Free interactions, hierarchical structure: Factors explaining replies attraction in online discussions

by Itai Himelboim and Stephen McCreery

Abstract

Given the opportunity to interact freely, individuals conform to a structure, in which a few actors attract a large and disproportionate number of ties or relationships. Drawing from literature on preferential attachment and scholarship about online discussions, this study examines patterns of replies, which are one aspect of the disproportionate attraction of replies in forums, as predicted by two factors: number of existing replies and content of posted messages. In two 2X2 experimental designs conducted via a custom developed online discussion platform, 198 subjects participated. Findings show an interaction, where the number of replies increased replies attraction only for the high–interest messages, illustrating the balance between the individual and group dynamics levels in evoking discussions.

Contents

Introduction

Integrative theoretical framework: What makes messages attract replies?

Methods: Tools, measurements, and procedures

Discussion

Conclusions

Introduction

In what has become known as the information society, individuals and organizations depend more and more on information flow to operate, interact, and survive (Castells, 1996). The online discussion forum was one of the first platforms to take advantage of the Internet for information exchange, and it is still one of the most popular venues. Like in many other networks, however, patterns of interactions in online discussions conform to a highly unequal distribution of replies across individuals and messages (Raban and Rabin, 2007). This study aims to help understand the dynamic that gives a few participants much more power than others to control the flow of discussions.

This study draws from and integrates two theoretical frameworks to explore replies attraction in online political forums. First, network literature on preferential attachments shows that across domains of research, connections are distributed among groups of nodes — individuals or objects — according to how many links — connections — they already have. So, the “rich get richer” (Barabási, et al., 2002;
Newman, 2001). This dynamic, together with the natural growth of networks, results in a highly unequal distribution of information, connections, and other qualities, where a few nodes can become more potent and influential than most nodes. Second, studies exploring discussion forums have identified two types of explanations for replies attraction — characteristics of authors or messages and history of replies from the group.

Aiming to explore why a few messages attract a large and disproportionate number of replies in online forums, this study suggests testing two factors that can explain replies attraction — interest in message content and number of existing replies. Because in existing forums it is not plausible to separate the two factors (messages with perceived high content quality will probably also have many replies), an experimental design will be used. Using a Web–forum tool that was developed for this study, two 2X2 designs are used to examine how well the two independent variables — interest in a message’s content and the number of replies it already has (or “popularity”) — predict the dependent variable — participants’ replies.

Free interactions, hierarchical structure

Across areas of research and life, network scientists have noticed that almost any group of elements — social actors or other — given the opportunity to interact freely with one another, displays similar patterns. A few elements, or nodes as they are frequently referred to, interact with, or are linked to, a large and disproportionate number of other nodes in their networks. Most other nodes, however, are connected to very few others, if at all. This positively skewed distribution of the number of connections (degrees) among nodes follows a specific pattern called power–law. Power–law models the growth of networks, assigning probabilities that depend on degree centrality, that is, the number of other nodes a network is connected to (for further discussion on power–law, see Faloutsos, et al., 1999; Aiello, et al., 2002; Newman, 2005).

Attempting to describe the mechanism that explains the formation of the power–law, scientists often refer to preferential attachment: New links are attached preferably to nodes that are already well connected (Newman, 2001). Consequently, already highly connected nodes increase their connectivity faster than their less connected peers. Power–law can form on large networks only. Preferential attachment is seen as an explanatory mechanism that, together with the natural growth of networks, leads to power–law degree distribution.

