MARTINA POCCHIARI

Managing Successful and Resilient Shared-Interest Communities

The Role of Digitization Technologies and Disruptive Events

Managing Successful and Resilient Shared-Interest Communities: The role of digitization technologies and disruptive events

Managing Successful and Resilient Shared-Interest Communities:
The role of digitization technologies and disruptive events

Beheren van succesvolle en veerkrachtige gemeenschappen met gedeelde belangen:
De rol van digitaliseringstechnologieën en ontwrichtende gebeurtenissen

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof. Dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on
Friday June 17^{th}, 2022 at 13:00hrs

by

Martina Pocchiari
born in Potenza, Italy

Erasmus University Rotterdam

Doctoral Committee

Committee members:	Dr. A. Lemmens Dr. H. Datta Prof. Dr. A. C. D. Donkers
Promotor:	Prof. Dr. Ir. G. H. van Bruggen
Copromotor:	Dr. J. M. T. Roos

Erasmus Research Institute of Management – ERIM
The joint research institute of the Rotterdam School of Management (RSM) and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam
Internet: www.erim.eur.nl

ERIM Electronic Series Portal: repub.eur.nl/
ERIM PhD Series in Research in Management, Number 552
ERIM reference number: EPS-2022-552-MKT
ISBN 978-90-5892-635-7

Design: FilippoBacci, iStock by Getty Images, photo ID 1271975501
Print: OBT bv, www.obt.eu

This publication (cover and interior) is printed on FSC® paper Magno Satin MC.

FSC
www.fsc.org
MIX
Papier van verantwoorde herkomst
FSC® C022586

Acknowledgements

You can't be what you can't see.

– Marian Wright Edelman

Growing up, I had no idea about what a university was, or how it worked; and no idea about what it meant, or what it took, to become a scholar. There was no example of academic excellence around me. There was no example of academic *anything*, for that matter – just some vague stories about high school teachers who would work at the local college (and usually just for the salary). In my formative years, all I could see were adults scrambling to recover from a disastrous financial crisis (a crisis that hit the South of Europe like a nuclear shock wave), and their kids on the side of the chaos, completely disillusioned with their present, and with very little faith in their future. It is only thanks to some truly precious, important people that, over time, I managed to visualize an academic future, and stretch the boundaries of my reality beyond what I thought was possible – sometimes, beyond what I thought was even imaginable. In this section, I want to thank those that allowed me to *see* what I wanted to *be*.

To my family – to my mamma, to my papà, to my Modi: grazie, thank you. You never put a limit to what I dared to dream. My ambitions were always met with enthusiasm, with 100% of your energy. Never, not once, with a hint of skepticism. As the boundaries of my reality expanded, I knew I would find you happy to explore them with me. With patience and love, you got to know the characters populating my life story, to ride the highs, and to swim through the lows with me. Thank you for walking along on the bumpy roads of life, and for making the journey so sweet, and so special. I look forward to walking all the new unexplored paths with you.

To the family I chose – to Francesco, Aldo and Giovanna: grazie, thank you. Giovanna and Aldo, your generosity, your big and open heart, and the joy you pour into sharing your life with me have completely changed the course of my adult life; thank you for welcoming me to your family. Mini: you have been on my side since day 1 (well, around day 4 to be precise?). You have seen, heard, experienced my

PhD trajectory like nobody else; and we have lived a thousand lives along with it. Through career changes, graduations, house moves, animal infestations, carbon monoxide leaks, conferences all over Europe, the plague, "angry birds", vibing and spinning, job market calls until 2AM, extremely bizarre humans, infuriating incompetence – *la magia della noia*, the blinding beauty of normal life – I am so happy you were by my side. I can't wait for our next adventures together. I love you.

To my mentors (in chronological order) – to Dr. Valentini, Prof. van Bruggen, Dr. Roos, Dr. Dover and Dr. Yildirim: thank you. Sara, you were the first educator to show me a path in uncharted territory while I was an undergraduate student; our conversations during your office hours changed everything for me. And even afterwards, whenever I asked for your help, you were there to give me the gift of your time, experience, and support. I will never forget that. Thank you.

Gerrit, together with Jason, you made the crazy bet to have me on board as your PhD student. Since then, I found support and guidance beyond the door of your office (and later, behind a Zoom link). Thank you for your trust in supporting my PhD journey. Jason, I peeked into your steampunk, chaotic, fascinating world for (at least) an hour a week for the last 5 years. I consider that my great privilege. You taught me so much of what I know today that I find it hard to quantify – that includes pop culture references, how to spell and pronounce many English words, and to choose compassion even when all the alternatives are easier. Thank you for your generosity, for your integrity, and for your grace in this amazing journey from student to scholar.

Yaniv, what an honor to be your co-author. Our brainstorming sessions and conversations have been continuously inspiring (yes, those include the talks about what chances I still have of becoming an astronaut or a jet pilot). Thank you for being a vocal supporter in my job market adventure, and for the unconditional encouragement in the past few years. And more recently, thank you and your family for being the most amazing hosts ever during my research visit. Pinar, working together on our common interests has been a peak moment in my PhD trajectory. You saw the "quick updates" on the projects, our recurring meetings, very long and (often way too detailed) project notes. But there was more that you didn't see: my absolute excitement about speaking to you, moments before the Zoom call would connect; and the way I take notes on your presentations, and on your reasoning and intuitions, just so I can learn, every time, a little more from your example. Thank you for being that example for me.

To my friends, present and past – thank you. Marianna, you were there on my first day of school (and I mean the first: September 1996). You have been there ever since, and I know I am the luckiest for that. You saw me transition from grade to grade, from city to continent, from kindergartener to professor. Your kind and loving presence has made the passing years lighter and warmer. No distance was ever too large for your words to reach me. Thank you for being there, thank you for always being my friend, my sister.

Gizem, thank you for sharing this incredible journey (and our office) with me; thank you to your beautiful family, for their kindness and warmth. It's been a long way from hiding under the desks, buying giant teddy bears for emotional support, watching Marvel movies I do not understand, brunching on the weekend in your home by the river, and wondering (really, seriously wondering) if we would ever make it. It was not easy, but we did it, friend; can't wait to see how you will grow into your new role as a professor. I am so proud of you. Ibrahim, thank you for sharing the quant adventure with me for the past years. I have been so lucky to be your colleague. Thank you for welcoming me into your family with Irem, for the coffee dates, and for the delicious Turkish breakfasts. You and Irem are brave, kind, beautiful people. I am so proud of you.

To every friend that has contributed to this achievement from undergrad to PhD: thank you. Thank you Margherita for our movie nights and talks; Adriana and Pepe for being the first to show me an entire world of international possibilities (and the bravery that must go along with ambition). FrancescoRipa (one word) for over 5 years of conversations, book recommendations, random texts and voice notes, and for reminding me to linger in the magic fabric of dreams from time to time. My friends from Unibo – Alessia and Yanxing Chen, who continued to cheer for me from a distance for almost 10 years. I miss you all dearly, I can't wait to see you again. To all the other friends, who have walked in (and sometimes, walked out) the door of my life in the past 5 years: thank you for sharing a bit of this challenging path with me. It was not easy; you made it easier. I keep you all close to my heart.

To my colleagues at all points in time: thank you. RSM Marketing Management faculty and the 2017-2022 PhD student cohorts, every single one of you has taught me something in the past 5 years. Thank you for your kindness, patience, altruism in sharing your knowledge with me. Pieter and Alina, one day in 2016 you suggested that I should consider doing a PhD. None of this would exist without you making that suggestion. Alina and Chris, Dan, Ana, Gabriele and Marta, the fun and delicious dinners and drinks we shared have been a highlight of my time at RSM. ERIM staff, and especially Jolanda and Annette: thank you for your super-heroic help, you make the magic happen. My future colleagues: I can't wait to share a new chapter of this exciting adventure with you.

All these incredible people contributed to turning this dissertation from a far, unattainable dream, to a tangible reality. They helped me to see what I could be. I hope – through my work, and starting with this dissertation – that I can also help someone else to see what they can be. In particular, I hope that many young people – and especially (self-identified) women, who are looking through the cracks of *any* periphery, both geographical and cultural, who are feeling distant and isolated – will be able to look at my path, see themselves into what's possible, and be inspired to pursue their ambitions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Shared-interest communities are social groups of people who gather around a common interest (Preece, 2001; Zander, 2018). In addition to providing people with a centralized source of information about their common interest, these communities are important hubs of knowledge, social support, socialization, and entertainment for consumers, brands, and institutions alike. For this reason, every day, millions of people resort to their shared-interest communities – both online and in-person – to meet, discuss, solve problems, and even manage disruptive situations of crisis or emergency, such as terrorist attacks (Burnap et al., 2014), natural or civil disasters (Eismann, Posegga, & Fischbach, 2016), financial instability (Racca, Casarin, Squazzoni, & Dondio, 2016; Romero, Uzzi, & Kleinberg, 2016), as well as product recalls and service failures (Hsu & Lawrence, 2016). Companies and brands also rely on shared-interest communities for marketing purposes. These *brand communities* are a specific type of "specialized, non-geographically bound shared-interest community, based on a structured set of social relations among the admirers of a brand" (Muniz & O'Guinn, 2001). Major brands invest between $500,000 and $10 million annually in the development of their brand communities, with returns on the investment spanning customer acquisition, satisfaction, retention, and purchase intentions (Manchanda, Packard, & Pattabhiramaiah, 2015; Millington, 2021).

Given the importance of shared-interest communities for businesses and society, research in marketing, economics, sociology, and computer science has investigated the factors that contribute to the success and sustainability of these social groups. Across disciplines, there is consensus that a key factor contributing to the success and to the continued existence of communities over time is the participation of their members (Iyer, Cheng, Brown, & Wang, 2020). The concept of community participation can be described as a spectrum – ranging from passive participation (frequently referred to as *lurking*, the simple act of accessing a community and consuming the content or goods produced by the group), to active contributions (such as liking, sharing, and posting content in online communities, and attending community events; e.g. Barger, Peltier, & Schultz, 2016; De Valck, Van Bruggen, & Wierenga, 2009). Understanding how communities can achieve sustainable levels of participation over time has been a major concern across fields. So far, the literature pointed out several important antecedents of community participation, including the benefits and needs that people satisfy when they participate in their communities (Kang, Tang, & Fiore, 2014a; Y. Wang & Fesenmaier, 2004), social influence and status-seeking (Zhou, 2011), pre-existing levels of activity (the "critical mass"; Marwell & Oliver, 1993).

The environments in which communities operate can also affect the way people participate in community dynamics – and these environments have greatly evolved over time. To start with, new technologies have lowered the cost of setting up communities online, contributing to the boom in popularity of *virtual communities* and hybrid community experiences (De Valck et al., 2009). As communities shifted to virtual settings, so did their experiences and aggregation occasions. The increased *digitization* of communities and their activities has supported the spread of digitized meeting formats, including virtual workshops, conferences, and social events organized around the communities' interests. The recent Covid-19 pandemic has only exacerbated this trend towards digitizing community experiences. Today, webinars, webcasts, and live chats about a common interest (as opposed to in-person meetings) have become widely accepted formats for community activities (Bevy, 2021).

Secondly, the nature of the relationship between external environments and communities online and offline is increasingly complex and interconnected. In 2021, Reddit communities of retail investors gained international media attention, when they coordinated a collective reaction to stock market information (Li, 2021). In that occasion, the coordinated efforts of millions of individual community members online effectively disrupted the global financial markets, caused nearly a billion-dollar accumulated loss for short-sell investors, and steered major policy changes in the regulation of stock market transactions (Fletcher & Aliaj, 2021). As another example, a deeply intertwined network of communities around conspiracy theories and far-right ideology was effectively used by rioters to coordinate the 2021 Capitol Hill insurrection (Heilweil & Ghaffary, 2021).

Both scientific evidence and recent events demonstrated that the complex environment of institutions, businesses, and technologies, and the daily dynamics of shared-interest communities are inevitably interconnected. However, it is still unclear how the emergence of new technologies and the occurrence of (often disruptive) external events relate to the success and sustainability of shared-interest communities. With three essays, in this dissertation, I shed light into the dynamics of shared-interest communities under the influence of changing technologies and potentially disruptive external events.

1.2 Research Questions

In this dissertation, I investigate the impact of digitization technologies and external events on several metrics related to shared-interest community success and sustainability. In evaluating community success and sustainability, I specifically focus on measures of *community participation* and *social network resilience*. To increase our understanding on these topics, I address the following research questions:

1. What is the impact of digitizing community activities on the participation intentions of community members? (*Chapter 2*)

2. How much does the impact of digitization on community participation differ across activity and community types? (*Chapter 2*)

3. What is the effect of a negative vs. positive shock to the shared purpose of an online community on members' engagement and social cohesion in the affected community? (*Chapter 3*)

4. To what extent does the community purpose shock affect core vs. periphery members differentially? (*Chapter 3*)

5. What is the effect of a brand crisis on the engagement and social network resilience of consumers in brand communities? (*Chapter 4*)

6. How does the effect of a brand crisis differ across consumers with different levels of loyalty, expertise, or attachment? (*Chapter 4*)

7. How does the response of the brand community to a brand crisis differ, depending on the type of crisis and the characteristics of the brand? (*Chapter 4*)

1.3 Outline of the Dissertation

Shared-interest communities generate functional, hedonic, and social-psychological benefits for their members, by offering community-organized activities (Kang et al., 2014a; Y. Wang & Fesenmaier, 2004). Increasingly, community organizers are offering digitized activities to their members. Digitized activities – which include webinars, webcasts, and live conferences – are often less expensive and more accessible than in-person activities (Bevy, 2021; The CMO Survey, 2021). At the same time, these digitized activities may not always provide the same degree of social and psychological benefits to the participants as their in-person counterparts (Cohn, Gesche, & Maréchal, 2018; H. F. Lin, 2007; Rothaermel & Sugiyama, 2001; Wirtz et al., 2013). The tension between convenience and meaningful social interactions may lead to higher or lower community participation. In Chapter 2, I investigate how increasing the extent of digitization of community activities impacts community participation, using data from the event-based community platform Meetup.com. Using structural causal models and causal random forests, I find that increasing the extent of activity digitization decreases members' intentions to attend such events. A counterfactual analysis shows that completely digitizing in-person activities causes an average 2.97% decrease in positive RSVPs. Furthermore, I find that the effect is heterogeneous across communities in different interest categories. This chapter contributes to the growing literature on the effects of digitizing human interactions on people's behavior in social groups. The chapter also informs community managers who need to evaluate the consequences of increasing the digitization of their communities.

Consumer-to-consumer activity in online communities has tangible consequences on brand shareholder value, and product category purchases (Algesheimer, Borle, Dholakia, & Singh, 2010; Manchanda et al., 2015; Mochon, Johnson, Schwartz, & Ariely, 2017). Online interactions become even more important during brand-related events that can make or break a customer community – such as product-harm crises, product launches, and "brandfest" events (Backhaus & Fischer, 2016; Cleeren, Van Heerde, & Dekimpe, 2013; Hsu & Lawrence, 2016). In Chapter 3, I assess the impact of external, community-related events – which act as negative vs positive shocks on the stated purpose of the community – on members' activity and social cohesion in online communities. In the empirical context of online sport communities, I leverage quasi-experimental conditions created by the outcomes of college basketball games, and integrate difference-in-difference models with social network analyses to show that (i) negative shocks to a community's purpose cause a decrease in activity compared to positive shocks; (ii) the decrease is attributable to members who belong to the "core" of the social networks; (iii) social cohesion is significantly affected by a negative purpose shock. In a series of heterogeneity analyses, I assess whether the disruptions to activity and cohesion can be mitigated by community managers. In particular, I evaluate two managerially relevant tools to address purpose-related shocks: expectations management and content moderation in the affected communities. This chapter supports community-facing professionals in maintaining their community in times of crisis, and in creating more value for their members during advantageous times.

In Chapter 4, I focus more specifically on the relationship between brand communities and the brand environment. Brand communities have an unparalleled power to integrate customer value with brand growth. Customers rely on brand communities to interact with each other, to connect with the brands they love, to solve problems, and to personalize their consumption experiences (Algesheimer et al., 2010; Bussgang & Bacon, 2020; Fournier & Lee, 2009; Manchanda et al., 2015). However, customers also resort to these communities to coordinate a negative collective crisis response (Ahluwalia, Burnkrant, & Unnava, 2000; Backhaus & Fischer, 2016; Hsu & Lawrence, 2016; Klein & Dawar, 2004; Luo, 2009). An uncontrolled reaction of online brand communities to brand crises can deteriorate brands' value and market performance, and push loyal and engaged consumers away from the brand social network. In this chapter, I assess the effect of brand crises on the volume of customer interactions in online brand communities, and the properties of the brand social network correlated with ease and speed of information spread. I use data from 300 brand communities on Reddit.com, and exploit the quasi-experimental exposure of community members to over 7000 brand crisis episodes reported by media channels between 2010 and 2019. In a series of difference-in-difference analyses, I find that brand crises (i) increase the weekly contributions of consumers in brand communities, and (ii) affect the patterns of information-sharing in the brand networks. Focusing on consumer types, I show that consumers who were active any time before the crisis effectively disengage from their brand communities following the crisis event – therefore, the average boost in brand-related activity is attributable to people who only activate after the crises. Furthermore, I show that the decrease in engagement is mitigated among consumers who had more experience, loyalty, or status within the brand community. Accordingly, I suggest that brand crises are a serious threat to the integrity of online brand communities, and that consumer loyalty and commitment has the potential to preserve the functioning of brand spaces online in the circumstances of serious reputation threats. The insights from this chapter support businesses and organizations managing online communities in situations of external stress and unexpected reputational threats.

1.4 Declaration of Contribution

In this section, I declare my contribution to the different chapters of this dissertation, as well as the contribution of my co-authors.

Chapter 2: The majority of the work in this chapter has been done independently by the author of this dissertation. The author formulated the research question, which was refined over time implementing the feedback from the promoter and co-promoter. The author also collected the data, reviewed and synthesized the literature, analyzed the data, and wrote the manuscript. The data for this study come from the website Meetup.com, a leading global community-building platform. The data were collected through Meetup's API, and include publicly available information about the website and its users. This chapter has been presented at numerous academic marketing conferences, and accordingly, the manuscript was improved over time by implementing the valuable feedback of the promoter, co-promoter, and other scholars in the field. The author of this dissertation is the first author of the article; the co-promoter is the co-author.

Chapter 3: The work in this chapter has been conducted in collaboration with the co-author of the paper, Dr. Yaniv Dover. The author of this dissertation and the co-author of the paper formulated together the research questions. The author of this dissertation reviewed and synthesized the available literature, collected the data, analyzed the data, and interpreted the results. Together, the author and the co-author wrote the manuscript for this article. The data for this chapter have been collected from two sources. One is the website DonBest.com, a leader in the provision of real time trading and odds information about North American sports. All the data collected from DonBest are archival and public. The second source is the Pushshift Reddit Archive API (Baumgartner, Zannettou, Keegan, Squire, & Blackburn, 2020). Pushshift is an archiving platform for data collected from the social media platform Reddit.com. The work in this chapter has been presented at various academic marketing conferences, and the manuscript has been improved thanks to the feedback of conference attendees and discussants. The author of this dissertation is the first author of the article; Dr. Yaniv Dover is the co-author.

Chapter 4: The majority of the work in this chapter has been done independently by the author of this dissertation. The author formulated the research question, reviewed and synthesized the available literature, collected the data, analyzed the data, and wrote the manuscript. The data for this chapter come from several different sources. The first source is the RepRisk ESG Risk Platform (RepRisk AG, n.d.). The RepRisk Risk platform is the world's largest database on environmental, social, governance (ESG), and business conduct risks. The second source is Crunchbase – a platform for business information about private and public companies (Crunchbase, 2021). The last source is the the Pushshift Reddit Archive API (Baumgartner et al., 2020). The author of this dissertation is the first author of the paper in this chapter; Dr. Pinar Yildirim and Dr. Abdullah Almaatouq are the co-authors.

1.5 Conclusions

This dissertation contributes to several streams of literature in marketing, network science, and economics. Across literature streams and disciplines, there is consensus that new technologies and external events have significant, profound effects on the way people interact with each other – especially in the context of shared-interest communities. However, empirical evidence on these effects is lacking. In this dissertation, I aim at filling several gaps in our understanding of the relationship between technology, external environment, and the internal dynamics of offline and online shared-interest communities. Chapter 2 contributes to literature investigating the impact of digitizing human interactions on economic behaviors – including cooperation and contribution to the public good – and the antecedents of active participation in shared-interest communities. Chapter 3 and 4 both contribute to literature in marketing and economics studying the consequences of negative publicity, reputation and status disruptions, and brand crises. In addition to these academic contributions, this dissertation has practical implications for companies and organizations working with community-facing channels. Finally, the insights from this dissertation point to numerous opportunities for future research, with the objective and wish to support the formation of valuable and resilient human connections in a changing world.

Chapter 2

The Effect of Digitizing Community Activities on Community Participation: Evidence from Meetup.com

2.1 Introduction

Shared-interest communities – social groups of people who gather around a common interest – are important sources of information, knowledge, social support, and entertainment for consumers, brands, and institutions alike. Major brands invest between \$500,000 to over \$10 million annually in brand community development, with returns on the investment spanning customer acquisition, satisfaction, retention, and purchase intentions (Manchanda et al., 2015; Millington, 2021). Over time, new technologies have lowered the cost of setting up community activities using digitized solutions – and the Covid-19 pandemic has only exacerbated the growing trend towards the complete digitization of community experiences. As a result, in recent years, webinars, webcasts, or live chats about a common interest (as opposed to in-person meetings) have become increasingly popular formats for community activities. While community digitization increases the reach and accessibility of community activities, it also entails several threats to the success and self-sustainability of the

Joint work with Dr. Jason M.T. Roos, Rotterdam School of Management, Erasmus University.

communities. Most importantly, digitized activities may fail to provide substantial social benefits from participation. These social benefits – which include exchanging emotional and physical support, socializing informally, and creating a shared identity – are among the main drivers of members' participation in their communities (Kang, Tang, & Fiore, 2014b; Y. Wang & Fesenmaier, 2004).

So far, the literature has suggested that people can extract social benefits from digitized interactions under specific circumstances – for example, when they have time to exchange information, present themselves selectively, and compare values (Walther, 1996). However, these circumstances are not always realized during digitized activities. Additionally, the literature has suggested that face-to-face interactions (as opposed to other interaction formats) are especially effective in creating solidarity, cohesion, and stronger social bonds between community members (Cohn et al., 2018; H. F. Lin, 2007; Rothaermel & Sugiyama, 2001; Wirtz et al., 2013). In sum, the impact of digitizing community activities on the participation of community members remains unclear. On the one hand, organizing digitized activities might extend the reach and lower the cost of community-building; on the other, it might also hinder community participation. Filling this knowledge gap with empirical evidence is, now, more important than ever. In the last two years, a growing number of companies and institutions have committed to substantially increase the digitization of their community activities, business operations, and workforce practices in the foreseeable future. Whether they achieve sustainable levels of participation in these digitized experiences will directly impact the success (and even the existence) of their communities, teams, and institutions (Bevy, 2021; Iyer et al., 2020).

This study aims at quantifying the impact of digitizing community activities on community participation. In particular, I quantify the effect of organizing digitized community activities – such as webinars, livestreams, or live chats, in contrast to in-person activities with comparable characteristics – on members' participation decisions. Furthermore, I quantify the differential effect of digitization across communities founded around different interests. To do so, I rely on detailed panel data from a leading community-building platform, Meetup.com. The data pertain to 118,326 events organized by 12,132 communities (called Meetup *groups*) in the first half of 2019 – before the Covid-19 pandemic forced many community events to become digitized. The estimation data include details about the groups (e.g. the size of their membership, category of interest), their events (e.g. detailed text descriptions, limits on attendance, attendance fees), their members (e.g., their past engagement with the group and its events), and the members' intentions to participate in future events (i.e. their *RSVPs*). Using the Meetup panel, it is possible to measure both event digitization and members' participation decisions about the community events. Since the data do not include a "digitization" variable, I use the events' text descriptions to measure the extent of digitization of each event. Specifically, two Support Vector Machines (SVMs) are trained to predict continuous probabilities of event digitization (versus in-person formats) based on the text that describes the events. Finally, individual RSVPs to the (differently digitized) events are used to measure members' participation decisions.

Identifying the causal impact of event digitization on members' participation choices is not trivial. The main identification threat comes from correlated unobservables, which may simultaneously affect both the likelihood that an event gets digitized, and the participation choices of community members. To address this concern, this study relies on a set of identifying assumptions. The identifying assumptions place a limit on the influence that any individual can exert on the demand for digitization, both in their groups and in the geographical market in which their groups operate. Relying on a set of relevant controls and fixed effects, I account for observed and unobserved factors that (i) relate to the market demand for digitization, and (ii) vary by group, event, member, and time. Conditioning on these important controls and fixed effects, I exploit the panel structure of the data to recover causal estimates, such that the effect of digitization is identified from observing repeated, within-member exposures to events with different probabilities of being completely digitized.

To estimate the effect of activity digitization on community participation, this study specifies a structural causal model (SCM) and several causal random forests

(CRFs – Athey & Wager, 2019; Wager & Athey, 2018). Using both parametric and non-parametric approaches entails several advantages. First, the SCM allows to perform counterfactual analyses, and to model non-responses as part of the choice problem. In the counterfactual analyses, I simulate a policy that forces all digitized events to have fully in-person formats. Second, the CRFs relax the functional form assumptions used for the SCM, and maximize the heterogeneity of the estimated group-level treatment effects. This allows us to perform a robustness check on the parametric results.

The results from the SCM provide two important insights: first, across all interest categories, people participate less in digitized events compared to similar in-person events. Second, perhaps most importantly, this effect is highly heterogeneous across interest categories. The parametric heterogeneity analysis indicates that digitization has the most detrimental effect in categories that may require high-frequency social interactions to generate value – such as sports, language courses, and socializing events. On the contrary, digitized events are equally or more attractive than in-person events in categories that might generate value even with lower-frequency interactions – such as music and concerts, career and business, and health and wellbeing, and photography. These insights are confirmed by the CRFs estimates for the conditional average treatment effect (CATE) of digitization on members' participation – which appear to be also highly heterogeneous. The average negative effect of digitization across interest categories is further characterized with a counterfactual policy evaluation. Under the simulated counterfactual policy, all events that were originally digitized are forced to turn into in-person meetings. The difference in counterfactual outcomes shows that, across these events, digitization causes an average 2.97% decrease in positive RSVPs, an average increase of 1.33% in negative RSVPs, and an average increase of 1.65% in non-responses.

This study contributes to several areas in the literature. In marketing and economics, I contribute to studies of the effectiveness of digitized human interaction and communication on economic behavior – such as cooperation, coordination, and contribution to a public good (e.g. Cohn et al., 2018; H. F. Lin, 2007; Short, Williams, &

Christie, 1976; Rothaermel & Sugiyama, 2001; Wirtz et al., 2013). These studies have investigated the role of digitizing human experiences in controlled lab experiments, or in specific empirical settings (e.g., non-profit virtual communities, or communities of wristwatch hobbyists and enthusiast), and have provided important initial evidence on the role of digitized vs. in-person interaction for community sustainability. I contribute to this literature by considering the impact of community digitization in the field, using a large sample of hundreds of communities and thousands of differently digitized activities. In marketing and sociology, this paper is related to studies on the impact of digitization on community success. Previous studies have touched upon the effect of increased community digitization in the context of single communities, organized either online or offline (Algesheimer et al., 2010; Dessart, Veloutsou, & Morgan-Thomas, 2015; Kang et al., 2014b; Wiertz & de Ruyter, 2007). This study complements this literature by assessing how digitized human interactions affect community participation across varying degrees of digitization, keeping the communities, members, and events as constant as possible.

Finally, this study has important managerial implications for marketing managers, community managers, and policy makers dealing with local communities (such as neighborhoods and workforce) as well as distributed communities (such as virtual or hybrid groups). Evaluating the consequences of digitizing community experiences has quickly become an urgent issue. Indeed, the Covid-19 emergency has dramatically impacted the demand for digitization in shared-interest communities, and investments in the digitization of customer experiences reached new heights during 2020 (The CMO Survey, 2021). This study provides a novel set of insights into the differential impact of digitizing experiences across different communities and activity types. In particular, the study suggests that idiosyncratic, category-specific norms and rules play the most important role in explaining the differential impact of digitization. Community-level interventions – such as nurturing and educating community members to the advantages of digitization, or highlighting the community-specific benefits from participating in digitized events – may help to mitigate the average negative impact of increased digitization on community participation.

The rest of the chapter is organized as follows. Section 2.2 describe the data and presents descriptive empirical analyses. In Section 2.3 I describe the methods, and provide details on the identification of the effects. In Section 2.4 I assess the impact of event digitization on members' participation decisions. I present the results of the parametric and non-parametric methods, and evidence of effect heterogeneity. Section 2.5 concludes.

2.2 Data and Descriptive Analysis

2.2.1 Background on Meetup

To estimate the impact of activity digitization on community participation, I collected data from Meetup.com, a leading global event-based community platform. Meetup is a community-building platform, launched in June 2002, that has experienced a dramatic growth over the past two decades. As of 2020, Meetup has reached over 49 million users in 230,000 Meetup communities and 193 countries. The primary goal of Meetup is to help users find and build local communities through the organization of events. Meetup also is widely used by companies and brands to build and maintain brand communities. Examples of brands relying on Meetup for community-building are Adobe, Google, Microsoft Azure, IBM, and Twitter (Meetup, 2020). To satisfy more business-oriented objectives, Meetup offers a paid Meetup Pro service, targeted at professional community managers and event organizers.

Users can join Meetup to create or join "groups." All Meetup groups are categorized into one of 33 interest categories, depending on the shared interest around which the group is formed. Examples of interest categories include dancing, social support, technology, and business. Meetup groups are primarily involved in organizing "events" related to their central interest. These events can have different formats, ranging from fully digitized to fully in-person. Examples of in-person Meetup events include workshops, product previews and tutorials, conferences, parties, dancing lessons, and book clubs. Examples of digitized Meetup events include webinars, live conferences, virtual discussion panels, and asynchronous video resources.

Each event has its own web page on Meetup.com. The event page includes details about the meeting – such as time and location – and an RSVP interface. Group members use the RSVP interface to communicate to the organizer whether they plan to attend the event. The RSVP interface has two buttons: "Yes" for a positive RSVP, and "No" for a negative RSVP. If an event has a limit on the number of attendees and the limit is reached, then the interface changes to "Yes" to "Waitlist". The button changes back to "Yes" if a new spot frees up. Although RSVPing is not compulsory for group members, it is strongly encouraged by both group organizers and the Meetup platform. Meetup emphasizes the importance of RSVPs for event management, and tries to support the organizers by encouraging group members to RSVP to upcoming events. One of these support initiatives is an RSVP reminder, automatically sent out by the platform to all group members 6 days before the scheduled event. These reminders are sent for all regular, non-recurring events, and for the first event in a recurring event series.

2.2.2 Data Collection and Data Structure

I collected public data describing Meetup groups, events, and members through the Meetup API between May 2019 and January 2021. The data cover a period of approximately six months, from January to June 2019. The data include the census of public Meetup groups primarily active in the 15 most populated cities in the U.S., according to the 2010 U.S. American Community Survey (U.S. Census Bureau, 2010). For each group, I collected information on the events organized during the January-June window. Finally, I collected member lists for each group – and for each group member, their RSVPs to the events. The resulting dataset has a panel structure, organized at the RSVP-event-group level. In the panel, I track time-varying and time-invariant information related to Meetup groups, events, and members' RSVPs recorded on the platform between Q1 and the end of Q2 of 2019.

Group Data. The group-level data include a full list of group members at the time of data collection, the interest category, whether the group has an active subscription to the Meetup Pro service, and the group's privacy options (i.e. whether the group is visible for non-members, and whether new members require the organizer's approval to join).

Event Data. The event-level data include the event date and time, the event creation timestamp, the event venue (if any), the event text description provided by the organizers, information about entry fees and RSVP limits, whether the event is part of a series of recurring meetings, and the number of members who were on the attendance waitlist at the time in which the event took place. The raw data do not include a field indicating if an event is digitized, but I leveraged the events' text descriptions to measure event digitization.

Event Digitization. To measure the extent to which an event has a fully digitized format, I relied on the events' text descriptions. In brief, I trained two support vector machines (SVMs) on the event text descriptions. One SVM predicts whether events have a "digitized" (vs non-digitized) format. The other, used to check the robustness of the first, predicts "in-person" (vs not in-person) events. The SVMs were trained on about 3000 cases, labeled as "digitized", "in-person", or both, by two independent raters. Any disagreement was resolved by a third rater not involved in the first labeling round. I then used the trained SVMs to predict extent of digitization of all the unlabeled events in the data. With 10-fold cross-validation, the SVMs achieved between 96% and 99% prediction accuracies. In Appendix A.1, I explain the measurement process more in detail and report descriptive statistics for the predicted cases.

To describe the format of each event, I used the continuous probability that an event is digitized, obtained from the SVM model predicting the "digitized" label. This "digitization" variable, ranging between 0 and 1, represents the accuracy of the predicted "digitization" label. For example, an event with digitization probability close to 0 will likely correspond to a completely non-digital event. The data also suggest that such an event would probably have a high probability of being in-person, as predicted by the "in-person" SVM model (Figure A.2). The vast majority of events is labeled consistently across the two SVM models (i.e., either high in-person probability and low digitization probability, or low in-person probability and high digitization probability). Only a small fraction of events (0.4%) has different labels predicted by the two SVMs (i.e. high in-person and digitization probability, or low in-person and digitization probability). Of these cases, a handful correspond to "hybrid" events, presenting both in-person and digital characteristics.

Average Past Digitization by Group. Using the event digitization variable, I measured the rolling average digitization of recent events offered by a group. This rolling average was calculated for each event at the time of its creation. This variable allows us to differentiate among groups that differ in their prior propensities to digitize events.

Member Data. The data include a rich description of members and their activity within groups, including the time in which the member joined a group, the time of their most recent visit to the group's page on Meetup.com, and the RSVPs to group events (including the time in which the RSVP was created). The data on group memberships and event participation allow us to derive additional important variables, including the outcome variable (members' participation in Meetup events), the tenure of different members in each group, the average attendance and exposure to past events, the number of other participants already planning to attend an event at the time of RSVP creation, and several metrics of event co-attendance among members of the same group. I elaborate on these variables below.

Member Participation in Events. I measured member participation in group events using the RSVP records from the Meetup data. Note that only positive or negative RSVPs are reported as observations in the panel, while non-responses to events are not automatically recorded in the Meetup API. However, combining the records of positive and negative RSVPs with the full list of group members, I could infer which members did not RSVP to an event. Combining the inferred non-responses with the positive and negative RSVPs, I created a categorical outcome variable, y, with three levels: $y = 1$ when the RSVP of a member to a group event is positive, $y = 0$ when the RSVP is negative, and $y = -1$ when the RSVP is missing.

Meetup emphasizes the importance of RSVPs for event management, and tries to support organizers by encouraging group members to RSVP to upcoming events. Given the importance of RSVPs to group organizers and the platform, I used both the presence (missing/non-missing) and value (positive/negative) of individual members' RSVPs to group events as measures of positive engagement in a community.[1]

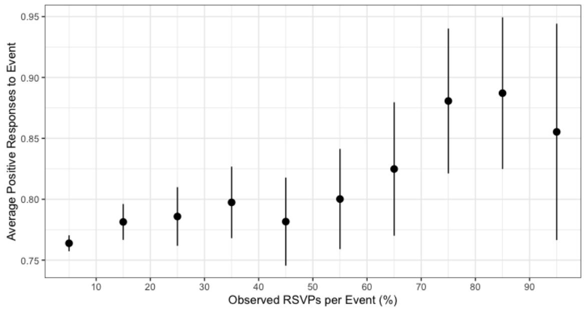
Event Awareness and RSVP Censoring. As discussed above, the measure of community participation is a categorical variable that includes a level for *non-response*, $y = -1$. When modeling members' RSVP decisions, I cannot assume that every missing RSVP is the consequence of a deliberate choice not to respond, because some members might not be aware of an event (e.g., because they did not visit the group's web page after the event was created). For some members, a missing RSVP (in place of a positive or negative RSVP) is neither the consequence of deliberation, nor a consequence of event digitization. Rather, it is the only available option. To identify which members are potentially unaware of each event, I constructed an *event awareness* indicator. The awareness indicator is a binary variable defined for each member-event pair. The indicator takes a value of 1 if a member was likely aware of the existence of the event at the most probable time of RSVP creation, and 0 otherwise. In Appendix A.2, I discuss how this variable can be constructed from information about the events (i.e. event creation time), the members (i.e. time of visits to group pages), and the RSVP timings (i.e. the time of RSVP creation, and the timing of the automatic RSVP reminder sent by Meetup.com). Using the awareness indicator, I classified 6.63% of the member-event observations in the original data set as *unaware members*. The unaware members were not included in the estimation sample.

After excluding non-responses from unaware members, the missing RSVPs carry useful information about the differential attractiveness of digitized events. If the missing RSVPs do not appear to be missing at random, then I can use their information to understand and model members' participation in events. To investigate the relationship between missing and observed RSVPs, I plotted data from a random sample of 18566 events (corresponding to 10% of the full data set) in Figure 2.1.

[1] An alternative measure of member participation, not available in the data, would be based on actual attendance records. Meetup provides a facility to record these data, but group organizers are highly inconsistent in how they record attendance (i.e., most organizers not recording attendance at all). Given the low reliability of the attendance records, I chose to focus on the much more reliable RSVP records.

The Figure shows a positive correlation between the share of positive RSVPs to an event and the share of non-responses. Such a pattern is likely to arise if individuals jointly decide whether to respond or not to an event – and if they respond, whether to respond positively or negatively.

