Is Distance Really Dead in the Online World? The Moderating Role of Geographical Distance on the Effectiveness of Electronic Word of Mouth

Vilma Todri, Panagiotis (Panos) Adamopoulos, and Michelle Andrews

Abstract
The authors investigate how the geographical distance between online users is associated with electronic word-of-mouth (eWOM) effectiveness. Their research leverages variation in the visibility of eWOM messages on the social media platform of Twitter to address the issue of correlated user behaviors and preferences. The study shows that the likelihood that followers who are exposed to users’ WOM subsequently make purchases increases with followers’ geographic proximity to the users. The authors propose social identification as a potential mechanism for why geographical distance still matters online in eWOM: because consumers may form a sense of social identity based on their physical location, information regarding the spatial proximity of users could trigger online social identification with others. The findings are robust to alternative methods and specifications, such as further controlling for latent user homophily by incorporating user characteristics and embeddings based on advanced machine-learning and deep-learning models and a corpus of 140 million messages. The authors also rule out several alternative explanations. The findings have important implications for platform design, content curation, and seeding and targeting strategies.

Keywords
electronic word of mouth, social media, geographical distance, deep learning

Electronic word of mouth (eWOM) plays a key role in shaping consumer attitudes and behaviors. More consumers digitally share information, research what others say about products and services, and rely on eWOM to gain knowledge or make decisions than ever before (Bloem 2017; Murphy 2018). While eWOM is becoming one of the most influential sources of information among consumers, the effectiveness of traditional communication channels employed by marketers is declining (Shaban 2019; Todri 2021; Todri, Ghose, and Singh 2020). These contrasting trajectories have motivated marketers to harness the power of eWOM. To do so, they are paying more attention to their user base and the corresponding eWOM episodes (Adamopoulos and Todri 2014, 2015a; Godes and Mayzlin 2009). At the same time, technological advancements and data opportunities in the online space enable marketers to access data such as user-generated content and user location (Adamopoulos, Todri, and Ghose 2020; Wedel and Kannan 2016). As the availability of location data in particular grows, marketers wonder whether location plays an important role in eWOM effectiveness. Some marketers believe that consumers may consider eWOM from farther distances as more widely accepted and universal and, thus, more valuable and informative (eMarketer 2019). Others believe that eWOM from closer distances may be perceived as more relatable and relevant and, thus, more trustworthy (Backaler 2018; Sharma 2018). Yet other marketers suppose that distance does not matter online and, thus, does not influence how consumers value eWOM (Quinn 2018).

The literature on the role of geographical distance in eWOM effectiveness is not decisive either. On the one hand, a common belief for the effectiveness of eWOM is that technology bridges the geographical distance between consumers. Information technology reduces communication costs
by decoupling the interaction process from geographic constraints. For instance, mobile devices enable shopping and communication between consumers independent from location, while the internet allows instant access to message exchanges and marketplaces. These reduced communication and access costs have led some scholars to pronounce the “death of distance” (Cairncross 2001) and “end of geography” (Graham 1998).

On the other hand, geographical distance may still play a role in online consumer behavior and WOM for several reasons (Goldfarb and Tucker 2019). These include location-specific goods as well as economic costs related to shipping, contracting, monitoring, enforcement, travel, and inconvenience (Adamopoulos, Ghose, and Tuzhilin 2021; Hortaçsu, Martínez-Jerez, and Douglas 2009). Local user preferences and spatially correlated social ties as well as limited regional availability of alternative product options can also lead to commonalities in product adoption among nearby consumers (e.g., Bell and Song 2007; Ma, Krishnan, and Montgomery 2014; Meyners et al. 2017). These reasons suggest that geography may still constrain the effectiveness of eWOM.

Our research investigates whether geographical distance is significantly associated with the effect of eWOM, beyond the aforementioned explanations. Specifically, we ask whether the geographical distance between pairs of familiar disseminators and receivers of online WOM messages in social media plays a role in driving recipients’ purchase behaviors or whether distance does not matter for eWOM beyond utilitarian reasons—such as transaction costs—and proxying for correlated user behaviors and preferences. We therefore examine this research question in an online setting where economic costs, such as those relating to contracting and travel, are not of concern.

To investigate whether geographical distance is associated with the effectiveness of eWOM, we use rich data from the social media platform of Twitter, where users disseminated messages about their real-world products (Dodri and Adamopoulos 2014). Our main identification strategy leverages variation in the visibility of these eWOM messages—enabled by a unique feature of the platform—causing some purchase messages to be visible and others to be invisible to followers (Adamopoulos, Ghose, and Todri 2018). We compare the purchase behaviors of followers for whom users’ purchase messages were included in their newsfeeds (i.e., the stream of tweets presented on the home screen of a user from accounts the user follows on Twitter) with the behaviors of followers for whom users’ purchase messages were not included in their newsfeeds due to this design feature, while employing an extensive set of controls (discussed in the “Econometric Model Identification” subsection). We also use the salience of user locations on Twitter to provide evidence for the potential mechanism. In our empirical setting, users are familiar with each other and thus aware of the location of their peers. Their location is often salient, as users may highlight this information in their profiles; user profile information is observable in peers’ timelines when hovering with the cursor over the profile picture, the username, or the person’s name.

We find that the relationship between eWOM and the likelihood that message recipients make a purchase strengthens as the geographical distance between disseminators and receivers decreases. This relationship is economically significant, managerially relevant, and robust to alternative methods and identification strategies. Social identity may explain why geographic proximity could increase the effectiveness of eWOM, beyond utilitarian reasons and proxying for correlated user behaviors, as consumers may form their social identities based in part on their geographic location (Forman, Ghose, and Wiesenberg 2008; Newman 1972; Twigger-Ross, Bonaitiu, and Breakwell 2003; Uzzell, Pol, and Badenas 2002). To provide evidence for this potential mechanism, we investigate, for instance, how the salience of geographic cues as well as conditions that strengthen the role of geographical location in the social identification process (Taylor, Gottfredson, and Brower 1985; Vezzali et al. 2015) further enhance the effectiveness of eWOM.

Our findings of geographical distance variation in eWOM effectiveness have important implications for both theory and practice. We contribute to the online WOM influence studies by demonstrating that despite technology’s promise, spatial distance continues to play an important role in a digital world (Ameri, Honka, and Xie 2019; Lovett and Staelin 2016). Specifically, we show how geographic proximity is significantly related to an increase in eWOM effectiveness. Geographic constraints thus tether the impact of electronic communications. This finding also contributes to the literature that shows how features of disseminators and receivers can drive eWOM outcomes (Baker, Donthu, and Kumar 2016; Godes and Mayzlin 2009) by identifying geographic proximity as a relational characteristic that can affect dyadic influence. Our work thus provides actionable strategies for boosting the effectiveness of online interpersonal communications. Our findings also imply that customizing seeding and targeting with more proximate connections or highlighting information and cues associated with social identity formation can increase the effectiveness of online advertising, product recommendations, social advertising, referral programs, and other marketing strategies and tools by leveraging local appeals and social identification.

Related Literature

Several studies demonstrate the importance of eWOM, documenting how it can be a major driver of consumer behaviors. For example, user opinions have been found to influence the consumption of movies (Chintagunta, Gopinath, and Venkatakrishnan 2010; Liu 2006), television shows (Godes and Mayzlin 2004; Lovett and Staelin 2016), video games (Zhu and Zhang 2010), and books (Chevalier and Mayzlin 2006; Li and Hitt 2008). Whereas these studies examine the direct impact of eWOM on purchase decisions, the present research extends this line of inquiry by focusing on conditions under which the effect of online WOM may be attenuated or accentuated.
Contextual Factors Affecting eWOM

Research has begun to investigate the contextual factors that impact the effectiveness of eWOM. These factors include characteristics of the product (Berger and Schwartz 2011), brand (Lovett, Peres, and Shachar 2013), or WOM message itself (Packard and Berger 2017). Scholars have also examined how certain features of disseminators (senders) and recipients shape the effect of eWOM. For instance, a sender’s brand loyalty (Godes and Mayzlin 2009), expertise (Aral and Nicolaides 2017), and identity disclosure (Forman, Ghose, and Wiesenfeld 2008) can each affect the influence of WOM messages. Recipient characteristics such as the number of ties to adopters, demographics (Katona, Zubecek, and Sarvary 2011), and product experience (Park et al. 2018) can also affect the influence of online communications. In addition, characteristics of the sender–recipient dyad, such as the tie strength (Aral and Walker 2014; Baker, Donthu, and Kumar 2016), similarity across personality traits (Adamopoulos, Ghose, and Todri 2018), and sociodemographic similarity (Fossen, Andrews, and Schweidel 2017), can also affect online WOM performance. We add to this stream of research by shedding light on an important factor that characterizes the pairwise relationship between senders and receivers of eWOM: the geographical distance between them.

