1-1-2023

The Emergence of Discursive Leadership In Online Commmunities

Chad Diaz II
University of Mississippi

THE EMERGENCE OF DISCURSIVE LEADERSHIP IN ONLINE COMMMUNITIES

A Dissertation

Presented in Partial Fulfillment of Requirements

For the degree of

Doctor of Philosophy in Business Administration

Management Information Systems

The University of Mississippi

Chad Philip Diaz II

December 2023

ABSTRACT

The purpose of this Dissertation is to provide a detailed rationale for the way in which I address the following research questions in my dissertation: “What characteristics of user discourse lead to the emergence of leadership in online communities?” To this end, I provide a literature review of studies on online communities, examine past definitions, and propose an integrated definition of online communities. Focusing on the specific domain of leadership in online communities, I conduct a literature review of early and seminal articles published in this domain with the objective of identifying the foundational theoretical frameworks and the factors that past research studies found as significant for the emergence of leadership in online communities. Further, I identify the gap in knowledge and operationalization of these factors and propose the theory-based model that addresses this gap that I will test empirically in my dissertation. In conclusion, I describe the methodology for testing my proposed model and provide my results.

DEDICATION

This dissertation is dedicated to the most important people in my life who have supported, encouraged, and inspired me throughout my academic journey.

My parents for their unwavering support. Without their help, I would not have been able to pursue my dreams and accomplish my goals. Their constant encouragement and love have been instrumental in shaping the person I am today.

To my sister, Alana, who has been my rock through both laughter and tears, thank you for always being there for me. Your belief in me has been constant, and your infectious humor has brought much-needed levity during the most challenging times.

I am forever grateful to my wife, who gracefully endured countless late nights. Your patience, understanding, and love have been invaluable to me.

To my grandmother, MawMaw Black, you believed it from the start, and your “Little Einstein” is now a PhD.

Finally, to my niece and nephew, Aleigha and Barrett, I hope that this dissertation serves as proof to them that they can do anything they set their minds to. I am grateful for the opportunity to be a role model in their lives and to show them that hard work and determination can lead to great things.

To all of you, I dedicate this dissertation with love and gratitude. Thank you for your endless support, encouragement, and belief in me.

ACKNOWLEDGMENTS

As a former PhD Candidate said in their dissertation acknowledgments, “It takes a village.” That is certainly true of this dissertation as well. This project could not have been completed without a host of people supporting me throughout the journey.

I would like to begin my acknowledgments by recognizing the contributions of my dissertation chair, Dr. Anthony (Tony) Ammeter. Dr. Ammeter was gracious enough to take on this position on very short notice. Even though it was not something that he sought out, he has always been available and willing to discuss my research even when his schedule was busy. The beginning thoughts of this dissertation came about in Dr. Ammeter’s seminar as we examined the literature and discussed the findings. Thank you for the many meetings, discussions, and your determination to see this research completed.

I first met Dr. Brian Reithel during my undergraduate studies. Dr. Reithel’s MIS 419 class was challenging but also fulfilling in showing me I had what it took to be an MIS professional. Dr. Reithel’s encouragement throughout that course built my confidence in my abilities. Continuing into my doctoral journey, Dr. Reithel’s encouragement and advice have greatly benefited me.

During my early years in the PhD program, I worked with Dr. Sumali Conlon. From the first day of her database course in undergrad all the way through my journey as a PhD candidate, I have learned so much from Dr. Conlon. In undergrad, she taught me how to think about SQL queries and format my data so it could be easily manipulated. This instruction saved me much

time and has been an invaluable part of my analyses in this research. I learned about Natural Language Processing research as her research and teaching assistant. Dr. Conlon showed me how fulfilling this area of research can be as we conducted many analyses of data together. Dr. Conlon and her husband, Dr. John Conlon, traveled with me to many conferences and showed me great hospitality. With Dr. Conlon and the body of her past research, I had knowledge that has been integral to this study.

I met Dr. Milorad Novicevic during my MBA. Dr. Novi taught his course with a method that I was not accustomed to and that I had not experienced before; his class was very much based on dialogue rather than discourse. Throughout that MBA course, I found that Dr. Novi challenged me and caused me to reflect on why I held the position I did. As I progressed through that course, I found Dr. Novi among the most insightful and deep-thinking people I have ever known. As I completed my MBA and moved on to my doctoral degree, I reconnected with Dr. Novi and began working with him on multiple projects. Throughout my academic career, he has been a friend, confidant, advisor, and collaborator. This research was only completed due to his advice, support, and countless hours of editing. I am forever indebted to him for his contributions.

As I went through the process of conducting my statistical analysis, I was introduced to Dr. Houston Lester. Dr. Lester has been an amazing resource of knowledge on statistical methods throughout the dissertation process. Thank you for your many reviews of my work and for always responding to my questions.

Dr. Bart Garner was one of the first professors I met in my undergraduate career at The University of Mississippi. From the first few sessions of his undergraduate courses, I found Dr. Garner to be insightful and brilliant in his methods of relaying information. Through those class

sessions, I learned much about programming and built upon the foundation laid by others. Without the massive amounts of programming knowledge relayed to me by Dr. Garner, I would not have the technical capability that I have today, and this project would not be possible.

To my friends, the friends I met as I began my second year in the PhD program and who I spent untold hours with studying, conducting research, and who ensured that I seldom had to have dinner alone, Caleb, Kaushik, Yankun, and Moriah. Thank you all for your friendship and companionship throughout this journey; you kept me sane on many occasions. I am happy to say that I consider each of you my lifelong friend. To Juawana Aubrey, thank you for always checking in to see how things were going, for your constant prayers, and belief in me. Although you are a more recent acquaintance, you showed confidence in me and for that, I am appreciative.

Finally, to my family, my parents-in-law, Steve and Pam, thank you for your support, your prayers, attending my events, and always encouraging me to see this work to the end. My Wife, Stella, my companion and friend, who married a PhD candidate, maybe not realizing how much of a commitment this would entail. Stella, you have been a great encouragement to me through it all; thank you for your prayers, your encouragement, and your belief in my abilities. My parents, Chad and Dena, supported me through their finances and never-failing confidence in my ability to succeed; I am forever grateful. To my sister and her husband, Alana and Michael. Alana is always available to chat when I am feeling down, cry with me when I am sad, and laugh with me and celebrate in the good times. She is not only my sister but also my forever friend. Michael, while he may not be aware, is the cause of much of my academic career. Since I first met him, he has been someone I admire and a role model for me. Thank you both for your

support and belief in me. To my aunts, Aunt Debbie and Aunt Diana, thank you for your support, faith in me, and many prayers.

Without the support of great people, I could not accomplish what I have accomplished. With their support and the help of God, I am what I am today. Thank you, all.

Sincerely,

Chad Diaz II

CONTENTS

CHAPTER I

INTRODUCTION

The internet has gone through changes over three generations; Web 1.0 (non-interactive), Web 2.0 (person to person interaction), and Web 3.0 (machine to machine interaction) (Andersen, 2007; Hiremath & Kenchakkanavar, 2016). During the first iteration of the internet (Web 1.0) websites on which static pages where authors posted their writings but could not engage in social interaction. A pivotal point in the timeline of internet development came with the second generation, Web 2.0 (Andersen, 2007; Hiremath & Kenchakkanavar, 2016; O'Reilly, 2012). Web 2.0 provided a virtual place where humans could interact with one another through computer-mediated communication platforms with Web 2.0, the internet became a place where the more people who used it. The more information was shared and the more valuable the platform became for each of the users. In Web 3.0, the internet became much more focused on machine-to-machine interactions that do not require any human input (Kreps & Kimppa, 2015). The shift from the internet in the era of Web 1.0 to the internet in the era of Web 2.0 opened up opportunities for the foundation of online communities, an example of this shift are sites found in the Wayback Machine that existed before the advent of Web 2.0 (~2003) and the sites that exist in the Web 2.0 era. Sites that existed before Web 2.0 were non-interactive and only a small number of users had the capability of posting their content on the internet (Aguiton & Cardon, 2007; Hiremath & Kenchakkanavar, 2016; O'Reilly, 2012). In the era of Web 2.0 websites allowed for computer-mediated human interaction through which users began sharing information on large scales and forming online communities.

The research studies of online communities have focused initially on social networks that are formed through the affordances introduced in Web 2.0. These studies examined social network structures by focusing on their characteristics. Centrality is a primary structural characteristic that social network researchers examine to assess how embedded a user is in an online community (Steven L. Johnson et al., 2015). Centrality is a measure of the distance between a user and the other users in a social network. When centrality is assessed, researchers can determine how long it would take for a message written by one user to be forwarded to all other users within the network. A second commonly examined structural feature of social networks is “coreness.” Coreness is the proposition that there can be a core group of users who are highly active within their group but who are not greatly involved with all other users of the community. The majority of researchers of online communities have focused on these structural features of social networking in the form of reciprocal posts and responses (Dewan et al., 2017; Levina & Arriaga, 2014; Safadi et al., 2021; Wasko & Faraj, 2005). These and other structural features are formed through a dialogue and/or interaction among users in an online community.

Examining the social networks in online communities is important because online communities have become a virtual place where knowledge is transferred through human interaction. Specifically they are the places where knowledge is transferred through dialog from one person to another (Samer Faraj et al., 2016), particularly as the internet transformed into an interactive platform with the advent of Web 2.0 (Samer Faraj et al., 2016). Recent studies focused on those who excel in knowledge transfer emerge as leaders in an online community. See Appendix A depicting research on leadership emergence in online communities. The seminal study researching the emergence of online leadership was conducted by Johnson et al. (2015).

These authors found that the structural aspects of social networks in online communities have more significant explanatory power for the emergence of leadership among the linguistic characteristics of the user's discourse in the online community. Specifically, structural measures such as centrality, coreness, and boundary spanning were found to have much higher influence on the emergence of leadership in the online communities than the communicative measures of discourse such as readability, vocabulary richness, prototypicality, and sentiment. However, most researchers have not examined discourse features of the communication-based factors that contribute to the emergence of leadership in those online communities that do not allow for threading in human interactions, and therefore do not allow for the application of social network analysis (e.g., The online communities formed by Amazon reviews). I address this knowledge gap in past research with this dissertation in which I examine which features of the communication-based factors contribute to the emergence of leadership in such online communities. For this examination, I have selected the context of the Amazon reviews of consumer products, specifically products for children which are very sensitive and of high relevance for parents. Therefore, as communication makes primary impact in this context, **I address the following research question in my dissertation.**

What characteristics of user discourse do lead to the emergence of leadership in online communities with unthreaded user communication?

To address the above research question, I have organized my dissertation as follows. First, I conduct a literature review of articles on online communities with the primary objective to identify which definitions of online communities have been used in the articles. To ensure rigor, I focus only on eight articles which are published in journals listed in the Financial Times Top 50, and on 100 articles, which are published in the journals in the Association for Information Systems' top eight journals list. I review the definitions proposed in these articles with the objective to develop an integrated definition of online communities. My proposal for this integrated definition is the first contribution of my dissertation. Second, I focus on reviewing the literature on the topic of leadership in online communities. The objective of this literature review is to identify the factors that past research studies found as significant for the emergence of leadership in online communities (See Appendix A). Third, I propose new operationalization for some of these factors and extend the theory-based model. The extension of this model is the second contribution of my dissertation because it is the first model with a continuous dependent variable capturing the emergence of leadership. Fourth, I describe the methodology used for testing my proposed model and provide my findings. Finally, I outline limitations and implications of my study, as well as indicate the fruitful nature of directions.

CHAPTER II

LITERATURE REVIEW OF ONLINE STUDIES

In the subsequent paragraphs, I review the seminal and prominent studies of online communities research focusing on those articles involving member discourse. In their article exploring the power of gifts in an open source online community Bergquist & Ljungberg (2001) looked at the posts of users and used qualitative methods to analyze the discussions and gift giving within the online community. The open-source communities studied are based on networking, discussion, and knowledge sharing among users. Social dynamics proved imperative to the workings of these online communities. The authors found that a relationship existed between social relationships between members and the creation of “heroes” (leaders) in the community. In particular, the authors found that the “hierarchy in the online community is a matter of receiving or giving more or less attention.” Their definition implies that online communities are loosely coupled, as there is no formal structure for controlling individual behavior. Although no formal structure exists, individuals do conform their behaviors to the socially acceptable community norms. Individuals who are new to the community are informed implicitly and explicitly of social norms within the community. In order to obtain specific knowledge from members of the community, particularly those with more tenure, new members are expected to conform to the community post etiquette and style. As in an online community social interaction is the main focus of the users. They get the information from each other through dialogue that is meaningful to them.

Butler (2001) studied online social structures that involved social interaction via online community that was facilitated using text-based email lists. The online communities studied were listservs that allowed for mass communication among users. The model of interest in this study related membership growth and loss to membership size which was hypothesized to lead to communication activity between users. It was found that, while online communities change the medium through which interactions take place, the communities must still balance the forces of "membership size and communication activity." Through the balancing of these two impactful influences, online communities can retain the ability to "provide benefits to existing members" while attracting new participants to the online community. The definition of an online community proposed by Butler views online communities as social structures that use internet-based resources to centralize communication. This research was centered on communication between users in the listservs studied by examining how their activity influenced growth or loss of members. Therefore, social interactions are paramount focus of this study.

Wasko & Faraj (2005) examined the factors that influence participants in online communities to contribute knowledge to discussions within the community. The context for this research was an online message board which is "similar to a bulletin board." Online message boards in this case are visible to all users and structured as conversations between the individual participants. As each user's identity is openly shared to all other users, it is not possible to post anonymously in this type of online community. Relational capital cognitive capital and structural capital are related to the contributions that will be made by users in the in community. The dependent variable (knowledge contribution) was created by examining posts and determining if knowledge was contributed by their author. The factors identified as influencing knowledge contribution were classified in four categories, which included individual motivations, structural capital, cognitive capital, and relational capital. The authors found that posts by contributors who were focused on being helpful as well as being well embedded in the network were more likely to contain knowledge contributions. The definition of online communities derived from this work highlights the communities as being “self-organizing, open systems that focus on shared practice and exist primarily through computer-mediated communication.” (Wasko & Faraj, 2005, p. 37) Based on this definition of online communities, they require social interactions to exist and those interactions between users are used to determine how knowledge will flow throughout the community.

Mayzlin (2006) proposed a theoretical model depicting the effects of firms using word of mouth advertising in chat rooms and online bulletin boards. While there is no specific community sampled for this paper, the author examines how users interact socially through online messaging services. Social interactions between consumer users and users hired by industry firms are the primary focus of this theoretical paper. The research concluded that theoretically firms can find a balance where the number of online messages is convincing when accounting for the word-of-mouth messages communicated by other firms within the online community. The model proposed by the author focuses entirely on the effects of promotional chat messages sent from the firm’s users to the broader user base of online communities. While there is no explicit definition of online communities proposed by the author in this research, the implicit definition of online communities emphasizes chat rooms and online bulletin boards where users interact socially with other users. Therefore, the online communities of interest in this research must have a social network component for the proposed model to be applicable.

In studying “firm-hosted commercial online communities,” Wiertz and De Ruyter (2007, p. 348), used social capital theory to explore how commitment and knowledge contribution are impacted by other factors. The community of interest in this research is structured as a discussion forum that has threads (conversations) that may involve multiple users. Within this community, the focus of the authors was on relationships between users and the social capital that develops based on those relationships. The authors collected self-reported data from the participants in an online community to analyze the amount of knowledge contribution, commitment of users to the community and quality of knowledge contributions. They found that the users in the online community, who felt committed to that community, who tended to involve themselves in online discussions and who found the information within the community useful, were likely inclined toward sharing knowledge with the community. With this research, the authors find support for the relevance of the social interactions and dialogues between users of the online community. They defined online communities as "online aggregations of [users] who collectively co-produce and consume content about a commercial activity that is central to their interest by exchanging intangible resources." (Wasko & Faraj, 2005, p. 349)This definition specifies that the users in online communities co-produce content. This definition excludes any online space which facilitates the transfer of information or knowledge as being an online community, while emphasizing the use of social interactions among user in the online community.

The lens of Ising theory was used by Wonseok & Sangyong (2007) to study how "patterns of interaction" in online open source software communities can be used to identify relationships between users in an online community. Ising theory "generally concerned with basic patterns in dynamic interactions among physical objects or economic agents." (Wonseok & Sangyong, 2007, p. 1087) Open-source software communities were sampled and analyzed upon downloading the emails exchanged between users. The authors examined almost 100,000 messages that were exchanged between the users in the community and found that "membership herding" which is the tendency for everyone to do what they perceive others to be doing is most significant when external influences are weak. This herding was more significant in larger networks that have a more random connective structure. The findings by Wonseok & Sangyong point to the possibility of mass exodus events for online communities as well as how a core group of active users being essential to the long-term success of the community. The authors posit online communities emphasize "dynamic interactions between users" and "common goals" of those in the community as this research addresses the social interactions and dialogues between users within the online community, the analysis of the data is focused on the social network created within the community.

