Saïd Business School
Research Papers

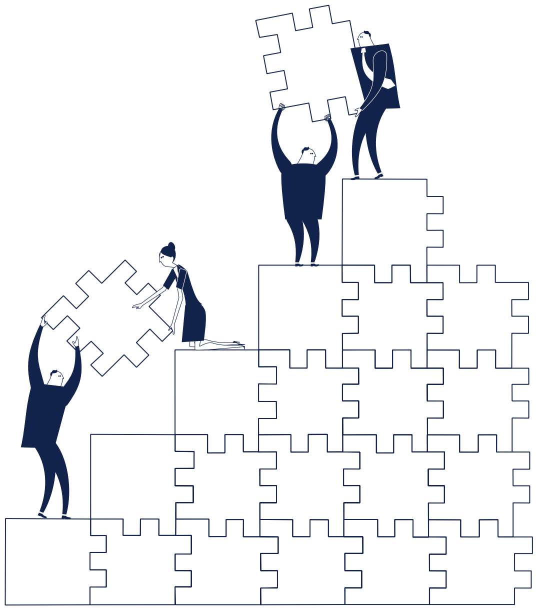
September 2015

Using Social Media Monitoring Data to Forecast Online Word-of-Mouth Valence: A Network Autoregressive Approach

Felipe Thomaz
Darla Moore School of Business, University of South Carolina

Andrew Stephen
Saïd Business School, University of Oxford

Vanitha Swaminathan
Joseph M. Katz Graduate School of Business, University of Pittsburgh



Saïd Business School RP 2015-15

USING SOCIAL MEDIA MONITORING DATA TO FORECAST ONLINE WORD-OF-MOUTH VALENCE: A NETWORK AUTOREGRESSIVE APPROACH

Felipe Thomaz
Assistant Professor of Marketing
Darla Moore School of Business, University of South Carolina
1014 Greene Street, Columbia, SC 29208
Tel: 803-777-5918
Email:Felipe.Thomaz@moore.sc.edu

Andrew Stephen
L'Oréal Professor of Marketing
Saïd Business School, University of Oxford
Email: andrew.stephen@sbs.ox.ac.uk

Vanitha Swaminathan
Professor of Business Administration & Robert W. Murphy Faculty Fellow in Marketing
Joseph M. Katz Graduate School of Business, University of Pittsburgh
344 Mervis Hall, Pittsburgh PA 15260
Tel 412-648-1579
Email: vanitha@katz.pitt.edu

September 2015

* Felipe Thomaz (Felipe.Thomaz@moore.sc.edu) is Assistant Professor of Marketing, University of South Carolina, Andrew T. Stephen (andrew.stephen@sbs.ox.ac.uk) is L'Oréal Professor of Marketing, Saïd Business School, University of Oxford, and Vanitha Swaminathan (vanitha@katz.pitt.edu) is Professor of Business Administration & Robert W. Murphy Faculty Fellow in Marketing at the Joseph M. Katz Graduate School of Business, University of Pittsburgh. The authors thank Jeff Inman, R. Venkatesh, Peter Zubcsek, and David Krackhardt for their assistance with this research and are grateful to the University of Pittsburgh and the INSEAD Alumni Fund for financial support. This paper appeared as an essay in Felipe Thomaz's doctoral dissertation.

ABSTRACT

Managers increasingly use social media for marketing research, particularly to monitor what consumers think about brands. Although social media monitoring can provide rich insights into consumer attitudes, marketers typically use it in a backward-looking manner—that is, to measure past online word-of-mouth (WOM) valence (i.e., sentiment). This article proposes a novel method for using social media monitoring in a forward-looking manner to forecast brands' future online WOM valence. The approach takes into account information on related brands based on the premise that consumers' attitudes toward one brand are likely relative to—and therefore associated with—attitudes toward other brands. The method infers associative relations between brands from social media monitoring data by observing which brands are mentioned at the same time in the same social media sources, thus enabling construction of time-varying brand "networks" for representing interdependencies between brands. The authors test six possible methods for capturing brand interdependencies (Jaccard, Dice, anti-Dice, correlation, normalized correlation, and Euclidean distance) and examine the relative performance of each alternative method with a view to identifying the best approach.

Keywords: social media, valence, Word-of-mouth, network autoregressive, brand interdependence, forecasting

Social media monitoring is a fast-growing and increasingly specialized area of marketing research. Firms use social media monitoring services to track brand and product mentions across various online social media sources, such as online social networking platforms (exemplified by Facebook and Twitter), blogs, and online discussion forums. These services typically provide firms with two types of brand-level time-series data: *volume*, which counts the number of times a given brand (or keyword, more generally) is mentioned in various social media sources, and *valence*, which quantifies the extent to which these brand mentions are positive or negative (i.e., sentiment). Many analytics companies provide this service (e.g., Crimson Hexagon, Conversition, Cymphony, Nielsen, Radian6), which firms view as valuable because it allows them to track consumer sentiment toward their brands and products. Compared with traditional marketing research methods for tracking brands over time (e.g., surveys), social media monitoring data are observational and unobtrusive, which makes them potentially more attractive from a research perspective and also cheaper to collect.

Key issues facing managers when using social media monitoring are the large volume of data that can be generated and the lack of a systematic approach for using the data to improve managerial decision making. For example, Apple products were mentioned 601 million times in social media in 2013 (http://www.thetechstorm.com/2014/01). Unfortunately, available data are often noisy, making it difficult to easily extract meaningful marketing insights. Another challenge is that because most brand-related conversations on social media are not controllable, managers’ actions tend to be more reactive than proactive.

A possible solution to the managerial problem of too much data and too little control is to develop forward-looking methods to identify trends in social media conversation that can then form the basis of proactive approaches for managing online brand sentiment. Currently, standard use of social media monitoring data, particularly online word-of-mouth (WOM) valence, is *backward looking* in the sense that managers use it to evaluate past performance. Although it is useful for managers to know, for example, that in the last three months positive mentions of their brand decreased and negative mentions increased, it would also be useful to know *in advance* that in the next three months, they can expect increasing negative and decreasing positive mentions. In other words, it would be helpful if managers could use social media monitoring data as an early-warning system to forecast consumer sentiment toward brands with reasonable accuracy.

This article shows how to use standard, commercially available, brand-level social media monitoring time-series data to build reasonably accurate models for forecasting WOM valence. A challenge when forecasting consumer attitudes toward brands, regardless of the data source, is that brands typically do not exist alone, neither in consumers' minds nor in social media. To some degree, brands are interdependent, related, or associated in various ways. Indeed, central to the concept of brand equity is the notion that the perceived value of a given brand is shaped by associations between it and competitive alternatives (Keller 1993). Thus, consumers' perceptions of brands are likely to be formed in a relative, not absolute, sense. This is consistent with the notion that consumers store information (including brand-related information) in cognitive associative networks, in which the "nodes" in these networks contain information (e.g., brands) and the links contain associative information between the nodes (e.g., similar brands; Krishnan 1996). Accordingly, prior research has emphasized the importance of understanding how brands are related to other brands (Henderson, Iacobucci, and Calder 1998) and have argued that interbrand or interproduct associations can be represented as networks and modeled using methods from network analysis literature (Goldenberg, Oestreicher-Singer, and Reichman 2012; Henderson, Iacobucci, and Calder 2002).

Consequently, a reasonably accurate valence forecasting model for a single brand will need to incorporate information from other brands, or at least take into account associative relations between brands in the same industry or product category. Although this can be done using traditional marketing research data (e.g., tracking surveys), it is likely to be expensive because firms would need to collect time-series data for many brands instead of just for their own brand and possibly a limited number of competitors. The cost of multibrand time-series WOM valence data from social media sources is typically substantially lower, making them a potentially viable data source for building WOM valence forecasting models.

This research proposes a method for building valence forecasting models that account for associative relations between brands to obtain better forecasts of positive and negative brand mentions in social media. Our method leverages standard, commercially available social media monitoring data to represent a brand's associations with other brands as a time-varying "network" in which brands are "nodes" and the strength of "ties" between pairs of brands is a function of the incidence of coexistence by brands in the same social media environments (e.g., social networking sites, blogs, forums) at the same time.[1] We show that the inclusion of these interbrand associations improves the accuracy of valence forecasts.

Because brands do not exist in a vacuum and consumer attitudes toward brands tend to be relative and comparative, we show that our ability to forecast valence of social media mentions can be improved. Specifically, a key contribution of this research is that we propose and test six alternative approaches for representing interbrand associations (Jaccard, Dice, anti-Dice,

---

[1] Although this is an imperfect proxy for brand relatedness or "connections," we show that it is sufficient, in that including this information significantly reduces forecast error in our empirical application.

correlation, normalized correlation, and Euclidean distance) and highlight which of these approaches delivers the greatest forecast performance improvements for online WOM valence versus forecasting models that do not factor in information from other brands. We also identify conditions under which managers can use these approaches.

The rest of the article proceeds as follows: We provide a background and literature overview in the next section, after which we outline our methodological approach and data. Next, we present our results. We conclude with a discussion of our findings and implications for further research.

BACKGROUND

Prior research has examined the implications of online WOM and brand mentions in social media channels with respect to both volume (i.e., number of mentions) and valence (i.e., sentiment or positivity/negativity of mentions). The three broad themes in this literature include the performance implications of social media mentions, research on posting behaviors, and understanding of social media data from a network perspective.

The first theme centers on the dynamics of online WOM and how it affects marketing outcomes, such as sales and new product adoption (e.g., Chevalier and Mayzlin 2006; Godes and Mayzlin 2004, 2009; Liu 2006; Moe and Trusov 2011; Schmitt, Skiera, and Van den Bulte 2011; Stephen and Galak 2012; Trusov, Bucklin, and Pauwels 2009). Related research has also attempted to link online brand/firm mentions to stock market performance (e.g., Luo 2007, 2009; McAlister, Sonnier, and Shively 2012; Tirunillai and Tellis 2011). Godes and Mayzlin (2004) link discussions of new television (TV) shows to TV show ratings. Chevalier and Mayzlin

(2006) examine online book reviews with respect to volume and valence and find that valence has a significant impact on retail sales, with negative valence having a stronger effect than positive valence. Liu (2006) analyzes movie box office revenues and finds that online review volume, rather than valence, drives revenues. Also in the movie context, Chintagunta, Gopinath, and Venkataraman (2010) find that valence is a key driver of box office revenues. Finally, in their meta-analysis, De Matos and Rossi (2008) find that WOM valence is associated with consumer loyalty and satisfaction. The valence of brand mentions in social media is clearly important because research has repeatedly shown that it is a key predictor of important marketing performance outcomes. However, while research has used online sentiment to forecast future buying behavior, sales (Sonnier, McAlister, and Rutz 2011), and firms' stock market performance (Tirunillai and Tellis 2012), to the best of our knowledge, no prior studies have focused on forecasting the valence of online WOM itself.

The second theme focuses on the factors driving posting behavior (Berger and Milkman 2012; Moe and Schweidel 2012; Toubia and Stephen 2013). Berger and Milkman (2012) demonstrate that positive content is more viral than negative content. They also suggest that the relationship between emotion and social transmission is more complex than valence alone. Moe and Schweidel (2012) examine factors that influence a person's decision to rate a product. Toubia and Stephen (2013) examine the motivations of people who post to a microblogging site (e.g., Twitter) and articulate two main types of utility that motivate these users to post content: intrinsic utility and image-related utility.