Figure 2.1: Average Response Rate and Positive Response Rate per Event (based on 18566 events).



Based on this evidence, I make two important assumptions about the RSVP variable. First, I assume that the outcome variable (RSVP existence and value) is an *ordered* categorical variable with three levels: non-response < negative response < positive response. Second, I assume that the non-responses in the estimation sample are the product of members' choices, and that the likelihood of not responding can be modeled in the same way as the likelihood of leaving a positive or negative RSVP.[2]

Imposing that RSVP values are ordered on a linear, cardinal scale implies certain limitations to the estimation of non-parametric effects of digitization on RSVP choices. Most importantly, the linear and cardinal order may not fully capture the true distance between choices perceived by community members. For example, from a participant's perspective, the values for non-response and negative responses may be perceived as relatively closer to each other, compared to the value for positive responses. As such, the value for non-responses may still be negative in the participant's perspective, but not as negative as -1. Another possibility is that the

[2]I also check the relationship between non-responses and positive RSVPs across the groups that contribute to the identification of the effect – namely, the groups that organize both digitized and in-person events. This analysis is reported in Appendix A.3. This analysis suggested is that non-response rates do not depend on event digitization. Accordingly, I do not model non-response differently by whether the event is digitized.

values for positive and negative responses may be perceived as relatively closer to each other, compared to the value for non-responses. Indeed, the act of responding, in itself, may place the values for positive and negative relatively further away from the perceived value of not creating any response at all. To assess the boundaries of these limitations, I adopted two additional coding rules for the outcome variable y, that complement the three-level solution. First, I coded $y^{\text{Bin1}} = 1$ when the RSVP of a member to a group event is positive, and $y^{\text{Bin1}} = 0$ when the RSVP is negative or missing. Second, I coded $y^{\text{Bin2}} = 1$ when the RSVP of a member to a group event is positive or negative, and $y^{\text{Bin2}} = 0$ when the RSVP is missing. I use these alternative measures to assess the robustness of the non-parametric empirical estimates to the specification of the outcome variable.

Members' Tenure. To measure the tenure of each member within their groups, I calculated the difference in days between the event creation date, and the first day of group membership.

Participation in Past Events. To measure member heterogeneity in average group engagement, I constructed two metrics of average participation in past events (i.e., events organized between January and March 2019). One metric is the percent of past group events the member participated in; the other is the total number of past group events the member was exposed to through the Meetup.com website.

Participants to the Same Events. Next to the past participation metrics, I provide an additional measure of heterogeneity in the attractiveness of different events from the same group. In particular, recall that I observe both the response value and the response timing for all member-event pairs. With this information, for each RSVP at the time of its creation, I measured the number of other group members who had already indicated a positive attendance decision.

Co-Attendance to Group Events. I used the positive RSVPs to define metrics of average co-attendance for each group member (and their peers) in the estimation sample. In particular, for each group member, I calculated (i) the total number of unique peers who also responded positively to the same events; and (ii) the share of peers who responded positively to common group events, even in absence of a positive RSVP from the focal group member.[3]

2.2.3 Estimation Samples

The full data set containing the group, event, and member information is a panel of 24 weeks organized at the member's RSVP-event-group level. To estimate the effects of interest, I split the panel in two parts. The first spans January 1 to March 21, 2019 (10 weeks), and is used to calculate three control variables: members' past exposure to group events, members' average participation in past group events, and the average digitization of past group events.

The second part spans March 22 to June 22, 2019 (14 weeks), and is used to estimate the effects of event digitization on members' participation intentions. To allow estimation of these effects, I excluded two types of events and RSVP records. First, I excluded events that have a single RSVP created by the group organizer. Second, I excluded RSVP records for group members who were potentially unaware of an event's existence, based on the awareness indicator (Section 2.2.2). The resulting panel contains 7,851,101 RSVP records, corresponding with 285,730 members, 118,326 events, and 12,132 Meetup groups. It spans 14 weeks, and includes data from groups located in 15 major U.S. metro areas, comprising 508 cities and municipalities, and serving 33 categories of interest.

To maintain computational tractability, I estimate the effect of interest on a subsample of the full data. The subsample was constructed by first drawing 500 random group identifiers from the full data, and then filtering the full data to only include group, event, and member information for those 500 groups.[4] The subsample includes 5,548 events (about 5% of the total event records), 19,267 members (7% of the total distinct members), and 705,502 RSVP records (9% of the total records).

[3] An alternative metric of co-attendance to group events, calculated for any group member, could be the share of group peers who participated in the same events. This metric can be calculated using the record of positive responses through the estimation panel.

[4] Choosing a random sample of groups can potentially minimize the risk of sample selection bias. On the other hand, given the skewed distribution of the digitization variable, a random sampling of groups would very likely result in a subset of events organizing an overwhelming majority of in-person events. One alternative to the random group sampling would be a stratified sample or oversampling on high digitization probabilities.

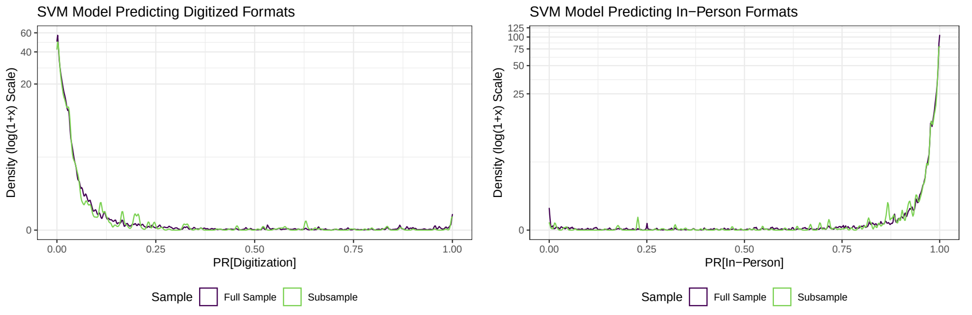
The subsample spans the same 14 weeks and 15 metro areas as the full sample, and covers 30 of 33 categories of interest. The 3 categories of interest not included in the subsample are "Lifestyle", "Cars and motorcycles", and "Paranormal".

2.2.4 Descriptive Evidence

In this section, I present descriptive evidence that event digitization relates to other event, group, and member-level features.

Event Digitization Within and Across Groups. On Meetup, the same group can organize different events with varying degrees of digitization and different other features. Event-level variation within the same group is important to estimate the causal effect of digitization on member participation, as it allows the same members to be exposed to events with different features, while keeping the group-level characteristics constant. Based on a dichotomization of the predictions from the Digitization SVM model using a 50% threshold, 403 groups in the full sample organized both digitized and non-digitized events (3.32% of all the groups). In the full sample, 94 groups organized exclusively digitized events (0.8% of the total), and 11,635 groups organized exclusively non-digitized events (95.9% of the total). In the subsample of 500 groups, 19 groups organized both digitized and non-digitized events (3.8% of all the groups), 4 groups organized exclusively digitized events (0.8% of the total), and 468 groups organized exclusively in-person events (95.4% of the total). Figure 2.2 shows the distribution of event format probabilities predicted by the two SVM models – one that predicts the probability of digitization, and one that predicts the probability of in-person events. Panel (a) shows the distribution of the predicted probabilities that events are digitized (versus not digitized, from the Digitization SVM model). Panel (b) shows the distribution of the predicted probabilities that events are in-person (vs not in-person, from the In-Person SVM model). The figure shows that the predictions from the two models are consistent with a situation in which in-person events are the norm. Furthermore, the figure suggests that there is a substantial overlap between the distribution of SVM predictions across estimation samples.

Figure 2.2: Distribution of SVM Probabilities by Predicted Event Format



(a) (b)

The within-group variation in event features suggests that the same Meetup members can be exposed to events with varying degree of digitization, within and across groups. Table 2.1 reports how many unique members were exposed to each combination of Digitized and In-Person formats, as predicted by the two SVM models. The descriptive statistics suggest that most Meetup members were exposed to only one type of event format – namely, in-person or non-digitized formats (column (3) in Table 2.1). The second largest group includes members who were exposed to both formats during the observation period (column (1)), while only a small minority of members were exposed only to digitized or non in-person events (column (2)).

Table 2.1: Members' Exposure to Digitized and In-Person Events

		N. Members Exposed to Event Formats		
Prediction Model	Sample	Both Formats (1)	Only Digitized/Not In-Person (2)	Only In-Person/Not Digitized (3)
SVM Digitization	Full	23364	555	261811
	Sub	987	53	18227
SVM In-Person	Full	24862	733	260135
	Sub	1462	70	17735

Note: the classification into "digitized" vs. "in-person" event class was performed using a 50% threshold for the predicted SVM probabilities.

How Does Participation Change Across Degrees of Digitization? In the previous section, I noted that thousands of Meetup members in the estimation sample were exposed to both digitized and in-person events. Here, I assess the extent of individual-level variation in *members' RSVPs* (both existence and value) across events with different digitization probabilities. Figure 2.3 visualizes the variation in RSVPs to digitized and in-person events in the estimation samples. The average response rate (which measures whether members created either a positive or negative RSVP, versus a non-response) varies with the degree of digitization in both samples. In the full sample, the average response rate to events with a probability of digitization greater than 50% is 8.56%, compared to an average response rate of 9.77% for events with a digitization probability lower or equal to 50%. In the subsample used for estimation, the average response rates are 9.92% (digitized) and 5.70% (non-digitized). The average *positive* response rate (conditional on the existence of a response) is also higher among events with higher degrees of digitization. In the full sample, the average positive response rate among events with a probability of digitization higher than 50% is 86.8% (versus 76.3% for non-digitized events). In the subsample, I record an 82.6% positive response average for digitized events, compared to 71.3% for non-digitized events.

Figure 2.3: Average Response Rates by Digitization Probability. The digitization probability is generated by the SVM model predicting Digitized vs Non-Digitized event labels. The shaded interval represents the 95% confidence interval of the mean response rate.

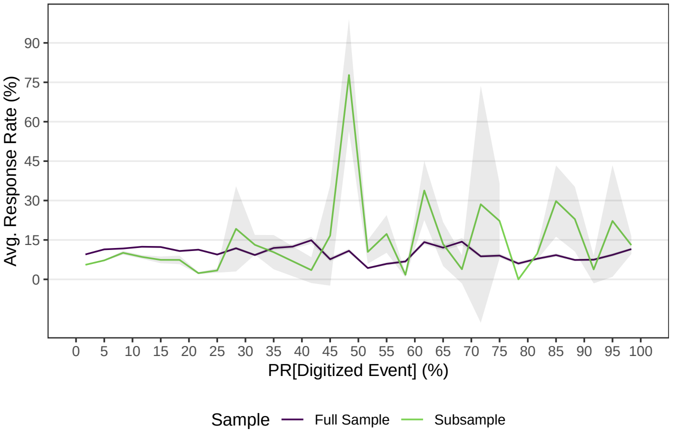
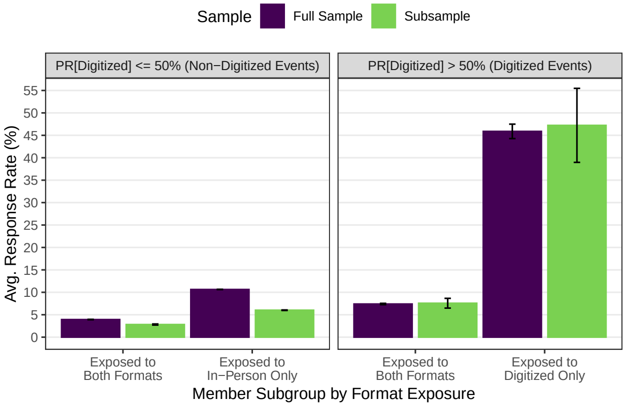


Figure 2.4 demonstrates how response rates vary among members who were exposed to different combinations of event types. Among people exposed to both dig-

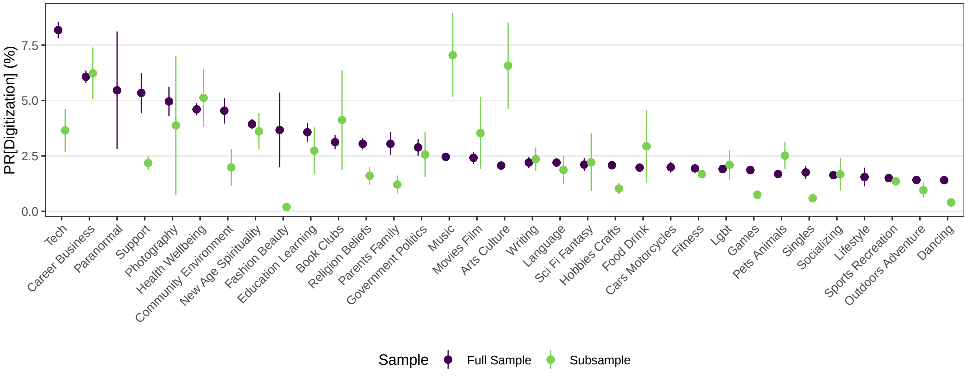
itized and in-person events, I measure a higher average response rate to digitized events: the average response rate to digitized events is 7.4% (7.6% in the subsample), against an average 4% response rate to in-person events (3% in the subsample). Furthermore, members exposed to only one event format throughout the panel – either in-person only, or digitized only – typically respond more often than members exposed to multiple event formats.

Figure 2.4: Response Rates and Digitization by Members' Exposure to Event Types



Groups, Events, and Digitization. The organizers' decision to digitize an event in their group may be influenced by several considerations, including cost-effectiveness, organizational flexibility, and the need for networking opportunities (6Connect, 2020). This suggests that events organized in certain interest categories (for example, business and career development, versus socialization and social support) may be more likely to get digitized. It also suggests that the effect of event digitization on members' participation may be differential across interest categories. Figure 2.5 shows that, in the estimation sample, the average event digitization indeed differs across interest categories. In particular, for example, events in the "technology" category in the full sample are more likely to have a digitized format than events in the "dancing" category – while digitized event formats are more likely to be observed in the "career and business", "music", and "arts and culture" categories in the subsample.

Figure 2.5: Digitization Probabilities from SVM Model Predicting Digitization, by Interest Category



Time, Geographies, and Event Digitization. Another factor that might influence event digitization is demand for digitization in the group's main location at a particular time. As Table 2.2 shows, the percentage of digitized events indeed varies across metro areas and weeks in the estimation sample. The metro area of Atlanta (GA) hosted events with the highest average digitization in the full sample, while the metro area of Chicago (IL) has the highest average digitization in the subsample. Across samples, the average digitization ranges between 0.66% and 5.17% over the observation period. Over the 14 weeks in the data, the weekly average digitization of events ranges between 2.42% and 3.99%.

Summary Statistics. Table 2.3 displays the summary statistics for all the variables used in the study, and the names of the matrix of features used in the models.

2.3 Estimating the Effect of Event Digitization

In this study, I aim at quantifying the effect of event digitization on members' participation in the events, measured by their RSVP decisions. At any given time, let Y_{ie} be the RSVP variable indicating the realized response value – which can be positive, negative, or missing. Let also D_e be the extent of digitization of e, and $\epsilon_{ie}, \epsilon_e$ be any unobserved factors affecting member i and event e. For member i exposed to event e from group g, I aim at estimating the causal effect of the extent of event digitization

Table 2.2: Percentage of Digitized Events by State and Week

Metro Area	Full Sample N. events	Full Sample N. groups	Full Sample Avg. Digitization (%)	Subsample N. events	Subsample N. groups	Subsample Avg. Digitization (%)
Atlanta	6823	665	4.68	193	21	1.26
Los Angeles	13715	1417	4.36	609	49	0.66
Detroit	1801	201	4.05	122	10	1.83
New York	23642	2167	3.97	1474	95	0.91
Chicago	6106	697	3.50	228	32	5.17
San Francisco	9848	1405	3.41	649	68	5.03
Houston	3643	353	3.14	130	10	3.43
Miami	2592	256	3.14	228	18	1.42
Dallas	5240	565	3.06	206	21	2.62
Boston	6503	755	2.90	269	32	1.44
Washington	11864	1184	2.55	478	60	3.35
Phoenix	6946	506	2.35	163	19	3.87
Seattle	9996	1092	2.24	231	35	1.73
Philadelphia	8001	750	1.78	534	26	1.17
Riverside	1606	119	1.60	34	4	1.45

(a)

Week (2019)	Full Sample N. events	Full Sample N. groups	Full Sample Avg. Digitization (%)	Subsample N. events	Subsample N. groups	Subsample Avg. Digitization (%)
12	5043	3196	3.12	247	136	2.47
13	8931	4936	3.48	399	202	3.61
14	8767	4869	3.28	407	199	3.06
15	9323	5069	3.40	456	216	2.97
16	8753	4850	3.25	395	204	2.89
17	9553	5298	3.73	463	229	3.47
18	8979	4863	3.56	439	227	2.92
19	8577	4763	3.44	398	188	2.82
20	9805	5407	3.23	456	231	3.01
21	8476	4575	3.27	432	201	3.00
22	8519	4609	3.45	398	193	2.56
23	9225	4982	3.31	407	198	2.73
24	9122	4885	3.35	432	203	2.42
25	5253	3479	3.96	219	132	3.99

(b)

Table 2.3: Descriptive Statistics of the Estimation Variables (Full Sample and Subsample of 500 Groups)

	Description	*Full Sample* Mean	*Full Sample* SD	*Subsample* Mean	*Subsample* SD
C_g	*Group Interest Categories (Binaries, $\{0,1\}$)*				
	Arts Culture	0.02	0.15	0.04	0.19
	Book Clubs	0.01	0.12	0.02	0.13
	Career Business	0.14	0.34	0.15	0.36
	Cars Motorcycles	0.01	0.07	0	0
	Community Environment	0.02	0.13	0.02	0.13
	Dancing	0.02	0.15	0.01	0.12
	Education Learning	0.02	0.16	0.02	0.15
	Fashion Beauty	0.00	0.06	0.00	0.04
	Fitness	0.02	0.15	0.03	0.18
	Food Drink	0.03	0.16	0.02	0.14
	Games	0.03	0.16	0.01	0.12
	Government Politics	0.01	0.09	0.01	0.10
	Health Wellbeing	0.06	0.24	0.08	0.27
	Hobbies Crafts	0.01	0.10	0.00	0.06
	Language	0.04	0.20	0.03	0.17
	Lgbt	0.02	0.13	0.01	0.10
	Lifestyle	0.00	0.03	0	0
	Movies Film	0.02	0.13	0.02	0.13
	Music	0.03	0.16	0.03	0.18
	New Age Spirituality	0.05	0.22	0.05	0.23
	Not Specified	0.00	0.01	0	0
	Outdoors Adventure	0.05	0.22	0.06	0.23
	Paranormal	0.00	0.03	0	0
	Parents Family	0.01	0.07	0.00	0.06
	Pets Animals	0.01	0.10	0.01	0.09
	Photography	0.02	0.12	0.01	0.11
	Religion Beliefs	0.02	0.13	0.02	0.13
	Sci Fi Fantasy	0.01	0.09	0.00	0.06
	Singles	0.01	0.12	0.02	0.13
	Socializing	0.07	0.26	0.08	0.27
	Sports Recreation	0.04	0.19	0.04	0.20
	Support	0.01	0.10	0.01	0.09
	Tech	0.19	0.39	0.19	0.39
	Writing	0.01	0.11	0.01	0.10
X_g	*Other Group Features*				
	Members (x1000)	0.15	0.24	0.17	0.29
	Is Open {0,1}	0.91	0.28	0.91	0.28
	Is Pro {0,1}	0.05	0.21	0.06	0.24

	Description	*Full Sample* Mean	*Full Sample* SD	*Subsample* Mean	*Subsample* SD
X_e	*Event Features*				
	Description Length (x10000 characters)	0.14	0.13	0.15	0.13
	Has Fee {0,1}	0.02	0.15	0.02	0.13
	Has Limits {0,1}	0.29	0.45	0.34	0.47
	Has Venue {0,1}	0.92	0.28	0.90	0.29
	Is Series {0,1}	0.06	0.24	0.07	0.25
	Morning {0,1}	0.22	0.42	0.22	0.45
	Avg. Digitization in Group	0.04	0.10	0.03	0.09
	Waitlisted Members (log1p)	0.03	0.24	0.05	0.31
	Event Digitization				
D_e	Pr[Digitized]	0.03	0.10	0.03	0.09
P_e	Pr[Non-Digitized]	0.97	0.11	0.98	0.09
M_{ie}	*Member-Event Features*				
	N. Positive RSVP (x10)	0.96	2.11	0.86	1.07
	Tenure (x10 years)	0.43	0.35	0.46	0.36
	RSVP Time (x10 years)	0.03	0.09	0.02	0.06
M_g	*Member-Group Features*				
	Avg. Past Response in Group	0.33	0.34	0.28	0.32
	N. Co-Attending Peers (log1p)	0.77	0.58	0.78	0.61
	Share of Co-Attending Peers (%)	0.16	0.10	0.16	0.10
	N. Past Events in Group (x100)	0.15	0.32	0.28	0.47

(D_e) on the likelihood of creating a (positive or negative) RSVP to the event (Y_{ie}). Therefore, the unit of analysis is the member of a Meetup group, deciding whether (and how) to RSVP to an upcoming event, given the extent of digitization of the event format. The outcome variable is the final RSVP value created by the member: a positive RSVP, a negative RSVP, or a missing RSVP. The treatment variable is the probability that the focal event has a fully-digitized format.

2.3.1 Identification Strategy

The model-free evidence describing the estimation sample highlighted several threats to the identification of the digitization effect. The main threat is that the effect may be potentially confounded: the decision to digitize an event and the decision to RSVP might be jointly influenced by several factors – including the unobserved market demand for digitization, time- and group-varying factors, and patterns of co-attendance among peers. To address this endogeneity concern, I include in the models a rich set of control variables at the member, event, and group-level. These controls explicitly account for the influence of market demand on RSVP choices and the extent of event digitization, and for the individual heterogeneity within each group. Throughout the rest of the paper, the controls are grouped in the term X_{ieg}, which includes (i) variables varying by group (group characteristics X_g and interest categories C_g); (ii) variables varying by event (event characteristics X_e); (iii) variables varying by member and event (M_{ie}); (iv) variables varying by member and group (M_{ig}).

In addition to controlling for observable confounders, I account for group-level and time-varying unobserved demand shocks using time- (τ_e), location- (ζ_m), and hierarchical group-level fixed effects (η_g). In the non-parametric models, I use group-level clustered standard errors. Finally, to identify the effect of interest, I need to rely on a few identifying assumptions:

Assumption 1. $corr(\epsilon_{ie}, D_e | X_{ieg}, \tau_e, \zeta_m, \eta_g) = 0.$

Assumption 1 states that the unobserved characteristics of any Meetup member are independent of the given extent of digitization of a Meetup event, conditional on controls and fixed effects. Under Assumption 1, no individual member in a group is so influential that their unobserved characteristics can directly change a group's decisions about event digitization.

Assumption 2. $corr(\epsilon_{ie}, \epsilon_e | D_e, X_{ieg}, \tau_e, \zeta_m, \eta_g) = 0$.

Assumption 2 states that the unobserved characteristics of any Meetup member are independent of the unobserved market demand for digitization, after controlling for observables and fixed effects. With Assumption 2, I assume that no Meetup member is so influential that, by themselves, they can shift the entire unobserved market demand for event digitization.

Assumption 3. $corr(\epsilon_e, \{\tau_e, \zeta_m, \eta_g\}) \neq 0$.

Finally, Assumption 3 states that the correlation between the unobserved demand for event digitization in any time and location, and the time-, market-, and group-varying fixed effects is not null. Under Assumption 3, the set of market-, group-, and time-varying fixed-effects capture most of the unobserved variation in market demand for digitized events. This final assumption can be relaxed if I assume that the contribution of any member i to the overall market demand is sufficiently small.

While there is a possibility that these assumptions may be violated by some individual member, the bias introduced by this violation would only affect the estimation for that individual member. As the reference group and market size grow larger, the bias from the violated assumptions would also potentially decrease. The set of control variables and identifying assumptions discussed so far, together, imply conditional unconfoundedness. In other words, the treatment assignment (the degree digitization of an event) is as good as random, conditionally on the selected covariates, and given the set of identifying assumptions (Rosenbaum & Rubin, 1983):

$$Y_{ie} \perp\!\!\!\perp D_e | X_{ieg}, \tau_e, \zeta_m, \eta_g \tag{2.1}$$

Parametric Identification. I recover the effect of D_e on Y_{ie} with heterogeneous parameter estimates in a Structural Causal Model, capturing the differential impact of digitization across interest categories. The parameters are identified from the data, since individuals are exposed to a sequence of events with varying extent of digitization within and across groups. Observing the sequential RSVP choices to differently digitized events within the same group, or across different groups over time, provides the necessary identifying variation from the data. For other event-level and group-level parameters, the identification arguments are in line to those for standard choice models using panel data, in which the RSVP decisions from the same member are observed over multiple time periods. The event-level parameters are identified both from members making RSVP choices about events with varying characteristics organized by the same group over time, and similar events organized by different groups at the same time. The group-level parameters are identified from the same members making RSVP choices, over time, about events organized by the different groups that they are members of.

2.3.2 Structural Causal Model

To measure the effect of event digitization on RSVP's, I specify a discrete choice model for the effect of event digitization on a member's RSVP. In particular, I construct a hierarchical Bayesian Structural Causal Model (SCM). In the SCM, each individual i decides whether to attend event e organized by group g. Y_{ie} is the RSVP value corresponding to this choice. I know that Y_{ie} includes the possibility to *not RSVP*. From Section 2.2.2 I also know that non-responses are part of the estimation sample, and that they are not missing at random. Therefore, I model the censoring on the RSVPs directly using a censoring parameter.

Choice Model. I define u_{ie} as the utility member i will receive from attending upcoming event e, which is organized at time t, by group g, located in market m. Based on the discussion in Sections 2.2.4 and 2.3.1, I anticipate the utility from attending this event could be influenced by factors that vary by market (m), time (t), group (g), member (i), and event (e). In particular, I allow the extent of event digitization, D_e to have a direct but differential impact on the utility from attending the event. In particular, I assume that the effect of event digitization is differential across groups operating in different interest categories (C_g). The drivers of member

i's RSVP decision are reflected in the following utility function:

$$\begin{aligned} u_{ie} &= v_{ie} + \epsilon_{ie} \\ v_{ie} &= D'_e * C'_g \beta_{DC} + C'_g \beta_c + X'_e \beta_e + X'_{ieg} \mu + \zeta_m + \tau_e + \eta_g \end{aligned} \tag{2.2}$$

where ζ_m, τ_e, and η_g are fixed effects that account for event demand varying by market, time, and group; C_g includes group g's category of interest; X_e includes event e's characteristics; X_{ieg} includes the identifying covariates; X_g includes group characteristics that inform the group intercept η_g; and ϵ_{ie} represents unobserved, idiosyncratic factors affect i's utility from event e.

For each event e, I jointly model individual i's decision to leave a positive or negative RSVP, or no response at all. As discussed in Section 2.2.4, I assume that the RSVP values are ordered in terms of their utility, such that censoring occurs among individuals who are not planning to attend. This specification requires two thresholds. The first, which I set to 0, separates positive RSVPs from negative RSVPs. The second, which I set to $L < 0$, separates non-responses from negative RSVPs. Equation 2.3 describes the relationship between utilities (u_{ie}), the thresholds (0 and L), and the observed outcomes (Y_{ie}):

$$Y_{ie} = \begin{cases} 1 & \text{if } u_{ie} > 0 \\ 0 & \text{if } L < u_{ie} \leq 0 \\ -1 & \text{if } u_{ie} \leq L \end{cases} \tag{2.3}$$

If the utility from the event is greater than zero ($u_{ie} > 0$), individual i will leave a positive RSVP. Otherwise, the outcome is determined by the censoring parameter L. When $u_{ie} \leq L$, the individual does not leave an RSVP. Otherwise, when $L < u_{ie} \leq 0$, the individual leaves a negative RSVP, indicating they will not attend. Assuming that ϵ_{ie} follows a standard normal distribution, the model implies the following ordered probit likelihood:

$$
\begin{aligned}
\mathcal{L}_{ie} &= \ell_{ie}(Y_{ie}) \\
\ell_{ie}(1) &= \Pr[v_{ie} + \epsilon_{ie} > 0] = 1 - \Phi(-v_{ie}) \\
\ell_{ie}(0) &= \Pr[L < v_{ie} + \epsilon_{ie} \leq 0] = \Phi(-v_{ie}) - \Phi(L - v_{ie}) \\
\ell_{ie}(-1) &= \Pr[v_{ie} + \epsilon_{ie} \leq L] = \Phi(L - v_{ie})
\end{aligned}
\tag{2.4}
$$

where Φ is the CDF of the standard normal distribution. Finally, I derive a Bayesian posterior distribution over the model parameters by specifying the following prior distributions:

$$
\begin{gathered}
\eta_g \sim N(X_g'\gamma_g, 1) \\
\beta_{DC}, \beta_c, \beta_e, \mu, \gamma, \zeta_m, \tau_e \overset{\text{iid}}{\sim} N(0,1) \\
L \sim N^{-}(0,1)
\end{gathered}
\tag{2.5}
$$

where N^{-} indicates a standard normal distribution truncated above at 0. During the estimation, I normalize the first element of the fixed-effect vectors τ_e and ζ_m to 0. The causal effect of interest is captured by the parameter estimates for β_{DC}. All the parameters to be estimated are summarized in Table 2.4.

Table 2.4: Estimated Parameters

Parameter	Description	Dimensions
β_{DC}	Digitization × Categories Heterogeneity	Vector Length N. Categories
β_c	Group-Varying Interest Categories	Vector Length N. Categories
β_e	Event Features	Vector Length N. Event Features
μ_1	Event-Varying Individual Features	Vector Length N. Event-Varying Individual Features
μ_2	Group-Varying Individual Features	Vector Length N. Group-Varying Individual Features
η_g	Group Intercepts	Vector Length N. Groups
τ_e	Time-Varying FE	Vector Length N. Weeks
ζ_m	Location-Varying FE	Vector Length N. Metro Areas
γ_g	Group Features	Vector Length N. Group Features
L	Censoring Threshold	Scalar

Parameter Estimation and Model Validation. I obtain the posterior parameter estimates using the shell interface to Stan (Stan Development Team, 2021). I use a No-U-Turn sampler of the variant of Hamiltonian Monte Carlo algorithm to sample from the posterior distribution of the model parameters. Finally, I validate the inferences from the Bayesian estimation using Simulation-Based Calibration (SBC; Talts, Betancourt, Simpson, Vehtari, & Gelman, 2018).

Counterfactual Policy Evaluation. Using the parameter estimates from the SCM, I evaluate the consequences of implementing a timely and highly-discussed policy on digitization: shifting all digitized events to be fully in-person. To construct the counterfactual estimation sample, first, I extracted the subset of events for which (i) the probability of digitization was higher than 50%; (ii) the probability of being in-person was lower than 50%; and (iii) the corresponding group had organized at least 1 in-person event in the past, such that in-person events were part of the consideration set for the community members. Then, for each event in the counterfactual subset, I simulated member's responses – first with D_e set to 0 (non-digitized), then with D_e set to 1 (fully digitized). I simulated the counterfactual scenarios using the entire posterior distribution of the structural parameters. Finally, I compared the distributions of positive, negative, and missing RSVPs across the two simulated scenarios in the counterfactual policy. The resulting difference in responses represents an average treatment effect on the treated (ATT). This counterfactual analysis provides important managerial implications about policies that increase or decrease the extent to which community activities take place in-person or in a digitized format.

2.3.3 Robustness Check: Causal Random Forests

I complement the Structural Causal Model with non-parametric Causal Random Forests (CRF; Wager & Athey, 2018). Using CRFs to complement the parametric analysis has several advantages. CRFs relax functional form assumptions on the structure of the unobserved errors, as well as on the distribution of group-level effects. The CRF algorithm also allows us to exploit and accurately reflect the heterogeneity in the available sample. Finally, the CRFs allow us to achieve all the desirable statistical properties of regression-based methods – such as asymptotic consistency – without committing to a parametric specification. The CRFs are based on the same set of causal relationships described in Section 2.3.1. In particular, D_e is the continuous treatment variable, Y_{ie} is the outcome variable, and X_{ieg} is the set of covariates described in Sections 2.2.4 and 2.3.1.

I estimate three versions of the CRF. The first is the "baseline" CRF, estimated using the same three-level, ordered and categorical outcome, $Y_{ie} \in \{-1, 0, 1\}$, as the one used in the SCM. The second CRF uses a binary coding for the outcome variable, such that $Y_{ie}^{bin1} = 1$ if member i responded positively to event e, and $Y_{ie}^{bin1} = 0$ otherwise. The last CRF uses a different binary coding for the outcome variable, such that $Y_{ie}^{bin2} = 1$ if member i responded positively or negatively to event e, and $Y_{ie}^{bin2} = 0$ if the member did not respond at all. The purpose of the second and third CRFs is to assess whether relaxing the assumption of a cardinal order for the outcome variable has sensible implications on the sign and magnitude of the estimated average treatment effects.

In implementing all the CRFs, I assume that there is considerable heterogeneity across groups, and that there could be unobserved group-level features that are treatment effect modifiers – such as strength of group leadership and susceptibility to social influence, or group norms. In the SCM model, I accounted for the sampling variability of potentially unexplained group-level effects using group-level hierarchical intercepts. In the CRFs, I adopt the approach of Athey and Wager (2019): assuming that the outcomes Y_{ie} of members within the same group may be arbitrarily correlated within a group (or cluster), I apply a group-level cluster-robust analysis. Since the treatment variable is continuous, I estimate a partial average treatment effect of event digitization on RSVP value, with the following estimator:

$$\hat{\tau} = E\left[\frac{Cov[D_e, Y_{ie}|X_{ieg}]}{Var[D_e|X_{ieg}]}\right] \tag{2.6}$$

I estimate $\hat{\tau}$ using the *grf* package in R (Tibshirani, Athey, & Wager, 2021).

2.4 Results and Discussion

2.4.1 Structural Causal Model

I report the results from the estimation of the Bayesian model defined by Equations 2.4 and 2.5 for the effect of event digitization on RSVP choices. As discussed, the RSVP indicator can take three ordered values, corresponding to positive ($Y_{ie} = 1$), negative ($Y_{ie} = 0$), and missing RSVP ($Y_{ie} = -1$). The likelihoods of Y_{ie} in Equation 2.4 are expressed in terms of individual utility from attending an event. Finally, the event digitization variable D_e is a continuous indicator of the probability that the event is fully digitized.

Model Diagnostics

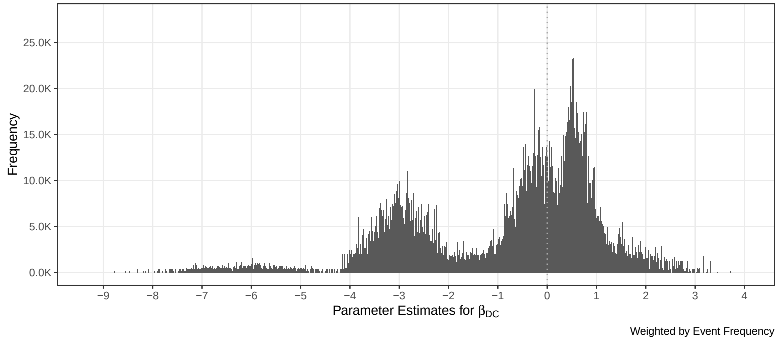
The model diagnostics indicate that I could successfully sample from the model's posterior distribution. The NUTS sampler did not report any divergences, meaning that the Hamiltonian Markov Chain has adequately explored the target distributions in Equation 2.4. The trace plots and diagnostics for the rest of the parameters are reported in Appendix A.5, and show that the MCMC chains explored the same region of parameter value. Finally, the SBC validations did not raise concerns on the model's ability to correctly recover the parameter estimates (Appendix A.4).

Parameter Estimates

Figure 2.6 demonstrates the posterior distribution of estimates for the parameters β_{DC}, capturing the heterogeneous effect of digitization across events in different interest categories. The average value of β_{DC}, weighted by event frequency, is equal to -0.601 (95% weighted C.I.: $[-0.644; -0.569]$) – suggesting a negative average impact of digitization on community participation. This corresponds to an average effect of digitization (weighted by event frequency and digitization probability) on the utility scale equal to -0.0037 (95% weighted C.I.: $[-0.0038; -0.0036]$). In addition, the distribution of parameter estimates suggests that digitization has a substantially heterogeneous impact on members' participation utility across events in different interest categories. This highly heterogeneous differential impact is not sensitive to weighting observations by groups or RSVPs frequencies (Figures A.11,A.12).

Descriptive evidence from the estimation sample supported the idea that events in some interest categories – such as technology, career, and business may be more or less attractive when they occur in a digitized format. Practically, the attractiveness of digitized events may change depending on cost-effectiveness, organizational flexibility, and the need for social support and networking opportunities (6Connect, 2020). Furthermore, the Technology Acceptance Model (TAM) and Uncertainty Reduction Theory (URT) suggest that the attractiveness of digitized events may vary with the extent to which prospective attendees are familiar with the digitized formats, and

Figure 2.6: Posterior Density of β_{DC} Estimates – Weighted by Number of Events per Category



well-informed about the event characteristics (Farzan, Lu, & Lin, 2016; H. F. Lin, 2007; Zhou, 2011). The parameter vector β_{DC} can help us explain the differential impact of digitization across all the interest categories available in the estimation sample. Figure 2.7 presents the posterior distribution of the heterogeneous parameter estimates distinguished by interest category.