Computer-mediated knowledge sharing has been explored in the literature specifically in regard the effect of "IT artifacts" (Ma Meng & Agarwal, 2007). Data used in the analysis of this research was gathered using surveys which were collected from the users of two online communities. These communities were a support group and a community for owners of a specific type of car. Each of these communities allowed for the social interaction among members of the community. The verification of identity in online communities was found to be significantly related to the existence of "member satisfaction and knowledge contributions." The Meng & Agarwal show that perceived identity verification leads to satisfaction with the online community and knowledge contribution. With the forces of information need, group identification, offline activity, and tenure also being significant. The results of this study show that when users feel that there is a method for identifying users within the online community, they will be more likely to participate in a meaningful way. The authors also implicitly define online communities as "computer-mediated coordination and collaboration." This definition as well as the theoretical base of the article require that there be a space in the online community for coordination and collaboration. To coordinate and/or collaborate with each other there must be an affordance in the online community for social interaction.

Through the lens of social disorganization theory, Chua, Wareham, and Robey (2007) analyzed the efforts of community members in bulletin board sites as they fought online fraud in online auctions. Data for this research was gathered from publicly available websites (Vigilante Community, Stamp Collectors Against Dodgy Sellers (SCADS), and Traderlist.com). The authors identified characteristics of online auctions that caused the community members to rally and fight against the alleged fraud. For the Vigilante community the primary auctions targeted were high-value, one-time purchase items such as computers or televisions. These “vigilantes” tended to be knowledgeable of the typical actions of known con artists within the online auction sites. The members of the vigilante site showed signs of attachment to the trading community with some members purporting 20,000 transactions (Chua et al., 2007).

In examining how online communities form and progress through levels of community development Chua et al. (2007) studied online auction fraud communities. Chua et al. examined the online communications of users in three online trading communities to gather data. The authors used qualitative methods to identify how online communities relate to formal authorities. The researchers found a wide range of relationships that they posit coincide with the level of community development. In their analysis the authors also identified how online fraud can be managed through relationships. It is suggested that formal authorities build relationships with users in the online community to allow for cooperation between the online community and the formal authorities who may be able to act in a more official manner. The implicit definition of online communities in this paper emphasizes web pages and bulletin boards which allow for users to interact with each other and share information through dialogues. By defining and studying the communities that they chose, the authors specifically connect their research to social networks within the online community. While vigilantes were seen as loyal to their unofficial community, online auction sites sometimes saw them as disruptive and suspended them from their sites. Within the trader community the vigilantes were sometimes appreciated by community members but at times derided for their methods in outing the criminals. In the SCADS community the members were remarkably similar to those in the vigilante community but differed in their relationship with the trader community. SCADS members tended to have a better reputation within the community but were still accused at times of "meddling." The third community studied, Traderlist, differed from the other two in that its members cooperated with official authorities and having a good relationship within the trading community. However, it is noted that the actions of the Traderlist community reduced the power of sellers and the overall credibility of the trading community (Chua et al., 2007). The findings of the research support the cooperation of authorities with online community members in order to identify fraudulent activity, and that through this cooperation online trading sites can create more satisfied user bases that will lead to stronger attachment.

In the article introducing the special issue of Organization Sciences, Sproull, Dutton, and Kiesler (2007) identified three types of online communities. The type is composed of people who gather only online as they have no "pre-existing" ties outside of the virtual environment. The second type of online community is made up of those with ties outside of the online environment through physical social interactions, while the third type is communities that start online but later form an offline community. These editors point out that none of the articles in the special issue neither examine cross community comparisons nor make significant theoretical contributions.

Ren, Kraut, and Keisler (2007) combine identity and bond theories to explain the design and building of online communities. The authors identify five dimensions of an online community design that are likely to influence the bond of online community members to the online community. The five dimensions identified are “newcomer socialization, discussion moderation, community size, the role of core members, and community goals at multiple levels.” The authors argue, newcomers to the community will bring new life and perspective but they will not quickly conform to the established norms within a particular community. However, online communities can incentivize newcomers to stay and participate by having robust content and low thresholds for reading content compared to writing new content in the community. Community designers must decide whether to allow off topic discussions within the community because “if a community is identity based it will tend to have less off topic discussions” (Ren et al., 2007). While communities “strive to grow by recruiting new members,” growth will likely lead to diversification of the community. Therefore, designers will need to determine how to fully support their users whether to promote the community dynamics through grouping users and capping the number of possible community members, or by using other methods. They should also focus on the members of an online community who are the most active are the “core members of the community” because these core members will submit an exceedingly high number of the posts within an online community, thus creating a power law distribution of posts. Upon identifying these factors of online community design, the authors propose a framework for designers.

To examine the impact of the disclosure of identifying information on sales, Forman et al. (2008) used Social Identity theory as a theoretical lens and test it analyzing data gathered from Amazon reviews. In this research study, the authors found that initial reviewers disclosing identifying information such as their name, location, etc. would influence subsequent reviewers to do the same. The authors also found not only that identifying information in the reviews affected the rate of sales such that people were influenced more by reviewers who were geographically close to them, but the study also found that review content other than identifying information had less effect than the identifying information on the intent to purchase.

In their introduction to the special issue of Information Systems Research the editors primarily focus on Usenet, Facebook, and social media platforms for online communities. These allow members to "create and maintain connections with friends and strangers" (Agarwal et al., 2008). The authors of the articles published in the special issue primarily focus on the social aspects of online communities.

Through an examination of online financial forums, Campbell et al. (2009) looked at conflict and identity in online communities. The researchers identified three roles within the online community, Big Man, Trickster, and Sorcerer. The role of Big Man gains status primarily based on wealth and status. Tricksters gain their status typically through the antics of the community member. The Sorcerer is a member who manipulates others in the community. Having analyzed the text and connections in the online community the researchers found a significant relationship between the interactions of the three member roles with each other and status gained or lost in the community.

In their study examining the participation of firms in online communities, Miller et al. (2009) developed a model for simulating the interrelation of demand, interpersonal communication, and the online community strategies pursued by firms. The authors identified three different strategy types that firms tend to use in online community engagement.

Having conducted a qualitative analysis of tweets during extreme events Oh et al. (2013) examined factors affecting the probability of rumors spreading during the extreme events. Upon manually coding tweets based on six factors, Rumor, Anxiety, Source Ambiguity, Content Ambiguity, Personal Involvement, and Social Ties, the authors, found that the majority of tweets during extreme events were ambiguous as to the source of their content. The overall results of their analysis indicated that content ambiguity does not significantly contribute to rumor mongering, however, source ambiguity did significantly affect rumor mongering.

In a study of the open source software development community SourceForge.net, Zhang et al. (2013) found that continued participation of the online community members was significantly impacted by community response rates. Community response rates are measured by how many responses were made to posts by the user. The authors also found that respondents to threads in the online community were less likely to be impacted by the community response rate. In their model, the authors included the length of post as a proxy for information contained in the post or response. However, the study included no linguistic analysis beyond the counting of words.

Bayus (2013) studied the crowdsourcing of product ideas from an online community. The data for their study was gathered from an online community run by Dell Computer in which users were able to propose and discuss product ideas that might later be implemented by the company. The researchers examined how previous number of generated ideas and the number of implemented ideas were related to the likelihood of the user suggesting a product idea that would eventually be adopted by Dell. The authors introduced entropy as a control variable in order to measure the diversity of ideas suggested by the users. They study found that past success in suggesting implemented ideas was negatively related to the likelihood of suggesting a product that would be implemented by Dell and that diversity of past commenting activity was positively associated with the likelihood of suggesting an idea that would be implemented.

Kane et al. (2014) conducted a study examining how online communities manage to meet the need for change while retaining knowledge. Using data from Wikipedia, the authors conducted a qualitative study of posts in the online community and by manually coding them. The authors identified several roles that users play in the cycles through which the community goes, as well as the types of posts that are made in the community. The study found that over time the community went through predictable cycles which the authors liken to the software development life cycle (Kane et al., 2014).

In their study of data collected from an online healthcare community, Yan & Tan (2014) studied which factors could significantly affect the latent health outcomes of community members. The authors used linguistic indicators to determine which posts in the online community contained content reflecting emotional support, informational support, and companionship content. The researchers found that these linguistic factors had a statistically significant impact on the participation of the user in the online community. Their model also included social network variables which had an even larger impact on the latent health outcomes of the users.

By analyzing data gathered from a university's online campus forum in the timeframe before and after an earthquake, Nan & Lu (2014) studied how self-organized online community members can form an "orderly crisis management process." The authors used complex adaptive systems theory as a lens through which they studied how a multi-level self-organized process can emerge from an unorganized community. By examining the phases of extreme event response in the online community, the researchers found that assembling affordances in the community had the effect of increasing the negative emotion, and information posts, while the verifying affordance caused an increase in subsequent information and self-reflection in posts. The meta voicing affordance in the community had a negative effect on subsequent action posts and the associating affordance increased subsequent action posts.

In the study of knowledge sharing in an online community, Hwang et al. (2015) studied interactions between users. The data were gathered from an online knowledge sharing community that was run by a Fortune 500 company. The authors studied several factors about the members of the online community and how they interacted with each other. The dependent variable in the study captured whether knowledge was shared in a post, the variables of interest were the status of the user, their similarities with others in the community, the member's experience, and their visibility in the community. The only textual variable in the study was that of knowledge sharing. To create the dependent variable, a subset of posts was manually tagged as containing or not containing knowledge sharing characteristics. The authors found that similarity and interactions with other users were significantly related to the likelihood of knowledge sharing being contained in a post.

Sociability, knowledge sharing, and social network features have been shown to be related to emergence of leadership in online communities (S. Faraj et al., 2015). The authors studied data gathered from three Usenet groups. Usenet groups are unmoderated and allow messages posted by their users to the entire mailing list as soon as they are sent. These groups have no formal structure or hierarchy. Usenet allows for responses to be made to any message by any member. The authors found that the social network variables were significantly related to emergence of leadership in the online community. The study also found that knowledge contribution had a statistically significant effect on the likelihood of being classified as a leader in the community.

My literature review of online communities indicates that many of the reviewed articles on online communities either explicitly or implicitly define online communities. Using different definitional elements. Based on my analysis of the definitional elements used in the previous definitions (Appendix A), **I propose the following definition of online communities.**

An online community is a loosely coupled network of members who are not formally organized and whose social interactions are computer mediated.

This definition fully captures the essence of what an online community consists of as an online community requires people to exist and it also requires computers to mediate the interaction between the people in that community which is informally organized characterized by the turnover in users.

CHAPTER III

LITERATURE REVIEW OF LEADERSHIP EMERGING IN ONLINE COMMUNITIES

Research studies examining emergence of leadership in online communities' research have focused on how networking influences the trust that users place in each other (Abbasi et al., 2018; Bapna et al., 2017; Johnson et al., 2015; Mayfield & Mayfield, 2017; Vaast et al., 2017). Mayfield and Mayfield (2017) detail how the methods of communication for leaders to emerge have changed over time. The change has evolved from 1972, when articles detailed how secretaries were integral to the communications received from leaders to the modern day when leaders communicate directly to their audiences using online communication platforms. The authors detailed how research of leaders has traditionally focused on a "psychological perspective" of leadership that is focused on the designation of formal leaders. Meanwhile various disciplines outside of communication research have advanced a perspective of discursive leadership where leaders emerge organically based on "communication" and discourse in the online community "discovery."

In examining trust as a signal of leadership in online communities, Bapna et al. (2017) explored how participants in online communities tend to put more emphasis on closeness of the relationship between themselves and others in the community. Embeddedness within a network, communication between users "tagged" in photos together and their perceived similarity were found to significantly influence the amount of trust that users would place in one another. Bapna et al., found that for users with many friends the most significant factor was being tagged in photos together, thus indicating that when users have direct interaction with other users, those interactions lead to trust between users within the community.

Affordances in online communities allow for connectivity between users as shown by Vaast et al. (2017). Affordances in online communities which were explored by Vaast et al. were primarily based on social collective, shared, connective connectedness. One type of affordance not centered on social ties are individualized affordances. Vaast et al. found that interdependence between users was a primary factor in amplifying the messages presented by users and therefore a primary driver for the emergence of leaders within the online community. These factors are indicative of social network structures within the online community.

Embeddedness, in the social network is a significant concept in online communities research. Safadi, Johnson, and Faraj (2021) found that not only are those who are centrally embedded in an online community are perceived as valuable to lead for their influence and knowledge but those users who are “socially embedded and epistemically marginal” are the most probable to contribute the necessary knowledge to the community. To capture venues of knowledge transfer from users who are not centrally important to those scattered in the social network of the community, Safadi et al. focus on social networking variables in an effort to explain how leaders in knowledge transfer emerge in the online community of interest.

The process of leadership emergence within online communities has been theorized from several different perspectives. Networked influence theory was used by Lee, Yang, Hsu, and Wang (2018) to examine the rise, maintaining, and demise of online communities over long periods of time. Lee et al. found that emergent leaders in the online communities which they studied contributed to the conversations within that community over a broad range of conversations and topics. The users who tend to become leaders in online communities tend to contribute to a "homogeneity of communications." These leaders build networks, and their interactions are influential on the perception within the communities that they lead. While the concept of homogeneity of communications is not a social structure variable in this study, it cannot be isolated from the social network structure to determine its true impact on the emergence of leaders.

Online communities not only affect the emergence of leadership between people on publicly available sites, but also within corporations (Riemer et al., 2015). Through a study of different "forms of user influence in Enterprise Social Networks (ESN)" Riemer et al. found that both network centrality (hierarchy) and communication have significant impacts on the influence of a user in ESN. While users who are centrally located with respect to the users with whom they are communicating tend to elicit higher numbers of responses there was found to be a more significant effect based on the communication of the users. However, Riemer et al. were unable to fully come to conclusions on how much communication affects emergence of leadership in ESN.

The discursive communication-based aspect of leadership emergence in online communities is a nascent topic in academic research (Johnson et al., 2015). In their empirical study, Johnson et al. (2015) attempted to address some of the gaps in knowledge about the emergence of leadership in online communities by examining not only the social structure but also a wide range of communication-based factors. Using a sample of users from three vBulletin based message board sites, the authors conducted surveys in order to determine who in the online community is considered a leader. The communication factors studied in their research included readability, vocabulary richness, external linking, and prototypicality of the messages posted. Although these factors were included in the study, most of the variance explained in the analysis is attributed to the influence of the social structure (formal leadership, centrality, coreness). Due to their use of surveys, Johnson et al. (2015) had, however, a dependent variable for leadership emergence that was not objective since it was solicited from the self-reports of the users. In addition, although the authors found support for their hypotheses that both social structure position and communication characteristics contribute leadership emergence. A strange result was that lower lexical diversity leads to emergence of leadership.

Overall, my review of the literature on emergence of leadership in online communities indicates that past studies have virtually all been focused on online communities that have a social network structure, thus suppressing the relevance of the communication structure. As the seminal article in this literature is Johnson et al. (2015), which was selected as the best paper at the 2016 Academy of Management Conference, I have contacted Dr. Steven Johnson and his associate Dr. Hani Safadi and informed them of my intended dissertation study aimed at examining emergence of leadership in which there are no socially networked interactions and the only resource for emergence of leadership is communication (i.e. discourse). Both Dr. Johnson and Dr. Safadi were supportive of my planned dissertation and extending their research into the emergence of leadership in online communities. By identifying which characteristics of communication are relevant, my dissertation makes the following contributions as the extension of their study. First, in my study I have a continuous, objective dependent variable as a measure of leadership emergence in the online community that does not allow for social networking. Second, I check for the potential boundary conditions (i.e., moderating effect of variables) that Johnson et al. (2015) did not check. Third, I extend the range of discursive variables influencing leadership characteristics in an online community that Johnson et al. (2015) did not include in their model.

CHAPTER IV

THEORETICAL FRAMEWORK

In online communities, leadership emerges typically spontaneously as a coordinating process (Faraj et al., 2011). The relational mechanism that exists in interactive social networking online communities is the threaded communication that governs the process, but this mechanism does not exist in non-interactive online communities in which user discourse is the relational mechanism. In these communities the online community members attend to signals of discourse acceptance indicating who is the informal leader deserving their following. Therefore, Gerpott et al. (2018) used evolutionary signaling theory of social interaction to explain the process of leadership emergence in online communities.

