
January 2017

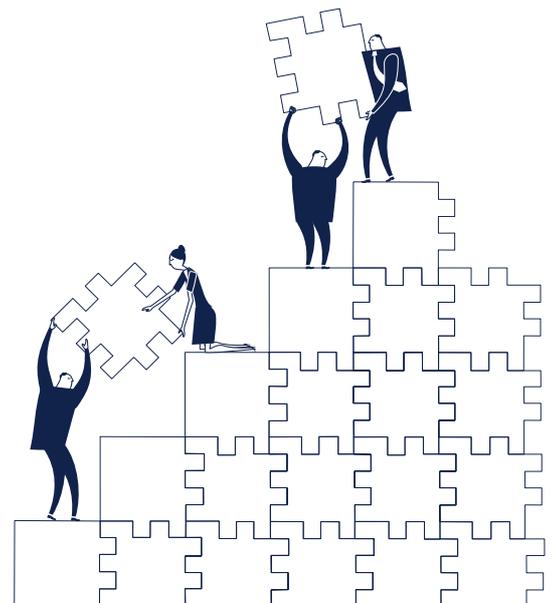
Pump It Out! The Effect of Transmitter Activity on Content Propagation in Social Media

Andrew T. Stephen
Saïd Business School, University of Oxford

Yaniv Dover
School of Business Administration, Hebrew University of Jerusalem

Lev Muchnik
School of Business Administration, Hebrew University of Jerusalem

Jacob Goldenberg
Arison School of Business, IDC Herzliya



Saïd Business School RP 2017-01

The Saïd Business School's working paper series aims to provide early access to high-quality and rigorous academic research. Oxford Saïd's working papers reflect a commitment to excellence, and an interdisciplinary scope that is appropriate to a business school embedded in one of the world's major research universities..

This paper is authorised or co-authored by Oxford Saïd faculty. It is circulated for comment and discussion only. Contents should be considered preliminary, and are not to be quoted or reproduced without the author's permission.

**PUMP IT OUT! THE EFFECT OF TRANSMITTER ACTIVITY ON CONTENT
PROPAGATION IN SOCIAL MEDIA**

Andrew T. Stephen

L'Oréal Professor of Marketing & Associate Dean for Research
University of Oxford, Saïd Business School, United Kingdom
Andrew.Stephen@sbs.ox.ac.uk

Yaniv Dover

Assistant Professor of Marketing
School of Business Administration, Hebrew University of Jerusalem, Israel
Yaniv.Dover@huji.ac.il

Lev Muchnik

Associate Professor of Information Systems
School of Business Administration, Hebrew University of Jerusalem, Israel
lev.muchnik@huji.ac.il

Jacob Goldenberg

Professor of Marketing
Arison School of Business, IDC Herzliya, Israel
jgoldenberg@idc.ac.il

Andrew T. Stephen is the L'Oréal Professor of Marketing and Associate Dean for Research at the Saïd Business School, University of Oxford, UK (Andrew.Stephen@sbs.ox.ac.uk). Yaniv Dover is Assistant Professor of Marketing at the School of Business Administration, Hebrew University of Jerusalem, Israel (Yaniv.Dover@huji.ac.il). Lev Muchnik is Associate Professor of Information Systems at the School of Business Administration, Hebrew University of Jerusalem, Israel (lev.muchnik@huji.ac.il). Jacob Goldenberg is Professor of Marketing at the Arison School of Business, IDC Herzliya, Israel (jgoldenberg@idc.ac.il). The authors thank Jonah Berger, Don Lehmann, Moshik Miller, Edith Shalev, Danny Shapira, Olivier Toubia, and Christophe Van den Bulte for their feedback on earlier versions of this manuscript.

**PUMP IT OUT! THE EFFECT OF TRANSMITTER ACTIVITY ON CONTENT
PROPAGATION IN SOCIAL MEDIA**

Abstract

People share billions of pieces of content such as news, videos, and photos through social media every day. Marketers are interested in the extent to which such content propagates and, importantly, which factors make widespread propagation more likely. Extant research considers various factors, such as content attributes (e.g., newness), source traits (e.g., expertise), and network structure (e.g., connectivity). This research builds on prior work by introducing a novel behavior-focused transmitter characteristic that is positively associated with content propagation in social media: *activity*, or how frequently a person transmits content. Evidence for this effect comes from five studies and different paradigms. First, two studies using data from large social media platforms (Twitter and LiveJournal) show that content posted by higher-activity transmitters—whom we refer to as “social pumps”—propagates more than content posted by lower-activity transmitters. Second, three experiments explore the mechanism driving this effect, showing that social media users receiving content from a social pump are more likely to retransmit it (a necessary behavior for achieving aggregate-level propagation) because they infer that content from a social pump is more likely to be current, and therefore more attractive as something to pass along through retransmission.

Social media has become a widely used communications tool for billions of people around the world. By mid-2015, three-quarters of U.S. adult Internet users regularly used social media (Perrin 2015); by 2016, 68% of the world's 3.4 billion Internet users were on social media (Kemp 2016); and Facebook, the world's largest social media platform, had 1.94 billion active monthly users by the end of March 2017 and averaged 1.28 billion active users per day (Facebook 2017). Indeed, billions of pieces of content are shared daily through social media. Given the popularity of social media, organizations of all sizes and types use social media in their attempts to spread information to their target audiences. Accordingly, marketers are interested in the extent to which content penetrates through the clutter and propagates and, critically, which factors make widespread propagation more likely. This has also become important outside of marketing, since information spreading in social media can impact public opinion, sometimes erroneously, on a large scale and with serious consequences (e.g., in 2016 in both the United Kingdom referendum on European Union membership and the United States presidential election).

Information propagation and content sharing in social media is undeniably complex and multiply determined. It is also increasingly relevant given the growing interest among both academics and practitioners in influencer marketing where firms select individual transmitters with desirable characteristics and "seed" content with them with the hope that they will share it on social media and it will propagate widely (e.g., Aral, Muchnik, and Sundararajan 2013; Chae, Stephen, Bart, and Yao 2017; Goldenberg, Han, Lehmann, and Hong 2009; Haenlein and Libai, 2013; Hinz, Skiera, Barrot, and Becker 2011; Stephen and Lehmann 2016). Although some facets and drivers of content propagation and social sharing in social media have been studied in consumer research (e.g., for reviews see Berger 2014, Lamberton and Stephen 2016), much

about social media consumer behavior is still not well understood and more work is needed (Stephen 2016). This research helps address this need by introducing a novel individual-level, behavior-focused transmitter characteristic that is positively associated with content propagation in social media: *activity*. Activity refers to the frequency with which an individual transmits (posts or shares) content (e.g., messages, photos, links, status updates) in social media. To the best of our knowledge, this characteristic has been largely overlooked by academics and practitioners, despite, as the current research suggests, its relatively strong effect on content propagation outcomes and relevant individual-level behaviors.

Instead of considering activity as a transmitter characteristic, extant studies typically consider factors associated with content attributes (e.g., newness), source or transmitter traits that tend to be more stable and/or unrelated to social media itself (e.g., expertise), and network structure (e.g., connectivity). A particular emphasis in the literature over the last decade has been on the intersection of transmitter characteristics and network structure, with a significant focus on transmitter characteristics associated with network position or connectivity, such as how many friends or followers a person has in their network (e.g., Goldenberg et al. 2009; Hinz et al. 2011; Iyengar, Van den Bulte, and Valente 2011; Stephen and Lehmann 2016; Watts and Dodds 2007). A transmitter's connectivity in a social network can be correlated with their potential for helping to widely propagate content because of the greater reach of so-called "social hubs" with their disproportionately high numbers of connections.

However, greater reach does not necessarily equate to greater content propagation in social media because the people who see the content—the receivers—must explicitly choose to retransmit (e.g., retweet, share) content if it is to have a chance of propagating. This explicit retransmission requirement is common in the vast majority of social media platforms, and means

that aggregate-level content propagation relies on individuals deliberately deciding to retransmit content. This has been largely ignored in extant research, however. Further, although not necessarily stated as such, the limits of transmitter reach are recognized by marketing practitioners, particularly within the fast-growing field of “influencer marketing.” For example, brands and marketing agencies are increasingly keen to use so-called “micro-influencers” (people who do not have extremely large numbers of social media connections or followers) because they realize that massive reach is not the only key to successful content propagation (e.g., Linton 2017).