The Impact of Geographical Distance in Various Commercial Contexts

Geographical distance has been shown to affect certain outcomes in multiple online and offline commerce settings. For example, geographical distance can affect online trade flows and volume due to costs related to shipping, contracting, monitoring, enforcement, travel, and inconvenience, as well as in cases of location-specific goods (Hortaçsu, Martinez-Jerez, and Douglas 2009). Such transaction costs can engender a “home bias” that leads consumers to prefer transacting with nearby others (Agrawal, Catalini, and Goldfarb 2015; Hortaçsu, Martinez-Jerez, and Douglas 2009; Lin and Viswanathan 2015). Similarly, geographical distance has been shown to correlate with product diffusion in the offline world due to imitation or direct observation, as physically close neighbors are more likely to adopt the same product (Bell and Song 2007; Choi, Hui, and Bell 2010; Forman, Ghose, and Goldfarb 2009). In this study, however, we investigate how the geographic proximity between eWOM message disseminators and receivers accentuates or attenuates the effectiveness of online WOM, whereas the extant explanations for the impact of geography—albeit in different online settings—are not applicable to the context of eWOM; similarly, other explanations that do not apply to our eWOM setting relate to the location specificity of products, distribution networks, or local network externalities.

Geographical Distance as a Potential Factor Affecting Online Persuasion

Recent studies hint at the potential role geographical distance may play in facilitating persuasion. For instance, lab studies find that when consumers have no identifying information about online reviewers, they assume that the reviewers are similar to them and so are as persuaded by them as they are by reviewers who appear similar to them, and more persuaded than by reviewers who appear dissimilar to them (Naylor, Lamberton, and Norton 2011). When ambiguous reviewers appear less similar, consumers are also less persuaded by their boastful reviews (Packard, Gershoff, and Wooten 2016). One of the many ways these studies manipulate cues to indicate reviewer similarity is to show that the reviewer appears to live in the same or a nearby city as lab participants. In instances of familiar peers, rather than ambiguous reviewers, field studies find that consumers are more persuaded the more recent or intense their relationships are with their peers (Aral and Walker 2014; Chen, Van Der Lans, and Phan 2017). One of the ways Aral and Walker (2014) measure relationship recency is to use as a proxy whether peers currently live in the same town. Whereas the aforementioned studies focus on how ambiguity about, trust in, or relationship with peers affects persuasion in settings without financial transactions or product purchases, our research focuses on whether the geographical distance between senders and receivers of online WOM messages is significantly related to eWOM effectiveness.

Prior Research on How Geographical Distance Impacts eWOM Effectiveness

Most relevant to our work are two studies examining whether geographical distance affects information diffusion and adoption. Specifically, Fossen, Andrews, and Schweidel (2017) examine how geographic proximity, measured as contiguous relationships between states, and sociodemographic proximity, measured as similarity in demographics between states, affect overall message propagation at the state level. They find that geographic proximity has no impact when sociodemographic similarity is accounted for. Because of data limitations, their operationalization of geographic proximity may mask the actual effect of distance. Importantly, they conduct an aggregate state-level analysis, noting that the “state-level nature of the data is a limitation” (p. 249) and “access to more disaggregated data would allow for a more granular analysis” (p. 264). Their analysis also prevents them from observing whether consumers are...
familiar with the message propagator or the social distance between peers (Manski 1993); familiarity and similarity between users as well as relationship strength have been shown to impact peer influence (Aral and Walker 2014; Naylor, Lamberton, and Norton 2011). Besides, diffusion may rely on a different mechanism than eWOM (Stephen and Lehmann 2016).

Meyners et al. (2017) is also very relevant to our work. These authors first conduct a descriptive study in the offline world and find that living nearby adopters of a cellular service provider is associated with faster switching to that provider. Because they do not directly observe WOM episodes, they conduct their analysis at an aggregate level, where it is not possible to account for homophily (i.e., the tendency of individuals to choose friends with similar tastes and preferences) or attribute WOM influence among the ties of each user (Manski 1993, 1999). In addition, they measure geographical distance as the average distance among all possible ties of each consumer. These limitations motivate them to conduct scenario-based experiments, in which they investigate the mediating role of perceived homophily and show that geographic proximity to an ambiguous online reviewer increases the likelihood of following the reviewer’s opinions, as geographic proximity is used as a cue for perceived similarity. However, the effects of geography are likely to be different in such lab settings due to the lack of familiarity and social ties with online reviewers, which play an important role in persuasion, as the extant literature shows (Aral and Walker 2014; Naylor, Lamberton, and Norton 2011; Packard, Gershoff, and Wooten 2016). In addition, while consumers may rely on any available cues to address these concerns of ambiguity and lack of familiarity, lab settings often provide few such cues beyond geography to inform their decisions (Naylor, Lamberton, and Norton 2011). Thus, it remains unclear whether geographic proximity can still play a role in facilitating eWOM influence in real-world settings, where users are familiar with and socially connected to each other and have access to an abundance of available cues.

In summary, we empirically investigate whether the geographical distance between individual pairs of familiar disseminators and receivers of organic eWOM in social media plays an important role in driving recipients’ subsequent purchase behaviors of physical goods. Table 1 outlines the related studies and additional important differences of our research.

### Empirical Setting and Data Description

Our empirical setting concerns a large-scale venture of American Express (the service provider) on the microblogging social media platform Twitter. This collaboration introduced a new purchasing service that was seamlessly integrated for two months into the social media platform (Twitter) to leverage users’ connections to stimulate eWOM. Specifically, the service enabled consumers to make purchases on the social media platform while simultaneously spreading the word about their purchases to their social media peers (receivers).

We further discuss this novel service by describing the data-generating process. In particular, the service provider first posts a short message (tweet) on the platform broadcasting the list of participating merchants and the corresponding products available for purchase. This announcement includes information about the product offerings (e.g., product, respective sale price) and the designated hashtags (i.e., a phrase or word preceded by a hash sign [#]) consumers must use to make a purchase. Consumers must have a microblogging account and sync their service provider account with their microblogging account. Once the service provider broadcasts the products available for purchase, users can purchase these products by posting a tweet that includes the designated hashtag. In addition to the necessary hashtag, consumers can choose to personalize the purchasing tweets that are (automatically) shared with their social media peers. Typically, such messages are posted on the users’ social media profiles, and their followers (i.e., those subscribed to their timeline) automatically receive these messages on their own (home) newsfeeds (for an explanation of how we use the variation in message visibility in our main identification strategy, see the “Econometric Model Identification” subsection and Figure A1 in the Web Appendix). The social commerce provider tracks the tweets containing the designated hashtag and pairs them with the responding product. After the purchase confirmation, the service provider bills the users and ships the product.

### Empirical Data

Our data set includes all the confirmed transactions generated through this purchasing process. Specifically, each transaction in our data set contains the message ID of the purchasing message, the message content, the designated product hashtag, the date and time the message was posted, the corresponding user ID, and whether the message was rendered visible or invisible to each of the user’s followers (i.e., included or not included in the follower’s newsfeed).

For each user, our data set also contains the screen name of the user on the social media platform, when they joined the platform, the set of the user’s followers and followees, all the messages users have posted on the platform since they joined, the geolocation of the user, and the self-reported description of the user’s profile. Our data set also contains the same information for users who chose not to make a purchase.

