Nurturing Online Communities: An Empirical Investigation

Sofia Bapna

sbapna@umn.edu

University of Minnesota, Carlson School of Management

Mary J. Benner

University of Minnesota, Carlson School of Management

Liangfei Qiu

University of Florida, Warrington College of Business

ABSTRACT

Online brand communities can be valuable to firms, but how do firms cultivate such communities? We find that engagement, that is, 'likes' in response to firm posts, in the online brand community is associated with subsequent growth in the community. We theorize that when individuals engage with firms' posts, the social media platform broadcasts such interactions to others, who are not necessarily part of the firm's online brand community. Such social diffusion of information about the interaction and the related firm content provides individuals with new information about the firm, based on which they may decide to join the brand community of the firm. We find that firm posts that convey firm credibility through product and industry knowledge; convey organizational achievements through information about firm milestones partnerships or awards; seek opinions; and convey promotions or offers, are associated with engagement. Such posts have a significantly greater effect on engagement for early stage brand communities, that is, for those in their first year and a half, than for later stage brand communities.

Keywords: Online community, social media, online brand community, sensegiving, influencing perceptions, monetary incentives

1. INTRODUCTION

Social media have become exceedingly popular among both individuals and businesses. For instance, Facebook, the largest online social network, has two billion members (Constine 2017), and 65 million business pages (Kaplan 2017) where firms create their own brand communities. The allure of such communities to businesses is the low financial cost of interacting with potential stakeholders and their ability to provide valuable benefits (Miller et al. 2009). Extant literature has established that brand communities are associated with information dissemination (Goh et al. 2013), increased brand awareness (Hoffman and Fodor 2010), brand building, positive word of mouth, increased ROI (Kumar et al. 2013), increased profitability (Rishika et al. 2013), customer satisfaction, and loyalty (Culnan et al. 2010).

To be able to provide benefits to firms, online communities must maintain both a large pool of members and a large volume of communication activity or content (Butler 2001). The former is important because members are a resource, and at least up to a point, larger membership provides greater benefits, as the number of possible interactions between members, and the audience size for announcements and visibility increases with membership (Butler 2001). However, growing online brand communities can be challenging. A Deloitte survey found that most business efforts to build online communities failed to attract a critical mass of members even in the face of substantial spending (Worthen 2008). Moreover, there is a wide variation in even Fortune 500 companies' ability to build viable online communities (Culnan et al. 2010). Surprisingly, there is almost no research that theorizes about the strategies that firms use to grow their online brand communities. We seek to fill this gap in research, and our first research question is to explore how firms grow their fledgling online brand communities through the firms' use of communications, in the form of posts.

We posit that when individuals engage with firm posts (e.g., by liking the post), it is likely to stimulate growth in the online community. This is because social media acts like a "giant word of mouth machine" accelerating and catalyzing the distribution of information (Gallaugher and Ransbotham 2010) through automatic notifications. Moreover, online community participants adjust their thoughts and actions in light of such information from socially relevant peers within the community (Miller et al. 2009). When individuals engage with firm content, such interactions are broadcast, by the social media platform, to others who are not necessarily part of the firm's online brand community (Kane et al. 2014). Such social diffusion of information about the interaction and the related post content provides individuals with new information about the firm, based on which they may decide to join the firm's brand community.

A consequent and important question is what types of firm posts are associated with engagement (likes, shares and comments) in online brand communities. Given that guidance from prior literature on this topic is limited, this is our second research question. Prior information systems literature provides high level direction on content strategies associated with engagement, including: client focused postings (Miller and Tucker 2013), a continuous supply of compelling content (Culnan et al. 2010), providing quality content and fostering member embeddedness (Porter and Donthu 2008). This literature has only recently begun to delve into what specific type of content is associated with engagement in brand communities. Lee et al. (2017) find that posts related to brand personality (e.g., humor, emotional appeals, casual banter, philanthropic efforts) have a positive effect while informative content (that can be used to optimize purchasing decisions for e.g., price, availability, promotions, product facts) by itself is associated with lower levels of engagement. Combining informative content along with brand personality content improves engagement levels. Moreover, Aral et al. (2013) in the introduction to a special issue on social media and business transformation make a call for research that "looks deeper into the questions of what type of firm communications provoke what type of response." Marketing literature posits that firm posts with informational and entertaining content affect engagement in online brand communities (De Vries et al. 2012) and an industry white paper (Cvijikj and Michahelles 2013) finds empirical evidence for this association, as well as for the association between posts about remuneration and engagement. This review of prior research suggests the opportunity for additional research to help us better understand firms' brand community posting strategies that are associated with engagement.

Gallaugher and Ransbotham (2010) argue that when used effectively, firm postings underpin brand positioning and perceptions. Thus, we draw from literature on sensegiving, which is the process of influencing the perceptions of others (Gioia and Chittipeddi 1991), to hypothesize about the type of information that firms might convey through posts to engage their community. Specifically, we draw from prior research on sensegiving through firms' use of symbolic actions (e.g., Zott and Huy, 2007) and hypothesize that posts that convey credibility, convey professional organizing, convey organizational achievements, and seek opinions are associated with engagement. An alternate approach is to engage individuals through communications about monetary incentives, for example saving money or winning a prize, which has been identified in prior literature as a motivator for engagement (Jarvenpaa and Tuunainen 2013; Muntinga et al. 2011). Thus, we also hypothesize about the positive effect of monetary incentives communicated by firms, through posts, on engagement. Posts that convey monetary incentives (e.g., offers, promotions or contests) are likely to entail a different financial cost to a firm than posts that influence perceptions through content that conveys credibility, conveys professional organizing, conveys organizational achievements, and seeks opinions. Hence, it is worthwhile contrasting the effectiveness of these two strategies.

We focus our study on firms' use of the brand pages (profile pages hosted by firms and which focus on the firm hosting the page) on the social networking site, Facebook, because of its widespread adoption. We collected weekly data on online retail firms that were founded in 2010, including their post content, and engagement, i.e. 'likes' in response to firm posts from social media users who see this content, and the growth in the firms' online communities as represented by their fan base each week. The 9,470 posts in our final sample spanned the period from when the firms joined Facebook, in 2010, to July 2012. We used Amazon's Mechanical Turk workers to manually analyze and categorize the content of posts along thirteen different dimensions corresponding to content that conveys credibility, conveys professional organizing, conveys organizational achievements, seeks opinions, and offers monetary incentives. This process included viewing linked videos and reading all linked articles and blog posts, which allowed for an in-depth analysis of the content of each post. We used the data to analyze the association between engagement and online brand community growth, and the association between different types of post content and engagement.

As is the case with much empirical social media research, tests of our hypotheses are vulnerable to endogeneity issues due to reverse causality and the possible bias from omitted variables. For example, it is conceivable that a firm's online community as well as their content posting strategy grows because of existing status or brand effects that persist from an earlier period. In addition, there may be omitted variables that impact one of the explanatory variables as well as our dependent variables, that is, the size of the online community and engagement. We address the first concern, at least in part, by using a sample that consists of new firms that are unlikely to have developed the history, brand, and status that is associated with well-established firms. Our use of panel data with firm fixed effects and firm age controls also addresses omitted variables that are both time invariant, or that change over time but stay constant across individual firms (Hsiao 2003). To address time varying omitted variables that are specific to particular firms, we control for events like securing funding, acquisitions or mergers, winning awards and introducing new product lines. Additionally, we use the Google trends search volume of each firm as a control for the level of interest in the firm (Geva et al. 2017), which further helps us account for time-varying individual firm characteristics (Chen et al. 2015). Finally, we use propensity score matching (PSM) to try to address endogeneity due to unobserved variables that might affect engagement as well as the explanatory variables.

It is also conceivable that our theory does not identify certain types of posts used by firms, which might be associated with engagement. To help us identify such latent topics that firms may be posting about, we use text mining and let the data speak for themselves. Specifically, we also use Latent Dirichlet Allocation (LDA), a type of topic modeling method based on text mining, to identify post topics in an automated and scalable manner (Blei et al. 2003). The LDA algorithm converts each post into a distribution of a predetermined number of algorithmically identified topics. The strength of this methodology is that each post is categorized into many topics with different weights, instead of simple binary categorization based on the presence or absence of each topic. We find consistent results after controlling for the topics generated from LDA, which is indicative of the robustness of our findings. Our approach of using an empirical model that combines explanatory variables that are informed by theory along with algorithmically identified controls that are constructed by allowing the data to speak for themselves provides cleaner identification of the parameter of interest.

We find that engagement is positively associated with online brand community growth. Our analysis also confirms a positive association between firm communications that influence perceptions as well as those that convey monetary incentives, and engagement. Specifically, we find that firm posts that influence perceptions including those that convey firm credibility through product and industry knowledge; convey organizational achievements through information about firm milestones; partnerships or awards; and seek opinions, are positively associated with engagement. The association between product and industry knowledge and engagement holds for different approaches to conveying this information, including conveying tips or suggestions, industry information, design origin information and collections or picks. In addition, we find that firm posts which convey monetary incentives in the form of promotions or offers are also positively associated with engagement.

We examine the heterogeneous effects of the different types of post content i) at different times in an online brand community's life cycle, and ii) on posts that are ex-ante likely to receive lower levels of engagement versus those that are ex-ante likely to receive higher levels of engagement. This helps us go beyond simply identifying that some posts are more likely to be associated with engagement. For all the post types that are significantly associated with engagement in our study, we find that the magnitude of their effect is significantly higher for early stage online brand communities, that is, those in their first year and a half, than for later stage online brand communities. Additionally, our quantile regressions suggest that a firm's post that is ex-ante likely to receive lesser engagement (e.g., a post about a niche product) might benefit from firm content that i) conveys organizational achievements through information about firm milestones, partnerships or awards ii) conveys product and industry knowledge through either design origin information or collections and picks. In contrast, a post that is ex-ante likely to receive higher levels of engagement (e.g., a post about a mass appeal product) might benefit from firm content that i) seeks others' opinions ii) conveys promotions or offers iii) conveys product and industry knowledge through content about the industry, for example, news, anecdotes, or feature articles about the industry.

Our empirical investigation of important online phenomena contributes to understanding how to apply the theory of sensegiving and symbolic actions to building online brand communities. Although prior studies provide theoretical insights into sensegiving and symbolic actions in general, the empirical implementation of these theoretical insights is still in its infancy, particularly in online settings. Practitioners cannot work only with the conceptual models. Instead, practice calls for an engineering approach: how to design concrete strategies to increase the engagement level of online communities? The answer to this question does not merely depend on how well we understand the general principles of sensegiving and symbolic actions, but also on how well we can bring this knowledge to bear on practical online community design and implementation. To bridge the gap between theory and practical implementation, we take a step further and investigate how firms use different social media strategies to nurture online communities. In particular, we study what types of postings by firms might influence engagement in their online brand communities.

We also contribute to prior research on sensegiving and symbolic actions. Petkova et al. (2013) suggest that seeking opinions and receiving feedback through “face-to-face” social interactions at interactive events such as conferences provide an opportunity for sensegiving. We find that seeking opinions in the social media context is also effective. Thus, we extend the boundary of this sensegiving action beyond face to face interactions to also include interactions via social media, which potentially allows for a much larger reach.

2. THEORY AND HYPOTHESES

In this section, we first hypothesize about the association between user engagement in the online brand community and growth in the community. We then hypothesize about the association between different types of firm posts that influence perceptions - as well as those that convey monetary incentives - and engagement. Figure 1 summarizes our hypotheses.

We draw from prior research on sensegiving and firms’ use of symbolic actions to convey symbolic meaning (e.g., Zott and Huy, 2007), to hypothesize about the association between four broad categories of posts that convey symbolic meaning and engagement. These include posts that convey credibility, convey professional organizing, convey organizational achievements, and seek opinions (corresponding to hypotheses 2-9). Symbolic meaning is evaluated according to subjective criteria like values, feelings, and predilections of the observer (Rafaeli and Vilnai-

Yavetz 2004b). Section 2.2 below describes how each of these post types convey symbolic meaning.

Additionally, we draw from advertising literature on the use of economic incentives (e.g., Muntinga et al. 2011) to hypothesize about the association between posts with content about monetary incentives and engagement. In contrast to symbolic meaning which is evaluated according to subjective criteria, intrinsic purpose is evaluated according to economic or performance criteria (Rafaeli and Vilnai-Yavetz 2004b) as is the case with posts conveying monetary incentives (corresponding to hypotheses 10 and 11, described in Section 2.3 below). Table A1 in Appendix 1 provides an example from each high-level category of post in our hypotheses.

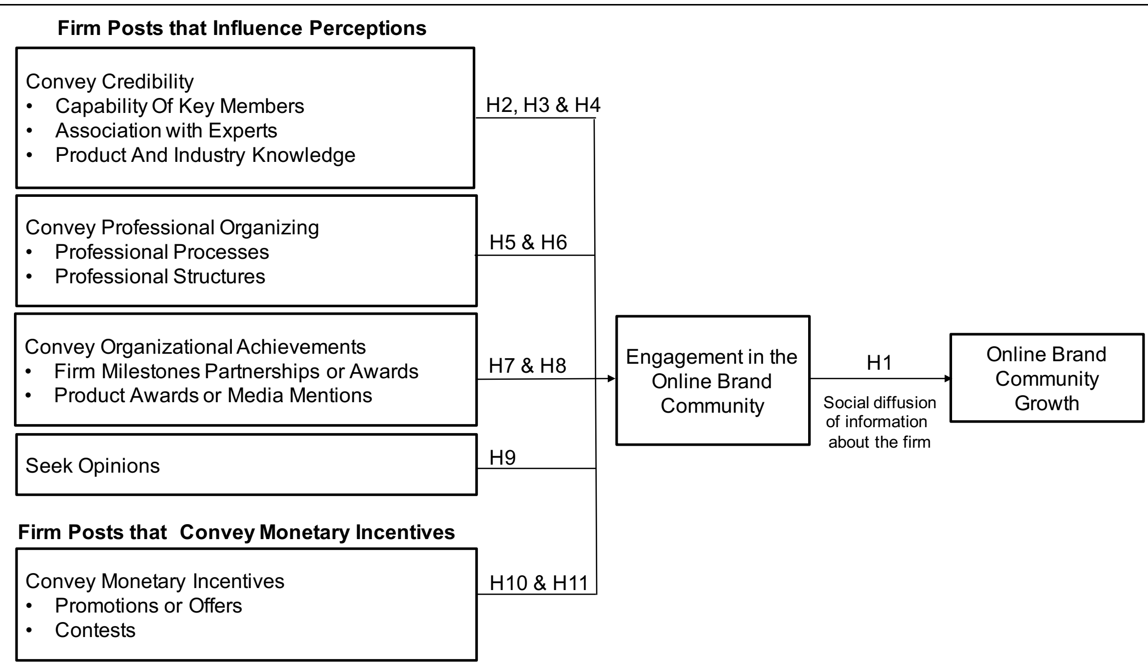

Figure 1. Summary of Hypotheses

2.1. Online Brand Community Growth

A community is "a group of people who are similar in some way" (Collins 2018). A brand community is "a specialized, non-geographically bound community, based on a structured set of social relations among admirers of a brand" (Muniz Jr. and O'Guinn 2001). Finally, an online brand community is "a specialized, non-geographically bound community, based upon social relationships among admirers of a brand in cyberspace" (Jang et al. 2008). Prior research (Habibi et al. 2014) indicates that social media based online communities (for e.g., groups in social media that are centered on a certain brand and which are initiated by brand managers) manifest all the characteristics identified in traditional brand communities.