Preferential attachment has been identified in many social networks, such as sexual relationships (de Blasio, et al., 2007), e–mail messages (Dodds, et al., 2003), Web hyperlinks (Huberman and Adamic, 1999; Barabási, et al., 2002; Broder, et al., 2000; Soares, et al., 2005), and citations in academic work (Newman, 2001). However, preferential attachment is not limited to social networks, as it has also been found in biology (Albert, 2005) and linguistics (Dorogovtsev and Mendes, 2001), to name a few. More specific to this study, Johnson and Faraj (2005) found that electronic knowledge networks — online platforms of information exchange such as forums — also follow preferential attachment. Only in large–scale networks can a power–law degree distribution be formed (Simon, 1955). Preferential attachment, however, can precede power–law, as it explains, in part, how this unique network structure is developed (Newman, 2001). Therefore, although existing literature tends to examine large datasets to identify preferential attachment, it is not necessarily limited to large networks.

The preferential attachment model, then, describes the growth of hubs, that is, a few highly connected individuals in a network. How influential these hubs are, especially in terms of public opinion, is a topic of much debate. The idea of a few, highly connected individuals who are likely to influence other persons in their immediate environment, can be traced back to Katz and Lazarsfeld’s (1955) theory of opinion leaders and Rogers’ (1995) theory of diffusion of innovations. Despite much theoretical and methodological criticism (see Weinmann [1994] for further discussion), the notion that trends and innovations are often initiated by a relatively small segment of opinion leaders in the population remains popular both in academic scholarship (see, for example, in marketing, Van den Bulte and Joshi, 2006; public health, Moore, et al., 2004; and, political behavior, Nisbet, 2006; Roch, 2005), as well as in popular books (Barabási, 2002; Gladwell, 2000) and marketing literature (Burson–Marsteller, 2001; Keller and Berry, 2003; Rand, 2004). Contradicting this trend, Watts and Dodds (2007) conducted a series of computer simulations of interpersonal influence processes, which undermined the power of hubs. He found that under many conditions that were considered, large cascades of influence were driven not by hubs, but by a critical mass of easily influenced individuals. Beyond the types and conditions under which hubs can influence their immediate social
networks, what makes some participants disproportionately more
connected than others is still underdeveloped. This study, then, is
concerned with the dynamics that leads to the rise of hubs.

Preferential attachment, then, is a probabilistic model suggesting a
dynamic where highly connected nodes have an advantage over others in
obtaining new connections. This model does not, however, explain what
makes the highly connected nodes more likely than others to obtain new
connections. Very broadly, then, two possible factors can explain why
nodes attract links: One is the unique individual characteristics of a node.
A node becomes connected in a network, it can be very intuitively
argued, because it has a quality or qualities that attract connections from
or with other nodes in the same network (naturally, these qualities will
differ across areas of research); a second explanation suggests that the
number of existing connections in a network (links) is the reason for a
node’s advantage in obtaining new connections. In a sense, the rich get
richer, because they are rich. Is it the node’s quality, the number of
connections it has, or a combination of the two that best predicts new
connections?

We next discuss literature about a specific domain of social interactions —
one online political forums. A later section will tie the theoretical context
provided by the preferential attachment literature with the online forums
literature discussed next, to construct research hypotheses.

Online discussion forums

Discussion forums exist on almost any topic imaginable. Attention to
messages and participants in these discussions is distributed in a highly
skewed way, as a few forums and messages posted to them attract a
large and disproportionate number of authors and replies (Raban and
Rabin, 2007). These few hubs can play unique social roles in forums.
Honeycutt (2005) studied hazing techniques in online communities and
found that "elite members" of the communities use their powers to
maintain boundaries and retain their power through, among other means,
controlling access from other community members and employing
threats. In political discussions, Himelboim, et al. (2009) found that
discussion catalysts, which constitute less than two percent of
participants, attracted more than half of the replies, influencing the
direction of discussions. Welser, et al. (2007) identified the role of
"answer people" — individuals whose dominant behavior is to respond to
questions posed by other users.

Only limited literature however explored what causes individuals to reply
in forums. Studies addressed two types of variables that predict how
likely a message is to attract replies — the characteristics of the message
itself or its author (a node) and the responses from others (links).