In addition, our data set contains information about the product offerings. The service provider collaborated with known retailers and offered various products for purchase (e.g., see Figure A2 in the Web Appendix). In particular, the products involve video game consoles and related accessories, electronics and sports equipment (e.g., high-definition
<table>
<thead>
<tr>
<th>Study (Chronological Order)</th>
<th>Research Focus</th>
<th>Context</th>
<th>Setting</th>
<th>WOM Type</th>
<th>Social Ties</th>
<th>Familiarity with Sender</th>
<th>Geographical Distance Measure</th>
<th>Analysis Level</th>
<th>Notes</th>
<th>Main Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naylor, Lamberton, and Norton (2011)</td>
<td>How ambiguous reviewers affect persuasion</td>
<td>Scenario-based</td>
<td>Online reviews</td>
<td>Indirect; active; artificial</td>
<td>None</td>
<td>No</td>
<td>Binary (same vs. different city)</td>
<td>Choice instance</td>
<td>Same vs. different city domicile is one of many reviewer variables manipulated to match that of a lab participant.</td>
<td>Consumers assume that reviewers with no identifying information have similar tastes as them, so they are similarly persuaded as they are by reviewers who appear similar to them.</td>
</tr>
<tr>
<td>Aral and Walker (2014)</td>
<td>How tie strength and embeddedness affect app adoption</td>
<td>Social media platform (Facebook)</td>
<td>Indirect; passive (automated notifications without variation)</td>
<td>Observed</td>
<td>Familiarity with the subject (not sender)</td>
<td>Binary (same vs. different town)</td>
<td>Notification episode</td>
<td>The authors note that they &quot;estimate how relationship-level covariates [same town] are correlated with the extent or impact of influence&quot; (p. 1363), and the &quot;results may have limited generalizability to … cases where there is a significant financial cost to adopting a product&quot; (p. 1366).</td>
<td>Consumers will adopt a free Facebook (movie review) app more quickly the more recent their relationship is and the more embedded they are with an existing user.</td>
<td></td>
</tr>
<tr>
<td>Packard, Gershoff, and Wooten (2016)</td>
<td>How boastful WOM affects persuasion</td>
<td>Scenario-based</td>
<td>Online reviews</td>
<td>Indirect; active; artificial</td>
<td>None</td>
<td>No</td>
<td>Binary (nearby vs. distant location)</td>
<td>Choice instance</td>
<td>Nearby vs. distant location is one of several variables used to manipulate how similar the reviewer appears to a lab participant.</td>
<td>Consumers are less persuaded by boosters’ WOM when trust cues are low.</td>
</tr>
<tr>
<td>Chen, Van der Lans, and Phan (2017)</td>
<td>How relationship characteristics in a social network affect diffusion</td>
<td>Microfinance program adoption</td>
<td>Potential offline WOM episode</td>
<td>Unobserved</td>
<td>Surveys</td>
<td>Yes</td>
<td>None</td>
<td>Aggregate</td>
<td>WOM occurs in geographically bounded networks (e.g., village, university), so geographical distance is not material.</td>
<td>Social (economical) relationships are the most (least) important drivers of adoption for a microfinance program. Relationship intensity (e.g., message volume) is the most important driver of information diffusion in a network.</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Study (Chronological Order)</th>
<th>Research Focus</th>
<th>Context</th>
<th>Setting</th>
<th>WOM Type</th>
<th>Social Ties</th>
<th>Familiarity with Sender</th>
<th>Geographical Distance Measure</th>
<th>Analysis Level</th>
<th>Notes</th>
<th>Main Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fossen, Andrews, and Schweidel (2017)</td>
<td>How social vs. geographic proximity affect diffusion</td>
<td>Message propagation</td>
<td>Social media platform (Twitter)</td>
<td>Direct; active</td>
<td>Unobserved</td>
<td>Unobserved</td>
<td>Binary (contiguous vs. noncontiguous state)</td>
<td>Aggregate</td>
<td>The authors note that “the state-level nature of the data is a limitation” (p. 249). The study does not control for advertising or other marketing activities and focuses on diffusion. Sociodemographic similarity propagates online message spread. Geographic proximity has no effect when accounting for sociodemographic similarity.</td>
<td></td>
</tr>
<tr>
<td>Meyners et al. (2017)</td>
<td>How geographic proximity affects adoption</td>
<td>Cellular service provider adoption</td>
<td>Potential offline WOM episode</td>
<td>Unobserved</td>
<td>Observed (aggregate)</td>
<td>Yes (aggregate)</td>
<td>Average distance from all peer adopters</td>
<td>Aggregate</td>
<td>The authors note their field study “did not have information on either the valence of the signals from the social network or the location of nonadopters” (p. 63). Averages distance across adopters, does not observe WOM, does not control for network externalities, and studies switch behavior. The lab studies have no homophily cues beyond age and gender, and no familiarity with or social ties to reviewers. Geographic proximity to adopters of a cellular provider is associated with faster switching to the provider. Geographic proximity to an ambiguous reviewer increases the likelihood of following the reviewer’s recommendation due to perceived homophily.</td>
<td></td>
</tr>
<tr>
<td>Present study</td>
<td>How geographical distance affects eWOM effectiveness</td>
<td>Product purchases of significant financial cost</td>
<td>Social media platform (Twitter)</td>
<td>Direct; active</td>
<td>Observed</td>
<td>Yes</td>
<td>Continuous (miles), categorical (state)</td>
<td>Choice instance</td>
<td>Our study has information on nonadopters and social ties; controls for homophily, advertising, and message content; and studies actual product purchases of significant financial cost in a social network where receivers are familiar with the WOM sender. Geographical distance is negatively associated with the likelihood that eWOM influences purchases, above and beyond other effects.</td>
<td></td>
</tr>
</tbody>
</table>
tablets, sports and action cameras), general-purpose gift cards, and fashion accessories (e.g., designer bracelets, handbags). These offerings from the social commerce service provider were available for purchase at a reduced price only through the platform (about a 25% discount, yielding an average retail price of $125).

Overall, our data set tracks the corresponding purchasing decisions of 132,995 social media users on Twitter with 1.4% of the users purchasing available products. The users are located across the continental United States, as illustrated in Figure 1, and on average follow 996 Twitter users and have 342 followers. Table 2 presents the summary statistics and description of the main variables and Figure 2 shows the corresponding correlations.

We enhanced our data set with a proprietary data set from the ad intelligence company Kantar Media, which includes the local (and national) advertising expenditures of each brand and for each product. We also supplemented our data set with additional information from the American Community Survey five-year estimates of the U.S. Census Bureau regarding local demographics.

**Empirical Methodology**

We model users’ purchase decisions as a function of eWOM message, sender, and recipient as well as relationship characteristics, including observed and latent homophily controls; homophily creates a natural correlation in behaviors that could be incorrectly interpreted as a causal effect (e.g., Aral, Muchnik, and Sundararajan 2009; Manski 1999). To further control for any unobserved confounds, we use in our research design the variation in the visibility of eWOM messages.

**Econometric Model Specification**

Consistent with prior literature (e.g., Meyners et al. 2017), we use a continuous-time single-failure survival model. In particular, we model how quickly users purchase a product, if any, using a Cox (1972) proportional hazard model and correcting for censoring of transactions that might have been intended to occur after the observation window (Kalbfleisch and Prentice 2011). Specifically, our main estimation equation for the decision of peer i (eWOM recipient) is as follows:

$$
\lambda_i(t) = \lambda_0(t) \exp(\text{Visible message}_i \beta^M + \text{Geographical distance}_ij \beta^D + \text{Visible message}_i \times \text{Geographical distance}_ij \beta^{DM} + X_{ij} \beta^X + M_i \beta^M + D_j \beta^{DS} + R_i \beta^R),
$$

where $\lambda_i(t)$ is the hazard of peer (follower) i of consumer j to...
purchase the same product as consumer j after 3 an eWOM message from j, \( \lambda_0(t) \) represents the baseline hazard, \( \text{Visible message}_j \) captures whether the eWOM message of consumer j (sender) was rendered visible to (i.e., included in the home newsfeed of) peer i (recipient), and \( \text{Geographical distance}_{ij} \) measures the physical distance of i from consumer j in (log) miles. The coefficient of interest is \( \beta_{DM} \) and captures the relationship between geographical distance and the effectiveness of eWOM, while the coefficient \( \beta_D \) accounts for spatially and nonspatially correlated user behaviors and preferences (e.g., homophily) that would have otherwise manifested as the association between geographical distance and the effectiveness of WOM; put simply, our research design leverages the variation in the visibility of eWOM messages, as influence can occur if and only if the eWOM message is rendered visible, whereas correlated user behaviors and preferences are extant even when the eWOM message is nonvisible. We also control for user-relationship (sender–recipient dyad) characteristics, \( X \); message characteristics, \( M \); disseminator characteristics, \( D \); and recipient characteristics, \( R \); as well as product fixed effects and geography (state, time fixed effects).

Model features and machine-learning methods. To construct the aforementioned user-relationship, message, and disseminator and recipient controls, we employed machine-learning techniques to leverage the vast amount of unstructured and structured data.

More specifically, the user-relationship (sender–recipient dyad) characteristics, \( X \), include controls for observed and latent pairwise user similarity and tie strength. This set of variables—in addition to the research design—allows us to capture the correlation in latent tastes and preferences between WOM message disseminators and recipients to better distinguish the relationship of interest. In particular, the user similarity between the disseminator and the recipient of a WOM message is measured based on (1) the similarity of topics discussed in social media posts using the results of a machine-learning model for natural language processing (NLP), as well as the overlap of the local communities as captured by the (2) Jaccard similarity coefficient of followers (Culotta and Cutler 2016) and (3) Jaccard similarity coefficient of followees (Culotta and Cutler 2016).

The NLP model we employ for this task is the latent Dirichlet allocation (LDA) model (Blei, Ng, and Jordan 2003).

