Bellwethers and the Emergence of Trends in Online Communities

Yasuaki Sakamoto (ysakamot@stevens.edu)
Elliot Sadlon (esadlon@stevens.edu)
Jeffrey V. Nickerson (jnickerson@stevens.edu)
Center for Decision Technologies
Howe School of Technology Management, Stevens Institute of Technology
Castle Point on Hudson, Hoboken, NJ 07030 USA

Abstract

Group-level phenomena, such as trends and congestion, are difficult to predict as behaviors at the group-level typically diverge from simple aggregates of behaviors at the individual-level. In the present work, we examine the processes by which collective decisions emerge by analyzing how some articles gain vast popularity in a community-based website, Digg. Through statistical analyses and computer simulations of human voting processes under varying social influences, we show that users' earlier choices have some influence on the later choices of others. Moreover, the present results suggest that there are special individuals who attract followers and guide the development of trends in online social networks.

Keywords: Trends, trendsetters, group behavior, social influence, social networks.

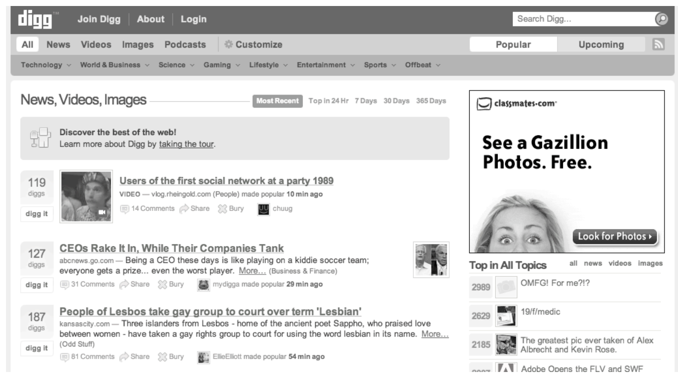

Figure 1: A snapshot of Digg homepage.

Background

An individual's decisions are often influenced by the context or the environment in which the decisions are made. For example, preferences between two choices can change when a third option is introduced (Tversky, 1972). At the same time, an individual's actions alter the environments. By modifying their surroundings, humans indirectly communicate to one another and influence one another's cognition. This stigmergic process results in emergent, group-level phenomena. For instance, our purchases influence and are influenced by the purchase patterns of others, giving rise to trends (Gladwell, 2000). Similarly, our driving behavior interacts with driving patterns of others to create congestion (Helbing, 2001).

Group-level phenomena are difficult to predict as they typically emerge spontaneously without our comprehension. Although we may be aware of the latest trends, we know little about how our beliefs and decisions contribute to the development of the trends, and we usually do not have the goal of influencing trends in our mind when we make a purchase. Furthermore, the behavior of the group and the behavior of individuals in the group often diverge (Axelrod, 1984). For example, a completely segregated world can emerge when each individual is only moderately biased toward living near members of own race, and thus no individuals intend to create a totally separate community (Schelling, 1971). Thus, despite the close relationship between the behavior of the individuals and the group, it is misleading to try to understand the behavior of the group from the behavior of the individuals or vice versa (Salganik, Dodds, & Watts, 2006).

To understand group phenomenon such as trends, we need to examine behavior at the level of both the individuals and the group. Although the cognition of individuals has been the focus of most cognitive scientists, there is a recent interest in the social (Levine, Resnick, & Higgins, 1993) and distributed (Hutchins, 1995) nature of learning and thinking in cognitive science (Goldstone & Janssen, 2005; Harnad & Dror, 2006). The past work in collective behavior tends to focus on people's behavior in controlled laboratory experiments (Gureckis & Goldstone, 2006), following the tradition of research in individual cognition. The present work complements the past work by examining the interaction between individuals' behavior and the performance of the group in a natural setting.

Trends in Digg

We advance that trends emerge as a result of the interaction between an individual's choice and the choices made by others. People are often influenced by the choices made by others (Cialdini & Goldstein, 2004). The decision of the individuals is inherently tied to the decision of the group. The number of a particular product left in a store and the crowding of a restaurant are reflections of the decision of the group based on the aggregate of individuals' decisions.