Characteristics of authors and messages. When participating in discussion
forums, users have control over the information they post as well as
general self-characteristics by which they choose to identify themselves.
For example, screen names and linguistic styles have been discussed as
cues that shape participants’ impressions (Jacobson, 1999). In Becker
and Stamp’s (2005) study on chat rooms, participants reported that
screen names with positive connotations attracted chat partners, whereas
negative ones were repelling. Heisler and Crabill (2006) found that e-mail
usernames do provide an opportunity to gather information about
senders. For example, names that were moderately descriptive were
perceived more favorably than very plain or very descriptive usernames.

Naturally, posting messages is the most basic way for participants to
characterize themselves. Focusing on the content of posted messages,
Fiore, et al. (2002) found that messages with simpler content were more
likely to seed new discussion threads than were complex messages.
Characteristics of the platform were also found to affect the amount of
activity. Wise, et al. (2006) showed that participants were more likely to
post to a Web site with "interactive messages" — messages that referred
to previous messages — than to forums with no interactive messages.

Responses of the group. Joyce and Kraut (2006) found that participants
who had their initial post replied to were more likely than others to
participate again, regardless of the content of the replies (positive,
negative, or neutral). Caspi, et al. (2003) studied an asynchronous online
learning discussion group and found that as group size increases, the
amount of group interaction — i.e., message postings — increases
accordingly. Beuchot and Bullen (2005) studied an online graduate
students forum and found that social or personal online interactions in the
forum increases participation and expands the depth of discussion.

Messages’ perceived qualities were also found to be related to the
amount of discussions they evoke. Fiore, et al. (2002) showed that
Usenet participants’ perceptions of trust in or respect for authors were
positively correlated with the number of replies a specific author received. However, authors who frequently posted messages were considered to be rude, and others were less interested in their messages. Jones, et al. (2002) also identified negative results associated with high levels of activity in forums: Higher proportions of active online discussion group users would end their active participation in larger, more overloaded discussion groups.

Integrative theoretical framework: What makes messages attract replies?

As discussed earlier, given the freedom to interact freely in a discussion forum, a few authors and messages will attract a large and disproportionate number of replies, whereas for most participants and messages they post, there will be little likelihood of attracting replies. In discussing the preferential attachment theorem, two competing explanations where suggested for the disproportionate amount of links grown out of the highly connected nodes — characteristics of nodes and the number of connections they have. In the context of online discussion forums, nodes can be authors or messages, and links are replies to posted messages. Indeed, studies on online discussions indicate that both message content (e.g., Wise, et al., 2006) and history of replies to a message or author (e.g., Joyce and Kraut, 2006; Caspi, et al., 2003) can influence replies’ attractions.

These two factors can have quite different implications, as participants have full control over the content they post but are much more limited and have less direct influence on the replies they attract. As the distribution of replies in forums is highly skewed, what role does the actual content play in the rise of these disproportionately popular messages? What role does a rich–get–richer effect type of dynamics, in which many replies leads to even more replies, play? Beyond the individual contribution of each of the two factors, the interaction between the two can also shed light on the rise of these highly replied to messages. In online discussion forums, however, these two potential factors — message content and number of replies — are unlikely to be independent. A message with content that is perceived to be of a higher quality by others in a discussion forum is also likely to have many replies. In “nature,” therefore, it is not likely to conduct research that examines the individual contribution of each predictor and possible or actual interactions between the two. In order to examine these two factors independently, an experimental design is applied, as discussed next in the methods section.

Both a message and an author can be conceptualized here as a node and become a unit of analysis for this study. Some of the reasons for replies attraction can be associated with the actor who posted them, its accumulated impression of previous postings, including reputations, and other actor–related characteristics. Other reasons are related directly to a single message itself — perceived interest of content and number of replies. Although participants might be a more intuitive choice for units of analysis, in order to separate successfully the two suggested predictors — posted content and replies — and potentially examine the interaction between the two, a message was selected as the unit of analysis. By making this choice, the history of replies an author has attracted to all message (via reputation, for example) cannot influence how likely new messages are to attract replies.