### Table 2. Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean/Median</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purchase</td>
<td>Whether the recipient of the message made a purchase</td>
<td>.014</td>
<td>.12</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Visible message</td>
<td>Whether the message was visible to the recipient</td>
<td>.77</td>
<td>.42</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Geographical distance</td>
<td>Geographical distance between sender and recipient</td>
<td>971.08</td>
<td>894.90</td>
<td>0</td>
<td>5,585</td>
</tr>
<tr>
<td>Number of followers</td>
<td>Number of followers of user</td>
<td>342</td>
<td>101,000</td>
<td>0</td>
<td>376,000</td>
</tr>
<tr>
<td>Number of followees</td>
<td>Number of followees of user</td>
<td>996</td>
<td>12,629</td>
<td>0</td>
<td>115,000</td>
</tr>
<tr>
<td>Reciprocal relationship</td>
<td>Whether the relationship between the users is reciprocal</td>
<td>.08</td>
<td>.27</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Number of interactions</td>
<td>Number of interactions between users</td>
<td>.26</td>
<td>6.10</td>
<td>0</td>
<td>1,612</td>
</tr>
<tr>
<td>Sentiment of message</td>
<td>Intensity of message advocacy</td>
<td>.21</td>
<td>.35</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Personalized message</td>
<td>Whether the message was personalized by the sender</td>
<td>.82</td>
<td>.38</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

*We report the median instead of the mean value.

Note: The values of the variables Number of followers, Number of followees, Reciprocal relationship, and Number of interactions correspond to the time of posting the WOM message.
LDA is a probabilistic generative NLP model that we use to model the user-generated content of each user in our data set (a document in our corpus) as a distribution over topics and every topic as a distribution over words in the English dictionary. We build this model on the corpus of all the messages of the users in our data set as the topics users discuss online reflect their latent interests (Weng et al. 2010); using the complete corpus instead of only the user messages during the observation window improves the inference of the NLP model. In particular, for the implementation of the LDA model, we used 139,850,033 messages in total. We also use a part-of-speech tagger/tokenizer developed specifically for Twitter (Owoputi et al. 2013) for more accurate tokenization, the removal of stopwords, symbols, typos and uninterpretable words (Son et al. 2019), and the creation of n-grams, while retaining online-specific textual features (i.e., hashtags, at-mentions, and emoticons) (Sammut and Webb 2017; Wang, McCallum, and Wei 2007). We also use Hoffman, Bach, and Blei’s (2010) online variational Bayes algorithm to efficiently estimate the LDA model on our corpus. In addition, instead of arbitrarily determining the value of the LDA parameter corresponding to the number of latent topics, we find the natural number of topics present in our corpus using the procedure and measure proposed by Arun et al. (2010), evaluated in terms of the Kullback–Leibler (1951) divergence measure; nonetheless, the findings are not sensitive to the number of topics. Finally, regarding the hyper-parameters of our model, we learn an asymmetric prior directly from our data. Beyond capturing the disseminator–recipient similarity with these metrics, we alternatively measure the similarity as a single standardized factor, based on the principal factors method, to avoid any potential multicollinearity; we present both sets of results.

In addition, our model specifications include constructs capturing whether the user relationship is reciprocal and (the log of) the number of interactions between the two users (Chen, Van Der Lans, and Phan 2017). Finally, we also control for the sociodemographic distance of the users—based on the difference in average age and percentages of male, Black or African American, Hispanic, and Asian-origin residents in the locations of the disseminator and recipient of the message using Census data at the zip code level (Fossen, Andrews, and Schweidel 2017)—as well as the time zone difference. Overall, the various metrics of user pair similarity capture both observed similarity (e.g., the number of user interactions) and latent similarity (e.g., common latent interests) to further control for potentially unobserved confounds and homophily.

The message characteristics, \( \mathbf{M} \), capture the sentiment of the message (i.e., intensity of WOM advocacy) as well as whether the message was personalized (i.e., explicit rather than implicit advocacy). The sentiment of the message (measured on a continuous scale between \(-1 \) and \(+1\)) provides a richer metric of the advocacy intensity of the sender compared with other simple metrics, such as lexicon-based scores. The main method we employed uses a publicly available commercial sentiment analysis mechanism based on deep learning (API Harmony 2021). Nonetheless, the results are robust to employing alternative machine learning methods for sentiment analysis.\(^4\) Moreover, the message controls also include whether a user account handle is mentioned in the message and whether the sender started the message with a period as the first character; starting a message with a period as the first character affects the visibility of the message and is a common norm among Twitter users when they want to explicitly make a message visible to all users and not only the account mentioned in the message.\(^5\) Finally, we also control for advertising expenditures of each brand during our observation period in the local region of the eWOM message recipient expressed in (logarithm of thousands of) U.S. dollars.

The disseminator and recipient characteristics, \( \mathbf{D} \) and \( \mathbf{R} \), include the user (opinion) leadership and expertise measures following the extant literature. We capture these disseminator and recipient characteristics to further control for factors that could bias the true relationship between geographical distance and WOM effectiveness due to omitted individual characteristics and preferences. The user expertise levels are measured on the basis of the standardized similarity of the timeline of a user with the corresponding timelines of the participating vendor and product employing the probabilistic NLP machine-learning model we previously described (Bhattacharya et al. 2014; Mottazi and Naumann 2013). The motivation for this measure is, for instance, that users who frequently tweet about technological trends and topics similar to those in the social media accounts of the specific product and the corresponding vendor are more likely to be perceived by their social media peers as experts in the area of technological products (Ito et al. 2015). In addition, we further control for correlated user preferences and interests (Blum and Goldfarb 2006; Ma, Krishnan, and Montgomery 2014) by including in the econometric specifications the latent interests based on the aforementioned LDA model (Adamopoulos, Ghose, and Todri 2018). The user (opinion) leadership is measured based on the additive smoothed ratio (Manning, Raghavan, and Schütze 2008) of followers to followees of the user. The additive smoothed ratio is frequently used in empirical studies to prevent this metric from being oversensitive to small-scale changes in the numbers of followees or followers (Adamopoulos, Ghose, and Todri 2018; De Veirman, Cauberghe, and Hudders 2017). Furthermore, we also control for the number of followers of each user, whether the user has a default profile on the platform, and how many months have elapsed since the user joined the social media platform. Finally, we also control for the age, gender, and income of the users based on the Census data (Fossen, Andrews, and Schweidel 2017).

\(^4\) For instance, the results are robust to supervised methods, according to which we first assign sentiment scores to a small number of messages and then build machine-learning models for predicting the scores for the rest of the messages. The scores estimated through different methods are similar, and the findings remain robust.

Econometric Model Identification

We next discuss the research design we use to further distinguish the effect of geographical distance on eWOM effectiveness from unobserved confounds and correlated user behaviors and preferences, such as homophily. Our research design is enabled by a unique feature of the platform that causes some WOM messages to be visible and other messages to be invisible (not included) on the newsfeeds of other peers, as described next.

Typically, a message posted by a user on the Twitter platform appears in the newsfeeds of all her followers, as the newsfeeds were not algorithmically curated during this venture. Thus, in our context, whenever a user makes a purchase, her social media peers are exposed to her advocacy toward the product if the purchase is visible in their (home) newsfeeds and their purchasing decisions might, in turn, be affected through such WOM episodes; the specific offers were available for purchase only through Twitter. The research design framework we employ leverages information on whether a message broadcasted on the platform was indeed rendered visible in the newsfeeds of other users or not (see Web Appendix Figure A1).

More specifically, on Twitter, a publicly broadcasted message may not be included in the newsfeeds of some followers in instances when a Twitter account handle (if any) appears in a specific position of the message (i.e., if any Twitter account user name following the “@” symbol appeared at the very beginning of the message). For instance, if a consumer purchases a product by posting the message “#BuyActionCamPack @AmericanExpress” (see Web Appendix Figure A2, Panel A), the message is visible on the newsfeeds of all her followers. However, if the same consumer purchases the same product by posting the message “@AmericanExpress #BuyActionCamPack” (see Web Appendix Figure A2, Panel B), then the message is visible only to the social media users who follow both this consumer and the mentioned account.

Interestingly, due to the design of Twitter during our observation window, if a consumer initialized the purchase process by clicking on the announcements of the product offerings or a post of another consumer, the handle of the corresponding account (e.g., “@AmericanExpress”) is automatically prepopulated in the purchasing message. Then, the consumer must click anywhere on the prepopulated field and manually write the purchasing message that includes the required designated hashtag. Notably, if the click happens to be recorded on the left of the prepopulated account handle, the cursor will be placed at the beginning of the field and the consumer can write the purchasing message starting after the mentioned account handle (see Web Appendix Figure A3, Panel B). Then, if the consumer wants to complete the transaction, she is likely to write a message similar to the one in Web Appendix Figure A2, Panel B, starting immediately with an account handle, and the tweet will not be included in the newsfeeds of her peers (recipients) who do not follow the mentioned account.