Moving away from, but complementary to prior work, we propose that a transmitter’s activity is positively associated with content propagation because it affects receivers’ retransmission decisions. We refer to this as the activity effect, and in this paper this effect is established and explained across five studies that consider the effect at two levels using multiple methods. First, in Studies 1 and 2 we use field data from two popular social media platforms (Twitter and LiveJournal) and find a positive effect of transmitter activity on aggregate-level content propagation. Content from higher-activity transmitters—whom we refer to as “social pumps”—is found to be more likely to propagate more than content from lower-activity transmitters. Second, in Studies 3, 4, and 5, we focus on the activity effect at the individual level and show that transmitter activity positively affects receivers’ content retransmission intentions. As we mentioned earlier, a receiver explicitly deciding to retransmit a piece of content is necessary for aggregate-level content propagation in social media. Thus, to understand behaviorally why the aggregate-level activity effect found in Studies 1 and 2 occurs, it is appropriate in Studies 3, 4, and 5 to focus on individual receivers’ retransmission decisions. Finally, note that our intention is not to refute or challenge drivers covered in prior work, such as

transmitter connectivity. Rather, by introducing transmitter activity we hope to expand theoretical and managerial understandings of the complexity of content propagation in social media.

Conceptual Framework

Our main hypothesis is that a social media transmitter's level of activity is positively associated with the extent to which their transmitted content will propagate in a given social media platform. Put simply, we predict a positive association between transmitter activity and aggregate-level content propagation, which we refer to this as the activity effect. For example, all else equal, suppose there are two individuals—Donald and Ronald—on Twitter who tweet the same piece of content, such as a link to a video or a news article. If Donald has a higher activity level (e.g., higher average number of tweets per day) than Ronald, we expect that the content Donald posted will propagate more in a given period of time compared to what Ronald tweeted.

This is distinct from a connectivity (social hub) effect, which is related to how many connections transmitters have and is concerned with a property of the network structure and a person's position in it. The number of connections a transmitter has are, in part, a consequence of their own behaviors (e.g., in "friending" people or attracting "followers"; e.g., Toubia and Stephen 2013), and in part a consequence of other people's actions (e.g., others choosing to "follow" them, "friend" them, or accept a transmitter's "friend" request). A social media user's activity, however, is entirely a consequence of their own actions because others do not post on a user's behalf. Thus, the hypothesized activity effect is associated with the dynamic behavior of an individual (as a content transmitter in social media) and not groups of interdependent,

networked individuals. Although in our studies we ensure that we control for the possibility of a connectivity effect (since prior work has extensively considered it), our focus is on activity, not connectivity. This also represents a point of departure from much of the recent literature, which, as mentioned earlier, has disproportionately focused on connectivity and network-related factors and has neglected to consider other transmitter characteristics, particularly those that are derived from social media users' behaviors.

To understand why transmitter activity could positively affect aggregate-level content propagation in this type of context, it is necessary to consider how transmitter activity might affect receivers' retransmission decisions. Moreover, it is important to note that transmitter activity is an easily observable user characteristic on most social media platforms, for their direct connections (e.g., followers) and for indirect connections through observation (e.g., followers of their followers). This observability matters because, as we argue next, the hypothesized positive effect of transmitter activity on retransmission (and, ultimately, aggregate-level content propagation) is based on content receivers inferring certain content characteristics from a transmitter's activity level. Activity, therefore, must be easily observable, or relatively straightforward to discern from looking, even in a cursory manner, over a social media user's posting history.

What is likely to affect a receiver's decision to retransmit a piece of content in social media? Prior work has considered related questions and has often associated the general notion of what people share and discuss in social settings with certain content characteristics. For example, Berger and Schwartz (2011) found that interesting (e.g., newer) products were more likely to be talked about. Also, information that is perceived as fresh and current should be preferred over content that is perceived as stale and old because of inherent desires for novelty

(Campbell, Mayzlin, and Shin 2013; Hirschman 1980; Moldovan, Goldenberg, and Chattopadhyay 2011; Rogers and Shoemaker 1971; Wu and Huberman 2007). This is also consistent with research showing that people post in social media or share word of mouth (WOM) more generally because of self-focused status-enhancement motives (e.g., De Angelis, Bonezzi, Peluso, Rucker, and Costabile 2012; Toubia and Stephen 2013), since a person is more likely to accrue social status from telling others something fresh than from telling them things that are already known.

A receiver's decision to retransmit content in social media should therefore be positively associated with the extent to which they think the content is fresh or current. Or, more precisely, the extent to which they think that their potential receivers (friends/followers) will perceive the content as fresh or current. People will be averse to retransmitting content that they think their followers might have already seen (i.e., stale content) because this could lower their social status. In various social contexts, including social media, being motivated by the perceptions of others has been linked to status seeking and peer recognition (e.g., Bughin 2007; Glazer and Konrad 1996; Lerner and Tirole 2002, 2005). Accordingly, since social media users' posting behaviors are likely influenced by how others perceive them, retransmitting content that risks others perceiving oneself unfavorably—which is a plausible consequence of retransmitting stale content—is something people will likely avoid doing. For this reason, freshness is a relevant content characteristic when considering social media users' content retransmission decisions.

A significant complication is that social media users typically cannot know whether a piece of content will be perceived as fresh by their audiences (e.g., followers on Twitter). While some content is objectively fresh at a particular point in time (e.g., news when it first breaks, live-broadcast video content), the vast majority of content shared in social media lacks this

property. Even if a potential retransmitter considers a piece of content as fresh, they may not know whether others will necessarily think the same. For a retransmission decision, therefore, freshness is a content characteristic that would need to be based on the perceptions of others—not a receiver’s own perception. This complicates the decision considerably, because retransmitters are unlikely to know what their followers or friends have already seen. In fact, in most situations this would be impossible to find out, and it would also be hard to guess because each follower or friend is connected to hundreds (if not thousands) of other entities, including other people as well as media outlets, commercial content producers, and companies, each of whom is another source of content.

In this situation, we contend that social media users might adopt a heuristic-based approach to overcoming this information asymmetry whereby they attempt to infer content freshness from the information that is available to them. Although not quite the same, and not focused on social media, the persuasion literature on source credibility has considered related problems (e.g., Chaiken and Maheswaran 1994; Hovland and Weiss 1951; Karmarkar and Tormala 2010; Petty, Cacioppo, and Goldman 1981; Tormala and Clarkson 2007). Pertinent to the current research, in this literature it has been shown that source characteristics are used as a basis for inferring message characteristics (e.g., a persuasive message’s believability might be inferred from the source’s level of expertise). Based on this, in our setting it seems plausible that a specific content characteristic—perceived freshness—could be inferred from a transmitter characteristic in a heuristic-type manner.

We propose that a transmitter’s activity is used as an indicator of content freshness in such an inference-formation process. The issue for a receiver who might retransmit content is that they are very unlikely to actually know how fresh content is. Instead, they will look to the

transmitter and consider whether that person is an experienced publisher, and if he has a reputation for posting fresh, current, up-to-date things. If they do, then they will infer that content characteristic and be more likely to retransmit. We expect that the higher a transmitter's activity, the more likely it is that people will think that they have a reputation for posting fresh content. This is because activity indicates the communication frequency between a transmitter and her receivers, and communication frequency has been linked to various dimensions of credibility in other contexts, such as trustworthiness in managerial settings (e.g., Becerra and Gupta 2003) and persuasiveness in advertising (e.g., Cacioppo and Petty 1980). In a decision context where content freshness is important, communication frequency may therefore be associated with perceptions of how credible a transmitter is as a provider of fresh, up-to-date, and current information.

In summary, we hypothesize the following relationships. First, transmitter activity, a construct that was not considered before in this context, is positively associated with aggregate-level content propagation in social media and individual-level intentions to retransmit content. Second, this will occur because, for content to propagate widely in social media, individuals receiving that content from a transmitter must decide to explicitly retransmit it, and retransmission is more likely if a receiver perceives the content to be more fresh or current. Third, because receivers do not know how fresh or current content is in the eyes of their friends/followers, they will heuristically infer this by looking at a transmitter's past activity. Thus, taken together, we expect that transmitter activity will positively affect retransmission decisions because receivers infer higher content freshness from higher transmitter activity. As such, the activity effect—which we empirically demonstrate and is new in the literature—is the

result of a mechanism in which social media users infer a relevant, behavior-influencing content characteristic (freshness) from a common transmitter characteristic (activity). Stated formally:

H1a: Transmitter activity is positively associated with the extent to which content posted on a social media platform by that transmitter will propagate. (Aggregate level effect)

H1b: The higher a transmitter's activity, the more likely a receiver is to intend or decide to retransmit that content in social media. (Individual level effect)

H2a: The higher a transmitter's activity, the more likely a receiver is to infer that content from that transmitter is fresh/current.