The preferential attachment theorem, as discussed earlier, suggests that those with many connections in the networks are more likely than others to obtain new connections. Bridging the theorem and online forums literature, this study thus attempts to make a modest contribution to understanding the dynamics that lead to the unequal distribution of replies among messages in discussion forums, by identifying the individual contribution of each factor — posted replies and characteristics of content — and possible interaction between them. As discussed earlier, multiple studies suggest content as influencing replies attraction (for example, Jacobson, 1999; Becker and Stamp, 2005; Heisler and Crabill, 2006; Wise, et al., 2006). A message’s content is operationalized here as the level of interest that participants have in it. This hypothesis may sound at first rather expected:

H1: A message with high–interest content will be more likely than a low–interest message to attract replies in a network.

Other studies, however, suggest that replies from the group as a predictor for replies attraction (e.g., Joyce and Kraut, 2006; Fiore, et al.,...
Examining the message interest factor independently from the second factor — number of replies — will allow us to investigate how much variation in replies attraction can be accounted for the message’s content, and how much for the number of replies it already has when a participant decides to respond to it. Keeping the two factors independent from one another will furthermore allow us to examine for interaction between the two predictors. The second hypothesis is therefore:

H2: A message with many replies will be more likely than a message with no replies to attract replies.

The proposed research hypotheses are not topic specific. For the purposes of this study, we suggest examining political discussion forums for their theoretical importance, as we briefly discuss next.

Online political forums

Forums for political discussion play a critical role in societies and, in particular, democracies (de Tocqueville, 1945). Political discussions have shifted in large numbers to computer-mediated discussion spaces like Usenet newsgroups, Web boards, and e-mail lists (Levine, 2000). Informed citizens are crucial for the effective operation of a democratic society (Siebert, et al., 1956; Picard, 1985). The Internet has been a cause for great enthusiasm among those who appreciate that it gives a large population access to a wide range of sources of information, and gives each person the potential to create new information or nominate new topics for public discussion. Corrado and Firestone [1] found that online discussions create a conversational democracy where “citizens and political leaders interact in new and exciting ways.” Hauben and Hauben (1997) suggested that online discussion groups allow citizens to participate within their daily schedules. Business leaders invoke the Internet as a guarantee of press freedom (McMillan, 2008). Researchers have documented the potential for effective collective action through the Internet (Bucy and Gregson, 2001; Mehra, et al., 2004; Kahn and Kellner, 2004). Because computer-mediated discussions provide an almost infinite canvas for new messages, Rheingold notes that if discussion boards are not an example for democratizing technology, then there is no such thing. Examining political forums could therefore not only shed light on the dynamics of interactions on online forums, but also could have interesting insights for online political discussions.

Methods: Tools, measurements, and procedures

Participants. All subjects were recruited from introductory classes at a southern university. Twenty-nine subjects participated in pre-test 1, 87 in pre-test 2, 92 in experiment 1, and 106 in experiment 2. Fifty-five percent of all participants were female. The average age was 20.95; 81 percent were juniors or seniors. On a 1–5 scale, the average interest in politics was 3.34 (SD=1.11). Participants were randomly assigned to groups. These and others individuals’ differences are not expected to affect the analysis.

Variables. The dependent variable was replying to a thread, and held one of two values — yes and no. One independent variable was a node’s individual quality, which was operationalized for this study as the level of interest in the content of a message. This variable had two values — low and high, based on findings from pre-test 2. The most popular message ($M=4.47$) and the least popular message ($M=1.77$), on 5-point Likert scales, were selected as high- and low-interest messages (see more details in the pre-test 2 section).