Figure 2.7: Posterior Density of β_{DC} Estimates – by Category. Including the 80%, 90%, and 95% Bayesian credible intervals. The colors indicate whether the 80% credible interval for a parameter includes 0.

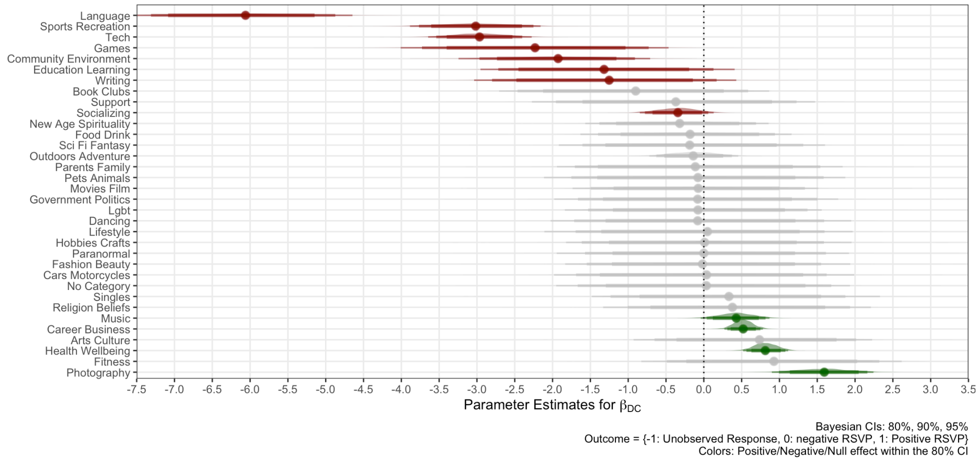


Figure 2.7 suggests that participation in digitized events differs considerably across categories of interest. The differential effect of digitization is negative, for

example, in events organized in the categories "sports and recreation" ($\beta_{D,Sports} = -3.00$, Std. Error= 0.005), "socialization" ($\beta_{D,Social} = -0.35$, Std. Error= 0.004), and "language" ($\beta_{D,Language} = -6.1$, Std. Error= 0.008). On the contrary, participation in digitized events is higher, on average, across events organized by groups in the category of "career and business" ($\beta_{D,Career} = 0.52$, Std. Error= 0.002), "photography" ($\beta_{D,Photo} = 1.60$, Std. Error= 0.005), "music" ($\beta_{D,Music} = 0.43$, Std. Error= 0.003), and "health and wellbeing" ($\beta_{D,Health} = 0.82$, Std. Error= 0.003). Overall, the results suggest that the category of interest in which a group operates is an important determinant of community participation, when events are digitized.

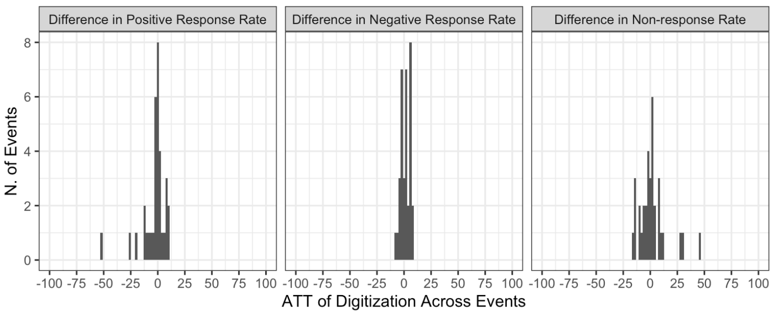
One possible explanation for this result is that event digitization may not be particularly suitable in categories that rely on one-to-one interactions, immediate social feedback, or taste-based discussions for the generation of shared benefits – such as playing sports, learning a new language, and making new friends. These results may support the expectation that digitized events provide relatively less social-psychological benefits to the attending members, and that coordination is more easily achieved in-person (Cohn et al., 2018; Koh, Kim, Butler, & Bock, 2007; Short et al., 1976). On the other hand, digitized events in categories that rely on functional exchanges to generate community benefits – such as "career and business" and "health" – may be relatively more attractive to the members, compared to the in-person alternatives. This result may indicate that digitized formats are at least as suitable as in-person formats when members are on the "receiving end" of functional value-generating activities – such as a skill-oriented webinar or a problem-solving task.

Counterfactual Analysis

So far, I established that the effect of digitization varies across interest categories. Next, I evaluate the impact of a highly-debated and managerially relevant counterfactual policy on RSVP choices. I simulate a counterfactual policy that forces all digitized events to be fully in-person. In the simulated scenario, I measure the difference in responses under completely digitized and completely in-person formats – i.e. an average treatment effect on the treated (ATT).

Figure 2.8 demonstrates the results of the counterfactual policy evaluation. The distribution of counterfactual effects across all relevant events suggest that digitizing community activities causes a decrease in the number of originally positive RSVPs (average -2.97%. 95% distribution limits $[-30.5\%, 9.02\%]$), and an increase in the number of originally negative RSVPs (average 1.33%. 95% distribution limits $[-5.52\%, 7.23\%]$). Additionally, completely digitizing activities causes an average 1.65% increase in non-responses (95% distribution limits across all relevant events $[-15.0\%, 29.2\%]$).

Figure 2.8: Distributions of Percentage Changes in Counterfactual RSVP Values (ATT) Due to Digitization.



Furthermore, Table 2.5 demonstrates that the counterfactual ATT of digitization varies across groups organizing events in different categories. For example, digitization has the most detrimental impact on positive response rates in the category "socializing" (mean 14.9% decrease in positive response rates), while it has a positive impact on positive response rates in the category "music" (mean 4.3% increase in positive response rates).

2.4.2 Robustness Checks: Causal Random Forests

The parametric specification in Eq. 2.2 assumes a functional form for the unobserved error term ϵ_{ie}. In this section, I relax this assumption and perform a non-parametric Causal Random Forest (CRF) analysis (Athey & Wager, 2019; Wager & Athey, 2018). In training the CRFs, I use the observed RSVP choice Y_{ie} as the outcome variable, the extent of event digitization D_e as a continuous treatment, and the set of variables X_{ieg} as the de-confounding covariates. Following Athey and Wager (2019),

Table 2.5: Percentage Changes in Counterfactual RSVP Values (ATT) Due to Digitization, by Category.

		% Difference in Non-Response Rate			% Difference in Negative Response Rate			% Difference in Positive Response Rate		
	N. Digitized Events	Mean	2.5% C.I.	97.5% CI	Mean	2.5% C.I.	97.5% CI	Mean	2.5% C.I.	97.5% CI
Music	9	-10.434	-14.158	-6.710	6.137	5.220	7.053	4.300	0.357	8.243
Tech	2	-1.205	-67.976	65.566	3.660	-26.327	33.647	-2.455	-39.239	34.329
Health Wellbeing	7	1.456	-1.976	4.887	-1.193	-3.841	1.455	-0.263	-1.943	1.417
Career Business	11	1.561	-1.347	4.469	-0.100	-1.685	1.485	-1.459	-5.306	2.387
Socializing	3	16.95	-11.383	45.283	-2.097	-6.57	2.377	-14.853	-38.918	9.211
Outdoors Adventure	1	28.83	-	-	-8.000	-	-	-20.830	-	-
Photography	1	45.38	-	-	6.25	-	-	-51.620	-	-

Note: The events from the categories included in the table had a predicted digitization probability higher than 50%, a predicted in-person probability lower than 50%, and had organized at least one in-person event in the past.

I train two separate causal random forests for improved precision. First, I train a pilot random forest on all the covariates in X_{ieg}. Then, I train a second forest using only the covariates with an above-average number of splits in the first forest. The second forest makes more splits on the most important features in low-signal situations (Athey & Wager, 2019). Additionally, the group identifiers are used to estimate the cluster-robust average treatment effects. Therefore, the forests assume that the outcomes Y_{ie} of members of the same group may be arbitrarily correlated within a group.

Baseline Forest: 3-Level Ordered Outcome

In the baseline CRF estimation, I used the same outcome as the SCM: a three-level ordered categorical outcome $Y_{ie} \in \{-1, 0, 1\}$. Additionally, the group identifiers were used to estimate the cluster-robust average treatment effects. Therefore, the forests assume that the outcomes Y_{ie} of members of the same group may be arbitrarily correlated within a group.

Using an overlap-weighted average treatment effect estimator, I find that the CATE of digitization in the training sample is equal to 0.052, and that the confidence interval associated with the CATE is suggestive of substantial heterogeneity (95% CI $[-0.074; 0.178]$). This result is consistent with the parametric insights of a highly heterogeneous average effect of digitization obtained from the SCM.

Alternative Forests with Binary Outcomes

To assess whether imposing a cardinal order on the outcome has a substantial impact on the CATE estimation, I trained two additional CRFs using different specifications of the outcome variable. I followed the same estimation procedure as the baseline CRF – first a pilot forest, then a final forest trained on relevant subsets of covariates. In the first alternative CRF, I used Y_{ie}^{bin1} (taking a value of 1 for positive RSVPs, and 0 otherwise). In the second alternative CRF, I used Y_{ie}^{bin2} (taking a value of 1 for positive or negative RSVPs, and 0 for missing ones).

Both alternative CRFs confirmed similar magnitudes for the overlap-weighted CATE, and estimated somewhat narrower confidence intervals compared to the baseline CRF. In the case of the first alternative forest, the $CATE^{bin1}$ equals 0.027 (95% confidence interval $[-0.026; 0.08]$). In the second case, $CATE^{bin2} = 0.02$ (95% confidence interval $[-0.042; 0.082]$). The magnitude and precision of these alternative CATEs suggest that the non-parametric estimation of the digitization effect is robust to different definitions of the event response outcome.

Non-Parametric CATE Heterogeneity

Following Athey and Wager (2019); Wager and Athey (2018), one heuristic for testing for heterogeneity in CRFs consists in grouping observations in two groups. The groups are formed according to whether the out-of-bag CATE estimates for the observations are above or below the median CATE estimate. Once these two groups are formed, the test for heterogeneity involves estimating average treatment effects in these two subgroups, separately, using a doubly robust approach. I run this heterogeneity test on the baseline CRF, and find that the difference in CATE between the high- and low-CATE groups is equal to -0.292. Furthermore, I find that the difference in CATE is significantly negative at the 90% confidence level (90% confidence interval $[-0.566; -0.018]$). This test provides additional evidence of the fact that the CATE of digitization is, indeed, substantially heterogeneous.

2.5 Conclusions

The digitization of human experiences is an increasingly attractive way of creating meaningful social connections among the members of interest-based communities. While digitization appears as a cheap and accessible alternative to in-person interactions, the lack of "human touch" may significantly impact the utility that people extract from their encounters. This is especially true in community contexts, in which establishing deep, trusting social bonds is the main reason why people participate in the community in the first place (Kang et al., 2014b; Y. Wang & Fesenmaier, 2004). As more and more marketing professionals, community managers, and policy-makers are evaluating the consequences of digitizing social encounters, the essential question is whether the digitization of events and activities has a detrimental impact on the chance that people will participate in the events. And more specifically, is digitization always detrimental (or helpful) towards community engagement, or does its impact vary across interests, communities, and activities?

In this study, I document that community participation is lower, on average, in digitized events than in in-person events. From the perspective of the average community member, digitizing community activities decreases the likelihood of creating positive RSVPs – which constitute a positive form of cooperation with the community organizers, and a commitment to respecting community norms. A counterfactual policy analysis quantified an average 2.97% decrease in the number of positive RSVPs due to implementing complete event digitization.

The empirical results also suggest that the impact of digitization on community participation is very heterogeneous. In particular, the differential parameter estimates for digitization indicate that the central interest of the community may be an important driver of heterogeneity. For example, digitized events organized in the "socialization" category are significantly less attractive to community members than their in-person counterparts. On the other hand, digitized events in the "career and business" category are at least as attractive – if not more attractive – than their in-person counterparts to community members. These insights were confirmed by a non-parametric Causal Random Forests model, which estimated a substantially heterogeneous conditional average treatment effect of digitization on members' participation.

From the perspective of community members, the results suggest that in-person community activities still generate considerable utility from participation – even though community managers are increasingly opting for greater community digitization. For a community organizer, this result suggests that – if community participation is an important objective or success metric for the community – digitized activities should not completely replace in-person activities. While digitized activities remain a viable, low-cost option to connect community members, the digitized activity formats probably generate a set of benefits for community members that do not necessarily correlate with active participation. Finally, the insights from the heterogeneity analysis suggest that the idiosyncratic category-level norms, rules, expectations, and social constructs play a very important role in explaining why different groups record higher or lower participation rates to their digitized events. Therefore, nurturing and educating community members to the advantages of digitization in the specific category – or highlighting the category-specific benefits from participating in digitized events – may be ways to mitigate the average negative impact of event digitization on community participation.

These results are in line with expectations from social presence theory, technology acceptance models, and uncertainty reduction theory – which would predict higher participation in community activities that feature in-person interactions (Farzan et al., 2016; Koh et al., 2007; H. F. Lin, 2007; Zhou, 2011). A lack of social presence may, in fact, generate communication weaknesses in community settings, while offline interactions help community members understand, trust, and identify with one another (Koh et al., 2007; H. F. Lin, 2007). Stronger solidarity and intimacy among community members, as a result, may encourage them to be more participative in community activities (Farzan et al., 2016). In addition, in 2019, digitized events were not the default format: community members may not have been as familiar and informed about the way digitized events were carried out. These high levels of uncertainty about highly digitized events may generate discomfort in community members, and in turn, result in lower participation intentions (Wirtz et al., 2013).

The insights from this study are extremely timely, as digitization is a pressing concern for marketing and community managers. On top of the organic growth over the past few years, digitized experiences dramatically gained more relevance during the COVID-19 pandemic. Among other things, the pandemic has forced marketing professionals to evaluate the balance between digitized and in-person activities. Meanwhile, marketing managers have continued to shift resources to building digital customer interfaces between 2020 and 2021 (investments in digital interfaces grew by 21.0% in February 2021 since June 2020 (The CMO Survey, 2021)). Similarly, community professionals predict that virtual events will continue to be essential even after the pandemic emergency (Bevy, 2021). The arguments and results from this study can help academics and managers formulating cautious predictions about the impact of community digitization in light of the Covid-19 pandemic. In particular, if the negative impact of digitized activities on participation is indeed due to a lack of social presence, lower coordination efficiency, higher uncertainty about digitized formats, and lower social-psychological benefits, then the current results could be interpreted and framed in three different scenarios related to the pandemic. First, early during the pandemic, fear and uncertainty regarding human-to-human virus transmission may have exogenously decreased the benefits from in-person community interactions. Furthermore, digitization technology was not yet as familiar and accepted as it is today – for example, the video-conferencing platform Zoom peaked at 300 million daily customers only in the three months to April 30 2020, signaling the recent and quick increase in popularity of this digitization tool (Sherman, 2020). Therefore, the average negative effect of activity digitization on community participation might have changed during the first half of 2020 – probably staying as negative at the beginning, and gradually becoming less negative over time. Between the second half of 2020 and the first half of 2021, more reliable scientific information on the Covid-19 virus reduced uncertainty around human transmission, and citizens and community members became more familiar with the digitization tools. In this second scenario, the results from this study – further informed by the Technology

Acceptance Model and the Uncertainty Reduction Theory – would predict that the negative effect of digitization on community participation should be mitigated. Possible mitigation mechanisms may be the increased familiarity with the new technology, and the presence of government regulations against in-person gatherings (Farzan et al., 2016; H. F. Lin, 2007; Zhou, 2011; Wirtz et al., 2013). Finally, from mid-2021 onward, as vaccination campaigns progressed and restrictions on in-person activities are gradually lifted, community members may start to attribute proportionally more value on face-to-face meetings. Therefore, on the one hand, the effect of digitizing community activities may become less negative, according to the current set of results and to the TAM predictions. However, the effect of *not digitizing* activities may increase proportionally more, as the benefits from face-to-face meetings are perceived as stronger and more urgent – a result suggested also by the counterfactual policy evaluation.

This study contributes to literature in marketing, operation science, and economics, investigating the impact of digitizing human interactions on economic behaviors – including cooperation and contribution to the public good (e.g. Cohn et al., 2018; H. F. Lin, 2007; Short et al., 1976; Rothaermel & Sugiyama, 2001; Wirtz et al., 2013). The results also contribute to literature in marketing and sociology investigating the antecedents of active participation in communities of interest (e.g. Dessart et al., 2015; Kang et al., 2014b; Y. Wang & Fesenmaier, 2004; Wirtz et al., 2013; Zhou, 2011). In particular, I add the extent of digitization of community activities to the list of potential antecedents of members' participation. To date and to the best of my knowledge, this is the most comprehensive study on digitization of community experiences taking into account multiple geographies, communities, interest categories, and event types. So far, marketing and sociology literature has focused on either single communities offering activities with varying degrees of digitization, or on multiple communities employing only one communication format – either in-person, or fully digitized (e.g. Dessart et al., 2015; Dutta-Bergman, 2005; Kang et al., 2014b; Koh et al., 2007; Ling et al., 2005; Y. Wang & Fesenmaier, 2004). However, community managers are increasingly resorting to various activity formats

– without necessarily committing to one digitized or in-person format – and will continue to offer a range of formats in the coming years (Bevy, 2021). The analyses showed that category-level idiosyncrasies can explain much of the heterogeneity in the effect of activity digitization. This result suggests that it is necessary to take into account more than one community when addressing research questions regarding community digitization. Methodologically, I contribute to literature in digital marketing with a framework to model digitization under endogeneity and censoring, and across multiple geographies and periods. Similar endogeneity and censoring concerns have been raised in different contexts, from pricing strategies to digitized entertainment in movie markets (e.g. Rooderkerk, Van Heerde, & Bijmolt, 2013; Yang, Anderson, & Gordon, 2021). Typically, these concerns are solved via instrumental variable estimation, or via randomized experiments in the field. However, in many digital marketing contexts – and especially when dealing with digitization of community experiences during a pandemic – resorting to instruments or RCTs can be both practically difficult and ethically problematic. In this study, I rely exclusively on observational data easily available to most community managers. The estimation strategy in this study can be extended to digital marketing problems that are based on comparable data generating processes, producing non-random treatment assignments and observable censored outcomes.

The digitization of human experiences offers many opportunities for future research. Future studies may expand the evaluation of digitization policies in a post-Covid reality, and compare how the shift to remote working and 100% digitized social activities has affected the reaction of community members to activity digitization. Relatedly, future research may assess if population density is an important confounding variable in the relationship between digitization and community participation in light of the Covid–19 emergency. Future work could also improve on the measurement of the digitization construct proposed in this study. An opportunity for future research is to train an NLP model to detect activity digitization from text in more sophisticated ways. A refined model of digitization detection would greatly help researchers and practitioners to understand what particular textual cues or constructs contribute the most to accurately predict event digitization. Also related to measurement, future research may study the effect of community digitization on additional outcomes from the community participation spectrum – which includes passive participation, referrals, moderation, and even negative and disruptive participation (Ardichvili, Vaugh, & Wentling, 2003; Dutta-Bergman, 2005; Brodie, Ilic, Juric, & Hollebeek, 2013; Kang et al., 2014b). In particular, it would be interesting to assess if members exploit the increased anonymity and the decreased inter-personality of digitized events to be more disruptive, or to choose negative forms of engagement. Finally, future studies could attempt to run well-designed field experiments, involving one or more communities willing to randomly expose their members to differently digitized events. After resolving any ethical concerns to assigning people to digitized or in-person situations, controlled experiments can provide unique insights into the mechanisms behind the effects recorded in the observational setting, and into how community members interact *with each other* during different types of events. From such studies, managers and policy makers learn valuable information about the boundaries of the digitization effects.

Chapter 3

The Role of Community Shared Purpose in Online Community Dynamics

3.1 Introduction

Digital platforms allow consumers and brands to interact with each other seamlessly, daily, and at a global scale. These networked interactions often occur within online communities – groups of people rallying around a common purpose, cause, or goal in a defined digital space (Armstrong & Hagel, 2000; Kozinets, 1999). More than ever before, consumers rely on online communities to share word-of-mouth and information (Ardichvili et al., 2003), nurture relationships with each other and with brands (Fournier & Lee, 2009), make collective decisions (Fletcher & Aliaj, 2021), and collaborate to achieve collective goals (Faraj, Kudaravalli, & Wasko, 2015). *Brand communities* are especially valuable in the digital landscape. Between 2019 and 2022, a vast majority of surveyed companies recognized that consumer interaction in brand communities online is critical to their business mission. The business value of brand communities is reflected in the fact that, in recent years, major brands invested between $500,000 and $10 million annually in online community management (The Community Roundtable, 2021; Millington, 2021). Brands that invest in the success

Joint work with Dr. Yaniv Dover, Hebrew University, Jerusalem Business School and Federmann Center for the Study of Rationality.

and resilience of their online brand communities report high returns on investment, increased brand awareness and loyalty, and a reduction in customer support costs (Bussgang & Bacon, 2020; Millington, 2021).

One particularly important factor affecting the success and resilience of online communities is their *shared purpose* – the common interest and basic reason for members to join and participate in the group (for example, in brand communities, the shared purpose of the community is the brand itself; Muniz & O'Guinn, 2001; Preece, 2001; Zander, 2018). The literature demonstrated that the shared purpose of a community is a foundational element in community formation (Dessart et al., 2015; Forsyth & Burnette, 2010; Preece, 2000). However, beyond community formation, the role of the shared purpose in sustaining successful dynamics remains unclear. More specifically, it is not known how much a *purpose disruption* – a positive or negative event that enhances or threatens the basic interests around which a community exists – would impact the dynamics related to community success and resilience.

Such purpose disruptions are common in digital platforms. For example, online brand communities sometimes face product-harm crises. These are events that can affect the brand's reputation and ability to deliver shared value, with consequences to brand equity and market performance (Backhaus & Fischer, 2016; Hsu & Lawrence, 2016). In online sports communities, fans are continuously confronted with positive and negative team performance events, usually following competitions and games. Sports team performance, as expressed in game outcomes, can spillover to the fans' perception of the team and affect churn rates in online sports communities (Zhang, Tan, & Lv, 2018). Finally, in online financial communities, investors often face sudden changes in the price of their stocks. These price shocks can affect participation activities within the online financial communities built around the affected stocks (Fletcher & Aliaj, 2021; Romero et al., 2016). These examples suggest an important connection between the shared purpose of a community, its internal dynamics, and its resilience to purpose disruptions. While prior research has theorized the existence of this connection (Preece, 2001; Tajfel, 1978; Zander, 2018), the empirical evidence supporting the theories is either correlational, or referring to very rare, atypical events (e.g., Racca et al., 2016; Rasmussen & Ihlen, 2017). Causal inference through field experiments, in organic communities and at a sufficiently large scale, is extremely difficult. Experiments that disrupt the shared purpose of a community – performed over a large number of real-life communities and members – are not only costly and complex, but often also ethically problematic (e.g., El-Sayed, Seemann, Scarborough, & Galea, 2013). Therefore, marketers and policy-makers lack empirical managerial insights on the management of online communities in the face of frequent purpose disruptions.

In this study, I address this gap in the literature by investigating how shocks to a community's purpose affect community dynamics. I focus on the effect of these shocks on community engagement, composition, and structure of social interactions. To do so, I leverage quasi-experimental conditions within hundreds of online communities on Reddit.com, one of the largest global platform hosting communities online. In particular, I use fan communities based around NCAA basketball teams, competing in the first division of the NCAA men's basketball league between 2015 and 2019. I collect and combine community data with team and game data, and construct an estimation panel tracking more than 196K Reddit users, 822K discussion threads, and 1.5M comments over 4 years. In this context, I assume that team losses-versus-wins are disruptions to purpose of the team communities. I then study the effect of purpose disruption on the following three dimensions related to the success and resilience of communities: (i) the volume of community interactions; (ii) the structure and cohesion of the social networks underlying these interactions; and (iii) the linguistic characteristics of the user-generated content shared in the communities.

There are several advantages to using NCAA online fan communities as an empirical setting. First, in terms of economic relevance, the value of the sports market is expected to reach $599.9 billion in 2025 – with digital platforms playing a critical role in driving engagement with fans and consumers (Kumar & Bhalla, 2021). Online sports communities are also a type of brand community (Yoshida, Gordon, Heere, & James, 2015). Like any other brand, sports teams have a marketing interest in maintaining a devoted consumer base in online communities. Sports communities online are often managed by companies, and are frequently used as marketing tools to establish true relationships with fans, to gather consumer insights, and to engage fans in campaign development and promotional efforts (Nelson, 2020). Second, this setting allows to leverage several measurement and modeling opportunities. The common purpose of fan communities is known and relatively well-defined: it is the common affinity to a specific team, and the shared interactions around it. Similarly to what happens in brand communities, this shared purpose should be closely associated with the level of success of the team itself. Therefore, losses and wins of the focal team effectively represent negative and positive shocks to the purpose of its fan community (Card & Dahl, 2009; Zhang et al., 2018). Additionally, game outcomes are observable, well defined in time, and have a relatively clear interpretation – wins are most likely a positive shock to the community's purpose, and losses are a negative shock.[1] In modeling the relationship between community purpose and community dynamics, sports games also offer the unique opportunity to account for unobserved expectations about the games or the teams in the estimation of treatment effects. In this study, I use large bookmaker prediction markets to collect, per each game, the odds of observing a win or a loss. Given the size and importance of bookmaker markets, consistent with the literature, I assume that market-predicted outcomes are efficient approximations of pre-game market expectations (Card & Dahl, 2011). Therefore, I exploit the pre-game predicted point spreads to capture unobserved expectations about game outcomes in the online fan communities. For example, if the market predicts equal winning odds for competing teams, it is reasonable to assume that the effect on the community of the specific game outcome will be perceived as "random-like" by the community members, and will not be the result of unobserved expectations. Furthermore, fans of NCAA teams heavily rely on online communities for their interactions online – similarly to consumers relying on brand communities. In this study, peer-to-peer fan discussions on Reddit are documented with high resolution: I observe the content of individual social interactions, and which member is interacting with which other member. This provides with the time-varying social net-

[1] Tied outcomes are not allowed in basketball, so game outcomes are either positive or negative shocks.

work structures underlying the fans' online interactions. Having reliable measures of the dynamic network structure allows to study how shocks to the community's purpose differently affect the behavior of the different parts of the community, as well as how social cohesion metrics are affected by external shocks. Finally, the large size and considerable breadth of the digital discussion communities allow to study relatively long time scales and large samples.

To estimate the effects of interest, I use a difference-in-difference framework, and rely on the bookmakers' prediction of game outcomes for the identification of the effects. I find that game outcomes have a significant effect on the dynamics of online communities – and therefore, that the purpose of communities plays an important role in their day-to-day existence, beyond community formation. More specifically, I find that a lost-vs-won game – that is, a negative-vs-positive purpose shock – decreases engagement within communities. I also find that the effect of negative shocks is absorbed differentially across sub-groups of community members. In particular, negative shocks mainly impede the activity of the community's core – the most highly connected, active, and central community members. This pattern of results is also reflected in the post-loss changes to social network structures of the affected communities. Negative shocks induce members to interact with fewer peers, and within smaller social cliques. In other words, the networks become more centralized and localized after a negative shock. Additionally, negative purpose shocks also negatively affect the influx of new members into the communities – suggesting that a positive shared purpose is important in the self-sustainability of online communities. In terms of user-generated content, I find that negative-vs-positive shocks also reduce the "energy level" in the discussion – measured through the magnitude of arousal in community discussions – and impede expressions of group affiliation. Discussions also exhibit a decreased focus on past events. Finally, I find that more unexpected negative shocks induce stronger disruptions to the activity and functioning of online communities. Moderation of the community's content also seems to correlate with a mitigation in the disruptive effect of negative shocks.

In sum, I find that the purpose of communities plays an important role in their ongoing existence and success, and its importance extends beyond community formation. The empirical results suggest that the state of a community's purpose fuels its social dynamics, especially during times of turmoil. It also seems that in this context, perhaps surprisingly, the core members of the community are not as committed to the long-term existence of the community. In the face of a negative shock, instead of investing in repairing the damage and maintaining cohesiveness, they disengage from the community. This suggests that managers working with such platforms should monitor the state of the purpose of their communities and consider incentivizing the engagement of core members, such that during crises they could continue to support the community. I also find that expectation management may be a helpful preventive or mitigating strategy for managers. Therefore, managers may consider playing a more proactive role in shaping expectations within the community, and practicing consistent community moderation.

With this study, I contribute to several literature streams. First, I contribute to the literature studying the impact of negative publicity and external disruptions on consumer behavior in digital platforms, including social networks, social media, and online communities (e.g., Ahluwalia et al., 2000; Dondio & Usher, 2017; Grégoire, Tripp, & Legoux, 2009; Hsu & Lawrence, 2016; Rasmussen & Ihlen, 2017). I complement these studies by expanding their empirical examinations to events beyond single, rare disruptions like large-scale natural disasters and geopolitical events. This expansion is important, since brands and organizations need to deal with shocks to the quality of their products, services, and reputations much more frequently than they have to manage natural disasters and financial crises. In the same stream of literature, I also contribute to the ongoing discussion on consumer response to threatening, negative events. I find evidence of "love-becomes-hate" effects – implying that highly loyal customers can develop an extreme negative attitude toward a brand (see Aaker, Fournier, & Brasel, 2004; Roehm & Brady, 2007) – and relate these effects to the pre-event levels of engagement and network position of community members.

Second, I contribute to studies in marketing and network science on the role of different types of community members in various community dynamics (Barberá et al., 2015; Bramoulle & Kranton, 2005; Borgatti & Everett, 2000; De Valck et al., 2009; Racca et al., 2016; Torres, Toral, Perales, & Barrero, 2011). I contribute to this stream by focusing not only on detecting core and periphery structures in the social networks, but also on the relationship between core-periphery status of community members and purpose disruptions. While previous studies adopted a descriptive approach to the core-periphery structures in social networks, I document the causal, differential effect that a relevant shock to the purpose of the community has on the separate behavior of core and periphery members.

Finally, I contribute to literature in marketing, sociology, and network science studying the factors leading to the success or failure of online communities, as well as the antecedents and consequences of consumer engagement in digital platforms (Brodie et al., 2013; Manchanda et al., 2015; Preece, 2001; Stam, 2009). Barger et al. (2016) reviewed the antecedents of consumer engagement in social networks, and found that consumer participation can be explained by factors at the level of the brand and product, of the consumer, of the content shared in the social networks, and on the platform itself. I expand this list of factors to also include the *environment* in which the social networks function, as I show that environmental shocks have an impact on community dynamics.

The rest of this paper is structured as follows. Section 3.2 provides an overview of relevant literature on digital platforms, online communities, and shocks to a community's purpose. Section 3.3 illustrates the empirical context and the institutional background, describes the data used in the study, and provides relevant descriptive statistics. Section 3.4 illustrates the methodology and the model used to estimate the effects of interest. Section 3.5 shows the results of the empirical analyses, while Section 3.6 investigates the robustness of the estimated results. Section 3.7 concludes.

3.2 Literature Review and Theory Development

Online Communities and Community Shared Purpose A large body of literature in marketing, information science, and management has established the importance and value of online communities for consumers, businesses, and for society in general. For consumers, participation in online communities can lead to making more informed purchase decisions (Algesheimer et al., 2010) and to the accrual of functional, hedonic, and social benefits and rewards (Kang et al., 2014b). For businesses, using online communities for marketing purposes can positively impact brand trust and commitment (Kang et al., 2014b), online and offline consumer purchase behavior (e.g. Algesheimer et al., 2010; Manchanda et al., 2015; Mochon et al., 2017), customer satisfaction, and can even cut customer support costs by several million dollars every year (Millington, 2021). For society, online communities represent important hubs to coordinate mutual support, collective efforts, and responses to impactful events, such as financial disruptions and crises (e.g. Fletcher & Aliaj, 2021; Racca et al., 2016; Romero et al., 2016), wars and terrorist attacks (e.g. Dondio & Usher, 2017; Jung & Park, 2014; X. Wang, 2016), product-harm crises, and product recalls (e.g. Backhaus & Fischer, 2016; Cleeren et al., 2013; Hsu & Lawrence, 2016).

The extant literature argues that, in theory, one of the main factors in the formation of successful and active communities is their *shared purpose*. The purpose of a community is the shared focus on a common interest that provides the basic reasons for members to join and belong to the group (Preece, 2001; Zander, 2018). Essentially, members invest their time and effort in a community, and, in return, they expect a series of benefits resulting from community participation (e.g., functional, hedonic, social, and psychological, see Y. Wang and Fesenmaier (2004)). The literature assumes that the shared purpose of a community is central to the definition of online communities, and as such, that it is a foundational element in community formation (for examples, see Dessart et al., 2015; De Souza & Preece, 2004; Dover & Kelman, 2018; Forsyth & Burnette, 2010; Preece, 2000). However, even though consumers report that communicating with others about their common interests is an important motivation for them (Hennig-Thurau, Gwinner, Walsh, & Gremler, 2004), the actual role of the shared purpose in sustaining community dynamics remains unclear.[2] To date, there is a lack of large-scale empirical insight into the relationship between community purpose and community dynamics – both for online communi-

[2]McGrath (1984) discusses a general perspective of shared purposes in groups. The author argues that social groups exist to generate common ideas around common purposes, select between them, and then execute the commonly agreed ones (Circumplex Model of Group Tasks).

ties, and for social groups in general. More specifically, there is limited evidence on whether disruptions to the community purpose have any effect on the sustainability and success of online communities. The answer to this question has important implications for marketers and managers of communities, as it should inform any policy designed to sustain the community and promote community engagement.

In marketing, prior related research has explored how consumers react to product or service failures in brand communities. Since the purpose of brand communities is the mutual interest for a brand, company, product or service, then product or service failures are effectively threats to the purpose of the respective communities (Muniz & O'Guinn, 2001). However, these studies focused on the effectiveness of different company- vs consumer-initiated recovery efforts (Schaefers & Schamari, 2016; Yuan, Lin, Filieri, Liu, & Zheng, 2020) and on the impact of post-crisis online engagement (Hsu & Lawrence, 2016) on brand equity and shareholder value, rather than studying how the performance failures affected the dynamics of the brand communities. Another relevant stream of literature has focused on the impact of negative brand publicity – which could also negatively impact the common purpose of established brand communities – on brand sales (Berger, Sorensen, & Rasmussen, 2010), attitudes towards the brand (Ahluwalia et al., 2000), and brand equity (Dawar & Pillutla, 2000). These studies did not investigate the actual dynamics of online brand communities in response to the negative shock to purpose – in this case, bad publicity.

Given the gap in the extant literature, I am interested in studying the following questions. First, do disruptions to the state of the shared purpose affect *any* aspect of the dynamics in online communities? Second, if purpose does play an important role, is it a negative or positive role? More specifically, if the community purpose is disrupted, does it hurt community engagement, or, on the contrary, does it rally members to invest in the long-term prospects of the community – and actually strengthen the group by increasing engagement? Third, do all community members react similarly to the effects of purpose disruptions, or do different members play different roles? The answers to these questions should deepen our knowledge about the functioning of online communities, and inform digital marketing strategies related to brand community spaces.

Online Community Dynamics: the Importance of Engagement and Social Cohesion Two important indicators of community dynamics over time are the level of members' engagement, and the level of social cohesion within the community. I am interested in observing the dynamic progression of these indicators over time, and how they are affected by disruptions to the community purpose.

Members expend time and effort when contributing to communities, and expect short- and long-term benefits from their contribution. Therefore, a common metric for community engagement is the level of community activity, typically measured as the volume of content and interaction between community members over time (Barger et al., 2016; Preece, 2001). Therefore, theoretically speaking, the state of the purpose in brand communities – i.e., the brand itself – should potentially impact community activity. For example, the extant literature suggests that, before and after brand crises, consumers contribute to digital platforms differently – mostly using platforms to search for information, connect with peers, and gain social support (Rasmussen & Ihlen, 2017). Similar patterns arise for online communities of investors following financial events of uncertainty (Racca et al., 2016), and following disruptive geopolitical events which altered the composition of an online finance community, as well as the sentiment of its user-generated content (Dondio & Usher, 2017). In sum, the available evidence suggests that disruptions related to the purpose of communities should play a role in their engagement dynamics, and that they should affect the type of benefits members receive from contributing to their communities.

Next to community engagement, social cohesion is another important aspect of healthy community dynamics. Social cohesion is defined as a resource shared by a group or society, that potentially interests both individual group members and the group as a whole (N. Lin, 2002). Social cohesion in groups, neighborhoods, and societies has been associated with feelings of trust, shared identity, awareness of needs, and commitment (Forrest & Kearns, 2001; Granovetter, 2018; Jenson, 1998). When social networks achieve higher social cohesion, people are more likely to engage in dynamics of strong interpersonal connection, and of knowledge-sharing and acquisition (Tortoriello, Reagans, & McEvily, 2012). In turn, these patterns of cooperation, prosociality, and shared identification are associated with higher levels of community engagement (Zhou, 2011; Y. Wang & Fesenmaier, 2004; Wirtz et al., 2013). Cohesion can be associated with higher commitment of its members to the long-term existence of the community and, so, with potentially higher incentive to invest in the community after a crisis. The above implies that cohesion in social networks is also important in determining how the network reacts to disruptions and shocks. Surveys showed that during non-routine situations, such as emergencies or rare negative events, a higher share of activity originated from dense networks of *core* nodes – central network agents connected to other important nodes. These core nodes became more active in non-routine situations to lend informal support (Hurlbert, Haines, & Beggs, 2000). In routine situations, on the other hand, members of more cohesive social networks were more likely to receive ongoing social support from peers (Hurlbert et al., 2000). Finally, social cohesion, as a correlate of perceived social support, is also important when social network members must cope with stressful situations (Thoits, 1995).