Firms use posts composed of text, photos, icons, and links to articles, blog posts or videos, to distribute information through their online brand communities. Prior work suggests that posts are artifacts around which people engage (Ellison and Boyd 2013). Artifacts provide people with points of reference (Rafaeli and Vilnai-Yavetz 2004a), and specifically, posts are a point of reference around which community members can engage by liking the post, commenting on it or sharing it (Cvijikj and Michahelles 2013; Lee et al. 2017).

When individuals engage with a firm's posts, others who are tied to the individual, but not necessarily to the firm, can see such interactions through notifications by the social networking platform (Kane et al. 2014). Thus, users of the platform are exposed to new information and stories about the firm, and may adjust their thoughts and actions in light of such information from socially relevant peers (Miller et al. 2009). Posting and liking occur frequently on social networking platforms. The average number of connections an adult has on Facebook is 338

(Smith 2014), and users reach 61% of their ties each month through their posts (Bernstein et al. 2013). Moreover, Facebook users like over 4 million posts every minute (Carey-Simos 2015).

The exposure to information through ties as a result of engagement in response to firm posts, serves as a trigger for activities on the site (Ellison and Boyd 2013), which may include joining the firm's community on the social networking site. Prior research supports the notion that platform provided notification of the actions of ties in a social networking community influences user behavior. For example, Bakshy et al. (2012) through their experiment find that automatic notification of an action like the sharing of a post by a friend influences similar behavior for both strong and weak ties, with the latter being the primary driver for the diffusion of novel information and influence. In addition, Aral and Walker's (2011) experiment indicates that automatic notifications to ties when people engage on a social networking site generate a 246% increase in peer influence.

Thus, when individuals engage with firm posts it may influence users of the social networking platform to join the firm's online brand community because they are exposed to new information about the firm through their peers. We propose:

H1: The greater the user engagement (i.e. "likes") in a firm's online brand community the more its online brand community (i.e. fan base) grows.

2.2. Online Brand Community Engagement – Influencing Perceptions

Posts in an online brand community form an initial artifact around which individuals may engage

(Ellison and Boyd 2013). Moreover, prior literature (e.g., Gallaugher and Ransbotham 2010; Miller et al. 2009) suggests that firms use posts as a means of influencing the perceptions of others. One way in which firms can do this is through the use of sensegiving communications. In prior research, sensegiving has been defined as the process of influencing the meaning construction of others (Gioia and Chittipeddi 1991). For example, firms can leverage their online brand communities to share stories such as those that convey firm credibility or firm achievements. Such stories can convey symbolic meaning (Zott and Huy 2007), which is evaluated according to subjective criteria like values, feelings, emotions and preferences of the observer (Rafaeli and Vilnai-Yavetz 2004), and thus may elicit engagement, that is, 'likes' in response to firm posts from social media users who see such content. We draw from Zott and Huy's (2007) categorization of symbolic actions used by firms to influence audiences including customers, employees, associates and investors, and propose that there are four broad categories of posts that firms may use to influence others' perceptions in the online domain as they seek to spur user engagement and grow their online communities. These are posts that convey credibility, convey professional organizing, convey organizational achievements, and seek opinions. We describe each of these in the next section.

2.2.1. Posts Conveying Firm Credibility

Credibility refers to "the quality of being believable or trusted" (Collins 2018). Past research has shown that firms have depicted themselves as being credible by: displaying the capability and commitment of their leadership (Zott and Huy 2007), using associations with prominent others (Higgins and Gulati 2003), and obtaining endorsements (Starr and MacMillan 1990). We define posts that convey firm credibility as those that fall into one of three subtypes:

Posts Conveying the Capability of the Founder or Key Employees: Specifically, these are posts that portray the founder(s) or key team members as capable company builders. Such capabilities can be displayed through posts that describe the founder(s) or key employees winning awards, being speakers at an event or conference, being interviewed by the media, or being mentioned in the media. From a symbolic perspective, such posts indicate future competence and capability of team members.

Posts Conveying Association with Industry Experts: These are posts showing content such as an interview with an external expert; product picks, tips or suggestions by an external expert; or a live Q&A session with an external expert. Symbolically these posts convey that established people recognize and want to associate with the firm.

Posts Conveying Product or Industry Knowledge: These are posts that provide educational or value added informative content related to the firm's products or the industry to which the firm belongs. Symbolically such posts portray the firm as being knowledgeable and having expertise (or having access to knowledge and expertise) about the products the firm sells and the industry in which it operates.

We expect that the use of firm posts that convey credibility are likely to be associated with engagement in the firm's online brand community. Thus, we propose:

H2: Firm post content conveying the capability of the founder or key employees is positively *associated with engagement in its online brand community.*

H3: Firm post content conveying association with industry experts is positively associated with engagement in its online brand community.

H4: *Firm post content conveying product and industry knowledge is positively associated with engagement in its online brand community.*

2.2.2 Posts Conveying Professional Organizing

Stable organizational structures are seen as a requirement for reliability and accountability, and organizations can attain such structures through institutionalization as well as by adopting standardized routines or processes (Hannan and Freeman 1984). Firms, therefore, convey the quality of their organizing efforts by showing that the firm has adopted professional structures and processes (Zott and Huy 2007).

We define posts that convey professional organizing as those that fall into one of the following two subtypes. The first subtype is posts that communicate the use of professional processes or procedures. For example, a post that includes a photograph of products being quality checked before they are shipped to a customer indicates that the firm employs professional processes. The second subtype constitutes posts that convey the existence of professional structures. An example of a post that conveys professional structures is an interview with an employee in which the employee describes her role in the firm. Symbolically, such posts depict the professional nature of the firm's structures and procedures, thus portraying the firm as being both professionally run and experienced. The use of firm posts that convey professional organizing are likely to be associated with engagement in the firm's online brand community. Hence, we propose:

H5: Firm post content conveying the use of professional processes is positively associated with engagement in its online brand community.
H6: Firm post content conveying the presence of professional structures is positively associated with engagement in its online brand community.

2.2.3 Posts Conveying Organizational Achievements

In previous studies, firms have been shown to use organizational achievements as a means of influencing the perceptions of others. For example, Rao (1994), in his study of the auto industry, discusses how winning certification contests helps to shape positive firm reputation. Zott and Huy (2007) find that firms use prototypes, awards, firm age and number of employees to signal organizational achievements, and thus influence others' perceptions about the firm.

We define posts that convey organizational achievement as posts that either i) convey milestones, partnerships and awards won by the firm, or ii) convey that the firm sells award winning products or those that are featured in the media. An example of a post that describes a milestone is a photograph of the firm's employees celebrating the firm's first anniversary. Symbolically, posts that convey information about milestones indicate that the firm has persisted over time or grown. Posts that convey information about partnerships or that the firm sells award-winning products symbolically indicate that the firm is recognized by established entities its environment.

We expect that the use of firm posts that convey organizational achievements are likely to be associated with engagement in the firm's online brand community. We therefore propose:

H7: Firm post content conveying milestones, partnerships and awards won by the firm is positively associated with engagement in its online brand community.
H8: Firm post content conveying that the firm sells award winning products or those that are featured in the media is positively associated with engagement in its online brand community.

2.2.4 Posts Seeking Opinions

Petkova et al. (2013) suggest that interactive events such as conferences that bring together actors from diverse professional, organizational, and geographical backgrounds provide a venue for sensegiving actions because such events allow firms to discuss their ideas and receive feedback.

Online communities offer a context for individuals to continuously express and access others' opinions (Miller et al. 2009). Such communities facilitate conducting a public discussion in which a firm can seek opinions. Individuals can view each other's opinions, interact with one another, and the firm can participate in the discussion. We define posts that seek others' thoughts or opinions as posts that show that the firm wants to engage in a conversation with their audience by asking questions. An example of such a post is sharing pictures of two different prints for a product and asking which one the reader prefers. On a social networking site, such questions are not rhetorical because readers are able to respond to the questions and are also able to view and react to the responses by others. The symbolic dimension of such posts is that the firm wants to connect with their audience and that the firm cares about what its audience thinks. These arguments suggest that the use of firm posts that seek opinions are likely to be associated with engagement in the firm's online brand community. Thus, we propose:

H9: Firm post content seeking opinions is positively associated with engagement in its online brand community.

2.3 Online Brand Community Engagement – Monetary Incentives

Monetary incentives have been identified in prior literature as a motivator for engagement in brand communities (Jarvenpaa and Tuunainen 2013; Muntinga et al. 2011). Such incentives can take the form of contests, or promotions and offers (e.g., Gallaugher and Ransbotham 2013). Jarvenpaa and Tuunainen (2013) find through their analyses of Finnair's brand community on Facebook that contests based on trivia or fun facts relating to the airline's destinations were associated with increased engagement in the online community. Additionally, an industry whitepaper (Cvijikj and Michahelles 2013) finds that contests in the form of sweepstakes are associated with increased engagement. Apart from contests, prior literature also recommends that firms incorporate promotions and offers into their social media communication strategy. Mandviwalla and Watson (2014), for example, observe that discounts and redemptions offered via social media create economic capital for stakeholders. Gallaugher and Ransbotham (2013) note that Starbucks, rated the top firm for social media engagement among 100 leading brands, broadcasts promotions (e.g., free pastry day) to its online brand community. We thus propose:

H10: Firm post content conveying firm related promotions or offers is positively associated with engagement in its online brand community.

H11: Firm post content announcing contests is positively associated with engagement in its online brand community.

3. EMPIRICAL SETTING, MEASURES AND METHODS

3.1. Empirical Setting

We tested our hypotheses with a sample of firms in the flash sales segment of the retail industry. Flash sales businesses operate by purchasing excess or out-of-season inventory at steep discounts from various brands and subsequently selling them online at deep discounts for a limited time (typically between a day and a week). We selected this sector because it is one of the fastest growing e-commerce segments (Cocotas 2012), and thus is an important sector for us to study. Moreover, the rapid growth of the segment since 2007 has given us a sample of several start-up firms founded around the same time and operating under the same business model. By using a sample that consists of new firms that are unlikely to have developed the history, brand, and status that is associated with well-established firms, we are able to alleviate the concern that a firm's posting strategy and outcomes examined, that is, engagement and online brand community size may all be affected by the firm's existing status or brand effects.

We drew our sample from two directories of flash sales firms, Lokango.com and FashionInvites.com, as well as the business press. The latter source was included to compensate for the time lag for some newly created firms to appear in the directories. The selected firms covered all the major categories of flash sales businesses, including apparel, household goods, travel, and life style items. The criteria for inclusion in our sample are that the firm was founded in 2010,[1] was in business at the time of data collection, has operations or an office location in the US or Canada, sells products in the US, and has a Facebook brand page.[2] We focused on North America to exclude confounding influences of the varying level of adoption and use of social media in different countries. We identified 23 firms that met our criteria, and obtained post content and likes in response to the post content from the Facebook profile pages of these firms.[3] We also collected the firms' weekly online community size (or fan count), from the date each firm joined Facebook until July 1, 2012.

We analyzed our data in two phases that correspond to the sequence of our hypotheses. In the first phase, we examined the association between engagement and community size. In the second phase, we examined the association between the different types of posts identified by our hypotheses and engagement. Our final sample consists of 15 firms and a total of 9,470 posts (we explain firm exclusions in 'Section 3.3. Second Phase - Measures and Methods.' The median number of posts for the firm in the final sample is 631 (mean 625). The median number of weeks of data for the firm in the final sample is 63 (mean 55). Only two of these firms had less than 50 weeks of data.

3.2. First Phase – Measures and Methods

3.2.1. Dependent Variable:

[1] In the case of firms with many subsidiaries, we include in the sample firms whose flash sales subsidiary was founded in 2010.
[2] For firms that operate both a regular retail business and a flash sales business, we included only those firms that had a separate Facebook profile page for the flash sales part of their business.
[3] The sample comprised of 12,689 posts that belonged to the 23 firms.

Our dependent variable is online community size. Community size is an indicator of online community growth (Butler 2001). We obtained the total Facebook community size per week for each firm in our sample from Wildfire.com, which was acquired by Google because of its preferred access to Facebook data (Baer 2013). This weekly data spans the period of our study.

3.2.2. Independent Variable:

Our main independent variable, that represents engagement, is the sum of the likes received each week in response to firm posts. There are three possible forms of engagement in response to firm posts on Facebook, namely likes, comments and shares (Cvijikj and Michahelles 2013; Lee et al. 2017). In our sample, these three forms of engagement are highly correlated.[4] Thus, including all three engagement measures in our model is likely to impact our results due to multicollinearity. We focused on likes instead of comments or shares because of the high frequency of likes. The average number of likes per post in our sample is 70. In contrast, the average number of comments and shares is 11 and 5, respectively. The total likes, comments and shares in response to firm posts, in our sample are 658,363, 99,495 and 46,025, respectively. Likes thus create a much larger number of digital impressions in individual's newsfeeds, through platform generated notifications, than shares or comments.

3.2.3. Model Specification:

---

[4] The correlation matrix can be found in Table A2 of Appendix 1. If all three measures of engagement are included in our estimation model then collinearity among the independent variables is likely be an issue, since the variance inflation factors (VIF) associated with *Likes* (Table A3, Appendix 1) was found to be equal to 5, which is the accepted threshold (VIF>= 5) that suggests multicollinearity (e.g., Aral and Walker 2014).

For the first phase of our analyses, we created a panel data set of firms where the time variable is the calendar week, and estimated the following model using the Arellano–Bond difference GMM estimator.

$$\begin{aligned}&\mathit{Log(Community\ Size)}_{jt} = \beta_1 \mathit{Log(Community\ Size)}_{j(t-1)} + \beta_2 \mathit{Log(Community\ Size)}_{j(t-2)} +\\ &\beta_3 \mathit{Log(Likes)}_{jt} + + \beta_4 \mathit{Log(UGC\ Count)}_{jt} + \beta_5 \mathit{Log(Firm\ Post\ Count)}_{jt} + \beta_6 \mathit{Google\ Trends}_{jt} +\\ &\beta_7 \mathit{Acquisition\ Merger}_{jt} + \beta_8 \mathit{New\ Product\ Line\ Added}_{jt} + \beta_9 \mathit{Financing\ Secured}_{jt} + \beta_{10} \mathit{Award}_{jt} +\\ &\beta_{11} \mathit{Controls}_{jt} + \varepsilon_{jt}\end{aligned}$$

We used the Arellano–Bond estimator because it is designed for situations such as ours, where the dependent variable is dynamic, that is, it is dependent on its own past values (Roodman 2009). In the equation above, $\mathit{Community\ Size}_{jt}$ is the size of the online brand community for firm j at the end of week t, and $\mathit{Community\ Size}_{j(t-1)}$ and $\mathit{Community\ Size}_{j(t-2)}$ are its lagged value by one and two weeks, respectively. The equation also includes the independent variable *Likes* corresponding to hypothesis 1, as well as the control variables[5] *Log*(*UGC Count), Log(Firm Post Count), Google Trends, Acquisition Merger, New Product Line Added, Financing Secured*, *Award* and firm age dummies. These control variables are discussed in the next sub-section on alternative explanations.