H2b: The more likely a receiver is to infer that content from a transmitter is fresh/current, the more likely they are to intend or decide to retransmit that content in social media.

These hypotheses are tested across a number of studies. Studies 1 and 2 provide support for H1a by demonstrating the existence of the activity effect with respect to aggregate-level content propagation using field data from two large social media platforms (respectively, Twitter and LiveJournal). Study 3 provides support for H1b with an individual-level experimental demonstration of the activity effect found in Studies 1 and 2 by showing the positive effect of transmitter activity on receivers' retransmission intentions. Finally, Studies 4 and 5, also experiments, provide support for H2a and H2b by showing that transmitter activity affects content freshness (but not some other content characteristics or transmitter characteristics that might otherwise affect content-sharing actions), and that perceived content freshness mediates the effect of transmitter activity on retransmission.

Study 1: Evidence of the Activity Effect in Twitter

The purpose of Study 1 is to test H1a by providing evidence of the aggregate-level activity effect. We use field data from Twitter that captures the sharing of links (URLs pointing to Internet content) in tweets, and includes information about the Twitter users who posted those tweets. We test whether links posted by transmitters with higher levels of prior tweeting activity tend to propagate more. We also control for various other factors that could affect this outcome.

Data and Method

This dataset contains a random sample of 2,461 active, non-commercial Twitter users who were observed over a 44-day period. Similar to Toubia and Stephen (2013), who used a very similar dataset and data collection procedure, the sampled users were pre-screened to ensure that they were non-commercial users; i.e., none were media organizations, celebrities, companies, or any other account that was, to the best of our knowledge, not an individual entity. For each user on each day we collected profile data describing their connectivity (number of followers) and activity (average number of tweets per day). Each day we also collected the text of their tweets from that day. No other information about these users or their tweets was available to us. The 2,461 users posted 114,711 tweets during our observation period.

Our dependent variable was content propagation, which we tracked using links to content that people posted in tweets. Of the collected tweets, 21,430 (18.7%) contained a link to some kind of content that existed outside of Twitter on the Internet (e.g., a news article, a video, a blog post). Our analysis used only the link-containing tweets since links could be tracked and propagation of those links could be measured. The unit of analysis was the linked-to content and

not the link itself, which is important because different links can link to the same piece of content (e.g., two “short URLs” such as <http://bit.ly/abc123> and <http://bit.ly/xyz789> could both link to the same piece of content).

With this in mind, the following steps were taken to compile our content propagation dataset for analysis. First, we screened the initial set of 114,711 tweets for links (text starting with “http://”) to find the set of 21,430 link-containing tweets. Second, since most links were short URLs, we used a service called LongURL to convert these links into their original links. This allowed us to create a list of unique pieces of linked-to content. Third, we used a service called BackTweets to obtain propagation data for each unique piece of content (original link). Given an address/URL, BackTweets provided data on the number of occurrences of that address across all tweets (not only those in our sample) within a 14-day period. We therefore knew, for each tweet, the number of times its linked-to content was shared in the Twitter network in the time before it was posted, and then 14 days after it was posted. Finally, we took this data to compile our dependent variable. For each piece of linked-to content mentioned in our sample of 21,430 link-containing tweets, content propagation was measured as the number of times that piece of content was mentioned in Twitter during the 14 days *after it was first mentioned* by a Twitter user (transmitter) in our sample. Content propagation (number of mentions in 14 days) was heavily skewed, ranging from 0 to 30,204, with a mean of 117.64 mentions ($SD = 1264.64$).

We also compiled data on various content characteristics that we used as control variables, since they could also affect the dependent variable even though they are not of interest in this study. Each piece of content was evaluated by three independent judges from Amazon Mechanical Turk who had no information about the transmitter and did not know the content was linked to from Twitter. Judges evaluated the linked-to content (e.g., a YouTube video) and not

the website hosting the content (e.g., YouTube.com) on three content characteristics: perceived quality, appeal, and freshness. Higher-quality and more-broadly-appealing content likely will propagate more, irrespective of transmitter characteristics, and should therefore be controlled for. We also wanted to check whether fresher content propagated more in line with our conceptualization. Judges rated each piece of content on five items (1 = strongly disagree, 5 = strongly agree; e.g., “This content is high quality,” “Many people would find this content appealing,” and “This content is current”). There was reasonable agreement between judges, and mean ratings across judges were used in the analysis. Judges also categorized content into one of three types: blog posts, news reports, and videos/photos. The proportions of tweets in the analysis that were blogs, news, and videos/photos were, respectively, 15%, 59%, and 26%.

Results

Before discussing results, we first consider potential endogeneity concerns and explain how they were addressed. Of particular concern is that transmitter activity, our main explanatory variable, is endogenous because it can be affected by a transmitter’s previous actions and the outcomes of those actions. For example, a transmitter’s activity today might be a function of how much something she posted yesterday has been propagating. We attempted to address this by temporally separating the transmitter activity (and connectivity) measures from content propagation. We used the first 28 days of our 44-day observation period for measuring these transmitter characteristics (activity = average number of tweets per day in days 1-28; connectivity = number of followers on day 28). The other 16 days (days 29 to 44) were used for measuring content propagation; i.e., we only analyzed content linked to in tweets posted between days 29 and 44. Note that the choice of day 28 as a cutoff was arbitrary and robustness checks

cutting the data at days 14 and 21 produced consistent findings. Additionally, we only considered content that had been introduced into Twitter for the first time; i.e., content that, at the time of it being posted, had never previously been mentioned in any tweets. This reduced the possibility that outside factors unrelated to transmitter and content characteristics could drive results, ensured that a piece of content's history in Twitter would not be a major factor (e.g., due to social proof), and allowed us to examine how characteristics of the transmitter who first introduced the content into Twitter affected that content's propagation over Twitter.

After accounting for these considerations, the results reported below are based on an analysis of 4,261 pieces of content for which we had complete data on content propagation, transmitter characteristics, and judge-rated content characteristics. Although this was substantially fewer pieces of content than were initially in our dataset, this is still a large number of observations and using this reduced set is preferable because of the abovementioned concerns. We regressed content propagation on transmitter activity, transmitter connectivity (to control for their reach), the three content characteristics (quality, appeal, and freshness), and dummy variables for content type. We applied natural log transformations to the transmitter variables because they were both heavily skewed. Since our dependent variable was a count variable, and because we had multiple observations per Twitter user in our sample, we used a random effects Poisson model with a transmitter random effect.

A series of models are reported in Table 1. In each model we found a significant positive effect of transmitter activity on content propagation (for the full model, "4" in Table 1: $b = .30$, $p < .01$). This is in line with our prediction; specifically, this supports H1a. Additionally, we found no evidence of a significant effect of transmitter connectivity (number of followers) on content

propagation, suggesting that, although it probably plays a role in other outcomes in social media, it does not affect outcomes that rely on individuals' retransmission behaviors, at least not here.

[INSERT TABLE 1 ABOUT HERE]

An alternative explanation for the activity effect that is unrelated to our theory is a selection process whereby higher-activity transmitters are disproportionately likely to select content that has more appeal, is better quality, and is therefore more likely to propagate. These two content characteristics could have driven the results if higher-activity transmitters were more likely to post content with these characteristics. We checked if activity was associated with the measured content characteristics by regressing the content characteristics of quality, breadth of appeal, and freshness on activity. There were no effects of activity on appeal and freshness on transmitter activity. The effect of activity on content quality was marginally significant ($p = .09$) and *negative*, which could suggest that higher-activity transmitters in fact might share content that is slightly lower in quality. Regardless, the positive effect of activity on content propagation did not appear to be due to higher-activity transmitters posting superior content. More evidence to reject this alternative explanation is provided in Study 4.

This first study, using real-world data from links propagating on Twitter, offers initial evidence in support of the activity effect, whereby digital content in social media from higher-activity transmitters tends to propagate more. Here we found that transmitter activity affected the propagation of links during a two-week diffusion window after being first introduced into Twitter. Perhaps surprisingly, considering the literature's focus on connectivity, the activity effect is much stronger than the connectivity effect in this case. We also found that the activity effect was not driven by high-activity transmitters selecting better content.

Discussion

One finding related that warrants additional consideration is the non-significant effect of transmitter activity on judge-perceived content freshness. One might think that this goes against our conceptualization, however this is not the case. The judges in this study had no transmitter information from which they could draw inferences about content freshness. In the absence of that information it is impossible to infer content freshness from a transmitter's level of activity. In fact, the absence of a relationship between these variables here suggests that higher-activity Twitter users do not post content that is actually more or less fresh. Thus, effects of content freshness on retransmission found in our later experimental studies are likely to be due to inferences about content freshness drawn from transmitter activity, not actual content freshness.