In particular, we examine whether individuals follow the choices made by all other individuals or imitate the behavior of bellwethers who set the trends. In the current work, we use as an example people's voting behavior in a social bookmarking website called Digg (digg.com). Digg users submit the URL of the web story they like. The submitted website initially remains in the "upcoming" section. Other users can "digg" or "bury" the website to vote for or against it. Digg employs a democratic, user-based ranking algorithm in promoting the articles. Once the website gains a certain number of supports within a certain timeframe, it is promoted to the "popular" section of Digg. Figure 1 shows the front page of Digg, which displays the stories that are recently promoted.

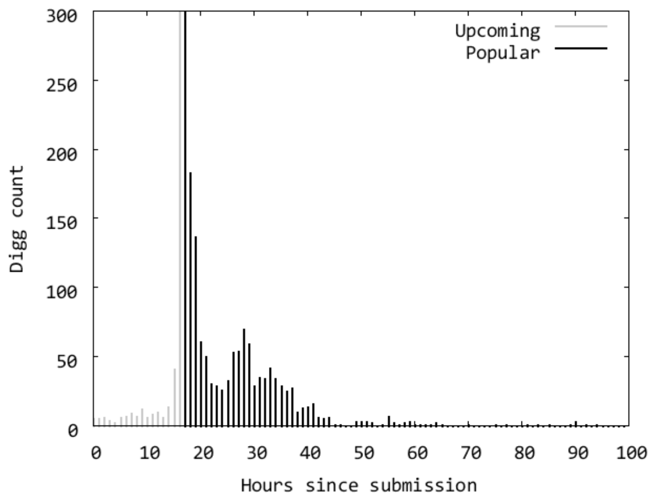

Figure 2: The time course of people's digging behavior for a promoted story.

In Digg, the total number of diggs received by an article so far is displayed next to each article. Thus, a user's environment in Digg is not simply a set of news stories to support but the distribution of other users' choices, allowing a user's decisions to interact with the others'. Similar environments can be found in transactional activities such as shopping although in these activities, unlike in Digg, participants usually don't intend to influence popularity.

Stories that become popular in Digg often receive a surge of diggs around the time of promotion. The time course of people's digging behavior shown in Figure 2 is typical of stories that become promoted. Because the promoted stories are moved to the front page and become more visible than upcoming stories, we examine the variability in popularity among promoted stories only.

We are interested in two main variables that may affect people's decision about supporting stories in Digg. The first variable is the number of early supporters. The second variable is the choices of trendsetters. Although the content of a story may play some role in its popularity, our focus in the present work is on characteristics associated with individuals and their behavior. After describing possible indicators of trends and our hypotheses, we present statistical and model-based analyses of the Digg data.

Possible Markers of Trends

Number of Early Supporters Some stories attract more supporters than others, creating trends. One reason why some stories receive more support than others may be that people simply favor stories that are supported by more people. People may follow the collective opinions of all individuals in the group, supporting stories with more existing adherents. If so the number of diggs collected by a story before the story is promoted (e.g., light bars in Figure 2) may predict the total number of diggs received by the story. Because each user can digg a story only once, the number of diggs for a story displayed next to the story is the number of supporters for the story. According to this account, initial random variations in the number of supporters between articles may lead to a dramatic difference in the subsequent popularity of the articles.

The idea that the number of previous supporters determines the preference of the subsequent decision maker is closely related to the principle of preferential attachment (Barabási & Albert, 1999) in social networks. In preferential attachment, nodes, or actors in a network, are added to the network successively, by connecting them to a small sample of existing nodes selected with probabilities proportional to their degrees, or the number of edges connecting the node to the other nodes. Thus, a node with many connections is more likely to acquire new connections than a node with only a few connections, leading to the "rich get richer" effect.

In Digg, there are two types of nodes: Users and stories. New connections are added between a user and stories. Using preferential attachment, each target for a user node's connections is sampled independently from a set of all existing story nodes with probability proportional to the popularity of the story node measured by its number of connected users. In this cardinality model, users are more likely to link to articles with more connections, or supporters, and thus this model follows the spirit of the number of supporters account.

There are psychological bases for choosing stories with more supporters. Social psychologists have shown that people are motivated to conform to the group's opinion (Deutsch & Gerard, 1955). Conformity may occur as a result of a pressure for social approval from others (Parks & Sanna, 1999), but sometimes it is found even privately (Asch, 1956; Sherif, 1935). Humans may have developed a natural disposition for agreeing with majority opinions, which operates even when the pressure for social approval is minimal as in the case of supporting stories in Digg.