The evolutionary signaling theory posits that the online community members are capable of capturing signals of competent leadership that are communicated by the member emerging as a leader in the online community. Specifically, online community members are posited to be able to attribute to a member the capacity for leadership based on the perceived quality of the communicated discourse.

The emergent leader that is successfully signaling discursive competence when her or she attracts recognition of the online community members. In other words, the successful outcome of this emerging leadership process is effective attraction of the community members' attention toward the emergent leader's most effective choices that result from the members' evolved monitoring of and learning from the emergent leader. The followers are attracted to the emergent discursive leader by comparing his or her discourse to the discourse communicated by other members of the online community. The characteristics of the emergent leader's discourse contribute to the follower's favorable comparative evaluation of the discourse's quality and the leader's commitment to communicate the best advice to the community members.

In non-interactive online review communities without social networking, which I examine in this dissertation, community members provide and acquire knowledge about specific products and services that are of interest to them as their users. As those communities are sustained by the content of the community members' posts, those members, whose posts are perceived as the highest quality, emerge as the online community leaders. As these communities thrive on self-organized voluntary participation, these emergent leaders are crucial for the sustainability of the online community.

While interactive online communities possess a social component manifested in threaded interactions, the non-interactive review communities do not possess this feature of social interactivity that has been shown as a significant resource for leadership emergence. In the review communities, leaders cannot rely on the use of the social and behavioral venues of interaction but must excel in the use of lexical features to create an impactful post that contributes reliable knowledge to the community. This impact must be recognized and endorsed by the community members.

In addition to evolutionary signaling theory, communication accommodation theory can be used to explain leadership emergence in online communities (Shepard, 2001). Communication accommodation theory (CAT) posits that the fit between a message and the targeted audience explains the effectiveness of a novel linguistic style in an online community (Lu et al., 2022). In other words, linguistic features of the members' message are likely to explain which one of the members will emerge as the online community leader. Specifically, "adjusting linguistic styles (such as linguistic complexity and sentiment) to fit the audience's frame can facilitate communication effectiveness and increase the likelihood of being recognized as a high quality contributor, which, in turn results in a higher reputation score for leadership determination" (Lu et al., 2022). Based on communication accommodation theory, the key feature of the communicated message is linguistic complexity, which reflects readability, diversity, and sentiment of the message (Lu et al., 2022).

The leader's proactive discourse is likely to be perceived as most helpful to the community members when the leader's discursive contributions are perceived in the follower's reviews as: a) longer; b) more complex; c) more positive; d) more frequent; and e) more diverse than those of the online community members that are not perceived as leaders. In the subsequent chapter, I present for my theory-based model and develop my hypotheses using evolutionary signaling theory and communication accommodation theories.

CHAPTER V

THEORY-BASED MODEL AND HYPOTHESIS DEVELOPMENT

In online communities that do not provide for social interactions between users, the content of a user's discourse is the only resource of potential value to the community. When users are evaluated by other users to contribute most valuable product-related information or knowledge to the online community they emerge as the online community leaders (Samer Faraj et al., 2016). Both evolutionary signaling and communication adaption theories posit that not only the knowledge that users present in their writings, but also the way that the information is communicated to other users affect the emergence of the user's perceived leadership status within the online community (Johnson et al., 2015). However, these theories explain different characteristics of the communicated posts.

Evolutionary signaling theory posits people put out signals that are subsequently viewed and evaluated by others. In online review communities, one signal that is quickly captured by readers of posts in the community is the volume of text in each review. The length of the text is one of the first signals that a user views when looking at a review in the online community. The length of a review has been shown in past research to be the key signal indicating the helpfulness of online reviews (Eslami et al., 2018). Review length can be considered a indicating the likelihood of valuable information that can be found in an online review (Eslami et al., 2018; Salehan & Kim, 2016).

As the length of the online review is considered as a textual indicator of leadership, I hypothesize:

H1: The longer a reviewer's post the more likely they are to emerge as a leader in the online community.

The responses of users to a post in an online community tend to be affected by the perceived complexity with which the post was written (Lu et al., 2013). Communication accommodation theory posits that readers of discourse will receive the message better if they are able to understand what is being communicated(Lu et al., 2022). Leaders who tailor their discourse will be more likely to be received by their target audience. In this way, leaders in online communities will signal the ability to write at a level of complexity that can be meaningfully received by a broad range of readers. When text is easier to read, the participants in the online community find it more informative and more relatable. Therefore, I hypothesize:

H2: The less complex a reviewer's post the more likely they are to emerge as a leader in the online community.

Natural language processing researchers have identified the relevance of the concept of text valence (Mohammad, 2016; Simmons et al., 2011). The valence of text indicates the extent to which the author is positive about the subject. Leaders are constantly modifying their communications to the expectations of their audience; leaders must consider that their communication will be positively received.

The audience within any specific community may value the positive sentiment of discourse differently and according to communication accommodation theory, leaders will need to adapt to fit the expectation of their intended audience (Lu et al., 2022). Therefore, I propose the following hypothesis:

H3: The more positive a reviewer's post the more likely they are to emerge as a leader in the online community.

An important factor influencing how text is received in an online community is how recognizable the text is to the reader (Johnson et al., 2015). When a writer uses words that are similar to the words typically used within an online community, they are accommodating the audience and will be more likely to be received by the reader. Based on communication accommodation theory, leaders will adapt their communication to fit their audience and the online community members will find the text more relatable and it will increase their perception of the writer as a leader (Lu et al., 2022). The concept of prototypicality examines how similar the text from one writer is to that of the community. When an individual writer's vocabulary is more aligned with the vocabulary used in the community, the text is viewed as prototypical. Thus, I hypothesize:

H4: The more prototypical a reviewer's post the more likely they are to emerge as a leader in the online community.

When users post in an online community, they may quickly lose the attention of their audience if they use a few major words as their only text. The use of few presumably powerful and long words may be received as a signal of lack of care by readers within the community. However, those that aspire to emerge as leaders in online communities should make their writing more readable by using more words that are short and easy to read. This construct vocabulary richness has been found by Johnson et al. (2015) to be "similar to but different from readability" because readability is focused on individual words and sentences and not on the text as a whole. Vocabulary richness is described by Johnson et al. (2015) as a measure of the number of distinct words that a writers use in their posts in the online community. Those with more expansive vocabularies tend to write posts using a larger variety of smaller words that are less confusing to their readers (Huffaker, 2009; Johnson et al., 2015). Therefore, vocabulary richness is focused on the entire text and on determining how many different words the writer has used. As vocabulary richness has been found in past research as a statistically significant predictor of leadership emergence in an online community beyond the significance of social network variables (Johnson et al., 2015), I propose the following hypothesis:

H5: The higher a reviewer's vocabulary richness the more likely they are to emerge as a leader in the online community.

Evolutionary signaling theory posits that leaders are constantly outputting signals that are then received by their audience. Due to the interrelated nature of the signals that are signaled by a leader, there is likely to be some intermingling of those signals in the mind of the receiver. The variables most likely to interact are Textual Diversity with Valence, Textual Diversity with

Complexity, and Richness with Valence. Therefore, I propose the following hypothesis which has three parts:

H6a: The interaction between Vocabulary Richness and Valence will have a positive effect on the emergence of leadership in the online community.

H6b: The interaction between Textual Diversity and Review Valence will have a positive effect on the emergence of leadership in the online community.

H6c: The interaction between Textual Diversity and Review Complexity will have a positive effect on the emergence of leadership in the online community.

The six hypotheses that I have proposed are depicted in my model which predicts the emergence of leadership in online communities (See Figure 1 below). The hypothesized model will use averages of each proposed construct to evaluate the values at the reviewer level. Averaging the data introduces the potential for a loss of information from variance from the average that could be gleaned from the data. However, by averaging the data we are given the ability to study the reviewers and how they emerge as leaders in the online community. By empirically testing this theory-based model, I aim to capture the effect of the hypothesized discursive characteristics of the posts in the online community on the emergence of leadership. In the subsequent chapters, I present the method that I used for this empirical testing.

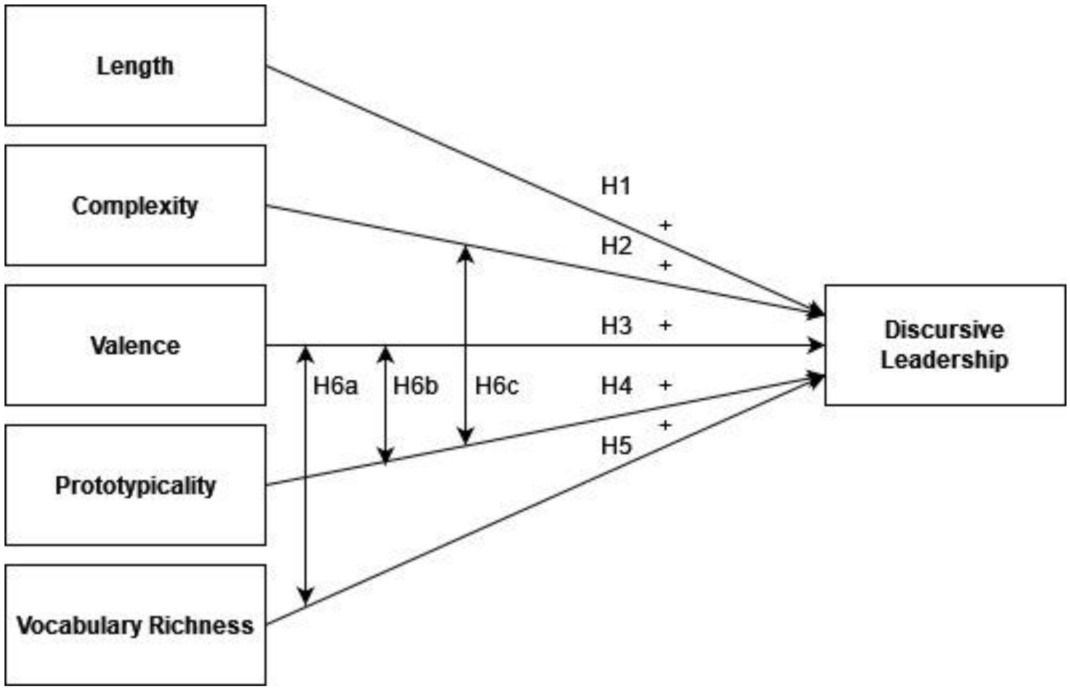
Length
Complexity
Valence
Prototypicality
Vocabulary Richness
Discursive Leadership
H1
H2
H3
H4
H5
H6a
H6b
H6c
+
+
+
+
+

Figure 1

CHAPTER VI

METHODS

6.1. Sample

The data used to test this model was gathered from Amazon.com during the time period from May 1996 to July 2014 by He & McAuley (2016) and McAuley et al., (2015) and is described in their research. From the dataset that contains approximately 143 million reviews, I have taken a subset data gathered by pulling reviews for products that have the first or second categorization of "Toys & Games." There are approximately two million reviews in this category. As in this sample, the majority of users have zero helpful votes from other users, this creates a power law distribution for the number of helpful votes for each review writer.

6.2. Data Preparation

In order to prepare the data for analysis I retrieved the data from the online repository with the help of the authors who gathered it as part of research at Stanford University (He & McAuley, 2016; McAuley et al., 2015). I then wrote a python script to extract and insert the data into a MySQL database for ease of data management and retrieval with Stata (Appendix B). From the gathered data I extracted the Toy category using SQL in the MySQL database. In total there were 2,252,771 reviews within the category. Of the reviews 461 were dropped due to parsing incorrectly, 12 records were dropped due to the ASIN (unique product identifier) not parsing, 7,209 records were dropped due to the Review Text being blank, 12,797 records were dropped due to not having a review time. After all data scrubbing efforts 2,232,292 reviews remained. In order to examine the data at the reviewer level the helpful ratings were summed up for each reviewer and the variables of interest were averaged which gives a total of 1,334,844 records which corresponds to one record for each reviewer.

6.3. Measures.

In this research emergent online leadership is measured by looking at the number of helpful votes that a member of the community has received. This will be used as the dependent variable of interest. In the Amazon review community, each post has a button that allows readers to mark the review as helpful to them. In the analysis for this dissertation, we will account for all posts in the community no matter the number of helpful votes the poster has received. In past research the community members studied have been filtered by a certain number of posts (Johnson et al., 2015) and by their reputation score (X. Lu et al., 2022). In this research I will use all valid observations in the analysis of emergent leadership in online communities. To better estimate the effect of the theoretically based measures on the emergence of leadership I summed the number of helpful votes for each reviewer and then created an average of each of the independent measures analyzed to create a single entry in the database for each reviewer. By bringing the analysis up to the reviewer level we can examine leadership for each reviewer overall and not at the individual review level.

Vocabulary richness is measured by determining the number of distinct words that each reviewer uses in their posts within the review community. This measure of vocabulary richness has been demonstrated in past research by Johnson et al. (2015) and by X. Lu et al.(2022). As in past research I measure diversity by using a Python script and creating a dictionary of all distinct words that a reviewer used in all of their posts within the online community. I then use this dictionary of words to compare to the words used in each individual post made by that reviewer to create an index that reflects the number of words used in an individual post but not used in any other post by that reviewer. Example reviews and the associated vocabulary richness scores can be seen in Table 1 below.

Review Text	Vocabulary Richness Score
I love it it was want I wanted. It was the best. I got to me real fast. Love it!	4.5
Would not recommend - I ordered the correct size but a different one arrived. Mistakes happen but this was too easy to do correctly.	176.33333

Table 1

To measure the prototypicality of any one poster I will use a Python script to analyze the text of the corpus. I will then compare each of the reviews to the dictionary created from the corpus. Once each post is compared to the entire text corpus it will be assigned a prototypicality score which will then be used in my analysis. Example prototypicality scores for two reviews can be found in Table 2 below.

Review Text	Prototypicality Score
...I LOVE IT...I LOVE IT...I LOVE IT...I LOVE IT...I LOVE IT... DOES THAT HELP :) VERY VERY HAPPY AND OH YEAH SO IS MY DAUGHTER LOL	.13445378
I bought this for my son because even thought he is almost 21 he still collects Hot Wheels. Just what we needed to keep them in a safe place.	.96428571

Table 2

Positivity of online posts can be measured in multiple ways. In my dissertation, I will use the method utilized by Johnson et al. (2015), AFINN. AFINN, as described by Johnson et al. (2015), uses a customized dictionary contains a list of words from the English language which have been classified based on their emotion. The dictionary used by AFINN has been customized so that it fits the language used in online communities (Johnson et al., 2015; Nielsen, 2011). When running the AFINN algorithm on the content of the posts gathered from the online community, each post will be assigned a score based on emotion portrayed by the words in the post. An excerpts and valence score of a highly positive and a less positive review can be seen in Table 3 below.

Review Text	Valence Score
…This outfit was absolutely Stunning for my doll. I have a Ashton Drake Collectable Doll who I believe has been retired. Her name is "Nina" and she is a Baby Doll of 16" But this outfit fit great. My doll can talk giggle turn her head and has wonderful soft eyes that open and close…	158
…To be able to carry hundreds of books on a few memory cards would have made the eBook worth its weight in gold. Instead to expand it's meager memory I'm required to box it up and send it along with $150 across county for 3-4 weeks for their 'technicians' to remove 4 screws and plug in a proprietary 32 MB memory module…	3

Table 3

Readability can be measured in multiple ways using formulas that have been widely used in the literature of natural language processing. Complexity of text is often measured by determining the readability of the text in question (Johnson et al., 2015; Y. Lu et al., 2013). Text becomes easier to read when "short, simple, familiar words" and "simple sentences" are used (DuBay, 2004). In my dissertation, I will use the Flesch-Kincaid Reading Ease formula to determine the readability of the text corpus (Johnson et al., 2015; Kincaid et al., 1975). Flesch-Kincaid is calculated using a formula that accounts for the length of sentences, words, and the syllables in each word. The formula produces a readability score that ranges from 1 to 100 with 100 considered the easiest text to read. In this research I will utilize a Python script to automate the analysis of the text of each review and assign it a readability score. Example reviews and their readability score generated using the Flesch-Kincaid method can be seen in Table 4 below.

Review Text	Readability Score
So much fun! We kept letting people jump in after watching - people who refuse board games at all costs were in the thick of things - on lookers want to help! No one's got the upper hand in this game since the phrases are so esoteric its all about creative phrase finishing and a little bit of bluffing. This is not a game for 2 people though since Amazon never tells you those details - 3 is minimum but do-able quite frankly the more the merrier!	3.14
It's easy enough to learn and get into to. If you child is at all into Pokemon (still) then this game will be a cool item for their collection.	75.88

Table 4

The length of a review can be found by a simple character count of the text contained in the review. I will use a SQL script within a MySQL database to create the count of characters in each review. This count will then be used as the measure of length.