Study 2: Evidence of the Activity Effect in LiveJournal

The novelty of the activity effect calls for a strong replication study, hence the main purpose of the second study is to provide more empirical support for the aggregate-level activity effect (i.e., H1a) using data from a different social media platform and with a larger sample size. The platform in this study is LiveJournal, which is one of the oldest and largest blog-based social media platforms in the world. In LiveJournal, each user maintains a blog that can be linked to other users' LiveJournal blogs. When a user links their blog to another user's blog they are following that blog in a similar fashion to following users on, for example, Twitter or Instagram. LiveJournal users typically share their daily experiences, political views, news, photographs, and videos on their blogs. The network structure of LiveJournal has been studied extensively (e.g.,

Brot, Muchnik, Goldenberg, and Louzoun 2016) and since its creation in 1999 has grown to be one of the most popular blogging networks.

Data and Method

The data for this study covers both network structure and publicly posted content propagation. The network data along with the detailed user profiles were collected by a crawler that iteratively traversed incoming and outgoing follower links starting from a small set of initial users. Simultaneously, all publicly accessible content posted by users in the network was recorded. Our analysis is based on content propagation data covering 83,502 separate posts made by 28,443 individual users over a 98-day period. Additionally, the data were inspected for patterns of malicious activity and the vast majority of “spammer” users (i.e., fake accounts set up purely for posting spam-like content) and other illegitimate users were not part of the sample.

Our dependent variable, content propagation, in this study is operationalized as the number of times an original post (as a piece of content) was “referenced” in later posts by users other than the user who authored it. Referencing is the primary way in which content (posts) spread through LiveJournal. The number of times a post is referenced is therefore a measure of its popularity within the network and, importantly, the extent to which that content propagated over the LiveJournal network during a given period of time. The number of references a post received ranged between 1 and 46 ($M = 3.05$, $SD = 4.77$, median = 2). For each post, this measure was taken at the end of the observation period.

For each post we knew who the author (transmitter) was as well as some information about that user and the post itself. The following variables were used to test for the activity effect: (i) activity, measured as the average number of posts per day made by a post’s transmitter

during our observation window; (ii) connectivity, measured as the number of users following a post's transmitter at the time of the post; and (iii) two observable post attributes used as indicators of a post's quality (number of words in post, number of images in post). Given the large number of posts in this dataset we were restricted in our ability to have quality (and other characteristics) judged as we did in study 1. However, in a pretest we evaluated the appropriateness of numbers of words and images as proxies for post quality. The pretest used 588 randomly selected posts from our sample, and presented them to 1,452 members of Amazon Mechanical Turk who were asked to look at the post and evaluate its quality (each post was evaluated by 2 or 3 judges). Quality was measured with five Likert-scaled items (1 = strongly disagree, 5 = strongly agree), e.g., "This is a high quality post" and "I like this post" ($\alpha = .89$). Numbers of words and images were positively correlated with quality (p 's $< .01$).

Results

Regression results are reported in Table 2. As in Study 1, a random effects Poisson model with a transmitter random effect was used to estimate the effect of transmitter activity on content propagation, controlling for transmitter connectivity and the two quality proxies. Both activity and connectivity were logged, as in Study 1. Consistent with H1a, and thus in line with Study 1, transmitter activity significantly positively affected content propagation ($b = .05$, $p < .001$). Unlike in Study 1, transmitter connectivity had a significant effect ($b = .17$, $p < .001$). The quality proxies also had significant effects on propagation, consistent with Study 1 (p 's $< .001$).

[INSERT TABLE 2 ABOUT HERE]

The findings of this study align with those of Study 1 in that they demonstrate the existence of the activity effect. Together, the first two studies provide evidence of the activity

effect on aggregate-level content propagation in two distinct social media platforms, Twitter and LiveJournal. Importantly, these two platforms have different styles of content and different ways in which content can be shared. Despite these differences, transmitter activity has a positive effect on content propagation in both cases, suggesting that this is a robust phenomenon and clearly one that is important to understand.

Discussion

A potential limitation of the data used in this study is that the time periods over which posts could generate references in some cases overlapped with the observation period used to measure transmitter activity and connectivity. This was due to how the web crawling software operated and the large size of the LiveJournal platform that made it difficult to capture user data at precise points in time repeatedly. This data limitation could be the source of an endogeneity bias in the above results because transmitter characteristics measures may not have always been perfectly temporally separated from the reference-generating dissemination process.

To address this concern we extracted a subset of the LiveJournal dataset that covered only the final 30 days of posting activity in our full dataset. All posts in this 30-day period were made after the time that the web crawler completed collecting activity and connectivity data. This guaranteed that the observed dissemination processes always started after we observed activity and connectivity and therefore there was no possibility of an endogenous feedback mechanism. As a robustness check, we estimated the same models and obtained very similar results (see Table 2 under “robustness”).

Study 3: Evidence of the Activity Effect on Receivers' Retransmission Intentions

The purpose of Study 3 is to demonstrate the individual-level activity effect (H1b) in a controlled experiment where it is directly shown that transmitter activity positively affects retransmission decisions. Based on our theory, this must occur at the individual level for the aggregate-level activity effect on content propagation to be possible. Additionally, in this study we provide some process evidence to offer an initial test of H2a and H2b (these are tested again in later studies). This is the first of a series of experimental studies that focus on the individual-level retransmission decision and how it is affected by transmitter activity.

Data and Method

One hundred and twelve members of Amazon Mechanical Turk who indicated that they were Twitter users completed this study in exchange for a small monetary payment. The sample was 45.37% female, with a mean (SD) age of 30.40 (10.07). Four participants were excluded from the analysis because they participated more than once, which resulted in a sample size of 108 Twitter users. We presented participants with a hypothetical Twitter user profile and asked them how likely they would be to retransmit ("retweet") a link transmitted by that person. We manipulated the transmitter's activity and also their connectivity. We expected activity to positively affect participants' retransmission decisions. Transmitter connectivity was also manipulated since it did have an effect on aggregate-level content propagation in Study 2 (but not in Study 1), so we wanted to further test it and, ideally, rule it out as an individual-level driver of retransmission.

Participants were randomly assigned to one of four conditions in a 2(transmitter activity: .07 vs. 12 tweets per day) by 2(transmitter connectivity: 6 vs. 693 followers) between-subjects design. Transmitter activity, or average number of tweets posted per day, was either .07 (low) or 12 (high), and transmitter connectivity, or number of followers, was 6 (low) or 693 (high). These levels were taken from the Twitter data used in Study 1 and correspond to the means of the first and fourth quartiles of these variables.

First, participants were asked to imagine that they followed a certain Twitter user (the transmitter), that this user “posted a tweet that contained a link (URL) to some content on the Internet,” and that they noticed this tweet in their Twitter “feed.” The transmitter characteristics (activity, connectivity) were presented to participants in a table similar to what is found on actual Twitter user profile pages. Participants were not shown any actual tweets or content, and were not given any information about the hypothetical linked-to “content on the Internet.” This was done so that only the two transmitter characteristics presented to participants could possibly affect retransmission decisions.

The dependent variable, retransmission intention, was measured by asking participants to indicate, from 0 to 100%, the probability that they would retransmit (“retweet”) the link to their followers. This measure of retransmission intention is consistent with WOM transmission measures used in prior research (e.g., Frenzen and Nakamoto 1993; Stephen and Lehmann 2016). Since the retransmission decision involved linked-to content, we also measured view intention (the probability of clicking on and viewing the linked-to content, 0 to 100%). This allowed us to control for general individual differences in tendencies to view content. To help test the hypothesized mechanism, we then also measured two content freshness-related perceptions on two seven-point scales (1 = strongly disagree, 7 = strongly agree; “this person

gets information sooner than others” and “the information this person posts is likely to be novel and fresh;” these items were presented randomly and masked by embedding them in a set of filler items). Finally, we took measures used for checking our manipulations.

Results

The activity and connectivity manipulations were checked by having participants rate the transmitter on three seven-point bipolar scales with respect to how they thought this user compared to other “non-commercial” (i.e., not companies, brands, organizations, celebrities) Twitter users in terms of activity (1 = “is less active in tweeting than the average user” to 7 = “is more active in tweeting than the average user”), connectivity (1 = “has many fewer followers than the average user” to 7 = “has many more followers than the average user”), and in general (1 = “is very rare, these characteristics are very uncommon” to 7 = “is very typical, these characteristics are very common”). Compared to participants in the low activity (connectivity) condition, participants in the high activity (connectivity) condition rated the target user higher on the activity (connectivity) item (activity $p < .001$; connectivity $p < .001$). There were no differences on the general comparison (p 's $> .21$), indicating that participants did not think their transmitter was unusual or extreme. Hence, our manipulations operated as intended.

