrics to characterize a distribution, they are not enough to identify polarization. For example, consider 10 individuals with final opinions distributed A) Uniformly at increments of 10 from 10 to 100 (at 10, 20, 30,…). B)Five people having the value of 24 and five the value of 86. Sample mean (55) and standard deviation (31) are equal for both scenarios. However, the case A shows little polarization of opinions, whereas case B shows significant polarization with all individuals lumping around two different opinion levels.

We therefore design a different metric to capture the lumpiness in opinions, which can point to polarization. The basic insight behind this metric is that lumpy distributions, those with peaks and valleys, are more likely to represent polarization than smooth distributions. Therefore one could measure and accumulate the absolute value of slopes of a smooth distribution to signify lumpiness. Formally, for a distribution f(x), between 0 and X, the metrics would be:

$$L(f) = \int_{x=0}^{X} \left| \frac{df(x)}{dx} \right| dx \qquad (8)$$

Such measure should however be modified to take into account the fact that we deal with discrete realizations of opinion in this study. We therefore need to develop a smoothed distribution based on the discrete realizations of opinions, and then calculated the above lumpiness metric. For that purpose we use a triangular kernel function, which generates the function f(.) by adding small triangles, with width u, around observed data points, and adding up those triangle together (we also normalize the function by *1/(2un)* so that L remains between 0 and 1). Formally:

$$f(x) = \frac{1}{2un} \sum_{i=1}^{n} \mathrm{Max}(0, 1 - |\frac{O_i - x}{u}|) \qquad (9)$$

Figure 4 provides a schematic view on how this metric is calculated.

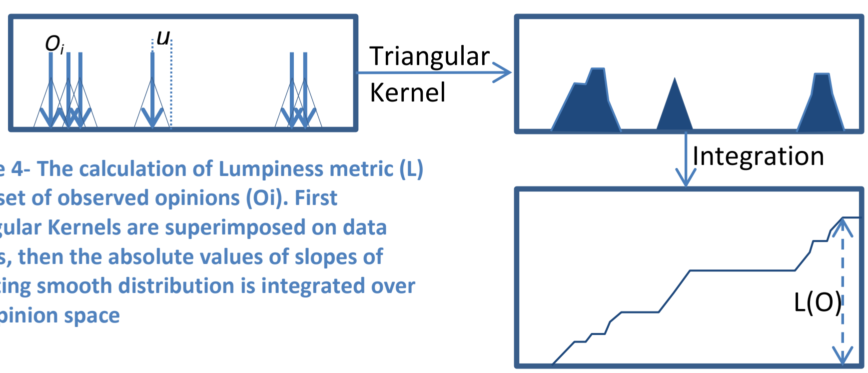

Figure 4- The calculation of Lumpiness metric (L) for a set of observed opinions (Oi). First triangular Kernels are superimposed on data points, then the absolute values of slopes of resulting smooth distribution is integrated over the opinion space

Figure 5 shows how different types of distributions are distinguished by standard deviation (S) and lumpiness (L) metrics together. Specifically, both metrics can not be very low, because low lumpiness points to opinions smoothly distributed on the space which will prohibit small S values. High standard deviation and low lumpiness point to smoothly distributed opinions, such as a uniform distribution. If L is high but S is low, we can deduce opinions are largely gathered in one location, which creates a single peak. If both S and L are high, we can expect two or more peaks on the opinion space, with valley's between them. In experiments that follow, S could change between 0 and 50 and L could change between 0 and 1. We also report the mean and standard deviation for final motivation of individuals over 25 simulations of synthetic communities with 50 members each.

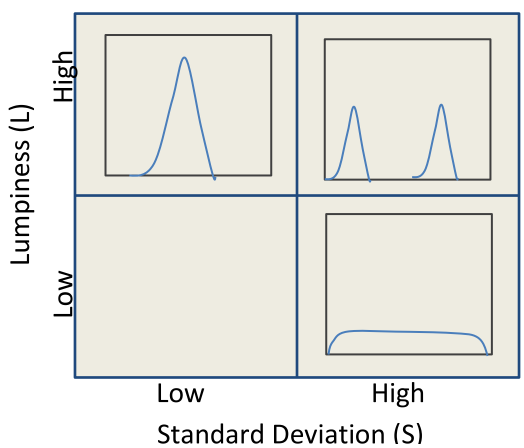

Figure 5-Different types of opinion distributions and the corresponding values for Lumpiness and Standard deviation of opinion.