To alleviate concerns about a high instrument count in the GMM estimations, we use a collapsed instrument set (e.g., Burtch et al. 2013) which is comprised of one instrument for each variable and lag distance, rather than one for each time period, variable, and lag distance. Such reduction

[5] $\mathit{Community\ Size}_{jt}$ for firm j is measured on day d at the end of week t. All the independent variables and controls are measured for days d-1 to d-7.

in instruments has been shown to produce more reliable results in scenarios where the instrument count is on the higher side (Mehrhoff 2009).

3.2.4. Alternate Explanations:

We have taken several steps to address alternative theoretical explanations that may trigger growth in the online community as well as alternative explanations that arise from the design of our empirical model.

i. Existing Online Community Size, Herding Effects and Autocorrelation: The size of our dependent variable, *Community Size*, at any point in time is a predictor of its future size (Butler 2001). Greater membership, at least to a point, implies greater benefits (for example information, influence, and social support), and it is these benefits that make it possible to attract and retain members (Butler 2001). Another argument that explains why current community size is a predictor of future community size is based on the herding phenomenon among online users (e.g., Duan et al. 2009) in which individuals converge to a uniform social behavior (Bikhchandani et al. 1998). Herding can occur because an individual, having observed the actions of others ahead of him, may "follow the behavior of the preceding individual without regard to his own information" (Bikhchandani et al.1992). Thus, the size of the existing online community conceivably impacts the extent of herding behavior, and therefore the number of new community members.

To control for these alternate explanations, we included the lagged dependent variables, $\textit{Community Size}_{j(t-1)}$ and $\textit{Community Size}_{j(t-2)}$, to our set of explanatory variables. The presence of the lagged dependent variables gives rise to autocorrelation. If OLS is used without correction when the errors co-vary, variances and standard errors for the OLS estimates of the coefficients may be biased upward or downward. This concern is addressed by using the Arellano–Bond (Arellano and Bond 1991) difference GMM estimator, which is designed for situations where the dependent variable is dynamic, that is, depends on its own past realizations (Roodman 2009). The Arellano–Bond estimator instruments the first-differenced lagged dependent variable with its past levels (Mileva 2007).

ii. User Generated Content: Online communities are typically co-created with others, and therefore consist of posts from individuals as well as those from the firm (Ellison and Boyd 2013). Previous literature indicates that user generated content (UGC), which refers to posts by individuals other than the firm, is positively correlated with economic outcomes (e.g., Duan et al. 2008).

UGC, like posts by the focal firm, can trigger platform provided notifications, which may be received by people who are tied to the individual that wrote the post, but not necessarily to the firm, and can thus contribute to growth in the online community. We therefore included the control *Log*(*User Generated Content)*, where *User Generated Content* is the count of the user generated posts for each firm in each week of our panel data set.

It is conceivable that individuals might be joining online communities in order to complain. If this is the case, then the sentiment of user generated posts is likely to be negative. To mitigate this concern, we conducted a sentiment analysis on user generated posts (see Appendix 2, Section 1) and found that the overall sentiment of user generated posts is positive, thus suggesting that individuals are not joining brands' online communities to complain.

iii. Firm Post Count: It is conceivable that a high frequency of firm postings might influence the frequency of engagement (even if the volume of engagement in response to each post is low), and this may be associated with growth in online brand community. We therefore included the control *log(Firm Post Count)*, where *Firm Post Count* is a count of the firm posts for each firm in each week of our panel data set.

iv. Omitted Stable Firm Characteristics and Macro Events Affecting all Firms of the Same Age: Omitted variables such as the sub-segment that the firm belongs to (for example, furniture or housewares) or macro events affecting all firms in the sample that are of a particular age (in terms of weeks since they joined Facebook), may impact one or more of the explanatory variables as well as the community size. The concern with omitted variables, both those that do not change over time for a particular firm, as well as those that change over time but stay constant across individual firms of a given age, is addressed by the use of panel data with firm fixed effects and firm age (that is, week since joining Facebook) dummies (Hsiao 2003).

v. Time Variant Events Specific to Particular Firms: It is conceivable that time variant changes or events that impact a firm, such as securing a round of venture capital funding or acquiring another company, may drive up both posting about such events as well as the size of the online community. The endogeneity problem due to such time varying omitted variables is addressed by adding four control variables, namely: *Funding Secured*, *Acquisition Merger*, *Award*, and *New Product Line Added*. We populated these variables by examining the content of all articles in the press (occurring during the sampling period), obtained through Factiva and LexisNexis for the firms in our sample. These are boolean variables that take the value one when one of these events is announced, and take the value zero otherwise. If the event is announced in advance of the event occurring (for example, an acquisition is announced one week before it is executed), the variable *Acquisition Merger* takes a value of one for the week when the event is announced as well as the week when the acquisition takes place. The controls *Funding Secured*, *Acquisition Merger*, *Award*, and *New Product Line Added,* and *Google Trends* help us account for time-varying individual firm characteristics (Chen et al. 2015).

Furthermore, we used the Google trends search volume (*Google Trends*) of each firm to control for interest and the attention firms might receive (Geva et al. 2017) outside Facebook. For example, if an award or partnership is announced in the news, the Google trends value of that firm that week is likely to reflect this event.

vi. Existing Status: It is conceivable that unmeasured existing status of a firm could drive growth in the online community, rather than the engagement level in the online brand community (in response to firm posts). We address this concern, at least in part, by employing a sample that consists of brand new firms that we have tracked since (or close to) the establishment of their Facebook site. As a result, these firms have not yet accumulated the history associated with well-established firms. Moreover, we capture existing status, at least in part, by the firm fixed effect and the use of the lagged dependent variable, *Community Size,* in our model.

vii. Forced Like: A forced "like" on Facebook is a situation in which some brands force visitors to become a fan or online community member in order to access content on the brand's Facebook profile page (Digital Marketing Glossary 2012). This is an alternate explanation that could potentially drive up community size. However, none of the firms in our sample had implemented the forced like feature during our data collection period.

viii. Paid Promotion of Posts: At the end of May 2012 Facebook launched a feature allowing users to promote posts (Gray 2012), so that it could potentially reach a greater percentage of their community on Facebook. Our data includes post until July 1, 2012. Thus, this feature could potentially affect the last five weeks of our data, if it was used by the firms in our sample. Given that this was a brand-new feature, its adoption may not be widespread in the first month after launch. Nevertheless, we test the robustness of our findings by re-estimating the model with a dataset that is limited to posts before the launch of this feature, that is, by excluding posts the last five weeks of our dataset.

3.3. Second Phase – Measures and Methods

In the second phase, we implement post level analyses to identify the association between the different types of posts that we hypothesize about, and engagement, that is, likes in response to these posts. A key distinguishing feature of our analyses is that we use LDA to identify topics for firm posts, that we may not have theorized about, and control for such alternate untheorized pathways to engagement. We also use PSM combined with OLS to try to address endogeneity concerns due to confounding variables that might affect both our dependent variable and explanatory variables.

3.3.1. Dependent Variable:

The dependent variable is the number of likes that each post receives. The sum of likes received each week, in response to posts by firms, is the independent variable for the first phase.

3.3.2. Independent Variables:

Based on the literature review and our theory we identified 13 dimensions (listed in Table 1) associated with our posts that reflect credibility, professional organizing, organizational achievements, seek opinions or offer monetary incentives. The first 11 of these dimensions are related to content that might influence perceptions (corresponding to hypotheses H2-H9). The last two dimensions correspond to content that includes monetary incentives (corresponding to hypotheses H10 and H11). Following Zaheer and Soda (2009), we invited a panel of three experts[6] to validate our categorization of posts that influence perceptions, using the Q-Sorting technique (Segars and Grover 1998). The Q-sorting technique is recommended when new scales are being developed (Segars and Grover 1998), as is the case with our study. The list of 11 dimensions was randomly ordered and provided to a panel of experts, who were asked to sort the dimensions into the four types of posts that might influence perceptions, namely those that reflect credibility, professional organizing, organizational achievements, and seek opinions. The instructions included options for panel members to say that a dimension did not belong to any of the types, or to say that a dimension belonged to multiple types. Based on the experts' comments we reworded the dimensions related to two categories. We then conducted another round of

[6] The experts included three tenured business school faculty, two of whom are experts in the area of social media, and the third has employed manual text analysis of content in their own published research.

surveys to validate our categorization; the percentage of correct classification and agreement was 100% for all three experts. Our independent variables were constructed by categorizing each Facebook post in our sample along these 11 dimensions. These dimensions and their mapping to the four post types and corresponding sub-types (from our hypotheses) are described in Table 1. Dimensions 1, 7, 8 and 9 in Table 1 are derived directly from Zott and Huy (2007) while the rest of the dimensions associated with influencing perceptions are new.

The Facebook application programming interface (API) was used to collect the text of all past posts for the firms in our sample, up to July 1, 2012. For some firms, we were not able to gather posts all the way back to the date the firm joined Facebook, possibly because of restricted access settings by some firms, or limits set by Facebook on historic data. This limitation combined with the fact that different firms joined Facebook on different dates resulted in an unbalanced panel of posts by each firm, that is each firm in the sample does not have posts for the same number of weeks.

Coders were asked to categorize each of the posts in our sample. This involved determining whether the post belonged to any of the 13 dimensions corresponding to our hypotheses. A web based questionnaire was given to the coders to perform the categorization process. Measurement error due to context effects like grouping of items (Kline et al. 2000) and item priming effects (Salancik 1984) was avoided by randomizing the order of questions (Tourangeau and Rasinski 1988) that categorize each post type. The questionnaire was peer reviewed to ensure principles of good item writing such as avoiding double-barreled questions, jargon, leading items, and negatively worded items. Further, the questionnaire was pre-tested via a "think aloud" (Sudman et al.1996; Qinag et al. 2012) and via a pilot test for 100 posts. We modified the questionnaire based on feedback from the think aloud and the pilot tests. While we use a questionnaire for our independent variables, we obtained our dependent variable from a different source, thus avoiding common source bias (King et al. 2007).

To allow for a granular and nuanced analysis of posts we opted for manual coding of posts and used Amazon's Mechanical Turk (AMT) to find our coders. These coders not only analyzed the text of the post, but went to Facebook to find the actual post (based on identifying information) and coded the post after reading articles linked to the post, watching videos embedded in the post, and scrolling through photos associated with the post. Since workers on AMT are significantly more diverse than workers from typical American college samples (Buhrmester et al. 2011) the probability of measurement error due to response biases such as social desirability, and acquiescence (Bagozzi and Yi 1991) is reduced by using AMT workers. In addition, a study by Buhrmester et al. (2011) indicates that the data obtained from workers on Mechanical Turk are "at least as reliable as those obtained via traditional methods" and that AMT can be used to obtain, "high-quality data inexpensively and rapidly." Further, workers on AMT receive negative ratings that are publicly displayed as part of the worker's profile if their work does not meet the expectations or quality standards of the job requestor.

To ensure consistency of understanding of the questions, terminology and task, coders were trained via videos that we posted online. Additionally, to filter out coders who had a poor understanding of the questions or the task we "pre-qualified" each coder by asking them to code a set of 15 test posts. The degree of agreement between coders was measured using Cohen's kappa, where the K value is interpreted as the degree of agreement between coders after taking into account probability (Cohen 1960). Literature on using kappa suggests that a coefficient of .61 indicates reasonably good overall agreement (Kvalseth 1989), so a coder was qualified if his or her categorization of the test posts resulted in an overall kappa coefficient of greater than .61.

A total of 15 coders were qualified using this metric. Qualified coders were given online access to categorize posts for this study and we were available by email to answer any questions. To ensure that inter coder reliability was maintained, we periodically selected random posts for each firm and checked the reliability of the coding by asking another coder to code the randomly selected posts. This scalable, manual process applicable to high volume content analysis that we developed for qualifying and managing workers, and for evaluating their output, is illustrated in Figure 2.

For reliability testing, 370 randomly selected posts, approximately 3% of our original sample, were examined by a second coder. The overall kappa coefficient for the 370 posts was 0.79. We also examined the kappa coefficient for inter-coder reliability for each firm. Three firms had a kappa coefficient of <0.6 indicating that the coders did not have reasonably good overall agreement. These three firms were dropped from our sample. An additional five firms were also dropped from our sample. Of these, two firms were dropped because post content was not available, possibly due to restricted access by the firms, and two others were dropped because significant post content was unavailable (because the firm deleted the content) after the firm had initially posted something. One firm was dropped because its product line was focused on Indian fashion and home décor, and thus the post content targeted a very narrow audience.

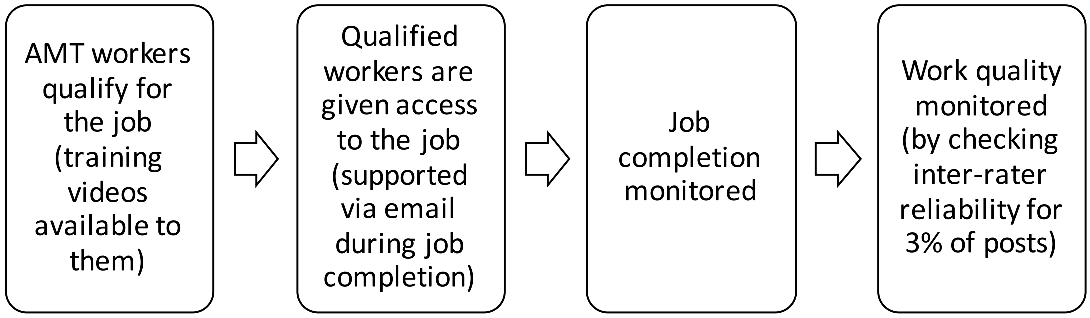

Figure 2. Process for Qualifying Workers and Coding Posts

Table 1			
Dimen-sion #	**Hypothesis #**	**Variable**	**Survey Question(s) / Description**
		Influencing Perceptions	
Credibility			
1	H2	*Capability of Key Members*	Does the post show an achievement by the founder(s) or one or more of the employees, such as the founder/employee winning an award, being a speaker at an event/conference, being interviewed by or mentioned in the media?
2	H3	*Association with Experts*	Does the post show either i) an interview with, ii) opinion/thoughts of, iii) suggestions/picks by, iii) invitation to an online (live) chat with an industry expert who is external to the firm?
3	H4 (operationa-lization 1)	*Product and Industry Knowledge: Tips or Suggestions*	Does the post provide either: i) commentary (tips/suggestions) on how to best use a particular product or product line that is sold by the firm, or ii) broad educational information such as tips, advice, or howtos?
4	H4 (operationa-lization 2)	*Product and Industry Knowledge: Industry Information*	Does the post provide any of the following that is specific to the industry in which the firm operates: i) feature article, ii) anecdote, or iii) news (including an opinion about the news)?
5	H4 (operationa-lization 3)	*Product and Industry Knowledge: Design Origin Information*	Does the post provide commentary (description or opinion) on what is definitional/characteristic about the designer, manufacturer, stylist, curator, or place of origin of the collection/product (that is sold by the firm)?
6	H4 (operationa-lization 4)	*Product and Industry Knowledge: Collection or Picks*	Does the post show a special 'collection' or 'picks'?
Professional Organizing			
7	H5	*Professional Process*	Does the post show a routine task being done by the firm or a team member?
8	H6	*Professional Structure*	Does the post profile information about one or more team members?
Organizational Achievements			

9	H7	*Organizational Achievements - Firm Milestone Partnership or Award*	Does the post show an achievement by the firm, e.g., receiving an award, establishing a partnership, or achieving a milestone?
10	H8	*Organizational Achievements - Product Award or Media Mention*	Does the post show a product that is sold by the firm being featured in the media, or indicate that the product is award winning?
Opinions			
11	H9	*Opinions*	Does the post ask a question or in some other way solicit the reader's thoughts or opinion?
Monetary Incentives			
12	H10	*Promotion or Offer*	Is the post about a sale event/offer related to the firm?
13	H11	*Contest*	Is the post an invitation or announcement to participate in a: contest, sweepstake, or give away? Or does it ask the user to take some action (other than an action that involves making a purchase) that may result in a reward?