The third theme explores network-related concepts to derive deeper insights into social media data. Netzer et al. (2012) use data from an online discussion forum and subject it to text-mining algorithms to build associative interbrand networks. They use these networks to show

how firms could employ this method to infer market structure. Similar to their approach, we use social media data to construct associative networks for brands. However, our research extends that of Netzer et al. in at least three ways. First, we build time-varying brand networks using co-occurrences of brands across a large number of social media sources. Second, our primary purpose lies in showing not how such networks can be inferred, but how future brand valence can be reliably forecast using such information. Third, we assess alternative methods for analyzing coaffiliation networks suggested in the literature on social networks (Borgatti and Halgin 2011). Importantly, our approach shows how marketing research insights can be improved (i.e., more accurate forecasts) without increasing firms' data requirements. Furthermore, we incorporate the notion of entropy, which is a measure of how concentrated or dispersed the "conversation" about a brand is across sources (Godes and Mayzlin 2004). We apply this concept to whether a brand's mentions in a given period are concentrated within a few *sources* (e.g., just on a particular Facebook page) or dispersed over many sources (e.g., Facebook, multiple discussion boards, blogs, and Twitter).

DATA AND METHOD

We use commercially available social media monitoring data from Nielsen's BuzzMetrics service. Our data set covers 77 consumer electronics and technology brands over 16 months from November 2009 to February 2011. Examples of brands included in the data set are Amazon.com, Apple, Motorola, and Sony. For each brand, Nielsen provided monthly counts of positive and negative brand mentions across various types of social media channels or sources.

Table 1 reports the descriptive statistics and correlations. We focus on the number of positive and negative mentions as indicators of brand sentiment that we attempt to forecast.

[INSERT TABLE 1 ABOUT HERE]

Nielsen uses proprietary algorithms to mine a large number of social media sources, from online social networks to blogs to discussion forums, for brand-related posts or mentions. The firm then analyzes the text using natural language-processing and sentiment analysis algorithms to identify which posts/mentions are predominantly positive, which are predominantly negative, and which are neutral. Thus, the raw data set provides monthly counts of positive and negative mentions by social media source for each brand.[2]

Social Media Source Data

In addition to the cross-sectional time-series data counting positive and negative mentions of brands by month, Nielsen provided data on the sources of the brand mentions in social media. For each month and each brand, we knew how many times it was mentioned across 7376 unique sources. In this context, a source is a specific social media site (indicated by a URL), such as a social networking site, a blog, or a discussion forum. The sources represented in the data range from relatively unknown blogs and forums to well-known destinations (e.g., mashable.com, endgadget.com, facebook.com, twitter.com). At the source level, there was large variance in monthly brand mentions, ranging from 0 mentions to 122,948 mentions.

An important caveat of the source data is that it does not break the brand mentions down by valence *within* each source. Thus, while we know, for each brand and each month, the number of positive and negative mentions *aggregated across all sources* and the number of mentions

[2] We did not perform sentiment analysis ourselves to determine the valence of the brand mentions but rather relied on the commercial data set from Nielsen (as managers would most likely do). This is appropriate because our focus is on using data that managers can readily access to forecast valence, not on validating the computer science and machine learning methods used to extract valence from social media monitoring data.

regardless of valence *for each source,* we do not have brand-source-level valence data. We use the source-level *volume* (but not valence) data to construct time-varying networks that describe how brands are related or similar to each other based on being mentioned in the same sources at the same time. Although this is not perfect, our network-construction method based on unvalenced source data still yields improvements in our valence forecasts. It should also be noted that the purpose of source data is to identify when brands are mentioned in the same "place" at the same time (co-occurrences), and we expect two brands to be more strongly associated if they are contemporaneously mentioned in the same source regardless of whether the mentions are of the same or different valences. By simply co-occurring, it is reasonable to assume that two brands are more closely associated with each other than two brands that do not co-occur.

Approach

We previously argued that brands and the associated consumer perceptions do not exist in a vacuum, but rather in an associative network in which the position of one brand is relative to the position of other similar and proximal brands. Given this conceptualization, our overall approach must be able to handle this implied lack of independence while taking advantage of the information contained in the relative position occupied by each brand. As such, we rely on spatial econometrics (Anselin 1988; LeSage 2008) and, in particular, network autoregressive models. While the basis of these approaches lies in the literal application of geographic spaces (e.g., zip codes, neighborhoods, traffic zones), their application has been generalized in a way that is useful in our context as well. For example, in marketing literature, Bradlow et al. (2005) present spatial models and discuss the generalization of maps to include demographic and psychographic information. Bradlow et al. (2005), LeSage (2008) as well as Van den Bulte and Wuyts (2007) agree that the notion of neighbor can be expanded to alternatively mean a peer institution or entity, where geographic distance is replaced by a measure of similarity/dissimilarity, with the former highlighting how this approach can be applied to social networks.

At its simplest, the approach we suggest follows the form

$$y = a + \rho W y + \beta X + e, \quad (1)$$

where y is a vector of our dependent variable, W is an *n*-by-*n* spatial weight matrix containing the positional information we are interested in, and ρ is the associated scalar parameter that reflects either the strength of spatial dependence or a decay in peer influence when considering social network structure and node-to-node distances (LaSage 2008). Lastly, X is the traditional vector of independent variables with the associated β coefficients, such that when spatial influence ρ equals 0, the equation becomes a standard linear regression.

As the differentiating component in this approach, the construction of W is a critical process. To arrive at the final weighting matrix, we begin with a matrix of associations/distances between entities, set the diagonal to zero such that each entity is not defined as a neighbor to itself, and then row-normalize such that each adds up to unity. The example shown in Figure 1, drawn from Bradlow et al. (2005), demonstrates this process in a simple definition of associations in which, given a specific network, neighbors are assigned a value of 1 and nonneighbors are assigned a value of 0.

[INSERT FIGURE 1 ABOUT HERE]

Again, in this example, we identify a neighbor or peer in a dichotomous way, such that an entity either is or is not a neighbor. However, Figure 1 also shows that this is not always the case, because we can infer from the graphical representation of the social network that the distances between each entity are not exactly equal (i.e., E_1 is closer or more similar to E_2 than to E_3); we can also use this information to identify neighbors and determine their impact in the final weight matrix. We next discuss the decision of which distance/similarity measures to use in the identification of neighbors/peers in the context of brands in social media.

Constructing Brand Networks from Source Data

As we alluded to previously, our intent is to remain agnostic to the final selected method to be used in the creation of our brand network (and subsequently our weighting matrix). However, our choice set of possible distance/similarity measures can be somewhat restricted by the way brand mention data on social media are collected and distributed to managers.

The source data counting brand mentions by source type and by month can be thought of as affiliation data that show how each brand is affiliated with each source (individual websites identified by unique URLs) in each month. In this context, a brand is affiliated with a source in a given month if that brand is mentioned in that source at least once in that month. Affiliation data can be represented mathematically by affiliation matrices, which are often used to summarize bipartite graphs or two-mode networks (Harary 1969). In our context, for period t we define an N-by-M affiliation matrix $\mathbf{A}_t$ to represent the affiliations between N brands and M sources. Element $a_{ikt} \geq 0$ is the number of times in period t that brand i is mentioned in source k (for i = 1 to N and k = 1 to M).

As mentioned previously, we need to infer "connections" or associations *between brands*, not between brands and sources. A measure of interbrand association must capture the extent to which online WOM overlaps across sources for each pair of brands. Following Borgatti and Halgin (2011), we can justify representing this type of association as a network tie from two perspectives. First, a brand-pair coexisting in a similar space of online discussion or conversation provides an opportunity for these brands to be at least loosely associated, if not directly compared, in the minds of consumers. Second, a brand-pair coexisting in a similar space of online discussion or conversation occurs because of some unobservable underlying relationship between the brands or unobservable characteristics that the brands have in common, which manifests in where and when consumers choose to discuss them online in social media.

To represent interbrand associations in a network form, we must convert the brand-by-source *affiliation* matrices $\mathbf{A}_t$ to brand-to-brand *association* matrices (see the graphical representation of matrix B in Figure 1). We identify six alternative approaches to summarize these data. To explain and illustrate these approaches, we borrow from previous work in sociology (Borgatti and Halgin 2011) that used information on event attendance to construct social networks of people. We use this analogy to explain the four categories described in the contingency table presented in Table 2.

[INSERT TABLE 2 ABOUT HERE]

To determine the connection between two specific attendees (identified as E_1 and E_2 in our contingency table), cell **a** contains the count of sessions and events in which both were simultaneously present, cells **b** and **c** provide counts of sessions and events in which only one member of the pair was present, and cell **d** gives a count in which both members were absent. With these categorized counts, we can construct various measures of similarity between the two attendees. For example, consider a situation in which the table only has observations in **a** and **d** (such that the pair is only seen together or not at all). In this case, we can construe the high incidence of overlapping session attendances as an indication of overlapping interests, high similarity between the pair, and high likelihood of an existing relationship.

We now turn to specific measures that can be used in network construction and their application to our *brand-by-brand* context. Note that each of these measures are for brand pairs

(i, j) in a single period of time (t). Thus, for our empirical application, each measure is computed in each month for all possible pairs of brands in the dataset.

Jaccard coefficient. Developed by its namesake in 1901, this approach takes the total count of common sources (i.e., overlaps)—denoted as a—and divides it by the total number of sources in which only one of the brands in the pair appears—denoted as b and c—plus the total number of sources in which both appear, i.e., a (Borgatti and Halgin 2011; Jaccard 1901). Put differently, for a pair of brands, their Jaccard coefficient in a given time period is the ratio of the number of times they were both mentioned in the same source(s) to the number of times one or both were mentioned. This measure captures the incidence of overlap between two brands given that the opportunity for overlap existed. Considering the contingency table (Table 2), for each period t, we can summarize the Jaccard coefficient for brands i and j as follows:

$$\text{Jaccard coefficient}_{ijt} = \frac{a}{(a+b+c)} \text{.}$$

Dice coefficient. Developed by Dice (1945), this measure puts greater weight on overlaps than on missed opportunities. Though similar to the Jaccard coefficient, the Dice coefficient differs by putting double weight on the count of overlapping occasions. The Dice coefficient captures similarity between two brands using the following formula:

$$\text{Dice coefficient}_{ijt} = \frac{2a}{2a+b+c} \text{.}$$

Anti-Dice coefficient. First used by Anderberg (1973), the anti-Dice coefficient provides a similarity measure that is also similar to the previous two calculations. It differs from the Dice coefficient in one important way, however, by instead putting double weight on the missed opportunities to overlap. Specifically, we compute the anti-Dice coefficient as follows for a given pair of brands i and j:

$$\text{anti-Dice coefficient}_{ijt} = \frac{a}{a+2(b+c)} \text{ .}$$

Some key characteristics of these three coefficients warrant discussion. First, none of these measures make use of the full information available in the contingency table shown in Table 2, as none use cell **d**, i.e., the count of sources (or websites, in this application) in which neither brand in the pair was present at a given time. Although we could argue that this incremental information has value and can assist in identifying neighboring/peer brands, it is also likely dependent on brand pairs. For example, a brand pair with excessive joint absences might indicate similarity but also might indicate a pairing of small, non-topical brands that do not elicit much online WOM. Second, each of these initial coefficients only uses a dichotomous treatment of our available data, by considering simply whether a brand had been mentioned or not in a specific source at a given time. Alternatively, we could also incorporate the volume of brand-related mentions in a specific source. Next, we discuss three additional measures that account for the extent of overlap along these lines.