To summarise, both engagement and cohesion are indicators of the "health" of community dynamics. If the shared purpose does play a role in the community dynamics, I expect to see a significant effect of disruptions to the purpose on these two indicators.

Different Roles of Community Members: Core vs Periphery Social networks can be generally decomposed into two subgroups, populated by different members: the *community core* and the *community periphery* (Borgatti & Everett, 2000). Up until now, the extant literature mainly discussed these two subgroups descriptively, and suggested that they differ in their incentives to expend effort in the community. On the one hand, core members are highly active members and typically interact regularly and heavily with each other, and as such, can be characterized by being densely connected (Borgatti & Everett, 2000; De Valck et al., 2009). Core members are commonly considered relatively expert and trustworthy, intrinsically motivated, and strongly committed and identified with the purpose of the community (Hunt, Bristol, & Bashaw, 1999; Racca et al., 2016). On the other hand, periphery members are sparsely connected relative to the core, and relatively isolated from the rest of their peers. The periphery is usually larger than the core, and most of the time, it includes lurkers and inactives (Borgatti & Everett, 2000; De Valck et al., 2009). The periphery also tends to include members who display more casual interest in the purpose, who are more extrinsically motivated, and whose primary attachment to the community might depend on particular circumstances rather than the purpose itself (Mahony, Nakazawa, Funk, James, & Gladden, 2002). There is some descriptive evidence that core and periphery members play different roles in a community. In that respect, the literature shows descriptively that these two subgroups contribute differently to the community's growth, create different content, and diffuse information and innovation in different ways (Barberá et al., 2015; Bramoulle & Kranton, 2005; De Valck et al., 2009; Racca et al., 2016; Torres et al., 2011). Finally, small-scale surveys show that core nodes display different activation patterns following negative, non-routine external events (Hurlbert et al., 2000). In general, the available evidence suggests that core and periphery extract different but overlapping benefits from participating in online communities. Our research setup provides a rare opportunity to study the different roles of each subgroup towards community success, following disruptions to the shared purpose of the community.

Community Purpose Disruptions and Community Dynamics If members' motivation to invest time and content in the community is tied to its shared purpose, I expect that multiple aspects of the community dynamics may be impacted when an external event threatens that purpose. A disruption to the purpose of the community could negatively affect the community, by either directly or indirectly (i) threatening the social identity of members, (ii) reducing the basic motivation to contribute, (iii) harming the emotional climate in the community, and (iv) hindering other aspects of community activity. If the disruption significantly harms the value that members extract from the community, members could choose to either leave the community, or stay and "repair the damage" by investing in it – for example, if they expect higher returns in the long run (Marwell, Oliver, & Prahl, 1988). Therefore, in case of a negative shock to the community purpose, I can expect three scenarios related to activity levels, social cohesion, and well-being in the community: (1) no effect, (2) negative effect, or (3) positive effect. The *no-effect* scenario may suggest that the shared purpose does not play an important role in day-to-day community dynamics. In that case, the shared purpose may merely be an excuse for the community to form, but bear no practical usefulness afterwards. The *negative effect* scenario – which implies that purpose shocks reduce engagement and cohesion – would provide evidence that the shared purpose plays a role in the everyday existence of online communities, and not only in their formation. In that scenario, as the value that members extract from the community is damaged, members may have a weaker motivation to participate – which could lead to a threat to the community's existence. Finally, the *positive effect* scenario – which implies that purpose shocks increase engagement and cohesion – would suggest that not only community benefits are tied to its purpose, but that members may be willing to endure a short-term loss of benefits in exchange for maintaining the community in the longer term. For example, members may be willing to spend more time and effort in the community – even if the shared purpose of the group is temporarily threatened – because they may expect higher returns from a stable community in the long term. Identifying which one of the three outcomes (no effect, negative, and positive effect) occurs in this context can provide insight into online community dynamics and their underlying drivers.

Community Core vs Periphery: Differential Impact of Community Purpose Shocks As discussed above, core members are more active and more strongly committed to the purpose of the community. Therefore, I expect that core members may experience purpose shocks as a more severe threat to the value they extract from participating in the community. However, it is unclear whether core members would be more *positively* or *negatively* affected by a purpose disruption than periphery members. Literature in marketing and management suggests that people with stronger identification and commitment to their favorite brands can either leverage their strong commitment to buffer negative effects of threatening events, or perceive the threats as too severe and experience an amplified negative impact (Khamitov, Grégoire, & Suri, 2020; Sharma, Sadh, Billore, & Motiani, 2021). On the one hand, there is evidence that consumers with higher levels of brand commitment tend to extensively counter-argue negative information, while endorsing positive information (Ahluwalia et al., 2000). On the other hand, the marketing literature has documented a "love-becomes-hate effect", for which customers with higher levels of brand identification, attachment, and commitment, engage in stronger desires for revenge and avoidance following performance failures (Aaker et al., 2004; Aggarwal, 2004; Roehm & Brady, 2007).

In sum, in evaluating the differential impact of purpose disruptions on core versus periphery members, I expect three reasonable scenarios. Under the first scenario, there would be no observed difference between the subgroups following a purpose disruption. This would suggest that changes to the purpose have a uniform impact on the different sub-groups of community members, and do not affect high-commitment members differently than low-commitment ones. Under the second scenario, core members would increase their activity in the community, while periphery members would not exhibit an equivalently strong change in behavior. This outcome would suggest a "repairing the damage" coping strategy by the high-commitment members. In this scenario, core members would invest time and effort to maintain stability, perhaps in expectation of higher returns in the long run. Under the third scenario, core members would disengage and decrease their activity in the community, while the periphery would experience a weaker effect. This scenario would suggest that a purpose disruption negatively affects all members, but affects high-commitment members more intensely. In contrast to the second scenario, the third one suggests that the core members may discount the long-term benefits of investing in the community to maintain stability. In other words, core members may not feel it is "worth" to invest time and effort in a disrupted community[3]. Any of these scenarios, if observed in the empirical estimations, would provide insight into the different roles these subgroups play in the community.

[3]For completeness, a fourth scenario would imply that periphery members increase their activity, while the core decrease. Based on the literature reviewed, this scenario seems to be of very low probability.

To summarise, in this paper, I ask the following questions: (i) Do the most important indicators of community dynamics rely on the state of community purpose for their sustainability?; (ii) If so, how does a disruption to the community purpose affect online community dynamics?; (iii) How do different members depend on the stability of the community's shared purpose? I leverage the context of online sport communities to empirically answer these questions. In the next section, I describe the empirical context and the data used in this empirical analysis.

3.3 Data

In this study, I estimate the effect of a shock to a community's purpose on several dimensions of community dynamics. I do that in the empirical context of online communities which are created around sports teams. Online sport communities are a specific form of brand community (Yoshida et al., 2015). Like brands, sport teams have intrinsic and financial interests in acquiring and maintaining a devoted fan base that will both consume their services, and be a source of continuous advocacy for the brand. Additionally, sport communities are often managed by commercial organizations, news outlets, companies, and brands[4]. Online sport communities are frequently used as marketing tools to establish true relationships with fans, to gather consumer insights, and to engage fans in campaign development and promotional efforts (Nelson, 2020).

Teams, games and communities data To collect relevant data on teams, games, and online communities, I focused on the NCAA Men's Basketball (NCAA-BB) Division 1, between November 2015 and March 2019, i.e., including four seasons. The NCAA is a collegiate athletic body, governing college basketball in the United States. In the most recent years (2021–2022), 358 colleges and universities competed in 32 Division 1 basketball conferences.

The team statistics and game information for all the teams competing across the four seasons were collected from the website DonBest.com – one of the largest suppliers of real-time betting data for North American sporting events (Crunchbase, 2021).

[4]Examples of online sport communities managed by brands and sport organizations are Fussball.de, operated by Deutsche Telekom, and Liverpool's LFC forum operated by DigitalSportsGroup.

I complemented the DonBest game data with information on pre-season team rankings, obtained from the Associated Press college basketball poll (Associated Press, n.d.).

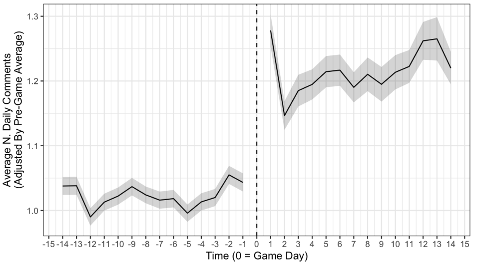
Based on the list of games and teams, I matched the teams with their respective online fan communities on the Reddit.com platform. Reddit is one of the biggest platforms to host online communities globally. Due to its global reach and popularity, Reddit is ranked as the 19^{th} most-visited website in the world, and the 7^{th} most-visited website in the U.S. in 2021 (Alexa Internet Ranking, 2021). On Reddit, millions of users create and participate in communities (called subreddits), organized around the shared interests of the users. In the subreddits, community members can share content with their peers by posting text, links, images, and videos. Peer-to-peer interactions within Reddit communities occur in the form of discussion threads – tree-like structures of posts and comments. Discussion threads can be as small as a single post, or a very large tree of posts and comments. For each community of fans, I used the Pushshift Reddit Archive (Baumgartner et al., 2020) to extract information on all community discussions occurred in the 15 days before and after each game date, as well as their meta data (e.g., creation date and thread structures). The resulting sample is a panel dataset, organized at the subreddit-game-day level. In total, the panel includes 244 subreddits, 196,456 Reddit users, 822,454 discussion threads, and 1,546,346 comments. Merged with the game data, the panel also includes 12,738 games, played by 259 teams over 484 game dates.

Community dynamics and purpose disruptions I quantified community activity using the daily count of posts and comments created by community members. To account for community-level factors (e.g., community size effects), I divided the daily count of community contributions by the average daily community activity recorded during the relevant pre-game period. The "negative purpose disruption" indicator is a binary variable that takes value 1 if the outcome of a game is a loss for the focal team, and 0 if the outcome is a win. Figure 3.1a demonstrates the average adjusted daily activity across communities in the two weeks before and after the games in the estimation sample. The Figure shows that the average community activity sharply increases after a game – and remains high for about 2 weeks – suggesting that community members gather to discuss the game outcomes online. Figure 3.1b, on the other hand, shows that the average post-game activity levels are usually lower after losses than after wins.

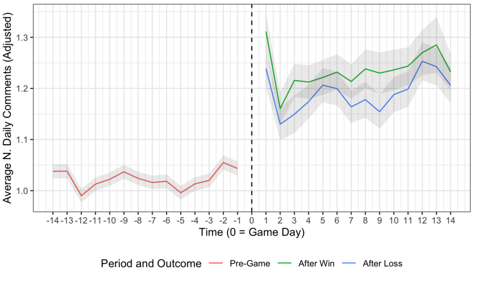
I also used text analysis tools to further characterize community engagement. In particular, I coded several dimensions of the text exchanged between community members using lexicon for norms of valence, arousal, and dominance (Warriner, Kuperman, & Brysbaert, 2013), the LIWC software (Pennebaker, Boyd, Jordan, & Blackburn, 2015), and the Hedonometer data (Dodds, Harris, Kloumann, Bliss, & Danforth, 2011).

Figure 3.1: Daily Average Community Activity (Adjusted), 2 Weeks Pre- and Post-Game Day

(a)



(b)



To measure the dynamic social network structures of the communities, I leveraged the peer-to-peer interactions recorded in the communities pre- and post-game. In particular, two community members were connected in the community's social network if they interacted directly with each other within a discussion thread. For example, if member A created a post or comment, and member B commented on member A's contribution, members A and B would be connected in the peer-to-peer network. Using this link formation rule, I quantified social network cohesion using three metrics: the average number of links per member (*average node degree*), the density of members' local networks (*local clustering coefficient*), and the number of closely-knit groups in the network (*cohesive blocks* – see marti2017network, moody2003structural). Table 3.1 provides summary statistics for the pre- and post-game social cohesion metrics used in this study.

Table 3.1: Average Social Network Metrics per Period (2 Weeks Pre- and Post-Game Day)

	Period		T-test	
Average per Period	Before	After	T-value	P-value
Degree	7.084	7.082	0.10	0.92
Clustering Coefficient	0.54	0.55	-4.25	$< .001$
N. Cohesive Blocks	23.46	23.48	-0.075	0.94
N. Cohesive Blocks (Adjusted)	1	1.29	-33.68	$< .001$

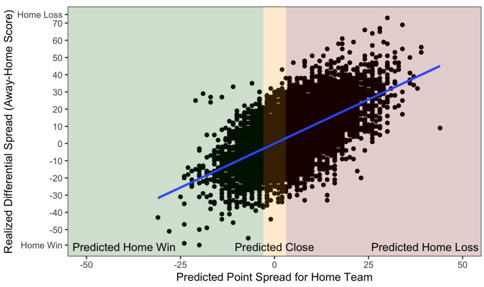
Core and periphery members I classified each community member in the data into *core* versus *periphery* using a binary indicator. To identify core-vs-periphery members, I used the rich core-periphery algorithm of Ma and Mondragón (2015). Across subreddits and games, 51% of the community members are classified as periphery members, 10% as core members, and 39% as “new members”, who only activated during the 2 weeks following the games.

Control variables I constructed several variables that I use as covariates and econometric controls: a binary indicator for whether the focal team was highly ranked in the relevant season (based on the Associated Press top-25 ranking); an indicator for the season period in which a game was played (i.e., first- or second-half of the season); an indicator for weekend or weekday game; and two count variables for the cumulative wins or losses accumulated during the current season, which reset after a winning or losing streak is interrupted.

One particularly important control variable is the closing point spreads, provided by independent bookmaker markets before each game. The bookmakers’ closing point spreads represent the expected likelihood (and magnitude) of a win or loss for a particular game, and are formulated and synthesized by multiple prediction markets and bookmakers. More specifically, the closing point spreads which I collected are not updated anymore after the game begins. Therefore, they represent the most accurate approximation of the pre-game market expectations about the outcome of a certain game. In the empirical analysis, I adopt a similar categorization as card2009family, dividing the predicted spreads into three regions: "predicted draw" (predicted point spread equal to 0); "predicted close outcome" (predicted point spread between ±4); and "clear predicted outcome" (predicted point spread larger than ±4). In order to verify the assumption that the point spreads are sufficiently unbiased predictors of game outcomes, I plot the relationship between the actual and predicted point spread in each game. Figure 3.2 shows that the realized spreads are somewhat noisier than the predicted ones, but the two are indeed strongly positively correlated.

Table B.1 in the Appendix provides summary statistics for all the variables used in the analyses of this study.

Figure 3.2: Market Predicted Outcome and Realized Outcome



3.4 Methods

The key objective of this study is to estimate the effect of shocks to a community's stated goal on the community's dynamics, captured by the levels of activity and social cohesion of community members. To estimate the effects of interest, I use a Diff-in-Diff specification. Technically, I aim at estimating the following equation:

$$Y_{ct} = \beta_T T_t + \beta_D D_c + \delta D_c \times T_t + \gamma X_{ct} + \eta_{ct} + \tau_t + U_{ct} \tag{3.1}$$

Here, Y_{ct} represents the outcome variables – respectively, the adjusted number of daily contributions shared in each community, a vector of social network metrics, or the percentage of words per contribution shared in each community. The binary indicator D_c represents the treatment assigned to a community – measured as the outcome of a game – and takes value 1 if the community was exposed to a loss, and 0 to a win. I assume that for each Reddit community, treatment exposure happens on the day of the game. In the estimations, I consider a time window of 30 days (15 days before and after each game), $t \in [-15, 15]$, in which I track activity, social network metrics, and user-generated content.[5] $T_t \in \{\text{Pre, Post}\}$ is a time indicator for whether the outcome variables are measured in $t \in [-15, 0)$ or $t \in (0, 15]$.[6] X_{ct} is a vector of control variables – including the closing point spread for the reference game, the team's ranking in the league and season, the cumulative losses for the team and season up to game day, and the binary indicators for seasonality. η_{ct}, τ_t are vectors of community-month-year and week-month fixed effects. The community-month and week-month fixed effects control for time-invariant differences in the underlying contribution levels and cohesion in each community. Additionally, these controls account for trends in the popularity of certain seasons, games, or communities. U_{ct} is an unobserved error term.

3.4.1 Identification Strategy

The parameter δ can be interpreted as an estimate of a causal effect of game results on the outcome variables (activity, content, and network metrics). To identify δ, I need to account for unobserved factors that correlate both with the outcome of a NCAA game, and with the likelihood that Reddit members discuss and interact online. This source of endogeneity may introduce bias in the estimation. To correctly identify δ, I need to assume that the error term is not differentially correlated with unobserved factors – in other words, that the consequences of unobserved factors do not affect Reddit members' behavior differentially after losses and wins, conditionally on the fixed effects and the controls. A particularly important unobserved confounder (that could differentially impact outcomes after losses and wins, and may not be necessarily captured by the fixed effects) is the prior expectations that Reddit members have about the game outcomes (Card & Dahl, 2011). In the estimation of δ, I account for Reddit members' expectations about the treatment by adding the bookmakers' closing point spreads in the vector of controls X_{ct}. Closing spreads are the most

[5] I run robustness checks with varying window sizes in Section 3.6.

[6] To exclude game-dependent residual variation not captured by the control variables, game days $t = 0$ are excluded from the analyses.

updated prediction of the outcome of a game, produced across multiple prediction markets – and, therefore, represent the most accurate approximation of pre-game market expectations. Following this discussion, formally I assume that:

Assumption 1. $corr(D_c \times T_t, U_{ct} | T_t, D_c, X_{ct}, \eta_{ct}, \tau_t) = 0.$

Note that, while the estimation panel includes games with varying realized score differences, I do not have a proper set of control games for which the score difference was exactly 0 (i.e. a "neutral" outcome). In presenting the results from the estimation, I discuss δ as the effect of a loss against the baseline of a win – after controlling for market expectations, and other time- and community-varying factors.

In the following section, I am going to assess the differential impact of game outcomes across losses vs wins on community activity, social network cohesion, and internal community structure. All the models are estimated using the *miceadds* package in R and the *lm.cluster* function, with cluster-robust standard errors at the month-year level (Robitzsch & Grund, 2021).

3.5 Results

In this section, I estimate the effects of negative-versus-positive disruptions to a community's purpose on several indicators of community dynamics, using the model specification in Eq. 4.2. If a game loss – as a negative shock to the community's shared purpose – boosts participation or cohesion among community members, then I expect the coefficient δ to be positive and significant. If, instead, losses have a detrimental effect on community participation and cohesion, then I expect that δ should be significant and negative. Finally, if δ is not significantly different from zero, then the interpretation is that the shocks to the purpose of the community have no effect on its dynamics.

I estimate this model on different outcome variables. In Section 3.5.1, I assess the impact of losses – as negative shocks to the communities' shared purpose – on the volume of daily community contributions. In Section 3.5.2 I investigate whether these negative shocks affect core and periphery members differently. In Section 3.5.3 I show to what extent negative shocks impact a variety of social network metrics measuring different aspects of network cohesion. In Section 3.5.4 I measure whether negative shocks change the type of user-generated content that is shared in the impacted communities. Finally, in Sections 3.5.5 and 3.5.5 I explore the heterogeneous roles of community expectations and content moderation in the effect of negative shocks, to the community purpose, on community dynamics.

3.5.1 The Effect of Negative-vs-Positive Events on Community Activity

Table 3.2 demonstrates that negative shocks to a community's purpose – i.e., game losses in respect to fan communities – cause a decrease in community activity relative to wins. The results suggest that, compared to the average pre-game activity levels and relative to wins, community members contribute fewer posts and comments in their communities after losses. This effect is statistically significant and robust to the inclusion of relevant controls and fixed effects (columns 1-3 in Table 3.2). The negative coefficient in column 3 implies an average 33.3% decrease in the number of daily contributions to the online communities over the post-game periods (column 3 of Table B.2).

These baseline results support the conclusion that the ongoing dynamics of online communities *are* related to the shared purpose of the community. This suggests that the community purpose may be acting as an ongoing "fuel" for community engagement and resilience, and not only as the starting point for community formation. Furthermore, in this context, it seems that negative shocks to the purpose likely reduce the benefits or value that members extract from participating in the community, and from interacting with peers. In other words, it seems that a threat to the purpose of the community does not encourage members to invest additional time and effort in nurturing the disrupted community, but to prefer to step back and disengage. Next, I study whether different sub-groups within the community react differently to purpose disruptions.

Table 3.2: Effect of negative events on adjusted daily contributions per subreddit

	Dependent Variable: Adjusted Daily Contributions		
	(1)	(2)	(3)
Loss × Post-Game	-0.0430^{***}	-0.0439^{***}	-0.0438^{***}
	(0.019)	(0.019)	(0.019)
Loss	0.000	0.001	0.002
	(0.000)	(0.001)	(0.005)
Post-Game Period	1.210^{***}	1.262^{***}	1.269^{***}
	(0.026)	(0.022)	(0.024)
Seasonality Controls	No	No	Yes
Team Popularity Controls	No	No	Yes
Predicted Point Spreads	No	No	Yes
Subreddit-month FE	No	Yes	Yes
Week-year FE	No	Yes	Yes
R^2	0.526	0.5262	0.5263
Num. obs.	297059	297059	297059

Robust standard errors clustered at the month-year level in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Estimating equation: $Y_{ct} = \beta_T T_t + \beta_D D_c + \delta D_c \times T_t + \gamma X_{ct} + \eta_{ct} + \tau_t + U_{ct}$. All specifications include subreddit-month and week-year fixed effects.
DV: New daily subreddit contributions divided by pre-game average contributions within subreddit. Treatment: Loss by focal team. Seasonality controls: first-half of season binary, weekend binary, number of cumulative losses in the season until game date. Team popularity controls: AP top-25 ranking binary. Predicted point spreads control: categorical point spread indicators – predicted draw, predicted close, clear predicted outcome.

3.5.2 Is There a Differential Effect for Core, Periphery, and New Members?

To extract more insight into the roles of different community members responding to purpose shocks, I evaluate whether the baseline decrease in contributions occurs differently for core versus periphery members. The results in Table 3.3 show that the activity from both the core and the periphery decreases after negative events; however, the reduction in activity for core members is steeper. This is true in terms of daily contributions, of number of daily active authors, and of daily contributions per author shared in the community (columns 1-2, 4-5, 7-8 in Table 3.3). In addition to affecting core and periphery members, negative purpose shocks also have an effect on the inflow of new members to the community. These "bandwagon fans" – or "lurkers" – are people who usually only become active after a team wins devalck2009. In particular, I find that negative shocks cause fewer "new" members to activate and contribute to the community for the first time after the event (columns 3 and 7 of Table 3.3). Interestingly, while fewer new members join the community after losses, they contribute more content per capita after negative events, compared to positive events (column 9 of Table 3.3).

Table 3.3: Negative Events and Community Contributions by Core, Periphery, and New Members per Subreddit

	Dependent Variables:								
	Daily Contributions			Daily Active Authors			Daily Contributions per Capita		
	Core (1)	Periphery (2)	New (3)	Core (4)	Periphery (5)	New (6)	Core (7)	Periphery (8)	New (9)
Loss × Post-Game	-0.038^{***}	-0.026^{***}	-0.293^{***}	-0.027^{***}	-0.018^{***}	-0.183^{***}	-0.033^{***}	-0.014^{**}	0.010^{***}
	(0.007)	(0.005)	(0.076)	(0.005)	(0.004)	(0.052)	(0.006)	(0.005)	(0.002)
Loss	0.012^{*}	0.002	-0.284^{***}	1.015^{***}	1.026^{***}	0.079	0.009	-0.002	-0.002
	(0.005)	(0.004)	(0.062)	(0.004)	(0.004)	(0.044)	(0.005)	(0.004)	(0.002)
Post-Game Period	0.540^{***}	0.499^{***}	10.707^{***}	-0.479^{***}	-0.609^{***}	7.793^{***}	0.740^{***}	0.832^{***}	1.279^{***}
	(0.005)	(0.004)	(0.063)	(0.003)	(0.003)	(0.035)	(0.005)	(0.004)	(0.002)
Seasonality Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Team Popularity Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Predicted Point Spreads	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subreddit-month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.242	0.209	0.351	0.272	0.292	0.380	0.357	0.687	0.893
Num. obs.	292812	292812	292812	292807	292807	292807	293758	293758	281153

Robust standard errors clustered at the month-year level in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$.
Estimating equation: $Y_{ct} = \beta_T T_t + \beta_D D_c + \delta D_c \times T_t + \gamma X_{ct} + \eta_{ct} + \tau_t + U_{ct}$. All specifications include subreddit-month and week-year fixed effects.
DVs: (1) Adjusted Core Daily Contributions; (2) Adjusted Periphery Daily Contributions; (3) Newly Active Members' Daily Contributions; (4) Adjusted Daily Active Core Authors; (5) Adjusted Daily Active Periphery Authors; (6) Daily Newly Active Members; (7) Adjusted Contributions per Core Member; (8) Adjusted Contributions per Periphery Member; (9) Contributions per Newly Active Member. The outcome adjustment is performed with respect to the pre-game period mean within member types (i.e. by subreddit, game, and member type sub-group).
Treatment: Loss by focal team. Seasonality controls: first-half of season binary, weekend binary, number of cumulative losses in the season until game date. Team popularity controls: AP top-25 ranking binary. Predicted point spreads control: categorical point spread indicators – predicted draw, predicted close, clear predicted outcome.

Note: the post-game coefficients for "new members" are, by construction, large and positive due to the "new member" definition. Since "new members" have 0 pre-game activity, and only activate post-game, the coefficient for post-game is naturally positive. Therefore, I only interpret the differential effect of lost-versus-won games on this cohort of community members.

This set of results suggests that a negative shock to the purpose of a community can have a disruptive effect on its composition. Specifically, the community as a whole is negatively affected by a purpose shock, but the core of the community experiences a stronger disruption compared to the periphery. This result is in line with the expectation that core members experience a stronger association with the state of the community's purpose. The results are also consistent with a scenario in which the core members disengage from the community because they do not perceive that investing in the community during times of crises would pay off in the long term. Finally, the results are also consistent with the "love-becomes-hate effect", according to which people with higher levels of commitment to a shared purpose engage in stronger desires for avoidance after performance failures. The dependency of periphery members on the integrity of the shared purpose appears weaker.

3.5.3 Do Negative Purpose Shocks Affect Social Network Cohesion?

I now turn to analyze the impact of negative shocks on the structure of the social networks underlying the online communities, focusing on network metrics that correlate with social cohesion. The effect of external shocks on network structure in real-life situations is under-researched, mainly due to the difficulty of finding appropriate empirical settings to perform causal inference. In particular, I measure the impact of negative purpose shocks on the average network degree, clustering coefficient, and number of cohesive blocks in the communities. I use the average degree as an indicator of how many distinct peers each community member interacts with, in the pre- vs. post-game period. The clustering coefficient measures how tightly-knit the interactions are, and how likely it is that the interactions occur within social cliques (i.e., between friends of friends). Finally, the number of cohesive blocks represents the number of distinct sub-groups of people which mainly interact within that sub-group, rather than interact with other sub-groups in the network. I use this measure to observe to what extent the community "tolerates" the existence of several distinct sub-groups during and following disruption.

The results in Table 3.4 show that after negative events, community members interact with fewer unique peers (column 1). Related to that, the analysis of the clustering coefficient shows that, instead of completely disengaging, members turn their efforts towards their closer-knit cliques (column 2). Finally, the number of sub-groups is reduced when the community is disrupted (column 3), suggesting that under a purpose disruption, the community only enables the existence of a smaller number of cohesive sub-groups.

In the context of this study, the findings regarding the communities' social networks suggest that the interactions and discussions among online community members are less diverse, involve fewer unique peers, and imply more cohesion within the ego networks of members. Furthermore, the results suggest that there are fewer separate sub-groups of peer-to-peer discussions within the communities. These findings have two implications. First, they show that the shared purpose of the community

Table 3.4: Negative Events and Network Cohesion Metrics

	Dependent Variables:		
	Degree Centrality (1)	Clustering Coefficient (2)	N. Cohesive Blocks (3)
Loss × Post-Game	-0.076^{*}	0.004^{**}	-0.036^{*}
	(0.109)	(0.003)	(0.023)
Loss	-0.206^{***}	0.000	-0.006
	(0.109)	(0.006)	(0.009)
Post-Game Period	6.322^{***}	0.562^{***}	1.339^{***}
	(0.368)	(0.007)	(0.027)
Seasonality Controls	Yes	Yes	Yes
Team Popularity Controls	Yes	Yes	Yes
Predicted Point Spreads	Yes	Yes	Yes
Subreddit-month FE	Yes	Yes	Yes
Week-year FE	Yes	Yes	Yes
R^2	0.307	0.604	0.746
Num. obs.	1592116	1592116	24412

Robust standard errors clustered at the month-year level in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Estimating equation: $Y_{ct} = \beta_T T_t + \beta_D D_c + \delta D_c \times T_t + \gamma X_{ct} + \eta_{ct} + \tau_t + U_{ct}$. All specifications include subreddit-month and week-year fixed effects.
DVs: (1) Average degree centrality *per user* in post-game period; (2) Average clustering coefficient *per user* in post-game period; (3) N. cohesive blocks *per community* in post-game period, adjusted by pre-game average. Treatment: Loss by focal team. Seasonality controls: first-half of season binary, weekend binary, number of cumulative losses in the season until game date. Team popularity controls: AP top-25 ranking binary. Predicted point spreads control: categorical point spread indicators – predicted draw, predicted close, clear predicted outcome.

does play a role in the day-to-day dynamics of the network structure of the community – and most likely, its resilience. Second, it seems that a disruption to the purpose of the community can cause its social network to be more *fragmented*, with a less diverse set of the ongoing social interactions.

3.5.4 The Effect of Negative Purpose Shocks on User-Generated Content

In this section, to gain further insight, I study the effect of purpose disruptions on the user-generated content (UGC) shared in the online communities. In particular, I assess the impact of negative purpose shocks on several dimensions of the user-generated text. The first dimension is *affect* – measured by coding valence, arousal, happiness, and the percentage of positive and negative words shared in the UGC. The other interesting dimensions are the use of words related to *cognitive processes*; the *time orientation* of the discussions (i.e. oriented towards the past, present, or future); and the use of words related to *group affiliation*. The objective of the UGC analysis is to observe whether the emotional climate in the community, expressed through the content and emotion shared in the UGC, is affected by purpose disruptions. This analysis should provide further insight into how community members try to cope with a purpose shock.

Affect, positive emotions, and emotion intensity

Table 3.5 reports the estimated effects of negative purpose shocks on the extent to which community members use positive and intense emotional words. I find that a negative purpose shock causes a reduction in the use of all forms of affect – i.e., valence, arousal, happiness, and positive emotions. However, only the decrease in arousal is statistically significant. In particular, members share −0.6% fewer emotionally intense words following a negative shock, compared to a positive one (column 3).

Table 3.5: Negative Events and Text Valence, Arousal, and Happiness Metrics

	Dependent Variable: Daily Average % Words per Contribution			
	Valence (1)	Arousal (2)	Happiness (3)	Positive Emotions (4)
Loss × Post-Game	−0.003	−0.006**	−0.002	−0.038
	(0.003)	(0.002)	(0.002)	(0.057)
Loss	−0.001	−0.002	−0.001	−0.018
	(0.003)	(0.002)	(0.002)	(0.037)
Post-Game Period	5.872***	4.078***	5.478***	6.016***
	(0.006)	(0.004)	(0.002)	(0.049)
Seasonality	Yes	Yes	Yes	Yes
Team Popularity	Yes	Yes	Yes	Yes
Predicted Point Spreads	Yes	Yes	Yes	Yes
Subreddit-month FE	Yes	Yes	Yes	Yes
Week-year FE	Yes	Yes	Yes	Yes
R^2	0.996	0.996	0.999	0.494
Num. obs.	277644	277644	277644	277644

Robust standard errors clustered at the month-year level in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Estimating equation: $Y_{ct} = \beta_T T_t + \beta_D D_c + \delta D_c \times T_t + \gamma X_{ct} + \eta_{ct} + \tau_t + U_{ct}$. All specifications include subreddit-month and week-year fixed effects.
DV: Percentage of Words in Text. Happiness is measured using the Hedonometer dictionary (). Treatment: Loss by focal team. Seasonality controls: first-half of season binary, weekend binary, number of cumulative losses in the season until game date. Team popularity controls: AP top-25 ranking binary. Predicted point spreads control: categorical point spread indicators – predicted draw, predicted close, clear predicted outcome.

Negative emotions and group affiliation

Table 3.6 presents the results for the percentage of words in UGC that relate to negative emotions and group affiliation. Although I find a minute increase in the percentage of sadness-related words, and a small decrease in anger-related words after a negative shock, none of these effects seem to be statistically significant (columns 1-2). As is the case for positive emotions, I do not observe a meaningful effect of disruption on valence. On the other hand, I do observe a significant effect of purpose disruptions on the use of language relating to affiliation and group-related words. Compared to after wins, community members share 5% fewer words related to sense of affiliation (column 3) and 3.5% fewer words related to sense of group after a loss (column 4).

Table 3.6: Negative Events and Negative Emotions Text Metrics

	Dependent Variable: Daily Average % Words per Contribution			
	Sadness (1)	Anger (2)	Affiliation (3)	Group (4)
Loss × Post-Game	0.003	-0.012	-0.050^{*}	-0.035^{**}
	(0.010)	(0.022)	(0.022)	(0.013)
Loss	0.001	$-0.026^{\dagger}$	-0.050^{**}	$-0.019^{\dagger}$
	(0.008)	(0.014)	(0.023)	(0.019)
Post-Game Period	0.477^{***}	1.000^{***}	3.031^{***}	1.261^{***}
	(0.013)	(0.044)	(0.048)	(0.043)
Seasonality	Yes	Yes	Yes	Yes
Team Popularity	Yes	Yes	Yes	Yes
Predicted Point Spreads	Yes	Yes	Yes	Yes
Subreddit-month FE	Yes	Yes	Yes	Yes
Week-year FE	Yes	Yes	Yes	Yes
R^2	0.092	0.111	0.441	0.228
Num. obs.	277644	277644	277644	277644

Robust standard errors clustered at the month-year level in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Estimating equation: $Y_{ct} = \beta_T T_t + \beta_D D_c + \delta D_c \times T_t + \gamma X_{ct} + \eta_{ct} + \tau_t + U_{ct}$. All specifications include subreddit-month and week-year fixed effects.
DV: Percentage of Words in Text. Treatment: Loss by focal team. Seasonality controls: first-half of season binary, weekend binary, number of cumulative losses in the season until game date. Team popularity controls: AP top-25 ranking binary. Predicted point spreads control: categorical point spread indicators – predicted draw, predicted close, clear predicted outcome.

These findings imply that, overall, purpose disruptions reduce the intensity of emotion in UGC, and the general community-oriented atmosphere. These results seem to be consistent with the overall disengagement from the community I observe – and especially the disengagement of the core members, the social leaders in the community.

Cognitive processes and temporal focus

Table 3.7 shows that negative shocks to a community's purpose, largely, do not affect the cognitive aspects of the interaction within the community (columns 1-3). On the other hand, negative shocks do seem to cause members to focus less on past events:

community members share 4% fewer words that focus on a past time after losses, than after wins (column 4).

Table 3.7: Negative Events and Text Metrics for Cognitive Aspects

	Dependent Variable: Daily Average % Words per Contribution					
	Cognitive Processing	Interrogation	Certainty	Focus on Past	Focus on Present	Focus on Future
	(1)	(2)	(3)	(4)	(5)	(6)
Loss × Post-Game	0.043	0.005	0.001	-0.040^{*}	0.032	-0.010
	(0.029)	(0.014)	(0.018)	(0.018)	(0.046)	(0.020)
Loss	0.108^{***}	0.004	-0.010	$0.032^{\dagger}$	-0.007	-0.005
	(0.036)	(0.011)	(0.012)	(0.012)	(0.056)	(0.018)
Post-Game Period	11.737^{***}	1.261^{***}	1.534^{***}	3.355^{***}	12.186^{***}	1.527^{***}
	(0.093)	(0.017)	(0.028)	(0.047)	(0.183)	(0.027)
Seasonality	Yes	Yes	Yes	Yes	Yes	Yes
Team Popularity	Yes	Yes	Yes	Yes	Yes	Yes
Point Spreads	Yes	Yes	Yes	Yes	Yes	Yes
Subreddit-month FE	Yes	Yes	Yes	Yes	Yes	Yes
Week-year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.852	0.398	0.364	0.590	0.849	0.394
Num. obs.	277644	277644	277644	277644	277644	277644

Robust standard errors clustered at the month-year level in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Estimating equation: $Y_{ct} = \beta_T T_t + \beta_D D_c + \delta D_c \times T_t + \gamma X_{ct} + \eta_{ct} + \tau_t + U_{ct}$. All specifications include subreddit-month and week-year fixed effects. DV: Percentage of Words in Text. Treatment: Loss by focal team. Seasonality controls: first-half of season binary, weekend binary, number of cumulative losses in the season until game date. Team popularity controls: AP top-25 ranking binary. Predicted point spreads control: categorical point spread indicators – predicted draw, predicted close, clear predicted outcome.

To summarise, after being exposed to a negative shock to their common purpose, online community members share less emotionally intense content, use language which expresses less connection with each other and with the group, and reduce mentions of past events, with no effect on mentions of present and future events. These findings are consistent with the findings regarding engagement and social cohesion - indicating an overall relatively minimal coping strategy and disengagement.

3.5.5 The Role of Expectations and Moderation in Online Communities

In this section, I assess some of the boundaries of the baseline effects, and further demonstrate the managerial relevance of the baseline findings. To do that, I focus on two aspects of community dynamics that could be particularly relevant for community managers: expectations management and content moderation.