2.3) Results

The model includes several parameters which could influence the modes of behavior observed. We therefore start from a plausible initial set of parameters which we found qualitatively consistent with real online communities, and then conducted sensitivity analysis around those values to assess the behavior of the model under a larger parameter space. Table 1 reports the parameter definitions and values used in the base case simulations and sensitivity analysis. When feasible, we included zero values of parameters in the sensitivity analysis to shut down different mechanisms all together and explore the impact. However note that full exploration of the parameter space requires concurrent variations in multiple parameters which we did not conduct. Moreover, to keep the different simulations comparable, we kept a few basic parameters unchanged. Yet we think even this one-dimensional sensitivity analysis provides interesting insights regarding the dynamics of online communities. These findings are discussed next.

Table 2 reports the results of the simulations, including the averages for five metrics (mean and standard deviation for opinion and motivation plus lumpiness of opinion) across batches of 100 simulations in 21 scenarios. The table also identifies if metrics are different from the base case at 99% (*) and 99.9% (**) confidence levels. Overall, four modes of behavior are identified in this synthetic online community. A sample of simulations with those modes is highlighted in Figure 6. First, in the *majority dominated (MD)* condition one group will converge together, remain active and dominate the community while others initially far from the majority opinion get discouraged and stop contributing to the community. Even if they continue reading stories, they will not see many stories that are consistent with their opinions, because the stories likely to be read (e.g. those on the front page) will be aligned with the increasingly narrow majority perspective. This is the most common mode of behavior in the model which we observed in the base case and many of the other settings. The second mode of behavior, *competitively polarized (CP)*, happens when multiple subgroups converge to a few points on the opinion space, while remaining far from each other. This equilibrium happens when individuals can be motivated to stay inside the system and continue contributing to it. This equilibrium becomes less viable if majority stories are promoted to the first page and therefore the minorities do not come across stories close to their opinion, which could bring them together. A third mode of behavior, *converged (C),* is when all individuals converge to the same opinion and thus keep a high level of motivation. This result is dependent on the existence of a population tolerant of opinions relatively far from their own. Finally, if opinions do not change, individuals could continue interacting in the online system with moderately compromised motivations (we call this condition *diversified (D)*). In Table 2 we also report the number of times each mode is observed within the 100 simulations with each parameter setting.