A related psychological reason for favoring stories with more supporters is that people have a tendency to imitate the behavior of others. For instance, paths are formed by following predecessors' trails (Goldstone, Jones, & Roberts, 2006). Similarly, culture is developed by adopting innovations that are already established by others (Dennett, 1995). Imitation allows people to be more efficient by trying out solutions that they would not have considered otherwise (Bandura, 1965). Frequently imitated solutions are often the useful ones, and thus people may develop the tendency to perceive solutions selected by more people to be the useful ones. Organizations indeed tend to imitate changes that are adopted most frequently by other organizations (Kraatz, 1998). Similarly, Digg users may imitate the most frequent digging behavior, namely supporting stories with more adherents.

Trendsetters What creates inequality in the number of supporters among stories? Besides random variation, a trend might develop if some individuals are more influential than others. Influential users are trendsetters who lead the way in finding interesting stories. People may be more supportive of a story favored by a bellwether.

Some marketing practitioners believe that influentials who can sway the opinions of many peers play a major role in creating global trends, and thus they try to identify such influentials to anticipate and capitalize on trends (e.g., Keller & Berry, 2003). In fact, it has been shown that people tend to be swayed by vocal group members, and thus a small subset of the group may control the opinion of a group (Asch,

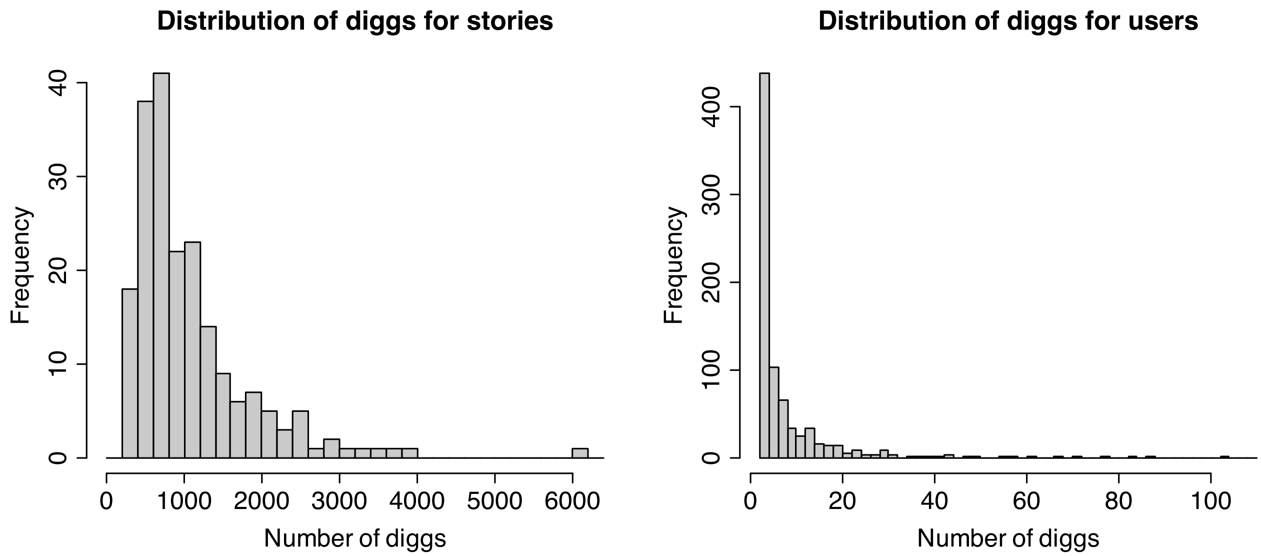

Figure 3: Distributions of diggs for the 200 stories and the 800 users are shown.

1951). Similarly, some Digg users may have a large influence on the Digg community. However, some studies suggest that infulentials exist but are not special in creating far-reaching trends (Watts & Dodds, 2007). Thus, it is unclear whether the popularity of stories in Digg is shaped by trendsetters influencing the opinions of followers.

Influentials are sometimes defined as individuals who have connections to a large number of others (e.g., Watts & Dodds, 2007). However, people weight opinions of other individuals differently (e.g., Friedkin & Johnsen, 1999), and thus influentials may be those who are trusted more than others. In the present work, trendsetters are those who attract many followers, suggesting that they are trusted by many users. To attract followers, trendsetters must digg before others. Early choices by the trendsetters may change the attractiveness of choices for subsequent decision makers. In business, the early entrant to a market often plays a major role in setting the tone for the future of the market (Zhang & Markman, 1998).