A second IV was a message’s number of links, which was operationalized as the number of replies, and held one of two values — low (no replies) and high (67 replies). The value selected for high number of replies (67) was based on a pre-test. Thirty-nine students were asked to participate in discussion forums (using the pretested messages, level of interest remained stable, and forums’ order of appearance was randomized across participants), where the number of messages in the forums ranged from zero to 100, with intervals of 10. Tracking whether students scrolled down to look at all messages, we found that, for threads of 60 messages or higher, more than 80 percent of participants did not scroll to view all messages. We interpret not viewing all messages in a thread as perceiving that thread to have many replies. Letting for a number larger than 60, to avoid round numbers that may look less realistic, 67 was selected arbitrarily.

A message, not its author, was used as an unit of analysis. In real life
forums, over time, messages that an individual posts, as well as replies these messages attract, can reflect on the author and potentially influence future replies attraction. However, within the experiment, a single root message appeared in each forum, disallowing its “author” to develop a reputation.

Research tool. For experiments 1 and 2, an online discussion platform tool was developed. This tool allows researchers to create threads, each which includes a first message (root message) and replies. Each thread shows the number of replies posted in it, whether the actual replies are visible or not, as explained next. One or more threads can be included in each forum. A forum and the threads in it appear on one Web page. Participants may post a reply to existing messages in any given thread or start a new thread, based on the thread’s definitions. A condition includes one or more forums. The research tool allows researchers to control for the number and content of discussion forums, threads, and replies that each subject is exposed to. It also allows to choose from a variety of restrictions for participation: Whether participants can reply to a root message (the first message of a thread) or to all levels of the thread; whether they can start a new thread in a forum, or just reply to existing messages; whether they can see replies, or only the number of replies posted in a thread; and whether or not they can see messages posted by their peers. The order of forums and threads within forums were randomized for each subject. The researcher can ask subjects to participate from a computer lab or from home. At the administrator level, all data related to subjects and messages are collected.

Many of the studies on online discussion forums, especially if taking a networks approach, have analyzed large existing datasets. We decided to use an experimental design for several reasons. First, the level of interest in the content of a message — one of the independent variables — is subjective and very likely to change from one population to another. Using an experimental design, we were able to draw from the same group of people — subgroups for the experiments and a subgroup for pretesting for messages’ levels of interest. If using existing discussion forums, it would not be feasible to examine perceived messages’ interest using people from that group, and especially not for all messages that appear in these discussions. Second, if using existing forums, history would have become a string–confounding variable. The perceived quality of previous messages and the number of replies they attracted in the past could have influenced the reputation (for example) of an author and potentially its reply attraction, regardless of the two independent variables in question. The experiments did not allow for such confounding variables, as the “author” associated with each manipulated message had no history and appeared only once. Third, in many online forums, participants interact off–line as well as online and are familiar with friends’ screen names. In the experiment, participants were instructed to not use their real names or other currently used usernames, so off-line interactions were less likely to affect their participation.

Four conditions were created for a 2X2 experimental design (low and high number of replies, and low and high interest in a message). Each condition included five forums — one of the four forums in the experiment and four “fake” forums. Fake forums were the same in all conditions and ranged in topic (movies, travel, cars, and cooking). The number of replies in each fake forum also ranged. Each forum appeared in a different Web page, and subjects moved from one forum to another by clicking a “next” button. The order of forums’ appearances was randomized. Each forum, for purposes of simplicity, included only one thread. Each thread included a root message. Replies were included in the appropriate conditions. Subjects were not allowed to create new threads and were limited to replying to the root message only. Subjects did not see each other’s contributions to the forums, in order to standardize the conditions they were exposed to.

Procedures. Subjects were assigned randomly to one of four conditions. Each group was asked to participate in a newly developed online discussion platform. Subjects were told that the goal was to test the platform for usability and technical bugs. Subjects were asked to treat the platform like any other discussion platform and participate in forums of their choice. Each subject created a username and selected the forums, one in each page. Once they reached the last page, they followed a link to an online survey, which included questions about their experience.