We examined retransmission intention using ANCOVA with activity and connectivity as the factors, and view intention as a covariate. Including view intention as a covariate means that our dependent variable is the probability of retransmission conditional on viewing the content first. Because we do not expect all social media users to blindly retransmit links without viewing the content, the conditional probability of retransmission is more relevant in practical terms than

the unconditional probability retransmission. Least-squares means for retransmission intention are plotted in Figure 1.

[INSERT FIGURE 1 ABOUT HERE]

In line with our prediction in H1b, and conceptually consistent with Studies 1 and 2, transmitter activity had a significant positive effect on retransmission intention ($F(1, 103) = 4.71$, $p = .03$). The main effect of connectivity and the activity-by-connectivity interaction were both not significant (p 's $> .40$). The covariate (view intention) had a positive and significant effect ($F(1, 103) = 18.40$, $p < .001$). Additionally, in a separate model without this covariate the activity effect (H1b) still holds.

We next tested the proposed mechanism. Our conceptual framework argues that the positive effect of transmitter activity on retransmission intention should be mediated by the perception that the content is fresher, and that this should be because people infer content freshness from transmitter activity. If this is true, then we should also expect receivers to think that a higher-activity transmitter is more likely to have access to fresh information, which would bolster their belief that the transmitter is a credible source of fresh content that is worth retransmitting. This implies a mediation model whereby transmitter activity affects the belief that the transmitter has access to fresher information, which in turn affects perceived content freshness (H2a) and, finally, retransmission intention (H2b).

We tested this using the two measured freshness-related items described above (perception that the transmitter gets information sooner than others and perceived content freshness) in a conditional indirect effects analysis (Hayes 2013, model 4). The results were as expected and in line with H2a and H2b. The indirect effect of transmitter activity on retransmission intention, first through the transmitter-related freshness perception and then

through content freshness, was positive (95% C.I. = [.02, 2.97]). Further, when we swapped the order of the mediating variables the indirect effect was not significant (95% C.I. = [-.18, 2.21]).

Discussion

The findings in this study support our conceptualization by showing that social media (Twitter) users are more likely to want to retransmit content from higher-activity transmitters. Further, results from the mediation analysis indicate that this effect operates through content freshness-related perceptions that are triggered by activity. Interestingly, our results shed light on why freshness might be inferred from transmitter activity, suggesting it is because people think that higher-activity transmitters are likely to have access to new information before others.

Study 4: Inferences About Content Freshness from Transmitter Activity

The purpose of Study 4 is to examine in greater detail the relationship between transmitter activity and perceived content freshness by further testing H2a. In addition, we test if transmitter activity triggers other inferences, and we explore what people infer from transmitter connectivity and demonstrate that it does not trigger inferences about content freshness.

Data and Method

Three hundred and fifteen members of Amazon Mechanical Turk, who indicated that they were Twitter users, completed this study in exchange for a small monetary payment. The sample was 58.55% female, with a mean (SD) age of 29.94 (8.48). Forty of these participants were excluded because they failed an attention and comprehension check in which, toward the

end of the experiment, participants had to recall the levels of transmitter activity and connectivity in their condition by responding to multiple-choice questions with choices corresponding to ranges of activity and connectivity. A participant failed and was excluded if the ranges they selected did not match the activity and connectivity levels that they saw in the study. This left us with a sample of 275 Twitter users.

As in Study 3, participants were randomly assigned to one of four conditions in a 2(transmitter activity: .07 vs. 12 tweets per day) by 2(transmitter connectivity: 6 vs. 693 followers) between-subjects design. The procedure was similar to study 3's in that participants were asked to imagine that they followed a certain Twitter user (the transmitter), that this user "posted a tweet that contained a link (URL) to some content on the Internet," and that they noticed this tweet in their Twitter "feed." The transmitter characteristics (activity, connectivity) were presented to participants in the same way as in Study 3. Once again, participants were not shown actual tweets or content, and were not given any information about the content so that any content-related inferences they might make would be based on only the provided transmitter characteristics.

Next, we measured three content and two transmitter characteristics with multiple items each on seven-point scales (1 = strongly disagree, 7 = strongly agree). These were measured as two blocks of items (content, transmitter) with the blocks presented randomly. These items measured participants' perceptions of content and transmitter characteristics that might be inferred from the provided transmitter information.

Content characteristics were measured by asking participants to indicate their agreement with statements describing the content that they thought the transmitter likely posts (i.e., "The content/information this person posts is likely to be..."). Three content characteristics, similar to

those tested in study 1, were measured: (i) freshness (11 items; $\alpha = .93$; e.g., new, fresh, current, up to date, stale [reversed], out of date [reversed]), (ii) appeal (6 items; $\alpha = .89$; e.g., appealing to most people, interesting, fascinating, exciting), and (iii) quality (3 items; $\alpha = .79$; high quality, better than most content shared through Twitter, carefully selected or chosen). The 20 items were presented in random order, and the items corresponding to each characteristic were averaged to form three composite items for content freshness, appeal, and quality.

Transmitter characteristics were measured by asking participants to indicate their agreement with statements describing the transmitter (i.e., “This Twitter user is likely to be...”). Two transmitter characteristics were measured: (i) social status (8 items; $\alpha = .93$; e.g., important, popular, a high social status person, higher up on the social ladder), and (ii) credibility (6 items; $\alpha = .90$; e.g., credible source of information, trustworthy, well informed, a good source of information). The 14 items were presented in random order and the items corresponding to each characteristic were averaged to form two composite items for transmitter status and credibility.

Results

We regressed each content and transmitter characteristic on manipulated transmitter activity, transmitter connectivity, and their two-way interaction. Activity and connectivity were dummy coded (0 = lower, 1 = higher). Results and correlations are reported in Table 3.

[INSERT TABLE 3 ABOUT HERE]

The results confirmed our prediction (H2a) and the previous study: transmitter activity had a significant positive effect on perceived content freshness ($b = .54$, $p = .003$), but the effect of connectivity and the activity-by-connectivity interaction were non-significant. Additionally, consistent with a marginally significant negative effect of transmitter activity on content quality

found in Study 1, activity had a significant negative effect on perceived content quality ($b = -.68$, $p < .001$). Activity did not affect any of the other variables. Thus, the positive effect of transmitter activity on retransmission and, ultimately, aggregate-level content propagation is unlikely to be because receivers believe that a higher-activity transmitter's content is of a higher quality or more appealing. It is also unlikely that they retransmit it because they believe it is from a higher-status or more-credible source.

Finally, we note that transmitter connectivity affected both appeal ($b = .84$, $p < .001$) and quality ($b = .50$, $p = .013$). Importantly, however, it did not affect perceived content freshness. Connectivity also had positive effects on both perceived transmitter characteristics (status, $b = 1.51$, $p < .001$; credibility, $b = .69$, $p < .001$). Clearly, transmitter connectivity does trigger some inferences, just not the one—content freshness—that makes receivers more inclined to retransmit.

Discussion

This study showed, consistent with our conceptualization and Study 3, that people think that content from higher-activity transmitters is fresher than content from lower-activity transmitters. Importantly, participants' evaluations of content freshness were unaffected by transmitter connectivity. Connectivity, however, positively affected perceptions of content appeal, content quality, transmitter status, and transmitter credibility. Combined with the null effect of connectivity on retransmission intentions in Study 3, this suggests that while social media users do infer something from a transmitter's level of connectivity, none of those inferred characteristics are likely to affect receivers' retransmission decisions.

Study 5: Additional Evidence of the Activity Effect and Freshness-Inference Mechanism

The purpose of Study 5 is to provide a final test of the theorized process (i.e., H2a and H2b). In particular, our goal is to demonstrate the robustness of our previous findings by varying some key elements of this study's design vis-à-vis Studies 3 and 4. This study's design differed from the previous ones in two ways. First, participants were exposed to content in this study. Second, retransmission was measured as a binary choice instead of on a 0 to 100% scale.

Data and Method

Two hundred and fourteen members of Amazon Mechanical Turk completed this study in exchange for a small monetary payment. The sample was 62.50% female, with a median age of 25-34 (in this study we measured age in ranges instead of exact years; 82.69% were between 18 and 34). Six participants were excluded because they said they were not active Twitter users, which was a requirement for participation. This left us with usable data from 208 participants.