To better understand the effect of the dependent variables at the reviewer level I will aggregate the results of each measure for each reviewer. In order to aggregate the results, I will take the sum of the helpful votes for each review that a reviewer has written. For each of the dependent variables I will take the average of the value over all reviews written by the individual reviewer. These aggregated values will be used in my statistical analysis.

6.4. Statistical Methods.

To measure leadership in the Amazon dataset I will use the number of helpful votes as the dependent variable in my analysis. The number of helpful votes is a count of the number of times the user's posts have been marked as helpful by readers within the online community. As such this number cannot be negative and is considered a discrete count variable (Williams, 2021; Hoffman, 2022). Due to the nature of count variables we need to check for and deal with extra zeroes (if present) in the dataset.

Counts are non-negative integers which can go up to any value but are never negative. While the counts are always positive the predicted counts may not always be positive. Hoffman (2021) shows that the link function (Equation 1) predicts the log of count as $\hat{y}i$.

$$\boldsymbol{g(\cdot) Log[E(y_i)] = Log(\hat{\mu}_i) = [model]}$$

Equation 1

The inverse link (Equation 2) is used to un-log $\hat{y}_i$ back to a count. Equation 1 above shows how the predicted count can be negative. Equation 2 above shows how we can transform the predicted log value back to a count variable for interpretation.

$$\boldsymbol{g^{-1}(\cdot) E(y_i) = \exp(\hat{y}_i)}$$

Equation 2

There are a few distributions that can be used when analyzing count value data. The Poisson Conditional Distribution is the most basic of these. For Poisson to be applicable the mean and the variance of the data must be equal as the distribution has only one parameter and it is equal to the mean and variance. There are a few issues presented with Poisson, the first is when the mean does not equal the variance (Hoffman, 2022). If the mean is less than the variance there may be "under-dispersion" and if the mean is greater than the variance there may be "over dispersion." The second comes when there are no zero values in the data. The third issue is when there are too many zero values in the data. Each of these problems require adjustments to the models in order to correctly predict the outcomes (Hoffman, 2022; Williams, 2021).

To address the first problem, variance being greater than the mean, we can "add a parameter that allows the variance to exceed the mean." (Hoffman, 2022). The addition of the parameter allows our distribution to become a Negative Binomial Distribution. The negative binomial distribution with mean μ and dispersion scale k is shown in Equation 3 below. When k is equal to zero the negative binomial distribution approximates the Poisson distribution. The negative binomial distribution can allow k to not equal zero and becomes a more relaxed form of the Poisson distribution that can better predict the outcome of the count variable in data with a greater skewness that otherwise follows a Poisson distribution (Hoffman, 2022).

$$\boldsymbol{Prob(y_i = y) = \frac{\Gamma(y + \frac{1}{k})}{\Gamma(y+1) * \Gamma(\frac{1}{k})} * \frac{(k\hat{\mu})^y}{(1 + k\hat{\mu})^{y+\frac{1}{k}}}}$$

Equation 3

In the dataset of interest for this dissertation we do have zero values so we will not need to deal with the second problem type, however, we will need to check if there are an excess number of zeroes and deal with that. In order to deal with there being too many zero values in the data we can use Zero-Inflated Negative Binomial (Hoffman, 2022; Williams, 2021). A zero-inflated negative binomial model deals with excess zeros by separating two types of zero values, expected and inflated through “Bernoulli and Negative Binomial” (Blevins et al., 2015; Corredoira & Rosenkopf, 2010; Hoffman, 2022; Soh, 2010). Zero-inflated negative binomial creates two models. One model (Equation 4) predicts the probability of the observation being an extra zero, and the second (Equation 5) predicts all other counts using a “link equal log” function (Hoffman, 2022; Williams, 2021).

$$\boldsymbol{Logit[p(y_i = extra\ 0)] = \beta_{oz} + \beta_{1z}(x_i)}$$

Equation 4

$$\boldsymbol{Log[E(y_i)] = \beta_{0c} + \beta_{1c}(x_i)}$$

Equation 5

Appendix B-Table 1 shows the overdispersion of the dataset that I will use for my statistical analysis. This can be seen in the standard deviation (square root of the variance). The standard deviation of my dependent variable is much larger than the mean (due to mean centering the mean may be negative for some variables, e.g., ASIN). Therefore, I need to consider controlling for this overdispersion. To do this I have conducted the following analyses using Stata 14. Appendix B-Table 2 in Appendix B shows the Poisson regression output for my dataset. In order to have a common basis to which I can compare the various statistical models I then ran estat ic in Stata to retrieve the AIC and BIC for this model (Appendix B-Table 3). AIC and BIC allow for the comparison of AIC and BIC allows us to determine which model is a better fit for the data, smaller numbers are better (Hoffman, 2022; Williams, 2021).

Second, I ran a negative inflated binomial statistical analysis with the results shown in Appendix B-Table 4. I then ran estat ic for this model to generate the AIC and BIC. Appendix B-Table 5 shows the results of that command. Based on a comparison of the AIC and BIC for the

Poisson regression (Appendix B-Table 3) and for the negative inflated binomial regression (Appendix B-Table 5), I found that the negative inflated binomial model is preferred over the Poisson regression.

Third, in order to account for possible excess zeros, I ran a zero-inflated negative binomial (zinb) regression. The results of the zinb regression can be found in Appendix B-Table 6. I then ran estat ic to retrieve the AIC and BIC for this model (Appendix B-Table 7). Based on the comparison of the AIC and BIC for the negative inflated binomial model (Appendix B-Table 5) and for the zero-inflated negative binomial model (Appendix B-Table 7) we find that the AIC and BIC for the zero-inflated negative binomial model are smaller showing that the zero-inflated model is a better fit for this data. Stata also returns the Vuong test for determining if the zero-inflated model is a better fit for the data than the negative binomial model (Appendix B-Table 6). Based on the significant result of this test (Pr>z=0.000) I found that based on this measure as well the zero-inflated negative binomial model is the best fit for this dataset.

To show the significance of the interaction terms that were added to the model I dropped them all from the model and ran the zero-inflated negative binomial model with only the main effect variables (Appendix B-Table 8). This model has an adjusted McFadden's R2 of 0.342 (Appendix B-Table 9). I then added each of the interaction terms to the model (Appendix B-Table 10-Appendix B-Table 28). By adding the interaction terms individually, we can see that the interaction terms cause the R2 to increase to 0.347 (Appendix B-Table 28) in the full model with all interactions (Appendix B-Table 27). I also found that the AIC and BIC decreased with the addition of the interaction terms. While all of the interaction terms were significant, I will only interpret the hypothesized interactions.

CHAPTER VII

RESULTS

Through my statistical analysis I found that all independent variables and interactions were significant at the $P<.01$ level except Linguistic Diversity and Linguistic Diversity by Vocabulary Richness. In the final model (Appendix B-Table 27) I have included two control variables, Star Rating and Count ASIN. Star Rating is a non-linguistic rating that each reviewer gives to the product as part of the review process on Amazon. Star rating is a common feature in online reviews and has been studied extensively in the literature. The star rating is included as a control since it does appear with the online review but is not considered a variable of interest in this research as it is not a textual indicator of leadership in the online community. The second control Count of ASIN is simply a count of the various products to which each reviewer has posted a review. Count ASIN is included as a control to account for the aggregation of the review variables to the reviewer level and any skewness that may occur for reviewers that have very few or many reviews in the online community.

The length of reviews is found to positively impact the perception of leadership in the online community (0.00235; $p<0.01$) which shows support for Hypothesis 1. The Linguistic Diversity of reviews is found to negative coefficient as related to leadership in the online community, however, it is not significant at conventional levels of statistical significance (-0.05373; $p=0.216$), which does not show support for Hypothesis 2. Textual Richness is shown in my analysis to be positively related to emergent leadership in the online community (0.00102; $p<0.01$) showing support for Hypothesis 3. Textual complexity is shown in the analysis conducted to negatively affect the perception of leadership (-0.00115; $p<0.01$) and shows support for Hypothesis 4. The valence of textual content is shown to negatively affect the perception of leadership (-0.01845; $p<0.01$) which does not show support for Hypothesis 5. The interaction of Vocabulary Richness with Valence (.0001; $p<0.01$) is shown to have a positive effect on leadership in the online community showing support for Hypothesis 6a. The interaction of Textual Diversity with Review Complexity is shown to have a positive effect on the emergence of leadership (.007; $p<0.01$), showing support for Hypothesis 6b. The interaction of Textual Diversity with Review Valence is shown to have a negative effect on leadership (-.12069; $p<0.01$), showing support for Hypothesis 6c.

CHAPTER VIII

DISCUSSION AND CONCLUSION

8.1. Discussion.

I have proposed six hypotheses and tested them based on my theoretical model. These hypotheses are based on both evolutionary signaling theory and communication accommodation theory. Evolutionary signaling theory is based on people giving signals which are then received by others. As the volume of text in a review is one signal that is perceived by readers of reviews in an online review community, hypothesis one is that longer reviews posted by a user will make them more likely to emerge as a leader in the online community. Through empirical testing, I have found support for this (H1) hypothesis because of a statistically significant relationship between longer reviews and the emergence of leadership in the online community.

Communication accommodation theory posits that as people communicate, they will adjust their discourse to accommodate their readers ability to understand. Ease of reading a post is important to the emergence of leaders in online communities. Therefore, writers who emerge as leaders will adjust their discourse such that it accommodates the understanding of their audience. Hypothesis two (H2) proposes that there is a relationship between less complex reviews and the emergence of leadership. I have not found statistically significant support for this hypothesis in my empirical study.

Valence has been found in past research to be a relevant factor in lexical analysis. Valence indicates the extent to which the writer is positive about the subject of his or her writing. Those who emerge as leaders will need to adapt their communication to positivity so that it appeals to their readers. Hypothesis three (H3) proposes that writers who have more positive posts will be more likely to emerge as leaders in the online community. In my empirical study I have not found significant statistical support for this hypothesis.

When text is recognizable to readers it causes them to better identify with the author. Recognizable text will be better received by readers. Leaders that emerge in an online community will adapt their communication use terms that are prototypical to those expected in the community. Hypothesis four (H4) proposes that when a writer's posts are prototypical and aligned with the vocabulary of the community, the writer will be more likely to emerge as a leader in the online community. I have found through my empirical study that this hypothesis is statistically supported.

The vocabulary of a writer and how it is used has been researched as an indicator of leadership emergence in online communities. Johnson et al. (2015) state that while vocabulary richness is related to the ease of reading text, but they are different notions. Readability is focused on individual words, while vocabulary richness is focused on the text as a whole. A rich vocabulary will lead to authors using a large number of smaller, easier to understand words which will be easier to understand and will lead to the signals given by the writer to be better received by their audience. Hypothesis five (H5) proposes that vocabulary richness will have a positive impact on the emergence of leadership in the online community. I have not found support for this hypothesis through my empirical study.

Based on evolutionary signaling theory, which posits that leaders are always putting forth signals which are perceived by their audience there is likely to be some interrelated signals between the factors which affect the emergence of leadership in online communities. Hypothesis six a, b, and c (H6) propose that there will be statistically significant interactions between the factors that lead to the emergence of leadership such that significant interactions will exist. I have found that there are statistically significant interactions between the factors in my theoretical model supporting these three hypotheses.

While hypothesis five which posits a connection between higher vocabulary richness and leadership and hypothesis three which proposes a positive relationship between valence and the emergence of leadership were not directly supported in my empirical study, the interaction between valence and vocabulary richness was found to have a significant positive relationship with the emergence of leadership in the online community. Therefore, as vocabulary richness and sentiment rise, they interact and have a statistically significant positive effect on the emergence of leadership in the online community.

In this dissertation, I examine how the discourse characteristics of the lexical content created by members of online communities with a lack of social network (i.e., interpersonal interactions via threaded dialog) features affect their emergence as discursive leaders in the community. Using data from Amazon.com reviews of toys and games, I have found that several theory-based linguistic features have a statistically significant effect on the emergence of discursive leadership in the online community reflecting the highest rated discursive contribution. I also found that among these lexical features the features with the most effect on discursive leadership emergence was the length of reviews. The richness of text in the reviews was found to be positively related to the emergence of leadership. Reviews with greater complexity were found to have a negative effect on leadership emergence.

To design a study focused on the lexical features of an online community, I eliminated online communities with the effects of social networking, eventually, I selected the online community involving reviews at Amazon.com. By using this online community, I have been able to analyze how the lexical features affect leadership without confounding interpersonal factors that often are the main contributors to leadership emergence in online communities.

8.2. Insights, Implications and Limitations.

Extant research on leadership in online communities has demonstrated that leaders emerging within online communities are imperative to the sustainability of the community, as they are crucial for the fostering of knowledge transfer within the community. While lexical features of discourse in online communities have been studied in literature focusing on online communities the findings on how lexical features independently effect the emergence of leaders have been overshadowed by the overwhelming influence of social networking factors. Most past research studies have shown that social networks have primary influence on the emergence of leadership in online communities suppressing the influence of other factors of the community to the minuscule level. By isolating lexical features from social networking factors, in my study, I have been able to gain deeper insight into how the discourse of users affects the ability of some of these to emerge as leaders within the online community. Therefore, in this dissertation I could examine how in online communities lexical features affect the emergence of leadership independent of the influence of social networks. My findings indicate which characteristics of discourse are impactful on the user's ability to emerge as leaders in the online community. By supporting these characteristics, online communities with no social networking interactions, will be able to continually nurture leaders which will sustain the community.

To frame my research study, I developed a cohesive definition of online communities. Research on online communities is replete with different definitions with broadly dispersed definitional terms. Due to the varied approaches to the study of online communities and the various types of communities studied authors have proposed multiple different definitions of an online community is. To frame my study, I have compiled definitions from past research and analyzed the definitional elements to develop a unified definition for online communities encompassing its self-organization, self-governing, and digital nature. Owing to this unified definition, future research programs examining online community will be facilitated.

With my study, I have extended Johnson et al's (2015) seminal study examining how leaders emerge in online communities owing to the lexical features of their discourse. The study also focused on social interactions as a contributing factor of leadership emergence in online communities, while my study has examined only how the lexical features of discourse affect the emergence of leaders in communities. Johnson et al. (2015) also removed a significant portion of data that was available in their reference community to normalize their data rather than using all of the data and accounting for the power law distribution. Many studies use statistical analysis techniques that do not completely capture the relationships found in power law distributed data (Bradley & Aguinis, 2023). In my study, I isolated discourse so that I could study it independently and thus furthered the body of knowledge for discursive leadership in online communities, as well as utilizing zero-inflated negative binomial regression to completely analyze the relationships found in the data due to the power law distribution.

By combining evolutionary signaling theory Faraj et al., (2011) and communication accommodation theory (CAT) Lu et al. (2022), I have built a theoretical basis to hypothesize lexical features as factors that influence the emergence of leadership in online communities. Specifically, I have developed a theory-based model that I tested empirically by examining the effect of theorized lexical features on leadership emergence.

I developed my model upon reviewing the literature and identifying a gap in the knowledge of how leaders emerge in online communities. Nearly all past research was primarily focused on social networking factors, which account for the majority of the statistical significance when studying leadership emergence in online communities. The lexical features of online community members' posts have been largely ignored as insignificant in comparison to the social networking factors, and therefore were used merely as control variables in these studies. Once I tested my theory-based model, I found statistical support for the theorized lexical features as factors of leadership emergence in online communities.

The data used for my study is unique relative to the data used in past research studies in that it comes from an online community that has no social networking features and provides an objective continuous dependent variable. By analyzing this data, I was able to determine that there are interactions between the variables in my theoretical model that moderate the effect of lexical features of text on the emergence of leadership.

My study has limitations that are introduced by limited metadata about the posts made in the online community sample used. For example, there is no information available if the posts of other users had been read before a user posted his or her review on a specific product. In addition, I have also not examined how often one user posts a review in conjunction with reviews from any other distinct reviewer. My study is also limited to one online community platform. Due to the nature of the data and obfuscation of user information, there is no way to contact users in the online community for further research on their demographic data, such as education level, age, and other biographical information which may affect the lexical features of their reviews in the online community. The data used in the study has been averaged to the reviewer level. Averaging data causes a potential for a loss of depth in the analysis.