Reliability was fairly high for the pretests (Cronbach’s α = 0.83, for pretest 1; Cronbach’s α = 0.88, for pre–test 2). Reliability for the experiments, although within the acceptable range, was lower (KR–20 = 0.73 for experiment 1 and KR–20 = 0.75 for experiment 2). The low number of participants in each measure (20 to 25) can explain these values.
The first pre-test was designed to identify the most and least popular topics in political and current affairs for the age group (18–21) that participated in this study. Fourteen topics, ranging from international, national, and local news and politics, to more specific topics of coverage, were rated by 29 subjects for their levels of interest on a 5-point Likert scale. The least popular topic was news from the city where the university resides \((M=2.8)\). The most popular topic was national news \((M=4.57)\). Both topics kept their level of popularity across gender.

Pre-test 2

The second pre-test was designed to identify the messages that participants showed the most and least interest in. Fourteen messages on each of the two topics that were identified in pre-test 1 were collected from existing online Web forums. Each message had a title sentence and a two to three sentences body message content. Eighty-nine subjects were asked to rate the likelihood that they would reply to an online forum for each of 28 messages, on a 5-point Likert scale. The most popular message addressed gas prices \((M=4.47)\), and the least popular message addressed the election of a new superintendent for the county’s school system \((M=1.77)\). Both messages kept their levels of interest across gender. These two messages were represented in the following studies as low- and high-interest messages.

Experiment 1

Ninety-two subjects were randomly assigned to four groups \((n_{\text{High replies, High interest}}=23; n_{\text{High replies, Low interest}}=21; n_{\text{Low replies, High interest}}=20; n_{\text{Low replies, Low interest}}=22)\). Subjects were able to see the content of the root (first) message — title and message body — of each thread, but only the number of replies each thread had. The content of replies in a thread was not revealed to avoid their influence on the perceived interest in the root message. Next, in experiment 2, replies’ contents would be visible.

Findings

Subjects in the condition of high interest messages and high number of replies were more likely than subjects in other conditions to post a reply to the thread (82.6 percent of 23), followed by subjects in the condition of high interest messages and low number of replies (50 percent of 23). See Table 1. When asked after the study, 77 percent of the subjects recalled that forums indicated the number of replies posted to each thread. The four groups were not significantly or meaningfully different in term of their level of interest in politics on a 5-point Likert scale \((M=3.28, SD=1.09)\).

A logistic regression model was used to test for the relationship between the two independent variables — message interest (low vs. high) and popularity (low vs. high) — and the dependent variable — replies \((y/n)\). In step one, the two independent variables were included, and in step two, the interaction between the two dependent variables was introduced to the model. In step one, only interest was significant \((\text{Wald test } = 15.72, p<0.001)\). For the entire model, Nagelkerke \(R^2 = 0.251\). The addition of the interaction effect to the model was significant \((\text{Wald } = 4.72, p<0.05)\), with a Nagelkerke \(R^2 = 0.059\) for the model The entire model’s R Square was 0.31. See Figure 2 for an illustration of the interactive effect.

Level of interest in a message, therefore, had the most effect on whether participants would reply to it. Number of replies posted to a message had a positive influence on the number of replies a message evoked only if it contained high interest content.
Experiment 2

Over 100 (106) subjects participated in the second experiment ($n_{\text{High replies, High interest}} = 27$; $n_{\text{HL}} = 27$; $n_{\text{LH}} = 27$; $n_{\text{LL}} = 25$). This study was identical to the first with one difference — subjects were allowed to see the content of the replies. The researchers added these replies. Revealing messages' contents can increase the strength of the popularity variable. However, it also increases the danger for an influence of the message interest in the replied messages' contents on the interest in the root message. To reduce such negative influence on the study, replies were short and constructed either as a repetition of one part of the original root message and an equal number of short messages of agreement (e.g., "he has a point") and disagreement (e.g., "no way").