Participants were randomly assigned to one of two between-subjects conditions for transmitter activity, using the same manipulation as in Studies 3 and 4. We did not manipulate transmitter connectivity here since it had been ruled out as a factor affecting retransmission and our hypothesized mechanism in previous studies. Participants were asked to examine some tweets posted by a "randomly chosen" Twitter user who was a person (i.e., non-commercial), and, as in Studies 3 and 4, they were presented with transmitter activity information in the form of a user profile. For realism and to mitigate a potential demand effect if the user's profile only mentioned activity, we still included transmitter connectivity in the stimuli, which was six followers in both conditions (this is not uncommon for real Twitter users, most of whom have

very few followers). Setting connectivity low was also necessary to prevent any potentially confounding transmitter status-related inferences from being made (see Study 4).

Participants were asked to consider retransmitting (“retweeting”) six tweets, all of which were about books. The tweets were presented in random order. The tweets came from a 3(genre: business, fiction, health) by 2(newness: recent release, non-recent release) within-subject design nested in the single-factor between-subjects design. Genre was manipulated to test the robustness of the effects to some variations in product category. Newness was manipulated to show that the freshness-mediated effect of activity on retransmission is due to inferences about freshness drawn from transmitter activity and not actual freshness (using product newness as a proxy).

Participants were told that they would be shown six tweets about books posted by a single, randomly selected Twitter user and were shown that user’s profile. They were then shown each of the six tweets. Each tweet was presented in the following format: “This book looks like it is worth reading (*genre*). *Title by Author.*” For example, the fiction/recent tweet was “This book looks like it is worth reading (fiction). *The Storyteller by Jodi Picoult.*” At the time of running this experiment, three recent bestsellers in their respective genres were used: (i) “Contagious: Why Things Catch On” (Jonah Berger; business), (ii) “The Storyteller” (Jodi Picoult; fiction), and (iii) “Salt Sugar Fat: How the Food Giants Hooked Us” (Michael Moss; health). The three older bestsellers in their respective genres that were similar to the three recent books were used in the non-recent conditions: (i) “The Tipping Point” (Malcolm Gladwell; business), (ii) “The Da Vinci Code” (Dan Brown; fiction), and (iii) “Fast Food Nation” (Eric Schlosser; health). For each tweet participants indicated their retransmission decision as a binary choice (not retweet = 0, retweet = 1).

Finally, we measured our mediator, perceived content freshness. This was measured at the transmitter level, i.e., once and not for each tweet. We used four seven-point scales (1 = strongly disagree, 7 = strongly agree) that asked participants to indicate agreement with the statement “The information this person posted was likely to be...” (i) novel, (ii) new, (iii) fresh, and (iv) unique ($\alpha = .84$). These items were similar to and based on the larger set of content freshness items used previously.

Results

Since participants each made six binary choices we examined their retransmission decisions using a multivariate random effects binary logit model, where a participant random effect accounted for individuals’ repeated choices. Each retransmission choice was regressed on transmitter activity (low = 0, high = 1), two dummy variables for genre (fiction, health; with business was the baseline), and a dummy variable for newness (non-recent release = 0, recent release = 1). Model parameter estimates are reported in Table 4.

[INSERT TABLE 4 ABOUT HERE]

Our main results are based on the “main effects” column in Table 4, which is the model without activity-by-genre and activity-by-interactions. As expected, there was a significant positive effect of transmitter activity on the decision to retransmit a tweet about a book ($b = .66$, $p < .001$), which supports H1b. The within-subject genre dummy variable for health had a significant positive effect ($b = .48$, $p < .04$), but the dummy variable for fiction did not ($p = .13$). Genre therefore did not appear to have a strong effect on retransmission, although retransmission was generally slightly higher for the health books than for the business and fiction books. Product newness—actual, as opposed to perceived, freshness—also had no effect ($p = .93$).

The other set of results in Table 4 are for a model that included interactions between transmitter activity and the two within-subject factors (genre and newness). The intention was to see if the activity effect was moderated by these contextual factors. We did not expect any of these interactions to be significant, and none were (p 's > .29). We had no reason to predict that genre would moderate the activity effect, and newness should also not interact with it since our freshness-related process is about perceived, not actual, freshness. Further, simply because a product is newly released does not necessarily imply that a tweet about it will be seen as either fresh or stale by a receiver's followers.

Finally, we tested the full process model using a conditional indirect effects analysis (Hayes 2013, model 4). This analysis confirmed, as in previous studies, that the effect of transmitter activity on retransmission operates through perceived content freshness (95% C.I. = [.07, .30]). As in the previous studies, this is in line with H2a and H2b. Although this estimation technique can be used for binary dependent variables, it technically does not accommodate repeated observations per participant using, for example, random effects. Thus, the conditional indirect effects analysis reported here has the caveat that it does not control for this characteristic of our data, although this is unexpected to be an issue since separate random effects regressions for the pathways in the mediation model produced robust results. Specifically, adding perceived content freshness to the random effects binary logit model showed a significant positive effect of freshness on retransmission choice ($b = .73, p < .001$) and a smaller but still significant effect of transmitter activity ($b = .57, p < .01$). And a random effects regression of perceived freshness on transmitter activity, with genre and newness dummy variables as covariates, showed a significant positive effect of activity on perceived content freshness ($b = .24, p < .001$).

Discussion

This final study replicated findings from our other studies and provided support for our hypothesized process. Importantly, unlike the previous studies, participants in this study made retransmission decisions with respect to actual content in the context of Twitter. A key finding of this study is that it did not appear to matter whether or not the content referenced a recently released product (i.e., a product that is actually fresh). Given that genre also did not appear to matter and that the only other information available was the transmitter's level of posting activity, the absence of moderation by product newness strengthens our argument that the content-freshness inference is drawn from transmitter activity and is unlikely to be affected by other contextual factors, even those that are seemingly relevant like newness because they are related to (actual) freshness.

General Discussion and Implications

Social media's popularity as a way for people to communicate with each other has grown over the past decade. However, research on how consumers use social media platforms to do various things and achieve certain goals remains relatively scant, particularly research that attempts to understand social media user behavior at the individual level (Lamberton and Stephen 2016; Stephen 2016). Although many things can be done on myriad social media platforms, a very common activity is content sharing and the widespread propagation of shared content in social media is an important outcome in settings inside business (e.g., marketing) and outside business (e.g., politics). What drives content propagation and information diffusion over the social networks underpinning most social media platforms is a critical question that has been

considered in extant research. However, as we discussed earlier, most prior work has considered either content-related factors or characteristics associated with a transmitter's network position, particularly their connectivity (e.g., number of friends/followers).

The current research instead has introduced a new type of transmitter characteristic—activity—as a predictor of aggregate-level content propagation. Transmitter activity, as we demonstrated at the aggregate level in Studies 1 and 2 with field data from Twitter and LiveJournal, is a positive driver of content propagation. Put simply, content posted by social media users with higher levels of activity—social pumps—appears to propagate more than content posted by users with lower levels of activity. This activity effect was found after controlling for some content characteristics and, importantly, transmitter connectivity. This is an important effect given that it expands the set of identifiable and measurable transmitter characteristics that are positively associated with content propagation in social media.

In addition to demonstrating this effect at the aggregate level, we also offered an explanation for how this effect operates in Studies 3, 4, and 5 with a series of experiments. Given that the aggregate outcome of content propagation occurs at the level of the network it is difficult to directly test an individual-level, behavioral mechanism driving the aggregate-level effect. However, our approach was to focus on an individual-level behavior that is necessary for an aggregate-level content propagation outcome on the majority of social media platforms: retransmission. As we explained earlier, for content to spread on platforms like Facebook, Instagram, and Twitter, as well as for other kinds of social media such as blog posts, users must deliberately pass it on to members of their networks by retransmitting it through some explicit user action (e.g., clicking “Share” or “Retweet” buttons). Thus, our strategy for uncovering an underlying behavioral mechanism for the activity effect was to focus on understanding why

receivers, as possible retransmitters, are more inclined to retransmit something when it comes from a higher-activity transmitter than when it comes from a lower-activity transmitter.

As we demonstrated across Studies 3, 4, and 5, the transmitter activity effect on retransmission appears to occur because receivers infer that content posted by higher-activity transmitters is likely to be fresher, i.e., more current and up-to-date. And, critically, this perceived content freshness positively affects their tendency to retransmit. We also showed that transmitter activity does not trigger other content- or transmitter-related inferences that affect retransmission behavior, and that the other previously studied transmitter characteristic of connectivity does not affect perceived content freshness (even though it affects other perceptions that are unrelated to retransmission decisions).