Community expectations First, I study the role of expectations in the way communities react to negative purpose shocks. Particularly, I assess whether expectations towards an event can mitigate the negative effects I observe on the community dynamics. I leverage the knowledge of the bookmakers' market predictions about the game outcomes, to partition the data into four subsets (A - D). Subset A includes all games for which the realized outcome was *opposite* to the predicted outcome, and for which the predicted point spread was larger than 3 points (i.e., a clear predicted outcome). I refer to this subset as the set of games in which I have *disconfirmed* expectations. For robustness, I also create subset B, in which I use a 5-point instead of a 3-point threshold. Subset C includes all games for which the realized outcome matched with the predicted outcome, and for which the predicted point spread larger than 3 points (*confirmed* expectations, clear predicted outcome). Again, I created subset D exactly as I did subset C, but using a 5-point threshold. I estimate Eq. 4.2 on each of these subsets, using the adjusted daily community contributions as the outcome. The estimations of the robustness subsets B and D are reported in Table B.4.

Table 3.8 reports the results of the estimations on subsets A and C. Column 1 implies that, when the community expectations are *disconfirmed*, a negative shock to the purpose of the community decreases engagement considerably. More specifically, losses cause a 2.7% decrease in the number of daily contributions generated in the online communities compared to wins. Column 3 of Table 3.8 shows that, in case of an *expected* loss, community contributions decrease, but at a lower rate than in case of an unexpected one. This last result suggests that the negative impact of losses on adjusted daily contributions is indeed mitigated when the community expectations are confirmed by the realized outcome.

These findings suggest that expectations may play a practical role in community management. When community members' expectations are met, negative shocks to their shared purpose still hinder the sustainability of community engagement, but to a lower extent than when their prior expectations are disconfirmed. This suggests that expectation management is an important tool for managers and marketers involved in the management of successful online communities.

Table 3.8: Negative Events and Community Activity – Disconfirmed vs. Confirmed Outcomes

	Dependent Variable: Adjusted Daily Contributions	
	Disconfirmed, ±3 points	Confirmed ±3 points
	(1)	(2)
Loss × Post-Game	-0.068^{***}	-0.038^{**}
	(0.016)	(0.014)
Loss	0.014	-0.006
	(0.014)	(0.011)
Post-Game Period	1.275^{***}	1.250^{***}
	(0.010)	(0.012)
Seasonality Controls	Yes	Yes
Team Popularity Controls	Yes	Yes
Predicted Point Spreads	Yes	Yes
Subreddit-month FE	Yes	Yes
Week-year FE	Yes	Yes
R^2	0.485	0.562
Num. obs.	118475	100652

Robust standard errors clustered at the month-year level in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.
Estimating equation: $Y_{ct} = \beta_T T_t + \beta_D D_c + \delta D_c \times T_t + \gamma X_{ct} + \eta_{ct} + \tau_t + U_{ct}$. All specifications include subreddit-month and week-year fixed effects.
DV: New daily subreddit contributions divided by pre-game average contributions within subreddit. Games for which predictions were: (1) disconfirmed, predicted spread > ±3 points; (2) disconfirmed, predicted spread > ±5 points; (3) confirmed, predicted spread > ±3 points; (4) confirmed, predicted spread > ±5 points. Treatment: Loss by focal team. Seasonality controls: first-half of season binary, weekend binary, number of cumulative losses in the season until game date. Team popularity controls: AP top-25 ranking binary. Predicted point spreads control: categorical point spread indicators – predicted draw, predicted close, clear predicted outcome.

Content moderation The last analysis I perform is related to an aspect of online communities which managers and marketers may readily influence: the extent and timing of content moderation. I create a new binary indicator, $M_{c,T<t_0}$, for whether the community received any content moderation in the pre-game period. The indicator takes value 1 if any community post or comment was removed by moderators in the pre-game period – signaling the fact that the moderators were monitoring and managing user-generated content – and 0 otherwise. Then, I estimate the heterogeneous effect of negative shocks – in presence and in absence of content moderation – using the following specification:

$$\begin{aligned} Y_{ct} = \quad & \beta_T T_t + \beta_D D_c + \beta_M M_{c,T<t_0} + \\ & + \delta_{DT} D_c \times T_t + \beta_{TM} M_{c,T<t_0} \times T_t + \beta_{DM} M_{c,T<t_0} \times D_c + \\ & + \delta M_{c,T<t_0} \times D_c \times T_t + \\ & + \gamma X_{ct} + \eta_{ct} + \tau_t + U_{ct} \end{aligned} \tag{3.2}$$

Where $M_{c,T<t_0}$ is the content moderation indicator. I aim at estimating parameter δ as capturing the heterogeneity in the effect of purpose disruptions across communities subject to content moderation prior to the shock, compared to communities without any moderation.

Table 3.9 shows that content moderation mitigates the baseline negative effect of losses on the adjusted daily contributions in the affected communities (column 1). Furthermore, the positive heterogeneous effect of content moderation on community participation is stronger when content moderation occurs closer to the event day (column 2). One possible explanation for this result is that, if content moderation is enforced prior to a negative event, members may feel that the community managers are committed to invest in the community. Community members may interpret the enforcement of content moderation as signal that, in spite of the threat to its purpose, there is a higher probability that the community will survive and provide them with long-term benefits. I cannot rule out alternative explanations – which include unobserved correlations between moderator characteristics and community-level variables. One of these alternative explanations could be that the communities in which moderators tend to be more active before games are also more intrinsically resilient to external shocks.

Table 3.9: Negative Events and Adjusted Daily Contributions per Subreddit

	Dependent Variable: Adjusted Daily Contributions	
	Moderation 15 days pre-game (1)	Moderation 48 hours pre-game (2)
Loss × Post-Game × Community Moderation	0.012	0.043^{*}
	(0.034)	(0.028)
Loss	0.003	0.001
	(0.006)	(0.005)
Post-Game	1.393^{***}	1.308^{***}
	(0.034)	(0.028)
Community Moderation	−0.000	0.003
	(0.004)	(0.005)
Loss × Post-Game	-0.052^{***}	-0.056^{***}
	(0.033)	(0.024)
Loss × Community Moderation	0.0004	−0.0003
	(0.002)	(0.002)
Post-Game × Community Moderation	-0.220^{***}	-0.169^{***}
	(0.034)	(0.028)
Seasonality Controls	Yes	Yes
Team Popularity Controls	Yes	Yes
Predicted Point Spreads	Yes	Yes
Subreddit-month FE	Yes	Yes
R^2	0.529	0.527
Num. obs.	297059	297059

Robust standard errors clustered at the month-year level in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Estimating equation: $Y_{ct} = \beta_T T_t + \beta_D D_c + \delta D_c \times T_t + \beta_M M_{c,T<t_0} + \beta_{TM} M_{c,T<t_0} \times T_t + \beta_{DM} M_{c,T<t_0} \times D_c + \delta_M D_c \times T_{t\,c,T<t_0} + \gamma X_{ct} + \eta_{ct} + \tau_t + U_{ct}$. All specifications include subreddit-month fixed effects.
Heterogeneity analysis: (1) Subreddits and games that received any moderation in the 15-day pre-game period ($M_{c,T<t_0} = 1$, otherwise $M_{c,T<t_0} = 0$); (2) Subreddits and games that received any moderation in the 48 hours preceding the game ($M_{c,k<T<t_0} = 1$, otherwise $M_{c,k<T<t_0} = 0$, and k = 2 days).
DVs: New daily subreddit contributions divided by pre-game average contributions within subreddit. Treatment: Loss by focal team. Seasonality controls: first-half of season binary, weekend binary, number of cumulative losses in the season until game date. Team popularity controls: AP top-25 ranking binary. Predicted point spreads control: categorical point spread indicators – predicted draw, predicted close, clear predicted outcome.

3.6 Baseline Results with Varying Window Specifications

In this section, I provide alternative estimates of δ from Eq. 4.2 on the volume of adjusted community contributions. Specifically, I aim at understanding the magnitudes of the main results, and the time permanence of the effects. I report what happens before and after a negative shock to the communities' purpose, for windows of 2 days and 7 days after the games, while keeping a fixed 15-day window before the games. I estimate the same specification as Eq. 4.2, for $t_0 - 15 < t < t_0 + k$. The results, presented in Table 3.10, show that the baseline negative effect of losses (reported in column 3 of Table 3.2) is robust across varying post-game time windows. More in detail, the effect of a loss on the volume of adjusted community contributions is larger in the 48 hours following the event (column 1 in Table 3.10). This time window likely captures the most immediate and intense reactions of the community members to the losses. Then, the negative coefficient decreases in magnitude over the 1 and 2 weeks following the losses (column 2 in Table 3.10 and column 3 in Table 3.2). These longer-term effects suggest that the disruptions caused by losses on the affected communities may loom for at least 2 weeks following the negative events.

Table 3.10: Average Adjusted Daily Contributions: Different Windows

	Dependent variable: Adjusted Average Daily Contributions	
	2 days post-game	7 days post-game
	(1)	(2)
Loss × Post-Game	-0.051^{***}	-0.043^{***}
	(0.018)	(0.016)
Loss	0.005	0.002
	(0.002)	(0.003)
Post-Game Period	1.252^{***}	1.265^{***}
	(0.022)	(0.023)
Seasonality Controls	Yes	Yes
Team Popularity Controls	Yes	Yes
Predicted Point Spreads Control	Yes	Yes
Subreddit-month FE	Yes	Yes
Week-year FE	Yes	Yes
R^2	0.643	0.592
Num. obs.	168026	218148

Robust standard errors clustered at the month-year level in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.
Estimating equation: $Y_{ct} = \beta_T T_t + \beta_D D_c + \delta D_c * T_t + \gamma X_c + \eta_{ct} + U_{ct}$. All specifications include subreddit-month and week-year fixed effects.
DV: New daily subreddit contributions divided by pre-game average contributions within subreddit. Treatment: Loss by focal team. Seasonality controls: first-half of season binary, weekend binary, number of cumulative losses in the season until game date. Team popularity controls: AP top-25 ranking binary. Predicted point spreads control: categorical point spread indicators – predicted draw, predicted close, clear predicted outcome.

3.7 Conclusion

In a series of analyses, I study whether disruptions to the purpose of online communities impact their levels of engagement, network structure, composition, and user-generated content. The results indicate that disruptions to a community's shared purpose damage the ongoing activity in the affected communities – both in terms of engagement and in terms of community composition. In particular, I find that the effect of the purpose disruption is not homogeneous across community members.

Negative purpose shocks impact the active leaders of the community (i.e., the community *core*) more strongly than the periphery, while also reducing community growth by obstructing the inflow of new members. I also find that negative purpose shocks significantly change the social network metrics correlated with social cohesion in the affected communities. These changes imply that, after negative purpose shocks, fewer community members contribute to the discussions, members interact with fewer of their peers, and community discussions occur more frequently within social cliques. The content of the discussions is also affected: after negative events, people share fewer emotionally intense words, fewer words related to sense of group and affiliation, and fewer words related to the "past times" in the communities. Finally, I show that the disruptive effects may be mitigated through management of the community's expectations regarding the negative event, and through the implementation of content moderation in the community discussions.

The results from this study have several implications for relevant stakeholders, including marketers, managers, and policy-makers working with digital platforms and community-facing channels. First, the findings suggest that stakeholders should monitor the state of the purpose of the community, and strive to keep it in a "healthy" state, as it seems to be fuelling the social dynamics in the community. Stakeholders may wish to constantly develop tactics that remind consumers about their shared interest – the common mission and vision that brought them together in the first place. A "healthy" state of the purpose also enables community growth by attracting new members. Second, the findings show that it may be important to nurture the relationship of the brand with the core members of the community. In times of crises, the motivation of the core members to keep contributing to the collective may be crucial for the community sustainability and growth. In line with prior research in marketing and network science, I suggest that any good-will that the core members hold towards the community or the brand could be useful during purpose-related disruptions. Third, stakeholders should be aware that expectation management can significantly reduce the detrimental consequences of purpose disruptions. the findings suggest that stakeholders may be able to mitigate disruptions by setting realistic expectations about the outcomes of a product crisis or brand failure, if they can. In essence, it seems that it is best that stakeholders keep their consumer communities informed, if they wish to keep leveraging the numerous benefits that come with these communities. I also find suggestive evidence that stakeholders may want to use user-generated content to counteract the negative effect of a purpose shock. These interventions may include promoting more positive affect, and promoting a sense of group-membership and affiliation within the community.

One limitation of this study is that I am only able to compare negative events to positive ones, without a neutral benchmark event. In future work, it would be ideal to compare negative (or positive) shocks against the *absence* of a disruption. This will allow to better isolate the effects of positive and negative events.

In terms of game timing, NCAA-BB Div. 1 teams typically play 2 games per week – one on Wednesday or Thursday, and one on Saturday or Sunday. The bi-weekly frequency of the matches could create problems in the estimation of the effects of interest, as the community members are exposed to possibly competing treatments (e.g., a win and a loss during the same week). In the estimation sample, this situation occurs for 46.6% of the team-match combinations. Therefore, the effects estimated in this paper may be a lower bound to the true effects if, for example, a competing win mitigates the negative effect of a loss during the same week. In the current version of this paper, I addressed this concern with a win (loss)-strike counter. However, the presence of competing treatments may require different modeling assumptions. In future versions of this study, I will address this threat more explicitly, and provide bounds to the effects currently estimated.

A limitation related to the time surrounding the games is that I focus on the 15 days preceding and following each event. In terms of social media time scales, and given the frequency of the matches, this time window should capture an interesting proportion of the activity. However, an investigation of longer time scales would enable a more complete picture of the long-term dynamics triggered by a purpose disruption, after accounting for match frequency.

Furthermore, I observe the correlations between community expectations and community dynamics, and between content moderation and dynamics. While I argue that the interplay between expectations and community dynamics is interesting per se, I can not claim that the effect I observe is causal. A future investigation of the causal mechanisms related to community expectations would be of great value to managers and policy-makers. Another threat to the analysis of community activity and content moderation is that it is unclear whether any team representatives or employees are active in the communities, or even act as moderators. If "team insiders" are present in the communities, the expectations of the online groups may systematically deviate from the expectations of the prediction markets, however efficient. Future studies should investigate the frequency with which "insiders" disseminate information in online communities prior to an otherwise exogenous event, in order to provide bounds on the effects estimated in this paper.

Finally, this study evaluates the impact of negative events in the empirical context of sport communities. Online communities exist in a wide variety of contexts and types. For the sake of completeness, future work should consider an analysis of online communities in other relevant contexts – such as brand-centered communities, communities of product users, knowledge-sharing communities, and communities of organizational teams working together towards a common goal.

Chapter 4

How Do Brand Networks Break in Face of a Crisis?

4.1 Introduction

Online communities have a tremendous importance in the life of consumers, brands, and organizations. Brands use online communities to achieve an array of marketing objectives – such as increasing brand awareness, attracting and retaining loyal and engaged consumers, and improving brands' financial performance (Algesheimer et al., 2010; Bussgang & Bacon, 2020; Fournier & Lee, 2009; Manchanda et al., 2015). Consumers use online communities to connect with the brands they love, to find like-minded people, to solve problems, and to personalize their consumption experiences (Fournier & Lee, 2009). Consumers also use online brand communities to coordinate collective responses to negative brand-related events. So far, the literature has suggested that customer interactions online following brand crises negatively impact brand shareholder value, consumers' brand share, and category purchases (Ahluwalia et al., 2000; Backhaus & Fischer, 2016; Hsu & Lawrence, 2016). However, the impact of brand crises on the behavior of consumers in online brand communities remains unclear. How do brand crises impact the functioning of consumer communities on-

Joint work with Dr. Pinar Yildirim, University of Pennsylvania, the Wharton School and the Leonard Davis Institute; and Dr. Abdullah Almaatouq, MIT Sloan School of Management, MIT Center for Computational Engineering and MIT Connection Science Research Initiative.

line? Which types of consumers are most negatively hit by a brand crisis? How well does information spread in a post-crisis brand network online? It is especially important to investigate these questions, as we recently saw that the coordinated efforts of online community members in response to negative events can go as far as disrupting global financial markets, and steer major policy changes (Fletcher & Aliaj, 2021).

To address these questions, I study the effect of brand crises on the volume and the structure of consumer interactions in online brand communities. First, I study how much brand crises impact the volume of consumer-generated content in online brand communities. Second, I assess how brand crises affect the structure of the network of consumer-to-consumer interactions, focusing on network metrics correlated with speed and ease of information spread. Additionally, I investigate how brand crises may have differential effects on the online behavior of more loyal, experienced consumers relative to less experienced ones. Finally, I explore the scope of the heterogeneity in the estimated effects, across news providers, companies, and crisis types.

I collect and combine data on 154 companies and brands (including their funding status, market relevance, size, and corporate governance), 7805 episodes of brand crises covered by media outlets between January 2010 and September 2019, and consumer-to-consumer discussions in 299 brand-related online communities. In the resulting panel dataset, I track all interactions between consumers in the brand communities for 180 days around any brand crisis – 90 days preceding and 90 days following the crisis events. This results in a panel of 13M posts and comments, generated by 1.9M unique brand community members. I further exploit the thread structure of the brand community discussions to construct social networks of information spread, based on peer-to-peer interactions. In particular, I construct a social network for each brand community, and each pre- and post-crisis period in the panel. In the information networks, community members are connected through a link when they directly commented on a post or comment created by a peer. I leverage the resulting discussion networks to measure how information spreads differently post-crisis, and which consumers occupy a high- or low-importance position in the information ecosystem – both in terms of community engagement before the crisis, and in terms of their position in the brand social networks.

Identifying the causal effect of a brand crisis on the behavior of consumers in online brand communities poses some challenges. There may be unobserved factors – seasonal or related to specific companies – correlated with the media coverage of the event, with the likelihood that certain brands engage in corporate misbehavior, and with the volume of consumer discussions online. To overcome the endogeneity challenges, I adopt a difference-in-difference approach. In assessing how a brand crisis affects consumers' behavior in online brand communities, I control for several crisis, news provider, and brand characteristics. I also control for company-month and week-of-month fixed effects to account for brand-specific, time-varying unobserved factors – such as product launches, concurrent advertising efforts, or other brand news covered in the same period. The identifying assumption is that unobserved determinant of consumers' engagement in brand communities did not differentially affect engagement among high- versus low-type consumers in the brand networks, after controlling for the company-month and week-of-month fixed effects, and for the rest of the covariates.

The results show that after a brand crisis, consumers' activity in the affected brand communities increases by 9.1% on average, compared to the pre-crisis period. However, the change in activity is significantly positive only due to the contributions of "bandwagon consumers" – consumers who become active in brand communities exclusively after a brand crisis is covered by the media. On the contrary, consumers who were already active prior to the crisis event significantly *decrease* their activity in the communities after a brand crisis. Additionally, the rate at which post-crisis activity decreases among this cohort of consumers is not homogeneous. Classifying community members based on their pre-crisis activity levels and their embeddedness in the brand networks, I show that high-type consumers (people who contributed intensely or occupied central network positions before the crisis crisis) contribute relatively more to the brand communities after a brand crisis, compared to low-type consumers. I also find that the effects on activity levels are reflected in the social networks of consumer-to-consumer interactions in the brand communities. This finding suggests that brand crises also significantly alter the ease and speed of information diffusion in brand networks. An average brand crisis causes a 1% increase in degree centrality, and a 0.2% increase in clustering coefficient across network members. Together, these results suggest that, after brand crises, information travels through a more diverse pool of consumers, and in more tight-knit discussion sub-groups.

In terms of user-generated content, I find that brand representation in online communities is also affected by brand crises. After a brand crisis, consumers in brand communities share more words related to negative emotions and conflict. In their discussions, consumers in post-crisis brand networks focus on discussing about past and present events, and use more words related to insight and cognitive processes. Importantly, high-type consumers (compared to low-type peers) share fewer words related to negative emotion, and more discussions on past and present events, using more words related to cognitive processes. Therefore, high-type consumers may act as "emotion regulators" and as contributors of informative content in post-brand crisis information networks.

The effect of brand crises on brand community engagement is differential across types of crises and types of companies. In particular, I find that brand crises have a detrimental effect on consumer activity online when the company operates as business-to-consumer (compared to business-to-business), and when the crises have direct consequences on the health and well-being of the consumers (compared to indirect consequences). The effect of brand crises on community engagement also depends on the intensity of media coverage. In particular, I measure a decrease in the volume of brand community discussions after an international news provider covers a crisis story, compared to crises covered by local or national news providers. Finally, the severity of the crisis consequences also has a differential impact on brand community engagement. I distinguish between high- and low-severity crises – where high severity depends on the gravity of the crisis consequences (for example, in terms of harm to people), the number of people involved, and the extent to which there was intention to harm. I find that, overall, more severe crises have a detrimental impact on community engagement.

To summarise, the results of this study suggest that brand crises are potentially disruptive for the online presence of the affected brands. The communities are effectively "taken over" by bandwagon consumers, while the consumers who were historically active in brand conversations tend to disengage with the brand. While the disengagement of valuable consumers is a threat to the value of brand communities online, an encouraging result is that loyalty preserves communities after a crisis: high-type consumers keep their engagement levels higher than the low-type consumers post-crisis, and regulate the emotional and informative content shared in the post-crisis information networks. These highly involved and experienced community members are also known to have higher lifetime value, resulting from their retention, loyalty, and engagement (Bussgang & Bacon, 2020).

Despite both academic literature and industry consider brand communities an important tool to achieve marketing objectives, there is still ambiguous evidence on the role of these spaces relative to brand crises. On the one hand, brand communities may generate a boost in word-of-mouth, sustained by an increase in brand attention and awareness (Berger et al., 2010; Backhaus & Fischer, 2016). On the other hand, consumers in brand communities may prefer to disengage from the brand involved in the transgression and with the online spaces associated with it (Aaker et al., 2004; Aggarwal, 2004; Roehm & Brady, 2007). The same ambiguity exists regarding the reaction of different types of community members to the same crisis information: the most involved, expert, and attached consumers in consumer-brand networks may be simultaneously more likely to punish and to defend a brand in response to a brand transgression (Aaker et al., 2004; Aggarwal, 2004; Kuchmaner, Wiggins, & Grimm, 2019; Roehm & Brady, 2007). With this project, I contribute to existing research on negative brand reputation and brand crises. I make a substantial contribution by evaluating the consequences of brand crises in the context of brand communities. I also expand on that literature by considering the differential effect of negative brand information of different types of consumers in consumer-brand social networks.

Furthermore, I make a methodological contribution, since I evaluate the impact of brand crises on consumer behavior under quasi-experimental conditions. While experimenting on large scale brand networks online would be ideal to estimate the effects of interest, running such experiments while inducing a reputational damage on existing brands is costly, complex, and ethically problematic (e.g., El-Sayed et al., 2013).

The rest of the paper is organized as follows: in Section 4.2 I provide the institutional background on the use of brand communities for marketing purpose, and on brand crises as disruptive events for consumers' behavior online. In Section 4.3 I provide a summary of the data used in this study. In Section 4.4 I detail the methodological frameworks for the empirical analyses. In Sections 4.5, 4.6, and 4.7 I present the main results, the heterogeneity analyses, and the robustness checks. Section 4.8 concludes.

4.2 Institutional Background & Literature Review

4.2.1 Online Brand Communities and Brand Crises

One of the most popular platforms for community building is Reddit.com. Reddit.com is a discussion website founded in 2005 as a network of communities fueled by user-generated content. To date, Reddit is one of the most important platforms for community formation: as of April 2021, Reddit counts over 52 million active users and more than 100 thousand communities (Reddit Inc., 2021). Thanks to its 50 billion monthly views and its sustained growth over time, Reddit ranks as the 19th most-visited website in the world, and the 7th most-visited website in the U.S. (Alexa Internet Ranking, 2021). Online communities hosted on Reddit – also called *subreddits* – are forums organized around a common interest. Popular examples of common interests on Reddit include breaking news, sports, TV fan theories, and animals. Reddit users can join subreddits to participate in the community discussions, and to receive updates about the discussions in their home feed. Although subreddits are all hosted on Reddit.com, each subreddit is customizable in its appearance, governance, and system of rules and norms. Reddit hosts numerous brand-related subreddits, such as a community for Dyson customers, a community about the partner program at Quora, a community dedicated to Macy's employees, and a subreddit discussing about Walgreens stores. Brand-related subreddits are typically created and managed by customers, brand advocates, or brand users. However, companies and brands can also engage on Reddit. Brands can get involved directly – for example, by creating subreddits about themselves – or indirectly – for example, by creating sponsored posts or headlines in relevant subreddits. Subreddit members can contribute to their subreddits in two ways. Members can create new "submissions", or they can create a "comment" on existing submissions. Creating a submission typically means posting stories, links, images, and videos to the community. Members can leave comments either to existing submissions, or to existing comments. Finally, subreddit members can use Reddit's voting system to show appreciation or disapproval towards the community's submissions and comments.

Subreddits are important online meeting points that people can use during and after crisis events. For example, during the global Covid-19 pandemics, concerned customers, business owners, and employees used Reddit communities to exchange mutual support and answer relevant questions about safety and store re-openings[1]. Subreddits are also used to coordinate collective responses to economic events. In January 2021, the news of a predicted price drop for GameStop stocks caught the attention of the Reddit community /r/wallstreetbets (Lyons, 2021). In reaction to the news, millions of retail investors coordinated a collective response on /r/wallstreetbets to disrupt the trading of GameStop stocks. As a result of this coordination effort, the Reddit investors eventually shorted the market, and caused severe financial consequences for several institutions and investors (Fletcher & Aliaj, 2021).

The Reddit communities of retail investors are only one example of how customers coordinate in online social networks to prepare and respond to external events and crisis information. People continuously rely on their online social networks after receiving crisis information – for example, during terrorist attacks (Burnap et al., 2014), natural or civil disasters (Eismann et al., 2016), financial instability (Racca

[1]For example, customers discussed on this thread about Macy's re-opening after the 2020 lockdowns

et al., 2016; Romero et al., 2016), as well as product recalls and service failures (Hsu & Lawrence, 2016). Coordinated responses to crises in online social networks can even mitigate risk and uncertainty, and enhance crisis response efforts. Access to online social networks during natural or civil disasters significantly affects information exchanges, and the quality and quantity of relevant communications (Lu & Yang, 2011). Following terrorist attacks, information collectively exchanged in online social networks can help stabilize the public response and reduce uncertainty (Jung & Park, 2014). During financial crises, online communities supply news and technical analyses to contrast market uncertainty (Racca et al., 2016). In addition to the information-related role, during crisis events online social networks favor the exchange of opinions to influence response policies, guarantee a platform to coordinate individual actions, and distribute emotional support to those who need it (Qu, Wu, & Wang, 2009).

Consumers also coordinate in online social networks in response to negative publicity and brand crises, with significantly harsh consequences for brands. In particular, the e-Word-of-Mouth generated in online social networks following brand crises has significant consequences on brand equity and post-event market performance. Negative information about brands and brand crises can induce negative changes in customer attitudes, brand evaluation, brand strength, purchase intentions, and financial returns and cash flows (Ahluwalia et al., 2000; Backhaus & Fischer, 2016; Hsu & Lawrence, 2016; Klein & Dawar, 2004; Luo, 2009). The levels of engagement in online brand communities are also predictive of how well brands can recover from negative publicity and crisis events (Yuan et al., 2020). In spite of the importance and the severe consequences of online customer activity towards brand equity, it still unclear how exactly brand crises affect consumer engagement in brand social networks, the patterns of information spread among consumers online, and the representation of the brand in relevant digital spaces. I fill this gap with the present study.

4.2.2 Brand Crises as Information for Customers

In the marketing literature, a "brand crisis" is an unexpected, well-publicized event, that threatens a brand's perceived ability to deliver expected benefits with potential negative effects for brand equity (Backhaus & Fischer, 2016). Brand crises can be extremely disruptive events for brands, and when managed improperly, a brand crisis may propel a firm into a severe, if not existential, crisis (Stäbler & Fischer, 2020). Therefore, it is important to understand the collective reaction of consumers in online brand communities to these disruptive events.

When news providers cover a brand crisis, it represents new brand-related information acquired from a news source by the consumers in brand communities. Literature on information processing suggests that new information about a brand can significantly affect consumers' behavior and their utility from contributing to a brand community (Festinger, 1957; Harmon-Jones & Mills, 2019). More specifically, the theory of cognitive dissonance suggests that brand crisis information may represent a dissonant cognition for brand community members. The negativity of the news is not congruent with a positive prior belief about the brand, which may have motivated customers to join the community in the first place (Festinger, 1957; Harmon-Jones & Mills, 2019). The mechanisms and intensity with which dissonant information affects consumer behavior in brand networks depend on the source credibility, on the information quality, and on the complexity of the issues (Brown, Broderick, & Lee, 2007; Wirtz et al., 2013). Theories on expected utility, impression formation, and negativity effects also suggest that negative brand information can affect the expected utility derived from consumers' choices (Staats, Kc, & Gino, 2018), the extent to which consumers are aware of the brand (Berger et al., 2010), and can be weighted relatively more than positive information in forming overall brand evaluations (Ahluwalia et al., 2000).

Literature in marketing also demonstrates that different customers might react differently to the same brand crisis information. Pre-crisis levels of commitment and identification with the community and/or with the brand are particularly important in evaluating the impact of the crisis on brand networks. Consumer types are typically distinguished on the basis of their pre-event levels of loyalty, commitment, or self-identification with the brand ("high-type" consumers versus "low-type" consumers; Ahluwalia et al., 2000; Cheng, White, & Chaplin, 2012). The enactment of differential coping strategies by high- vs. low-type consumers after brand crises has been documented. However, there is no empirical convergence on how the behavior of different types of consumers should be differently affected by the same brand crisis information, nor empirical evidence of these differential effects in brand communities.

When the behavior of these different segments of consumers is considered, one subset of available evidence suggests that high-type consumers—consumers with higher levels of pre-crisis commitment or identification with the brand—should react less negatively to brand crisis information than low-type consumers. Marketing literature shows that the effects of negative publicity on brand attitudes are mitigated among high-commitment customers, who are more likely to counter-argue the negative information while supporting any positive evidence (Ahluwalia et al., 2000). Similarly, the negative consequences of a product-recall crisis on brand equity are mitigated when customers have stronger positive expectations about a brand, again through a mechanism of selective information processing (Dawar & Pillutla, 2000). Customer experience and expertise with products involved with brand-related crises also moderates customers' response to the crisis information, with prior experience acting as a source of bias in the process of updating existing beliefs (Kalra, Li, & Zhang, 2011). Finally, consumers with stronger attitudes or commitment towards a brand are also more likely to use their prior knowledge to mitigate the disruptive effect of negative brand information (Cheng et al., 2012).

A different set of marketing studies on information processing and consumer-brand relationships suggests that negative brand information may represent a stronger dissonant cognition for high-type members, since the negativity of the news may be too far from – and clash too violently with – their prior beliefs about the brand (Harmon-Jones & Mills, 2019). High-type consumers are also more likely to tie the brand's performance with an interpretation of their own performance, and interpret negative brand information as an intense personal failure (Cheng et al., 2012). This external threat to high-type consumers' positive self-view has, therefore, consequences on their behavior. High-type consumers were also shown to frequently occupy central and embedded positions in social networks of consumer-to-consumer interactions, and to feel stronger psychological ownership towards the brand and their social network (Assmann, Sandner, & Ahrens, 2009; Kuchmaner et al., 2019; Huffaker, 2010; Park & Cho, 2012). The high embeddedness and psychological ownership of high-type consumers in consumer-brand networks can then give rise to inherent conflicts – in which high-type consumers are simultaneously more likely to punish and to defend a brand in response to a brand transgression (Kuchmaner et al., 2019).

In sum, the available evidence suggests that, in response to a brand crisis, I can expect the following: (1) brand crisis information affects consumer activity in online brand communities; (2) brand crisis information affects the patterns of consumer interactions and information spread in brand networks; (3) brand crisis information has a differential effect across customer types.

With respect to the consumer types, I can expect that brand community members who are more loyal or more experienced ("high-type" consumers) may react to brand crisis information by either reducing their exposure to the community – for example, in an attempt to reduce the psychological discomfort and threat to their positive self-view – or by increasing their activity in the community – for example, to fix a dissonant world, to counter-argue negative publicity, to update their beliefs, or to share their expertise in time of need. On the other hand, low-loyalty or low-experience members ("low-type" consumers) may face three scenarios. First, low-type consumers do not change their behavior post-crisis, given that the difference between their pre- and the post-crisis brand beliefs may be too small to induce a change. Second, low-type consumers increase their activity, perhaps in an attempt to retrieve more information and reduce their post-crisis uncertainty. Third, low-type consumers decrease their engagement in the brand communities, as the brand crisis may be too severe to be offset by their weak prior expectations about the brand, or by temporarily high information needs.

Finally, I can expect that high-type consumers experience higher levels of psychological discomfort, a stronger threat to self-views, or a stronger drive to fix their dissonant cognition and share their expertise, compared to low-type consumers. Therefore, I can expect differences in the magnitude of the reaction to the brand crisis information between high- and low-type consumers – with a stronger reaction originating from the high-type, compared to the low-type consumers.

4.3 Data

For the empirical analyses in this study, I use data on (i) brand crises, (ii) companies and brands involved in the brand crises, and (iii) online brand communities. To construct a sample of companies and their brand crises, I obtained a comprehensive list of corporate social irresponsibility (CSI) events from the RepRisk ESG Risk Platform. The list includes details about companies that engage in environmental, social, and governance risks and business misconduct, with information about the nature of the misconduct, and about the media outlets that covered and reported the crisis events (RepRisk AG, n.d.). For this study, I focus on companies that headquartered in two large English-speaking countries, the United States and the United Kingdom. I complement the RepRisk dataset with the Crunchbase private company dataset, which includes additional information about the market positioning and performance of the companies such as the main product category that a company offers, the number of employees, a ranking measuring the prominence of each company in the Crunchbase dataset, and the number of funding rounds in which the companies participated. The resulting data include information about 154 companies in 21 product categories and 7805 brand crises. Combining an automated script with manual checks by independent raters, I then identified 299 Reddit communities related to these 154 companies.

Starting from the list of 299 Reddit communities and 7805 brand crisis dates, I generated a dataset of daily posts and comments created by Reddit members in the respective Reddit brand communities. The community activity and interactions records were collected through the Pushshift Reddit Archive (Baumgartner et al., 2020). The data tracks brand community discussions over the three months preceding and following each relevant brand crisis date. The database includes a total of 13M contributions created by 1.9M unique Reddit members between December 2009 and October 2019.

Finally, I combined the brand crisis, company, and community databases to generate estimation datasets. In particular, I created two types of estimation datasets.

The first type is organized and aggregated at the community-crisis-week level. The dataset resulting from this aggregation is a panel, in which I track the total weekly community activity and the average weekly network metrics for 5 weeks before and after any crisis date. The second type of dataset is organized and aggregated at the community-crisis-week and consumer-type level. In this second type of panel, I track total weekly activity and average weekly network metrics by consumer type, for 5 weeks before and after each crisis event. In the rest of the section, I provide additional information and descriptive statistics for the estimation panels.

4.3.1 Brand Crisis Data

The estimation panels include information on 7805 brand crises occurred between January 2010 and September 2019. Figure C.1 describes the number of brand crises recorded in each week and year in the estimation panels. The estimation data distinguish the crisis events by type of issue, how many countries were involved, the severity of the crisis consequences, and the reach and novelty of the news source covering the crises. Among the crisis characteristics, the type of issue is an important quantity to observe. First, the type of issue may determine whether the brand crisis receives more or less media coverage (Stäbler & Fischer, 2020). Second, different issues may also trigger different responses from the community, or prompt different members to modify their social network of interactions. For example, moral information shocks may increase the recruiting rates for protest groups, while financial information shocks cause social networks to become more close-knit and averse to out-group interactions (Jasper & Poulsen, 1995; Romero & Kleinberg, 2010). The RepRisk ESG Risk dataset includes 32 possible types of crisis issues. In Appendix C.1, I map the 32 crisis types into fewer orthogonal factors through an exploratory factor analysis, and classify them into crises that have a direct versus an indirect effect on consumers. Figures C.4a and C.4b in Appendix C.1 show that each brand crisis can involve multiple types of issues – although the majority of events only trigger one type of issue.

In addition to the type of issue raised by the brand crisis, there are other important crisis aspects that can affect consumer behavior in brand networks. For example, brand crises that entail more severe consequence on environment, society, and governance can impact the behavior of brand community members much differently than brand crises with mild societal impact. The same is expected for brand crises that gain international media attention versus local coverage only. The RepRisk ESG Risk dataset distinguishes between 3 levels of severity for the brand crises, and 3 levels of reach for news sources – where 1 is the lowest level and 3 is the highest. The crisis severity is determined by RepRisk based on the consequences of the crisis, the extent of people involved in triggering the crisis, and the cause of the crisis (e.g. accidents, negligence, intent, or systematic incident). The reach of the news source represents an indicator of the influence or readership of the source in which the crisis was published. Low influence sources include, for example, local media, smaller NGOs, or blogs. Medium influence sources include national and regional media, international NGOs, and state, national, and international government bodies. High influence sources include a few international media – such as the NY Times and the BBC (RepRisk AG, n.d.). Finally, the crisis information in the panels also includes 2 levels for the novelty of the crisis coverage by media sources. The highest level of novelty in the ESG dataset (level 2) denotes that a company was linked to a particular issue in a particular country for the first time.

4.3.2 Company Data

The sources of data about company characteristics in the estimation panels are the RepRisk ESG database, and the Crunchbase private company dataset. Combined, the two sources provide information about the size of a company (in number of employees), the main product category in which a company operates, whether a company went through any funding rounds, and the company's rank according to Crunchbase prominence algorithms. The rank score for a company is determined by Crunchbase, and is based on the amount of community engagement, funding rounds, media coverage of the company, and mergers and acquisitions. A small value for rank means a higher prominence in the Crunchbase records. A high company rank often translates into higher visibility among journalists and investors using the platform (Crunchbase Rank, 2019). In the estimation panels, the company rank was centered and standardized to have mean equal to 0 and standard deviation equal to 1. Finally, the funding rounds variable was transformed into a binary indicator, taking value 1 when the company received any funds from investors, and 0 otherwise. Figures C.5 and C.6 show the distribution of company characteristics in the estimation panels.