Thus, the visibility of each WOM message to the peers of a consumer depends on (1) whether an account is mentioned (e.g., @AmericanExpress or another account), (2) the exact position of the account handle (whether it is at the very beginning of the message or not), (3) what social media accounts the consumer’s peers follow, (4) whether a user click will be recorded more to the left or more to the right of the screen, and (5) whether an account handle will be prepopulated by the platform. Thus, the visibility of the messages is affected by platform design factors (e.g., prepopulating accounts) rather than certain user characteristics exploited by a curation algorithm. Importantly, with regard to the visibility of eWOM messages, the recipient does not control the visibility of these messages from the users she follows; this is important because the purchase decisions we study are the recipients’ decisions.

Because our main identification strategy assumes that the visibility of eWOM messages and the geographical distance of consumers are not endogenous in this study, we also empirically investigate whether the visible and nonvisible message groups differ across the pretreatment variables in our model based on the normalized differences tests (Imbens and Rubin 2015) that provide scale-invariant measures of the size of the differences. The normalized differences range from −.1768 to .1692, indicating that there are no significant differences between the two groups in observable characteristics, further enhancing the validity of the identification strategy; differences of .25 or less indicate a good balance between the two groups (Imbens and Rubin 2015). Similarly, kernel distributions, quantile-quantile plots, and the orthogonality test (Imbens and Rubin 2015) further confirm the validity. We also examine the geographical distance in the same way, reaching the same conclusion.

Overall, these unique features of Twitter induce an important variation in the visibility of messages, enabling us to examine the behaviors of peers in treatment (i.e., visible message) and comparison (i.e., nonvisible message) groups in a potential outcomes framework. Thus, differences in purchases between treatment and control groups can be attributed to the corresponding WOM messages and their characteristics, addressing the issue of correlated user behaviors and preferences. In this respect, our research is also relevant to the stream of work that has leveraged the variation in the visibility of advertising messages to estimate the effect of online ads (Adamopoulos, Ghose, and Todri 2018; Ghose, Singh, and Todri 2017; Ghose and Todri-Adamopoulos 2016).

We further enhance our identification strategy using only observations corresponding to dyadic relationships and social media peers who did not receive messages (either visible or invisible) from multiple disseminators (Adamopoulos, Ghose, and Todri 2018; Aral and Walker 2012). In addition to taking
advantage of this nonintrusive research design, we also avoid any observer biases; the subjects are unaware of being part of the study and, thus, do not alter their behavior in anticipation of the study. Nevertheless, we also control for differences in the pairwise relationships between users by employing an extensive set of variables and fixed effects in our data-rich setting. Table A1 in the Web Appendix presents additional identification strategies.

**Empirical Results**

**Main Results**

Table 3 presents the results of the main econometric specifications of the eWOM effectiveness model. In particular, Model 1 constitutes our baseline specification as it models eWOM effectiveness based on the constructs of dyadic similarity and relationship strength between the disseminator and recipient of the eWOM message (i.e., pairwise user similarity, reciprocity of users’ relationship, and number of user interactions), eWOM message advocacy (i.e., personalized message and sentiment of message), and user characteristics (i.e., expertise and leadership). Then, Model 2 introduces the notion of geographical distance, and Model 3 adds the information of eWOM message visibility and leverages the corresponding variation to further distinguish the relationship between geographical distance and eWOM effectiveness (βPM) from correlated behaviors and homophily among users. That is, Model 3 disentangles the relationship between geographical distance and eWOM effectiveness from the correlational effect of geographical distance reported by Model 2; the eWOM influence is transmitted through visible messages only, whereas correlation in user behaviors and preferences (i.e., homophily) is present even with invisible messages. Then, Model 4 further controls for whether a user account was mentioned in the eWOM message and if the message was explicitly made visible while introducing additional disseminator and recipient controls (i.e., number of followers, leadership, default profile, time on the platform, interests, age, gender, and income) and other controls (i.e., the difference in average age and percentage of male, Black or African American, Hispanic, and Asian-origin residents in the locations of the disseminator and recipient of the message, the log of brand advertising expenditures in USD (in 1,000s) in the location of the recipient of the message, and the time zone difference between the locations of the disseminator and recipient of the message). Finally, Model 5 controls for the specific product mentioned in the eWOM message as well as state and day fixed effects.

According to the results presented in Table 3, we find that the coefficient of the variable capturing the relationship between geographical distance and WOM effectiveness is negative and statistically significant (Visible message × Geographical distance: .945, p < .01, Model 5); all reported coefficients correspond to hazard ratios (HRs) representing the increase (HR > 1) or decrease (HR < 1) in purchase likelihood associated with each attribute. (Table A2 in the Web Appendix also shows the coefficients of the control variables.) Beyond the effect of interest, we also find a negative and statistically significant spatial homophily–based effect of geographical distance (Geographical distance: .954) as well as a positive and statistically significant effect of similarity (homophily). Interestingly, these findings show that despite the “death of distance” postulated in the literature (Cairncross 1997), geographical distance is negatively associated with the effectiveness of eWOM. Thus, our research is the first to establish that geographical distance has a negative relationship with eWOM outcomes even among familiar social media peers, in addition to the previously known homophily-based effect of geographical distance.

Moreover, the coefficients of all the other variables are in accordance with what one would expect as well as the extant literature on WOM (e.g., Aral and Walker 2014; Baker, Donthu, and Kumar 2016; Fossen, Andrews, and Schweidel 2017). Specifically, we find that the increased user similarity and strength of relationship between users (User similarity: 1.264; Reciprocal relationship: 4.527) as well as more intense WOM advocacy (Sentiment of message: 1.556; Personalized message: 1.041) are associated with higher levels of purchase likelihood after exposure to WOM (Visible message: 1.608). Similarly, users with higher product expertise (User expertise: 1.191) seem more persuasive; thus, their followers are associated with a higher purchase likelihood after being exposed to their advocacy.

**Economic Significance and Managerial Relevance**

The relationship of interest is also of economic significance, as a decrease of 10 miles in the distance between users accentuates the effectiveness of eWOM by 12.78% based on the aforementioned coefficients; similarly, an increase of 100 miles in the distance corresponds to a decrease of 25.56%, and an increase of 1,000 miles corresponds to a decrease of 38.34%. For instance, for a recipient living in New York City, the relationship is reduced by 24.39% when the eWOM message originates from a sender in Philadelphia, and 38.82% from Miami, compared with the same message from a sender in New York City.

We further assess the economic significance of the findings by measuring the out-of-sample performance of the models. Specifically, we use a holdout evaluation scheme with an 80/20 random split of data and evaluate the models in terms of Harrell’s C concordance coefficient, which measures the likelihood of correctly ordering survival times for pairs of senders and recipients of eWOM messages; the concordance measure is similar to the Mann–Whitney–Wilcoxon test statistic as well as the area under the receiver operating characteristic curve. The results show that our model achieves a predictive performance of .840. Thus, it outperforms the baseline by a large margin, as the baseline performance corresponds to a value of .5. This statistically significant difference further illustrates the managerial relevance of the findings, as they can enhance seeding and targeting strategies (Hinz et al. 2011; Sun et al. 2021).

We further quantify the (dollar) value of this increase in out-of-sample performance (Provost and Fawcett 2013). To
conduct this calculation (Provost and Fawcett 2013), we use estimates of the cost of targeting (e.g., promoting eWOM messages) and the average product price (Goldfarb and Tucker 2011); the cost of this type of targeting on Twitter is estimated to be $1.35 based on WebFX (2020), while the average product price in our data set is $125. Combining these data reveals that our model suggests a profit of $.85 per targeted user, which corresponds to a 9% increase over the baseline of not using the information of geographical distance (i.e., $.78), while for random targeting the corresponding profit is only $.008.

### Potential Mechanism

Our findings are surprising, as in such an empirical setting the products are not location-specific, there are no transportation fees for consumers, and there is no contracting or potential conflict or ambiguity between senders and recipients of WOM messages (Hortaçsu, Martínez-Jerez, and Douglas 2009; Meyners et al. 2017). Thus, to fully understand our findings, we delve into a likely underlying mechanism of the identified effect and conduct additional analyses that allow us to assess the likelihood of this potential mechanism.