8.3 Conclusions.

Throughout the history of the internet, we have seen multiple generations, Web 1.0, 2.0, and 3.0. Each of these generations of the internet has introduced new ways in which this technology is thought about and interacted with. In Web 1.0 the internet primarily consisted of static web pages that could not be interacted with but that provided a place where publishers of information could present their work. With the advent of Web 2.0 we saw the introduction of the interactive web. In Web 2.0 websites became places where people could interact with each other and form communities. The communities on the internet became more informative as they gained more users. As more users joined the online communities, the communities became more valuable to each user which in turn attracted more users. As technology has evolved, we have seen a shift on the internet from people interacting to a place where machines interact with each other, this generation of the internet is often called Web 3.0. The machine-to-machine interactions of Web 3.0 do not need human interaction or input. Using tools such as the Wayback Machine we can see examples of both Web 1.0 and Web 2.0, while Web 3.0 is generally invisible to us as there are no graphical representations of the machine-to-machine interactions.

As the internet evolved from Web 1.0 to Web 2.0 we saw a shift from it being a place where relatively few people were seen and heard as they had the skill to publish their pages, to a place where anyone can be seen and heard in the broad range of sites that allow for self-publishing. These sites created in Web 2.0 have allowed for mass computer mediated human interactions which allow for the sharing of information and opinions between users. In this environment users have shared vast amounts of knowledge with audiences that far surpass what would be possible without the online community.

The affordances provided by Web 2.0 have been the primary focus of research on online communities. Studies in the literature have examined online communities by focusing on the characteristics of their social structures. One of the primary features of that social structure is called centrality. Centrality is the distance between one user and the other users within the online community. By assessing centrality, researchers can determine the amount of time required for a post by one user to reach all other users within the online community. Coreness is a second commonly studied feature of the social structure of online communities. Coreness is the concept that there is a core group of users who are very active, however, they are not often interacting with the other users of the online community. Most of the research in online communities has focused on these features of social networks which are formed through interaction and dialogue within the community.

The study of social networks in online communities is valuable as these online communities have become a primary place for human interaction. These communities are places where knowledge is transferred between users. Recent research has focused on users of online communities that are good at sharing knowledge within the online community and how they emerge as leaders among the users. Seminal studies such as Johnson et al. (2015) have examined leadership emergence and found that the structural aspects of social networks have strong statistical significance when compared to the linguistic features of the posts in the community. However, the research in online communities has not focused on lexical features in the absence of social networks in order to isolate and fully examine how those lexical features affect the emergence of leadership in online communities.

Research on online community leadership has been almost exclusively focused on examining the effects of social networks, while communities are often conducive to social networking, discourse a likely significant factor when there are no threaded discussions of participants in online communities. In this dissertation, my review of literature on online communities identified a wide range of definitions of an online community, indicating that there be a synthesis of extant definitions into one integrated definition that can serve as a basis for future research. I have examined the definitional elements of the past definitions and proposed an integrated definition that fully captures the meaning of online communities with and without social networking of users.

I have also reviewed past research on online communities' leadership to determine whether the body of knowledge fully explains how leaders emerge. In my review, I found that virtually all the past research has focused on social networks and the role of social interactions on the emergence of leaders in online communities. While lexical variables are often included in these studies, they are included as an afterthought and analyzed as a control variable, while the social variables explain most of the variance in past studies. In my research, I have integrated evolutionary signaling theory with communication accommodation theory to propose a theoretical model of discourse constructs and how they affect leadership emergence in online communities without social networking. I have developed my model for using data from an online community with no social interactions to examine whether discourse plays a statistically significant role for the emergence of online leaders. My study is also the first study with a continuous dependent variable for measuring leadership emergence in an online community.

My findings will further the knowledge of online communities and the leadership within them as online community creators can find out how to support their users and encourage them to learn to communicate better so that they can emerge as leaders in the online community. My findings will provide a basis for future research on leadership in online communities using my integrated definition of online communities and my theoretical model for empirical tests.

8.4. Future Directions.

My literature review has examined the studies published in the most prestigious peer reviewed journals. My study provides further knowledge and opens future research by further studying and adding new input variables to the theoretical model testing in my proposed model. By expanding the model with my new lexical variables, researchers may be able to explain more of the variance in the dependent variable of leadership emergence.

While I conducted the empirical test of my theoretical model in only one online community (Amazon reviews), future research studies should examine the reviews in other online communities where there are minimal social networks. One potential online community for testing my model is the Yelp online community. While Yelp is an online review community, with limited social networking opportunities separate from the reviews. Yelp also allows for a wider range of reactions from readers of online reviews, thus providing richer data reported by the participants in the online community.

While my study used averaged data to focus on the reviewers, future studies should further explore using multi-level analysis methods to study how each review affects the emergence of leadership for the reviewer. While I focused on the reviewer and how they emerge as a leader by looking at their complete body of reviews, it is possible that there is much rich data that could be gleaned from looking at individual reviews. The individual reviews can then be analyzed using multi-level analysis to feed into the emergence of leadership.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abbasi, A., Zhou, Y., Deng, S., & Zhang, P. (2018). Text analytics to support sense-making in social media: A language-action perspective. *MIS Quarterly: Management Information Systems*, *42*(2), 427–464. https://doi.org/10.25300/MISQ/2018/13239

Agarwal, R., Gupta, A. K., & Kraut, R. (2008). The interplay between digital and social networks. *Information Systems Research*, *19*(3), 243–252. https://doi.org/10.1287/isre.1080.0200

Aguiton, C., & Cardon, D. (2007). The Strength of Weak Cooperation: an Attempt to Understand the Meaning of Web 2.0. *Communication & Strategies*, *65*(65), 51–65. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=0F396C11C3193C650D612309C65A2400?doi=10.1.1.99.3845&rep=rep1&type=pdf%5Cnhttp://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.99.3845%5Cnhttp://papers.ssrn.com/sol3/papers.cfm?abstract_id=1009070

Andersen, P. (2007). *What is Web 2.0?: ideas, technologies and implications for education*. http://21stcenturywalton.pbworks.com/f/What is Web 2.0.pdf

Bapna, R., Gupta, A., Rice, S., & Sundararajan, A. (2017). Trust and the strength of ties in online social networks: An exploratory field experiment. *MIS Quarterly: Management Information Systems*, *41*(1), 115–130. https://doi.org/10.25300/MISQ/2017/41.1.06

Bauer, J., Franke, N., & Tuertscher, P. (2016). Intellectual property norms in online communities: How user-organized intellectual property regulation supports innovation. *Information Systems Research*, *27*(4), 724–750. https://doi.org/10.1287/isre.2016.0649

Bayus, B. L. (2013). Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community. *Management Science*, *59*(1), 226–244.

https://doi.org/10.1287/mnsc.1120.1599

Bergquist, M., & Ljungberg, J. (2001). The power of gifts: Organizing social relationships in open source communities. *Information Systems Journal*, *11*(11), 305–320. https://doi.org/10.1046/j.1365-2575.2001.00111.x

Bradley, K. J., & Aguinis, H. (2023). Team Performance: Nature and Antecedents of Nonnormal Distributions. *Organization Science*, *34*(3), 1266–1286. https://doi.org/10.1287/orsc.2022.1619

Butler, B. S. (2001). Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures. *Information Systems Research*, *12*(4), 346–362. https://doi.org/10.1287/isre.12.4.346.9703

Butler, B. S., & Wang, X. (2012). The cross-purposes of cross-posting: Boundary reshaping behavior in online discussion communities. *Information Systems Research*, *23*(3 PART 2), 993–1010. https://doi.org/10.1287/isre.1110.0378

Campbell, J., Fletcher, G., & Greenhill, A. (2009). Conflict and identity shape shifting in an online financial community. *Information Systems Journal*, *19*(5), 461–478. https://doi.org/10.1111/j.1365-2575.2008.00301.x

Chua, H., Wareham, J., & Robey, D. (2007). The role of online trading communities in managing internet auction fraud. *MIS Quarterly: Management Information Systems*, *31*(4), 759–781. https://doi.org/10.2307/25148819

Dewan, S., Ho, Y.-J., & Ramaprasad, J. (2017). Popularity or proximity: Characterizing the nature of social influence in an online music community. *Information Systems Research*, *28*(1), 117–136. https://doi.org/10.1287/isre.2016.0654

DuBay, W. H. (2004). The principles of readability: A brief introduction to readability research. *Impact Information*, *949*, 1–72. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Principles+of+Readability#1

Eslami, S. P., Ghasemaghaei, M., & Hassanein, K. (2018). Which online reviews do consumers find most helpful? A multi-method investigation. *Decision Support Systems*, *113*(March), 32–42. https://doi.org/10.1016/j.dss.2018.06.012

Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. (2011). Knowledge collaboration in online communities. *Organization Science*, *22*(5), 1224–1239. https://doi.org/10.1287/orsc.1100.0614

Faraj, S., Kudaravalli, S., & Wasko, M. (2015). Leading collaboration in online communities. *MIS Quarterly: Management Information Systems*, *39*(2), 393–412. https://doi.org/10.25300/MISQ/2015/39.2.06

Faraj, Samer, von Krogh, G., Monteiro, E., & Lakhani, K. R. (2016). Online community as space for knowledge flows. *Information Systems Research*, *27*(4), 668–684. https://doi.org/10.1287/isre.2016.0682

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, *19*(3), 291–313. https://doi.org/10.1287/isre.1080.0193

Forman, Chris, Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, *19*(3), 291–313. https://doi.org/10.1287/isre.1080.0193

Greul, A., West, J., & Bock, S. (2018). Open at birth? Why new firms do (or don't) use open innovation. *Strategic Entrepreneurship Journal*, *12*(3), 392–420. https://doi.org/10.1002/sej.1282

He, R., & McAuley, J. (2016). Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. *25th International World Wide Web Conference, WWW 2016*, 507–517. https://doi.org/10.1145/2872427.2883037

Hiremath, B. K., & Kenchakkanavar, A. Y. (2016). An Alteration of the Web 1.0, Web 2.0 and Web 3.0: A Comparative Study. *Imperial Journal of Interdisciplinary Research, 2*(4), 2454–1362. http://www.imperialjournals.com/index.php/IJIR/article/view/327/320

Hoffman, L. (2022). *Generalized Linear Models for Count Outcomes and Zero Inflated Count Outcomes*.

Huffaker, D. A. (2009). *Dimensions of Leadership and Social Influence in Online Communities A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS for the degree DOCTOR OF PHILOSOPHY Field of Media , Technology and Society School of Communicat*. *December*, 1–159.

Hwang, E. H., Singh, P. V., & Argote, L. (2015). Knowledge sharing in online communities: Learning to cross geographic and hierarchical boundaries. *Organization Science*, *26*(6), 1593–1611. https://doi.org/10.1287/orsc.2015.1009

Johnson, S.L., Safadi, H., & Faraj, S. (2015). The emergence of online community leadership.

Information Systems Research, *26*(1), 165–187. https://doi.org/10.1287/isre.2014.0562

Johnson, Steven L., Safadi, H., & Faraj, S. (2015). The emergence of online community leadership. *Information Systems Research*, *26*(1), 165–187. https://doi.org/10.1287/isre.2014.0562

Kane, G. C., Johnson, J., & Majchrzak, A. (2014). Emergent life cycle: The tension between knowledge change and knowledge retention in open online coproduction communities. *Management Science*, *60*(12), 3026–3048. https://doi.org/10.1287/mnsc.2013.1855

Kincaid, J. P. J., Jr, R. F., Rogers, R. L., Chissom, B. S. B., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. B. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel (No. RBR-8-75). Naval Technical Training Command Millington TN Research Branch. *Naval Technical Training Command Millington TN Research Branch*. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA006655

Kreps, D., & Kimppa, K. (2015). Theorising Web 3.0: ICTs in a changing society. *Information Technology & People*.

Lee, J. Y. H., Yang, C. S., Hsu, C., & Wang, J. H. (2019). A longitudinal study of leader influence in sustaining an online community. *Information and Management*, *56*(2). https://doi.org/10.1016/j.im.2018.10.008

Levina, N., & Arriaga, M. (2014). Distinction and status production on user-generated content platforms: Using Bourdieu's theory of cultural production to understand social dynamics in online fields. *Information Systems Research*, *25*(3), 468–488. https://doi.org/10.1287/isre.2014.0535

Lu, X., Jiang, J., Head, M., & Yang, J. (2022). The Impact of Linguistic Complexity on Leadership in Online Q&A communities: Comparing Knowledge Shaping and Knowledge Adding. *Information and Management*, *59*(6), 103675. https://doi.org/10.1016/j.im.2022.103675

Lu, Y., Jerath, K., & Singh, P. V. (2013). The emergence of opinion leaders in a networked online community: A dyadic model with time dynamics and a heuristic for fast estimation. *Management Science*, *59*(8), 1783–1799. https://doi.org/10.1287/mnsc.1120.1685

Mayfield, J., & Mayfield, M. (2017). Leadership communication: Reflecting, engaging, and innovating. In *International Journal of Business Communication* (Vol. 54, Issue 1). https://doi.org/10.1177/2329488416675446

Mayzlin, D. (2006). Promotional chat on the internet. *Marketing Science*, *25*(2), 155–163. https://doi.org/10.1287/mksc.1050.0137

McAuley, J., Targett, C., Shi, Q., & Van Den Hengel, A. (2015). Image-based recommendations on styles and substitutes. *SIGIR 2015 - Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 43–52. https://doi.org/10.1145/2766462.2767755

Meng, M., & Agarwal, R. (2007). Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information Systems Research*, *18*(1), 42–67. https://doi.org/10.1287/isre.1070.0113

Meng, Ma, & Agarwal, R. (2007). Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information Systems Research*, *18*(1), 42–67. https://doi.org/10.1287/isre.1070.0113

Miller, K. D., Fabian, F., & Lin, S.-J. (2009). Strategies for online communities. *Strategic Management Journal*, *30*(3), 305–322. https://doi.org/10.1002/smj.735

Mohammad, S. M. (2016). Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text. *Emotion Measurement*, 201–237. https://doi.org/10.1016/B978-0-08-100508-8.00009-6

Nan, N., & Lu, Y. (2014). Harnessing the power of self-organization in an online community during organizational crisis. *MIS Quarterly: Management Information Systems*, *38*(4), 1135–1157. https://doi.org/10.25300/MISQ/2014/38.4.09

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *CEUR Workshop Proceedings*, *718*, 93–98.

O'Reilly, T. (2012). What is web 2.0?: Design patterns and business models for the next generation of software. *The Social Media Reader*, *4578*, 32–52.

Oh, O., Agrawal, M., & Rao, H. R. (2013). Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. *MIS Quarterly: Management Information Systems*, *37*(2), 407–426. https://doi.org/10.25300/MISQ/2013/37.2.05

Ray, S., Kim, S. S., & Morris, J. G. (2014). The central role of engagement in online communities. *Information Systems Research*, *25*(3), 528–546. https://doi.org/10.1287/isre.2014.0525

Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization Studies*, *28*(3), 377–408. https://doi.org/10.1177/0170840607076007

Riemer, K., Stieglitz, S., & Meske, C. (2015). From Top to Bottom: Investigating the Changing Role of Hierarchy in Enterprise Social Networks. *Business and Information Systems Engineering*, *57*(3). https://doi.org/10.1007/s12599-015-0375-3

Ross, D. A. R. (2007). Backstage with the knowledge boys and girls: Goffman and distributed agency in an organic online community. *Organization Studies*, *28*(3), 307–325. https://doi.org/10.1177/0170840607076000

Rullani, F., & Haefliger, S. (2013). The periphery on stage: The intra-organizational dynamics in online communities of creation. *Research Policy*, *42*(4), 941–953. https://doi.org/10.1016/j.respol.2012.10.008

Safadi, H., Johnson, S. L., & Faraj, S. (2021). Who contributes knowledge? Core-periphery tension in online innovation communities. *Organization Science*, *32*(3). https://doi.org/10.1287/orsc.2020.1364

Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, *81*, 30–40. https://doi.org/10.1016/j.dss.2015.10.006

Shepard, C. A. (2001). *Communication accommodation theory. The new handbook of language \& social psychology, ed. by WP Robinson and H. Giles*, *33--56*. New York: Wiley.

Silva, L., Goel, L., & Mousavidin, E. (2009). Exploring the dynamics of blog communities: The case of MetaFilter. *Information Systems Journal*, *19*(1), 55–81. https://doi.org/10.1111/j.1365-2575.2008.00304.x

Simmons, L. L., Mukhopadhyay, S., Conlon, S., & Yang, J. (2011). A computer aided content analysis of online reviews. *Journal of Computer Information Systems*, *52*(1), 43–55.