An important contribution of the current research is the set of findings pertaining to inferred content freshness. It appears that social media users consider certain types of information about content transmitters to be indicative of the types of content those people post. Interestingly, social media users infer content characteristics from aspects other than content, in this case, a transmitter characteristic. This is the case in cases where there is no information about the content available or, more realistically, even when the content itself can also be assessed. As we suggested, when on the receiving end of content, social media users do not necessarily know a piece of content's fitness for retransmission, so they instead use heuristics to help make their retransmission decisions. We focused on inferring content freshness from a transmitter's level of activity, though it is conceivable that other content characteristics might be inferred from other transmitter characteristics (e.g., some evidence for this was seen in Study 4 with respect to connectivity), and that some of these might affect social media user behaviors other than retransmission.

A number of practical implications can be derived from our findings. For firms trying to virally spread information across social media platforms (influencer marketing), our findings suggest that a person's social media posting activity should be an important factor considered when searching for "influencers" or "seeds" for these campaigns and marketing programs. Presently, firms tend to select on the basis of a person's connectivity (since this corresponds to reach). This is not unreasonable, but if marketers hope for content to not only be seen but also be passed along so that it propagates widely, activity should also be considered.

Another interesting practical implication is that activity may prove to be more generally implementable by marketers. This is because a social media user's activity is a function of their own behavior and not contingent on what others are doing. Thus, if a firm identifies potential seeds or influencers who possess other desirable characteristics but are not highly active, they could take steps to increase their activity levels (e.g., through incentives, helping them with posting, training). Connectivity, however, is relatively harder to change or influence without resorting to unethical and inauthentic practices (e.g., "buying" followers). This is because a social media user's connectivity, and their position in the underlying network more generally, depends on not only their actions but also the actions of others in the network.

Finally, often firms do not have a complete map of the network, hence many connectivity-related, centrality-based characteristics cannot be calculated. This brings into question the practicality of using connectivity for influencer or seed selection. Activity, however, is a very simple measure that does not require any information about the underlying network. This arguably makes activity more generally useful in practice.

In conclusion, across five studies combining analyses of large real-world social media data and experiments, we demonstrate and explain the activity effect on content propagation.

This research contributes to the literature on how individuals use social media and provides insights into both an important aggregate-level outcome in content propagation and a critical individual-level behavior in retransmission. More research on other social media behaviors, inferences, and transmitter characteristics is needed, however, to develop a fuller understanding of how people use social media. We hope this research encourages work along these lines.

References

- Aral, Sinan, Lev Muchnik, and Arun Sundararajan (2013), "Engineering Social Contagions: Optimal Network Seeding and Incentive Strategies," *Network Science*, 1 (2), 125-153.
- Becerra, Manuel and Anil K. Gupta (2003), "Perceived trustworthiness within the organization: The moderating impact of communication frequency on trustor and trustee effects," *Organization Science*, 14 (1), 32-44.
- Berger, Jonah (2014), "Word of Mouth and Interpersonal Communication: A Review and Directions for Future Research," *Journal of Consumer Psychology*, 24 (4), 586-607.
- Berger, Jonah and Katherine L. Milkman (2012), "What Makes Online Content Viral?" *Journal of Marketing Research*, 49 (2), 192-205.
- Berger, Jonah and Eric M. Schwartz (2011), "What Drives Immediate and Ongoing Word of Mouth?" *Journal of Marketing Research*, 48 (5), 869-880.
- Brot, Hilla, Lev Muchnik, Jacob Goldenberg, and Yoram Louzoun (2016) "Evolution Through Bursts: Network Structure Develops Through Localized Bursts in Time and Space," *Network Science*, 4 (3), 293-313.
- Bughin, Jacques R. (2007), "How companies can make the most of user-generated content," *The McKinsey Quarterly* (August).
- Cacioppo, J. T. and Petty, R. E. (1980), "Persuasiveness of communications is affected by exposure frequency and message quality: A theoretical and empirical analysis of persisting attitude change," In J. H. Leigh & C. R. Martin (Eds.), *Current issues and research in advertising* (pp. 97-122). Ann Arbor, MI: University of Michigan Press.

- Campbell, Arthur, Dina Mayzlin, and Jiwoong Shin, (2013), "Managing Buzz," working paper, Yale University.
- Chae, Inyoung, Andrew T. Stephen, Yakov Bart, and Dai Yao (2017), "Spillover Effects in Seeded Word-Of-Mouth Marketing Campaigns," *Marketing Science*, forthcoming.
- Chaiken, Shelly and Maheswaran, Durairaj (1994), "Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment," *Journal of Personality and Social Psychology*, 66 (3), 460-473
- De Angelis, Matteo, Andrea Bonezzi, Alessandro M. Peluso, Derek D. Rucker, and Michele Costabile (2012), "On Braggarts and Gossips: A Self-Enhancement Account of Word-of-Mouth Generation and Transmission," *Journal of Marketing Research*, 49 (4) 551-563.
- De Vries, Lisette, Sonja Gensler, and Peter S.H. Leeflang (2012), "Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing," *Journal of Interactive Marketing*, 26 (2), 83-91.
- Facebook (2017), *Company Info* (accessed 5/14/17, available at <http://newsroom.fb.com/company-info>).
- Glazer, Amihai and Kai Konrad (1996), "A signaling explanation for private charity," *American Economic Review*, 86 (4), 1019-1028.
- Goldenberg, Jacob, Sangman Han, Donald R. Lehmann, and Jae Weon Hong (2009), "The Role of Hubs in the Adoption Process," *Journal of Marketing*, 73 (2), 1-13.
- Haenlin, Michael and Barak Libai (2013), "Targeting Revenue Leaders for a New Product," *Journal of Marketing*, 77 (3), 65-80.
- Hayes, Andrew F. (2013), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford Press.
- Hinz, Oliver, Bernd Skiera, Christian Barrot, and Jan U. Becker. (2011), "Seeding Strategies for Viral Marketing: An Empirical Comparison," *Journal of Marketing*, 75 (6), 55-71.

- Hirschman, Elizabeth C. (1980), "Innovativeness, Novelty Seeking, and Consumer Creativity," *Journal of Consumer Research*, 7 (3), 283-295.
- Hovland, Carl I. and Walter Weiss (1951), "The Influence of Source Credibility on Communication Effectiveness," *Public Opinion Quarterly*, 15 (4), 635-650.
- Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente (2011), "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science*, 30 (2), 195-212.
- Karmarkar, Uma R. and Zakary L. Tormala, (2010), "Believe me, I have no idea what I'm talking about: The effects of source certainty on consumer involvement and persuasion," *Journal of Consumer Research*, 46, 1033-1049.
- Kemp, Simon (2016), *Digital in 2016: We Are Social Annual Report* (accessed 9/26/16, available at <http://wearesocial.com/uk/special-reports/digital-in-2016>).
- Lamberton, Cait and Andrew T. Stephen (2016), "A Thematic Exploration of Digital, Social Media, and Mobile Marketing: Research Evolution from 2000 to 2015 and an Agenda for Future Inquiry," *Journal of Marketing*, 80 (6), 146-172.
- Lerner, Josh and Jean Tirole (2002), "Some Simple Economics of Open Source," *Journal of Industrial Economics*, 50 (2), 197-234.
- Lerner, Josh and Jean Tirole (2005), "The Economics of Technology Sharing: Open Source and Beyond," *Journal of Economic Perspectives*, 19 (2), 99-120.
- Linton, Pauline (2017), "Building an Influencer Marketing Strategy," *Edelman Digital* (accessed 5/14/17, available at <http://www.edelman.com/post/building-an-influencer-marketing-strategy>).
- Moldovan, Sarit, Jacob Goldenberg, and Amitava Chattopadhyay (2011), "The different roles of product originality and usefulness in generating word-of-mouth," *International Journal of Research in Marketing*, 28 (2), 109-119.
- Perrin, Andrew (2015), "Social Media Usage: 2005-2015," *Pew Research Center* (accessed September 12, 2016), [available at <http://www.pewinternet.org>]

- Petty, Richard E., John T. Cacioppo, and Rachel Goldman (1981), "Personal involvement as a determinant of argument-based persuasion," *Journal of Personality and Social Psychology*, 41, 847-855.
- Rogers, Everett and F. Shoemaker (1971), *Communication of Innovations*. New York, NY: Free Press.
- Stephen, Andrew T. (2016), "The Role of Digital and Social Media Marketing in Consumer Behavior," *Current Opinion in Psychology*, 10, 17-21.
- Stephen, Andrew T. and Donald R. Lehmann (2016), "How Word-of-Mouth Transmission Encouragement Affects Consumers' Transmission Decisions, Receiver Selection, and Diffusion Speed," *International Journal of Research in Marketing*, 33 (4).
- Tormala, Zakary L. and Joshua J. Clarkson (2007), "Assimilation and contrast in persuasion: The effects of source credibility in multiple message situations," *Personality and Social Psychology Bulletin*, 33, 559-571.
- Toubia, Olivier and Andrew T. Stephen (2013), "Intrinsic Versus Image-Related Motivations in Social Media: Why Do People Contribute Content to Twitter?" *Marketing Science*, 32 (3), 90-102.
- Watts, Duncan J. and Peter Sheridan Dodds (2007), "Influentials, Networks, and Public Opinion Formation," *Journal of Consumer Research*, 34 (4), 441-458.
- Wu, Fang and Bernardo A. Huberman (2007), "Novelty and collective attention," *Proceedings of the National Academy of Sciences*, 104 (45), 17599-17601.