4.3.3 Community Data

The source of data about brand community engagement and brand networks is the Pushshift Reddit Archive (Baumgartner et al., 2020). These archival data include information about posts and comments created by Reddit members in the context of community discussions happening in the subreddits. Each posts or comment in the community data has a unique identifier, a timestamp with the time of creation, the name of the subreddit in which the contribution was created, and the Reddit username of the author. Therefore, the Pushshift dataset is organized at the subreddit-member-contribution level. Using the subreddit information, I linked each of the 299 Reddit communities in the Pushshift archive to one of the companies in the company dataset. Table C.5 shows that, in the estimation sample, most brand communities on Reddit are centered around companies operating in the “software” sector, followed by companies active in the “retail”, “media”, and “technology” sectors.

The structure of the Pushshift dataset allows us to develop a measure of weekly activity generated by community members around each crisis date. Furthermore, the archival data allows us to measure weekly social network characteristics, based on the networks of consumer-to-consumer interactions in community discussions.

Community Activity First, using the contribution id’s and their creation timestamps, I construct an individual-level measure of community activity. For each community and crisis event combination, community member, and for each week in the data, I measure individual-level weekly activity as the sum of the posts and comments created by that member in a given week. I use this individual activity measure to distinguish consumer types. In particular, for each community and crisis event, I classify community members as “high-type” or “H-type” if their weekly activity in the pre-crisis period exceeded the average weekly activity from all the members in the pre-crisis period. Therefore, H-type members achieve above-average pre-crisis contribution levels, while L-type members remain below-average. Based on this criterion, 21% of the Reddit members in the estimation sample are classified as H-type.

In addition to the individual-level metrics, I also calculate the weekly activity for each community around a brand crisis event. I calculate the weekly community activity by summing all the posts and comments created in a given week by all the community members. Based on this measure, Reddit members generate 367.3 new posts and comments every week. Furthermore, Figure C.7 provides evidence that, controlling for community and week fixed effects, the average weekly activity in brand communities responds to a brand crisis – and that community activity is higher in the weeks that follow it.

Social Networks of Information Spread To measure the changes to the structure of the brand networks following a brand crisis, I generate weekly undirected networks of information spread. The networks are generated for each community and brand crisis combination, and for each week in the data. The network formation rule for these information networks requires that a link is created between two community members if one of them created a comment in reply to the other's post or comment. Therefore, any two members are connected if they directly interacted with each other under a common thread.[2] I use the resulting network of members discussing over the same threads to calculate a series of relevant network metrics correlated with ease and speed of brand-related information spread, and overall network resilience to external events. In particular, for each node (i.e., each member) in a network, I calculate its degree centrality, clustering coefficient, and closeness centrality.

The *degree centrality* captures the connectedness of the average member in the group. Connectedness is an informative measure as it is frequently used as a correlate of ease and speed of information spread (Jalili & Perc, 2017) associated with the

[2] An alternative link formation rule could enforce directed networks of information spread, since I have information on the contribution ID's and timestamps in which content was generated in the communities. With a directed graph, I could compute users' in-degree and out-degree, to better characterize the patterns of information creation and spread across consumer types. I will explore this alternative link formation rule in a future version of this study.

diffusion of viral brand-related content (Bampo, Ewing, Mather, Stewart, & Wallace, 2008). A high value degree centrality suggests greater interaction between a given member and the rest of the community. A higher degree can be desirable from the point of the firm, when a brand community is instrumental to the diffusion of product and brand information. In this study, the degree centrality of a community member indicates the number of different peers who the member interacts with in the same community discussion. I calculate the average weekly degree centrality of the community as the sum of node-level weekly degree centralities, over the number of members active in that particular week in the brand network. I use the weekly community degree as an aggregate measure of network connectedness and exposure to diverse information.

The *clustering coefficient* is a measure of the density of the ego-networks of brand community members in their community discussions. Highly clustered neighborhoods in discussion threads can potentially suppress the information spread across a brand network (Easley, Kleinberg, et al., 2010). In the context of this study, highly clustered communities may impede the diffusion of relevant post-crisis brand information. Similarly to the degree centrality, I calculate the average weekly clustering coefficient as the sum of weekly local clustering coefficients divided by the number of members in the weekly brand network.

Finally, the *closeness centrality* of a node is a measure of centrality in a network, calculated as the reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the graph. The measure of closeness centrality takes *lower* value for the Reddit members that are closer to the centre of the discussions networks. These members are directly connected to many other peers in the brand network, and therefore are able to spread information more efficiently than others. Central members might have better access to brand information, be more influential in spreading pre- and post-crisis information in their communities, and don't need to rely on many other people to access information or knowledge (Banerjee, Chandrasekhar, Duflo, & Jackson, 2018; Jalili & Perc, 2017). I measure the average weekly closeness centrality as the sum of individual closeness centrality, over the number of members in the weekly brand network.

Using the individual-level network metrics, I also classify consumer types based on their centrality and embeddedness in the brand networks. In particular, I classify members as H-type degree, H-type clustering, or H-type closeness if, respectively, their average individual degree, clustering, or closeness is above-average during the pre-crisis period. Classifying members according to this criterion, I observe that 30% of the members in the estimation sample are H-type degree, 37% are H-type clustering, and 31% are H-type closeness.

Table C.7 presents the summary statistics for all the variables used throughout the empirical analyses.

4.4 Empirical Framework

The objective for this study is to measure the impact of brand crises on consumer engagement and social network structures in online brand communities. To do so, I model the consumer contributions to online brand communities (and the resulting social network characteristics) with a Difference-in-Difference framework.

4.4.1 Modeling Members' Contributions to Online Brand Communities

Community-Level Utility I consider a set of brand communities B indexed by b. In each discrete period t, given a crisis date T_s, members of the community b choose to contribute and obtain utility given by:

$$u_{bt} = \mathbb{I}(t > T_s)_{bt}\delta + X_{bt}\beta_1 + X_b\beta_2 + \varepsilon_{bt} \tag{4.1}$$

where $\mathbb{I}(t > T_s)_{bt}$ is a dummy indicating the incidence of a crisis; X_{bt} is a matrix of information-related features (e.g. number of issues triggered, type of issue, number of countries involved, etc); and X_b is a matrix of brand-specific features (e.g. Crunchbase funding, product sector). ε_{bt} captures unobservable time-varying, brand community-level factors influencing the utility from contributing to the brand community, and is Type-1 extreme value and i.i.d. distributed.

Consumer-Level Contribution Utility I consider a set of brand communities B indexed by b, and a set of consumer groups N indexed by i, which vary depending on their status or type within community b (i.e., high- vs low-type group). In each discrete period t, consumer group i chooses to contribute to brand community b and obtains utility given by:

$$u_{ibt} = \mathbb{I}(i = \mathrm{H})_{ibt}\beta_1 + \mathbb{I}(t > T_s)_{bt}\beta_2 + \mathbb{I}(t > T_s)_{bt} \times \mathbb{I}(i = \mathrm{H})_{ibt}\delta + X_{bt}\beta_3 + \beta_4 X_b + \varepsilon_{ibt}$$

where $\mathbb{I}(t > T_s)_{bt}$, X_{bt}, and X_b are the same quantities introduced in the community-level utility; $\mathbb{I}(i = \mathrm{H})_{ibt}$ is an indicator for type of consumer group, based on the pre-crisis levels of brand attitudes, experience, commitment, and/or network embeddedness of the consumers in the group (the indicator function returns 1 for high-type consumers, and 0 otherwise). ε_{ibt} captures unobservable time-varying, individual-brand level factors influencing the utility from contributing to the brand community, and is Type-1 extreme value and i.i.d. distributed.

4.4.2 Identification

Models with Fixed Effects Based on the utility specification above, I aim to estimate the following equation for the effect of brand crises on community activity and social network metrics:

$$Y_{bt} = \mathbb{I}(t > T_s)_{bt}\delta + X_{bt}\beta_1 + X_b\beta_2 + \gamma_{b,m(t))} + \zeta_t + \varepsilon_{bt} \tag{4.2}$$

and the following equation for the differential effect of brand crises on community activity and social network metrics depending on the type of consumer:

$$\begin{aligned} Y_{ibt} &= \mathbb{I}(i = \mathrm{H})_{ibt}\beta_1 + \mathbb{I}(t > T_s)_{bt}\beta_2 + \\ &+ \mathbb{I}(t > T_s)_{bt} \times \mathbb{I}(i = \mathrm{H})_{ibt}\delta + \\ &+ X_{bt}\beta_3 + X_b\beta_4 + \gamma_{b,m(t))} + \zeta_t + \varepsilon_{ibt} \end{aligned} \tag{4.3}$$

In Eq. 4.2 and Eq. 4.3, Y_{bt} and Y_{ibt} are the outcome variables, alternatively measuring community contributions or social network metrics. $\mathbb{I}(t > T_s)_{bt}$ is the brand crisis indicator; $\mathbb{I}(i = \mathrm{H})_{ibt}$ is the consumer type indicator; X_{bt} is the matrix of information-related features (e.g. number of societal issues involved, type of issue (performance vs value based), number of countries involved, etc); and X_b is the matrix of brand-specific features (e.g. Crunchbase funding, product sector). $\gamma_{b,m(t)}$ is a company-month fixed effect, and ζ_t is a week of the month fixed effect. The fixed effects account for brand-specific, unobserved time-varying factors. These factors may include time-dependent attitudes towards the brands, or intensity of marketing communication during a month. ε_{ibt} captures unobservable time-varying, individual-brand level factors that contribute to the utility from participating in the community, and is Type-1 extreme value and iid distributed.

In Eq. 4.2, δ is the coefficient of interest, capturing the baseline main effect of brand crisis information. I estimate δ as the effect of the crisis information on the contribution behavior of consumers in brand communities using an OLS specification.

In Eq. 4.3, β_1 is the baseline effect of the impact of consumer type on the utility from contributing to the brand community. δ provides an estimate of the moderating effect of the consumers' pre-crisis type (high- versus low-type) on the relationship between brand crises and consumer behavior in brand networks. The net effect of consumer type is based on the combination of baseline consumer type (β_1) and the post-crisis change across consumer types (δ) on the behavior of consumers in brand networks, $\beta_1 + \delta\mathbb{I}(t > T_s)_{bt}$. If high-type consumers contribute more than low-type consumers in online brand communities after a brand crisis, then I will observe that $\beta_1 + \delta\mathbb{I}(t > T_s)_{bt} > 0$.

The effect parameter δ is identified thanks to two sources of variation in the data. First, I can identify parameter δ by observing within-community variation in treatment assignment over time. In fact, I observe the same communities being exposed (or not exposed) to different brand crisis information over the observation period. Second, the parameter is identified thanks to between-communities variation in exposure to crisis information, since I observe different communities being exposed (or not exposed) to different crisis information over the same observation period. Assuming common trends before and after the exposure to crisis information, the differences framework produces an unbiased estimate of the effect of brand crisis information on community activity. Finally, to recover an unbiased parameter estimate for δ, I need to make an assumption motivated by the SUTVA (Imbens & Rubin, 2015). Throughout the estimation, I need to assume that no individual Reddit member can, by themselves, influence whether information about a particular brand crisis is reported by media outlets.

4.5 Results

4.5.1 Community-Level effects of Brand Crises

In this section, I discuss the results from specification 4.2 for the adjusted weekly contributions to brand communities (Table 4.1). I also report the results for the relevant network resilience metrics (Table 4.2). All tables report the coefficients with robust standard errors clustered at the product category-week level.

Brand crises and community activity

The primary objective of this study is to understand how online brand communities are impacted by brand crises. In this section, I address this question by measuring the impact of brand crises at the community level. Table 4.1 shows that, following a brand crisis, the volume of weekly consumer contributions to brand communities increases ($\delta_{Contrib} = 0.091$, p-value < 0.001). An increase in community activity alone is not surprising, but the magnitude of the activity is informative: following a brand crisis, consumers generate 9.1% more contributions every week (column 3, Table 4.1). The significant and large effects indicate that, indeed, consumers hear about and respond to brand crises, and the data capture more than noise.

While it is clear that there is higher activity in brand communities, the nature of the increase is unclear. On the one hand, the increased activity may be generated by brand-loyal consumers becoming more active – the “high-type” consumers, presumably defending the brand. On the other hand, the boost in activity may be due to brand-strangers, people flooding the brand discussion boards for information,

Table 4.1: Brand Crises and Weekly User Contributions to Brand Communities

	Weekly Contributions (Log(1+x) Scale)		
	(1)	(2)	(3)
Brand Crisis	4.611***	1.323***	0.091***
	(0.023)	(0.039)	(0.009)
Crisis Controls	No	No	Yes
Company Controls	No	No	Yes
Product Category	No	No	Yes
Company-Month FE	No	Yes	Yes
Week of Month FE	No	Yes	Yes
R^2	0.418	0.694	0.852
Num. obs.	430093	430093	430093

Note: Robust standard errors clustered at the level of the product category and week in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Specification tested: $Y_{bt} = \mathbb{I}(t > T_s)_{bt}\delta + X_{bt}\beta_1 + X_b\beta_2 + \gamma_{b,m(t)} + \zeta_t + \varepsilon_{bt}$.
Treatment variable $\mathbb{I}(t > T_s)_{bt}$: brand crisis occurrence indicator. *Outcome variable* Y_{bt}: Weekly contributions to subreddits created by community members (Log(1+x) scale). *Crisis control variables* X_{bt}: crisis severity, news novelty, media reach, number of countries affected, number of issues raised by the crisis, type of crisis issue. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset. *Product category*: main product category in which the company operates. *Fixed effects* $\gamma_{b,m(t)}$, ζ_t: company-month and week of month fixed effects.

rumors, or personal takes on the crisis event. It is also not clear if the internal structure of the social networks in the brand communities is changed in any way by the crisis event. I will answer these questions in the next two sections.

Brand crises and effects on community network

While the literature in marketing informs us about the overall effects of brand crises on consumer behavior (e.g., Ahluwalia et al., 2000; Backhaus & Fischer, 2016; Hsu & Lawrence, 2016; Klein & Dawar, 2004; Luo, 2009), their effects on consumer-to-consumer interactions in online brand communities have been little documented. Therefore, another objective of this study is to clarify whether brand crises have the disruptive potential to shake the structural stability of brand networks, and their information-spread potential. I address this objective in this section, in which I discuss how a brand community's network structure changes following a crisis. In particular, I investigate any changes in network metrics relating to the ease and speed of information spread.

Table 4.2 reports the results of the social network analysis. Similar to the analysis of weekly consumer contributions, all summary tables report coefficients with robust standard errors clustered at the product category-week level. Columns 1-3 show that the average network degree across brand communities slightly increases after a brand crisis. The positive coefficient estimate corresponds to an increase in average network degree of about 1%, compared to the pre-crisis period. The sign of the effect is robust to the inclusion of fixed effects (column (2)) and various controls (column (3)). Overall, the increased average network degree suggests that consumers discuss with more peers following a community crisis, and that information can spread more easily across members in the affected communities. Whether any post-crisis information spread is desirable is questionable from the firm's perspective.

In columns (4)-(6), I find that, following a crisis, the average clustering coefficient across brand networks also increases by 0.2%. The magnitude of the effect of brand crises on the average clustering coefficient in brand networks is affected by the type and quantity of controls and fixed effects included in the analysis; however, the sign of the effect remains positive throughout (columns (4) and (5) in Table 4.2). An increase in clustering coefficient suggests that, while consumers discuss brand information with more distinct peers, they tend do keep these discussions in (slightly) closer-knit subgroups. More practically, in the post-crisis brand networks, it is slightly more likely that two consumers who discussed with a third peer will also discuss with each other – rather than engaging with another, completely distinct consumer.

Finally, I also find that the average closeness centrality in brand networks does not significantly change compared to the pre-crisis period (columns 7-9). The direction and significance of the effect is sensitive to the inclusion of fixed effects and control variables (column (7) compared to columns (8) and (9)). A change in closeness centrality would indicate a change in the speed and breadth of information spread, due to a flatter hierarchy in the community. As a brand crisis fades, a flattened structure could turn to the advantage of the brand, if they can manage the process of controlling the content communicated adequately. However, I do not find evidence of such process in this analysis.

4.5.2 Effects of Crises on Community Members

So far, I found that brand crises can generate a boost in consumer activity online, and can significantly impact the network of consumer-to-consumer relationships as people rely on brand communities to discuss the crisis event. The next question was

Table 4.2: Brand Crises and Weekly Average Network Metrics

	Weekly Average Network Degree			Weekly Average Clustering Coefficient			Weekly Average Closeness Centrality		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Brand Crisis	2.909^{***}	0.820^{***}	0.033^{***}	0.352^{***}	0.094^{***}	0.004^{***}	0.022^{***}	$0.001^{\dagger}$	-0.000
	(0.029)	(0.021)	(0.007)	(0.002)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)
Crisis Controls	No	No	Yes	No	No	Yes	No	No	Yes
Company Controls	No	No	Yes	No	No	Yes	No	No	Yes
Product Category	No	No	Yes	No	No	Yes	No	No	Yes
Company-Month FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Week of Month FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
R^2	0.273	0.460	0.590	0.364	0.619	0.740	0.021	0.054	0.070
Num. obs.	430093	430093	430093	430093	430093	430093	430093	430093	430093

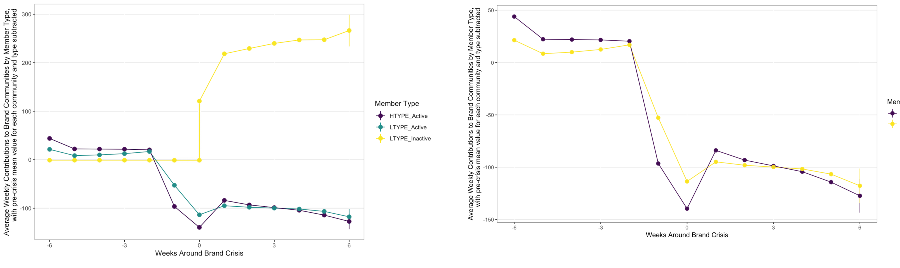
Note: Robust standard errors clustered at the level of the product category and week in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Specification tested: $Y_{bt} = \mathbb{I}(t > T_s)_{bt}\delta + X_{bt}\beta_1 + X_b\beta_2 + \gamma_{b,m(t))} + \zeta_t + \varepsilon_{bt}$.
Treatment variable $\mathbb{I}(t > T_s)_{bt}$: Brand crisis occurrence indicator. *Outcome variables* Y_{bt}: Columns (1-3): Weekly average degree centrality. Columns (4-6): Weekly average clustering coefficient. Columns (7-9): Weekly average closeness centrality. *Crisis control variables* X_{bt}: crisis severity, news novelty, media reach, number of countries affected, number of issues raised by the crisis, type of crisis issue. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset. *Product category*: main product category in which the company operates. *Fixed effects* $\gamma_{b,m(t)}, \zeta_t$: company-month and week of month fixed effects.

whether the boost in activity can be attributed to any particular member type – high-type (brand-loyal or highly connected consumers), low-type, or brand-strangers (inactive members, people who join brand discussions only after a crisis event). In this section, I dive into the analysis of crisis response by consumer type. In this section, I label community members as "High" versus "Low" types; first, based on the intensity of their activity in the community, and then based on their position and status in the brand social networks.

Tenure & intensity of activity in community. The first classification of consumer types is with respect to the tenure and intensity of activity in the community. Specifically, I compare the post-crisis response of the members who were (i) active before the crisis, but whose activity levels were below-average (*L-types*), and (ii) active before the crisis, whose activity levels were above-average (*H-types*). A third consumer type includes the community members who were *not active* in a community until the occurrence of the brand crisis (*L-type, Inactive*). Since the latter lack pre-crisis activity data, I do not compare their pre- and post-crisis activity levels in the brand communities. Figure 4.1 summarizes the change in the average activity following a brand crisis that can be attributed to the three types of consumers. Panel 4.1a clearly shows that the *Inactive* consumers flood the brand community following a crisis and become the most active users. Put differently, the brand community is under a siege of individuals who are presumably neither loyal nor committed to the

Figure 4.1: Weekly Community Contributions by Member Type

(a) (b)



brand, since they were not interested in the community in the months preceding the crisis. When I look at the consumers who were previously active in the community, surprisingly, I find that both H-types and L-types are "quieter", both compared to their previous levels of activity and compared to the Inactives (Panel 4.1b). While the lay intuition may suggest that brand crises should be discussed heavily by the loyal and committed consumers of the brand, and a wrong-doing by the brand can be defended by its loyal consumers, this does not seem to be the case in practice. This is the key empirical finding I deliver in this paper: the previously most active, committed, and loyal members of brand communities become the silent-most consumers post a crisis.

I measure the differential effect of brand crises across consumer types more formally in Table 4.3, estimating the specification in Eq. 4.3. The positive and significant interaction coefficients in columns (3-4) suggest that loyalty and tenure can potentially preserve the community under stress. While the main effect of a crisis continues to be negative, the H-type consumers keep sustainable levels of engagement, and their contributions increase at faster rates compared to the L-types. On the other hand, I also notice that the sign of the interaction effect is susceptible to the inclusion of product category controls, and to company- and crisis-level controls. The sudden flip in sign and the significance of the effects across specifications may be suggestive of heterogeneity across brands – and more specifically, across product categories – and across crisis types.

Table 4.3 demonstrated that above-average contributors (H-type consumers, based

Table 4.3: Brand Crises and Weekly Contributions by Member Type

	Weekly Contributions from H-Type vs. L-type Contributors (Excluding Inactives, Log(1+x) Scale)			
Specification:	(1)	(2)	(3)	(4)
Brand Crisis × H-Type	-3.381^{***}	-1.316^{***}	0.390^{***}	0.397^{***}
	(0.022)	(0.043)	(0.005)	(0.005)
Brand Crisis	2.782^{***}	0.721^{***}	-0.988^{***}	-0.996^{***}
	(0.021)	(0.045)	(0.009)	(0.009)
H-Type	4.150^{***}	2.085^{***}	0.379^{***}	0.372^{***}
	(0.024)	(0.044)	(0.005)	(0.005)
Crisis Controls	No	No	No	Yes
Company Controls	No	No	No	Yes
Product Category	No	No	Yes	Yes
Company-Month FE	No	Yes	Yes	Yes
Week of Month FE	No	Yes	Yes	Yes
R^2	0.547	0.655	0.789	0.800
Num. obs.	860186	860186	860186	860186

Note: Robust standard errors clustered at the level of the product category and week in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Specification tested: $Y_{ibt} = \mathbb{I}(i = \text{H})_{ibt}\beta_1 + \mathbb{I}(t > T_s)_{bt}\beta_2 + \mathbb{I}(t > T_s)_{bt} \times \mathbb{I}(i = \text{H})_{ibt}\delta + X_{bt}\beta_3 + X_b\beta_4 + \gamma_{b,m(t)}) + \zeta_t + \varepsilon_{ibt}$.
Treatment variable $\mathbb{I}(t > T_s)_{bt}$: Brand crisis occurrence indicator. *Outcome variable* Y_{ibt}: Weekly contributions to subreddits created by H vs. L-type authors (Log(1+x) scale). *Moderator* $\mathbb{I}(i = H)_{ibt}$: Type of member indicator: {1=H-type; 0=L-type}, based on above vs. below-average pre-crisis activity level.
Crisis control variables X_{bt}: crisis severity, news novelty, media reach, number of countries affected, number of issues raised by the crisis, type of crisis issue. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset. *Product category*: main product category in which the company operates. *Fixed effects* $\gamma_{b,m(t)}, \zeta_t$: company-month and week of month fixed effects.

on pre-crisis activity) tend to contribute more to their brand communities than L-type consumers, after a crisis event. Now, I show that brand crises also affect the social network characteristic of H- vs L-type consumers differentially. Earlier, I noted that brand crises cause an increase in the average weekly degree in the centrality and clustering of the brand networks – that is, brand information spreads to more and more diverse consumers, and consumers tend to form slightly more discussion sub-groups than prior to the crisis. Table 4.4 shows that this baseline increase is not observed among H-Type consumers. Rather, the social networks of H-Type consumers are potentially disrupted by brand crises. Columns (3) and (6) in Table 4.4 show significantly negative interaction coefficients for average weekly degree and clustering coefficient among H-Type members. Similarly, the positive coefficient for inverse closeness centrality in column (9) suggests that H-Type consumers lose their central positions in the brand networks relatively more than L-type consumers. Therefore, albeit trying to maintain sustainable activity levels, H-Type consumers are less central, less embedded in the network, and less influential than L-type consumers in the information networks after a brand crisis. Contrary to the interaction effects on activity, the signs of the differential effects of consumer types on social network metrics are robust to the inclusion of company- and crisis-level controls.

Table 4.4: Brand Crises and Weekly Average Network Metrics by Member Type

	Weekly Average Network Degree			Weekly Average Clustering Coefficient			Weekly Average Closeness Centrality		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Brand Crisis × H-Type	-2.373^{***}	-1.257^{***}	-0.207^{***}	-0.367^{***}	-0.176^{***}	-0.036^{***}	-0.011^{***}	0.003^{***}	0.006^{***}
	(0.027)	(0.042)	(0.013)	(0.002)	(0.004)	(0.001)	(0.000)	(0.001)	(0.001)
Brand Crisis	2.596^{***}	1.482^{***}	0.432^{***}	0.363^{***}	0.173^{***}	0.032^{***}	0.004^{***}	-0.010^{***}	-0.013^{***}
	(0.024)	(0.041)	(0.012)	(0.002)	(0.004)	(0.001)	(0.000)	(0.000)	(0.001)
H-Type	3.462^{***}	2.346^{***}	1.296^{***}	0.357^{***}	0.167^{***}	0.026^{***}	0.014^{***}	0.000	-0.003^{***}
	(0.038)	(0.051)	(0.021)	(0.002)	(0.004)	(0.001)	(0.001)	(0.000)	(0.001)
Crisis Controls	No	No	Yes	No	No	Yes	No	No	Yes
Company Controls	No	No	Yes	No	No	Yes	No	No	Yes
Product Category	No	No	Yes	No	No	Yes	No	No	Yes
Company-Month FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Week of Month FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
R^2	0.452	0.485	0.566	0.516	0.596	0.688	0.014	0.033	0.043
Num. obs.	860186	860186	860186	860186	860186	860186	860186	860186	860186

Note: Robust standard errors clustered at the level of the product category and week in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Specification tested: $Y_{ibt} = \mathbb{I}(i = \mathrm{H})_{ibt}\beta_1 + \mathbb{I}(t > T_s)_{bt}\beta_2 + \mathbb{I}(t > T_s)_{bt} \times \mathbb{I}(i = \mathrm{H})_{ibt}\delta + X_{bt}\beta_3 + X_b\beta_4 + \gamma_{b,m(t)}) + \zeta_t + \varepsilon_{ibt}$.
Treatment variable $\mathbb{I}(t > T_s)_{bt}$: Brand crisis occurrence indicator. *Outcome variables* Y_{bt}: Columns (1-3): Weekly average degree centrality. Columns (4-6): Weekly average clustering coefficient. Columns (7-9): Weekly average closeness centrality.
Moderator $\mathbb{I}(i = H)_{ibt}$: Type of member indicator: {1=H-type; 0=L-type}, based on above vs. below-average pre-crisis activity level.
Crisis control variables X_{bt}: crisis severity, news novelty, media reach, number of countries affected, number of issues raised by the crisis, type of crisis issue. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset.
Product category: main product category in which the company operates. *Fixed effects* $\gamma_{b,m(t)}, \zeta_t$: company-month and week of month fixed effects.

Status within Community. After looking at consumer types from the perspective of activity levels, I classify community members with respect to their position (or status) within the brand networks, using centrality and clustering metrics. In this section, I distinguish between community members who were (i) active in the community before the crisis, but whose social network metrics were below-average (*L-types*), and (ii) active in the community before the crisis, whose social network metrics were above-average (*H-types*).

Columns (1-3) in Table 4.5 demonstrate a similar pattern of results as those shown for high- vs low-activity consumers. In particular, the baseline negative impact of a brand crisis on community activity is mitigated among consumers who are highly central or influential in their brand networks. Once again, H-type consumers tend to maintain higher levels of engagement within their brand communities following a brand crisis, compared to L-type consumers.

Table 4.5: Brand Crises and Weekly Contributions by Member Type According to Social Network Status

	Weekly Contributions from H vs. L-Network Position (Excluding Members Only Activated by the Events, Log(1+x) Scale)		
Member Type:	H vs. L-Degree (1)	H vs. L-Clustering (2)	H vs. L-Closeness (3)
Brand Crisis × H-Type	0.354^{***}	0.135^{***}	0.174^{***}
	(0.004)	(0.004)	(0.018)
Brand Crisis	-0.872^{***}	-0.737^{***}	-0.610^{***}
	(0.009)	(0.009)	(0.015)
H-Type	-0.013^{*}	-0.351^{***}	-0.321^{***}
	(0.005)	(0.004)	(0.021)
Crisis Controls	Yes	Yes	Yes
Company Controls	Yes	Yes	Yes
Product Category	Yes	Yes	Yes
Company-Month FE	Yes	Yes	Yes
Week of Month FE	Yes	Yes	Yes
R^2	0.795	0.797	0.767
Num. obs.	860186	860186	860186

Note: Robust standard errors clustered at the level of the product category and week in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Specification tested: $Y_{ibt} = \mathbb{I}(i = \mathrm{H})_{ibt}\beta_1 + \mathbb{I}(t > T_s)_{bt}\beta_2 + \mathbb{I}(t > T_s)_{bt} \times \mathbb{I}(i = \mathrm{H})_{ibt}\delta + X_{bt}\beta_3 + X_b\beta_4 + \gamma_{b,m(t))} + \zeta_t + \varepsilon_{ibt}$.

Treatment variable $\mathbb{I}(t > T_s)_{bt}$: Brand crisis occurrence indicator. *Outcome variable* Y_{ibt}: Weekly contributions to subreddits created by H vs. L-type authors (Log(1+x) scale). *Moderator* $\mathbb{I}(i = H)_{ibt}$: Type of member indicator: {1=H-type; 0=L-type}, based on (1) above vs. below-average network degree pre-crisis; (2) above vs. below-average clustering coefficient pre-crisis; (3) above vs. below-average closeness centrality pre-crisis.

Crisis control variables X_{bt}: crisis severity, news novelty, media reach, number of countries affected, number of issues raised by the crisis, type of crisis issue. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset. *Product category*: main product category in which the company operates. *Fixed effects* $\gamma_{b,m(t)}, \zeta_t$: company-month and week of month fixed effects.

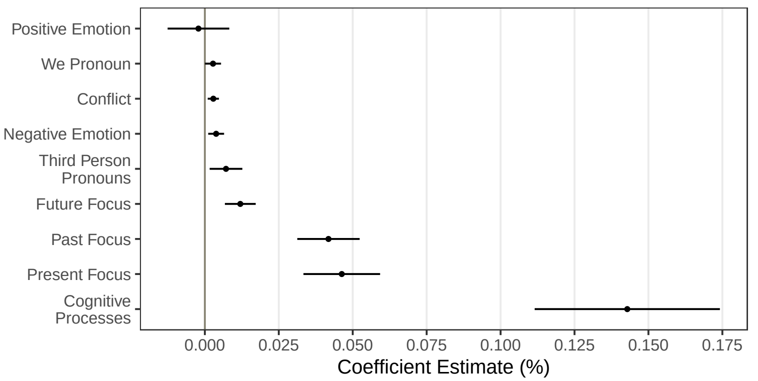
4.5.3 Changes in Community Discussion Content

Thus far, I focused on the changes in activity and network structure of the brand communities affected by brand crises. An equally significant factor to consider in understanding how crises affect brand communities is the change in the content of the consumer-to-consumer discussions. On the one hand, the changes in content may be steered by the faults of the brand, and take a negative tone. On the other hand, if the previously brand-loyal community members take to defend the brand, the conversations may also be more positive. Furthermore, as consumers may perceive brand crises as a dissonant cognition and a threat to their identity, community content may take in-group versus out-group connotations. Finally, in processing the occurrence and the consequences of a crisis, consumers may discuss in brand communities either using factual, informative statements, or resorting to speculation and wishful thinking. A more nuanced analysis is which type of consumers share certain classes of content.

In this section, I investigate the linguistic characteristics of the brand community discussions following a brand crisis. In this content analysis, I adopt a lexicon-based approach. In particular, I process the text shared in each Reddit post with the Linguistic Inquiry and Word Count (LIWC) software, based on the 2022 LIWC dictionary (Pennebaker, Francis, & Booth, 2001). For each post, the LIWC software calculates the weekly percentage of words reflecting several emotions, thinking styles, and parts of speech. In addition to the weekly word count, I focus on four main linguistic dimensions of the community contributions: positive and negative emotions (including words related to positive sentiment, negative sentiment, and conflict); in-group vs. out-group expressions (including the use of first vs third-person pronouns); cognitive processes; and time orientation (including words related to past, present, and future tense).

Figure 4.2 reports the coefficients and 95% confidence intervals for each linguistic category, estimated with a multivariate OLS model specified as in Eq. 4.2. The results of the content analysis suggest that, post-crisis, consumers in brand communities share more words related to negative emotions and conflict, while the share of positive words does not significantly vary. Furthermore, consumers center their post-crisis discussions around past and present events, and less about future events. Finally, the post-crisis discussions contain more words related to cognitive processes (such processes include reflection, insight, and causality). Taken together, the results suggest that consumers resort to brand communities mostly to share and collect information, to generate insight, to think, and to reflect about past and unfolding events – and perhaps less to regulate their emotions. This is an encouraging result for brands involved in crisis events: the sudden influx of brand strangers measured through the activity levels might have implied an emotional threat to the stability of the group. The average results suggest that the communities may be resilient to such “take-over” – at least on an emotional level.

Figure 4.2: Brand Crises and Weekly % User-Generated Content. Multivariate OLS coefficients including crisis, company, and product category controls, and company-month and week-of-month fixed effects.



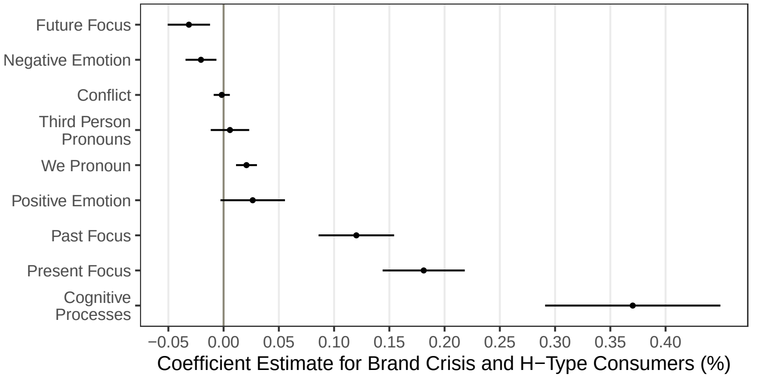
Note: the coefficients are estimated with a multivariate OLS model. Error bars represent 95% confidence intervals with robust standard errors clustered at the level of the product category and week. Specification tested: $Y_{bt} = \mathbb{I}(t > T_s)_{bt}\delta + X_{bt}\beta_1 + X_b\beta_2 + \gamma_{b,m(t))} + \zeta_t + \varepsilon_{bt}$. *Treatment variable* $\mathbb{I}(t > T_s)_{bt}$: Brand crisis occurrence indicator. *Outcome variables* Y_{bt}: Average weekly share of words (%) per contribution. *Crisis control variables* X_{bt}: crisis severity, news novelty, media reach, number of countries affected, number of issues raised by the crisis, type of crisis issue. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset. *Product category*: main product macro-category in which the company operates. *Fixed effects* $\gamma_{b,m(t)}$, ζ_t: company-month and week of month fixed effects.

User-Generated Content and Member Types

After assessing the average changes in user-generated content in brand communities after a brand crisis, I turn to analyze any differences in content created by high- versus low-type consumers.

Figure 4.3 reports the coefficients and 95% confidence intervals for high-type consumers. The coefficients are estimated with a multivariate OLS model, including an interaction term for consumer type, as specified in Eq. 4.3. Figure 4.3 suggests that H-type consumers share fewer words related to negative emotions compared to L-types, and marginally more positive emotion words. At the same time, H-types do not engage in more or less conflict than the L-Types. H-type consumers also share more words related to in-group expressions (first-person plural pronouns) than out-group expressions (third-person pronouns), compared to the L-Types. H-Type consumers also reflect more on past and present events, and share more words related to cognitive processes: the coefficients for cognitive processes and past and present time orientation are consistently positive and significantly different from zero.

Figure 4.3: Brand Crises and Weekly % User-Generated Content by Member Type. Multivariate OLS coefficients including crisis, company, and product category controls, and company-month and week-of-month fixed effects.