Correlation. This approach uses full information on overlapping presence, opportunities for overlap, and joint absence of pairs of brands. It also preserves information (by using all available information) and is more straightforward because it is based on Pearson correlation coefficients. For this measure, each period's brand-by-source affiliation matrix $\mathbf{A}_t$ is converted to a brand-by-brand (N-by-N) association matrix $\mathbf{B}_t$ such that off-diagonal element b_{ijt} in $\mathbf{B}_t$ (for $i \neq j$ and $b_{ijt} = b_{jit}$) is the Pearson correlation coefficient computed across all M sources for brand i and j in period t:

$$\mathrm{b}_{ijt} = \frac{\sum_{k=1}^{M}(a_{ikt}-\overline{a_{it}})(a_{jkt}-\bar{a}_{jt})}{\sqrt{\sum_{k=1}^{M}(a_{ikt}-\bar{a}_{it})^2}\sqrt{\sum_{k=1}^{M}(a_{jkt}-\bar{a}_{jt})^2}} ,$$

where $\bar{a}_{it} = \frac{1}{M}\sum_{k=1}^{M} a_{ikt}$ is the mean number of mentions per source for brand *i* in period *t*. A higher correlation (b_{ijt}) between a pair of brands means a higher degree of association between them in that period due to a greater extent of overlapping mentions (or non-mentions) across sources. This accounts for being mentioned or not mentioned in the same sources at the same time *and* the volume of mentions. Importantly, this measure takes into account instances in which a brand-pair is jointly absent, which is not taken into account by the three methods introduced earlier.

Normalized correlation matrix. Another correlation-based approach to summarizing the brand-by-source matrix is a normalized correlation matrix. Normalized correlation first normalizes by source/website, such that the volume of all messages within each source adds to unity, and then generates correlations between brands as per the previously discussed coefficient. This measure accounts for the "large room effect" in which one singular large or important source forces a relationship between brands where none may exist. This may be important in our context given that the "popularity" of sources in which brands are mentioned is non-uniform. Reverting to our example of attendees at a conference, consider the relative value of the information obtained from the overlap of two people given their joint presence at a luncheon, a social mixer, and a special session with 15 other attendees. Assuming that a large proportion of conference attendees also attend both the luncheon and the mixer, the fact that our two attendees were also present is of relatively little value in determining how similar or different they truly are; their presence in these two large initial areas has little to do with how similar they are and more to do with the popularity of the environment. However, their joint presence in a small special session provides relatively more information. In this case, the small membership enables us to infer that the coexistence in this environment is more indicative of stronger similarity.

In our brand-by-brand context, this measure penalizes colocation in large “rooms”—sources—such as Facebook or Twitter or very popular blogs. In these websites, nearly all brands are mentioned, thus facilitating colocation in which no underlying relationships or similarities might exist. Instead, this measure rewards colocation in smaller and more specialized sources of WOM to determine similarity between brands.

Euclidean distance. The final measure tested here is the Euclidean distance between pairs of brands using the difference of within-source volume of messages, aggregated across all the sources with available data. We can calculate this value as

$$Euclidean\ distance_{ijt} = \left\{\sum_{a=1}^{p}(x_{iat} - x_{jat})^2\right\}^{1/2}.$$

where each value of x_i is the total number of messages about brand ***i*** in source ***a*** at time ***t***. This approach differs from those previously presented in that it is an indication of distance, or dissimilarity, rather similarity. This measure would be problematic in cases in which investigators are preoccupied with the ultimate estimated value of the ρ coefficient associated with W matrix arising from this measure. In those cases, they should use the inverse of the measure provided for consistent comparisons. However, in our study, given the focus on forecast accuracy rather than the interpretation of estimated coefficients, the reversal in meaning associated with this measure is of no consequence.

Though present in the notation, and briefly discussed in our first measure, it is important to stress that for each of these alternative measures, the consideration of brand similarity and colocation is also time specific. As such, each measure will yield an individual network time series, with brand relational data specific to each month available in our data.

We note one important caveat. Because of data restrictions, our brand association network does not necessarily imply that a brand-pair is mentioned in *precisely* the same post (e.g., when a consumer compares a Sony TV with a similar Samsung TV). Rather, it simply indicates that a pair of brands was mentioned in the same place at approximately the same time (i.e., in the same discrete period). We acknowledge that this is not a perfect measure of association between brands; however, as we show subsequently, incorporating this information into forecasting models is enough to reduce forecast error significantly. Furthermore, this is the best we can do with standard social media monitoring data provided by companies such as Nielsen and without more thorough text-mining analysis (see Netzer et al. 2012). In any case, this “same-time but not necessarily exact-same-place” situation is arguably more conservative.

FORECASTING MODELS

Model Considerations

We now develop a forecasting model that accounts for characteristics specific to our data set that are also likely to be present in many commercial social media monitoring data sets, such as multibrand, multiperiod dynamic panel data sets. First, we account for brand-level unobserved heterogeneity in our panel data set because we have multiple observations for each brand. We use a straightforward random-intercept specification with brand random effects. An alternative would be brand fixed effects; however, with 77 brands, this specification is unlikely to be as efficient as the random-effects specification. In any case, we checked both random- and fixed-effects specifications for each of our models and confirmed (with Hausman tests) that the random-effects specification is preferred in all cases.

Second, the forecasted (dependent) variables of positive and negative WOM valence are counts of the numbers of posts of each valence, and these vary over time. Thus, they are time-series count variables. We must determine whether each is a stationary or evolving time series, consistent with standard practice in multivariate time series modeling. If these variables are stationary, they should be modeled as counts (i.e., nonnegative integers); however, if evolving, they should be transformed (first-differenced), and thus they will have different distributional properties. According to an augmented Dickey–Fuller unit-root test for panel data using a variety of lags for robustness (0-, 1-, and 2-period lags), both positive and negative valence are stationary (*p*s < .001).

Third, given that the dependent variables are stationary, they should be modeled as counts. Two typical distributions used to model count data are Poisson (assuming equidispersion; i.e., mean = variance) and negative binomial (allowing for overdispersion; i.e., mean ≤ variance). From the unconditional means and variances of these two variables, they appear overdispersed. This suggests that a negative binomial model will be more appropriate; however, we tested both random-effects Poisson and random-effects negative binomial models.

Fourth, for an excessive number of observations, the dependent variables may equal zero, which would necessitate using a zero-inflated model. We estimated regular and zero-inflated random-effects Poisson and negative binomial models. Vuong tests suggested that a zero-inflated random-effects Poisson model is preferable to a regular random-effects Poisson model ($Z = 4.51$, $p < .001$). However, a regular random-effects negative binomial model is preferable to a zero-inflated version ($Z = –3.68$, $p < .001$).[3] From these comparisons, we use a random-effects

[3] For these Vuong tests, a positive (negative) test statistic indicates that a zero-inflated model is preferable (not preferable).

negative binomial model (without zero inflation) because it allows us to handle overdispersion without added model complexity due to zero inflation.

Model Specification

To use our dynamic panel of brands, we employ a negative binomial network autoregressive model with random effects. In simple terms, our model regresses the number of positive (or negative) mentions of a brand in the current period on a series of lagged predictors. Importantly, because we are interested in forecasting a brand's online WOM valence, we consider the impact of (1) a brand's prior-period valence (i.e., time autoregressive effects); (2) the prior-period valence of networked brands (i.e., network autoregressive effect), which makes use of the spatial econometric approach discussed previously (Bradlow et al. 2005; LeSage 2008); and (3) the entropy of conversation on current-period valence.

Specifically, for the random-effects negative binomial model, let y_{it} be the number of positive (or negative) mentions of brand *i* in month *t* with the following probability distribution function:

$$P(y_{it}) = \left(\frac{\emptyset}{\emptyset + \lambda}\right)^{\emptyset} \frac{\Gamma(\emptyset + y_{it})}{\Gamma(y_{it} + 1)\Gamma(\emptyset)} \left(\frac{\lambda}{\emptyset + \lambda}\right)^{yit},$$

where λ is the mean, the variance is $\lambda + \lambda^2/\phi$, and ϕ is the dispersion parameter. We use a generalized linear model to form the conditional mean of y_{it} as a function of predictors $\mathbf{X}_{it}$, such that $\log \lambda = \alpha_{\boldsymbol{i}} + \mathbf{X}_{it}\beta$. Consistent with prior research, we use gamma random effects (α_i) for brand-specific deviations from the mean, ensuring a nonnegative count (Arora, Allenby, and Ginter 1998). We perform the estimation using maximum likelihood.

By populating $\mathbf{X}_{it}$ with the time-lagged values of both positive and negative counts, we arrive at our baseline Model 1, which is a simple autoregressive prediction of future valence (with error e_{it} and brand random effect u_{oi}). Given the time granularity of the data (i.e., monthly intervals), we examine both a single lagged period and the value of adding a second lagged period. We estimate Model 1 twice—once when Y_{it} is populated with the count of positive messages for brand *i* at time *t* and once for negative messages.

$$\text{Model 1: } Y_{it} = \beta_0 + \beta_1 Pos_{it-1} + \beta_2 Pos_{it-2} + \beta_3 Neg_{it-1} + \beta_4 Neg_{it-2} + u_{0i} + e_{it}.$$

To take full advantage of the network of brands and the actual relational values captured in the measures discussed previously, we augment Model 1 by employing a network autoregressive approach. As such, Model 2 adds as additional predictors the first and second lags of valence (i.e., positive and negative counts) of all other brands in the network. However, the impact of these entries on each brand being forecast is adjusted by the strength of the relationship between the brands using the previously described brand network measures. We accomplish this by incorporating the weighting matrix $\mathbf{W_t}$ as a multiplier of the various vectors containing lagged valence information by each brand, consistent with network autoregression, general spatial econometrics models, and our previous discussion on this overall approach and methodology. Given the number of relational measures we consider in this study, six alternative weighting matrices are possible (Jaccard, Dice, and anti-Dice, correlation, normalized correlation, and Euclidean distance); as such, we estimate Model 2 separately for each alternative Y_{it} (positive or negative message counts, as Model 1) and for each variant of $\mathbf{W_t}$, resulting in 12 total models. Model 2 takes into account not only the brand's own previous valenced information but also that of other brands, given their interdependencies (thus incorporating time lags and spatial lags; see LeSage 1999).

$$\text{Model 2: } Y_{it} = \beta_0 + \beta_1 Pos_{it-1} + \beta_2 Pos_{it-2} + \beta_3 Neg_{it-1} + \beta_4 Neg_{it-2} + \beta_5 W_{t-1} Pos_{jt-1} + \beta_6 W_{t-2} Pos_{jt-2} + \beta_7 W_{t-1} Neg_{jt-1} + \beta_8 W_{t-2} Neg_{jt-2} + u_{0i} + e_{it}.$$

Finally, given the large number of sources (websites) tracked and available in our data, we can move beyond within-source WOM dynamics and consider the potential value of incorporating information arising from the across-source WOM dynamics. Essentially, we consider how the conversation about a specific brand is evolving and distributed across all visible sources and account for its possible impact on future WOM valence. Therefore, Model 3 adds lagged entropy as an additional predictor. Godes and Mayzlin (2004) introduce entropy as a construct that captures the extent to which mentions of a brand in a month are concentrated in a small number of sources or equally dispersed over a large number of sources. If all mentions are from a single source in a month, entropy equals 0. Entropy increases and approaches 1 as the number of sources with mentions increases and the overall volume of mentions becomes evenly distributed across all sources. Model 3 then captures the brand's autoregressive component, the impact of the network of brands and their interdependencies (again calculated with the six alternative relational measures and subsequent weighting matrices), and a measure of the across-source dynamics of the conversation about each brand, estimated such that Y_{it} is populated by counts of positive messages about brand *i* at time *t* and then again for negative counts.