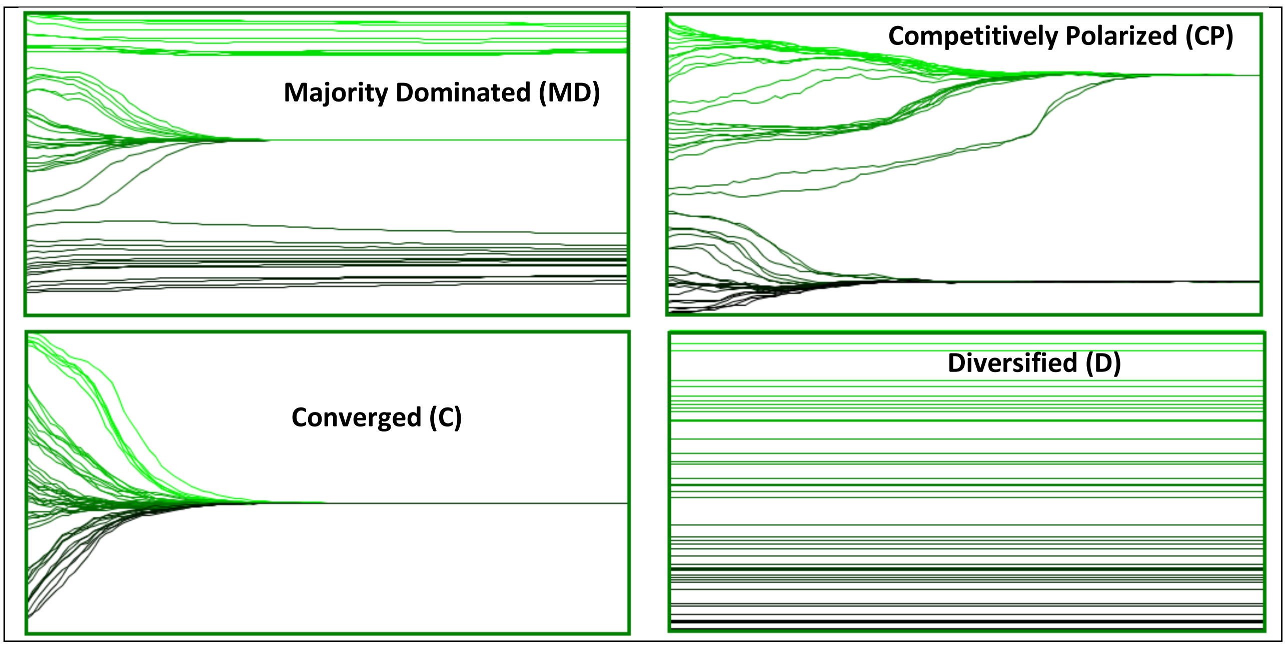

Figure 6- Sample simulations with the four basic modes of behavior.

In Table 2 the base case (first row) shows relatively high levels of both lumpiness and opinion variance, though neither are at their maximum (1 and 50 respectively). These simulations typically have the Majority Dominated mode of behavior: people close to majority opinion come together as they post stories that reinforce each others' motivations and opinions. Those with minority views gradually lose interest because the community is dominated with stories that promote opinions far from theirs, stop contributing and reading stories, and therefore their motivation further deteriorates. The high variation in the standard deviation of motivation shows this dichotomy. The moderately high levels of lumpiness and standard deviation of opinion show the coming together of a majority, plus scattered presence of opinions different from the majority in other parts of the opinion space.

Table 1-Parameter definitions and values in the base case and low/high values used for sensitivity analysis.

Parameter	Explanation	Base	Low	High
f	How long a story remains, in average	5	1	10
s	Fractional move rate towards read stories. (eq. 2)	0.005	0	0.02
h	The impact a posting will have on individual's motivation (eq. 5)	1	0	20
w	Strength of motivation change with reading stories (eq. 4)	2	0	4
t	Threshold of story distance from opinion for changing attitude towards stories (eq. 1,2,4)	20	10	40
c	The maximum incremental change in motivation possible with reading (eq. 4)	20	0	50
p	Power parameter for effect of motivation on posting (eq. 6)	2	0	4
k	Power parameter for effect of motivation on reading (eq. 7)	0.5	0	2
$a.d$	Fraction of maximum motivation attained in equilibriu, in the absence of effect of stories	0.5	0.2	0.9
y	The importance of votes in a story on its reading probability	0.5	0	2
n	Number of people in population	50	Not Varied	
v	The standard deviation of posted opinion around poster's mean opinion	10	Not Varied	
r	The (maximum) posting rate by posters (eq. 6)	1	Not Varied	
e	The (maximum) story reading rate for people (eq. 7)	50	Not Varied	
x	The maximum probability of voting for a story (eq. 1)	0.2	Not Varied	
u	The width of triangular kernel function for calculating lumpiness (eq. 9)	10	Not Varied	

The sensitivity analysis results point to a few insights that further elucidates the key dynamics at play. First, note that average opinions are unchanged in all the analysis. Basically all the mechanisms influencing opinion levels are symmetric in this model, therefore over many runs, no bias is introduced into simulations. Some bias in average opinion could evolve in a single simulation due to random realizations which may create biased initial conditions or push for opinions in one direction or another. Moreover, the changes in the average life of a story (f) have limited impact on dynamics.