Sproull, L., Dutton, W., & Kiesler, S. (2007). Introduction to the special issue: Online communities. *Organization Studies*, *28*(3), 277–281. https://doi.org/10.1177/0170840607076013

Vaast, E., Safadi, H., Lapointe, L., & Negoita, B. (2017). Social media affordances for connective action: An examination of microblogging use during the Gulf of Mexico oil spill. *MIS Quarterly: Management Information Systems*, *41*(4). https://doi.org/10.25300/misq/2017/41.4.08

Wang, X., Butler, B. S., & Ren, Y. (2013). The impact of membership overlap on growth: An ecological competition view of online groups. *Organization Science*, *24*(2), 414–431. https://doi.org/10.1287/orsc.1120.0756

Wasko, M. M., & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *Social Capital & Knowledge Contribution SPECIAL*, *29*(1), 35–57. https://doi.org/10.2307/25148667

Wiertz, C., & De Ruyter, K. (2007). Beyond the call of duty: Why customers contribute to firm-hosted commercial online communities. *Organization Studies*, *28*(3), 347–376. https://doi.org/10.1177/0170840607076003

Williams, R. (2021). *Models for Count Outcomes*.

Wonseok, O., & Sangyong, J. (2007). Membership herding and network stability in the open source community: The ising perspective. *Management Science*, *53*(7), 1086–1101. https://doi.org/10.1287/mnsc.1060.0623

Xu, B., Xu, Z., & Li, D. (2016). Internet aggression in online communities: a contemporary deterrence perspective. *Information Systems Journal*, *26*(6), 641–667. https://doi.org/10.1111/isj.12077

Yan, L., & Tan, Y. (2014). Feeling blue? Go online: An empirical study of social support among patients. *Information Systems Research*, *25*(4), 690–709. https://doi.org/10.1287/isre.2014.0538

Zhang, C., Hahn, J., & De, P. (2013). Continued participation in online innovation communities: Does community response matter equally for everyone? *Information Systems Research*, *24*(4), 1112–1130. https://doi.org/10.1287/isre.2013.0485

APPENDIX

APPENDIX A – DEFINITIONS OF ONLY COMMUNITY

Year	Journal	Authors	Article	Definition of Online Community
2001	Information Systems Journal	Magnus Bergquist & Jan Ljungberg	The power of gifts: Organizing social relationships in open-source communities	The [online] community is primarily a loosely coupled network of individuals with no organizational forces in terms of economy or management that can force other individuals to behave in a certain way.
2001	Information Systems Research	Brian S. Butler	Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures	Online social structures use Internet-based e-mail and a server (i.e., a list server or listserv) to centrally maintain a mailing list that enables individuals to broadcast text messages to the other members.

2005	MISQ	Wasko & Faraj	Why should I share? Examining social capital and knowledge contribution in electronic networks of practice	A self-organizing, open activity system focused on a shared practice that exists primarily through computer-mediated communication.
2006	Marketing Science	Dina Mayzlin	Promotional Chat on the Internet	Online communities are defined as chat rooms and online bulletin boards.
2007	Organization Studies	Caroline Wiertz and Kode Ruyter	Beyond the Call of Duty: Why Customers Contribute to Firm-hosted Commercial Online Communities	Commercial online communities [are] firm-hosted online aggregations of customers who collectively co-produce and consume content about a commercial activity that is central to their interest by exchanging intangible resources.

2007	Management Science	Wonseok Oh, Sangyong Jeon	Membership Herding and Network Stability in the Open Source Community: The Ising Perspective	Virtual communities in which dynamic interactions take place in the pursuit of common goals.
2007	Information Systems Research	Meng Ma, Ritu Agarwal	Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities	Computer-mediated coordination and collaboration
2007	MISQ	Chua, Wareham, & Robey	The Role of Online Trading Communities in Managing Internet Auction Fraud	Online communities are loosely defined as online bulletin boards and online trading sites.

2007	Organization Studies	Sproull, Dutton, & Kiesler	Introduction to the Special Issue: Online Communities	Online communities consist of people who may be connected through ties outside the online environment or have no pre-existing ties and who may later establish offline relationships.
2007	Organization Studies	Ren, Kraut, and Kiesler	Applying Common Identity and Bond Theory to Design of Online Communities	An Internet-connected collective of people who interact over time around a shared purpose, interest, or need.
2007	Organization Studies	Drew and Ross	Backstage with the knowledge boys and girls: Goffman and distributed agency in an organic online community	Chat rooms, message boards, and announcement pages

2008	Information Systems Research	Forman, Ghose, and Wiesenfeld	Examining the Relationship Between Reviews and Sales	Voluntary collectivities whose members share a common interest or experience and who interact with one another primarily over the Internet
2008	Information Systems Research	Agarwal, Gupta, and Kraut	The interplay between digital and social networks	Digitally enabled social networks
2009	Information Systems Journal	Campbell, Fletcher, and Greenhill	Conflict and identity shape shifting in an online financial community	Participation and presence in an online forum
2009	Information Systems Journal	Silva, Goel, and Mousavidin	Exploring the dynamics of blog communities: The case of MetaFilter	Online blog communities

2009	Strategic Management Journal	Miller and Lin	Strategies for online communities	Online communities consist of people who engage in computer-supported social interaction
2012	Information Systems Research	Butler and Wang	The cross-purposes of cross-posting: Boundary reshaping behavior in online discussion communities	Discussion spaces that consist of arrangements that structure, constrain, and enable particular types of interaction among individuals
2013	Information Systems Research	Zhang, Hahn, and De	Continued participation in online innovation communities: Does community response matter equally for everyone?	Voluntary participation, the relatively free flow of information, and far less hierarchical control and coordination than seen in firm.

2013	Organization Science	Wang, Butler, and Ren	The impact of membership overlap on growth: An ecological competition view of online groups	Virtual spaces where globally distributed people can interact around a shared purpose
2013	Research Policy	Rullani and Haefliger	The periphery on stage: The intra-organizational dynamics in online communities of creation	Online communities that are composed of individuals and firms that share a common interest, a sense of belonging, a shared language, rules for participation and governance, mechanisms to manage intellectual property rights, and an explicit purpose for the cumulative creation of knowledge.

2014	Information Systems Research	Levina and Arriaga	Distinction and status production on user-generated content platforms: Using Bourdieu's theory of cultural production to understand social dynamics in online fields	A social space engaging agents in producing, evaluating, and consuming content online that is held together by a shared interest and a set of power relations among agents sharing this interest.
2014	Information Systems Research	Yan and Tan	Feeling blue? Go online: An empirical study of social support among patients	Virtual platforms to bring together patients with shared interests to communicate with and help each other
2014	Information Systems Research	Ray, Kim, and Morris	The central role of engagement in online communities	Discussion forums of strangers whose communal identity revolves around professional interests

2015	Organization Science	Hwang, Singh, and Argote	Knowledge sharing in online communities: Learning to cross geographic and hierarchical boundaries	A virtual space where information needs can be presented in the form of natural language
2015	MIS Quarterly: Management Information Systems	Faraj, Kudaravalli, and Wasko	Leading collaboration in online communities	Fluid objects where boundaries are ever changing, and roles are temporary, existing only in the moment.
2015	Organization Science	Hwang, Singh, and Argote	Knowledge sharing in online communities: Learning to cross geographic and hierarchical boundaries	OCs [Online communities] are collective spaces of knowledge flows characterized by a continuous morphing and are mutually constituted by digital technologies and participants

2015	MISQ	Faraj, Kudaravalli, Wasko	Leading collaboration in online communities	A distributed group of virtually connected individuals united by a common goal or purpose
2016	Information Systems Research	Bauer, Franke, and Tuertscher	Intellectual property norms in online communities: How user-organized intellectual property regulation supports innovation	"Social units with shared common values that bind them
2016	Information Systems Journal	Xu, Xu, and Li	Internet aggression in online communities: a contemporary deterrence perspective	Internally, distinguish them externally, and provide an
2016	Information Systems Research	Faraj, von Krogh, Monteiro, Lakhani	Online community as space for knowledge flows	Entity they can identify with"

2018	Strategic Entrepreneurship Journal	Greul, West, and Bock	Open at birth? Why new firms do (or don't) use open innovation	Social network communities

Variable	Obs	Mean	Std. Dev.	Min	Max
sumhelpful~s	1,334,844	2.566298	29.34843	0	23796
c_avgstarr~g	1,334,844	-2.35e-07	1.300857	-3.052494	.9475064
c_avgrevie~h	1,334,844	-2.76e-06	319.5201	-312.9004	4686.1
c_avgdiver~e	1,334,826	-1.41e-08	.1064707	-.7876675	.1888425
c_avgrichn~e	1,334,844	-1.95e-06	41.61022	-52.61651	955.7168
c_avgreadi~e	1,334,844	-1.61e-06	40.3036	-4680.088	143.1422
c_avgsenti~e	1,334,844	-1.62e-07	6.061622	-80.23277	151.7672
c_countasin	1,334,844	-3.76e-08	2.983771	-.6723243	1523.328
starBYlength	1,334,844	-43.17582	421.4948	-14264.61	4433.477
starBYdivs~e	1,334,826	.020892	.1425933	-.74632	2.063554
starBYrich~s	1,334,844	3.528229	54.20909	-2917.319	739.7342
starBYease	1,334,844	4.25579	56.01412	-4354.254	14285.94
starBYsent~t	1,334,844	3.121175	8.577187	-249.594	199.1226
lengthBYdiv	1,334,826	-25.13599	80.12096	-2968.01	171.0971
lengthBYri~s	1,334,844	-2882.92	10593.19	-228775.3	432109.6
lengthBYease	1,334,844	-6625.011	56754.31	-4388007	904853.2
lengthBYse~t	1,334,844	710.2037	6344.649	-335817.3	578551.8
divBYrichn~s	1,334,826	2.115846	5.332438	-160.3524	164.0021
divBYease	1,334,826	1.732496	8.737309	-867.8235	2356.261
divBYsenti~t	1,334,826	-.186166	1.054281	-61.45168	28.13679
richnessBY~e	1,334,844	185.4965	1328.866	-65822.66	245493.5
richnessBY~t	1,334,844	-20.80508	218.2761	-13602.49	8888.254
easeBYsent~t	1,334,844	-40.39206	583.4339	-92229.13	57913.5

APPENDIX B – STATISTICAL ANALYSIS STEP TABLES

Table 1

Table 2

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	1,334,826	.	-6413745	23	1.28e+07	1.28e+07

Note: N=Obs used in calculating BIC; see **[R] BIC note**.

Table 3

```
Generalized linear models                          No. of obs      =    1334826
Optimization     : ML                              Residual df     =    1334803
                                                   Scale parameter =          1
Deviance         =  1072130.786                    (1/df) Deviance =   .8032127
Pearson          =  12908204.96                    (1/df) Pearson  =   9.670494

Variance function: V(u) = u+(3.4393)u^2            [Neg. Binomial]
Link function    : g(u) = ln(u)                    [Log]

                                                   AIC             =   3.119538
Log likelihood   = -2081997.358                    BIC             =  -1.78e+07
```

sumhelpfulvotes	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
c_avgstarrating	-.1263275	.0018045	-70.01	0.000	-.1298642	-.1227908
c_avgreviewlength	.0035238	.0000223	157.71	0.000	.00348	.0035676
c_avgdiversityscore	-1.129381	.0529698	-21.32	0.000	-1.2332	-1.025562
c_avgrichnessscore	.0032553	.0001272	25.58	0.000	.0030059	.0035047
c_avgreadingease	-.0020258	.0000795	-25.47	0.000	-.0021817	-.0018699
c_avgsentimentscore	-.0212518	.0004369	-48.65	0.000	-.0221081	-.0203956
c_countasin	.2006909	.0011577	173.36	0.000	.1984219	.20296
starBYlength	.0000641	6.45e-06	9.94	0.000	.0000515	.0000768
starBYdivscore	-.1835895	.0219543	-8.36	0.000	-.2266191	-.1405598
starBYrichness	-.0008008	.0000445	-18.01	0.000	-.0008879	-.0007136
starBYease	-.0000494	.0000398	-1.24	0.215	-.0001274	.0000287
starBYsentiment	-.0031873	.000241	-13.23	0.000	-.0036596	-.002715
lengthBYdiv	.0077187	.0000838	92.15	0.000	.0075546	.0078829
lengthBYrichness	-4.79e-06	7.05e-07	-6.80	0.000	-6.17e-06	-3.41e-06
lengthBYease	1.07e-06	6.04e-08	17.74	0.000	9.53e-07	1.19e-06
lengthBYsentiment	-9.09e-06	5.27e-07	-17.25	0.000	-.0000101	-8.06e-06
divBYrichness	-.0223701	.0010087	-22.18	0.000	-.0243471	-.020393
divBYease	.0100984	.0006312	16.00	0.000	.0088613	.0113356
divBYsentiment	-.1489442	.0044508	-33.46	0.000	-.1576676	-.1402208
richnessBYease	-.0000586	2.53e-06	-23.18	0.000	-.0000635	-.0000536
richnessBYsentiment	.0002967	.0000131	22.62	0.000	.000271	.0003224
easeBYsentiment	-7.03e-06	3.63e-06	-1.93	0.053	-.0000141	9.75e-08
_cons	.6007447	.0025786	232.97	0.000	.5956907	.6057987

Note: Negative binomial parameter estimated via ML and treated as fixed once estimated.

Table 4

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	1,334,826	.	-2081997	23	4164041	4164319

Note: N=Obs used in calculating BIC; see **[R] BIC note**.

Table 5

```
Zero-inflated negative binomial regression      Number of obs     =  1,334,826
                                                Nonzero obs       =    547,978
                                                Zero obs          =    786,848

Inflation model = logit                         LR chi2(22)       =  225260.24
Log likelihood  =  -1462413                     Prob > chi2       =     0.0000
```

sumhelpfulvotes	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sumhelpfulvotes						
c_avgstarrating	-.0467454	.0015028	-31.11	0.000	-.0496908	-.0437999
c_avgreviewlength	.002346	.0000163	143.67	0.000	.002314	.002378
c_avgdiversityscore	-.0818348	.0435167	-1.88	0.060	-.1671259	.0034563
c_avgrichnessscore	.0008838	.0001019	8.67	0.000	.0006841	.0010835
c_avgreadingease	-.0011363	.0000654	-17.36	0.000	-.0012646	-.0010081
c_avgsentimentscore	-.0170849	.0003874	-44.10	0.000	-.0178442	-.0163255
c_countasin	.0871366	.0005411	161.05	0.000	.0860761	.0881971
starBYlength	.0000318	4.74e-06	6.69	0.000	.0000225	.0000411
starBYdivscore	-.1233828	.0189537	-6.51	0.000	-.1605315	-.0862342
starBYrichness	-.0005928	.0000411	-14.42	0.000	-.0006734	-.0005122
starBYease	-.0000377	.0000276	-1.36	0.173	-.0000918	.0000165
starBYsentiment	-.003551	.0001769	-20.08	0.000	-.0038976	-.0032043
lengthBYdiv	.0043118	.0000649	66.43	0.000	.0041846	.0044391
lengthBYrichness	3.85e-06	5.62e-07	6.84	0.000	2.74e-06	4.95e-06
lengthBYease	6.57e-07	4.58e-08	14.36	0.000	5.68e-07	7.47e-07
lengthBYsentiment	-5.91e-06	3.96e-07	-14.94	0.000	-6.69e-06	-5.14e-06
divBYrichness	-.0008912	.0008865	-1.01	0.315	-.0026287	.0008464
divBYease	.0077762	.0005249	14.81	0.000	.0067474	.0088051
divBYsentiment	-.1064825	.0037201	-28.62	0.000	-.1137739	-.0991912
richnessBYease	-.0000395	2.26e-06	-17.47	0.000	-.000044	-.0000351
richnessBYsentiment	.0002045	.0000123	16.57	0.000	.0001803	.0002287
easeBYsentiment	6.91e-06	2.52e-06	2.75	0.006	1.98e-06	.0000118
_cons	1.465166	.0020961	698.99	0.000	1.461057	1.469274
inflate						
sumhelpfulvotes	-66.96553	47381.06	-0.00	0.999	-92932.14	92798.21
_cons	33.44522	23636.5	0.00	0.999	-46293.25	46360.14
/lnalpha	-.0249201	.0020535	-12.14	0.000	-.0289449	-.0208954
alpha	.9753878	.002003			.97147	.9793214

```
Likelihood-ratio test of alpha=0: chibar2(01) =  5.7e+06 Pr>=chibar2 =  0.0000
Vuong test of zinb vs. standard negative binomial: z =  1782.50  Pr>z = 0.0000
```

Table 6

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	1,334,826	-1575043	-1462413	26	2924878	2925192

Note: N=Obs used in calculating BIC; see **[R] BIC note**.