$$\text{Model 3: } Y_{it} = \beta_0 + \beta_1 E_{it-1} + \beta_2 Pos_{it-1} + \beta_3 Pos_{it-2} + \beta_4 Neg_{it-1} + \beta_5 Neg_{it-2} + \beta_6 W_{t-1} Pos_{jt-1} + \beta_7 W_{t-2} Pos_{jt-2} + \beta_8 W_{t-1} Neg_{jt-1} + \beta_9 W_{t-2} Neg_{jt-2} + u_{0i} + e_{it}.$$

To evaluate model performance, we first examine model fit using likelihood-based metrics (–2 log-likelihood, Akaike information criterion, and Bayesian information criterion). We then examine forecasting performance using root mean square error (RMSE) and adjusted mean absolute percentage error (MAPE). We compute the unadjusted version of MAPE as

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{O_t - F_t}{O_t}\right|,$$

where O_t is the observed value in period t and F_t is the forecasted value from the model for period t; there are T periods in total. However, this is infeasible here (as is often the case) because we observe $O_t = 0$. Following Voronin and Partanen (2012), we instead use adjusted MAPE, where we replace O_t with the three-period moving average value of O_t. Doing so eliminates the division-by-zero problem, except in the case of three consecutive periods of zero values, which we never observe.

RESULTS

Recall that Model 1 is the brand-only autoregressive model, including lagged positive and negative mentions; Model 2 is the network autoregressive model that incorporates information from other brands; and Model 3 is the network autoregressive model with the addition of WOM entropy. In this section, we briefly discuss the results, focusing primarily on model comparisons and forecasting performance. Again, we focus on the six alternative weight matrices based on the six different approaches for estimating interdependencies that we outlined previously. Furthermore, we have two dependent variables—positive and negative mentions of brands. We measure forecast error using MAPE and RMSE across the 77 brands being forecast, as per the formulae described previously.

Model Fit: In Sample Performance

We compare the in-sample performance of the six approaches for both positive and negative mentions using MAPE, which we calculated using the difference between observed and forecasted responses, as described previously. Before doing so, we compare these results with a straw-man model that does not benefit from the network autoregressive approach we suggest. In our baseline (naïve) model, where the expected current value is equal to the previous observed value ($E[Y_{T1}] = Y_{T0}$), the MAPE for positive mentions is 18.13%, and the MAPE for negative mentions is 21.63%. We now describe the results using the six alternative approaches for calculating brand interdependencies. We can compare the MAPE results of the six methods for the network autoregressive model.

MAPE (Positive Valence). Focusing on positive valence, the anti-Dice method offers the lowest error (3.7%) and therefore is superior to the other methods. However, for positive valence all three discrete methods (Jaccard, Dice, anti-Dice) offer lower error than the nondiscrete methods (correlation, normalized correlation, Euclidean distance). Among the nondiscrete methods, focusing on positive valence, we find that the correlation method outperforms the other two methods in terms of error reduction (4% MAPE for correlation, 4.3% for normalized correlation, and 4.1% for Euclidean distance). These results are summarized in Figure 2.

[INSERT FIGURE 2 ABOUT HERE]

MAPE (Negative Valence). Regarding negative valence, focusing on the nondiscrete methods, we find that correlation outperforms both normalized correlation and Euclidean distance in terms of MAPE (4% for correlation, 4.2% for normalized correlation, and 4.5% for Euclidean). For the discrete methods, the Dice method has a slight edge over the Jaccard and anti-Dice methods (4.5% MAPE for Dice, 4.6% for Jaccard, and 4.7% for anti-Dice). As Figure 2 shows, the correlation method outperforms the other five methods for negative valence.

RMSE (Positive Valence). We begin by examining the results of the basic autoregressive model with one lagged period (Model 1a), a basic autoregressive model using two lagged periods (Model 1b), and the model augmented with our network autoregressive approach (Model 2). Figure 3 provides the RMSE for each model specification.

[INSERT FIGURE 3 ABOUT HERE]

As the figure 3 shows, for positive valence, the addition of a second lagged period improves model performance for the basic autoregressive model (RMSE drops from 688 to 679); however, the network regressive approach (Model 2) outperforms Models 1a and 1b regardless of association methodology used in the network construction (though there are differences in RMSE across the six methods).

Recall that of the six methods used to represent brand interdependencies, three relied on continuous data (correlation, normalized correlation, and Euclidean distance), while the other three used discrete data on co-occurrences (Jaccard, Dice, and anti-Dice coefficients). Compared with the best RMSE from Model 1 (679), among the continuous methods, the correlation method decreases RMSE to 591, the normalized correlation to 634, and Euclidean distance to 604.

The results for the three discrete methods reveal a similar pattern, but with better overall results than the three methods based on continuous data. The Jaccard method decreases RMSE to 569, the Dice method decreases it to 584, and the anti-Dice approach performs the best of all, with an RMSE of 558. Recall that the anti-Dice approach overweights instances in which brands do not overlap (i.e., their unique areas for conversation among social media sources).

RMSE (Negative Valence). Figure 4 summarizes the results of the models using negative valence as the dependent variable. As the figure shows, the addition of a second lagged period for positive and negative sentiment (moving from Model 1a to Model 1b) provides a reduction in RMSE from 460 to 400.

[INSERT FIGURE 4 ABOUT HERE]

However, the comparison of Model 1b and Model 2 and its various network variations is not as straightforward in the case of negative brand mentions. Although the addition of a network autoregressive component can improve RMSE, it is not universally better than the base autoregressive model, and it is again dependent on the methodology used to derive brand interdependencies. In this case, the versions of Model 2 using networks built on continuous data provide consistent improvements over Model 1b. For negative valence, similar to positive valence, the correlation method provides the greatest reduction, with an RMSE of 352, while the normalized correlation approach shows an RMSE of 374, and the Euclidian method comes close to Model 1b with an RMSE of 395. Conversely, for the discrete methods, the results are not consistently better than the base autoregressive model. The Dice approach is the only one with a decrease in RMSE, at 396. The Jaccard approach has a slightly high RMSE of 407, and the anti-Dice method has an RMSE of 414.

Models Including Entropy

Next, we compare the results of our network autoregressive model (Model 2) with an augmented version of the same, but we incorporate an entropy measure (Model 3) while retaining the six different network specifications for each. Figures 5 and 6 depict the comparison between each version of Model 2 and Model 3, for positive and negative mentions, respectively.

[INSERT FIGURES 5 AND 6 ABOUT HERE]

Positive Valence. As Figure 5 shows, the inclusion of the entropy measure is universally detrimental to model performance in the case of positive WOM valence, regardless of the methodology used in the construction of the brand network. For the model based on correlations, RMSE increases to 780, while it increases to 734 for normalized correlation; the Euclidean approach generates a similar, albeit smaller, increase in RMSE to 679. The models relying on discrete data for brand interdependencies do not fare much better. The Jaccard approach generates an increase in RMSE to 659, while the Dice and anti-Dice methods generate increases to 695 and 634, respectively.

Negative Valence. Conversely, the addition of the entropy measure is universally beneficial in models of negative mentions, regardless of network construction, though the extent of improvement varies. For example, the correlation-based model experiences a reduction in RMSE to 328, while the RMSE for the normalized correlation and Euclidean approaches drops to 364 and 388, respectively. Among the discrete-based methods, Jaccard generates a reduction in RMSE to 377, Dice has the largest difference from Model 2, with an RMSE of 365, and the anti-Dice approach reduces RMSE to 386.

These findings suggest that including information on brand interdependencies through association networks and network autoregression is a good strategy in general; however, the specific measures of association used need to be empirically compared, and a single measure of association between brands is unlikely to be sufficient. Instead, employing multiple measures and comparing the results of their forecasts to determine consistency is preferable. In our data, this approach suggests different strategies for forecasting positive and negative valence messages for brands. For positive messages, we observe the smallest error in the network autoregressive model using the anti-Dice approach to network construction; for negative messages, a network autoregressive model using a correlation-based network while accounting for conversation entropy is preferable.

Predictive Fit: Out-of-sample Performance

We now turn to forecasting future WOM sentiment. To test the usefulness of the models previously discussed, we employ each to forecast one, two, and three months ahead out-of-sample, first for positive WOM and then for negative WOM, while iterating across the network construction methods for the models that use a weighting matrix (Models 2 and 3). The resulting 2 × 3 × 3 × 6 design (WOM valence × models × forecast length × network construction) provides a fairly comprehensive picture of which approaches prove the most useful, depending on the sentiment and forecasting period. Web Appendix A presents the resulting RMSEs and MAPEs for positive messages, while Web Appendix B provides a graphical illustration of the MAPEs. Similarly, Web Appendix C presents the RMSEs and MAPEs for negative messages, while Web Appendix D provides the corresponding graphical illustration of the MAPEs.

As expected from our in-sample analysis, forecasting positive online WOM works better if the models do not include entropy information (i.e., the dispersion of conversation about the brand across all visible sources of conversation). However, the addition of the network autoregressive component (Model 2) dominates the simpler brand autoregressive model (Model 1) across nearly every approach to brand network construction, with consistently better predictive performance. Only the models using the network autoregressive component from a network built on normalized correlations exhibit significantly worse predictive fit. In this approach, we deemphasized connections being made from the large common sources shared between brands (e.g., Facebook, Twitter). Not only does this evidence suggest that such correction is unnecessary, but it can also be harmful when attempting to forecast positive sentiment.

For the successful models, those incorporating time and network autoregression (Model 2) using Euclidean distance for network construction provide the lowest error when forecasting one and two months out-of-sample, with MAPEs of 4.48% and 4.68%, respectively. Compared with a naive forecast in which the expected value of the forecast sentiment equals that observed in the previous period ($E[Y_{T1}] = Y_{T0}$), we find a tremendous reduction in error—the naive method provides a MAPE of 18.13% for positive messages across the same 77 brands (equivalent to a 75% reduction in error if applying Model 2 instead of forming a naive expectation). Even against the purely time autoregressive forecasting model (Model 1), we observe an improvement over the observed MAPEs of 4.77% and 5.48% for one and two months, respectively (a 6% improvement over the time-only model for one month, and a 15% improvement for two months).

When extending predictions to a third month, Model 2 with the anti-Dice network construction provides the lowest error, with a MAPE of 4.12%. This is notable because it marks a departure not only in network construction but more specifically in the type of measure being used (the anti-Dice measure is dichotomous, in that it only accounts for presence/absence of the brand in a source rather than the overall strength of that presence with a count of messages). However, if a manager wanted to create networks from a single measure, the continued use of the Euclidean networks used for the first and second forecast would not yield much additional error, with a MAPE of 4.37% for the same third-month forecast. For ease of application, the Euclidean distance–based brand network might provide the best forecasts for positive sentiment.