Findings

A logistic regression model showed a major effect for message interest (Wald test = 13.38, $p<0.001$) with a Nagelkerke $R^2 = 0.336$ for the model. Although values in the high-popularity conditions are a bit higher than in low-popularity conditions, popularity was not a significant predictor for posting replies (Table 2; Figure 2). Low-interest message conditions
forums remained lowest in terms of evoking replies, regardless of popularity. Unlike in experiment 1, however, popularity showed no effect at the high interest message threads. Here again, the four groups were not significantly or meaningfully different in term of their levels of interest in politics on a 5-point Likert scale ($M=3.31$, $SD=1.15$).

Table 2

<table>
<thead>
<tr>
<th></th>
<th>Low-Interest Message</th>
<th>High-Interest Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>Popularity L</td>
<td>7.41 percent (N=25)</td>
<td>62.96 percent (N=27)</td>
</tr>
<tr>
<td>Popularity H</td>
<td>24 percent (N=27)</td>
<td>70.37 percent (N=27)</td>
</tr>
</tbody>
</table>

* Values are percentages in cells and the number of subjects in that condition.

*For the logistic regression model: Wald test = 13.38, p<0.001, $R^2 = 0.336$.

Figure 2

Experiment 2 - Reply attraction by popularity and interest in a message

* Values are percentages in cells and the number of subjects in that condition.
Discussion

The major finding of this study is the interaction between the two independent variables — the number of replies and the interest in a message’s content. Root messages with high interest content attracted more replies than those with low interest content. However, the number of existing replies affected the number of replies posted, but only for the message that was perceived as interesting. In other words, to attract many replies, a message should have qualities that are considered to be more attractive than those of others. To become one of the very few messages that attracts a large and disproportionate number of replies in a discussion, message quality is not enough. The message should have already established a form of recognition in a group, which in this case takes the form of replies. Furthermore, as a discussion forum grows, a few messages that are perceived as having high quality will continue to attract many more replies, in part, simply because of their “popularity.” Such dynamics can advance our understanding regarding the emergence of a highly disproportionate distribution of replies across messages and, by proxy, across participants.

Findings, then, illustrate the balance between the individual and group levels in evoking discussions. Predicting replies’ postings in a forum depends first and foremost on the individual participants — the content posted and participants’ interest in it. The network dynamics level started playing a role only when the requirement at the individual level was fulfilled. The role of existing group recognition of messages, however, should not be underestimated. In times when anyone with a computer or a cellular phone with Internet access can publish information, having valuable information is not enough for its successful dissemination. Whether or not a message has established recognition regarding the value of its published content is key to understanding how a few become extremely successful in attracting readers, participants, contributors and other types of recognition over the Internet. Equally important, it explains why most messages, and therefore individuals, are much less successful in attracting audiences and participants.

The research hypotheses for this study did not call for network analysis. However, results can make a modest contribution to network literature by laying a conceptual framework for future studies on preferential attachments. In this study, two predictors for messages were examined. They are content — characteristics of a message, and number of replies — connections made by others in the group. As the interactions within this study were limited, power–law could not have been formed. However, since literature regards preferential attachment as a dynamic that together with the growth of a network leads to power–law, it can be expected in smaller groups. Findings have identified a rich–get–richer dynamic and therefore can provide limited support for the preferential attachment dynamics in small groups. Future studies can extend this conceptualization to other forms of social interactions and networks and may investigate whether the interactions found in this study can be replicated elsewhere.