According to the trendsetter account, the diggers' decisions are not simply based on the stories and the number of existing supporters, but they are influenced by who supports the story. Instead of selecting the story nodes directly as in the cardinality model, a user may select another user as a target node, with probabilities proportional to the number of stories the other user has previously supported. Interestingly, this model suggests that the "rich get richer" process of preferential attachment underlies the growing popularity of trendsetters. After target users are determined, users are randomly connected to stories within the neighborhood of the target users. In this model, the "preferential attachment" to stories result because people preferentially attach to the trendsetters. Whereas connections are chosen from all existing story nodes in the cardinality model, connections are chosen from the story nodes within the neighborhood of an influential user in the trendsetter model (cf. Steyvers & Tenenbaum, 2005).

Current Hypotheses on Trend Development

We examine two hypotheses in the present work. One hypothesis is that the number of initial supporters shapes the development of trends. We predict that people favor stories that are supported by more people, and thus the number of supporters at an early stage should correlate positively with the number of supporters at a later stage. The basic idea behind this prediction is consistent with the "rich get richer" process of preferential attachment. We compare the statistics generated by the cardinality model with the statistics observed in the Digg data.

The other hypothesis is that trendsetters exist who affect other users' decisions. We test if a version of a preferential attachment model that incorporates the idea that trendsetters shape the behavior of others can account for the observed data better than the cardinality model. We identify bellwethers who are followed by many individuals from the results of the computer simulation. If people follow the trendsetters, they should often digg on the same stories. In addition, trendsetters should digg early to lead the way in creating trends.

Method and Results

Data Collection

Two hundred web stories listed as popular, and data for all of the diggs for each story, were gathered at random using Digg's API (apidoc.digg.com). We focused on the promoted stories because these stories are moved to the front page when promoted and become more visible than upcoming stories. The promoted stories and upcoming stories may be qualitatively different.

Data for a website include its title, description, time of submission, time of promotion, and username of the person who submitted the website. Data for a digg include the date and time the digg was made, the story on which it was made, and the username of the person who made it. The 200 websites had a total of 216,148 diggs. Of all of these diggs, 44,673 were from distinct users. From this collection of users, 800 that had dugg two or more of the 200 stories were selected at random.

Analyses

The left side of Figure 3 shows that most of the 200 stories had around 1,000 diggs and a few had over 3,000 diggs ($M = 1080.74, SD = 791.89$). The right side of Figure 3 shows that the 800 users varied in the number of diggs they made ($M = 7.94$, $SD = 11.42$). Whereas many users had only a few diggs, a handful of people made a large number of diggs. Our main interests were the effects of the number of early diggs and the trendsetters on people's digg behavior.

Number of Supporters If the number of existing supporters is an indicator of a story's popularity, there should be a significant positive correlation between the number of supporters at an early stage and the total number of supporters. Although as predicted there was a significant correlation between the number of early diggs and the total number of diggs ($r = .14$, $p < .05$), only a small portion of variance was accounted for by this variable. The number of early supporters likely plays only a small role in the development of popularity.

Our model-based analysis also suggests that people's voting behavior is influenced by more than just the number of early supporters. We simulated the 800 users' digg behavior for the 200 stories. In the simulation, we start with 200 story nodes and 800 user nodes. At each time step, n_i new connections are added between user node i and story node j, where n_i is determined for each user node i to match the observed distribution of the users' diggs as shown in the right side of Figure 3.

In the cardinality model, the probability $P_{ij}(t)$ of connecting user i to story j at time t is proportional to the popularity of j measured by its number of connections,

$$P_{ij}(t) = \frac{k_{\text{story}_j}(t)}{\sum_{m=1}^{200} k_{\text{story}_m}(t)}, \tag{1}$$

where $k_{\text{story}_j}(t)$ is the degree of story node j at time t. Users cannot digg the same story twice. When all available stories have no connections (i.e., $P_{ij}(t) = 0$), the user is connected to a story randomly from all story nodes.