We hypothesize that the negative relationship between geographical distance and eWOM effectiveness could be due to the identification processes of social media users. Specifically, a user who resides near the sender of the message is likely to share a common social identity with the sender based on their geographic proximity (Forman, Ghose, and Wiesenfeld 2008; Jacobs and Munis 2020; Newman 1972; Twigger-Ross, Bonaiuto, and Breakwell 2003) and thus might be more susceptible to WOM influence originating from this (local) sender. Conversely, a recipient who resides farther away from the

---

6 The social identity theory holds that an individual’s social identity is formed by the perception of belongingness to a community (Ashmore, Deaux, and McLaughlin-Volpe 2004; Tajfel 1974). Common social identity (between the message sender and recipient) can be used as a heuristic to make judgments and guide actions (Chaiken 1987; Chaiken and Maheswaran 1994). Thus, social identity (group identification) has a powerful influence on human behavior (e.g., Atefi et al. 2018; Jenkins 2014; Shang, Reed, and Croson 2008).
disseminator of the WOM message is not likely to share the same location-based social identity and thus is less likely to be persuaded (Cialdini 2016; Coleman, Deutsch, and Marcus 2014; Smith and Hogg 2008) by mere exposure to eWOM advocacy.

**Location-based social identity activation.** To empirically assess this potential underlying mechanism, we first examine the moderating effect of the salience of the geographical distance from the source of WOM. Salience is activating common social identity identification (Forman, Ghose, and Wiesenfeld 2008; Heere et al. 2011; Hinds and Mortensen 2005; Lin and Viswanathan 2015; McGarty et al. 1994; Turner 1982), and thus, we empirically test the likely underlying mechanism by examining the moderating effect of the salience of the sender’s location on the impact of geographical distance on the effectiveness of WOM. If the relationship is accentuated when the geographic proximity of the source of the WOM message is more salient, this would provide empirical support for the hypothesized mechanism of common social identity, as the salience of the location—and thus the salience of the geographic proximity—enhances the social identification processes of the recipient (Forman, Ghose, and Wiesenfeld 2008; Heere et al. 2011; Hinds and Mortensen 2005; Lin and Viswanathan 2015; McGarty et al. 1994; Turner 1982). This would also provide additional empirical evidence in favor of the main identification strategy. Alternatively, if more salient location cues attenuate the relationship, this would provide evidence against the hypothesized mechanism.

According to the results presented in Table 4, we find a negative and significant moderating effect of the salience of the sender’s location on the impact of geographical distance on the effectiveness of WOM; the salience of location variable corresponds to whether the location of the WOM sender is explicitly mentioned in her profile. That is, the relationship between geographic distance and eWOM effectiveness is even more negative when the distance is more salient. This finding indicates that common social identity is a likely mechanism for the identified relationship. The results are robust to alternative econometric specifications.7

**Location-based social identity prominence.** We also examine the likelihood of the hypothesized mechanism in additional ways. For instance, we examine the effectiveness of eWOM under conditions that strengthen the role of geographical location in the social identification process. Specifically, increased political homogeneity in the local area of the recipient of the eWOM message is likely to enhance the importance of the location-based social identity of the recipient as it increases the salience and significance of individuals’ social identity due to political entities operating at geographic levels (e.g., precinct, county, state) and the characteristics of the local information environment (e.g., increased number of times an individual is reminded of the local identity, positive perceptions of the local community) (Jacobs and Munis 2020; Roccas and Brewer 2002; Taylor, Gottfredson, and Brower 1985). As a result, a pronounced location-based social identity of the recipient is likely to engender biases based on geographical distance, accentuating the relationship. Thus, if the relationship is accentuated when the political homogeneity in the local area of the recipient of the WOM increases, this would provide empirical support for the potential mechanism of social identification. Conversely, the opposite would provide empirical support against this potential mechanism.

Based on the results in Table A3 in the Web Appendix, we find a negative and significant moderating role of the political homogeneity in the local area of the recipient; we have collected data from the MIT Election Lab (https://electionlab.mit.edu) on political voting patterns at the precinct level for the 2016 elections and measure political homogeneity on the basis of the percentage of voters that would need to switch from the majority party to the minority party for the two parties to have equal votes. Put simply, the negative relationship between geographical distance and eWOM effectiveness is amplified when location-based social identity might be more pronounced due to increased political homogeneity. This finding lends support to the hypothesized mechanism of social identity. The results are also robust to alternative specifications.

Similarly, we also examine the moderating role of exogenous hardships in the local area of the recipient of the WOM message. If there have been significant local community hardships or natural disasters, then geography-based common social identity is likely to be more prominent for the residents of the affected area (Vezzali et al. 2015). Thus, if the relationship is accentuated when the geographic proximity of the source of the WOM message is combined with local community hardships for the recipient, this would provide additional support for the potential mechanism of common identity; we measure local community hardships using (exogenous) deaths related to extreme weather events in the location of the recipient of the message during the last five years prior to our observation window based on data from the National Oceanic and Atmospheric Administration. According to the results in Web Appendix Table A4, we find that the relationship between geographical distance and the effectiveness of eWOM is even more negative for users for whom location-based social identity is likely more pronounced due to local community hardships, lending empirical support to the hypothesized potential mechanism.8

---

7 These alternative specifications include examining the moderating role of the salience of common location (e.g., same city, same county) or the moderating role of the salience of common location on the association between geographic proximity (instead of geographical distance) and eWOM effectiveness; the geographic proximity variable is estimated as the inverse of the geographical distance variable.

8 The results are highly robust to alternative econometric specifications such as measuring local hardships over different time horizons or on the basis of weather-related property damages instead of deaths. Similarly, the results are robust to including in the moderating effect of local hardships whether the disseminator and the recipient of the eWOM message reside in the same local area or to examining the effect of geographic proximity instead of distance.
Location-based social identity versus other peer effects. We also examine whether the estimated moderating role of geographical distance is above and beyond other potential peer effects. Table A5 in the Web Appendix presents the corresponding results controlling for both the interaction between user similarity and visible message and the interaction between sociodemographic distance and visible message. Our findings remain highly robust, further illustrating that the estimated moderating effect of geographical distance is above and beyond other peer effects, including actual and perceived homophily. The results are also robust to alternative specifications, such as including additional interactions.

Ruling Out Additional Alternative Explanations

In addition to the aforementioned evidence and identification strategies, we assess various alternative explanations. Table 5 presents an overview of these, with the main ones discussed next and additional ones discussed in Web Appendix B.

Local marketing promotion effects. One may be concerned that the results might be driven by unpaid or organic marketing effects in the local region of the eWOM message recipient. To evaluate this, we supplement our data set with local web search trends for each product from Google Trends (Archak, Ghose, and Ipeirotis 2011). Table A6 in the Web Appendix presents the corresponding results controlling for both local marketing expenditures and ad response (via search behaviors) of the local audience. The results remain robust, alleviating concerns that local marketing promotion activities drive the results. The results are also robust to alternative specifications, such as using national Google Trends and advertising expenditures or estimating separate models for each potential confound.

Local user-preferences effects. We also examine the robustness of the findings to alternative specifications, such as controlling for homophily based on the overlap in brands that each social media user follows on the platform. Table A7 in the Web Appendix