3.3.3. Model Specification:

In the baseline specification, we estimate the following regression model:

$$\log(Like_i) = \beta_0 + \beta_1 Capability\ Of\ Key\ Members_i + \beta_2 Association\ With\ Experts_i + \beta_3 Tips\ or\ Suggestions_i + \beta_4 Industry\ Information_i + \beta_5 Design\ Origin\ Information_i + \beta_6 Collection\ or\ Picks_i + \beta_7 Professional\ Process_i + \beta_8 Professional\ Structure_i + \beta_9 Firm\ Milestone\ Partnership\ or\ Award_i + \beta_{10} Product\ Award\ or\ Media\ Mention_i + \beta_{11} Opinion_i + \beta_{12} Promotion\ or\ Offer_i + \beta_{13} Contest_i + \beta_{14} Controls + \varepsilon_i,$$

In the equation, the dependent variable $Like_i$ is the number of likes in response to firm post *i*. The independent dummy variables $Capability\ Of\ Key\ Members_i$ - $Contest_i$ correspond to dimensions 1 – 13 in Table 1 (which correspond to H2-H11). If post *i* conveyed a particular dimension, the value was set to 1, otherwise it was set to 0. $Controls$ include (i) firm dummies, (ii) posting time dummies (year, month and day dummies), (iii) two article related dummies –

one indicating whether the post included a link to an article by the firm, the other indicating whether the post included a link to an article by someone external to the firm, iv) two video related dummies – one indicating whether a post included a link to video content that was created by the firm, the other indicating whether the post included a link to video content by someone external to the firm, (v) a dummy variable indicating whether a post included an image, (vi) Google trends search volume for the firm in the week of post *i* (*Google Trends*), and (vii) firm age, that is, week since joining Facebook (*Firm Age*). These control variables are discussed in the next section.

3.3.4. Alternate Explanations:

As with the first phase of our analyses, we have taken several steps to address alternative explanations that may trigger engagement as well as alternative explanations that arise from the design of our empirical model.

i. Using Controls to Address Omitted Variables: The following control variables follow the explanation in the first phase analyses: i) omitted stable firm characteristics are addressed by the use of panel data with firm fixed effects ii) time variant events specific to particular firms are addressed by using the Google trends search volume (*Google Trends*) of each firm as a control.

Additionally, macro events affecting all firms in the sample during a particular time period (for example, increases in the use of social media more generally over time, or peoples' use of social media over the holiday season) during a particular time period are addressed through the use of time dummies. Finally, the *Firm Age* control helps address the issue that arises from the data constraint that some firms in our sample do not have data from when the firm joined Facebook.

In our context, the truncation problem may bias our estimation if firms post different types of posts in different stages of their life cycles. Controlling for *Firm Age* should alleviate the concern that there are some firms in our sample that do not have data from when the firm joined Facebook.

ii. Latent Post Topics: It is conceivable that our theory does not identify certain types of posts used by firms, which might be associated with engagement. To help us identify such latent topics that firms may be posting about, we used text mining and let the data speak for themselves. Thus, we were able to more fully exploit the text information of firms' posts.

Following Blei et al. (2003), we used LDA to discover latent topics (e.g., Singh et al. 2014), and further controlled for them in our regression. The LDA is a Bayesian statistical and information retrieval technique. The input to the LDA topic model is a set of documents, which in this study is the text of firms' posts. The output of LDA is a predefined number (*K*) of topics from these posts. In addition, LDA provides a posterior topic distribution over all *K* topics for each post. In other words, for each post *i*, the LDA model outputs a *K*-vector, $< T_{i,1}, T_{i,2}, \cdots T_{i,K} >$, where $T_{i,k}$ represents the weight of topic *k* associated with post *i*. Also, for each post *i* and topic *k*, $0 \leq T_{i,k} \leq 1$ and $\sum_{k=1}^{K} T_{i,k} = 1$. The larger the weight of a given topic, the more likely it is that the post is associated with that topic.

We ran the LDA model with $K = 5$, 10, and 15 topics, and found that 5 is the optimal number of latent topics in terms of the posterior log-marginal likelihood (Nam et al. 2017). A list of the most frequent words associated with each latent topic ($K = 5$) is shown in column 2 of Table 2.

Looking at the keywords for each topic, we observe that the first four topics reflect keywords pertaining to products sold by the firms in our sample. The last topic included keywords related to monetary incentives such as contests and offers. We labeled each topic based on the keywords associated with each topic. Column 3 of Table 2 shows these topic labels.

We illustrate the output of the LDA model using a few examples. For each post i = 1, 2, 3 in Figure 3, the LDA model outputs a vector, $< T_{i,1}, T_{i,2}, T_{i,3}, T_{i,4}, T_{i,5} >$, where $T_{i,k}$ represents the weight of topic k associated with post j. Post 1 is associated with topics 1 (Home Decor & Furniture) and 5 (Offer & Contest), and the weight on topic 5 (Offer & Contest) is much greater than that on Topic 1 (Home Decor & Furniture). This suggests that post 1 is more related to topic 5 (Offer & Contest). Post 2 is associated with only topic 4 (Clothing), and its weight on this topic is 1. Post 3 is associated with topics 1 (Home Decor & Furniture), 4 (Clothing), and 5 (Offer & Contest), and the weight on topic 1 (Home Decor & Furniture) is much greater than the weight on topics 4 (Clothing) and 5 (Offer & Contest).

We included topics 1, 2, 3 and 4 as control variables, and re-estimated our regression model. Topic 5 was our reference group. Including all five topics $< T_{i,1}, T_{i,2}, T_{i,3}, T_{i,4}, T_{i,5} >$ in our regression model would lead to perfect collinearity since $\sum_{k=1}^{5} T_{i,k} = 1$.

Table 2. Keywords Associated with Each Topic

Topic #	Keywords	Label
1	Flash, art, style, modern, furniture, pieces, beautiful, space, designs, designer, fun, products, unique	Home Decor & Furniture
2	Dinnerware, flatware, spice, kitchen, rack, steel, jar, stainless, enjoy, serving, cutlery, gourmet	Housewares
3	Hotel, travel, beach, resort, spa, pool, vacationist, stay, luxury, best, city, 4-star, free, book, secret	Travel & Leisure
4	Think, happy, weekend, shop, open, wear, summer, dress, giveaway, daily, accessories, shoes	Clothing
5	Shipping, enter, contest, post, offer, week, chance, winner, friends, time, credit, know, tomorrow, free	Offer & Contest

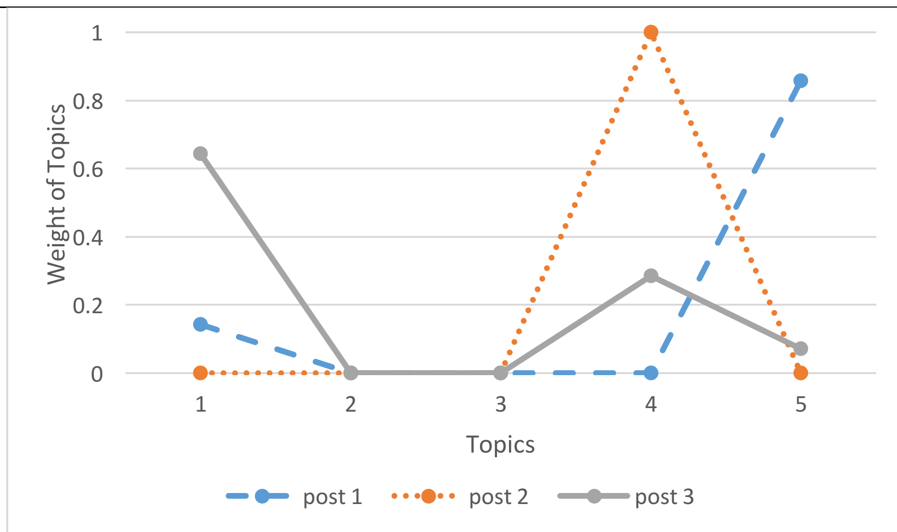

Figure 3. Examples of Weight of Topics for Different Posts

iii. Paid Promotion of Posts: In late May 2012, Facebook launched a feature allowing users to promote posts (Gray 2012). Similar to the first phase analyses, we test the robustness of our findings by re-estimating the model with a dataset that is limited to posts before the launch of this feature, that is, by excluding 634 posts that correspond to the last five weeks of our dataset.

iv. Unobserved Variables - Robustness Check Using Propensity Score Matching: Confounding variables that might affect the dependent variable, engagement, as well as the explanatory variables, could bias our results. Beyond the controls and robustness tests we describe above, to further address such endogeneity concerns, we also tested the robustness of our results using PSM combined with a regression method (OLS). Combining PSM with regression models has been increasingly adopted in information systems literature (e.g., Goh et al. 2013; Li 2016).

Following Goh et al. (2013) and Li (2016), we constructed proper control and treatment groups of matched posts by using PSM and ensured that these groups were comparable on observable characteristics. Treated posts included posts that conveyed at least one of the 13 dimensions listed in Table 1, which correspond to H2-H11. Control posts included posts that did not convey these dimensions. We used the nearest neighbor matching algorithm to match[7] each treated post to the most similar post in the control group (closest propensity score) in terms of the following post characteristics: (i) firm dummies, (ii) posting time dummies (year, month and day dummies), (iii) dummies indicating whether a post included a link to an article by the firm, or by someone external to the firm, iv) dummies indicating whether a post included a link to video content that was created by the firm, or by someone external to the firm, (v) a dummy variable indicating whether a post included an image, (vi) Google trends search volume for the firm in the week of post *j* (*Google Trends*), and (vii) firm age, that is, week since joining Facebook (*Firm Age*). To obtain the predicted propensity scores we ran a logistic regression with the

[7] Before matching, we randomized the order of posts to make sure that their order did not affect the subsequent matching.

aforementioned post characteristics. We re-estimated our post level regression model using the new sample created by the PSM procedure.

4. RESULTS – FIRST PHASE

4.1. Descriptive Statistics and Correlations:

Table 3 provides descriptive statistics for our variables, while Table 4 provides correlations for the variables. To assess possible multicollinearity among model covariates, we calculated variance inflation factors for the covariates. We did not find collinearity among the independent variables to be an issue since the variance inflation factors (VIFs) associated with each variable were found to be less than or equal to 1.1, which is well below the conventionally accepted threshold (VIF< 5), indicating that multicollinearity is not an issue in our model and does not significantly impact our findings (e.g., Aral and Walker 2014).

Table 3. Descriptive Statistics

	Obs	Mean	Std. Dev.	Min	Max
Community Size	829	41397.97	56109.58	279	252633
Likes	829	796.55	2548.95	0	36662
UGC Count	829	13.70	29.67	0	503
Firm Post Count	829	11.45	10.25	0	95
Google Trends	829	38.36	24.31	0	100
Acquisition Merger	829	0.01	0.09	0	1
New Product Line Added	829	0.00	0.07	0	1
Financing Secured	829	0.00	0.07	0	1
Award	829	0.00	0.07	0	1

Table 4. Correlation Matrix

	Community Size	Likes	UGC Count	Firm Post Count	Google Trends	Acquisition Merger	New Product Line Added	Financing Secured
Likes	0.66							
UGC Count	0.16	0.03						
Firm Post Count	-0.03	0.13	0.05					
Google Trends	0.27	0.26	0.04	0.09				
Acquisition Merger	0.17	0.07	0.10	-0.02	0.06			
New Product Line Added	0.07	0.02	0.15	-0.02	0.06	-0.05		

Financing Secured	-0.02	0.01	0.03	0.17	0.01	-0.03	-0.03	
Award	0.15	0.00	0.01	-0.11	0.01	-0.05	-0.04	-0.03

4.2 Results:

Table 5 presents the results of our empirical tests of the effect of engagement on online community size. We find that *Log(Likes)* is positively associated with online *Community Size*, which supports hypothesis 1. In addition, we find that the control variables *L.Log(Community Size)* which is the first lag of *Community Size*, and *Log(UGC Count)* are positively associated with online *Community Size*.

Table 5, Model 2 shows the results when we excluded posts after Facebook introduced the paid promotions feature. The re-estimation produced coefficients that were similar in terms of sign and statistical significance to the main model (Table 5, Model 1).

Table 5. Impact of Engagement on Online Community Size

	(1)	(2)
		Excluding the last 5 weeks of data
	Log(Community Size)	Log(Community Size)
L.Log(Community Size)	1.12***	1.12***
	(0.14)	(0.14)
L2.Log(Community Size	-0.19	-0.19
	(0.13)	(0.13)
Log(Likes)	0.01***	0.01***
	(0.002)	(0.003)
Log(UGC Count)	0.04***	0.04***
	(0.01)	(0.01)
Log(Firm Post Count)	0.01	0.004
	(0.01)	(0.01)
Google Trends	0.0001	0.0001
	(0.0001)	(0.0001)
Acquisition Merger	0.01	0.01*
	(0.01)	(0.01)
New Product Line Added	-0.01	-0.01
	(0.01)	(0.01)
Financing Secured	0.04	0.04

	(1)	(2)
	(0.04)	(0.04)
Award	0.01	0.01
	(0.01)	(0.02)
Firm Fixed Effects	Yes	Yes
Firm Age Dummies	Yes	Yes
Observations	784	729

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Model (1) Instruments for first differences equation
Standard: D.(Log(likes) Log(UGC) Log(FirmPosts) GoogleTrends AcquisitionMerger NewProductLine FinancingSecured Award weekNumber_1-weekNumber_114
GMM-type: (missing=0) L(1/92).(L.Log(Community Size) L2.Log(Community Size)) collapsed
Number of instruments = 210

4.3. Robustness:

The robustness test in Appendix 2 Section 2 shows that the dynamic model is correctly specified and that our model is based on instruments that are valid.