The middle bar graph in Figure 4 shows the distribution of diggs across stories generated by the cardinality model. Because we controlled the number of diggs for each user node to match the observed count, the means for diggs are equal in the model and the observed data. As can be seen in Figure 4, when the user's decision is determined solely by the number of existing supporters, a few stories result in a large number of diggs with many stories having only a few diggs. The model was not able to account for the observed distribution well ($RMSD = 47.14$). The distribution of diggs in the cardinality model had much higher variability than the distribution observed in the 200 stories we sampled ($SD = 67.76$ vs. $SD = 23.17$).

Trendsetters Next, we examined whether diggers follow the opinion of the trendsetters. We simulated the idea that people follow trendsetters. In the trendsetter model, user i first find another user l, with the probability $P_l(t)$ proportional to the popularity of user node l measured by its number of connections,

$$P_l(t) = \frac{k_{\text{user}_l}(t)}{\sum_{m=1}^{800} k_{\text{user}_m}(t)}, \tag{2}$$

where $k_{\text{user}_l}(t)$ is the degree of user node l at time t. Once user node l is determined, user i is connected randomly to story j that is supported by user l. When no stories are available, a user is connected to a story randomly from all story nodes.

The right graph in Figure 4 displays the trendsetter model's output. With the simple following mechanism, the trendsetter model ($SD = 23.9, RMSD = 7.5$) does a much better job of accounting for the observed distribution of diggs across stories than the cardinality model.

We identified 20 individuals who were followed, or chosen as targets, most often in the simulation as potential trendsetters. If trendsetters attract followers, trendsetters should digg the same stories as others more often than expected by chance. If two users digg independently of each other, the probability $P(k \mid m,n)$ of the two users supporting k stories in common given that one user diggs m of the 200 stories and the other diggs n of the 200 is

$$P(k \mid m,n) = \frac{m!(200m)!n!(200n)!}{200!k!(mk)!(nk)!\Big(200(m+n)+k\Big)!}, \tag{3}$$

where $m \geq k$, $n \geq k$, and $m+n \leq 200+k$. Out of the 319,600 possible user pairings, 1348 pairs of users had diggs on common stories significantly more often than chance ($p < .05$), involving 136 users. All of the trendsetters we identified were in this list.

Trendsetters should also digg earlier than the other users. The trendsetters we identified made significantly more diggs before the stories became promoted than the others (9.85 vs. 0.24), $t(19) = 2.1, p < .05$. The trendsetters also made significantly more total diggs than the others (52.7 vs. 6.79), $t(19) = 7.1, p < .01$.

Discussion

We examined two variables that might affect the development of popularity. The number of diggs before the story was promoted was weakly yet significantly correlated with the overall popularity of the story. The cardinality model could not account for the observed data well.

The ability of the trendsetter model to account for the observed data suggests that there are bellwethers who appear to be successful in generating trends. It is unlikely that the trendsetters appear to be successful because they are quite active and digging everything. If simply being active can attract followers, then a model in which a user is randomly linked to stories should be able to account for the observed pattern. This model failed to fit the observed data.