Table 4. Estimation Results of eWOM Effectiveness Model with the Moderating Effect of the Salience of Geographical Distance.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>User similarity</td>
<td>1.244***</td>
<td>1.240***</td>
<td>1.238***</td>
<td>1.256***</td>
<td>1.260***</td>
</tr>
<tr>
<td></td>
<td>(.009)</td>
<td>(.009)</td>
<td>(.009)</td>
<td>(.010)</td>
<td>(.010)</td>
</tr>
<tr>
<td>Reciprocal relationship</td>
<td>5.868***</td>
<td>5.378***</td>
<td>5.525***</td>
<td>4.881***</td>
<td>4.606***</td>
</tr>
<tr>
<td></td>
<td>(.317)</td>
<td>(.290)</td>
<td>(.298)</td>
<td>(.272)</td>
<td>(.258)</td>
</tr>
<tr>
<td>Number of interactions</td>
<td>1.015</td>
<td>.979</td>
<td>.976</td>
<td>.989</td>
<td>.969</td>
</tr>
<tr>
<td></td>
<td>(.034)</td>
<td>(.033)</td>
<td>(.033)</td>
<td>(.034)</td>
<td>(.035)</td>
</tr>
<tr>
<td>Sentiment of message</td>
<td>1.540***</td>
<td>1.614***</td>
<td>1.642***</td>
<td>1.781***</td>
<td>1.568***</td>
</tr>
<tr>
<td></td>
<td>(.107)</td>
<td>(.112)</td>
<td>(.118)</td>
<td>(.138)</td>
<td>(.124)</td>
</tr>
<tr>
<td>Personalized message</td>
<td>1.041***</td>
<td>1.046***</td>
<td>1.043***</td>
<td>1.050***</td>
<td>1.035***</td>
</tr>
<tr>
<td></td>
<td>(.003)</td>
<td>(.003)</td>
<td>(.004)</td>
<td>(.005)</td>
<td>(.008)</td>
</tr>
<tr>
<td>User expertise</td>
<td>1.204***</td>
<td>1.183***</td>
<td>1.156***</td>
<td>1.581***</td>
<td>1.172*</td>
</tr>
<tr>
<td></td>
<td>(.030)</td>
<td>(.029)</td>
<td>(.028)</td>
<td>(.129)</td>
<td>(.099)</td>
</tr>
<tr>
<td>User leadership</td>
<td>1.006***</td>
<td>1.006***</td>
<td>1.004***</td>
<td>1.010***</td>
<td>.998</td>
</tr>
<tr>
<td></td>
<td>(.001)</td>
<td>(.001)</td>
<td>(.001)</td>
<td>(.002)</td>
<td>(.002)</td>
</tr>
<tr>
<td>Geographical distance</td>
<td>.891***</td>
<td>.947***</td>
<td>.961*</td>
<td>.965*</td>
<td>.965*</td>
</tr>
<tr>
<td></td>
<td>(.007)</td>
<td>(.019)</td>
<td>(.021)</td>
<td>(.021)</td>
<td>(.021)</td>
</tr>
<tr>
<td>Visible message</td>
<td>1.430***</td>
<td>1.588***</td>
<td>1.783***</td>
<td>1.783***</td>
<td>1.783***</td>
</tr>
<tr>
<td></td>
<td>(.168)</td>
<td>(.192)</td>
<td>(.214)</td>
<td>(.214)</td>
<td>(.214)</td>
</tr>
<tr>
<td>Visible message \times Geographical distance</td>
<td>.949**</td>
<td>.950**</td>
<td>.918***</td>
<td>.918***</td>
<td>.918***</td>
</tr>
<tr>
<td></td>
<td>(.021)</td>
<td>(.021)</td>
<td>(.020)</td>
<td>(.020)</td>
<td>(.020)</td>
</tr>
<tr>
<td>Visible message \times Geographical distance \times Salience of location</td>
<td>.920***</td>
<td>.948***</td>
<td>.974*</td>
<td>.974*</td>
<td>.974*</td>
</tr>
<tr>
<td></td>
<td>(.011)</td>
<td>(.012)</td>
<td>(.013)</td>
<td>(.013)</td>
<td>(.013)</td>
</tr>
<tr>
<td>Additional user controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Additional message controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Additional controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Product fixed effects</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Geography fixed effects</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Time fixed effects</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-22,422.8</td>
<td>-22,312.8</td>
<td>-22,280.2</td>
<td>-22,024.7</td>
<td>-21,751.1</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>2,763.5</td>
<td>2,983.4</td>
<td>3,048.5</td>
<td>3,559.6</td>
<td>4,106.8</td>
</tr>
<tr>
<td>N</td>
<td>132,955</td>
<td>132,955</td>
<td>132,955</td>
<td>132,955</td>
<td>132,955</td>
</tr>
</tbody>
</table>

Notes: eWOM effectiveness analysis with the moderating effect of salience of geographical distance. The salience of location variable corresponds to whether the location of the disseminator is explicitly mentioned in the profile of the disseminator. Additional table notes as in Table 3.
### Table 5. Alternative Explanations

<table>
<thead>
<tr>
<th>Alternative Explanation</th>
<th>Rationale</th>
<th>Identification Strategy</th>
<th>Table(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Location-Related</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location characteristics</td>
<td>Location-based characteristics may affect purchase likelihoods</td>
<td>Main identification strategy including geography fixed effects</td>
<td>All</td>
</tr>
<tr>
<td>Spatially correlated user preferences</td>
<td>User interests and brand preferences may be spatially correlated</td>
<td>Main identification strategy including latent user interests</td>
<td>All</td>
</tr>
<tr>
<td>Local time-difference effects</td>
<td>Time zone differences are correlated with geographical distance and may relate to differences in users' activities or moods</td>
<td>Main identification strategy including time zone differences</td>
<td>All</td>
</tr>
<tr>
<td>Local weather conditions</td>
<td>Local weather conditions may affect consumers' activities and moods</td>
<td>Additional weather controls</td>
<td>A9</td>
</tr>
<tr>
<td>Small-city effects</td>
<td>Geographic distances are shorter in smaller and more remote locations, where the demand for products sold online might be higher due to potentially limited availability of other products</td>
<td>Main identification strategy including geography fixed effects</td>
<td>All</td>
</tr>
<tr>
<td>Local marketing effects</td>
<td>Local marketing effects may affect purchase likelihoods</td>
<td>Additional ad controls</td>
<td>All</td>
</tr>
<tr>
<td><strong>User-Related</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homophily</td>
<td>Correlated behaviors among similar (across observed characteristics) peers</td>
<td>Main identification strategy including user similarity controls</td>
<td>All</td>
</tr>
<tr>
<td>Latent or unobserved homophily</td>
<td>Correlated behaviors among similar (across latent or unobserved characteristics) peers</td>
<td>Main identification strategy including user latent similarity controls</td>
<td>All</td>
</tr>
<tr>
<td>User characteristics</td>
<td>User characteristics may affect purchase likelihoods</td>
<td>Main identification strategy including user controls</td>
<td>All</td>
</tr>
<tr>
<td>Local demographic effects</td>
<td>Sociodemographic distance between users may affect purchase likelihoods</td>
<td>Main identification strategy including sociodemographic controls</td>
<td>All</td>
</tr>
<tr>
<td>Income-level effects</td>
<td>Income levels may affect where users select to live; as such, geographical distance may be correlated with income levels</td>
<td>Main identification strategy including income controls</td>
<td>All</td>
</tr>
<tr>
<td><strong>Message-Related</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Message content</td>
<td>Message content characteristics may affect purchase likelihoods</td>
<td>Main identification strategy including message controls</td>
<td>All</td>
</tr>
<tr>
<td>Nonrandom message visibility</td>
<td>Message visibility may not be random, despite provided evidence</td>
<td>Propensity-score matching</td>
<td>A13</td>
</tr>
<tr>
<td><strong>Product-Related</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product characteristics</td>
<td>Product characteristics may affect purchase likelihoods</td>
<td>Main identification strategy including product fixed effects</td>
<td>All</td>
</tr>
<tr>
<td>Marketing promotions</td>
<td>Advertising or other marketing activities may affect purchase likelihoods</td>
<td>Main identification strategy including ad controls</td>
<td>All</td>
</tr>
</tbody>
</table>

(continued)
presents the corresponding results. The results remain robust, further corroborating our findings.

**Small-city effects.** Another potential alternative explanation is that the results are driven by disseminator–recipient pairs located in small and remote locations as in such locations distances are in general shorter and demand for products sold online is higher due to the limited availability of other product alternatives (Bell and Song 2007; Meyners et al. 2017). We assess this alternative mechanism by repeating the analysis excluding any observations that correspond to small and remote locations, as determined by the Census (i.e., locale assignments). Table A8 in the Web Appendix presents the results; the results remain robust.

**Local weather conditions effects.** We also assess the alternative explanation that the results are driven by the local weather conditions affecting the moods and activities of users (Ghose and Todri-Adamopoulos 2016; Li et al. 2017). We assess this potential mechanism by controlling for the temperature, precipitation, and sunshine levels in the location of the recipient using data from the National Oceanic and Atmospheric Administration. Web Appendix Table A9 presents the corresponding results; the results remain robust.

**Robustness Checks and Alternative Identification Strategies**

We also undertake an extensive set of tests to assess the robustness of the results and further strengthen the findings, as discussed next; see Table 5 and Web Appendix B for additional details.

**Extended econometric specifications.** First, to enhance the employed identification strategy and examine the robustness of the findings, we further control for latent user characteristics by tapping into the social network structure and recent deep-learning advances. Specifically, we use the method of DeepWalk, a deep-learning method for graphs (Perozzi, Al-Rfou, and Skiena 2014), to learn the latent representations of the users and their similarity and further account for both network structure roles and latent user homophily. Table A10 in the Web Appendix presents the corresponding results. The results corroborate our findings. The results also remain robust to employing alternative deep-learning methods, such as the node2vec method (Grover and Leskovec 2016).

We also repeat the analysis including multiple user similarity measures. In particular, the similarity measures correspond to the similarity levels between disseminators and recipients based on (1) the Jaccard coefficient of their followers, (2) the Jaccard coefficient of their followees, (3) the topics discussed in social media posts using the results of the LDA model, (4) the intrinsic brand and product preferences of the users based on the overlap in brands that each social media user follows on the platform, (5) the demographic information at the corresponding geographic locations (i.e., average age and percentages of male, Black or African American, Hispanic, and Asian-origin residents based on Census data), and (6) the latent characteristics of the users based on the deep-learning methods for representation learning; in addition to (7) the reciprocity of the relationship and (8) the number of interactions between the users. Table A11 in the Web Appendix presents the corresponding results; the results remain robust.