In the context of discussion forums, findings have important social implications. Previous studies have illustrated how popular participants influence discussions and other participants. Honeycutt (2005) showed that “elite members” in online communities use their powers to control access from other community members and employ threats. Himelboim, et al. (2009) showed that the less than one percent of the population that attracted more than half of the replies in political forums influenced discussion topics by importing news articles from elsewhere on the Web. Whereas the unit of analysis in this study was a message, those who post them can clearly benefit from such highly popular messages. It has been established that in online forums, a few individuals and messages attract a large and disproportionate number of replies (for example, Raban and Rabin, 2007). This study illustrates the advantage a message’s popularity has, but it also shows its limitations. As long as popular content presented to the group is considered valuable in one way or another, it will evoke more discussion than other valuable, but less popular content. To the extent to which the findings can be generalized from a message to its author, one’s influence on a group holds as long as the author continues to provide valuable content. Once a participant stops making valuable contributions to the group, its previous group recognition may lose its impact.
The failure of the second experiment to replicate the interaction found in the first illustrates the methodological complexity of separating the two independent variables. In naturally occurring networks, hubs are most likely to be considered attractive in terms of forming new ties or links with them, both for their number of existing links and their unique individual characteristics. In order to understand the dynamics, and therefore the implications, we needed to isolate the two variables so that they were independent from one another. In the first experiment, we presented only the number of replies posted to a thread to avoid an effect of replies’ content on interest in the root message. In Experiment 2, subjects were allowed to see the content of the replies to make the experiment platform more similar to real–life discussion forums. Findings, however, did not show significant effect of number of replies. One interpretation is that revealing the content of the posted replies to subjects contaminated the individual effect of the root message’s content.

In other words, replies that were posted by the researcher in order to manipulate one independent variable — number of replies — might have projected on the second independent variable — interest in a message, losing the unique contribution of the first variable — interest in a message. To overcome this limitation of the study, further methodological effort is needed to keep the two independent variables orthogonal.

Limitations. This study is limited in terms the operationalization of the independent variables as nominal variables. Interest in a message had two values — low and high. Because the number of links in the network seemed to have an effect only for the high interest messages, the time at which popularity started to affect the interaction remains unclear. As we selected for this study only the messages that were ranked very low or very high in terms of participants’ interest in them, we do not know how the number of replies affects threads with less extreme levels of interest. In this study, an individual node’s characteristics in a discussion forum were conceptualized as a message. Future studies can choose a different conceptualization, such as the author’s reputation from previous interactions and other messages she posted. The design used in this study is limited to replies to a root message. Future studies could examine full threaded discussions. Further research can also focus on what motivates discussants to reply to the messages that have a higher number of replies, in addition to the contents of the message (e.g., social influence and reputation).

Conclusions

Given the opportunity to interact freely, individuals and social actors tend to follow a pattern of highly unequal distribution of connections, in which a few individuals (or “hubs”) are disproportionally more connected than others. A rich literature, popular and academic, has been devoted to examine and debate the importance of these hubs (see, for example, Nisbet, 2006; Roch, 2005; Barabási, 2002; Gladwell, 2000; and, Watts and Dodds, 2007). This study examined the initial dynamics that leads to this unique pattern of interactions. It identified two types of predictors of replies’ attraction in online discussions — individual qualities of a message and existing connections (here, replies) it has in the group. Findings suggest that in online political forums, messages with many existing replies are more likely to obtain new replies, first and foremost due to the unique qualities of the message content. Arguing that participants tend to post replies to the most popular posts because they are popular is accurate only if the popular messages have individual qualities that are considered attractive by other members in the network. These findings are clearly limited to online forums. However, this study suggests a theoretical and methodological framework to examine the growing connectivity of a few highly connected participants that can be applied to examining dynamics of many other social interactions.

About the authors

Itai Himelboim is an assistant professor in the Department of Telecommunications at the University of Georgia’s Grady College in Athens.
E–mail: itai [at] uga [dot] edu

Stephen McCreery is a doctoral student at the University of Georgia’s Grady College in Athens.
E–mail: mccreery [at] uga [at] edu

Note
References


R. Kahn and D. Kellner, 2004. "New media and Internet activism: From the 'Battle of Seattle' to blogging,' New Media & Society, volume 6, number 1, pp. 87–95.


---

**Editorial history**

Received 20 April 2011; accepted 2 March 2012.

---

Copyright © 2012, *First Monday*.

Free interactions, hierarchical structure: Factors explaining replies attraction in online discussions

by Itai Himelboim and Stephen McCreery

*First Monday*, Volume 17, Number 3 - 5 March 2012

doi:10.5210/fm.v17i3.3533

---