Note: the coefficients are estimated with a multivariate OLS model. Error bars represent 95% confidence intervals with robust standard errors clustered at the level of the product category and week. Specification tested: $Y_{ibt} = \mathbb{I}(i = \mathrm{H})_{ibt}\beta_1 + \mathbb{I}(t > T_s)_{bt}\beta_2 + \mathbb{I}(t > T_s)_{bt} \times \mathbb{I}(i = \mathrm{H})_{ibt}\delta + X_{bt}\beta_3 + X_b\beta_4 + \gamma_{b,m(t)}) + \zeta_t + \varepsilon_{ibt}$. *Treatment variable* $\mathbb{I}(t > T_s)_{bt}$: Brand crisis occurrence indicator. *Outcome variable* Y_{bt}: Average weekly share of words (%) per contribution. *Moderator* $\mathbb{I}(i = H)_{ibt}$: Type of member indicator: {1=H-type; 0=L-type}, based on above vs. below-average pre-crisis activity level. *Crisis control variables* X_{bt}: crisis severity, news novelty, media reach, number of countries affected, number of issues raised by the crisis, type of crisis issue. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset. *Product category*: main product macro-category in which the company operates. *Fixed effects* $\gamma_{b,m(t)}, \zeta_t$: company-month and week of month fixed effects.

Recall that, from the analysis of weekly community activity, the engagement of H- and L-type consumers was most negatively impacted by a brand crisis event. The content analysis results suggest that, in spite of the damage to their engagement levels, the active "brand loyal" consumers may play an important "emotion regulation" function in post-crisis brand communities. The active H-type consumers do not give in to negative emotion (and actually, share fewer negative words than the L-types); they keep the conversations slightly more positive and more group-oriented; they keep the focus on the group in the present; and they spend their efforts in the community to reflect and elaborate on what happened with other peers. From the companies' perspective, active H-type consumers may represent important allies in the post-crisis brand networks, and a great asset to manage or mitigate the crisis aftermath in online brand communities.

4.6 Heterogeneity

In the main analyses, I measured a baseline negative effect of brand crises on the engagement of consumers in brand communities, and significant post-crisis alterations in the structures of the brand networks. I also found that the effect of brand crises is differential across types of consumers. In this section, I expand the main analysis to document the heterogeneous crisis effects across important brand-related and crisis-related dimensions. Research on brand transgressions, service failures, and product-harm crises has conceptualized negative brand events according to the type of issues they trigger, their context, sector, or type of company, the amount of media coverage they received, and the severity of their consequences (Khamitov et al., 2020; Stäbler & Fischer, 2020). I am going to investigate the heterogeneous effect of brand crises on consumer engagement in brand communities along each of these dimensions.

4.6.1 Company and Crisis Types

First, I document evidence of heterogeneous effects of the average brand crisis on weekly brand community activity, based on the type of crisis experienced and the type of company engaging in corporate social irresponsibility. In terms of crisis type, I distinguish between crises that had a potentially direct versus an indirect impact on customers. As for company type, I consider whether the company operates mostly as business-to-consumer (B2C), business-to-business (B2B), or both (B2C+B2B). The classifications of crises into direct and indirect, and of companies into B2B and B2C, are detailed in Appendices C.1 and C.2. For the heterogeneity analysis, I estimate a modified version of Eq. (4.3), which includes an interaction term between the brand crisis indicator and the company-type or crisis-type indicators.

Table 4.6 demonstrates the results from the heterogeneity analysis at the company-type level. Particularly, column (3) shows negative interaction coefficients for δ_{B2C} and $\delta_{\text{B2B,B2C}}$. The negative and statistically significant coefficient for δ_{B2C} suggests that brand crises have a disruptive effect on the viability of brand communities online when they affect companies operating as B2C, compared to companies only operating as B2B. While brand crises still have a negative effect on the brand communities of companies operating as both B2C and B2B, the effect is not statistically significant.

Table 4.6: Brand Crises, Type of Company, and Weekly Contributions to Brand Communities. Including crisis and company controls, and company-month and week-of-month fixed effects.

	Weekly Contributions (Log(1+x) Scale)		
	(1)	(2)	(3)
Baseline: Only B2B			
Brand Crisis × Only B2C	-4.890^{***}	-4.423^{***}	-1.133^{***}
	(0.046)	(0.066)	(0.027)
Brand Crisis × B2C+B2B	-0.017	-0.018	-0.016
	(0.018)	(0.017)	(0.016)
Brand Crisis	4.962^{***}	4.497^{***}	1.204^{***}
	(0.047)	(0.067)	(0.028)
B2C	4.358^{***}	3.950^{***}	0.749^{***}
	(0.024)	(0.059)	(0.043)
B2C+B2B	0.767^{***}	0.835^{***}	0.725^{***}
	(0.032)	(0.031)	(0.029)
Crisis Controls	No	No	Yes
Company Controls	No	No	Yes
Company-Month FE	No	Yes	Yes
Week of Month FE	No	Yes	Yes
R^2	0.776	0.787	0.842
Num. obs.	429387	429387	429387

Note: Robust standard errors clustered at the level of the product category and week in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Specification tested: $Y_{bt} = \mathbb{I}(t > T_s)_{bt}\beta_1 + M_{bt}\beta_2 + \mathbb{I}(t > T_s)_{bt} \times M_{bt}\delta + X_{bt}\beta_3 + X_b\beta_4 + \gamma_{b,m(t))} + \zeta_t + \varepsilon_{bt}$.
Treatment variable $\mathbb{I}(t > T_s)_{bt}$: Brand crisis occurrence indicator. *Outcome variable* Y_{bt}: Weekly contributions to subreddits (Log(1+x) scale). *Moderator* M_{bt}: Company type indicators: B2C-only indicator, B2B+B2C indicator. B2B-only is used as a baseline.
Crisis control variables X_{bt}: crisis severity, news novelty, media reach, number of countries affected, number of issues raised by the crisis, type of crisis (direct-only, direct+indirect. Indirect-only used as a baseline). *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset. *Fixed effects* $\gamma_{b,m(t)}$, ζ_t: company-month and week of month fixed effects.

After measuring heterogeneity across types of companies, I also assess how the effects of brand crises vary at the crisis level. The results in Table 4.7, column (3) demonstrate that brand crises have a significant negative impact on brand community engagement when the crisis *directly* impacts the final consumers, as opposed to having an indirect impact.

Table 4.7: Types of Brand Crises and Weekly User Contributions to Brand Communities

	Weekly Contributions (Log(1+x) Scale)		
	(1)	(2)	(3)
Baseline: Only Indirect Impact			
Brand Crisis × Direct Impact	-4.541^{***}	-2.561^{***}	-0.329^{***}
	(0.037)	(0.069)	(0.028)
Brand Crisis × Direct and Indirect Impact	−0.004	−0.022	−0.008
	(0.026)	(0.033)	(0.023)
Brand Crisis	4.628^{***}	2.671^{***}	0.415^{***}
	(0.029)	(0.063)	(0.020)
Direct Impact	4.437^{***}	2.454^{***}	0.466^{***}
	(0.029)	(0.062)	(0.022)
Direct and Indirect Impact	0.120^{***}	0.148^{***}	-0.262^{***}
	(0.022)	(0.029)	(0.020)
Crisis Controls	No	No	Yes
Company Controls	No	No	Yes
Company-Month FE	No	Yes	Yes
Week of Month FE	No	Yes	Yes
R^2	0.685	0.738	0.841
Num. obs.	429387	429387	429387

Note: Robust standard errors clustered at the level of the product category and week in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Specification tested: $Y_{bt} = \mathbb{I}(t > T_s)_{bt}\beta_1 + M_{bt}\beta_2 + \mathbb{I}(t > T_s)_{bt} \times M_{bt}\delta + X_{bt}\beta_3 + X_b\beta_4 + \gamma_{b,m(t))} + \zeta_t + \varepsilon_{bt}$.
Treatment variable $\mathbb{I}(t > T_s)_{bt}$: brand crisis occurrence indicator. *Outcome variable* Y_{bt}: Weekly contributions to subreddits (Log(1+x) scale). *Moderator* M_{bt}: Indicators for type of brand crisis issue (direct impact only, both direct and indirect impact. Indirect-only is used as a baseline).
Crisis control variables X_{bt}: number of countries affected, number of issues raised by the crisis. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset, B2C indicator, B2C+B2B indicator. *Fixed effects* $\gamma_{b,m(t)}, \zeta_t$: company-month and week of month fixed effects.

The observed heterogeneity is in line with the conceptualization of brand crises as either events that *directly* impact consumers, or as situations in which the consumers *indirectly* witness the event – for example, through media or word-of-mouth (Khamitov et al., 2020). There are a few possible explanations for this empirical result. For example, as consumers are directly and practically affected by the consequences of a crisis (instead of only being affected in their values, principles, or broader environment), they may be busy with dealing with those consequences. In turn, they may disengage from the brand communities to a greater extent than consumers who are only affected indirectly by a crisis event. Alternatively, a brand crisis that has immediate consequence for the final consumers may represent an information too dissonant with both the prior image that consumers had of the brand, and with their own set of values and morals. As a result, it may be too psychologically costly for consumers to engage with the culprit brand – and much more psychologically inexpensive to distance themselves from the brand and its associated community. It is unclear whether the disengagement of consumers directly impacted by a brand crisis event is transitory or long-lasting.

4.6.2 Media Coverage

Next, I analyze the heterogeneous effects of media coverage of the brand crisis event. Literature on corporate crises has extensively demonstrated that media coverage is one of the most important factors shaping the pace, depth, and length of a brand crisis and its consequences for the relevant company (Stäbler & Fischer, 2020). Here, I classify media coverage of a crisis brand on the reach of media, and label it *high* if the news was covered by national and international media, and *low* if there was only coverage by local media.

Table 4.8 shows that the heterogeneous effect of a brand crisis that received high media coverage is negative and significant (column 3). This result implies that more national and international attention drawn on brand crises hurts consumer participation in online brand communities.

In Table 4.1, I found that brand crises, on average, increase the activity in brand communities by 9.1%. With the arrival of new brand-relevant information, consumers are discussing and sharing information regarding brands and products among themselves. I see that the interaction with media reach is negative, indicating that for brand scandals that are covered by media outlets of greater reach, the boost in community activity is lower. While I cannot explicitly test the reason for the negative sign of interaction, I can suggest two explanations that are consistent with this sign. First, the heterogeneity may indicate that traditional media and UGC in brand communities are substitutes – the information generated by media outlets serves a similar purpose in informing community members about the brand/products, and when there is more information generated by traditional media, the need to generate information by the members is lower. A second explanation has to do with the selection among media outlets of the types of crises and types of brands they prefer to cover. If, for instance, larger and more international media outlets pay attention to legal and accounting scandals – where the average consumer may have less to contribute to the discussion compared to a product quality problem – I can expect the direction of the relationship to be negative. Regardless of the explanation, the negative interaction is important, as it suggests that media coverage does not necessarily parallel the change in activity in the brand communities.

Table 4.8: Brand Crises, Media Reach, and Weekly User Contribution to Brand Communities

	Weekly Contributions (Log(1+x) Scale)		
	(1)	(2)	(3)
Brand Crisis × High Media Reach	-4.391^{***}	-3.224^{***}	-0.133^{***}
	(0.032)	(0.045)	(0.023)
Brand Crisis	4.458^{***}	3.286^{***}	0.199^{***}
	(0.030)	(0.044)	(0.020)
High Media Reach	4.580^{***}	3.391^{***}	0.175^{***}
	(0.024)	(0.040)	(0.018)
Crisis Controls	No	No	Yes
Company Controls	No	No	Yes
Product Category	No	No	Yes
Company-Month FE	No	Yes	Yes
Week of Month FE	No	Yes	Yes
R^2	0.746	0.767	0.852
Num. obs.	430093	430093	430093

Note: Robust standard errors clustered at the level of the product category and week in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Specification tested: $Y_{bt} = \mathbb{I}(t > T_s)_{bt}\beta_1 + M_{bt}\beta_2 + \mathbb{I}(t > T_s)_{bt} \times M_{bt}\delta + X_{bt}\beta_3 + X_b\beta_4 + \gamma_{b,m(t)}) + \zeta_t + \varepsilon_{bt}$.
Treatment variable $\mathbb{I}(t > T_s)_{bt}$: Brand crisis occurrence indicator. *Outcome variable* Y_{bt}: Weekly contributions to subreddits (Log(1+x) scale). *Moderator* M_{bt}: High media reach indicator $\in \{0, 1\}$. Low media reach: local media, smaller NGOs, local government bodies, blogs. High media reach: national and regional media, international NGOs, state, national, and international government bodies, international media – e.g. the NY Times, BBC.
Crisis control variables X_{bt}: number of countries affected, number of issues raised by the crisis, type of crisis issue. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset. *Product category*: main product category in which the company operates. *Fixed effects* $\gamma_{b,m(t)}, \zeta_t$: company-month and week of month fixed effects.

4.6.3 Severity of Brand Crisis

Finally, I turn to the heterogeneous effect of high-severity brand crises on consumer activity in brand communities. Crisis severity is determined by RepRisk as a function of three dimensions: first, the consequences of the risk incident (e.g., the gravity of crisis consequences on people's safety, such as injury or death); second, the extent of the crisis impact (e.g., one person, a group of people, or a large number of people); and third, the cause of the risk incident (e.g. an accident, negligence, or intent).

Table 4.9 demonstrates that both the heterogeneous effect and the main effect of high crisis severity on consumer activity in brand communities are negative and significant (column 3). The negative heterogeneous effect coefficient implies a decrease in consumer contributions in the brand communities due to high-severity crises. The stark decrease is in contrast with the 9.1% increase in community activity after an average brand crisis – suggesting that high-severity crises have a bigger, disruptive effect on brand community engagement compared to an average brand crisis. This result is in line with marketing studies demonstrating that it is harder to recover from high-magnitude failures (De Matos, Henrique, & Alberto Vargas Rossi, 2007), and the severity of a brand crisis has a strong correlation with consumer dissatisfaction and negative responses towards the brand (Tsarenko & Tojib, 2015). The result is also in line with the discussion on dissonant cognitions associated with higher media coverage, and with the direct-vs-indirect impact of the crisis consequences on consumers. More severe, direct, high-coverage crisis events may trigger higher psychological discomfort in community members, who may have had a prior positive attitude towards the brand, and may prefer to disengage from the community to recover from the discomfort (Festinger, 1957).

Table 4.9: Brand Crises, Crisis Severity, and Weekly User Contribution to Brand Communities

	Weekly Contributions (Log(1+x) Scale)		
	(1)	(2)	(3)
Brand Crisis × High Crisis Severity	-4.573^{***}	-1.562^{***}	-0.047^{*}
	(0.035)	(0.047)	(0.023)
Brand Crisis	4.636^{***}	1.619^{***}	0.105^{***}
	(0.024)	(0.027)	(0.010)
High Severity	4.471^{***}	1.472^{***}	-0.085^{***}
	(0.030)	(0.042)	(0.019)
Crisis Controls	No	No	Yes
Company Controls	No	No	Yes
Product Category	No	No	Yes
Company-Month FE	No	Yes	Yes
Week of Month FE	No	Yes	Yes
R^2	0.511	0.707	0.852
Num. obs.	430093	430093	430093

Note: Robust standard errors clustered at the level of the product category and week in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Specification tested: $Y_{bt} = \mathbb{I}(t > T_s)_{bt}\beta_1 + M_{bt}\beta_2 + \mathbb{I}(t > T_s)_{bt} \times M_{bt}\delta + X_{bt}\beta_3 + X_b\beta_4 + \gamma_{b,m(t)}) + \zeta_t + \varepsilon_{bt}$.
Treatment variable $\mathbb{I}(t > T_s)_{bt}$: Brand crisis occurrence indicator. *Outcome variable* Y_{bt}: Weekly contributions to subreddits (Log(1+x) scale). *Moderator* M_{bt}: High crisis severity indicator $\in \{0, 1\}$. Indicator equals 1 when the severity level is 2 or 3. Indicator equals 0 when the severity level is 1.
Crisis control variables X_{bt}: number of countries affected, number of issues raised by the crisis, type of crisis issue. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset. *Product category*: indicators for main product macro-category in which the company operates. *Fixed effects* $\gamma_{b,m(t)}$, ζ_t: company-month and week of month fixed effects.

4.7 Robustness Checks with Varying Time Windows

In this section, I offer an alternative estimation of the impact of brand crises – both baseline and differential across types of consumers – in a way that is qualitatively similar to the visualization in Figure C.7. Specifically, I measure community activity and network structure, before and after the occurrence of a brand crisis, for windows of 1, 2, 3, and 4 weeks before or after the date in which the crisis was reported by media outlets. More specifically, I estimate the following specification:

$$Y_{bt} = \mathbb{I}(t > T_s)_{bt}\delta + X_{bt}\beta_1 + X_b\beta_2 + \gamma_{b,m(t))} + \zeta_t + \varepsilon_{bt} \tag{4.4}$$

for $t_0 - 6 < t < t_0 + k$ or $t_0 - k < t < t_0 + 6$.

These results, presented in Table 4.10 show that the coefficients for the brand crisis effect are consistent with the main results. In particular, I notice that the baseline increase in activity does not seem constant over the weeks following the crisis event. The coefficients in columns 1-4 suggest a peak in activity at around 3 weeks post-crisis, consistent with the average adjusted activity plot in Figure C.7. Another important insight from the analysis with varying windows is the fact that the effect is strongest if I only include 1 week pre-event (column 4). This insight may indicate two potential issues: (i) there may anticipation of the crisis news among the community members – including the spread of rumors – or (ii) there may be a lag in the coverage of the crisis news by the media outlets, up to a week before the actual crisis news date, such that the community reacts earlier than the reported crisis date (as suggested also by weeks -1 to +1 in Figure C.7). Finally, I notice the gradual decrease in the magnitude of the effect, as well as the negative sign when I only include 4 weeks pre-crisis. Together with the positive baseline measured with the main specification, this may indicate that, fixing community, company, and week, the pre-crisis activity might peak at -4 weeks, and then decrease until -6 weeks.

Next, I run the same robustness analysis on the network structure outcomes. In this case, I notice that the strongest increase across metrics occurs when I only include 1 week post crisis (columns 1, 5, 9 in Table 4.11), while the magnitude of

Table 4.10: Brand Crises and Weekly User Contributions to Brand Communities: Different Windows with Controls and Fixed Effects

	Weekly Contributions, Log(1+x) Scale							
	Fix pre-, vary post-crisis window				Vary pre-, fix post-crisis window			
	+1 week (1)	+2 weeks (2)	+3 weeks (3)	+4 weeks (4)	-1 week (5)	-2 weeks (6)	-3 weeks (7)	-4 weeks (8)
Brand Crisis	0.005	0.112^{***}	0.170^{***}	0.156^{***}	0.532^{***}	0.169^{***}	0.053^{***}	-0.002
	(0.011)	(0.010)	(0.009)	(0.009)	(0.015)	(0.011)	(0.010)	(0.009)
Crisis Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Company Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Product Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Company FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week of Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.849	0.852	0.854	0.855	0.852	0.853	0.855	0.856
Num. obs.	294901	335708	376515	417322	257613	298420	339227	380034

Note: Robust standard errors clustered at the level of the product category and week in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Specification tested: $Y_{bt} = \mathbb{I}(t > T_s)_{bt}\delta + X_{bt}\beta_1 + X_b\beta_2 + \gamma_{b,m(t))} + \zeta_t + \varepsilon_{bt}$.
Treatment variable $\mathbb{I}(t > T_s)_{bt}$: brand crisis occurrence indicator. *Outcome variable* Y_{bt}: Weekly contributions to subreddits (Log (1+x) scale). Columns (1-4) include respectively 1, 2, 3, and 4 weeks of post-crisis observations, 6 weeks of pre-crisis observations. Columns (5-8) include respectively 1, 2, 3, and 4 weeks of pre-crisis observations, 6 weeks of post-crisis observations.
Crisis control variables X_{bt}: crisis severity, news novelty, media reach, number of countries affected, number of issues raised by the crisis, type of crisis issue. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset. *Product category*: main product category in which the company operates. *Fixed effects* $\gamma_{b,m(t)}, \zeta_t$: company and week of month fixed effects.

the effects slightly decreases over the following 4-6 weeks. This may be an indication of patterns of information spread being immediately disrupted by the crisis event, and then re-adjusting over the following 1-2 months post-crisis. From Table 4.12), I learn that only including 1 week pre-crisis makes the change in degree centrality non-significant (column 1) – again, potentially signaling a lag in the circulation of the crisis news, reflected in the volume of connections created in the community discussions.

Table 4.11: Brand Crises and Weekly Network Metrics: Fixed Pre-Crisis, Different Post-Crisis Windows with Controls and Fixed Effects

	Weekly Network Degree				Weekly Network Clustering Coefficient				Weekly Network Closeness Centrality			
	Fix pre-, vary post-crisis window											
	1 week (1)	2 weeks (2)	3 weeks (3)	4 weeks (4)	1 week (5)	2 weeks (6)	3 weeks (7)	4 weeks (8)	1 week (9)	2 weeks (10)	3 weeks (11)	4 weeks (12)
Brand Crisis	0.043^{***}	0.037^{***}	0.036^{***}	0.036^{***}	0.005^{***}	0.004^{***}	0.004^{***}	0.004^{***}	0.000	-0.000	-0.000	-0.000
	(0.009)	(0.008)	(0.008)	(0.007)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Crisis Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Company Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Product Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Company-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week of Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.591	0.591	0.590	0.591	0.739	0.739	0.740	0.740	0.067	0.066	0.066	0.066
Num. obs.	294901	335708	376515	417322	294901	335708	376515	417322	294901	335708	376515	417322

Note: Robust standard errors clustered at the level of the product category and week in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Specification tested: $Y_{bt} = \mathbb{I}(t > T_s)_{bt}\delta + X_{bt}\beta_1 + X_b\beta_2 + \gamma_{b,m(t))} + \zeta_t + \varepsilon_{bt}$.
Treatment variable $\mathbb{I}(t > T_s)_{bt}$: brand crisis occurrence indicator. *Outcome variable* Y_{bt}: Columns (1-4) weekly average degree; columns (5-8) weekly average clustering coefficients; columns (9-12) weekly average closeness centrality. Columns 1-4, 5-8, 6-12 include respectively 1, 2, 3, and 4 weeks of post-crisis observations, 6 weeks of pre-crisis observations.
Crisis control variables X_{bt}: crisis severity, news novelty, media reach, number of countries affected, number of issues raised by the crisis, type of crisis issue. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset. *Product category*: main product category in which the company operates. *Fixed effects* $\gamma_{b,m(t)}, \zeta_t$: company-month and week of month fixed effects.

Table 4.12: Brand Crises and Weekly Network Metrics: Fixed Post-Crisis, Different Pre-Crisis Windows with Controls and Fixed Effects

	Weekly Network Degree				Weekly Network Clustering Coefficient				Weekly Network Closeness Centrality			
	Fix post-, vary pre-crisis window											
	-1 week (1)	-2 weeks (2)	-3 weeks (3)	-4 weeks (4)	-1 week (5)	-2 weeks (6)	-3 weeks (7)	-4 weeks (8)	-1 week (9)	-2 weeks (10)	-3 weeks (11)	-4 weeks (12)
Brand Crisis	0.014	0.018^{*}	0.018^{*}	0.023^{**}	0.004^{***}	0.004^{***}	0.004^{***}	0.004^{***}	0.002^{**}	0.001	0.000	0.000
	(0.011)	(0.008)	(0.008)	(0.008)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Crisis Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Company Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Product Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Company-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week of Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.592	0.591	0.591	0.592	0.743	0.742	0.742	0.741	0.066	0.066	0.066	0.066
Num. obs.	257613	298420	298420	339227	257613	298420	339227	380034	257613	298420	339227	380034

Note: Robust standard errors clustered at the level of the product category and week in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Specification tested: $Y_{bt} = \mathbb{I}(t > T_S)_{bt}\delta + X_{bt}\beta_1 + X_b\beta_2 + \gamma_{b,m(t)}) + \zeta_t + \varepsilon_{bt}$.
Treatment variable $\mathbb{I}(t > T_S)_{bt}$: brand crisis occurrence indicator. *Outcome variable* Y_{bt}: Columns (1-4) weekly average degree; columns (5-8) weekly average clustering coefficients; columns (9-12) weekly average closeness centrality. Columns 1-4, 5-8, 6-12 include respectively 1, 2, 3, and 4 weeks of pre-crisis observations, 6 weeks of post-crisis observations.
Crisis control variables X_{bt}: crisis severity, news novelty, media reach, number of countries affected, number of issues raised by the crisis, type of crisis issue. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset. *Product category*: main product category in which the company operates. *Fixed effects* $\gamma_{b,m(t)}$, ζ_t: company-month and week of month fixed effects.

Finally, I repeat the above exercise for the heterogeneous effect of brand crises across types of consumers in the brand networks. From Table 4.13, I notice that the strongest positive differential impact of brand crises on H-type consumers occurs when I include only 1 or 2 weeks pre-crisis (column 5). This additional evidence helps to understand and measure the magnitude of the dip in activity recorded descriptively in Figure 4.1b, at weeks -1 and -2 prior to the media coverage of the crisis.

4.8 Discussion and Conclusions

In this study, I assessed how brand crises affect the engagement and social network resilience in online brand communities. I constructed a unique panel at the subreddit-crisis-week level, using nearly 10 years of information about brand community discussions on Reddit, and brand crisis information from the RepRisk ESG database and the Crunchbase company database. Using a Difference-in-Difference specification, I found that following a brand crisis, members of online brand communities create more contributions. This result suggests that brand crises affect the utility that people obtain from contributing to brand communities after a negative brand event. The increase in consumer engagement, however, is attributable to the activity of “bandwagon consumers” – people who only engage in brand communities after the occurrence of negative publicity. For community members that were already

Table 4.13: Brand Crises and Weekly Contributions by H-Type v. L-Type Members: Different Windows with Controls and Fixed Effects

	Weekly Contributions from H-Type vs. L-Type Contributors (Excluding Members Only Activated by the Events, Log(1+x) Scale)							
	Fix pre-, vary post-crisis window				Vary pre-, fix post-crisis window			
	1 week	2 weeks	3 weeks	4 weeks	-1 week	-2 weeks	-3 weeks	-4 weeks
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Brand Crisis × H-Type	0.391^{***}	0.401^{***}	0.400^{***}	0.399^{***}	0.425^{***}	0.420^{***}	0.412^{***}	0.405^{***}
	(0.005)	(0.005)	(0.005)	(0.005)	(0.008)	(0.006)	(0.006)	(0.005)
H-Type	0.358^{***}	0.362^{***}	0.367^{***}	0.370^{***}	0.344^{***}	0.349^{***}	0.357^{***}	0.364^{***}
	(0.005)	(0.005)	(0.005)	(0.005)	(0.008)	(0.006)	(0.005)	(0.005)
Brand Crisis	-1.072^{***}	-1.022^{***}	-1.001^{***}	-0.994^{***}	-0.481^{***}	-0.816^{***}	-0.919^{***}	-0.966^{***}
	(0.011)	(0.010)	(0.010)	(0.009)	(0.015)	(0.011)	(0.010)	(0.009)
Crisis Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Company Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Product Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Company-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week of Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.812	0.807	0.804	0.801	0.770	0.781	0.789	0.795
Num. obs.	589802	671416	753030	834644	515226	596840	678454	760068

Note: Robust standard errors clustered at the level of the product category and week in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Specification tested: $Y_{ibt} = \mathbb{I}(i = \mathrm{H})_{ibt}\beta_1 + \mathbb{I}(t > T_s)_{bt}\beta_2 + \mathbb{I}(t > T_s)_{bt} \times \mathbb{I}(i = \mathrm{H})_{ibt}\delta + X_{bt}\beta_3 + X_b\beta_4 + \gamma_{b,m(t)}) + \zeta_t + \varepsilon_{ibt}$.

Treatment variable $\mathbb{I}(t > T_s)_{bt}$: Brand crisis occurrence indicator. *Outcome variable* Y_{ibt}: Weekly contributions to subreddits created by H vs. L-type authors (Log(1+x) scale). Columns (1-4) include respectively 1, 2, 3, and 4 weeks of post-crisis observations, 6 weeks of pre-crisis observations. Columns (5-8) include respectively 1, 2, 3, and 4 weeks of pre-crisis observations, 6 weeks of post-crisis observations. *Moderator* $\mathbb{I}(i = H)_{ibt}$: Type of member indicator: {1=H-type; 0=L-Type}, based on above vs. below-average pre-crisis activity level.

Crisis control variables X_{bt}: crisis severity, news novelty, media reach, number of countries affected, number of issues raised by the crisis, type of crisis issue. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset. *Product category*: main product category in which the company operates. *Fixed effects* $\gamma_{b,m(t)}, \zeta_t$: company-month and week of month fixed effects.

highly active and embedded in the brand networks *before the crisis event*, a brand crisis has a negative impact on their contribution levels. The rate at which their engagement decreases, however, is not constant and depends on the experience, status, or loyalty of the consumers. Consumers with higher experience, status, or loyalty in their communities ("High-type" consumers) keep their contribution levels relatively higher than the "Low-type" consumers. The high-type consumers also share different content in the post-crisis brand networks, compared to low-type consumers. In particular, the high-types share fewer words associated with negative emotions, more words related to positive emotions and in-group expressions, and are more focused on discussing past and present events using words related to cognitive processes.

In terms of brand network resilience and patterns of information spread, I found that after brand crises, information travels through a more diverse pool of consumers, and in more tight-knit discussion sub-groups. Finally, I documented that brand crises have the most detrimental effect on brand community activity when companies operate in the finance and health sectors, when the crises create legal and ethical issues, after an international media outlet covers a crisis story, and when crises have more severe consequences in terms of harm caused, number of people involved, and intention to harm.

The study offers a number of insights for managers. First, the study documents that brands' lapse of CRS is also costly, as it harms pre-existing online brand communities. Second, I found that the structure of these communities changes significantly following brand misconduct. In terms of network structure, conversations among consumers become more close-knit and diverse. In terms of community composition, brand crises threaten the representation of brands online, as brand-strangers take over the conversation in their dedicated online social networks. Finally, the study informs managers about what to expect, and how much to rely on, the highly engaged consumers of brand communities ex-post a crisis. In addition to its marketing relevance, this project has broad societal implications. This paper addresses the general problem of collective reaction to crises. Citizens resort to online and offline communities whenever a crisis hits them – including financial crises, natural disasters, and governmental crises (Fischer, 2018; Jung & Park, 2014; Rasmussen & Ihlen, 2017). Therefore, the insights from this study can support the numerous societal parties dealing with crisis situations, including financial institutions, non-profit organizations, and humanitarian first-responders. These parties can rely on the insights about social network disruptions to understand how people the directly affected by a crisis engage with their social networks. During a natural or government crisis, a deeper understanding of citizens' behavior in their social networks would mean a more efficient and effective crisis response.

While, to my knowledge, the study is the first to focus on the impact of brand scandals on hundreds of online consumer-brand networks, it is not without limitations. The study captures online brand networks, and I do not have data on the offline brand communities. The effect of crises on offline communities can differ relative to online communities, and I cannot make generalizations without a proper empirical investigation. Future research can extend the study in this direction. Currently, this study does not include placebo tests, to analyze the presence of any obvious violations of the identifying assumptions – especially the parallel trends assumption. These additional robustness checks are ongoing, and will be included in future versions of the paper. In the analysis of brand crisis effects on consumer types, I record a change in sign when I introduce product category and crisis controls. This is suggestive of heterogeneous behavior of H- vs L-Type consumers in brand networks. In ongoing analyses, I am exploring the source and extent of that heterogeneity. Finally, this study provides qualitative discussions of the possible behavioral mechanisms driving the results, informed by the empirical findings. It will be important, in future research, to examine which behavioral channels are driving the results – so as to make sharper managerial recommendations for all the practitioners and policy-makers working with community-facing brands.

Chapter 5

Conclusions

5.1 Introduction

In this dissertation, I aimed at investigating the impact of digitization technologies and external, disruptive events on the success and sustainability of shared-interest communities.

In Chapter 2, I focused on the impact of digitizing community activities on the decisions to participate in the activities by the community members. A priori, the effect of digitizing community experiences is ambiguous. On the one hand, digitized solutions for community interactions – such as webinars, live conferences, and asynchronous media – dramatically cut the costs to run these activities, and allow members to remotely participate in the life of their community. On the other hand, literature in marketing and economics suggests that in-person interactions are ideal situations to form deep, meaningful social connections among like-minded individuals (Cohn et al., 2018; Koh et al., 2007). This wealth of social and psychological benefits from in-person activities is a main driver of participation in shared-interest communities (Kang et al., 2014b; Y. Wang & Fesenmaier, 2004). To solve this ambiguity, I collected data about over 12,000 local and virtual communities and 180,000 community events from Meetup.com. Using machine learning algorithms for natural language processing and text classification, I measured the extent to which each community event occurred with a digitized or in-person format. Then, I modeled the impact of event digitization on the participation intentions of community members with two complementary methodologies. First, I used a parametric Structural Causal Model (SCM) to derive the likelihood of RSVPing to a community event with a given extent of digitization. Then, I employed non-parametric Causal Random Forests (CRFs) to assess the robustness of the parametric estimates, while relaxing any functional form assumptions, and fully exploiting any heterogeneity in the estimated average treatment effect of digitization. The results from the study suggest that increasing the extent of activity digitization decreases members' intentions to attend such events. A counterfactual analysis showed that completely digitizing in-person activities causes a median 2.97% decrease in positive RSVPs. I also found that the effect is heterogeneous across communities operating in different interest categories. Chapter 2 contributes to literature studying the effect of digitizing human interactions on people's behavior in social groups, and informs community managers in their efforts to measure and balance the consequences of increasing digitization in their communities.

In Chapter 3, I focused on the effect of negative vs. positive shocks to the shared purpose of an online community on several aspects of healthy community dynamics. Literature in marketing, economics, and network science agrees that there is an important relationship between a community's purpose and its internal dynamics. However, there is little empirical evidence that details what this relationship looks like. Existing theories even suggest a connection between the common purpose of a group and its existence (Preece, 2001; Tajfel, 1978; Zander, 2018), but most empirical studies supporting these theories are either observational, or small scale, or focus on unique, specific events that could potentially disrupt a community's purpose (e.g., Racca et al., 2016; Rasmussen & Ihlen, 2017). To shed more light on the relationship between community purpose and community dynamics, I leveraged pseudo-experimental conditions, created by the outcomes of college basketball games, within hundreds of online communities on Reddit.com–the largest global platform hosting communities online. In particular, using fan communities around NCAA teams as the empirical setting, I collected data on 244 Reddit communities, revolving around 259 teams competing in the first division of the NCAA men's basketball league between 2015 and 2019. The community data include information about more than 196K Reddit users, 822K discussion threads, and 1.5M comments. In addition to the community data, I collected team and game data from the platform DonBest.com, which include information about more than 12.7K games played over 484 game dates. Using a difference-in-difference framework, I found that game outcomes, acting as exogenous shocks on the purpose of the sport communities, have a significant effect on the community dynamics. Therefore, I showed that the purpose of communities plays an important role in their day-to-day existence. I found that a lost-vs-won game – i.e., a negative-vs-positive shock – decreases activity and engagement within communities. I also found that the effect of negative shocks is absorbed differentially within the community: the most highly connected, active, and central community members are most negatively hit by the negative purpose shocks. Furthermore, I showed that negative shocks induce social networks of fans to become more centralized and localized. In terms of user-generated content, I demonstrated that negative purpose shocks reduce the "energy level" in the fans' discussions, and impede expressions of group affiliation. Finally, in a series of heterogeneity and subgroup analyses, I found that the baseline effects are sensitive to the prior expectations of the community members. Overall, I concluded that the purpose of communities plays an important role in their ongoing existence and success. The empirical results suggested that the state of a community's purpose fuels its social dynamics, especially during times of turmoil.

In Chapter 4, I continued the investigation on the effects of external disruptions on community dynamics. In this Chapter, I focused more specifically on brand communities, and I showed that brand communities online are significantly affected by brand crisis events. So far, the literature has suggested that customer interactions online following brand crises negatively impact brand shareholder value, consumers' brand share, and category purchases (Ahluwalia et al., 2000; Backhaus & Fischer, 2016; Hsu & Lawrence, 2016). However, the impact of brand crises on the behavior of consumers in online brand communities remains unclear. To fill this gap, I collected data on about 300 brand communities hosted on Reddit.com. I complemented the community data with information about over 7,000 brand crisis events, reported by media outlets between 2010 and 2019. In addition to the crisis data, I also collected information on the companies and brands involved in the brand crises (i.e., the same brands for which I observed the Reddit discussion communities). In the resulting panel dataset, I tracked all interactions between consumers in the brand communities for 180 days around any brand crisis – 90 days preceding and 90 days following the crisis events. This resulted in a panel of 13M posts and comments, generated by 1.9M unique brand community members. I further exploited the thread structure of the brand community discussions to construct bipartite social networks of consumer interactions. I leveraged the data on discussions and social networks to measure which consumers occupy a high- or low-importance position in the community – both in terms of community engagement before the crisis, and in terms of their position and status in the brand social networks (*High* versus *Low-type* consumers). Using a difference-in-difference framework, I showed that after a brand crisis, consumers' activity in online brand communities increases by 9.1%. However, the change in activity is significantly positive only due to the contributions of "bandwagon consumers" – consumers who become active in brand communities exclusively after a brand crisis is covered by the media. On the contrary, consumers who were already active prior to the crisis event significantly *decrease* their activity in the communities after a brand crisis. In terms of consumer types, I showed that high-type consumers contribute relatively more to the brand communities after a brand crisis, compared to low-type consumers. The findings on the structure of brand networks suggested that brand crises significantly alter the ease and speed of information diffusion. An average brand crisis causes a 1% increase in degree centrality, a 0.2% increase in clustering coefficient across network members, and a 0.1% decrease in average inverse closeness centrality. After brand crises, information travels through a more diverse pool of consumers, and in more tight-knit discussion sub-groups. Finally, I showed that the effect of brand crises on brand community engagement is differential across types of crisis impact, brands operating as B2C versus B2B, intensity of media coverage, and severity of the crisis consequences.

5.2 Theoretical Contributions

This dissertation contributes to various literature streams in marketing, economics, and information and network sciences.