Regarding the forecasts for negative sentiment, we find additional value in the inclusion of information on the dispersion of conversation as captured by the entropy measure. This information, included in Model 3, is almost a universal improvement over the simpler Model 2, which includes the time and network autoregressive components. Exceptions to this rule only occur in attempts to forecast a single period (month) forward, which, when combined with our first observation on entropy, provides evidence that the dispersion of conversation has important informational content for negative sentiment beyond the immediate term.

Regarding the results of all negative sentiment forecasts, the version of Model 3 using the anti-Dice network construction clearly provides a nearly perfect single-method approach to this forecast, offering consistently low MAPEs for one (5.73%), two (6.36%), and three (6.15%)

months out. The only improvement to this suite of predictions comes from the one-month forecast of Model 2 under the Jaccard coefficient network construction, with an error of 5.72%; however, given the small difference in errors, it is likely best to treat the anti-Dice version of Model 3 as the best alternative.

Compared with a naive predictive model in which we expect the current valence to equal the last period's ($E[Y_{T1}] = Y_{T0}$), we again observe dramatic reduction in predictive error. Such a naive model produces a MAPE of 21.63% across all 77 studied brands, and the best predictive model from our stable of options provides as much as a 73.5% reduction in error. Even if we do not expect managers to use a truly naive model for predictions, the full model (Model 3) provides an 11% reduction in error compared with the simple time autoregressive model (Model 1).

Another notable observation from the results of negative sentiment forecasting involves which network construction approach yields the best results. Although the anti-Dice method provided the best and most consistent results over time, with only the Jaccard coefficient as a potential alternative, both options are dichotomous simplifications of the data available, concerned simply with the presence/absence of the brand, given a specific source and unit of time. Considering that the forecasting of negative sentiment is likely to be of higher managerial importance because of the accompanying possibility of intervention arising from this knowledge, the finding that the best model comes with the lowest associated data cost is encouraging.

DISCUSSION

In this article, we proposed and empirically validated a relatively straightforward method for using brand-level, social media monitoring time-series data that extends what managers have typically been doing with this relatively new source of data. Managers often use social media monitoring only for backward-looking "listening" purposes; if they do use it for forward-looking purposes, such as forecasting future "buzz" (or valence of that buzz), their models tend to be naive or do not account for interdependence or associations between brands in a common product category or industry. Our findings show that accounting for these types of "connections" or "linkages" between brands can substantially improve forecasting ability with respect to social media WOM valence. Overall, our findings show that it is possible for managers to build reasonably reliable models to forecast positive and negative brand social media mentions and that the accuracy of these models improves when accounting for interbrand associations.

Our research identifies six alternative approaches for modeling the linkages between brands (i.e., brand network effects). We contrast three methods that use continuous data—correlation, normalized correlation, and Euclidean distance—and three methods that use discrete data—Jaccard, Dice, and anti-Dice. While forecasting positive valence, the Euclidean method outperforms the other two methods that rely on continuous data, demonstrating improvements in the network autoregressive approach (Model 2) over just the consideration of time autoregression (Model 1). The anti-Dice method is also effective in reducing errors of positive-valenced data, showing particular improvements in the network autoregressive approach while forecasting three months out of sample. Interestingly the addition of information on WOM dispersion (Model 3) is not helpful in reducing forecast error for positive-valenced messages.

While forecasting negative valence, evidence shows that the discrete approaches to modeling brand network effects dominate their continuous alternatives. In fact, the anti-Dice approach essentially provides the best alternative construction for the forecast of negative messages regardless of forecast length (one, two, or three months out of sample) provided the network autoregressive approach is followed and information on WOM dispersion is included (Model 3). The only alternative approach that challenges this specification would be the network autoregressive (Model 2) version utilizing the Jaccard coefficient, but restricted to a single month forecast out of sample, and even then these two alternatives exist within 0.01% error of each other. Interestingly, the value of including information on WOM dispersion (Model 3) is evident when considering the forecast of negative messages.

Comparison of all six methods yields important insights, with the anti-Dice method outperforming the other five methods in the domain of negative valence while also providing the best long term forecast in the domain of positive valence. The anti-Dice method overweights missed opportunities, suggesting that it is important for managers to examine places where the brand is not co-located with competing brands, to contextualize the valenced comments.

Our focus was on the valence of WOM about brands in social media, not the volume (total count) of those mentions. As we discussed previously, extant research suggests that WOM valence is often a more important predictor of consumers' actions (e.g., purchasing) than WOM volume. Managers also likely care more about forecasting valence than volume because how they react to expected changes in positive and negative valence is likely to be different. Furthermore, valence is important in the social media monitoring context because it gives managers insights into consumers' attitudes toward brands, as expressed by consumers' opinions in social media posts. Because consumers' opinions are unlikely to be generated in a vacuum, but instead are relative to other brands and products, considering how brands are connected with each other is also important.

This study is not without limitations. First, we do not have data on the text of consumers' brand-related posts in social media sources and thus cannot validate the valence data we received from Nielsen. However, Nielsen is one of the world's leading sources of social media monitoring data and a market leader in this industry, which gives us confidence in the validity of the data.

Second, we were only able to examine the valence of the posts and not other potentially important variables. As natural language-processing and sentiment analysis methods become more sophisticated, it will be possible not only to quantify positive and negative mentions of brands but also to capture more specific variables, such as mentions by underlying topic or theme or mentions by specific type of emotion (e.g., happiness- or anger-related mentions). Managers could forecast such variables to determine, for example, whether aspects of their brand positioning are picked up consistently over time in social media posts. For example, do consumers perceive Samsung and Apple differently, depending on the emotions expressed in social media mentions? We expect that our method for forecasting WOM valence could be applied to forecasting other types of social media mentions given the appropriate data.

Third, we examined a single general product category (consumer electronics and technology). Because our focus was on proposing and testing a method and not on testing a specific theory, the generalizability of the research is less of a concern. Nevertheless, it would be fruitful to test our method on other product categories, and we encourage researchers to do so. We selected the consumer electronics industry because it is a broad category with general appeal. In the United States, the consumer electronics industry was worth approximately $91 billion in 2012 (Research and Markets 2012). A caveat is that our method is perhaps most applicable to product categories that have more general appeal and are talked about in social media more frequently.

Finally, although our method for using source information to identify which brand-pairs are associated with each other is straightforward, it is not perfect. As we mentioned previously, inferring that two brands are associated with each other does not mean that they are closely related even if they were mentioned *jointly* in social media posts at the same time and in the same place. In contrast with Netzer et al. (2012), who use online forum post data and text-mining methods, we could not identify whether brands were mentioned jointly or just in the same sources in the same month. Thus, our network is an imperfect estimate of the true underlying associative relations between brands. Nevertheless, we showed that even when these inferences are imperfect, they still improve forecast accuracy. We expect that if we used brand networks based on finer-grained co-mentions (e.g., with full-text data for each post), our forecasts would improve further. Thus, while this is a limitation of the current work, it suggests that our results are lower-bound estimates of what is possible. We hope that future research considers these and other related issues in an attempt to find valuable forward-looking uses for social media monitoring data.

REFERENCES

Anderberg, Michael (1973). *Cluster Analysis for Applications.* Academic Press.

Anselin, Luc (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Arora, Neeraj, Greg M. Allenby, and James L. Ginter (1998), "A Hierarchical Bayes Model of Primary and Secondary Demand," *Marketing Science*, 17 (1), 29-44.

Berger, Jonah and Katherine L. Milkman (2012), "What Makes Online Content Viral?" *Journal of Marketing Research*, 49 (2), 192-205.

Borgatti, Stephen P. and Daniel S. Halgin (2011), "Analyzing Affiliation Networks," in *The Sage Handbook of Social Network Analysis*, John Scott and Peter J. Carrington, eds. London: Sage Publications, 417-33.

Bradlow, Eric T., Bart Bronnenberg, Gary J. Russell, Neeraj Arora, David R. Bell, Sri Devi Duvvuri, et al. (2005), "Spatial Models in Marketing," *Marketing Letters*, 16 (3-4), 267-78.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345-54.

Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010), "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets," *Marketing Science*, 29 (5), 944-57.

De Matos, Celso A. and Carlos A.V. Rossi (2008), "Word-of-Mouth Communications in Marketing: A Meta-Analytic Review of the Antecedents and Moderators," *Journal of the Academy of Marketing Science*, 36 (4), 578-96.

Dice, Lee R. (1945). Measures of the Amount of Ecologic Association between Species. *Ecology* **26**, 297-302.

Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545-60.

——— and ——— (2009), "Firm-Created Word-of-Mouth Communication: Evidence from a Field Test," *Marketing Science*, 28 (4), 721-39.

Goldenberg, Jacob, Gal Oestreicher-Singer, and Shachar Reichman (2012), "The Quest for Content: How User-Generated Links Can Facilitate Online Exploration," *Journal of Marketing Research*, 49 (4), 452-68.

Harary, F. (1969), *Graph Theory*. Reading, MA: Addison-Wesley.

Henderson, Geraldine R., Dawn Iacobucci, and Bobby J. Calder (1998), “Brand Diagnostics: Mapping Branding Effects Using Consumer Associative Networks,” *European Journal of Operational Research*, 111 (2), 306-327.

———, ———, and ——— (2002), "Using Network Analysis to Understand Brands," in *Advances in Consumer Research*, Vol. 29, Susan M. Broniarczyk and Kent Nakamoto, eds. Valdosta, GA: Association for Consumer Research, 397-405.

Jaccard, Paul. *Etude comparative de la distribution florale dans une portion des Alpes et du Jura*. Impr. Corbaz, 1901.

Keller, Kevin Lane (1993), “Conceptualizing, Measuring, and Managing Customer-Based Brand Equity,” *Journal of Marketing*, 57 (1), 1-22.

Krishnan, H. Shanker (1996), “Characteristics of Memory Associations: A Consumer-Based Brand Equity Perspective,” *International Journal of Research in Marketing*, 13 (4), 389-405.

LeSage, James P. (1999), *The Theory and Practice of Spatial Econometrics*. Toledo, OH: University of Toledo.

——— (2008), “An Introduction to Spatial Econometrics,” *Revue d'économie Industrielle*, (123), 19-44.

Liu, Yong (2006), “Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue,” *Journal of Marketing*, 70 (3), 74-89.

Luo, Xueming (2007), “Consumer Negative Voice and Firm-Idiosyncratic Stock Returns,” *Journal of Marketing*, 71 (3), 75-88.

——— (2009), “Quantifying the Long-Term Impact of Negative Word of Mouth on Cash Flows and Stock Prices,” *Marketing Science*, 28 (1), 148-65.

McAlister, Leigh, Garrett Sonnier, and Tom Shively (2012), “The Relationship Between Online Chatter and Firm Value,” *Marketing Letters*, 23 (1), 1-12.

Moe, Wendy W. and David Schweidel (2012), “Online Product Opinions: Incidence, Evaluation, and Evolution,” *Marketing Science*, 31 (3), 372-86.

——— and Michael Trusov (2011), “The Value of Social Dynamics in Online Product Ratings Forums,” *Journal of Marketing Research*, 48 (3), 444-56.

Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), “Mine Your Own Business: Market-Structure Surveillance Through Text Mining,” *Marketing Science*, 31 (3), 521-43.

http://www.researchandmarkets.com/reports(2012), “Consumer Electronics in the United States,” Research and Markets Report, (accessed June 5, 2013), [available at www.researchandmarkets.com].

Schmitt, Philipp, Bernd Skiera, and Christophe Van den Bulte (2011), “Referral Programs and Customer Value,” *Journal of Marketing*, 75 (1), 46-59.