Table 7

```
Zero-inflated negative binomial regression      Number of obs     =  1,334,826
                                                Nonzero obs       =    547,978
                                                Zero obs          =    786,848

Inflation model = logit                         LR chi2(7)        =  204640.99
Log likelihood  =  -1472722                     Prob > chi2       =     0.0000
```

sumhelpfulvotes	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sumhelpfulvotes						
c_avgstarrating	-.0359247	.0012717	-28.25	0.000	-.0384172	-.0334322
c_avgreviewlength	.0009058	7.67e-06	118.15	0.000	.0008907	.0009208
c_avgdiversityscore	-1.535832	.0271565	-56.55	0.000	-1.589058	-1.482606
c_avgrichnessscore	.0007693	.0000524	14.69	0.000	.0006666	.000872
c_avgreadingease	-.0016643	.0000381	-43.72	0.000	-.0017389	-.0015896
c_avgsentimentscore	-.009215	.0002782	-33.13	0.000	-.0097602	-.0086698
c_countasin	.0970626	.0005544	175.09	0.000	.095976	.0981491
_cons	1.387852	.0016428	844.82	0.000	1.384632	1.391071
inflate						
sumhelpfulvotes	-67.51809	53976.59	-0.00	0.999	-105859.7	105724.6
_cons	33.74779	27354.05	0.00	0.999	-53579.21	53646.71
/lnalpha	.0129434	.0020298	6.38	0.000	.0089651	.0169216
alpha	1.013028	.0020562			1.009005	1.017066

```
Likelihood-ratio test of alpha=0: chibar2(01) =  6.1e+06 Pr>=chibar2 =  0.0000
Vuong test of zinb vs. standard negative binomial: z =  1901.32  Pr>z = 0.0000
```

Table 8

Measures of Fit for zinb of sumhelpfulvotes

Log-Lik Intercept Only:	-2.239e+06	Log-Lik Full Model:	-1.473e+06
D(1334815):	2945444.955	LR(8):	1532727.625
		Prob > LR:	0.000
McFadden's R2:	0.342	McFadden's Adj R2:	0.342
ML (Cox-Snell) R2:	0.683	Cragg-Uhler(Nagelkerke) R2:	0.708
AIC:	2.207	AIC*n:	2945466.955
BIC:	-1.588e+07	BIC':	-1.533e+06
BIC used by Stata:	2945600.102	AIC used by Stata:	2945466.955

Table 9

```
Zero-inflated negative binomial regression      Number of obs     =  1,334,826
                                                Nonzero obs       =    547,978
                                                Zero obs          =    786,848

Inflation model = logit                         LR chi2(8)        =  217447.16
Log likelihood  =  -1466319                     Prob > chi2       =     0.0000
```

sumhelpfulvotes	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sumhelpfulvotes						
c_avgstarrating	-.0361044	.0012572	-28.72	0.000	-.0385684	-.0336403
c_avgreviewlength	.0017194	.0000107	160.09	0.000	.0016984	.0017405
c_avgdiversityscore	-1.431047	.025947	-55.15	0.000	-1.481902	-1.380192
c_avgrichnessscore	.004077	.0000607	67.20	0.000	.0039581	.0041959
c_avgreadingease	-.0012718	.0000363	-35.00	0.000	-.001343	-.0012006
c_avgsentimentscore	-.0080354	.0002725	-29.49	0.000	-.0085694	-.0075013
c_countasin	.0899569	.0005399	166.63	0.000	.0888988	.0910151
lengthBYrichness	.0000299	2.68e-07	111.65	0.000	.0000293	.0000304
_cons	1.444814	.001723	838.54	0.000	1.441437	1.448191
inflate						
sumhelpfulvotes	-67.57019	54389.27	-0.00	0.999	-106668.6	106533.4
_cons	33.79989	28159.65	0.00	0.999	-55158.09	55225.69
/lnalpha	-.0108596	.0020452	-5.31	0.000	-.0148681	-.0068511
alpha	.9891991	.0020231			.9852419	.9931723

```
Likelihood-ratio test of alpha=0: chibar2(01) =  6.1e+06 Pr>=chibar2 =  0.0000
Vuong test of zinb vs. standard negative binomial: z =  1840.02  Pr>z = 0.0000
```

Table 10

```
Measures of Fit for zinb of sumhelpfulvotes

Log-Lik Intercept Only:     -2.239e+06   Log-Lik Full Model:         -1.466e+06
D(1334814):                2932638.784   LR(9):                     1545533.796
                                         Prob > LR:                       0.000
McFadden's R2:                   0.345   McFadden's Adj R2:               0.345
ML (Cox-Snell) R2:               0.686   Cragg-Uhler(Nagelkerke) R2:      0.711
AIC:                             2.197   AIC*n:                     2932662.784
BIC:                        -1.589e+07   BIC':                       -1.545e+06
BIC used by Stata:         2932808.036   AIC used by Stata:         2932662.784
```

Table 11

Zero-inflated negative binomial regression

Number of obs	=	1,334,826
Nonzero obs	=	547,978
Zero obs	=	786,848
LR chi2(10)	=	221843.88
Prob > chi2	=	0.0000

Inflation model = logit
Log likelihood = -1464121

sumhelpfulvotes	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sumhelpfulvotes						
c_avgstarrating	-.0347709	.0012493	-27.83	0.000	-.0372195	-.0323224
c_avgreviewlength	.0023183	.0000146	159.26	0.000	.0022898	.0023468
c_avgdiversityscore	.0858548	.0342873	2.50	0.012	.018653	.1530566
c_avgrichnessscore	.0008372	.0000751	11.14	0.000	.0006899	.0009844
c_avgreadingease	-.0014152	.0000466	-30.39	0.000	-.0015065	-.001324
c_avgsentimentscore	-.0079694	.0002678	-29.76	0.000	-.0084943	-.0074445
c_countasin	.0874731	.0005353	163.40	0.000	.0864239	.0885223
lengthBYdiv	.0035709	.0000585	61.00	0.000	.0034562	.0036856
lengthBYrichness	9.22e-06	4.00e-07	23.03	0.000	8.43e-06	.00001
lengthBYease	4.05e-07	2.63e-08	15.43	0.000	3.54e-07	4.57e-07
_cons	1.465988	.0017465	839.38	0.000	1.462565	1.469411
inflate						
sumhelpfulvotes	-67.59776	54609.94	-0.00	0.999	-107101.1	106965.9
_cons	33.82746	28583.53	0.00	0.999	-55988.86	56056.52
/lnalpha	-.0194465	.0020513	-9.48	0.000	-.0234668	-.0154261
alpha	.9807414	.0020118			.9768064	.9846923

Likelihood-ratio test of alpha=0: chibar2(01) = 5.8e+06 Pr>=chibar2 = 0.0000
Vuong test of zinb vs. standard negative binomial: z = 1822.45 Pr>z = 0.0000

Table 12

Measures of Fit for zinb of sumhelpfulvotes

Log-Lik Intercept Only:	-2.239e+06	Log-Lik Full Model:	-1.464e+06
D(1334812):	2928242.065	LR(11):	1549930.514
		Prob > LR:	0.000
McFadden's R2:	0.346	McFadden's Adj R2:	0.346
ML (Cox-Snell) R2:	0.687	Cragg-Uhler(Nagelkerke) R2:	0.712
AIC:	2.194	AIC*n:	2928270.065
BIC:	-1.590e+07	BIC':	-1.550e+06
BIC used by Stata:	2928439.526	AIC used by Stata:	2928270.065

Table 13

```
Zero-inflated negative binomial regression      Number of obs     =  1,334,826
                                                Nonzero obs       =    547,978
                                                Zero obs          =    786,848

Inflation model = logit                         LR chi2(11)       =  222020.52
Log likelihood  =  -1464033                     Prob > chi2       =     0.0000
```

sumhelpfulvotes	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sumhelpfulvotes						
c_avgstarrating	-.0312888	.0012765	-24.51	0.000	-.0337906	-.0287869
c_avgreviewlength	.0023415	.0000146	160.05	0.000	.0023129	.0023702
c_avgdiversityscore	.0725031	.0342553	2.12	0.034	.005364	.1396422
c_avgrichnessscore	.0008072	.000075	10.76	0.000	.0006602	.0009541
c_avgreadingease	-.0014451	.0000467	-30.93	0.000	-.0015366	-.0013535
c_avgsentimentscore	-.0102417	.000318	-32.21	0.000	-.0108649	-.0096185
c_countasin	.0874592	.0005352	163.40	0.000	.0864102	.0885083
lengthBYdiv	.0038045	.0000606	62.80	0.000	.0036857	.0039232
lengthBYrichness	8.85e-06	4.00e-07	22.14	0.000	8.07e-06	9.63e-06
lengthBYease	4.26e-07	2.68e-08	15.90	0.000	3.73e-07	4.78e-07
lengthBYsentiment	3.25e-06	2.47e-07	13.16	0.000	2.77e-06	3.74e-06
_cons	1.467714	.0017493	839.02	0.000	1.464285	1.471143
inflate						
sumhelpfulvotes	-67.60104	54637.31	-0.00	0.999	-107154.8	107019.6
_cons	33.83074	28635.78	0.00	0.999	-56091.26	56158.92
/lnalpha	-.019774	.0020515	-9.64	0.000	-.0237948	-.0157532
alpha	.9804202	.0020113			.9764861	.9843702

```
Likelihood-ratio test of alpha=0: chibar2(01) =  5.8e+06 Pr>=chibar2 =  0.0000
Vuong test of zinb vs. standard negative binomial: z =  1819.13  Pr>z = 0.0000
```

Table 14

Measures of Fit for zinb of sumhelpfulvotes

Log-Lik Intercept Only:	-2.239e+06	Log-Lik Full Model:	-1.464e+06
D(1334811):	2928065.427	LR(12):	1550107.153
		Prob > LR:	0.000
McFadden's R2:	0.346	McFadden's Adj R2:	0.346
ML (Cox-Snell) R2:	0.687	Cragg-Uhler(Nagelkerke) R2:	0.712
AIC:	2.194	AIC*n:	2928095.427
BIC:	-1.590e+07	BIC':	-1.550e+06
BIC used by Stata:	2928276.992	AIC used by Stata:	2928095.427

Table 15

```
Zero-inflated negative binomial regression      Number of obs     =  1,334,826
                                                Nonzero obs       =    547,978
                                                Zero obs          =    786,848

Inflation model = logit                         LR chi2(12)       =  222043.02
Log likelihood  =  -1464021                     Prob > chi2       =     0.0000
```

sumhelpfulvotes	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sumhelpfulvotes						
c_avgstarrating	-.0311244	.001277	-24.37	0.000	-.0336272	-.0286216
c_avgreviewlength	.0023752	.0000163	146.03	0.000	.0023433	.0024071
c_avgdiversityscore	.1977444	.0432492	4.57	0.000	.1129775	.2825114
c_avgrichnessscore	.0004913	.0001003	4.90	0.000	.0002948	.0006879
c_avgreadingease	-.0014395	.0000467	-30.80	0.000	-.0015311	-.0013479
c_avgsentimentscore	-.0102446	.0003179	-32.23	0.000	-.0108677	-.0096215
c_countasin	.0878012	.0005409	162.33	0.000	.0867411	.0888614
lengthBYdiv	.0037699	.0000611	61.70	0.000	.0036501	.0038896
lengthBYrichness	.0000103	5.12e-07	20.20	0.000	9.34e-06	.0000114
lengthBYease	4.21e-07	2.67e-08	15.77	0.000	3.69e-07	4.73e-07
lengthBYsentiment	3.29e-06	2.46e-07	13.35	0.000	2.81e-06	3.77e-06
divBYrichness	.0041139	.0008686	4.74	0.000	.0024114	.0058164
_cons	1.463045	.0020066	729.11	0.000	1.459112	1.466978
inflate						
sumhelpfulvotes	-66.96259	47351.44	-0.00	0.999	-92874.07	92740.15
_cons	33.44229	23577.06	0.00	0.999	-46176.75	46243.63
/lnalpha	-.0197649	.0020514	-9.63	0.000	-.0237856	-.0157442
alpha	.9804292	.0020112			.9764951	.9843791

```
Likelihood-ratio test of alpha=0: chibar2(01) =  5.7e+06 Pr>=chibar2 =  0.0000
Vuong test of zinb vs. standard negative binomial: z =  1807.66  Pr>z = 0.0000
```

Table 16

Measures of Fit for zinb of sumhelpfulvotes

Log-Lik Intercept Only:	-2.239e+06	Log-Lik Full Model:	-1.464e+06
D(1334810):	2928042.920	LR(13):	1550129.660
		Prob > LR:	0.000
McFadden's R2:	0.346	McFadden's Adj R2:	0.346
ML (Cox-Snell) R2:	0.687	Cragg-Uhler(Nagelkerke) R2:	0.712
AIC:	2.194	AIC*n:	2928074.920
BIC:	-1.590e+07	BIC':	-1.550e+06
BIC used by Stata:	2928268.589	AIC used by Stata:	2928074.920

Table 17

```
Zero-inflated negative binomial regression      Number of obs     =  1,334,826
                                                Nonzero obs       =    547,978
                                                Zero obs          =    786,848

Inflation model = logit                         LR chi2(13)       =  222212.65
Log likelihood  =  -1463937                     Prob > chi2       =     0.0000
```

sumhelpfulvotes	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sumhelpfulvotes						
c_avgstarrating	-.0311074	.0012768	-24.36	0.000	-.0336099	-.0286049
c_avgreviewlength	.0024049	.0000164	146.73	0.000	.0023728	.0024371
c_avgdiversityscore	.1694237	.0432686	3.92	0.000	.0846189	.2542286
c_avgrichnessscore	.0006631	.0001011	6.56	0.000	.0004649	.0008613
c_avgreadingease	-.0008446	.0000649	-13.02	0.000	-.0009718	-.0007174
c_avgsentimentscore	-.0104912	.0003179	-33.01	0.000	-.0111142	-.0098683
c_countasin	.0877264	.0005409	162.18	0.000	.0866662	.0887865
lengthBYdiv	.0038702	.0000611	63.31	0.000	.0037504	.00399
lengthBYrichness	.0000104	5.11e-07	20.32	0.000	9.38e-06	.0000114
lengthBYease	8.66e-07	4.17e-08	20.78	0.000	7.85e-07	9.48e-07
lengthBYsentiment	3.57e-06	2.42e-07	14.78	0.000	3.10e-06	4.05e-06
divBYrichness	.0029247	.0008713	3.36	0.001	.0012171	.0046324
divBYease	.0055622	.0004247	13.10	0.000	.0047298	.0063945
_cons	1.461751	.0020086	727.76	0.000	1.457815	1.465688
inflate						
sumhelpfulvotes	-66.96304	47352.38	-0.00	0.999	-92875.92	92742
_cons	33.44274	23578.96	0.00	0.999	-46180.46	46247.35
/lnalpha	-.0198281	.0020512	-9.67	0.000	-.0238483	-.0158079
alpha	.9803672	.0020109			.9764338	.9843163

```
Likelihood-ratio test of alpha=0: chibar2(01) =  5.7e+06 Pr>=chibar2 =  0.0000
Vuong test of zinb vs. standard negative binomial: z =  1807.52  Pr>z = 0.0000
```

Table 18

Measures of Fit for zinb of sumhelpfulvotes

Log-Lik Intercept Only:	-2.239e+06	Log-Lik Full Model:	-1.464e+06
D(1334809):	2927873.296	LR(14):	1550299.283
		Prob > LR:	0.000
McFadden's R2:	0.346	McFadden's Adj R2:	0.346
ML (Cox-Snell) R2:	0.687	Cragg-Uhler(Nagelkerke) R2:	0.712
AIC:	2.193	AIC*n:	2927907.296
BIC:	-1.590e+07	BIC':	-1.550e+06
BIC used by Stata:	2928113.070	AIC used by Stata:	2927907.296