5. RESULTS SECOND PHASE

5.1. Descriptive Statistics and Correlations:

Table 6 provides descriptive statistics for our variables, while Table 7 provides correlations for the variables. To assess possible multicollinearity among model covariates, we calculated VIFs for the covariates. We did not find collinearity among the independent variables to be an issue since the VIFs associated with each variable were found to be less than or equal to 1.4, which is well below the conventionally accepted threshold (VIF< 5), indicating that multicollinearity is not an issue in our model and does not significantly impact our findings (e.g., Aral and Walker 2014).

Table 6. Descriptive Statistics for Post Level Analyses (Second Phase)					
	Obs	Mean	Std. Dev.	Min	Max
Capability of Key Members	9470	0.0029	0.053	0	1
Association with Experts	9470	0.017	0.13	0	1
Tips or Suggestions	9470	0.054	0.23	0	1
Industry Information	9470	0.034	0.18	0	1
Design Origin Information	9470	0.053	0.23	0	1
Collection or Picks	9470	0.070	0.25	0	1
Professional Process	9470	0.0049	0.070	0	1
Professional Structure	9470	0.0015	0.12	0	1
Firm Milestone Partnership or Award	9470	0.0058	0.076	0	1
Product Award or Media Mention	9470	0.0087	0.093	0	1
Opinion	9470	0.14	0.44	0	1
Promotion or Offer	9470	0.12	0.38	0	1
Contest	9470	0.061	0.24	0	1

Table 7. Correlation Matrix for Post Level Analyses (Second Phase)

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
1. Capability of Key Members	1											
2. Association with Experts	0.02											
3. Tips or Suggestions	-0.004	0.11										
4. Industry Information	0.09	0.16	0.08									
5. Design Origin Information	0.005	0.15	0.07	0.02								
6. Collection or Picks	-0.007	0.03	-0.02	-0.02	0.05							
7.Professional Process	0.08	0.01	-0.003	0.03	0.003	-0.007						
8. Professional Structure	0.11	0.22	0.26	0.15	0.15	0.004	0.09					
9. Firm Milestone Partnership or Award	0.13	0.03	-0.006	0.09	0.007	0.001	0.03	0.11				
10. Product Award or Media Mention	0.10	0.13	0.003	0.11	0.06	0.03	0.03	0.06	0.06			
11. Opinion	-0.009	0.02	0.03	0.04	-0.08	-0.01	-0.01	0.03	-0.03	0.01		
12. Promotion or Offer	-0.020	-0.01	-0.02	-0.04	0.24	-0.06	-0.02	0.004	-0.01	0.02	-0.17	
13. Contest	-0.005	-0.02	-0.02	-0.02	-0.05	-0.06	-0.02	-0.02	0.02	-0.01	0.01	-0.06

5.2. Results:

We present the results of the models that test hypotheses H2-H11 in Table 8. First, we present the results of the baseline model using observational data (Table 8, Model 1). Second, we report results of the model that uses the same observational data, but with controls included for the latent topics identified using LDA (Table 8, Model 2). Third, we report the results of the regression model using the sample created with PSM (Table 8, Model 3). In the estimations, we use robust standard errors (clustered at the firm level) to deal with cluster correlations and unknown heteroskedasticity in the error terms.

We adopt a conservative approach to verifying our hypotheses – we confirm a hypothesis only if the coefficients are positive and statistically significant across all three model specifications described above (Sala-i-Martin 1997). Thus, a hypothesis is confirmed only when the statistical relationship is robust. We find mostly consistent results across models 1, 2 and 3 in Table 8. The coefficients of *Tips or Suggestions, Industry Information, Design Origin Information, Collection or Picks, Firm Milestone Partnership or Award, Opinion,* and *Promotion or Offer* (corresponding to dimensions 3, 4, 5, 6, 9, 11 and 12 in Table 1) are positive and statistically significant in all three models. The first four of these variables are different operationalizations or approaches to conveying product and industry knowledge (which corresponds to H4). These results provide support for hypotheses H4, H7, H9 and H10. Effect size can be interpreted directly from the coefficients. For instance, based on Model 2 of Table 8, posts conveying *Design Origin Information* are associated with a 21% higher level of likes than those that do not convey *Design Origin Information*.

Among the independent variables that correspond to our hypotheses, the only one that does not exhibit a consistent statistical pattern across all three models is *Professional Process* (corresponding to H5). It is statistically significant in Model 1, but is not significant in Models 2 and 3 of Table 8. Since this suggests that the statistical relationship between posts conveying *Professional Process* and *log(Likes)* is not robust, H5 is not supported.

Table 8. Impact of Different Types of Firm Posts on Engagement (log(Likes))

	(1) OLS	(2) OLS (with LDA topics)	(3) PSM + OLS	(4) Excluding the last 5 weeks of data
Capability of Key Members	-0.68	-0.65	-0.52	-0.70
	[0.58]	[0.63]	[0.48]	[0.58]
Association with Experts	-0.11	0.0089	0.13	-0.059
	[0.11]	[0.11]	[0.12]	[0.11]
Tips or Suggestions	0.15**	0.14**	0.18***	0.11***
	[0.058]	[0.062]	[0.047]	[0.032]
Industry Information	0.32***	0.20***	0.25***	0.31***
	[0.082]	[0.024]	[0.061]	[0.083]
Design Origin Information	0.26***	0.21***	0.28***	0.24***
	[0.052]	[0.057]	[0.058]	[0.052]
Collection or Picks	0.71***	0.54***	0.67***	0.71***
	[0.047]	[0.037]	[0.10]	[0.048]
Professional Process	0.56***	0.24	0.37	0.57***
	[0.17]	[0.25]	[0.33]	[0.18]
Professional Structure	0.088	0.10	0.098	0.085
	[0.10]	[0.12]	[0.099]	[0.10]
Firm Milestone Partnership or Award	0.71***	0.79***	0.66***	0.69***
	[0.15]	[0.20]	[0.17]	[0.15]
Product Award or Media Mention	0.21*	-0.20	0.097	0.20
	[0.13]	[0.13]	[0.21]	[0.13]
Opinion	0.50***	0.25***	0.41***	0.48***
	[0.032]	[0.039]	[0.043]	[0.032]
Promotion or Offer	0.32***	0.30***	0.39***	0.28***
	[0.040]	[0.045]	[0.072]	[0.040]
Contest	0.10*	0.12*	0.14	0.071
	[0.060]	[0.066]	[0.17]	[0.061]
Firm Video	-0.58***	0.36	-0.014	-0.51**
	[0.19]	[0.35]	[0.38]	[0.18]
External Video	-0.57***	-0.72***	-0.75***	-0.52***
	[0.15]	[0.16]	[0.21]	[0.16]
Firm Article	-1.17***	-0.68***	-1.07***	-1.20***

	[0.045]	[0.055]	[0.093]	[0.046]
External Article	-0.58***	-0.37***	-0.59***	-0.63***
	[0.055]	[0.065]	[0.094]	[0.056]
Image	0.96***	0.73***	1.03***	0.91***
	[0.037]	[0.042]	[0.10]	[0.038]
Firm Age	0.030***	0.028***	0.025***	0.033***
	[0.0011]	[0.0014]	[0.0024]	[0.0012]
Google Trends	0.021***	0.016***	0.015***	0.019***
	[0.00067]	[0.00082]	[0.0016]	[0.00068]
Topic 1		0.54***		
		[0.061]		
Topic 2		-0.77***		
		[0.062]		
Topic 3		0.27***		
		[0.10]		
Topic 4		0.23***		
		[0.065]		
Firm dummies	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes
Constant	-2.77***	-2.56***	-2.27**	-3.02***
	[0.20]	[0.26]	[0.31]	[0.21]

Robust standard errors in brackets *** p<0.01, ** p<0.05, * p<0.1

With respect to effect size comparisons for posts that convey monetary incentives versus those that influence perceptions, we find (based on Table 8 Model 2) that the coefficients on post about *Firm Milestone Partnership or Award* (t = 224.25) and posts that convey product and industry knowledge through *Collection or Picks* (t = 392.91) are significantly higher than the coefficient on *Promotion or Offer* (p-values < 0.01). The coefficients on *Opinion* (t = -123.24) and the three types of posts that convey product and industry knowledge, namely: *Tips or Suggestions* (t = -240.06), *Industry Information* (t = -252.11), and *Design Origin Information* (t = -151.32) are significantly lower than the coefficient on *Promotion or Offer* (p-values < 0.01).

5.3. Robustness Tests:

Table 8, Model 4 shows the results based on observational data when we excluded posts after Facebook introduced the paid promotions feature. The re-estimation produced coefficients that are consistent with Model 1 in Table 8 (that is, the model based on observational data, without controls for the latent topics). These results are indicative of the robustness of our findings.

The following additional robustness tests are included in the appendices. First, Table A4 in Appendix 1 shows that many firms (in our sample of 15 firms) were using the different types of posts, and thus mitigates concerns that particular post types were concentrated in a small number of firms. Additionally, Table A5 in Appendix 1 shows the share of posts for each firm in our sample, and mitigates concerns that our results are driven by one or two firms with a significant share of all the posts. Second, Appendix 2 Sections 3 and 4 provide details regarding robustness tests related to the PSM procedure. These include: i) *t*-tests of equality of means, which confirm that the treatment and the control groups are balanced on observable characteristics (Appendix 2 Section 3), ii) graphs of the propensity score distributions for the treatment and control groups, which show that our data meet the common support assumption (Appendix 2 Section 3), and iii) a Rosenbaum bounds sensitivity analysis, which shows that the PSM estimates in our study are robust to the presence of unobserved confounders (Appendix 2 Section 4). Third, Appendix 2 Section 5 provides details regarding a placebo test, which shows that our results are unlikely to be driven by random chance.

6. HETEROGENOUS EFFECTS

In order to offer a dynamic perspective on how firm posts influence engagement, we examine the heterogeneous effects of different types of post content. Our first set of analyses examines whether particular types of firm content are more effective at stimulating engagement earlier in the life cycle of an online brand community, but become less effective as time goes by. Our second set of analyses uses quantile regressions to inform us about what types of content are likely to be effective in the case of posts that are ex-ante likely to have a larger number of likes (e.g., posts about mainstream products with broad appeal), versus posts that are ex-ante likely to have a smaller number of likes (e.g., posts about niche products). These analyses help us go beyond simply identifying that certain categories of firm posts are more likely to be associated with engagement.

6.1. Subsample Analysis

We conducted subsample analysis to identify if different types of post content are likely to be more effective earlier in a firm's life cycle versus later, or vice versa. We split our data into two subsamples as follows: early posts were classified as those posted by a firm in the first 1.5 years of a firm's online brand community (that is *Firm Age* <= 78 weeks). Content posted after 1.5 years was considered late posts. Another reason that we separate our subsample based on this criterion is that 78 weeks is the median *Firm Age* in our sample. We repeated our post level analyses with each of these subsamples separately. Results for early and late subsamples can be found in Table 9 Models 1 and 2, respectively. The results of both models are consistent, in direction and significance, with our main analyses with the entire sample (shown in Table 8, Model 2). Interestingly, for the post types that are significantly associated with engagement, the magnitude of effect for these post types are significantly higher (p value < 0.05) for early posts in comparison to late posts. T-tests associated with these results can be found in Appendix 1, Table A6. These results suggest that the post types that were found to be significantly associated with engagement have a significantly greater effect on engagement in the first year and a half of creating an online brand community, than later on in the community's life cycle.

Since the two broad strategies that we hypothesized about, namely influencing perceptions and monetary incentives, are likely to entail different financial costs to the firm, we compare the effectiveness of the two strategies. We find that the results for early posts (Table 9, Model 1) are consistent, in direction and significance level, to those for the entire sample (detailed at the end of section 5.2). Late posts (Table 9, Model 2) also show the same results with one exception. The coefficient on *Opinion* is significantly greater than the coefficient on *Promotion or Offer* (p-values < 0.01).

The empirical results suggest that (in the case of both online brand communities that are earlier in the life cycle as well as those that are later in their life cycle) posts about firm milestones, partnerships, or awards, and posts that convey product and industry knowledge through information about collection or picks have a bigger effect on engagement than posts about promotions or offers. In contrast, three types of posts that convey product and industry knowledge, namely those with tips or suggestions, industry information, and design origin information have smaller effects on engagement than posts about promotions or offers. Moreover, for online brand communities that are earlier in the life cycle, that is in their first year and a half, posts that seek opinions have a smaller effect than those about promotions or offers. For communities that are later on in their life cycle (that is past the one-and-a-half-year point), the effect of seeking opinions is significantly higher than the effect of promotions or offers.

Table 9. Subsample Analysis - Impact of Different Types of Firm Posts on Engagement

	(1)	(2)
	Early Posts	Late Posts
Capability of Key Members	-0.52	-0.85
	[0.38]	[0.63]
Association with Experts	0.017	-0.028
	[0.11]	[0.21]
Tips or Suggestions	0.18***	0.11**
	[0.058]	[0.053]
Industry Information	0.34***	0.12***
	[0.085]	[0.031]
Design Origin Information	0.42***	0.11***
	[0.067]	[0.038]
Collection or Picks	0.69***	0.20***
	[0.089]	[0.078]
Professional Process	0.30	0.18
	[0.52]	[0.26]
Professional Structure	0.28	0.08
	[0.24]	[0.15]
Firm Milestone Partnership or Award	0.88***	0.32***
	[0.26]	[0.11]
Product Award or Media Mention	0.074	0.16
	[0.11]	[0.25]
Opinion	0.44***	0.21***
	[0.041]	[0.052]
Promotion or Offer	0.54***	0.18***
	[0.048]	[0.061]
Contest	0.26	0.045
	[0.18]	[0.085]
Firm Video	0.39	-0.12
	[0.52]	[0.14]
External Video	-0.33	-0.69***
	[0.22]	[0.17]
Firm Article	-0.56***	-0.42***
	[0.096]	[0.072]
External Article	-0.39***	0.0068
	[0.064]	[0.15]
Image	0.56***	0.78***
	[0.049]	[0.066]
Firm Age	0.012***	0.071***
	[0.0013]	[0.0042]
Google Trends	0.014***	0.023***
	[0.00084]	[0.0017]
Topic 1	0.11	0.59***
	[0.074]	[0.090]
Topic 2	-0.99***	-0.68***
	[0.083]	[0.088]
Topic 3	0.16	0.13

	[0.14]	[0.18]
Topic 4	0.16**	0.26**
	[0.073]	[0.12]
Firm dummies	Yes	Yes
Time dummies	Yes	Yes
Constant	0.038	-8.24***
	[0.29]	[0.53]

Robust standard errors in brackets *** p<0.01, ** p<0.05, * p<0.1

6.2. Heterogeneous Effects Using Quantile Regressions

We examined whether the magnitude of the effect of different types of firm posts might vary significantly at different quantiles of the distribution of engagement. We used quantile regression analysis because it is particularly useful when the conditional distribution of the dependent variable is not symmetric and does not have a "standard" shape (Koenker and Hallock 2001), as is the case with our dependent variable (likes). While our main regression model estimated how different types of firm posts affect the conditional mean of engagement, the quantile-regression analysis estimates the impact of different types of firm posts on the entire distribution of the engagement variable, and thus allowed us to identify heterogeneous effects of the different post types.