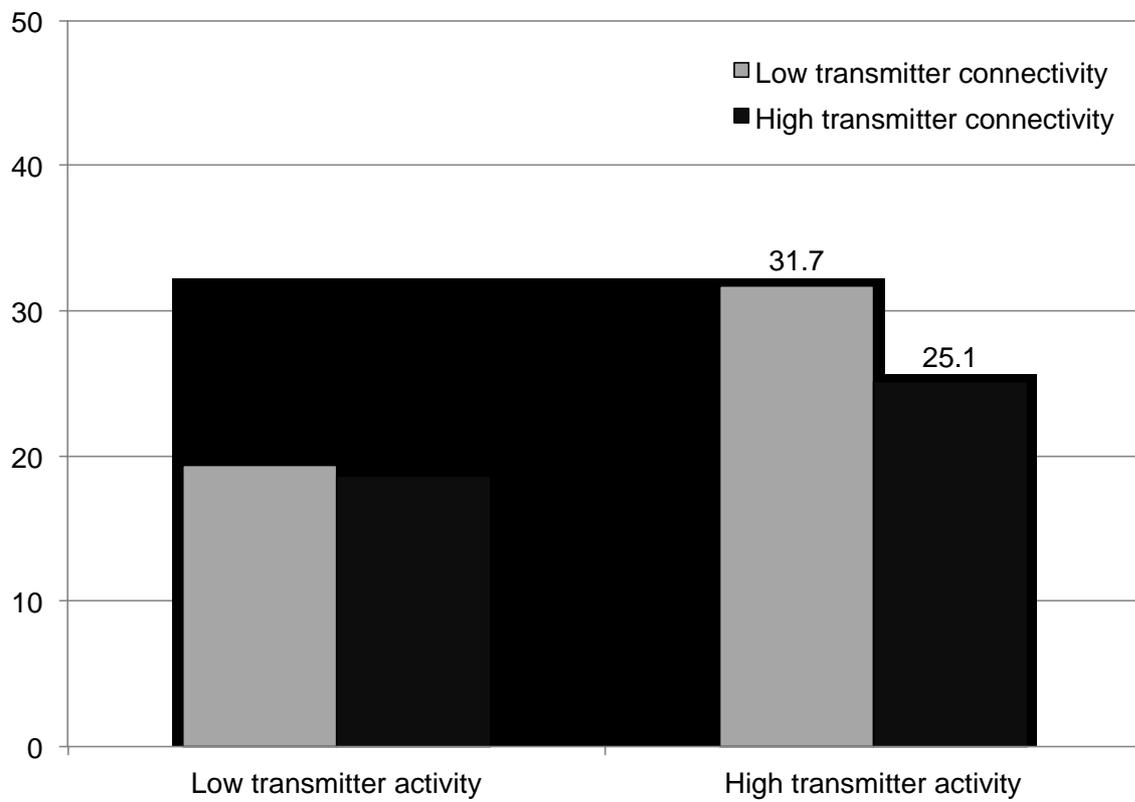
Figure 1: Retransmission intentions, 0-100% (Study 3)

Table 1: Effect of activity on content propagation in Twitter (Study 1)

Parameter	1	2	3	4	5
	Estimate (t-value)	Estimate (t-value)	Estimate (t-value)	Estimate (t-value)	Estimate (t-value)
Intercept	.83 (1.84)	.96 (5.65)**	1.07 (2.38)*	-3.49 (-7.76)**	-3.64 (22.75)**
log(Activity)	–	.29 (2.91)**	.30 (2.73)**	.30 (3.01)**	.28 (3.11)**
log(Connectivity)	.08 (.80)	–	-.03 (-.31)	-.04 (-.39)	–
Content: Quality	–	–	–	.52 (26.00)**	.52 (26.02)**
Content: Broadly appealing	–	–	–	.41 (51.25)**	.41 (50.95)**
Content: Current	–	–	–	.42 (46.67)**	.42 (46.73)**
Type: blog post ^a	.06 (6.00)**	.06 (6.02)**	.06 (6.01)**	.26 (25.98)**	.26 (26.01)**
Type: news report/article ^a	.84 (8.40)**	.84 (8.39)**	.84 (8.42)**	.83 (83.10)**	.83 (82.92)**
Random effect variance	5.41 (10.82)**	5.24 (10.69)**	5.26 (10.74)**	5.04 (10.72)**	5.02 (10.68)**
Number of tweets	4,261	4,261	4,261	4,261	4,261
-2 log likelihood	1,411,840	1,411,811	1,411,819	1,384,844	1,384,832
BIC	1,411,882	1,411,853	1,411,869	1,384,919	1,384,899

* $p < .05$, ** $p < .01$. ^a Dummy variable. The base level for content type was video/photo.

Table 2: Effect of activity on content propagation in LiveJournal (Study 2)

Parameter	Main: Estimate (t-value)	Robustness: Estimate (t-value)
Intercept	1.10 (203.88) ^{***}	1.04 (82.52) ^{***}
log(Activity)	.05 (7.31) ^{***}	.13 (9.90) ^{***}
log(Connectivity)	.17 (25.60) ^{***}	.09 (6.02) ^{***}
Number of words	.04 (7.35) ^{***}	.02 (1.17)
Number of images	.03 (7.12) ^{***}	.04 (3.16) ^{**}
Random effect variance	7.10 (202.86) ^{***}	6.16 (83.08) ^{**}
Number of posts	83,502	13,807
-2 log likelihood	309,126	49,884
BIC	309,194	49,941

* $p < .05$, ** $p < .01$, *** $p < .001$. All parameter estimates are standardized.

Table 3: Effects of activity on perceived content and transmitter characteristics (Study 4)

Parameter	Content Characteristics			Transmitter Characteristics	
	Freshness	Appeal	Quality	Social Status	Credibility
	Estimate (t-value)	Estimate (t-value)	Estimate (t-value)	Estimate (t-value)	Estimate (t-value)
Intercept	3.68* (28.61)	3.59* (30.33)	4.10* (28.72)	2.61* (21.52)	3.38* (26.45)
Activity	.54* (3.04)	.12 (.72)	-.68* (-3.44)	-.03 (-.20)	.10 (.58)
Connectivity	.30 (1.67)	.84* (5.05)	.50* (2.51)	1.51* (8.88)	.69* (3.88)
Activity x Connectivity	.18 (.71)	-.26 (-1.11)	-.24 (-.85)	.26 (1.10)	-.10 (-.42)
Correlations					
Appeal	.63*				
Quality	.41*		.68*		
Social status	.41*		.55*		.37*
Credibility	.57*		.73*		.66*

* $p < .01$.

Table 4: Effects on retransmission decision (Study 5)

Parameter	Main Effects	Main Effects and Interactions
	Estimate (t-value)	Estimate (t-value)
Intercept	-2.71** (-11.45)	-2.86** (-8.75)
Transmitter activity (0 = low, 1 = high)	.66** (3.63)	1.09** (2.72)
Genre: fiction (0 = no, 1 = yes)	.37 (1.52)	-.51 (-1.11)
Genre: health (0 = no, 1 = yes)	.48* (2.13)	1.04** (3.04)
Newness (0 = non-recent, 1 = recent)	-.02 (-.09)	-.25 (-.87)
Activity x Genre: fiction		-.26 (-.45)
Activity x Genre: health		-.35 (-.82)
Activity x Newness		.39 (1.04)
Participant random effect variance	.99** (24.98)	.99** (24.97)
Number of observations	1,248	1,248
-2 log likelihood	6,348.22	6,615.84
BIC	6,395.44	6,551.68

* $p < .05$, ** $p < .01$.