The case with s=0 is an interesting comparison point. With this parameter setting opinions remain constant throughout, leading to uniform distributions of opinions at the end, which entail larger S and fairly low L values compared to the base case (see Figure 5). The motivations of the population also reduce slightly given that people in the majority do not move together and thus do not provide as much positive feedback for each other as in the base case. Similarly, motivation distribution remains less polarized and has lower standard deviation than base. Increasing the speed of opinion adjustment (s=0.02) allows those in minority to also converge more to each other, increasing the lumpiness factor and polarization. In effect, if people are more responsive to the story items they read, we could expect more polarization in an online community.

The impact of story posting on motivation (h) is limited. Variations in this parameter do not change any of the opinion polarization metrics significantly. There is a small but statistically significant impact on motivation if posting has a significant impact on motivation. It is interesting that the reinforcing loop of posting stories, leading to more motivation, which leads to further posting, does not have a significant impact on the dynamics in light of much more notable impact of story reading on motivation. In effect, even if a few individuals with minority view get their energy from their own contribution to the system and stay motivated, this is not enough to change the overall polarization trends.

The parameters controlling the impact of story reading on motivation (w and c) are more influential. Specifically, in the absence of this impact (c=0), individual motivation remains around the mean value of 50 (note very low variation in final motivation), and all individuals continue reading and posting stories. As a result they converge to a few (usually two or three) opinion peaks (high L and S). If reading has a stronger impact on motivation than the base case (c=50 condition), the only major difference is further polarization of motivation (high standard deviation of motivation), where all people end up at very high or very low levels of motivation depending on whether they join the majority or not. The impact of shape parameter, w, follows the same logic. However note that w=0 points to a stronger impact of reading on motivation (see equation 4) in which if a story further than t from current opinion is read, then the full negative impact on motivation is realized, and if the story is closer than t, then full positive impact is realized. Such step change in reaction is probably not realistic, but the tests reveal that when readers categories stories in a black and white fashion (low w), more variance in motivation will follow and the majority ends up dominating the online community. When people require a story to be very far or very close to their own to have an impact on their motivation, then different sub-groups continue to be active in the system even though they converge to a few different peaks in the opinion space.

The threshold for categorizing a story as consistent with one's opinion is an influential parameter. If this threshold is large enough, then most (in many cases all) participants gradually converge to the same opinion, their motivation goes up, and continue interacting within the system. In contrast smaller thresholds further reduce motivation because most stories are considered contrary to one's opinion and therefore reduce motivation. In practice, this parameter maps into the level of curiosity and tolerance for different ideas among the members of an online community. Based on these results we hypothesize that communities with higher tolerance among the base population may stay along longer and have a more energized user base.

It is interesting to note that if individuals continue posting regardless of their motivation (p=0), they will degrade everybody's motivation (because now there are a lot of un-pleasant items in the system). This setting will therefore lead to more convergence in motivations across individuals which keeps their activity level close to each other. The side effect of this setting is therefore the co-existence of stories with different opinion content in the community which helps different subgroups to converge to a few local peaks (thus very large L values). On the other hand, cutting the link from motivation to reading (k=0), has very limited influence on the dynamics of the system. In contrast, strengthening this link leads to a general decline in both motivation and the lumpiness of the opinions, as more individuals give up interacting with the system before they converge to the majority opinion.

The steady state level of motivation (a.d) influences the motivation mean at the end, and through that could lead to increased participation by the minority and thus increased lumpiness of final opinions. Therefore in practice factors that can keep people motivated to stay in the system, while helpful for overall participation, can not resolve the polarization tendency in the system. Finally, the more viewership of stories depends on their current votes (larger y), the lower is the average motivation (because those in the minority do not see any stories of their liking) and lumpiness (because having lost their motivation, those individuals will not read any more and will stop converging to their peers on the opinion space).