Sonnier, Garrett P., Leigh McAlister, and Oliver J. Rutz (2011), “A Dynamic Model of the Effect of Online Communications on Firm Sales,” *Marketing Science*, 30 (4), 702-716.

Stephen, Andrew T. and Jeff Galak (2012), “The Effects of Traditional and Social Earned Media on Sales: A Study of a Microlending Marketplace,” *Journal of Marketing Research*, 49 (5), 624-39.

Tirunillai, Seshadri and Gerard J. Tellis (2012), “Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance,” *Marketing Science*, 31 (2), 198-215.

Toubia, Olivier and Andrew T. Stephen (2013), “Intrinsic vs. Image-Related Utility in Social Media: Why Do People Contribute Content to Twitter?” *Marketing Science*, 32 (3), 368-92.

Trusov, Michael, Randolph E. Bucklin, and Koen Pauwels (2009), “Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site,” *Journal of Marketing*, 73 (5), 90-102.

Van den Bulte, Christophe, and Stefan Wuyts (2007) *Social Networks in Marketing*. MSI Relevant Knowledge Series.

Voronin, Sergey and Jarmo Partanen (2012), "A Hybrid Electricity Price Forecasting Model for the Finnish Electricity Spot Market," paper presented at the 32st Annual International Symposium on Forecasting, Boston (June 24-27).

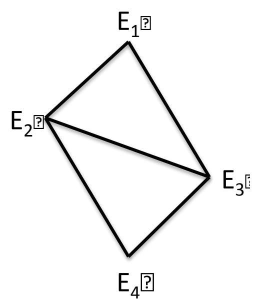
Table 1
DESCRIPTIVE STATISTICS AND CORRELATIONS

		M	SD	1	2	3	4	5	6
1	Positive mentions	1,105	2,504	-					
2	Negative mentions	380	1,131	.722	-				
3	Lagged positive mentions	1,121	2,527	.960	.679	-			
4	Lagged negative mentions	381	1,121	.683	.940	.703	-		
5	Volume	3,991,527	5,671,532	-.002	.049	-.003	.040	-	
6	Entropy	.012	.566	.056	.086	.052	.087	.172	-

Table 2
CONTINGENCY TABLE

		E_1	
		Present (1)	Absent (0)
E_2	Present (1)	a	b
	Absent (0)	c	d

Figure 1
SAMPLE NETWORK, NEIGHBOR IDENTIFICATION MATRIX B, AND WEIGHT MATRIX W



$$B = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \quad \text{and} \quad W = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix}$$

Figure 2
MODEL FIT FOR POSITIVE AND NEGATIVE VALENCE ACROSS METHODS (MAPE)

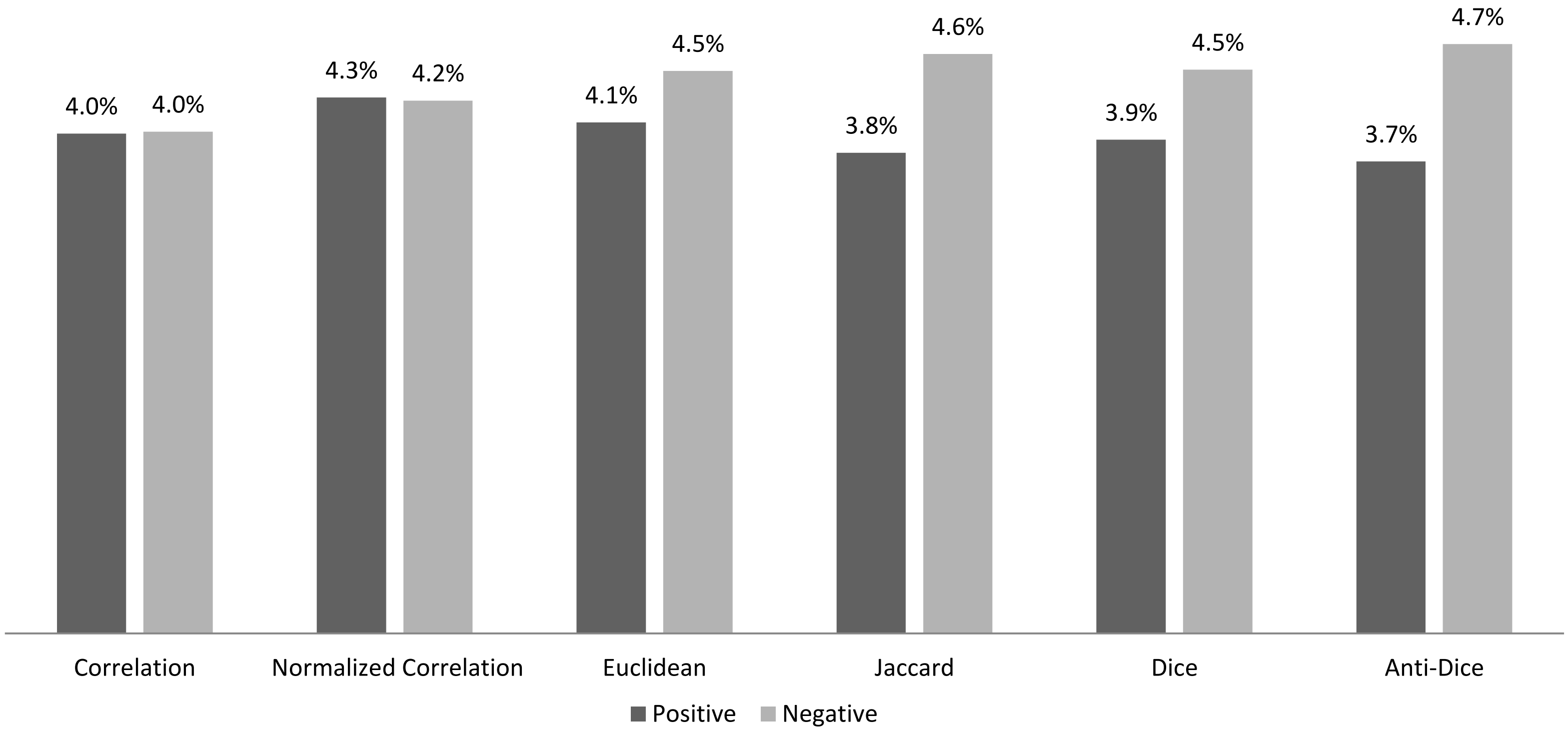


Figure 3
MODEL FIT FOR POSITIVE VALENCE: AUTOREGRESSIVE MODELS WITH ONE- AND TWO-PERIOD LAGS VERSUS NETWORK AUTOREGRESSIVE MODELS USING VARIED WEIGHT MATRICES (RMSE)

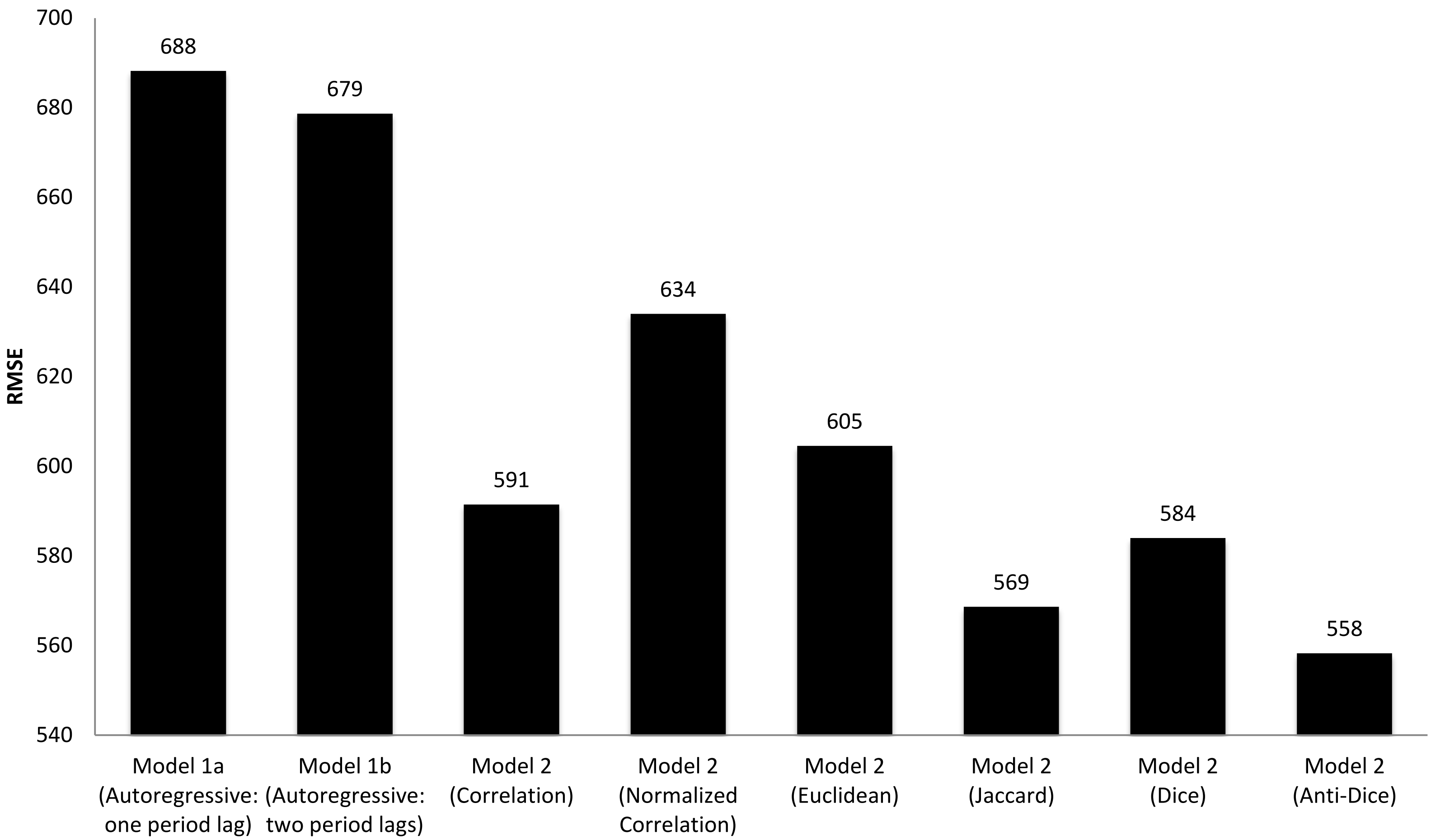


Figure 4
MODEL FIT FOR NEGATIVE VALENCE: AUTOREGRESSIVE MODELS WITH ONE- AND TWO-PERIOD LAGS VERSUS NETWORK AUTOREGRESSIVE MODELS USING VARIED WEIGHT MATRICES (RMSE)