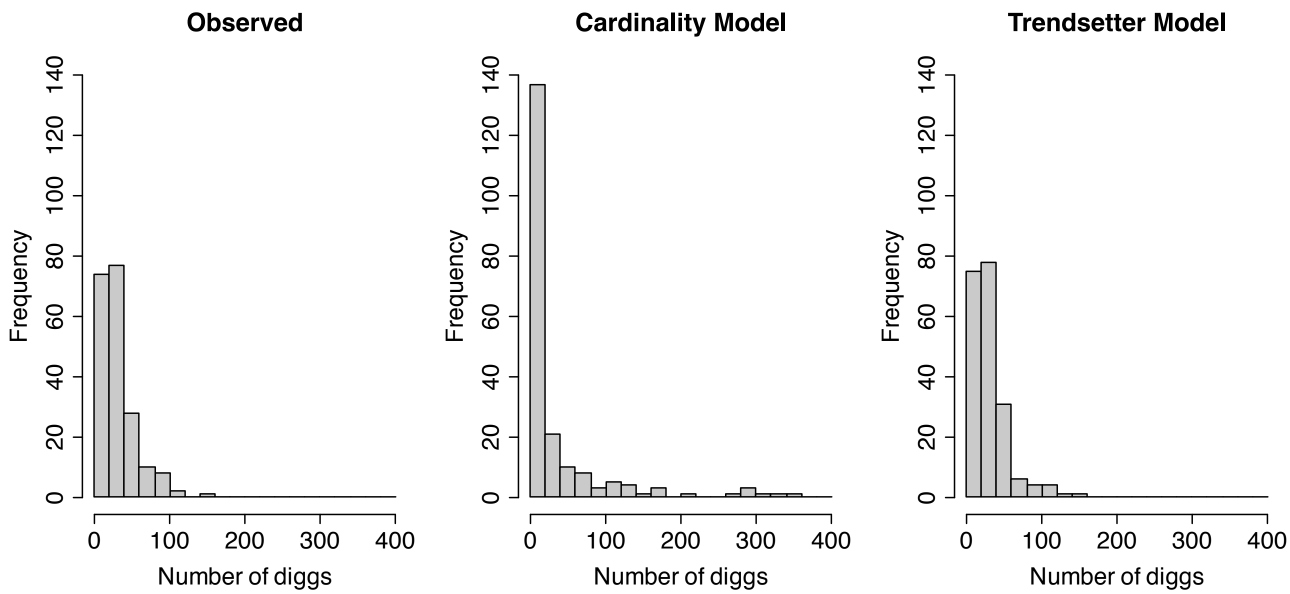

Figure 4: Distribution of 800 users' diggs for the 200 stories observed and generated by the cardinality and trendsetter models.

The bellwether may be an individual who monitor stories and submit the interesting stories to Digg as soon as the stories are published elsewhere. Other people will later find these stories interesting and digg. It is possible that the followers are responding only to the content of the stories, not to the bellwether. Stories may be inherently interesting or not, and perhaps everyone can recognize an interesting story. This account suggests that whether a story is submitted by an influential individual or not does not affect the story's popularity. In contrast, the trendsetter account predicts that the same story will become more popular when it is submitted by a bellwether than by other individuals. The followers treat the bellwether as a trusted editor and read whatever the editor likes. These accounts can be experimentally tested by manipulating the story content and the status of individuals.

Websites submitted to Digg focus on cultural and social happenings, but Digg occasionally receives advertising. In December 2006, a user submitted a small business's webfront to Digg. The next day it was promoted to popular status, and the business owner reported to the Wall Street Journal that more product was sold during that week than otherwise would have been in a year. An interesting question is whether the same outcome is obtained if the webfront is submitted by a different user. We think trendsetters will lead to more profit.

Some individuals appear to strive for achieving bellwether status in Digg. How might it be intentionally achieved? One way to become influential may be to position yourself in the "right" place in the network. The position in a communication network plays an important role in influencing other people to adopt knowledge (Chwe, 1999). There is a finding suggesting that people who are most central in a network tend to be the most influential in disseminating knowledge (Valente & Davis, 1999). On the other hand, deviant individuals are judged as more influential and more behaviors of deviant individuals are remembered (Taylor, Fiske, Etcoff, & Ruderman, 1978). One strategy may be to join the network by linking to individuals who are central and then gradually differentiate yourself from those individuals by noticing stories that others do not and creating a new niche in the ecology.

Although the Digg users are not generally friends with one another, informal conversations with some of the users revealed that they do communicate over the phone and in chat forums outside of the Digg community. We may be able to correctly predict who are or will become friends in Digg based on the similarity in users' digging behavior. Similar people do tend to gather together (Hornsey & Hogg, 2000) and stay together once grouped (Stangor, 2004). In Digg, the similarity between two persons can be based on the number of diggs shared and not shared by them. Our results suggest that some users do digg the same stories more often than expected by chance. Conversely, similarities among stories can be measured by the number of diggers shared by the stories. These similarity measures can be used to create clusters of people and stories to further analyze trends using larger units than each individual and story.

Our work suggests that bellwethers who can shape the development of trends exist in online communities. Some researchers hold that influentials are different from media stars who have stronger influences (Watts & Dodds, 2007). Are trendsetters in Digg media personalities? Are there discontinuities between influentials and stars? Can bellwethers become stars? To fully understand these issues, we need to conduct controlled experiments on collective behavior (e.g., Kearns, Suri, & Montfort, 2006), coupled with analyses of people's natural behavior. In addition, we need to examine cognition of individuals and the group in a single framework (e.g., Sun, 2006). Changes in the environment, such as a variation in the order in which people digg, can dramatically influence the overall behavior of the community even when the preferences of individuals do not change. Furthermore, an attempt to predict trends from preferences of individuals often fails because collective behaviors, such as trends, often diverge from simple aggregates of the behaviors of individuals and are difficult to predict by nature (Salganik et al., 2006). Thus, it is important to examine the interaction between the behavior of the group and the behavior of individuals within the group.