Table 2- Results of the analysis including base case and sensitivity results. Five performance metrics including mean and standard deviation of opinion and motivation as well as lumpiness of opinion are reported. Variations from base case are

identified at 95/99 percent significance levels by */. Also reported are the frequency of different modes of behavior (Majority Dominated (MD), Competitively Polarized (CP), Converged (C), and Diversified (D)) in each experimental setting.**

Parameter	Value	$\mu_{Opinion}$	$\sigma_{Opinion}$	$\mu_{Motivation}$	$\sigma_{Motivation}$	$L_{Opinion}$	MD	CP	C	D
Base Case		51	25	58	34	0.78	99	1	0	0
f	1	50	26	56	34	0.76	100	0	0	0
f	10	50	25	58	34	0.78	99	1	0	0
s	0	50	29**	41**	27**	0.31**	4	0	0	96
s	0.02	50	25	55	34	0.87**	86	14	0	0
h	0	50	25	59	34	0.79	97	3	0	0
h	20	50	26	60	34	0.79	97	3	0	0
w	0	50	28**	47**	43**	0.67**	100	0	0	0
w	4	50	24	45**	10**	0.92**	63	37	0	0
t	10	51	28**	26**	14**	0.67**	100	0	0	0
t	40	51	12**	89**	15**	0.95**	41	27	32	0
c	0	50	26	51**	1**	0.96**	22	78	0	0
c	50	49	27*	55	40**	0.72**	99	1	0	0
p	0	49	27**	24**	8**	0.89**	88	12	0	0
p	4	50	28**	51**	35**	0.68**	98	1	0	1
k	0	51	25	55	37**	0.8	97	3	0	0
k	2	50	26*	44**	11**	0.58**	89	0	0	11
a.d	0.2	50	27*	31**	18**	0.76	99	1	0	0
a.d	0.9	50	23*	71**	35	0.83**	94	6	0	0
y	0	50	24	61	33	0.79	99	1	0	0
y	2	50	29**	45**	34	0.7**	96	4	0	0

3- Discussion

What are the implications of this analysis for real world online communities? In terms of social welfare, two competing criteria influence the interpretation of the results. On the one hand, users utility gained from the online communication (motivation in our model could be a proxy for utility) should be considered. On the other hand polarization, by reducing commonalities among community members can increase conflict and forestall effective decision making in a society through democratic processes. The polarization of American politics in the recent years is a case in point. As a result conditions that lead to majority dominance, while providing relatively high motivation levels to majority members are conducive to polarization. In effect, members with minority views are sidelined and will likely seek to join other online communities which conform to their own opinions and biases. Such dynamic can lead to balkanization of public opinion and reduce the interaction among people with different opinions. The experience of Balatarin over its last two years of existence comes closest to this mode of behavior. The diversity of users and opinions expressed on the community has declined as a subset of users, mostly from those protesting against the current government, has come to dominate the discourse, posting, and voting on the system.

Competitively polarized communities increase the interaction among competing views, but may indeed reduce the utility gained from such interaction (because members are forced to view opposing opinions regularly), and thus not be stable over the long run. Balatarin's first two and half years (2006-2009) could be seen as coming close to this mode. In this period the pro- and anti-government users were both active on the system, posting and voting for stories closer to their own view. Discussions on this polarized community included many heated exchanges at the beginning, but gradually shifted towards isolating the pro-government voices, i.e. the majority dominated mode. In fact some of the data from Balatarin provides additional evidence for these trends. Specifically, in Balatarin individuals could vote both positive and negative for stories. However negative votes were typically used only when the individual had a very strong objection to a story. Therefore individuals with more negative votes could be seen as those finding themselves in the minority. By breaking the community into two groups (Figure 7), those with higher (a) and lower (b) fraction of negative votes (above 10% and below 1%), we can see how the two groups differ in their activity level over time. As expected from the majority dominance argument, the group in minority (with more negative votes) becomes less active after some time in the middle of the period of study. While very rough, this evidence suggests a similar dynamic of polarization and majority dominance could have been at play in Balatarin.