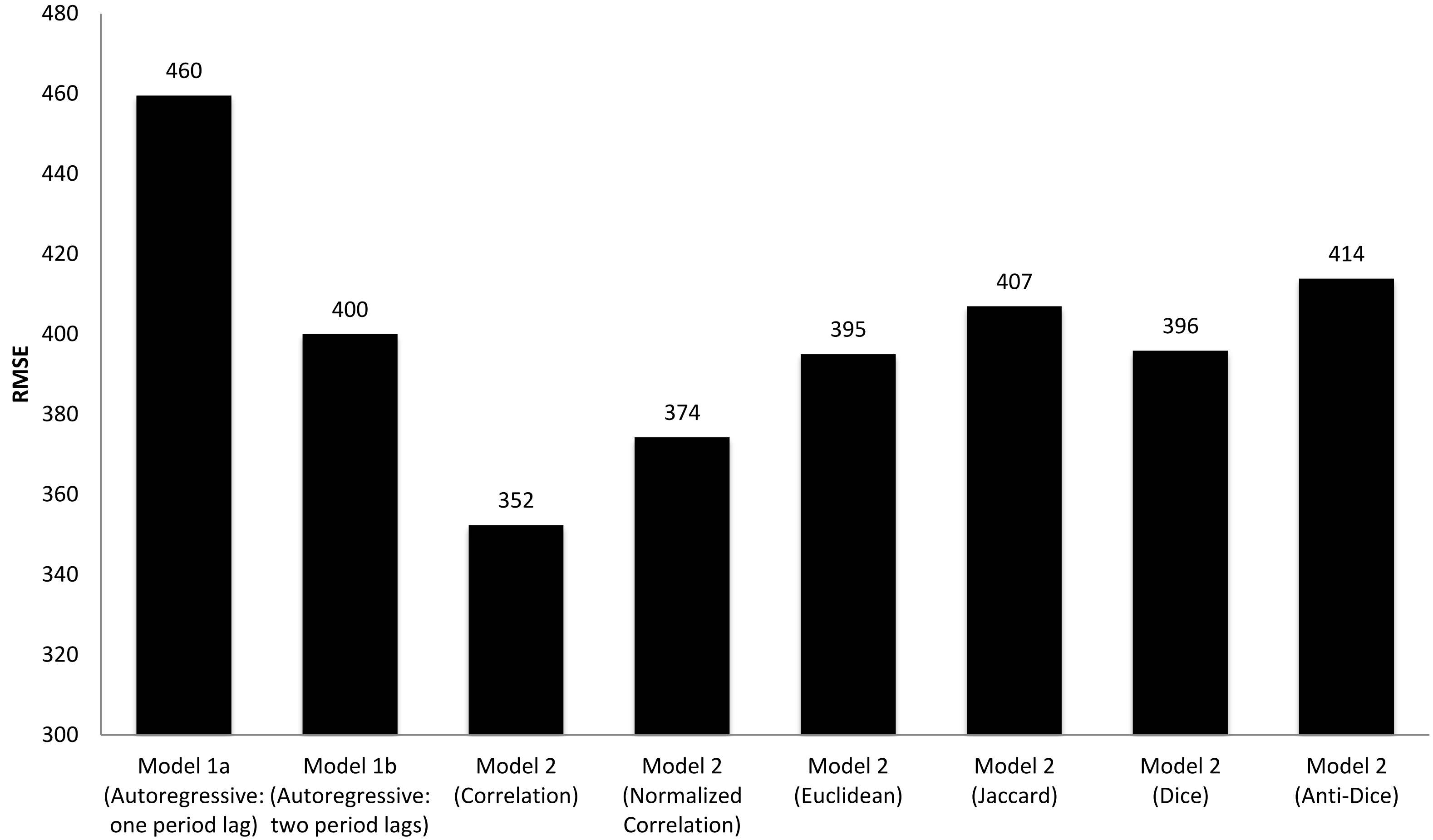


Figure 5
MODEL FIT FOR POSITIVE VALENCE: THE INCLUSION OF ENTROPY TO NETWORK AUTOREGRESSIVE MODELS (RMSE)

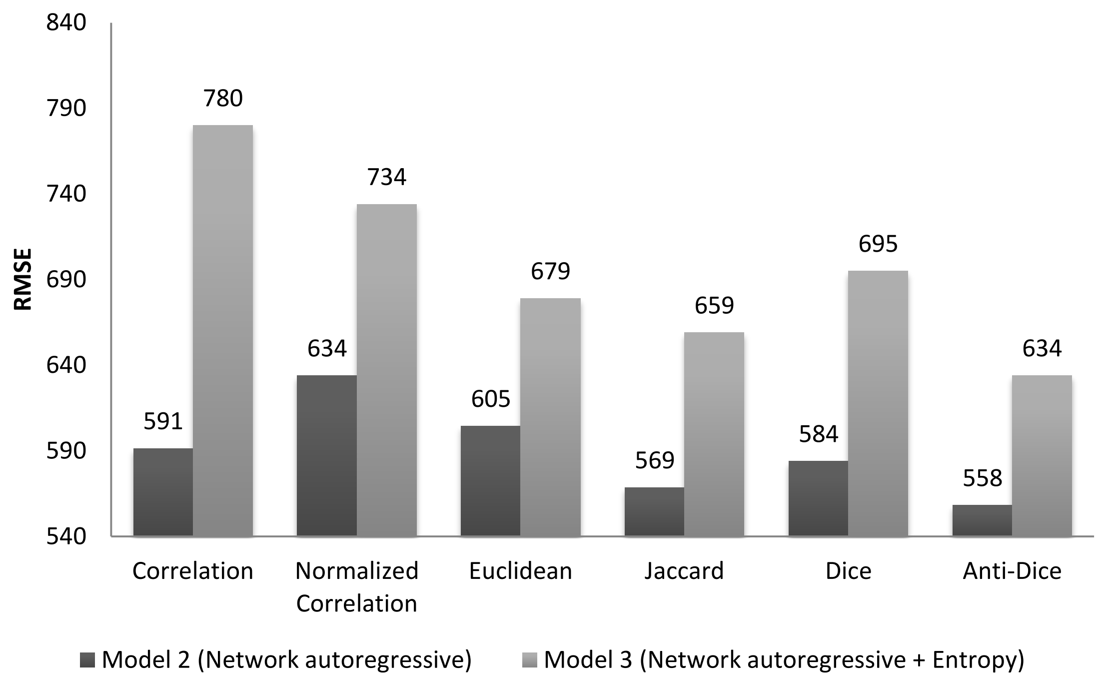
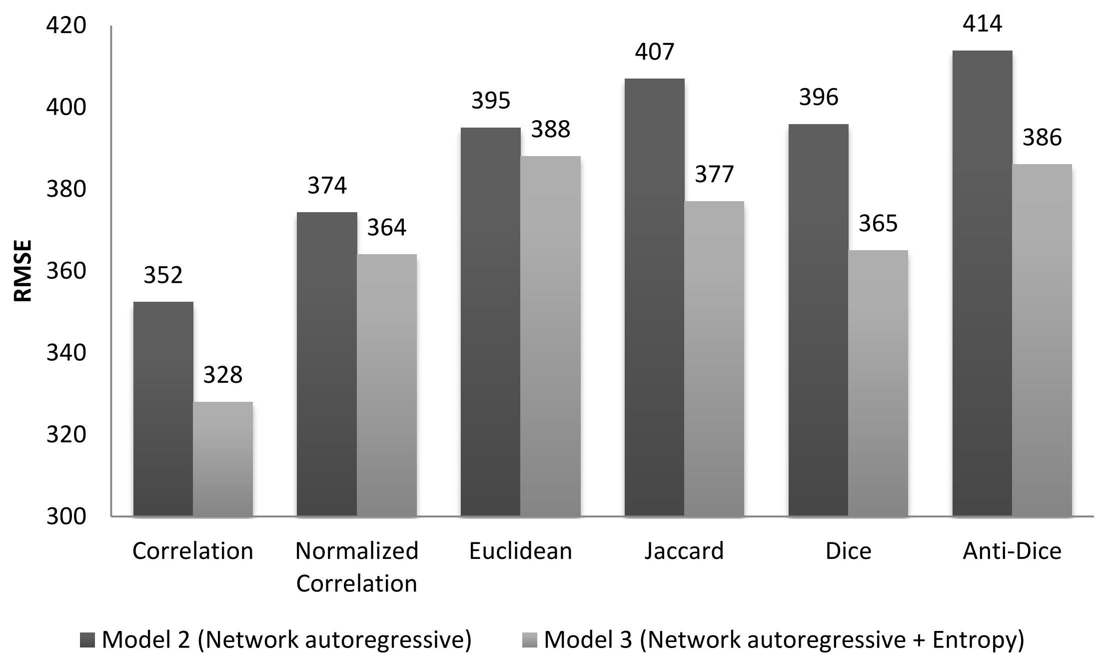


Figure 6
MODEL FIT FOR NEGATIVE VALENCE: THE INCLUSION OF ENTROPY TO NETWORK AUTOREGRESSIVE MODELS (RMSE)



Appendix A
RMSEs AND MAPEs FOR POSITIVE MESSAGES, ACROSS THREE DIFFERENT MODEL SPECIFICATIONS AND SIX ALTERNATIVE METHODS OF NETWORK CONSTRUCTION (N = 77 BRANDS)

A: RMSEs for Positive Messages Forecasted to One, Two, and Three Months

	Correlation			**Normalized Correlation**			**Euclidean**			**Jaccard**			**Dice**			**Anti-Dice**		
	Time 1	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***
Model 1	710	816	768	710	816	768	710	816	768	710	816	768	710	816	768	710	816	768
Model 2	670	757	661	1857	1342	1148	667	696	651	700	752	629	694	756	641	700	739	614
Model 3	761	768	686	1555	943	852	737	705	650	787	751	681	781	749	683	789	749	678

B: MAPEs for Positive Messages Forecasted to One, Two, and Three Months

	Correlation			**Normalized Correlation**			**Euclidean**			**Jaccard**			**Dice**			**Anti-Dice**		
	Time 1	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***
Model 1	0.0477	0.0548	0.0516	0.0477	0.0548	0.0516	0.0477	0.0548	0.0516	0.0477	0.0548	0.0516	0.0477	0.0548	0.0516	0.0477	0.0548	0.0516
Model 2	0.0450	0.0508	0.0444	0.1247	0.0901	0.0771	0.0448	0.0468	0.0437	0.0470	0.0505	0.0422	0.0466	0.0508	0.0431	0.0470	0.0496	0.0412
Model 3	0.0511	0.0515	0.0460	0.1044	0.0633	0.0572	0.0495	0.0473	0.0436	0.0529	0.0504	0.0458	0.0524	0.0503	0.0459	0.0530	0.0503	0.0455

Notes: The highlighted cells show the lowest percentage forecast error for that particular period.