For our quantile models, we focused on the seven types of post content, corresponding to our confirmed hypotheses (that is, post types that were found to be significantly associated with engagement). In Figure 4, the solid line connects the parameter estimates of the coefficients of the quantile regressions, and the shaded area represents the 95% confidence interval of the parameter estimates. The OLS estimates (shown in Table 8, Model 1) are plotted as horizontal dashed lines in this figure.

We find that for the covariate *Tips or Suggestions*, the heterogeneity across quantiles is small, so the OLS estimates in Table 8, Model 1 are reasonable approximations for quantile estimates. However, for all other covariates, the quantile estimates exhibit significant heterogeneity across quantiles. For instance, for *Design Origin Information,* we find that the quantile estimates are significantly larger at the low end than those at the high end (ranging from approximately 0.3 at the 0.1th quantile to approximately 0.1 at the 0.8th quantile). To examine whether the observed differences among the estimates are statistically significant across quantiles, we further conducted interquantile tests of the equality of pairwise treatment effects (Koenker and Hallock 2001). More specifically, we tested whether the quantile estimate in the 0.1th quantile is the same as the quantile treatment effect in the 0.8th quantile for *Design Origin Information*. The tests show that the treatment effect at the 0.1th quantile is significantly different from that at the 0.8th quantile (p value < 0.05). Figure 6 indicates that the OLS estimates tend to underestimate the impact of posts with *Design Origin Information* at the low quantiles of the distribution of likes, and tend to overestimate the impact at the high end.

The implication of the empirical result is that *Design Origin Information* exhibits significant heterogeneity across quantiles: The effect of *Design Origin Information* on low quantiles of the distribution of likes is greater than that on high quantiles. In other words, although the average impact of posts with *Design Origin Information* on likes (the impact on the conditional mean) is 0.26 (Table 8, Model 1), for posts that are ex-ante likely to have a larger number of likes (e.g., posts about mainstream products with broad appeal), having *Design Origin Information* is less effective in terms of increasing engagement (the magnitude of the impact is not statistically different from 0 at the 0.8th quantile). However, for posts that are likely to have a smaller number of likes (e.g., posts about niche products), having *Design Origin Information* is more effective (the magnitude of the impact is approximately 0.3 at the 0.1th quantile).

Examining the results for *Industry Information*, leads us to a different inference. For posts that are ex-ante likely to have a larger number of likes (e.g., those about mainstream products), having the *Industry Information* dimension is more effective (the magnitude of the impact is approximately 0.4 at the 0.8th quantile), but for posts that are ex-ante likely to have a smaller number of likes (e.g., posts about niche products), having the *Industry Information* dimension is less effective (the magnitude of the impact is not statistically different from 0 at the 0.1th quantile).

T-tests confirm that the quantile estimates for *Collection or Picks* and *Firm Milestone Partnership or Award* exhibit the same pattern as *Design Origin Information*. In contrast, t-tests confirm that the quantile estimates for *Opinions* and *Promotion or Offer* exhibit the same pattern as *Industry Information*.

These empirical findings provide nuanced insights about the heterogeneous effects of post content, which can help inform managers' posting strategies.

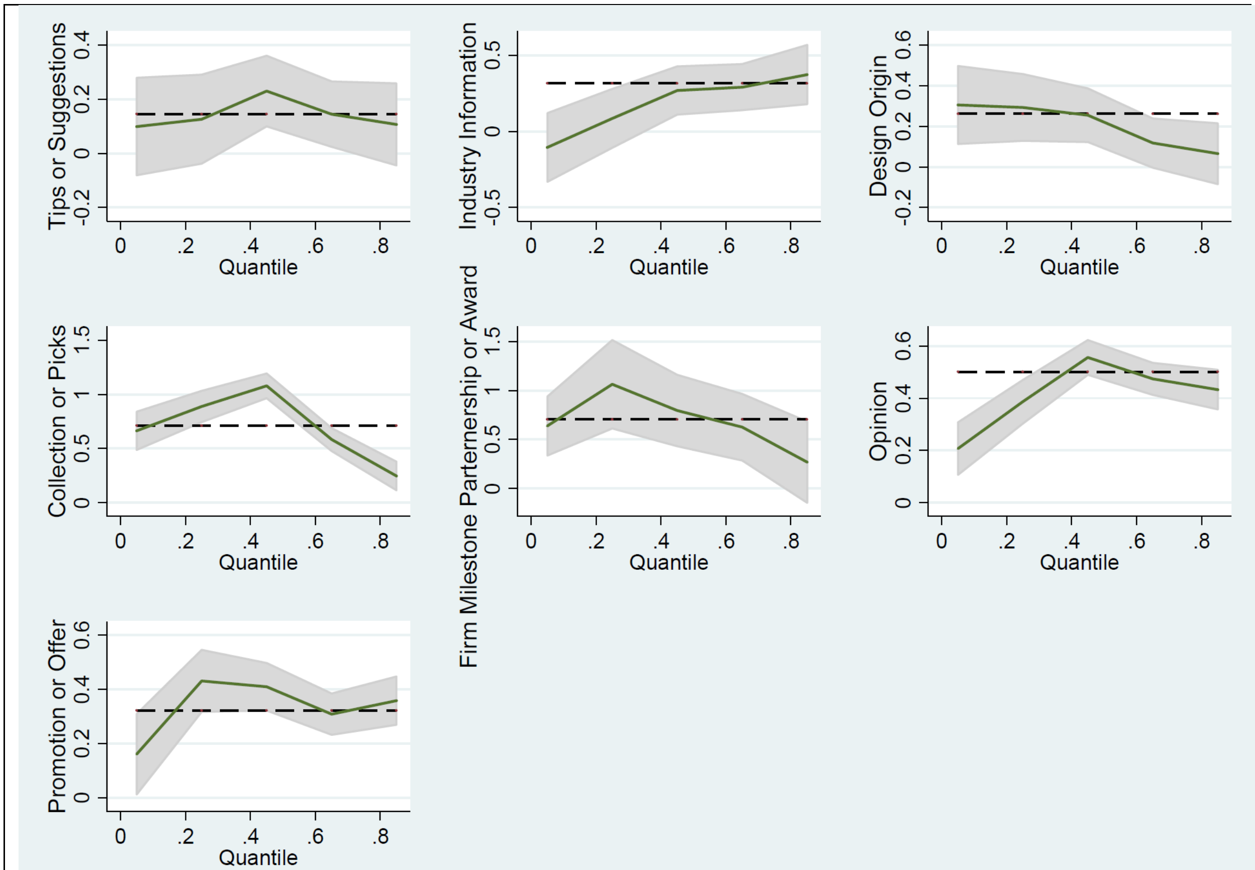

Figure 4. Quantile Treatment Effects of Different Types of Firm Posts on Engagement

7. DISCUSSION

Several researchers have established that online brand communities provide value to firms, including information dissemination, increased brand awareness, brand building, positive word of mouth, increased ROI, increased profitability, customer satisfaction and loyalty (Goh et al. 2013; Hoffman and Fodor 2010; Kumar et al. 2013; Rishika et al. 2013; Culnan et al. 2010). For an online community to provide benefits it needs a large member base (Butler 2001). Culnan et al. (2010) note that "simply creating a presence on a popular platform such as Twitter or Facebook won't guarantee that an organization will attract customers to its page", and the difficulty in cultivating brand communities has been documented (e.g., Worthen 2008; Culnan et al. 2010). Yet there is little prior research that addresses the question of how firms can grow their fledgling online brand communities. The present paper seeks to fill this gap in research. We theorize that firm posts that stimulate engagement may kindle subsequent growth in the community. When individuals engage with firm content, such interactions are broadcast, by the social media platform, to others who are not necessarily part of the firm's online brand community. Such social diffusion of information about the interaction and the related content posted by the firm, provides individuals with new information about the firm, based on which they may decide to join the online brand community of the firm. Using data from the retail industry, we provide empirical evidence for the association between engagement and growth in online brand community.

Further, we explore a related question, that is, what types of postings by firms might influence engagement in their online brand communities. Information systems researchers have only recently begun to explore this topic. Lee et al. (2017) investigate the effect of content related to brand personality and informative content on engagement. Since firm postings are believed to underpin brand positioning and perceptions (Gallaugher and Ransbotham 2010) we draw from sensegiving literature on influencing perceptions (Gioia and Chittipeddi 1991), to hypothesize about the post content that might be associated with engagement in online brand communities. In addition, since monetary incentives have been identified in prior literature as a potential motivator for engagement (Jarvenpaa and Tuunainen 2013; Muntinga et al. 2011), we also hypothesize about the effect of monetary incentives communicated via firms posts, on engagement. Our empirical analyses indicate that firm posts that influence perceptions including those that: convey firm credibility through product and industry knowledge, convey organizational achievements through information about firm milestones, partnerships or awards, and seek opinions are associated with engagement. We find that the association between product and industry knowledge and engagement holds for different approaches to conveying this information, including conveying tips or suggestions, industry information, design origin information, and collections or picks. In addition, we find that firm posts which convey monetary incentives in the form of promotions or offers are associated with engagement.

Posts that convey monetary incentives about offers and promotions are likely to entail a different financial cost to the firm than posts that influence perceptions through content that conveys product and industry knowledge, conveys organizational achievements, and seeks opinions. Thus, we contrast the effectiveness of these two strategies, and find that posts about firm milestones partnerships or awards, and posts that convey product and industry knowledge with information about collection or picks have a bigger effect on engagement than posts about promotions or offers. In contrast, three types of posts that convey product and industry knowledge, namely those with tips or suggestions, industry information, and design origin information have smaller effects on engagement than posts about promotions or offers.

Based on the aforementioned findings, our study contributes to understanding how to apply the theory of sensegiving and symbolic actions to building online brand communities. We further contribute to this literature by examining the heterogeneous effects of the different types of post content on engagement. We find that for all the post types in our study that are significantly associated with engagement, the magnitude of effect is significantly higher when these types of posts are used in the first year and a half of creating an online brand community. Furthermore, the results of quantile regressions suggest that a firm's post that is ex-ante likely to receive lesser engagement (e.g., a post about a niche product) might benefit from firm content that i) conveys organizational achievements through information about firm milestones, partnerships or awards, and ii) conveys product and industry knowledge through either design origin information or collections and picks. In contrast, a post that is ex-ante likely to receive higher levels of engagement (e.g., a post about a mass appeal product) might benefit from firm content that i) seeks others' opinions ii) conveys promotions or offers, or iii) conveys product and industry knowledge through content about the industry.

Finally, we contribute to theory on sensegiving and symbolic actions. Prior work (Petkova et al. 2013) posits that seeking opinions and receiving feedback through "face-to-face" social interactions at interactive events such as conferences provide an opportunity for sensegiving. Our study extends the boundary of this sensegiving action beyond face to face interactions to also include interactions via social media, which potentially allows for a much larger reach.

This study is relevant to managers and timely because firms are rapidly adopting profiles (brand community pages) on social networking sites. While several social networking analytics sites provide firms with up to the minute statistics such as community size and engagement level, the analytics sites leave it to the firm to determine what leads to better or worse performance on these metrics. This study provides nuanced insights into the type of post content that is associated with engagement, and shows that this engagement in turn is associated with online brand community growth. Firms can use this research to guide their posting strategy. Additionally, managers can benefit from our dynamic perspective on engagement. All the post types in our study that are significantly associated with engagement, are more effective when used in the first year and a half of creating an online brand community. Furthermore, we help inform managers' posting strategies by suggesting what post content is likely to be more effective for a post that is ex-ante likely to receive lesser engagement (e.g., a post about a niche product) versus a post that is ex-ante likely to receive higher levels of engagement (e.g., a post about a mass appeal product).

The study examines firms in the flash sales segment of the retail industry, a fast growing (Cocotas 2012) and intensely competitive e-commerce segment (Rao 2015), and focuses on the postings of firms since they join Facebook. Moreover, the study uses a manual text analyses approach (as opposed to relying solely on text mining), which limits our sample size. These factors affect the generalizability of our findings. We expect that our findings may be generalizable to firms in the retail industry that are in the early phases of growing their online communities. However, the findings of this study may not generalize to other businesses such as retail businesses with long established online brand communities, business-to-business firms, firms in the service industry, or brick and mortar firms. Future research could investigate the social media strategies adopted by such firms to gain members or to maintain their online community membership, and also examine differences between these settings and our setting.

A strength of this study is the manual examination of posts. Studies that analyze the content of textual data from social media typically use text-mining tools because the volume of data involved makes manual analysis difficult. For instance, Goh et al. (2013) used commercial text mining tools to capture the informative and persuasive nature of content shared on Facebook. Text mining, however, has many limitations (Mandviwalla and Watson 2014). For example, it does not analyze text embedded in images, or consider what the image itself conveys, nor does it allow the analysis of the content of videos linked to posts. Also, text mining does not analyze icons and symbols incorporated in posts. Depending on the way the text mining data is collected and configured, the analysis may not include content linked to posts such as blogs and articles, as well as captions associated with photos. Further, fine-grained analysis of text can be challenging when using text-mining tools. For example, in our study, we needed to differentiate between three types of awards that may be mentioned in post content, that is, awards to founders or key employees, awards to the firm, and whether the firm sells award-winning products. To allow for a granular and nuanced analysis of posts, we opted for manual rating of posts and used Amazon's Mechanical Turk to find our raters. These raters not only analyzed the text of the post but went to Facebook to find the actual post (based on identifying information) and rated the post after reading articles linked to the post, watching videos embedded in the post, and scrolling through photos associated with the post. The process we developed for qualifying workers and monitoring their output, illustrated in Figure 2, enables us to overcome the previously mentioned limitations of text mining tools and results in nuanced content analysis for a large volume of data.

A further strength of our study is that our empirical approach combines explanatory variables that are informed by theory with algorithmically identified controls that are constructed by allowing the data to speak for themselves. We use LDA to algorithmically identify topics that firms may be posting about, but which our theory may not have identified. This approach provides cleaner identification of the parameter of interest when using large data sets from social media, such as those in our study.

Our study is not without limitations. First, firms can pay companies to acquire thousands of community members within days. According to an article in the public press, such companies hire college or school students to create fake profiles on social networking sites and then utilize these profiles to increase the fan base of their client's firms (Anver 2013). Our study is unable to control for this directly. However, if firms undertake these practices, it would bias against us finding any consistent results, since we would be unlikely to see systematic influences of particular types of posts on engagement. Moreover, to the extent that this practice is ongoing for a particular firm, it would be a stable characteristic of the firm, controlled in our fixed effect. It is also important to note that Facebook has never permitted the purchase or sale of Facebook Likes, and their website indicates that detection of such activity could result in action against the firm by Facebook. Thus, using Facebook as our social media setting helps mitigate this issue. In addition to Facebook's warning, its team dedicated to examining such complaints is likely to serve as a deterrent for the firms in our study. Second, our sample is limited to firms that were founded in 2010. If there is something special about firms founded in this particular year that does not apply to firms that were founded in other years, then our analyses would not be able to account for this difference.

References:

Anver, J. 2013. "Money can't buy love, can buy 'likes'," *The Times of India* http://articles.timesofindia.indiatimes.com/2013-07-14/deepfocus/40569369_1_facebook-likes-facebook-page-followers, Accessed Jan, 2014.