Table 19

Zero-inflated negative binomial regression

		Number of obs	= 1,334,826
		Nonzero obs	= 547,978
		Zero obs	= 786,848
Inflation model = logit		LR chi2(14)	= 223641.12
Log likelihood = -1463222		Prob > chi2	= 0.0000

sumhelpfulvotes	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sumhelpfulvotes						
c_avgstarrating	-.0252163	.0012854	-19.62	0.000	-.0277357	-.0226969
c_avgreviewlength	.0023686	.0000164	144.86	0.000	.0023365	.0024006
c_avgdiversityscore	-.0182375	.043516	-0.42	0.675	-.1035273	.0670524
c_avgrichnessscore	.0009857	.0001015	9.71	0.000	.0007868	.0011846
c_avgreadingease	-.0010588	.0000655	-16.16	0.000	-.0011872	-.0009304
c_avgsentimentscore	-.0183394	.0003787	-48.42	0.000	-.0190817	-.0175971
c_countasin	.0879051	.0005415	162.34	0.000	.0868438	.0889664
lengthBYdiv	.0038638	.0000607	63.68	0.000	.0037448	.0039827
lengthBYrichness	9.66e-06	5.08e-07	19.00	0.000	8.66e-06	.0000107
lengthBYease	5.64e-07	4.26e-08	13.23	0.000	4.80e-07	6.47e-07
lengthBYsentiment	-6.86e-06	3.51e-07	-19.55	0.000	-7.55e-06	-6.17e-06
divBYrichness	-.0004053	.0008726	-0.46	0.642	-.0021155	.001305
divBYease	.0027707	.0004363	6.35	0.000	.0019155	.0036259
divBYsentiment	-.1052093	.0027832	-37.80	0.000	-.1106642	-.0997544
_cons	1.454012	.0020155	721.43	0.000	1.450062	1.457962
inflate						
sumhelpfulvotes	-66.96338	47359.85	-0.00	0.999	-92890.57	92756.64
_cons	33.44308	23593.95	0.00	0.999	-46209.86	46276.74
/lnalpha	-.0221282	.0020521	-10.78	0.000	-.0261502	-.0181062
alpha	.9781148	.0020072			.9741887	.9820567

Likelihood-ratio test of alpha=0: chibar2(01) = 5.7e+06 Pr>=chibar2 = 0.0000
Vuong test of zinb vs. standard negative binomial: z = 1797.23 Pr>z = 0.0000

Table 20

Measures of Fit for zinb of sumhelpfulvotes

Log-Lik Intercept Only:	-2.239e+06	Log-Lik Full Model:	-1.463e+06
D(1334808):	2926444.826	LR(15):	1551727.754
		Prob > LR:	0.000
McFadden's R2:	0.347	McFadden's Adj R2:	0.347
ML (Cox-Snell) R2:	0.687	Cragg-Uhler(Nagelkerke) R2:	0.712
AIC:	2.192	AIC*n:	2926480.826
BIC:	-1.590e+07	BIC':	-1.552e+06
BIC used by Stata:	2926698.704	AIC used by Stata:	2926480.826

Table 21

```
Zero-inflated negative binomial regression      Number of obs     =  1,334,826
                                                Nonzero obs       =    547,978
                                                Zero obs          =    786,848

Inflation model = logit                         LR chi2(15)       =  223964.04
Log likelihood  =  -1463061                     Prob > chi2       =     0.0000
```

sumhelpfulvotes	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sumhelpfulvotes						
c_avgstarrating	-.0255177	.0012854	-19.85	0.000	-.0280371	-.0229983
c_avgreviewlength	.0023521	.0000163	144.21	0.000	.0023201	.002384
c_avgdiversityscore	-.0324616	.0434412	-0.75	0.455	-.1176049	.0526816
c_avgrichnessscore	.0009273	.0001012	9.16	0.000	.000729	.0011257
c_avgreadingease	-.0011206	.0000651	-17.23	0.000	-.0012481	-.0009931
c_avgsentimentscore	-.0180976	.000379	-47.75	0.000	-.0188404	-.0173548
c_countasin	.087909	.0005415	162.34	0.000	.0868477	.0889704
lengthBYdiv	.0042842	.0000642	66.77	0.000	.0041585	.00441
lengthBYrichness	4.97e-06	5.65e-07	8.79	0.000	3.86e-06	6.07e-06
lengthBYease	7.00e-07	4.26e-08	16.45	0.000	6.17e-07	7.84e-07
lengthBYsentiment	-6.32e-06	3.51e-07	-18.03	0.000	-7.01e-06	-5.64e-06
divBYrichness	-.0021844	.0008819	-2.48	0.013	-.0039128	-.0004559
divBYease	.0078738	.0005162	15.25	0.000	.0068622	.0088855
divBYsentiment	-.1010751	.0027923	-36.20	0.000	-.1065478	-.0956024
richnessBYease	-.0000405	2.27e-06	-17.83	0.000	-.0000449	-.000036
_cons	1.453954	.0020171	720.81	0.000	1.45	1.457907
inflate						
sumhelpfulvotes	-66.96333	47360.93	-0.00	0.999	-92892.69	92758.76
_cons	33.44303	23596.13	0.00	0.999	-46214.12	46281
/lnalpha	-.0226707	.0020523	-11.05	0.000	-.0266932	-.0186483
alpha	.9775843	.0020063			.97366	.9815245

```
Likelihood-ratio test of alpha=0: chibar2(01) =  5.7e+06 Pr>=chibar2 =  0.0000
Vuong test of zinb vs. standard negative binomial: z =  1794.26  Pr>z = 0.0000
```

Table 22

Measures of Fit for zinb of sumhelpfulvotes

Log-Lik Intercept Only:	-2.239e+06	Log-Lik Full Model:	-1.463e+06
D(1334807):	2926121.904	LR(16):	1552050.675
		Prob > LR:	0.000
McFadden's R2:	0.347	McFadden's Adj R2:	0.347
ML (Cox-Snell) R2:	0.687	Cragg-Uhler(Nagelkerke) R2:	0.712
AIC:	2.192	AIC*n:	2926159.904
BIC:	-1.590e+07	BIC':	-1.552e+06
BIC used by Stata:	2926389.886	AIC used by Stata:	2926159.904

Table 23

```
Zero-inflated negative binomial regression      Number of obs     =  1,334,826
                                                Nonzero obs       =    547,978
                                                Zero obs          =    786,848

Inflation model = logit                         LR chi2(16)       =  224055.50
Log likelihood  =  -1463015                     Prob > chi2       =     0.0000
```

sumhelpfulvotes	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
sumhelpfulvotes						
c_avgstarrating	-.0251041	.0012861	-19.52	0.000	-.0276249	-.0225834
c_avgreviewlength	.0023495	.0000163	144.18	0.000	.0023175	.0023814
c_avgdiversityscore	-.0521046	.0434525	-1.20	0.230	-.1372698	.0330607
c_avgrichnessscore	.0010242	.0001016	10.08	0.000	.000825	.0012234
c_avgreadingease	-.0011578	.0000651	-17.78	0.000	-.0012855	-.0010302
c_avgsentimentscore	-.0184875	.0003813	-48.49	0.000	-.0192347	-.0177402
c_countasin	.0879543	.0005417	162.38	0.000	.0868927	.0890159
lengthBYdiv	.0043335	.000064	67.72	0.000	.0042081	.0044589
lengthBYrichness	4.48e-06	5.63e-07	7.96	0.000	3.38e-06	5.59e-06
lengthBYease	6.58e-07	4.28e-08	15.38	0.000	5.74e-07	7.42e-07
lengthBYsentiment	-7.22e-06	3.62e-07	-19.98	0.000	-7.93e-06	-6.52e-06
divBYrichness	-.0019743	.0008804	-2.24	0.025	-.0036999	-.0002487
divBYease	.0073879	.0005173	14.28	0.000	.0063739	.0084019
divBYsentiment	-.1201163	.0034286	-35.03	0.000	-.1268362	-.1133965
richnessBYease	-.00004	2.26e-06	-17.68	0.000	-.0000444	-.0000355
richnessBYsentiment	.0001022	.0000107	9.58	0.000	.0000813	.0001231
_cons	1.45286	.0020193	719.50	0.000	1.448903	1.456818
inflate						
sumhelpfulvotes	-66.96327	47360.98	-0.00	0.999	-92892.79	92758.86
_cons	33.44297	23596.23	0.00	0.999	-46214.32	46281.2
/lnalpha	-.0227404	.0020522	-11.08	0.000	-.0267627	-.0187182
alpha	.9775162	.0020061			.9735923	.9814559

```
Likelihood-ratio test of alpha=0: chibar2(01) =  5.7e+06 Pr>=chibar2 =  0.0000
Vuong test of zinb vs. standard negative binomial: z =  1792.91  Pr>z = 0.0000
```

Table 24

```
Measures of Fit for zinb of sumhelpfulvotes

Log-Lik Intercept Only:     -2.239e+06   Log-Lik Full Model:          -1.463e+06
D(1334806):                2926030.437   LR(17):                     1552142.142
                                         Prob > LR:                        0.000
McFadden's R2:                   0.347   McFadden's Adj R2:                0.347
ML (Cox-Snell) R2:               0.687   Cragg-Uhler(Nagelkerke) R2:       0.712
AIC:                             2.192   AIC*n:                      2926070.437
BIC:                        -1.590e+07   BIC':                        -1.552e+06
BIC used by Stata:         2926312.524   AIC used by Stata:          2926070.437
```

Table 26

```
Zero-inflated negative binomial regression      Number of obs     =  1,334,826
                                                Nonzero obs       =    547,978
                                                Zero obs          =    786,848

Inflation model = logit                         LR chi2(17)       =  224062.01
Log likelihood  =  -1463012                     Prob > chi2       =     0.0000
```

sumhelpfulvotes	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sumhelpfulvotes						
c_avgstarrating	-.0251336	.0012862	-19.54	0.000	-.0276544	-.0226127
c_avgreviewlength	.0023491	.0000163	144.19	0.000	.0023172	.0023811
c_avgdiversityscore	-.0537334	.0434576	-1.24	0.216	-.1389088	.031442
c_avgrichnessscore	.0010229	.0001016	10.07	0.000	.0008238	.0012221
c_avgreadingease	-.0011514	.0000652	-17.67	0.000	-.0012791	-.0010237
c_avgsentimentscore	-.0184541	.0003815	-48.38	0.000	-.0192018	-.0177064
c_countasin	.0879615	.0005417	162.39	0.000	.0868999	.0890232
lengthBYdiv	.0043517	.0000643	67.70	0.000	.0042257	.0044777
lengthBYrichness	4.46e-06	5.63e-07	7.93	0.000	3.36e-06	5.56e-06
lengthBYease	6.21e-07	4.56e-08	13.62	0.000	5.31e-07	7.10e-07
lengthBYsentiment	-6.99e-06	3.74e-07	-18.69	0.000	-7.73e-06	-6.26e-06
divBYrichness	-.0019616	.0008804	-2.23	0.026	-.0036871	-.000236
divBYease	.0074352	.0005184	14.34	0.000	.0064192	.0084512
divBYsentiment	-.1206913	.0034406	-35.08	0.000	-.1274348	-.1139479
richnessBYease	-.00004	2.26e-06	-17.70	0.000	-.0000444	-.0000356
richnessBYsentiment	.0001028	.0000107	9.63	0.000	.0000819	.0001237
easeBYsentiment	6.12e-06	2.40e-06	2.55	0.011	1.43e-06	.0000108
_cons	1.452915	.0020192	719.55	0.000	1.448957	1.456872
inflate						
sumhelpfulvotes	-66.96335	47361.38	-0.00	0.999	-92893.57	92759.64
_cons	33.44305	23597.03	0.00	0.999	-46215.89	46282.77
/lnalpha	-.0227618	.0020522	-11.09	0.000	-.0267841	-.0187395
alpha	.9774953	.0020061			.9735714	.981435

```
Likelihood-ratio test of alpha=0: chibar2(01) =  5.7e+06 Pr>=chibar2 =  0.0000
Vuong test of zinb vs. standard negative binomial: z =  1792.91  Pr>z = 0.0000
```

Table 27

```
Measures of Fit for zinb of sumhelpfulvotes

Log-Lik Intercept Only:      -2.239e+06   Log-Lik Full Model:          -1.463e+06
D(1334805):                 2926023.930   LR(18):                     1552148.650
                                          Prob > LR:                        0.000
McFadden's R2:                    0.347   McFadden's Adj R2:                0.347
ML (Cox-Snell) R2:                0.687   Cragg-Uhler(Nagelkerke) R2:       0.712
AIC:                              2.192   AIC*n:                      2926065.930
BIC:                         -1.590e+07   BIC':                        -1.552e+06
BIC used by Stata:          2926320.120   AIC used by Stata:          2926065.930
```

Table 28

APPENDIX C – COMPARISON OF DIAZ, 2022, LU ET AL., 2022, AND JOHNSON ET AL., 2015

Author	**Diaz, 2023**	**Lu et al., 2022**	**Johnson et al., 2015**
Theory	Evolutionary Signaling Theory and Communication Accommodation Theory	Communication Accommodation Theory	Functional Leadership Theory, Leader-Member Exchange Theory, Shared Leadership Theory, and Communication as Constitutive of Organizing
Dependent Variable	Continuous Objective	Continuous Objective	Ordinal Subjective
Outcome Variable	Online Leadership	Online Leadership	Online Leadership
Independent Variable Focus	Lexical Features	Lexical and Social Features	Lexical and Social Features
Distribution	Power Law	Poisson	Normal Distribution
Analysis Method	Zero-Inflated Negative Binomial	Negative Binomial	Logistic Regression

Table 1

VITA

Chad P. Diaz II

Ph.D in Management Information Systems

The University of Mississippi

chadpdiazii@me.com

Journal Articles:

Lugar, C., Meuser, J. D., Novicevic, M. M., Johnson, P. D., Ammeter, A. P., & Diaz, C. P. (2023). Retaining Self-initiated Expatriates: Systematic Reviews and Managerial Practices. In Research in Personnel and Human Resources Management (pp. 93-125). Emerald Publishing Limited.

Academic Cases:

Diaz, C., Novicevic, M. M., & Popoola, I. T. (2021). Putting Bitter for Sweet. SAGE Publications: SAGE Business Cases Originals

Refereed Conference Proceedings:

Diaz, Chad, II, & Conlon, Sumali (2017). “Analyzing Crib and Related Product Safety Through Text Analysis of Online Reviews.” Decision Sciences Institute Conference, Little Rock, AR, March (8-10).

Conlon, Sumali & Diaz II, Chad (2017). “Do They Say the Same Thing Forever? - Tracking Long Term Consumers’ Opinions of Products Through Analysis of Online Reviews.” Decision Sciences Institute Conference, Little Rock, AR, March (8-10).

Conlon, S, Abrahams, A & Diaz II, C (2017). “Big Data Analysis of Customers’ Opinions on Automotive Purchasing Service.” Decision Sciences Institute Annual Meeting, Washington, DC.

Diaz II, C, & Conlon, S (2017). “Analysis of Consumer Reviews and Comparison of Safety Factors for Products and Their Related Items.” Decision Sciences Institute Annual Meeting, Washington, DC.

Conlon, Sumali & Diaz II, Chad (2018). “Big Data Sentiment Analysis of Customers' Reviews, and the Prediction of Companies' Stock Performance.” Decision Sciences Institute Annual Meeting, Chicago, IL.

Diaz II, C, & Conlon, S (2018). “The Effect of Social Media Posts by Gun Violence Survivors on Firm Value Of Corporate Sponsors of Pro-gun Right Organizations.” Decision Sciences Institute Annual Meeting, Chicago, IL.

Conlon, Sumali & Diaz II, Chad (2018). “Predicting Customers’ Demand for Products using Sentiment Analysis Techniques.” Decision Sciences Institute Annual Meeting, Chicago, IL.

Diaz II, C, & Conlon, S (2019). “Automated Detection of Objectively Identified Spam Product Reviews.” Decision Sciences Institute Annual Meeting, New Orleans, LA.

Conlon, Sumali & Diaz II, Chad (2019). “Big Data Sentiment Analysis of Customers' Reviews, and the Prediction of Companies' Stock Performance.” AMCIS Annual Meeting, New Orleans, LA.

Teaching Experience Summary: Semester-Length Courses Taught:

- Introduction to Management Information Systems (MIS 309) – This course serves as an introduction to information systems principles for all active undergraduate business majors at the University of Mississippi. Major themes include systems analysis, data management, and security.
- Applied Systems Analysis and Design (MIS 317) – This course is an applications course that is focused on further developing the concepts and techniques learned by students in the first course of a two part series (MIS 307). The course is technical in nature and was based in C# programming.

Education Summary:

Master of Business Administration

The University of Mississippi, 2017

Bachelor of Science

-Major: Management Information Systems

The University of Mississippi, 2016

Other Professional Work Experience:

- July 2023 – Present: Sr. Manager, Configuration Implementation, Corelogic
- November 2022 – July 2023: Manager, Configuration Implementation, Corelogic
- April 2022 – November 2022: Associate Manager, Configuration Implementation, Corelogic
- June 2021 – April 2022: Configuration Analyst I, Next Gear Solutions (Acquired by Corelogic)
- August 2016 – July 2023: Graduate Assistant – MIS, The University of Mississippi, Oxford, MS.

- January 2017 – September 2017: Help Desk Representative, Next Gear Solutions, Oxford, MS.
- January 2014 – May 2014: Math and Sciences Tutor, Northeast Mississippi Community College, Booneville, MS.
- August 2011 – July 2013: Customer Service Manager, Walmart #0105, Corinth, MS
- February 2010 – August 2011: Wireless Sales Associate, Walmart #0105, Corinth, MS