References

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men* (pp. 177–190). Pittsburgh, PA: Carnegie Press.

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs*, *70 (Whole no. 416)*.

Axelrod, R. (1984). *Evolution of cooperation*. New York: Basic Books.

Bandura, A. (1965). Behavioral modification through modeling procedures. In L. Krasner & L. P. Ulmann (Eds.), *Research in behavior modification: New development and implications* (pp. 310–340). New York: Rinehart and Winston.

Barabási, A. L., & Albert, R. (1999, October 15). Emergence of scaling in random networks. *Science*, *286*, 509–512.

Chwe, M. S.-Y. (1999). Structure and strategy in collective action. *American Journal of Sociology*, *105*, 128–156.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*, 591–621.

Dennett, D. C. (1995). *Darwins dangerous idea*. New York: Touchstone.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informative social influences upon individual judgment. *Journal of Abnormal and Social Psychology*, *51*, 629–636.

Friedkin, N. E., & Johnsen, E. C. (1999). Social influence networks and opinion change. *Advances in Group Processes*, *16*, 1–29.

Gladwell, M. (2000). *The tipping point*. London: Little Brown and Company.

Goldstone, R. L., & Janssen, M. (2005). Computational models of collective behavior. *Trends in Cognitive Sciences*, *9*, 424–429.

Goldstone, R. L., Jones, A., & Roberts, M. E. (2006). Group path formation. *IEEE Transactions on System, Man, and Cybernetics*, *36*, 611–620.

Gureckis, T. M., & Goldstone, R. L. (2006). Thinking in groups. In S. Harnad & I. Dror (Eds.), *Distributed cognition: Special issue of pragmatics & cognition (vol. 14)* (pp. 293–311). Amsterdam, The Netherlands: John Benjamins.

Harnad, S., & Dror, I. (2006). *Distributed cognition: Special issue of pragmatics & cognition (vol. 14)*. Amsterdam, The Netherlands: John Benjamins.

Helbing, D. (2001). Traffic and related self-driven many-particle systems. *Review of Modern Physics*, *74*, 1067–1141.

Hornsey, M. J., & Hogg, M. A. (2000). Intergroup similarity and subgroup relations: Some implications for assimilation. *Personality and Social Psychology Bulletin*, *26*, 948–958.

Hutchins, E. (1995). How a cockpit remembers its speed. *Cognitive Science*, *19*, 265–288.

Kearns, M., Suri, S., & Montfort, N. (2006, August 11). An experimental study of the coloring problem on human subject networks. *Science*, *313*, 824–827.

Keller, E. B., & Berry, J. (2003). *The influentials*. New York: Free Press.

Kraatz, M. S. (1998). Learning by association? Interorganizational networks and adaptation to environmental change. *Academy of Management Journal*, *41*, 621–643.

Levine, J. M., Resnick, L. B., & Higgins, E. T. (1993). Social foundations of cognition. *Annual Review of Psychology*, *44*, 585–612.

Parks, C. D., & Sanna, L. J. (1999). *Group performance and interaction*. Boulder, CO: Westview Press.

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006, February 10). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, *311*, 854–856.

Schelling, T. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, *1*, 143–186.

Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology*, *27*, 1–60.

Stangor, C. (2004). *Social groups in action and interaction*. New York: Psychology Press.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*, 41–78.

Sun, R. (2006). Prolegomena to integrating cognitive modeling and social simulation. In R. Sun (Ed.), *Cognition and multi-agent interaction: From cognitive modeling to social simulation.* Cambridge: Cambridge University Press.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*, 778–793.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*, 281–299.

Valente, T. W., & Davis, R. L. (1999). Accelerating the diffusion of innovations using opinion leaders. *Annals of the American Academy of Political and Social Science*, *566*, 55–67.

Watts, D. J., & Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of Consumer Research*, *34*, 441–458.

Zhang, S., & Markman, A. B. (1998). Overcoming the early entrant advantage via differentiation: The role of alignable and nonalignable differences. *Journal of Marketing Research*, *35*, 413–426.