Aral, S., and Walker, D. 2011. "Creating social contagion through viral product design: A randomized trial of peer influence in networks," *Management Science* (57:9), pp. 1623-1639.

Aral, S., Dellarocas, C. and Godes, D., 2013. "Introduction to the special issue—social media and business transformation: a framework for research," *Information Systems Research* (24:1), pp. 3-13.

Aral, S., and Walker, D. 2014. "Tie strength, embeddedness, and social influence: A large-scale networked experiment," *Management Science* (60:6), pp. 1352-1370.

Arellano, M., and Bond, S. 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The Review of Economic Studies* (58:2), pp. 277-297.

Baer, J. 2013. 4 "Reasons Google Bought Wildfire," *Convince & Convert* http://www.convinceandconvert.com/social-media-software/4-reasons-google-bought-wildfire/, Accessed Jan 2014.

Bagozzi, R. P., and Yi, Y. 1991. "Multitrait-multimethod matrices in consumer research," *Journal of Consumer Research* (17:4), pp. 426-439.

Bakshy, E., Rosenn I., Marlow, C., and Adamic, L. 2012. "The role of social networks in information diffusion," *Proceedings of the 21st international conference on World Wide Web, Association for Computing Machinery,* April, pp. 519-528.

Bernstein, M. S., Bakshy, E., Burke, M., and Karrer, B. 2013. "Quantifying the invisible audience in social networks," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Association for Computing Machinery,* April, pp. 21-30.

Bikhchandani, S., Hirshleifer, D., and Welch, I. 1992. "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of Political Economy* (100:5), pp. 992-1026.

Bikhchandani, S., Hirshleifer, D., and Welch, I. 1998. "Learning from the behavior of others: conformity, fads, and informational cascades," *Journal of Economic Perspectives* (12:3), pp. 151-170.

Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent dirichlet allocation," *Journal of Machine Learning Research* (3:Jan), pp. 993-1022.

Brown, A. D. 1994. "Politics, symbolic action and myth making in pursuit of legitimacy," *Organization Studies* (15:6), pp. 861-878.

Buhrmester, M., Kwang, T., and Gosling, S. D. 2011. "Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data?" *Perspectives on Psychological Science* (6:1), pp. 3-5.

Burtch, G., Ghose, A., and Wattal, S. 2013. "An empirical examination of the antecedents and consequences of contribution patterns in crowd-funded markets," *Information Systems Research* (24:3), pp. 499-519.

Butler, B. 2001. "Membership size, communication activity, and sustainability: A resource-based model of online social structures," *Information Systems Research* (12:4), pp. 346-362.

Carey-Simos G. 2015. "How much data is generated every minute on social media?" Wersm. http://wersm.com/how-much-data-is-generated-every-minute-on-social-media/, Accessed July 2017.

Chen, H., De, P. and Hu, Y.J., 2015. "IT-enabled broadcasting in social media: An empirical study of artists' activities and music sales," *Information Systems Research* (26:3), pp.513-531.

Cocotas, A. 2012. "Flash sales will be a $6 billion market by 2015," *Business Insider, BI Intelligence*. http://www.businessinsider.com.au/special-report-flash-sales-2012-1, accessed on April 13, 2012.

Cohen, J. 1960. "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement* (20:1), pp. 37-46.

Collins. 2018. collinsdictionary.com. Accessed June, 2018.

Constine J. 2017. "Facebook now has 2 billion monthly users …. and responsibility," Tech Crunch. https://techcrunch.com/2017/06/27/facebook-2-billion-users/, Accessed July 2017.

Cvijikj, I.P. and Michahelles, F., 2013. "Online engagement factors on Facebook brand pages," *Social Network Analysis and Mining* (3:4), pp.843-861.

Culnan, M.J., McHugh, P.J. and Zubillaga, J.I., 2010. "How large US companies can use Twitter and other social media to gain business value," *MIS Quarterly Executive* (9:4), pp. 243-259.

De Vries, L., Gensler, S. and Leeflang, P.S., 2012. "Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing," *Journal of Interactive Marketing* (26:2), pp. 83-91.

Digital Marketing Glossary. 2012. "What is Facebook forced like definition?" http://digitalmarketing-glossary.com/What-is-Facebook-forced-like-definition, Accessed May 2014.

Duan, W., Gu B., and Whinston A. B. 2009. "Informational cascades and software adoption on the internet: an empirical investigation," *MIS Quarterly* (33:1), pp. 23-48.

Duan, W., Gu, B., and Whinston, A. B. 2008. "Do online reviews matter? An empirical investigation of panel data," *Decision Support Systems* (45:4), pp. 1007-1016.

Ellison, N. B., and Boyd, D. 2013. "Sociality through social network sites," in *The Oxford Handbook of Internet Studies*, W. H. Dutton (ed.), Oxford University Press, Oxford, pp. 151-172.

Gallaugher, J. and Ransbotham, S., 2010. "Social media and customer dialog management at Starbucks," *MIS Quarterly Executive* (9:4), pp. 197-212.

Geva, T., Oestreicher-Singer, G., Efron, N. and Shimshoni, Y., 2015. "Using forum and search data for sales prediction of high-involvement products," *MIS Quarterly* (41:1), pp. 65-82.

Gioia, D. A., and Chittipeddi, K. 1991. "Sensemaking and sensegiving in strategic change initiation," *Strategic Management Journal* (12:6), pp. 433-448.

Goh, K. Y., Heng, C. S., and Lin, Z. 2013. "Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content," *Information Systems Research* (24:1), pp. 88-107.

Gray, L. 2012. "Facebook Launches Promoted Posts, Pay As Little As $5 For More Fans To See Page Content," *Social Fresh.* http://socialfresh.com/facebook-promoted-posts-go-live/, Accessed May 2014.

Habibi, M. R., Laroche, M., and Richard, M. 2014. Brand communities based in social media: How unique are they? Evidence from two exemplary brand communities," *International Journal Of Information Management* (34:2), 123-132.

Hannan, M. T., and Freeman, J. 1984. "Structural inertia and organizational change," *American Sociological Review* (49:2), 149-164.

Heinrich, C., Maffioli, A., and Vazquez, G. 2010. "A primer for applying propensity-score matching," Inter-American Development Bank.

Higgins, M. C., and Gulati, R. 2003. "Getting off to a good start: The effects of upper echelon affiliations on underwriter prestige," *Organization Science* (14:3), 244-263.

Hoffman, D.L. and Fodor, M., 2010. "Can you measure the ROI of your social media marketing," *MIT Sloan Management Review*, (52:1), pp. 41-49.

Hsiao, C. 2003. *Analysis of Panel Data,* Cambridge University Press.

Jang H., Olfman L., Ko I., Koh J. and Kim K., 2008. "The Influence of On-Line Brand Community Characteristics on Community Commitment and Brand Loyalty", *International Journal of Electronic Commerce*," (12:3), pp. 57-80.

Jarvenpaa, S.L. and Tuunainen, V.K., 2013. “How Finnair socialized customers for service co-creation with social media,” *MIS Quarterly Executive* (12:3), pp. 125-136.

Kane, G.C., Alavi, M., Labianca, G.J. and Borgatti, S., 2014. “What’s different about social media networks? A framework and research agenda”, *MIS Quarterly* (38:1), pp. 275-304.

Kaplan D. 2017. “Facebook starts 2017 with 65 million local business pages.” GeoMarketing. http://www.geomarketing.com/facebook-starts-2017-with-65-million-local-business-pages, Accessed July 2017.

King, WR., Liu, C. Z., Haney, M. H., and He, J. 2007. “Method effects in IS survey research: an assessment and recommendations,” *Communications of the Association for Information Systems* (20:30), pp. 457-482.

Kline, T. J., Sulsky, L. M., and Rever-Moriyama, S. D. 2000. “Common method variance and specification errors: A practical approach to detection,” *The Journal of Psychology* (134:4), pp. 401-421.

Koenker, R., and Hallock, K. 2001. “Quantile regression: An introduction,” *Journal of Economic Perspectives* (15:4), pp. 43-56.

Kumar, V., Bhaskaran, V., Mirchandani, R., and Shah M. 2013. “Creating a measurable social media marketing strategy: Increasing the value and ROI of intangibles and tangibles for hokey pokey,” *Marketing Science* (32:2), pp. 194-212.

Kvalseth, T. O. 1989. “Note on Cohen's kappa,” *Psychological Reports* (65:1), pp. 223-226.

Lee, D., Hosanagar, K. and Nair, H., 2016. “Advertising content and consumer engagement on social media: evidence from Facebook,” *Management Science*, forthcoming.

Li, X. 2016. “Could deal promotion improve merchants' online reputations? The moderating role of prior reviews,” *Journal of Management Information Systems* (33:1), pp. 171-201.

Mandviwalla, M., and Watson, R. 2014. “Generating capital from social media,” *MIS Quarterly Executive* (13:2), pp. 97-113.

Mehrhoff, J. 2009. “A solution to the problem of too many instruments in dynamic panel data GMM,” Deutsche Bank Discussion Series 1: Economic Studies 31.

Mileva, E. 2007. “Using Arellano-Bond dynamic panel GMM estimators in Stata,” *Economic Department, Fordham University*.

Miller, K., Fabian, F., and Lin, S. 2009. “Strategies for online communities,” *Strategic Management Journal* (30:3), pp. 305-322.

Miller, A. R., and Tucker, C. 2013. “Active social media management: The case of health care,” *Information Systems Research* (24:1), pp. 52-70.

Muniz, A. M. and O’Guinn T.C., 2001. “Brand Community,” *Journal of Consumer Research*, (27:4), pp. 412-432.

Muntinga, D.G., Moorman, M. and Smit, E.G., 2011. “Introducing COBRAs: Exploring motivations for brand-related social media use,” *International Journal of Advertising* (30:1), pp.13-46.

Nam, H., Joshi, Y. V., and Kannan, P. K. 2017. “Harvesting brand information from social tags,” *Journal of Marketing*, forthcoming.

Petkova, A. P., Rindova, V. P., and Gupta, A. K. 2013. “No news is bad news: Sensegiving activities, media attention, and venture capital funding of new technology organizations,” *Organization Science* (24:3), pp. 865-888.

Porter, C., and Donthu, N. 2008. “Cultivating Trust and Harvesting Value in Virtual Communities,” *Management Science* (54:1), pp. 113-128.

Qinag, L., Maggitti, P. G., Smith, K. G., Tesluk, P. E., and Katila, R. 2012. “Top management attention to innovation: The role of search selection and intensity in new product introductions,” *Academy of Management Journal* (56:3), pp. 893-916.

Rafaeli, A., & Vilnai-Yavetz, I. 2004a. “Emotion as a connection of physical artifacts and organizations,” *Organization Science* (15:6), pp. 671-686.

Rafaeli, A., and Vilnai-Yavetz, I. 2004b. “Instrumentality, aesthetics and symbolism of physical artifacts as triggers of emotion,” *Theoretical Issues in Ergonomics Science* (5:1), pp. 91-112.

Rao, H. 1994. “The social construction of reputation: Certification contests, legitimation, and the survival of organizations in the American automobile industry: 1895–1912,” *Strategic Management Journal* (15:S1), pp. 29-44.

Rao, L., 2015. “Why Flash Sales Are In Trouble,” Fortune. http://fortune.com/2015/12/16/flash-sales-trouble/. Accessed on August 29, 2017.

Rishika, R., Kumar, A., Janakiraman, R., and Bezawada, R., 2013. “The effect of customers' social media participation on customer visit frequency and profitability: an empirical investigation,” *Information Systems Research* (24:1), pp.108-127.

Roodman, D., 2009. “How to do xtabond2: An introduction to difference and system GMM in Stata,” *The Stata Journal* (9:1), pp. 86–136.

Sala-i-Martin, X. X. 1997. “I just ran two million regressions,” *American Economic Review* (87:2), pp. 178-183.

Salancik, G. R. 1984. "On priming, consistency, and order effects in job attitude assessment: With a note on current research," *Journal of Management* (10:2), pp. 250-254.

Santos, F. M., and Eisenhardt, K. M. 2009. "Constructing markets and shaping boundaries: Entrepreneurial power in nascent fields," *Academy of Management Journal* (52:4), pp. 643-671.

Segars, A. H., and Grover, V. 1998. "Strategic information systems planning success: an investigation of the construct and its measurement," *MIS Quarterly* (22:2), pp. 139-163.

Singh, P. V., Sahoo, N., and Mukhopadhyay, T. 2014. "How to attract and retain readers in enterprise blogging?" *Information Systems Research* (25:1), pp. 35-52.

Smith A. 2014. "6 new facts about Facebook," Pew Research Center. http://www.pewresearch.org/fact-tank/2014/02/03/6-new-facts-about-facebook/, Accessed July 2017.

Starr, J. A., and MacMillan, I. 1990. "Resource cooptation via social contracting: Resource acquisition strategies for new ventures," *Strategic Management Journal* (11:4), pp. 79-92.

Sudman, S., Bradburn, N. M., and Schwarz, N. 1996. *Thinking about answers: The application of cognitive processes to survey methodology*, Jossey-Bass.

Tourangeau, R., and Rasinski, K. A. 1988. "Cognitive processes underlying context effects in attitude measurement," *Psychological Bulletin* (103:3), pp. 299-314.

Worthen, B. 2008. "Why Most Online Communities Fail," *Wall Street Journal*, July 16.

Zaheer, A., and Soda, G. 2009. "Network evolution: The origins of structural holes," *Administrative Science Quarterly* (54:1), pp. 1-31.

Zott, C., and Huy, Q. N. 2007. "How entrepreneurs use symbolic management to acquire resources," *Administrative Science Quarterly* (52:1), pp. 70-105.

APPENDIX 1

Table A1. An Example from Each High-Level Category of Firm Post Type

Post Type	Post Content
Conveys Credibility (product and industry knowledge)	“There several ways a wine can be considered "green." Have you tried one? How did it taste?” (with a link to a blog article titled “How Can I Find “Green” Wines?)
Conveys Professional Organizing (professional structure)	“It’s time for the newest round of team member introductions… meet our new Social Media Manager <name deleted for privacy>! We’re so happy to have her as part of our team.”
Conveys Organizational Achievements (firm milestone partnership of award)	“In less than 24 hours Fab is going to be honored with not one, not two, but THREE Webby awards! We're beyond excited and extremely grateful to be receiving such high honors.”
Seeks Opinions	“Would you rather: Nautical or Bohemian? Choose 1.” (with an image of the two prints shown side by side)
Conveys Monetary Incentives (promotions or offers)	“It's official - $0 shipping on EVERY offer from 1-5pm! What will you pick up?”