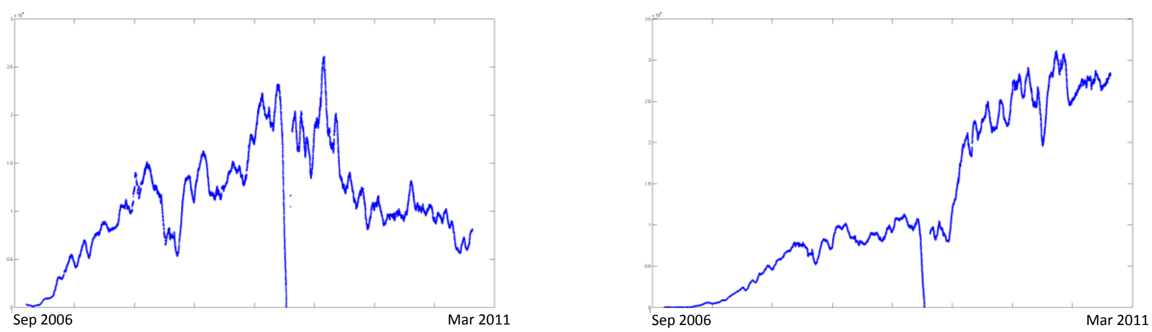

Figure 7- Aggregate patterns of activity by all individuals with more than 10% (a) and less than 1% (b) negative votes.

In practice, not all communities are created equal in terms of capabilities and appeal of different design features. Therefore potential members may tolerate some inconvenience of reading about opposing views because they want to stay in a better designed, more feature rich, or more populous online community. For example Facebook has been growing fast and keeping a diverse user base, partly due to its design and feature richness. Another mechanism online communities may be drawing upon to host diverse communities is to create customized online spaces that conform to the user's interest, rather than providing the same experience for all members. This mechanism, in effect, could parallel the majority dominance condition where multiple smaller communities could result from reorganizing the larger online communities into several customized sub communities, with limited interaction among them.

The convergence condition is less frequent in our experiments, and requires assumptions about individual characteristics such as tolerance, which are not controlled by the designers of online communities. Nevertheless this condition, if feasible, offers both high utility and no polarization. The rarity of such outcomes may partly be due to the modeling assumptions used here. Specifically, we assume none of the opinions is inherently more appealing than any other (e.g. empirically more valid, biologically more attractive). While extreme social constructivist views may espouse such a theory of reality (Berger and

Luckmann 1966; Grint and Woolgar 1992), we think there are inherent differences in the appeal of different opinions to human beings, and those differences are very relevant in the evolution of the distribution of opinions in a community. We assumed a uniform appeal for all opinions for theoretical convenience assumption, but empirical studies need to consider how to include such variations. Moreover, convergence condition is not without its own pitfalls. The risk with such homogeneous communities includes, among others, the lack of variation in opinions to challenge dominant assumptions and the potential for group think.

Overall, our analysis shows that individual tendencies to seek conforming evidence and avoid conflicting ideas can create robust feedback mechanisms likely to lead to polarization and balkanization of public opinion in online communities. These mechanisms are reinforced by different filtering and ranking methods that promise to bring the most appealing items to the view of the user, thus creating either a majority-dominated online environment (such as Diggs.com), or a bubble of similar opinions customized around each user (e.g. Netflix). Yet some form of ranking and filtering is indispensible to avoid information overload. Therefore a major design challenge for forthcoming online communities is to find ranking, filtering, and motivating methods which keep users engaged, but also provide them with a healthy portion of high-quality opposing views to stimulate divergent thinking and debate.

The current study is only a first step which explores a few relevant mechanisms theoretically. Empirical grounding of this research in data is an important next step. Opinions in online communities can be mapped and tracked over time using qualitative or algorithmic coding. Formation, activation, and dissolution of sub communities can be tracked and used to inform more realistic modeling work and analysis. The decisions to join or leave the community should be included as part of the study. More concrete design parameters, such as karma points (incentive points granted to online community members for taking socially beneficial actions, which could be spent elsewhere in the community), could be explicitly modeled and used for optimizing the design of online communities. Taking on these research tasks require a broad analytical toolbox that draws, besides dynamic modeling, on network analysis, statistical estimation, and working with large scale datasets.