Appendix C

RMSEs AND MAPEs FOR NEGATIVE MESSAGES, ACROSS THREE DIFFERENT MODEL SPECIFICATIONS AND SIX ALTERNATIVE METHODS OF NETWORK CONSTRUCTION (N = 77 BRANDS)

A: RMSEs for Negative Messages Forecasted to One, Two, and Three Months

	Correlation			**Normalized Correlation**			**Euclidean**			**Jaccard**			**Dice**			**Anti-Dice**		
	Time 1	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***
Model 1	572	617	614	572	617	614	572	617	614	572	617	614	572	617	614	572	617	614
Model 2	525	624	576	757	643	622	535	584	551	506	598	557	513	605	563	508	595	554
Model 3	528	594	565	694	611	604	544	571	547	510	568	548	522	580	554	506	562	544

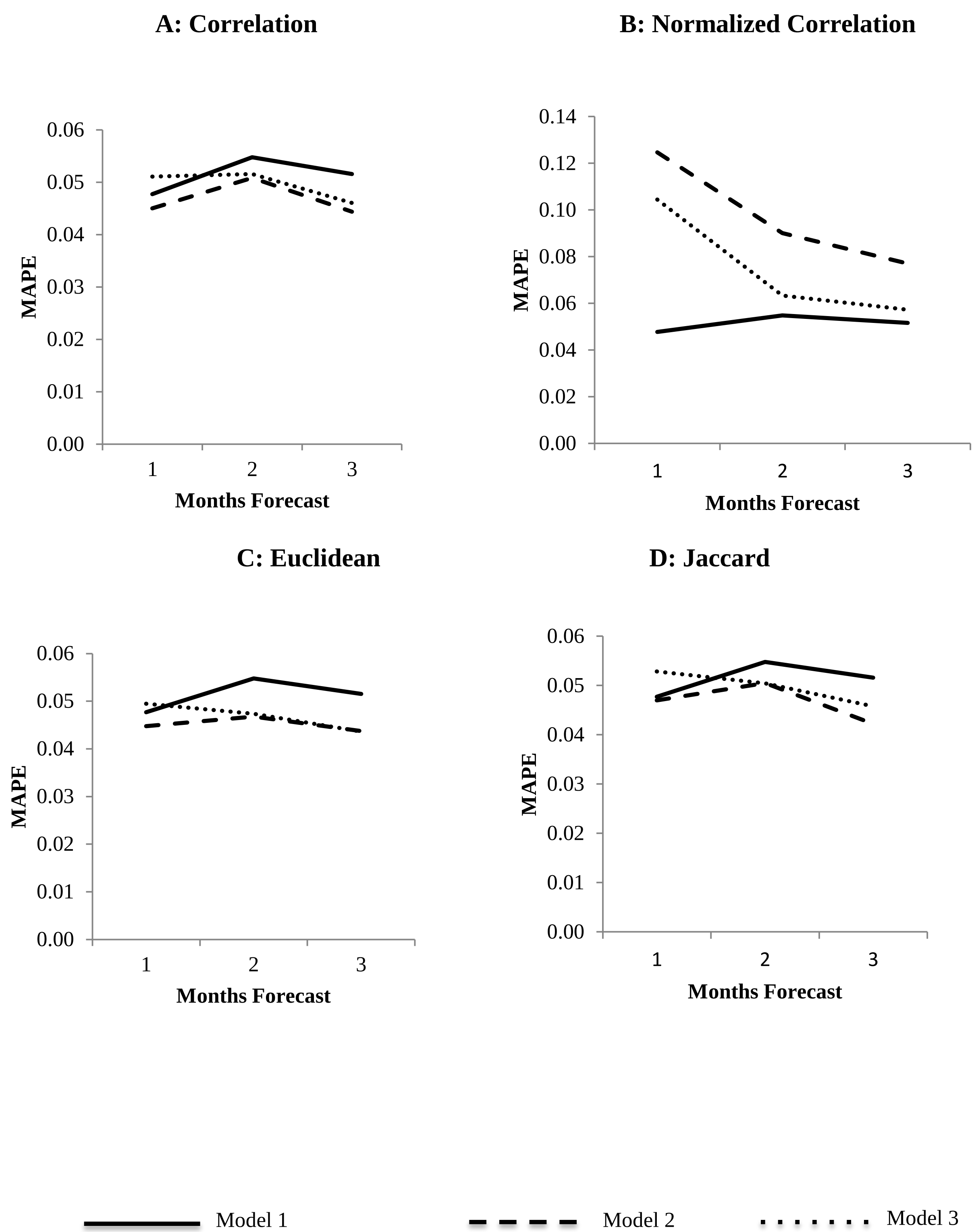
B: MAPEs for Negative Messages Forecasted to One, Two, and Three Months

	Correlation			**Normalized Correlation**			**Euclidean**			**Jaccard**			**Dice**			**Anti-Dice**		
	Time 1	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***	***Time 1***	***Time 2***	***Time 3***
Model 1	0.0647	0.0697	0.0694	0.0647	0.0697	0.0694	0.0647	0.0697	0.0694	0.0647	0.0697	0.0694	0.0647	0.0697	0.0694	0.0647	0.0697	0.0694
Model 2	0.0593	0.0706	0.0651	0.0856	0.0728	0.0703	0.0606	0.0661	0.0623	0.0572	0.0676	0.0630	0.0580	0.0684	0.0636	0.0575	0.0674	0.0627
Model 3	0.0598	0.0672	0.0639	0.0785	0.0691	0.0684	0.0616	0.0646	0.0619	0.0577	0.0643	0.0620	0.0590	0.0656	0.0627	0.0573	0.0636	0.0615

Notes: The highlighted cells show the lowest percentage forecast error for that particular period.

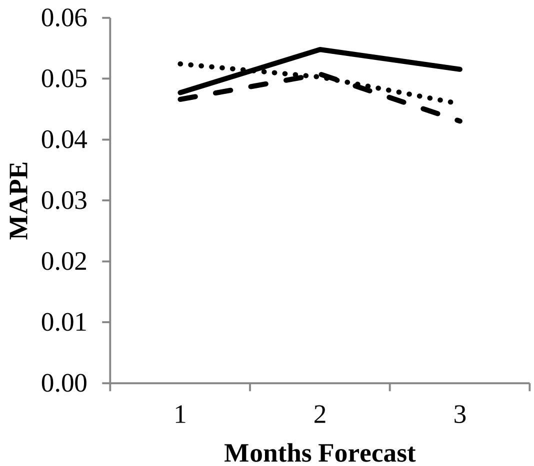
Appendix B

FORECAST ERRORS (MAPEs) FOR POSITIVE MESSAGES (N = 77 BRANDS)

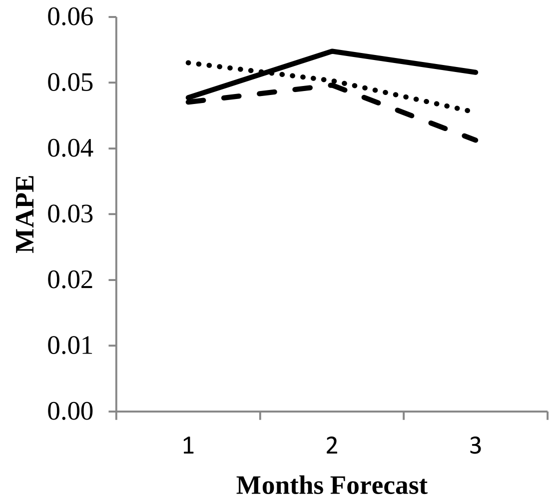


Appendix B (cont..)

E: Dice

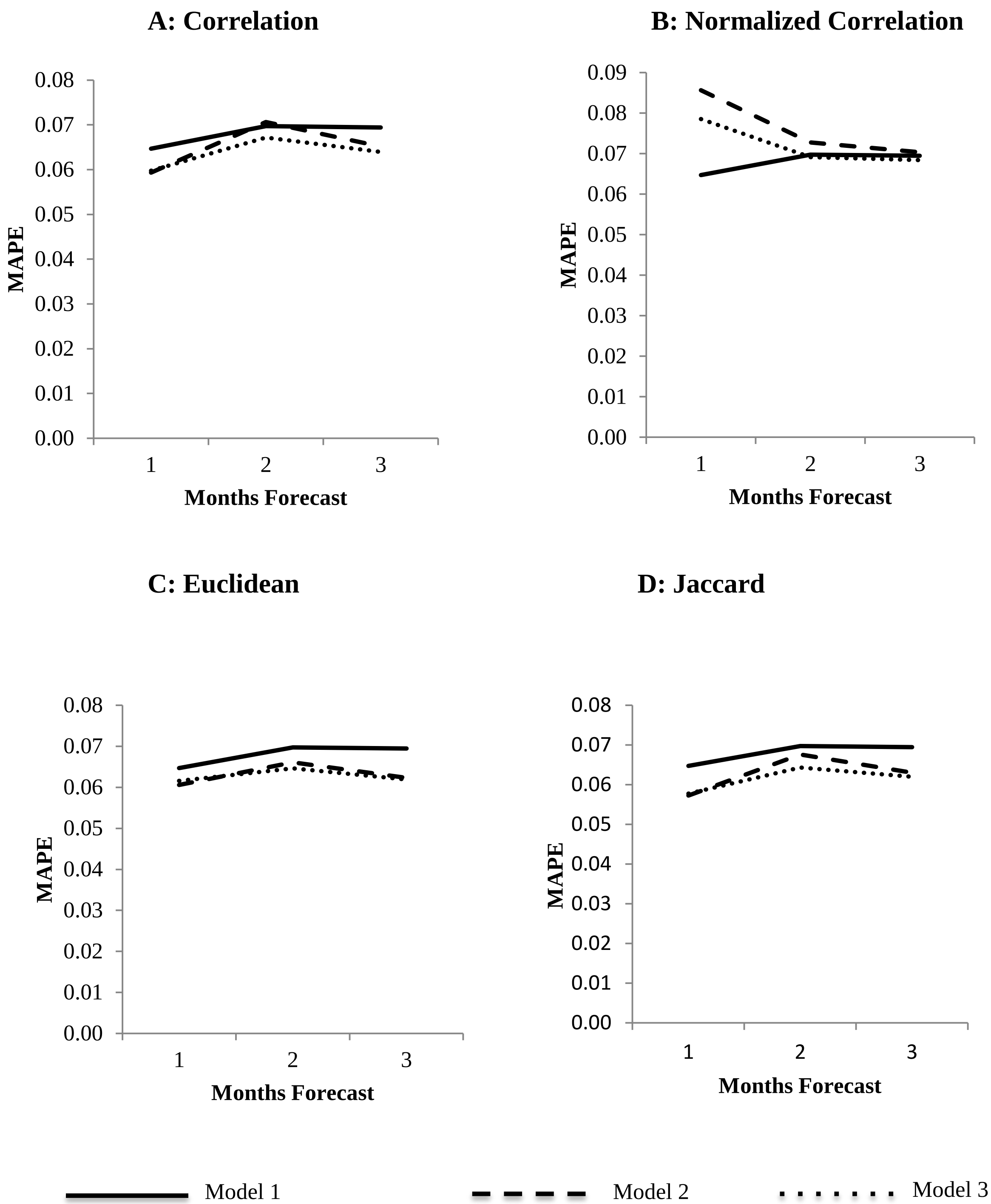


F: Anti-Dice



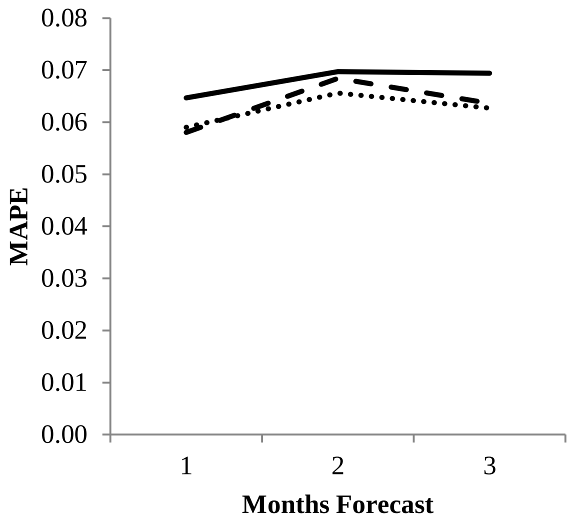
Appendix D

FORECAST ERRORS (MAPEs) FOR NEGATIVE MESSAGES (N = 77 BRANDS)



Appendix D (cont..)

E: Dice



F: Anti-Dice