**Alternative identification strategies.** We also examine additional alternative identification strategies to control for any potentially remaining differences between the visible and nonvisible messages; Table A1 in the Web Appendix presents a summary of the different identification strategies. First, we enhance our identification strategy following the covariate adjustment method of Imbens and Rubin (2015). Table A12 in the Web Appendix presents the corresponding results. The results remain robust; the results are also robust to including additional covariate interactions.

Moreover, as an alternative identification strategy, we combine propensity-score matching with the main research design. In particular, we model the propensity of each message to be rendered visible using all the variables that describe the users’ relationship and the message characteristics
as well as the geographical distance between the users.\textsuperscript{9} We conduct the matching based on the propensity scores before estimating again the same econometric models (for additional details, refer to the corresponding table notes). For this robustness check, we use one-to-one matching with replacement and a caliper of .05, yielding a standardized mean (median) absolute difference of .009 (.007) across all the variables, which ensures that covariate balance has been successfully achieved (Austin 2011); the density distributions of the propensity scores also indicate significant overlap and common support. As shown in Table A13 in the Web Appendix, the results remain robust. The results are also robust to nearest-neighbor matching with the generalized Mahalanobis distance.

Furthermore, as an additional alternative identification strategy, we build latent variable models where the sender–recipient similarity is latent and measured based on the various similarity features. Web Appendix Table A14 shows the corresponding results. Model 1 corresponds to the aforementioned latent variable model, while Model 2 combines the latent variable model with propensity-score matching estimating the model over the matched sample. The results of all the aforementioned alternative models are highly consistent and further corroborate our findings.

Finally, as an alternative strategy, to estimate the relationship between geographical distance and eWOM effectiveness, we also use a logit model (Train 2003) examining whether—rather than how quickly—a user purchases a product. As Web Appendix Table A15 shows, the results remain robust.

\textbf{Falsification tests.} We supplement these robustness checks with falsification tests to further assess whether the previous models are picking up spurious effects. As shown in Web Appendix Table A16, the results indicate our findings are not a statistical artifact of the specifications.

Overall, the findings remain highly robust to various alternative identification strategies, econometric specifications, robustness checks, and falsification tests. Figure 3 illustrates the corresponding estimated effects across the specifications.

\section*{Discussion and Implications}

In this study, we investigate the relationship between geographical distance and the effectiveness of eWOM. Specifically, we examine whether the geographical distance between familiar disseminators and receivers of eWOM messages plays an important role—beyond utilitarian reasons and proxying for consumer tastes—in driving recipients’ subsequent purchase behaviors. Our results show that the relationship between eWOM and the likelihood that message recipients subsequently also make product purchases significantly strengthens as the spatial proximity between disseminators and receivers grows.

\section*{Implications for Theory}

Our findings help advance understanding of conditions that affect online WOM performance. Many of the characteristics previously shown to impact eWOM outcomes relate to the product, brand, or message (Lovett, Peres, and Shachar 2013; Packard and Berger 2017). We contribute by illustrating the role of the important but often overlooked construct of geographical distance in eWOM effectiveness. In showing how geographical distance is still associated with the effectiveness of online WOM in the absence of geography-specific transaction costs between unambiguous users, we demonstrate how the social force field of geography can tether the potential of eWOM. That is, despite the promise of technology to reduce communication barriers and the proclaimed “death of distance” (Cairncross 2001; Graham 1998), we find that geographic constraints persist online in unexpected ways. Therefore, our results also help address the debate on whether and how geographical distance still matters online (Goldfarb and Tucker 2019) by showing that it can shape the influence of eWOM.

We also contribute to the theory of eWOM examining why geographical distance is associated with eWOM effectiveness. Specifically, we find evidence that social identification may explain why the influence of online WOM is negatively related to the distance between WOM message disseminators and receivers. That is, our results suggest consumers are susceptible to online information and cues related to social identification as they can, in turn, enhance message persuasiveness. Thus, information and cues relating to social identity can be agents of eWOM influence. Whereas much of the literature on the role of geographic distance in e-commerce and other online settings offers economically driven explanations for the impact of geography, our study proposes behavioral bias relating to social identification may be an underlying mechanism that drives the relationship between geographic distance and eWOM. This finding highlights the need for future research to study additional non–economically driven explanations that can induce such biases.

\section*{Implications for Practice}

Our findings have important implications for managers as well. For instance, a controversial argument in the industry is that solely characteristics of the disseminators catalyze the adoption of behaviors and products and thus much of marketing efforts to engineer WOM in social media focus on identifying such characteristics. However, our findings indicate that marketers should expand their focus to take into account the disseminator–recipient pairings and understand that factors pertaining to these pairs can be significantly related to the effectiveness of eWOM. In particular, our results suggest

\footnotesize{\textsuperscript{9} Following extant literature, we allow the matching to be affected by both senders’ and recipients’ characteristics because they can decrease the potential bias and variance of the estimates (Brookhart et al. 2006), but the results remain robust if only characteristics of the sender or recipient are used. The results also remain robust when we extend the propensity-matching model variables to also match consumers based on the corresponding region.}
geographical distance matters in online WOM and, thus marketers can readily take advantage of how geographical distance is associated with eWOM persuasion. Marketers may thus adopt data-driven strategies to selectively promote eWOM episodes according to the proximity of such episodes to each consumer, or to strategically engineer such episodes based on geolocation information. Interestingly, although research has begun to identify pairwise characteristics between senders and receivers that shape the influence of eWOM, such as tie strength and similarity across personality traits.

Figure 3. Hazard ratios (HRs) with 95% confidence intervals (whiskers) representing the percentage increase (HR > 1) or decrease (HR < 1) in postpurchase hazard across estimated models.
Beyond promoting or engineering geographically proximate eWOM episodes, marketers may also benefit from promoting and/or engineering episodes containing other social identity cues. The likely connection between eWOM outcomes and social identity suggests that firms may also consider other cues relevant to social identity formation to further boost the success of interpersonal communications and WOM messages, as enhancing social identification may significantly increase message persuasion and user engagement in the online world.

Our findings also have important implications for the design of viral marketing campaigns and ad content. Specifically, brands may boost the persuasiveness of their marketing campaigns by infusing into their content local cues or other identification triggers to induce consumers’ social identification processes. Relatedly, marketers are beginning to leverage users’ connections on social networks to develop and deliver marketing communications as part of their social advertising efforts. Our research suggests that they could further improve the effectiveness of these strategies by selecting geographically proximate connections to their targets.

Furthermore, going beyond advertising strategies, the implications of our work also provide actionable guidelines for optimizing the delivery of digital content. In particular, our findings can help platforms increase the effectiveness of their content curation and ranking algorithms by incorporating information on content location or source origin and by factoring geolocation into their determination of which user-generated content to disseminate. For instance, content generated by spatially proximate consumers may draw more attention due to identification processes and thereby increase the effectiveness of content provision. In a similar vein, social media platforms may also consider incorporating location information in other functions. For instance, social media platforms may incorporate such information into various other machine-learning algorithms, such as their whom-to-follow recommendations. In addition, our findings could also be used by marketers and platforms to better predict the diffusion of information, products, and user behaviors in social media (Adamopoulos and Todri 2015b).

Lastly, deepening our understanding of the factors that can attenuate or accentuate the effectiveness of eWOM has important implications that extend to public policies. For instance, revealing how geographic proximity is positively associated with eWOM effectiveness is critical for the development of effective public policies to induce positive behavioral changes, such as voter turnout, civic engagement, and public health actions.

Limitations and Future Research

While our work makes important strides in understanding how geographic proximity is related to eWOM effectiveness, we acknowledge certain limitations, which mostly stem from data availability issues. For instance, we examine the relationship between geographical distance and eWOM in a single social media platform because the service provider launched this venture on only one platform. Future research could examine whether the observed relationship manifests differently on other platforms. Moreover, we did not manipulate the visibility of the messages on Twitter because the venture did not alter the functionality of the platform in any way; future research could consider directly manipulating the visibility of the messages. Similarly, we did not manipulate the geographical distance of users from their followers. In addition, while we capture actual purchases in our data, we do not capture other consumer behaviors that could indicate interest in the products, such as online searches, as this type of information was not available to us. It would be interesting for future research to further examine such potential effects. Future research could also further examine and validate the underlying mechanisms. While we provide evidence that social identification may account for the relationship, future work may conduct experiments to verify this. Lastly, we do not observe in our data private communications between individuals due to privacy reasons and ethical concerns. Nevertheless, we hope these limitations provide avenues for future research that can deepen understanding of the critical role geographic proximity plays in eWOM and other online settings.

Associate Editor

Jacob Goldenberg

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

References


Choi, Jeonghye, Sam K. Hui, and David R. Bell (2010), “Spatiotemporal Analysis of Imitation Behavior Across New Buyers at an Online Grocery Retailer,” *Journal of Marketing Research, 47* (1), 75–89.


Todri et al.