References:

Alesina, A., R. Baqir, et al. (1999). "Public goods and ethnic divisions." Quarterly Journal of Economics **114**(4): 1243-1284.

Antweiler, W. and M. Z. Frank (2004). "Is all that talk just noise? The information content of Internet stock message boards." Journal of Finance **59**(3): 1259-1294.

Aronson, E. and J. Mills (1959). "The effect of severity of initiation on liking for a group." Journal of Abnormal and Social Psychology **59**: 177-181.

Aronson, E. and S. Patnoe (1997). The Jigsaw Classroom: Building Cooperation in the Classroom. New York, NY, Longman.

Berger, P. L. and T. Luckmann (1966). The social construction of reality; a treatise in the sociology of knowledge. Garden City, N.Y.,, Doubleday.

Bimber, B. (1998). "The Internet and political transformation: Populism, community, and accelerated pluralism." Polity **31**(1): 133-160.

Boisjoly, J., G. J. Duncan, et al. (2006). "Empathy or antipathy? The impact of diversity." American Economic Review **96**(5): 1890-1905.

Buyukozkan, G., A. Baykasoglu, et al. (2007). "Integration of Internet and web-based tools in new product development process." Production Planning & Control **18**(1): 44-53.

Chen, H. M., B. Gu, et al. (2008). Melting-pot or homophily? An empirical investigation of user interactions in virtual investment related communities. McCombs Research Paper Series No. IROM-05-08.

Cialdini, R. B. (2009). Influence : science and practice. Boston, Pearson Education.

Das, S. R. and M. Y. Chen (2007). "Yahoo! for Amazon: Sentiment extraction from small talk on the web." Management Science **53**(9): 1375-1388.

Delong, J. B., A. Shleifer, et al. (1990). "Noise Trader Risk in Financial-Markets." Journal of Political Economy **98**(4): 703-738.

Festinger, L. (1957). A theory of cognitive dissonance. Evanston, Ill.,, Row.

Fung, A. and E. O. Wright (2001). "Deepening democracy: Innovations in empowered participatory governance." Politics & Society **29**(1): 5-41.

Gilovich, T. (1991). How we know what isn't so : the fallibility of human reason in everyday life. New York, Free Press ;.

Grint, K. and S. Woolgar (1992). "Computers, Guns, and Roses - Whats Social About Being Shot." Science Technology & Human Values **17**(3): 366-380.

Grossman, L. K. (1995). The electronic republic : reshaping democracy in the information age. New York, NY, Viking.

Gurin, P., A. Hurtado, et al. (1994). "Group Contacts and Ethnicity in the Social Identities of Mexicanos and Chicanos." Personality and Social Psychology Bulletin **20**(5): 521-532.

Levine, P. (2002). Can the Internet Rescue Democracy? Toward and On-Line Commons. Democracy's Moment: Reforming the American Political System for the 21st Century. R. Hayduk and K. Mattson. Lanham, MD, Rowman & Littlefield**:** 121-137.

McLuhan, M. (1994). Understanding media : the extensions of man. Cambridge, Mass., MIT Press.

McPherson, M., L. Smith-Lovin, et al. (2001). "Birds of a feather: Homophily in social networks." Annual Review of Sociology **27**: 415-444.

Negroponte, N. (1995). Being digital. New York, Knopf.

Prandelli, E., G. Verona, et al. (2006). "Diffusion of Web-based product innovation." California Management Review **48**(4): 109-+.

Rheingold, H. (1991). "The Great Equalizer." Whole Earch Review **6**(Summer).

Simon, H. A. (1957). Models of man: social and rational; mathematical essays on rational human behavior in a social setting. New York,, Wiley.

Sunstein, C. R. (2002). "The law of group polarization." Journal of Political Philosophy **10**(2): 175-195.

Toffler, A. (1980). The third wave. New York, Morrow.

Van Alstyne, M. and E. Brynjolfsson (2005). "Global village or cyber-balkans? Modeling and measuring the integration of electronic communities." Management Science **51**(6): 851-868.

Westen, T. (1998). "Can technology save democracy?" National Civic Review **87**(1): 47-56.