Table A2. Correlation Matrix

	Likes	*Comm-ents*	*Shares*	*UGC Count*	*Firm Post Count*	*Google Trends*	*Acqui-sition Merger*	*New Product Line Added*	*Finan-cing Secured*
Comments	0.72								
Shares	0.84	0.63							
UGC Count	0.03	0.14	0.13						
Firm Post Count	0.13	0.17	0.02	0.05					
Google Trends	0.26	0.21	0.22	0.04	0.09				
Acquisition Merger	0.07	0.04	0.09	0.10	-0.02	0.06			
New Product Line Added	0.02	0.05	0.23	0.15	-0.02	0.06	-0.05		
Financing Secured	0.01	0.02	-0.03	0.03	0.17	0.01	-0.03	-0.03	
Award	0.00	-0.01	0.00	0.01	-0.11	0.01	-0.05	-0.04	-0.03

Table A3. Collinearity Diagnostics

	VIF	SQRT VIF	Tolerance	R-Squared
Likes	4.97	2.23	0.20	0.80
Comments	2.16	1.47	0.46	0.54
Shares	4.28	2.07	0.23	0.77
UGC Count	1.10	1.05	0.91	0.09
Firm Post Count	1.12	1.06	0.90	0.10
Google Trends	1.08	1.04	0.92	0.08
Acquisition Merger	1.03	1.02	0.97	0.03
New Product Line Added	1.20	1.09	0.84	0.16

Financing Secured	1.04	1.02	0.97	0.03
Award	1.02	1.01	0.98	0.02
Mean VIF 1.90				

Table A4. How Many Firms Had Each Type of Post

Post Type	Number of Firms
Capability of Key Members	9
Association with Experts	11
Tips or Suggestions	15
Industry Information	12
Design Origin Information	12
Collection or Picks	13
Professional Process	8
Professional Structure	7
Firm Milestone Partnership or Award	10
Product Award or Media Mention	11
Opinion	15
Promotion or Offer	15
Contest	12

Table A5. The Share of Posts for Each Firm

Firm	Share of Posts
Firm 1	6.02%
Firm 2	8.30%
Firm 3	6.65%
Firm 4	5.16%
Firm 5	7.05%
Firm 6	7.61%
Firm 7	10.79%
Firm 8	2.66%
Firm 9	5.20%
Firm 10	8.56%
Firm 11	3.52%
Firm 12	10.78%
Firm 13	4.10%
Firm 14	10.70%
Firm 15	2.89%

Table A6. Comparison of Coefficients in the Subsample Analysis

	t-statistic	*p*-value
Tips or Suggestions	86.70	0.00
Industry Information	236.63	0.00
Design Origin Information	391.65	0.00
Collection or Picks	402.93	0.00
Firm Milestone Partnership or Award	193.03	0.00
Opinion	338.00	0.00
Promotion or Offer	451.33	0.00

APPENDIX 2

1. Sentiment Analysis for UGC

It is conceivable that individuals might be joining online communities in order to complain. If this is the case, then the sentiment of user generated posts is likely to be negative. To mitigate this concern, we conducted a sentiment analysis on user generated posts and found that the overall sentiment of user generated posts is positive, thus suggesting that individuals are not joining brands' online communities to complain.

The sentiment analysis of user generated posts was conducted using Long-Term Short Term Memory (LSTM) networks, a deep learning technique. LSTMs are a type of network that "remembers" previous data and makes decisions based on that knowledge. These networks are especially relevant in sentiment analysis because each word in a sentence has meaning based on the surrounding words, that is, previous and upcoming words (Mousa and Schuller 2017). Our analyses generated a sentiment value for each user generated post. A post's sentiment value can range from minus one to plus one. If the sentiment value is greater than zero, the post's sentiment is positive, otherwise it is negative. The more positive the sentiment of the post, the greater the sentiment value. Table A7 shows the descriptive statistics of sentiment values of user generated posts. We find that the mean and median of sentiment values are 0.75 and 1, respectively. Additionally, 92.29% of user generated posts have a sentiment value larger than zero, suggesting that most user generated posts have positive sentiment, thus suggesting that individuals are not joining brands' online communities to complain.

Table A7. Descriptive Statistics of Sentiment Values of User Generated Post						
	Number of User Generated Posts	Mean	Std. Dev.	Min	Max	Median
Sentiment	22658	0.75	0.41	-0.59	1	1

2. Arellano – Bond Tests for Autocorrelation

We test whether the dynamic panel model is correctly specified. The Arellano – Bond tests for autocorrelation are reported in Table A8, and indicate that there is no serial correlation in the first-differenced disturbances.[8] In addition, both the Sargan and Hansen tests do not reject the null hypothesis (*p* value > 0.87 for both tests).[9] These test results suggest that our model specifications are based on instruments that are valid (Roodman 2009; Mileva 2007). Further, our results show that the coefficient for the lagged *L.Log(Community Size)* variable is positive and significant. This indicates that *Community Size* in the previous period is a good predictor of current *Community Size*, and hence pertinent to our model.

Table A8. Arellano-Bond Test for AR(1) and AR(2) in First-difference		
Order	z	Prob > z
AR(1)	-1.92	0.055
AR(2)	-0.58	0.564

3. Accessing the Quality of the PSM

We implemented the following tests to assess the quality of the PSM conducted (results of the regression model that uses the dataset generated by the PSM are shown in Table 8, Model 3).

[8] The Arellano-Bond test for autocorrelation has a null hypothesis of no autocorrelation, and is applied to the differenced residuals. We marginally reject the hypothesis at AR (1) (*p* value = 0.055). The test for AR (1) process in first differences is usually expected to reject the null hypothesis (Mileva 2007) and the test for AR (2) in first differences is more important (Mileva 2007; Roodman 2009). We do not reject the null hypothesis at AR(2) (*p* value = 0.564), implying that the model is not misspecified.

[9] The null hypothesis for these tests is that the instruments as a group are exogenous.

First, we performed *t*-tests of equality of means before and after the matching to check whether observable characteristics are balanced across posts in the treatment and control groups. The left panel of Table A8 shows clear evidence of covariate imbalance between groups before matching. After matching, the differences of means are no longer statistically significant (right panel of Table A8), suggesting that matching helped reduce the bias associated with differences in observable characteristics. In other words, the treated and control posts are more similar in terms of these characteristics. Figure A1 shows the standardized percentage bias for each covariate before and after matching, and indicates that the covariate imbalance has been reduced after matching. This is consistent with the results of the *t*-tests shown in Table A9.

Second, prior matching literature suggests that a common support or overlap condition is a critical assumption in matching (e.g., Heinrich et al. 2010). Checking the overlap or region of common support between treatment and control groups can be done through a visual inspection of the propensity score distributions for both the treatment and control groups (Heinrich et al. 2010). Figure A2 shows the propensity score histogram for the treatment and control groups, and reveals a clear overlapping of the distributions between the two groups, indicating that our data meet the common support assumption.

Table A9. Differences in Mean Before and After Matching

	Before Matching					After Matching				
	Mean Treated	Mean Control	% bias	*t*-statistic	*p*-value	Mean Treated	Mean Control	% bias	*t*-statistic	*p*-value
Firm Article	0.14	0.089	16.72	5.35***	0.00	0.14	0.14	0.32	0.07	0.95
External Article	0.17	0.058	36.93	13.53***	0.00	0.17	0.15	6.31	1.17	0.24
Firm Video	0.0021	0.0026	-1.11	-0.30	0.77	0.0021	0.0010	2.24	0.58	0.56
External Video	0.022	0.0066	12.92	4.99***	0.00	0.022	0.019	2.61	0.49	0.63
Image	0.87	0.83	10.44	2.94***	0.00	0.87	0.88	-2.02	-0.48	0.63
Firm Age	76.11	74.04	8.12	2.33**	0.02	76.11	75.86	1.03	0.23	0.82
Google Trends	31.32	38.86	-29.51	-8.45***	0.00	31.32	29.45	7.31	1.67*	0.096

Note: *** p<0.01, ** p<0.05, * p<0.1

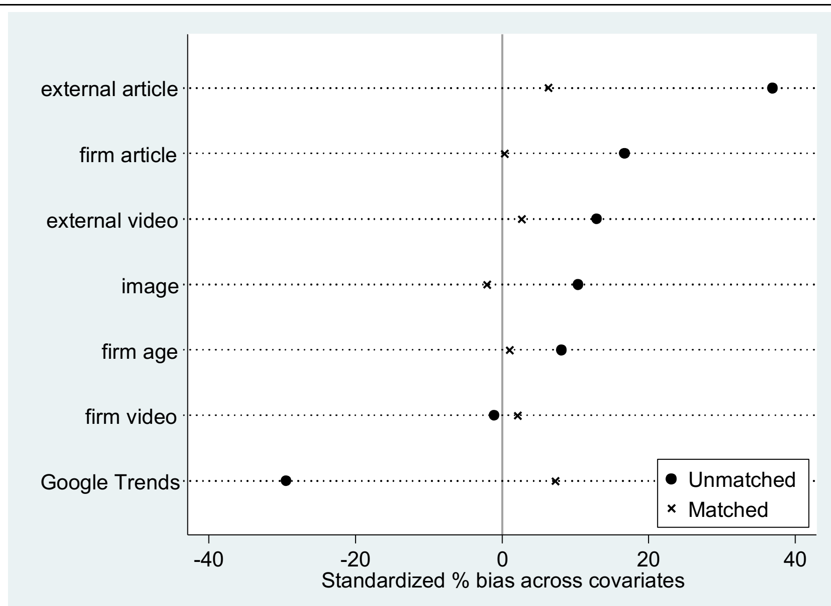

Figure A1. Standardized Percentage Bias for Each Covariate Before and After Matching

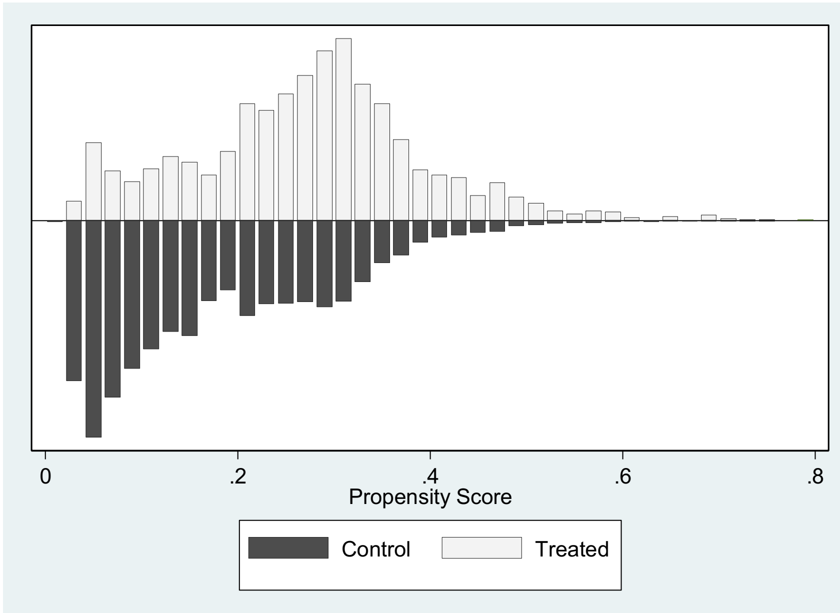

Figure A2. Propensity Score Histogram by Treatment Status

4. Rosenbaum Bounds – Sensitivity Analysis for PSM

We use PSM to try to address endogeneity concerns due to unobserved variables that might affect engagement as well as the explanatory variables. In our PSM analysis (Section 3.3.4.iv), we constructed control and treatment groups of matched posts and ensured that these groups were comparable on observable characteristics. However, it is possible that treatment and control groups differ on some unobserved characteristics (such as firm performance) even after matching on observed characteristics, leading to a hidden bias.

Rosenbaum's method of sensitivity analysis assesses if the estimates that are based on PSM are robust to the possible presence of an unobserved confounder. The sensitivity analysis provides a statement about the magnitude of the hidden bias that would need to be present to explain the associations actually observed (Rosenbaum 2002).

The sensitivity analysis considers several possible values of Γ, and shows how inferences might change based on the value of Γ. Γ is the odds ratio of treatment assignment, which is a measure of the degree of departure from a study that is free of hidden bias. Table A10 reports p-values from Wilcoxon signed rank tests for the averaged treatment effect while setting the level of hidden bias to different values of Γ.

When $\Gamma = 1$, we assume the absence of unobserved selection bias. In this case, both the upper and lower bounds of p values are zero (sig+ = 0, sig- = 0) indicating that the impact of post type is significant when there is no hidden bias. When $\Gamma = 8$, the upper bound of *p* value is greater than 5% (sig+ = .22, sig- = 0). Thus, we infer that for the impact of post type to disappear, the odds of differential assignment (to treatment and control group) due to unobserved factors is about 8. Intuitively, this means that for the impact of post type to disappear, the unobserved confounder has to cause a post to be eight times as likely as another post to receive treatment (assuming the two posts have the same observable characteristics). An Γ value of 8 is very large (Keele 2010, Guo and Fraser 2010), and it is well above the threshold values used in prior studies for robust PSM (e.g. Keele 2010, Wei and Lin 2017).

It is worth noting that the Rosenbaum bounds are "worst-case" scenarios. An insignificant upper bound *p* value for $\Gamma = 8$ does not mean that there is no true effect of post type when $\Gamma = 8$. This result means that the confidence interval for the effect of post type would include zero if an unobserved variable causes the odds ratio of treatment assignment to differ between treatment and control groups by 8.

In summary, this analysis suggests that the PSM estimates in our study are robust to the presence of unobserved confounders.

Table A10. Rosenbaum Bounds for PSM: Range of Significant Levels for the Signed Rank Statistic

Γ	Sig+	Sig-
1	0	0
2	2.4e-12	0
3	1.6e-10	0
4	3.9e-04	0
5	0.0026	0
6	0.012	0
7	0.036	0
8	0.22	0

Γ: is the odds of differential assignment due to unobserved factors,
Sign+: upper bound significance level, and
Sign-: lower bound significance level.

5. Placebo Test

We conducted a placebo test to check if our results could be driven entirely by chance. Following the technique implemented by Bertrand et al. (2004), we randomly assigned post types to all the posts in our sample and re-estimated our post level regression equation (described in Section 3.3.3. of the main manuscript). We repeated this exercise 1000 times. Since the randomly assigned post types are fake, a significant “effect” at the 5% level should be found at most 5% of the time (50 times). From the 1000 runs, we find that the fraction of simulations in which the null hypothesis is rejected is 4.2% of the time (42 times). These results suggest that our results are unlikely to be driven by random chance, and thus help reduce concerns regarding identification of the effects described.

References

Bertrand, M., Duflo, E., and Mullainathan, S., 2004. “How much should we trust differences-in-differences estimates?.” *Quarterly Journal of Economics*, (119:1), pp. 249-275.

Guo, S., and Fraser, M. 2010. Propensity Score Analysis: Statistical Methods and Applications. Los Angeles, CA: Sage Publications.

Keele, L. 2010. An overview of rbounds: An R package for Rosenbaum bounds sensitivity analysis with matched data. Working Paper, Accessed link:

http://www.personal.psu.edu/ljk20/rbounds%20vignette.pdf.

Mousa, A., and Schuller, B., 2017. Contextual bidirectional long short-term memory recurrent neural network language models: A generative approach to sentiment analysis. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (1), pp. 1023-1032.

Rosenbaum, P. R. 2002. Observational Studies. 2nd ed. New York: Springer.

Wei, Z., and Lin, M. 2017. "Market mechanisms in online peer-to-peer lending," *Management Science*, forthcoming.