The insights from Chapter 2 contribute to literature in marketing, operation science, and economics, investigating the impact of digitizing human interactions on economic behaviors – including cooperation and contribution to the public good (e.g. Cohn et al., 2018; H. F. Lin, 2007; Short et al., 1976; Rothaermel & Sugiyama, 2001; Wirtz et al., 2013). To this stream, I contribute with out-of-lab evidence that digitizing human interactions has an impact on the consequential decisions of people to engage in such interactions. The results from Chapter 2 also contribute to literature in marketing and sociology investigating the antecedents of active participation in communities of interest (e.g. Dessart et al., 2015; Kang et al., 2014b; Y. Wang & Fesenmaier, 2004; Wirtz et al., 2013; Zhou, 2011). In particular, I provide evidence that the extent of digitization of community activities is another potential antecedent of members' participation in shared-interest communities. Furthermore, to date, Chapter 2 is the most comprehensive study on digitization of community experiences, that takes into account multiple geographies, communities, interest categories, and event types. Indeed, marketing and sociology literature has typically focused on either single communities offering activities with varying degrees of digitization, or on multiple communities employing only one communication format – either in-person, or fully digitized (e.g. Dessart et al., 2015; Dutta-Bergman, 2005; Kang et al., 2014b; Koh et al., 2007; Ling et al., 2005; Y. Wang & Fesenmaier, 2004). However, industry evidence suggests that marketing and community managers are increasingly resorting to various activity formats – without necessarily committing to solely digitized or in-person options – and will continue to offer a range of formats in the coming years (Bevy, 2021).

With Chapters 3 and 4, this dissertations contributes to existing research on negative brand reputation and brand crises. Literature in marketing has explored the reaction of consumers to product or service failures through valenced word-of-mouth in brand communities. These studies focused on the effectiveness of different company- vs consumer-initiated recovery efforts (Schaefers & Schamari, 2016; Yuan et al., 2020), or on the impact of community engagement following a product-harm crisis on brand equity and shareholder value (Hsu & Lawrence, 2016), rather than on the effect of the event itself on the dynamics of the consumer groups. Another stream of literature has focused on the impact of negative brand publicity – which can potentially hinder the common purpose of established brand communities – on brand sales (Berger et al., 2010), attitudes towards the brand (Ahluwalia et al., 2000), and brand equity (Dawar & Pillutla, 2000), but not on the dynamics of brand communities online. In relation to these literature streams, Chapters 3 and 4 make two substantial contributions. First, the two chapters offer empirical evidence of the impact of the disruptive, brand-related events on the dynamics, success, and resilience of the brand communities, rather than other outcomes. Second, Chapters 3 and 4 contribute to the marketing literature on negative publicity and external disruptions on consumer behavior in social networks, social media, and online communities (e.g. Ahluwalia et al., 2000; Dawar & Pillutla, 2000; Dondio & Usher, 2017; Hsu & Lawrence, 2016; Racca et al., 2016; Rasmussen & Ihlen, 2017). The chapters contribute to these streams by considering crises and community purpose shocks in hundreds of consumer communities, and across different empirical contexts – including brand communities and communities of sport fans. I also complement the existing literature by expanding the empirical examinations beyond single, rare disruptions. Brands and organizations need to deal with shocks to the quality of their products, services, and reputations much more frequently than they have to manage natural disasters and financial crises. This is especially true when consumers are heavily and promptly interconnected through online social networks. With this dissertation, I evaluate events that are close, in principle, to frequent reputation disruptions for a company or brand.

This dissertation also makes several methodological contributions. Chapter 2 contributes to literature in digital marketing, offering a framework to model digitization effects under endogeneity and censoring, and across multiple geographies and periods. Similar endogeneity and censoring concerns have been raised in different contexts, from pricing strategies to digitized entertainment in movie markets (e.g. Rooderkerk et al., 2013; Yang et al., 2021). Typically, these concerns are solved via instrumental variable estimation, or via randomized experiments in the field. However, in many digital marketing contexts – and especially when dealing with digitization of community experiences during a pandemic – resorting to instruments or RCTs can be both practically difficult and ethically problematic. In this study, I rely exclusively on observational data easily available to most community managers. The estimation strategy in this study can be extended to digital marketing problems that are based on comparable data generating processes, producing non-random treatment assignments and observable censored outcomes. Furthermore, in Chapters 3 and 4, I exploit quasi-experimental conditions that strengthen the causal claims made in the studies that external, disruptive, threatening events can significantly impact internal community dynamics. Literature in marketing and network science has attempted to establish credible links between shared-interest communities, their internal dynamics, and the external environment in which they operate (e.g., Preece, 2001; Tajfel, 1978; Zander, 2018). However, the overwhelming majority of empirical evidence to support these theories is either correlational, or small scale, or refers to very rare or specific events, often not exogenous to the community dynamics (e.g., Racca et al., 2016; Rasmussen & Ihlen, 2017). Causal inference through field experiments (at least, in organic communities and at a sufficiently large scale) is extremely difficult. Additionally, experiments that disrupt the common purpose of a community – performed over a large number of real life communities and members – are not only very costly and complex, but often also ethically problematic (e.g., El-Sayed et al., 2013). To fill this gap, Chapters 2 and 3 propose two different quasi-experimental settings and two different identification strategies, such that these relationships can be studied and quantified using a causal approach.

5.3 Managerial Contributions

The results from this dissertation have several implications for marketers, managers, and policy-makers working with community-facing channels.

Chapter 2 has direct implications for companies, institutions, and policy-makers evaluating the balance between in-person and digitized experiences. For a community organizer, the results from Chapter 2 suggest that – if community participation is an important objective – digitized activities should not completely replace in-person activities. While digitized activities remain a viable, low-cost option to connect community members, the digitized formats probably generate a set of benefits for community members that do not necessarily correlate with active participation. The insights from Chapter 2 also suggest that the idiosyncratic category norms, rules, expectations, and social constructs play a very important role in explaining why different communities record higher or lower participation rates to their digitized events. An important managerial implication from this dissertation is that (i) nurturing and educating community members to the advantages of digitization, and (ii) highlighting the category-specific benefits from participating in digitized events, may be ways to mitigate the potential disutility from participating in digitized community events.

The implications from Chapters 3 and 4 are especially relevant in the context of brand-related events. Throughout the dissertation, I considered examples of brands whose online presence is disrupted by sudden, negative publicity – such as a product-harm crisis, a product failure, or an episode of corporate social irresponsibility. The findings from Chapters 3 and 4 suggest that, under those circumstances, managers may want to communicate with the leaders or "brand ambassadors" of their online communities. The communication could be centered, for example, around nurturing relationships with core/high-type members in anticipation of a negative event (to manage their expectations) or around strengthening the ties with the core/high-type users after a negative event occurs (to a faster recovery of community dynamics and growth). The results from Chapter 3 also suggest that prior expectations are important for the mitigation of disruptive outcomes. When the community clearly expects a negative outcome from a certain event, the damage to community dynamics is mitigated. Therefore, the findings from Chapter 3 suggest that managers may be able to mitigate disruptions by setting more realistic expectations about the outcomes of a product crisis or brand failure, before the event even transpires to the public. In terms of user-generated content, both Chapter 3 and 4 indicate that managers may have an opportunity to comfort and reduce the negativity induced by the events – for example, through firm communications that encourages and induces positive affect and promotes a sense of membership. Finally, the results from Chapter 3 concerning new members suggest that people join and contribute to communities when the "common denominator" is not threatened. Managers facing negative events may remind consumers about their shared interest – the common mission and vision that brought them together in the first place. Also, managers should be aware of how important the common interest is to the smooth operation of the community and spend effort in maintaining it.

In addition to its marketing relevance, this dissertation has broad societal implications. Chapter 2 taps into one of the most fundamental questions of the last two years, and certainly, a pressing concern for the upcoming period: what are the consequences of digitizing human interactions? Under which circumstances is this digitization detrimental to the utility from participating in social experiences? The insights from Chapter 2 are grounded in the fact that digitization is a serious, practical concern not only for marketing and community managers, but also for institutions and governments. Indeed, on top of their organic growth in popularity, digitized experiences dramatically gained more relevance during the COVID-19 pandemic. Among other things, the pandemic has forced marketing professionals, workplaces, and governments to evaluate the balance between digitized and in-person activities. Meanwhile, marketing managers have continued to shift resources to building digital customer interfaces between 2020 and 2021. Indeed, investments in digital interfaces grew by 21.0% in February 2021 since June 2020 (The CMO Survey, 2021). Similarly, community professionals predict that virtual events will continue to be essential even after the pandemic emergency (Bevy, 2021). Therefore, the arguments and results from Chapter 2 can give academics and managers empirical tools to evaluate the impact of community digitization in light of the Covid-19 pandemic. Chapters 3 and 4 address the general problem of collective reaction to crises. Citizens resort to online and offline communities whenever a crisis hits them – including financial crises, natural disasters, and governmental crises (Fischer, 2018; Jung & Park, 2014; Rasmussen & Ihlen, 2017). Therefore, the insights from both chapters can support the numerous societal parties dealing with crisis situations, including financial institutions, non-profit organizations, and humanitarian first-responders. These parties can rely on the insights from this dissertation about social network disruptions to understand how people the directly affected by a crisis engage with their social networks. During a natural or government crisis, a deeper understanding of citizens' behavior in their social networks would mean a more efficient and effective crisis response.

5.4 Suggestions for Future Research

The findings and limitations of this dissertation suggest numerous avenues for future research on the behavior of consumers and citizens in shared-interest communities.

In Chapter 2, the consequences of community digitization were evaluated in a pre-Covid situation. In the context of community digitization, future studies may expand the evaluation of digitization policies in a post-Covid reality, and compare how the shift to remote working and 100% digitized social activities has affected the reaction of community members to activity digitization. For example, future research may assess if population density is an important confounding variable in the relationship between digitization and community participation in light of the Covid–19 emergency. Future work could also improve on the measurement of the digitization construct proposed in this dissertation. Future studies may train alternative language processing models to detect activity digitization from text. A refined model of digitization detection would greatly help researchers and practitioners to understand which elements of language contribute to accurately predict digitization. Also related to measurement, future research may study the effect of community digitization on additional outcomes from the community participation spectrum – which includes passive participation, referrals, moderation, and even negative and disruptive participation (Ardichvili et al., 2003; Dutta-Bergman, 2005; Brodie et al., 2013; Kang et al., 2014b). Future work could assess the interplay between anonymity, digitization, and negative forms of community engagement – such as verbal assaults and targeted virtual violence. Finally, in Chapter 2, I recommended the design and implementation of field experiments to evaluate the effectiveness and consequences of digitized events on community participation. After resolving any ethical concerns to assigning people to digitized or in-person situations, controlled experiments can provide unique insights into the mechanisms behind the effects recorded in the observational setting, and into how community members interact *with each other* during different types of events. From such studies, managers and policy makers could gain from learning about the boundaries of the digitization effects.

Beyond the topic of digitization, there are several opportunities to study the relationship between community resilience and community dynamics in relation to the environment in which the community operates. In Chapter 3, I evaluated the impact of negative shocks to a community's purpose against positive ones of similar nature. In future work, it would be ideal to compare negative shocks against the *absence* of a disruption, preferably in a context in which it is still possible to adequately control for the prior expectations of the community. Furthermore, in Chapters 3 and 4, I tracked the behavior of community members and the relevant metrics of resilience and success for 30 to 180 days around disruptive events. In future work, it will be crucial to understand whether the disruptions to the communities keep lingering for longer periods. Understanding the permanence of structural disruptions in the communities following an external, threatening event is fundamentally important to plan crisis response strategies.

Finally, this dissertation covered an array of empirical settings – from 33 interest categories on Meetup, to NCAA teams on Reddit.com, to brand communities in more than 20 product categories and economic sectors. Future work should consider to extend the analyses in this dissertation to other practically important contexts – including knowledge-sharing communities, gaming, gambling, and other potentially addictive situations, and corporate or organizational teams working together towards a common objective.

Chapter A

Appendix to Chapter 2

A.1 Measuring Event Digitization

The raw data from the Meetup API do not include any field describing event formats (i.e. online, in-person, hybrid). Therefore, I had to define a measure of event digitization based on the information available in the raw data. As a source of digitization information, I exploited the *event descriptions* created by the event organizers. Event descriptions are visible to group members, and are crafted to inform perspective attendees about the event format (online, offline, or hybrid). Additionally, descriptions typically provide details on how to join the event location, and describe which activities will be performed for the duration of the event. In practice, group members can use the event descriptions to evaluate the event attractiveness, and form a decision about their event participation intentions. I used the *event venue* field as an additional source of information. The event venue field typically contains the address of the location in which in-person events take place, or the name of the platform used to host digitized events.

To extract information from the text, I created a list of non-empty event descriptions, using the events organized by the groups in the sample from the group creation date until June 2019. I processed the text to remove HTML tags, trailing whitespaces, English stopwords, phone numbers, punctuation, and special characters. I then used the cleaned description text as an input for two Support Vector Machine

(SVM) classifiers.

A.1.1 Training Set

To train the SVM classifiers, I created a training set of labeled descriptions as follows:

Step 1: Matching Keywords As a first step, I defined two vectors of keywords that could potentially indicate that an event format was completely digitized or completely offline. The digitized-event keywords were “online event”, “remote meeting”, “webinar”, “gotomeeting”, “webcast”, and “remotely”. The offline-event keywords were “space provided”, “breakfast served”, “coffee served”, “seats”, “snacks”, “drink”, “drinks”, “meet greet”, “doors”, “indoor”, “outdoor”, and “entrance”. Additionally, I defined a vector of event locations that clearly indicated that the event format was completely digitized. This final vector contained the words “http://”, “https://”, “online”, “computer”, “webinar”, “anywhere”, “your house”, “iphone”, “webcast”, and “your computer”. Finally, I obtained a list of tools used to organize digitized community activities from Spinks (2020), and appended the list of tools to the digitized-event keyword vector. I then selected, filtering from the full list, the event descriptions that contained both the digitized-event keywords and the digitized-location keywords, excluding the descriptions that contained the offline-event keywords. This gave us a first set of digitized event descriptions. I also selected the event descriptions that contained the offline-event keywords, excluding the descriptions that contained any of the digitized-event or digitized-location keywords. Finally, I isolated the event venues that contained the keywords “http://”, “https://”, “online”, “computer”, “webinar”, “anywhere”, “your house”, “iphone”, “webcast”, and “your computer”. I excluded from the training set all the events that were initially identified as offline if they had one of these keywords listed as their event venue.

Step 2: Random Sampling and LDA Topic Model The use of specific keywords to select training cases may introduce bias in the labeling process. To address this concern, I added to the training set a random sample of 200 labeled online event descriptions, a random sample of 1000 labeled offline event descriptions, and a random sample of 3000 unlabeled event descriptions. To further reduce the case selection bias, I trained an LDA topic model with 32 topics on all the available event descriptions. I identified one topic (topic 9) containing events with digitized formats, and I added the corresponding cases to the training set.

Step 4: Labeling Training Cases I employed 2 research assistants (RAs) to independently label the training cases. The independent RAs classified the events depending on whether the text descriptions were describing activities with a "Digital/Virtual" format (i.e., people in the group met in a digitized, digital, online activity), and/or an 'In-Person" format (i.e., people in the group met face-to-face during the activity). The RAs labeled the event descriptions with the class that most appropriately described the activity format ("Digital/Virtual" and/or "In-Person"). The events could be labeled as both "Digital/Virtual" and "In-Person" – in that case, the activity would be typically described as "Hybrid". When the two RAs chose different classifications for the same description, a third independent rater who was not previously involved in the classification task resolved the disagreements. The labeling phase resulted in a training set with 2851 cases, of which 158 classified as "Digital/Virtual", 2679 classified as "In-Person", and 14 classified as both ("Hybrid").

A.1.2 SVM Predictions

I trained two Support Vector Machines (SVMs) on the set of labeled cases. I trained the first SVM using the "In-Person" label, and the second SVM using the "Digital/Virtual" label. Then, I let the two SVMs predict the most likely class of all the remaining unlabeled event descriptions (respectively "In-Person" versus "Not In-Person", and "Digital/Virtual" versus "Not Digital/Virtual"). This prediction step resulted in four new variables for each event description: (1) "In-Person" prediction label: most likely class of the text description ("In-Person" or not "In-Person") based on the SVM model trained on the "In-Person" label; (2) Probability associated with the "In-Person" (or not "In-Person") predicted class; (3) "Digital/Virtual" prediction label: most likely class of the text description ("Digital/Virtual" or not "Digital/Virtual") based on the SVM model trained on the "Digital/Virtual" label;

(4) Probability associated with the "Digital/Virtual" predicted class.

A.1.3 SVM Performance

Using 10-fold cross-validation, the two SVM models achieved between 96.6% and 99.3% prediction accuracy. Table A.1 reports the 10 prediction accuracies resulting from the cross-validation for each of the two models. The "In-Person" model achieved an average 98.5% prediction accuracy, while the "Digital/Virtual" model achieved an average 97.9%.

Table A.1: SVM 10-fold Cross-Validated Prediction Accuracies

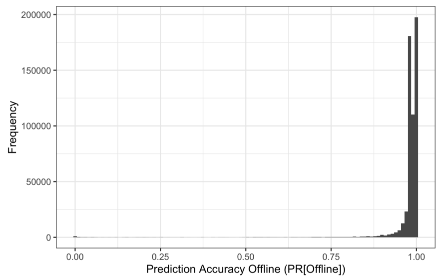
	Prediction Accuracy (%)	
CV Fold	In-Person SVM	Digital/Virtual SVM
1	97.67	97.28
2	99.27	98.90
3	98.15	97.78
4	97.86	97.86
5	97.98	96.64
6	98.93	97.52
7	98.01	98.34
8	99.29	98.23
9	98.98	98.31
10	99.34	99.01
Average	98.55	97.99

A.1.4 Prediction Descriptive Statistics

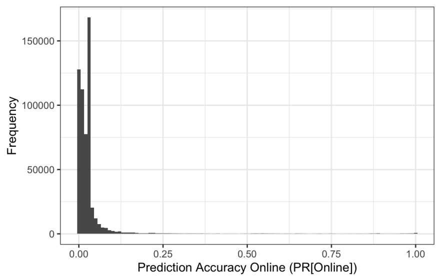
Figure A.1 shows the distribution of predictions generated by the two SVM models. The left panel demonstrates that the "In-Person" model predicted that most events occur with an in-person format. The right panel shows that the predictions from the "Digital/Virtual" model are consistent with the previous result, and that most events have a non-digital/non-virtual format.

Table A.2 describes which labels were attributed to each event in the panel. The vast majority of the events (99.6% of the total) were labeled consistently across prediction models. A small fraction of events (0.4%) were labeled differently by each SVM model – the 0.2% of the event was labeled as both "In-Person" and "Digital/Virtual", and the 0.2% was labeled as neither. Inspecting a random sample of event descriptions, the inconsistent labels can be explained in three ways. One type

Figure A.1: Distributions of Prediction Accuracies from In-Person SVM Model (a) and Digital/Virtual SVM Model (b)



(a)



(b)

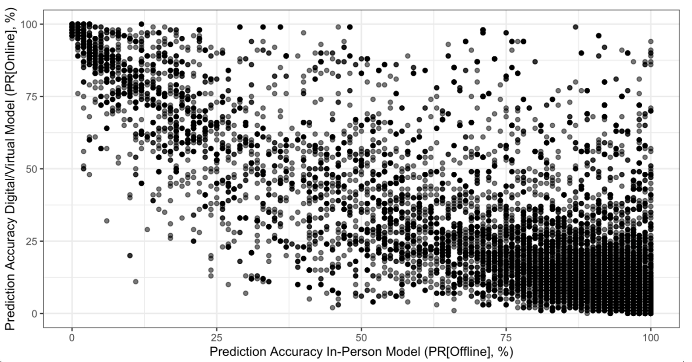
of inconsistency derives from a misclassification – one of the two labels is correct, and the other is incorrectly classified. In this case the prediction accuracies are informative, and the label with highest prediction accuracy is typically the right one. The second type of inconsistency derives from events that actually have blended formats. These events are typically in-person, but offer a virtual live stream, real-time videos, or asynchronous digital material. The last type of inconsistency describes events with little or no information, and reflects the low classification confidence of either or both SVM models.

Table A.2: Classification Labels from the In-Person and the Digital/Virtual SVM Models

Predicted Label				
In-Person SVM	Digital SVM	N	Total	(%)
Not In-Person	Not Digital/Virtual	900	562061	0.002
In-Person	Digital/Virtual	1397	562061	0.002
Not In-Person	Digital/Virtual	4236	562061	0.008
In-Person	Not Digital/Virtual	555528	562061	0.988

Finally, figure A.2 shows that, overall, the predictions from the two models appear highly correlated, and that the majority of the predictions are concentrated in the upper-left (Not In-Person, Digital/Virtual) and bottom-right (In-Person, Not Digital/Virtual) regions of the plot.

Figure A.2: Prediction Accuracies from In-Person SVM Model (x) and Digital/Virtual SVM Model (y). The correlation coefficient between x and y is equal to -0.908 ($t = -1627.5, \mathrm{df} = 562059, \text{p-value} < 0.001$



A.2 Event Awareness Indicator

In this paper, I make a distinction between members who were potentially unaware of an event, and members who were potentially aware. I distinguish the two types of members with an indicator variable.

To construct the awareness indicator, I relied on two variables available for each member-event pair in the sample: the *event series* indicator and the *timestamp of last group visit*. The event series indicator is *true* if the event is part of an event series – a set of events that repeat with fixed frequency (every week, every two weeks, or every month). The *event series* indicator is *false* if the event is a regular event.

For regular events, I also exploited a feature active on Meetup in 2019. In 2019, Meetup sent RSVP reminders, via email and to all group members, 6 days before the scheduled event date. Because of the reminder, I set the event awareness variable to 1 for all members who did not RSVP to a regular event. I also imputed their potential time of response, and set it at a date corresponding to 6 days before the scheduled event date. If a member who did not RSVP to a regular event visited the group after the event creation date and before the 6-day threshold, then I imputed the time of response as the most recent time at which the member visited the group.

For event series, if a member did not RSVP to an event, but visited the group after the event was created, I assumed that this member *decided* not to RSVP to the event. Therefore, I assumed that the person was aware of the existence of the event, and set the event awareness variable to 1. I also imputed their potential time of response, and set it at a date corresponding to 24 hours before the event occurrence. This is the last time window in which the member could have made a decision about RSVPing.

Finally, if a member did not RSVP to a recurring event, and did not visit the group before the event creation date, then I assumed that this member was potentially unaware of the event. I set the event awareness variable to 0, and did not impute their potential time of response. Figure A.3 summarises the event awareness measurement procedure.

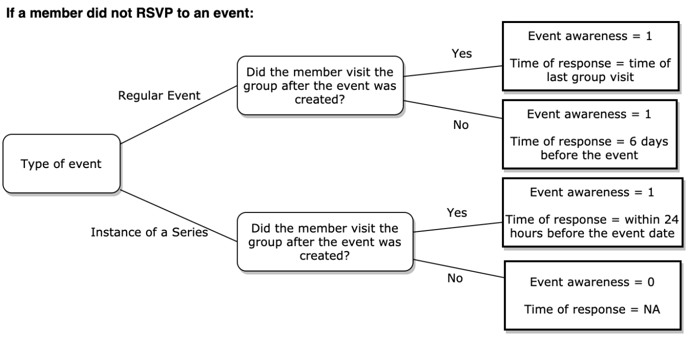


Figure A.3: Event Awareness Operationalization – Decision Flow

A.3 Non-Responses and Event Digitization

To complement the descriptive analysis of non-responses from Section 2.2.2, I check how non-response rates vary with positive RSVPs and event digitization. For this descriptive analysis, I focus only on the groups that organize both digitized and in-person events.

Figure A.4 shows that, in those groups, the relationship between non-response rate per event and positive response rate per event does not vary across digitized and non-digitized events ($N = 254$, $t = -0.87$, $df = 119.31$, p-value $= 0.3843$). The implication of this analysis is that I can model the non-response choices in the same way as positive and negative responses, without making additional assumptions about their relationship with event digitization.

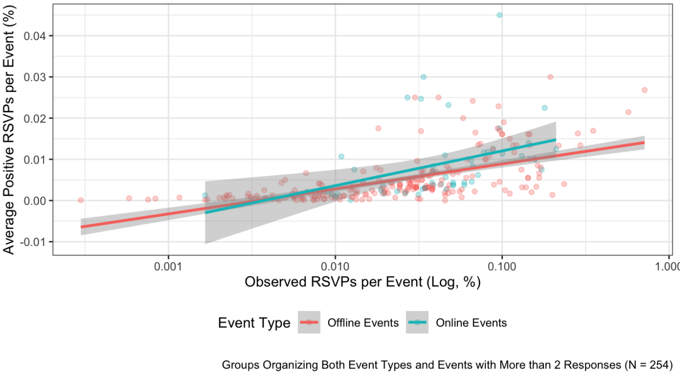


Figure A.4: (Log) Response Rate, Positive Response Rate and Event Digitization

A.4 Simulation-Based Calibrations Results

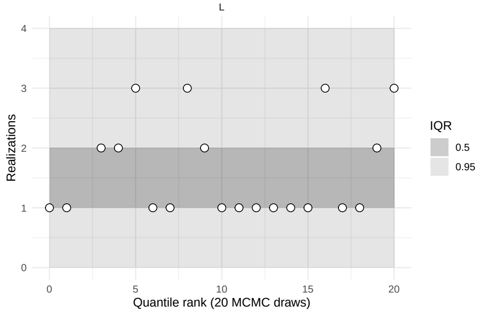
I verified the correctness of the Bayesian SCM using Simulation-Based Calibration (SBC), a procedure for validating inferences from Bayesian algorithms (Talts et al., 2018). In particular, I rely on the histograms of the rank statistics to understand if the analysis has been correctly implemented.

Any deviations from uniformity in the SBC rank histogram can indicate bias or mis-calibration of the computed posterior distributions. Uniformly distributed rank statistics are consistent with a correctly specified model. Spikes at the boundaries of the SBC histogram indicate that posterior samples possess non-negligible autocorrelation. Finally, symmetric,inverse-U-shaped distribution indicates that the computed data-averaged posterior distribution is overdispersed relative to the prior distribution (light red). This implies that on average the computed posterior will be wider than the true posterior.

Here, I report the results from running the SBC algorithm on a random sample of 20 groups from the dataset, assuming that the SCM model reflects the true data generating process. I calibrate the SBC algorithm setting $N = 32$ and $L = 20$. Figures A.5 (vector or scalar parameters), A.6 (parameter matrices), and A.7 (fixed effects and intercepts) show that the SBC rank histograms for all the parameters show no significant deviations from uniformity. This result indicates no issues with the model specification.

Figure A.5: SBC Rank Histogram for the SCM Parameter Vectors or Scalars Specified in Section 2.3.2

(a) L - Censoring Threshold



(b) β_{DC} - Heterogeneous Event Digitization Parameter

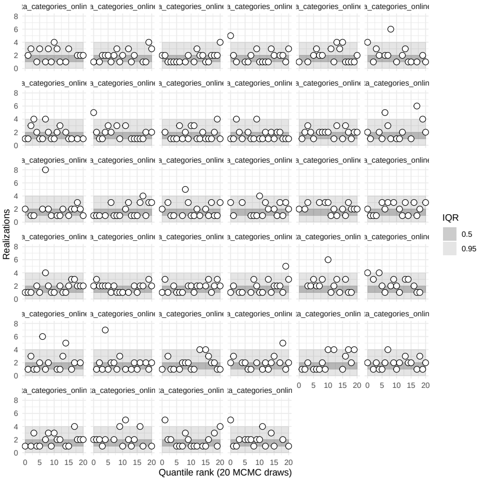
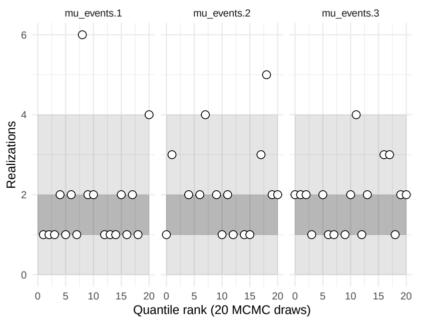
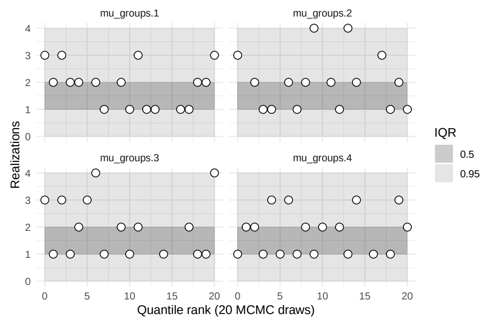


Figure A.6: SBC Rank Histogram for the SCM Parameter Matrices Specified in Section 2.3.2

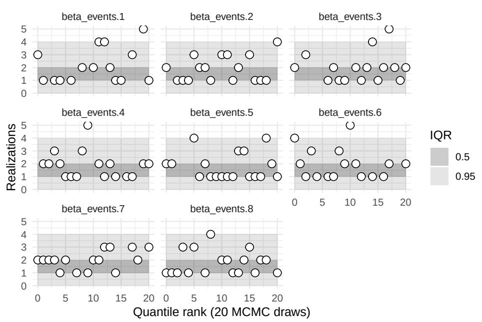
(a) μ_2 - Member-Event Features Parameters



(b) μ_3 - Member-Group Features Parameters



(c) β_e - Event Features Parameters



(d) γ - Group Features Parameters (including group categories)

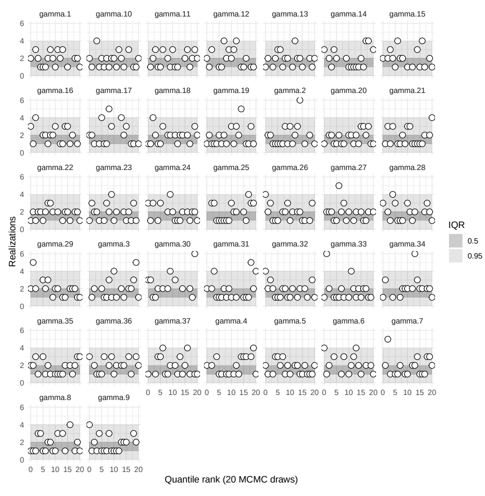
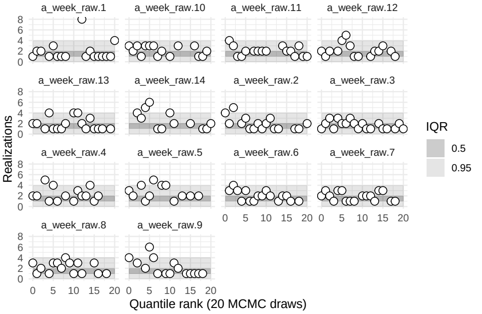
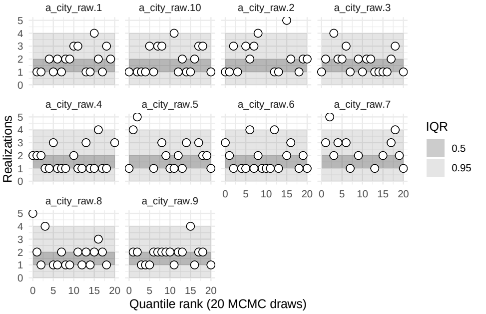


Figure A.7: SBC Rank Histogram for the SCM Fixed Effects and Group Intercepts Parameters Specified in Section 2.3.2

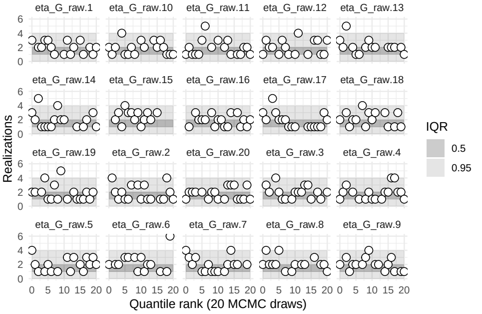
(a) τ - Time-Varying FE



(b) ζ - Location-Varying FE



(c) η - Group Intercepts



A.5 MCMC Diagnostics

A.5.1 Energy Diagnostics

The energy diagnostic for HMC quantifies the heaviness of the tails of the posterior distribution. The energy diagnostics can identify overly heavy tails that are also challenging for sampling. Figure A.8 show that there is large overlap between the π_E and the $\pi_{\Delta E}$ histograms, which indicates no sampling challenges due to overly heavy distribution tails.

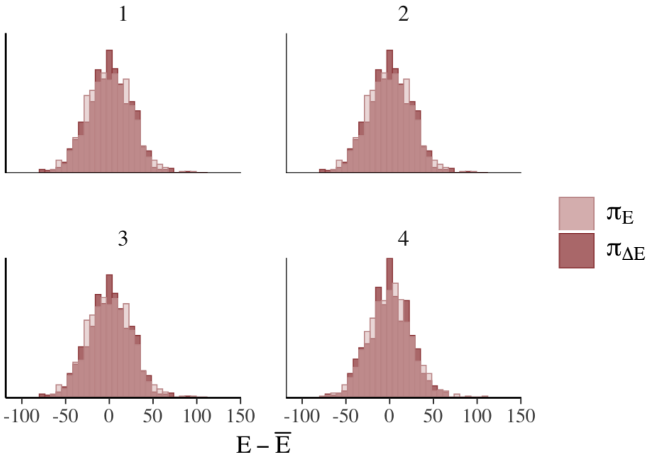


Figure A.8: Energy Diagnostic Plot for the No-U-Turn-Sampler (NUTS)

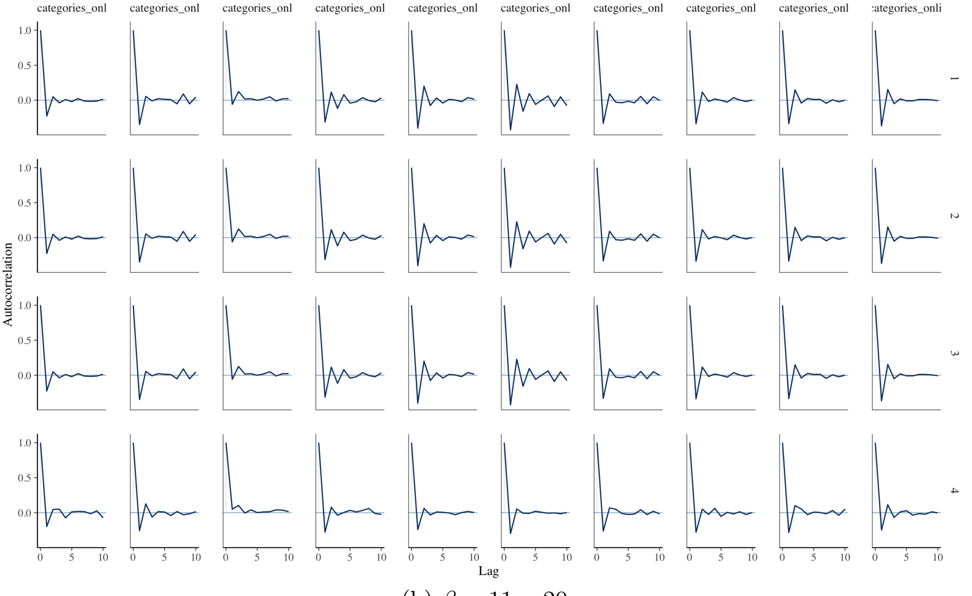
A.5.2 Autocorrelation

Positive autocorrelation between iterations during the MCMC estimation signals that the chains tend to stay in the same area between iterations, and that there may be no convergence (or slow convergence) of sample mean towards true mean.

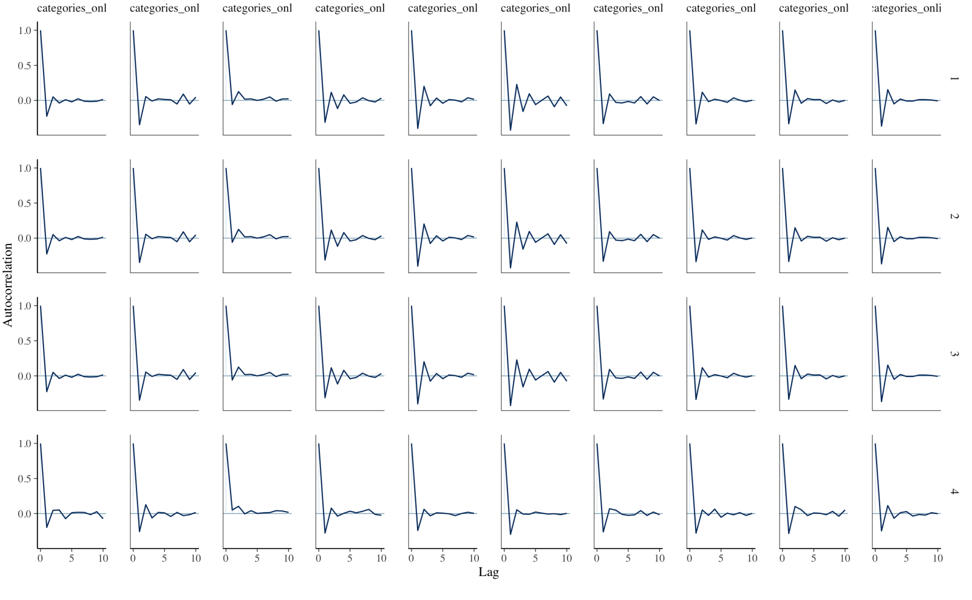
Figure A.9 shows the autocorrelation analysis for the causal effect parameter of interest. As I go further along the chains the values become less correlated, occasionally achieving negative autocorrelation.

Figure A.9: Autocorrelation Diagnostic Plot for β_{DC}

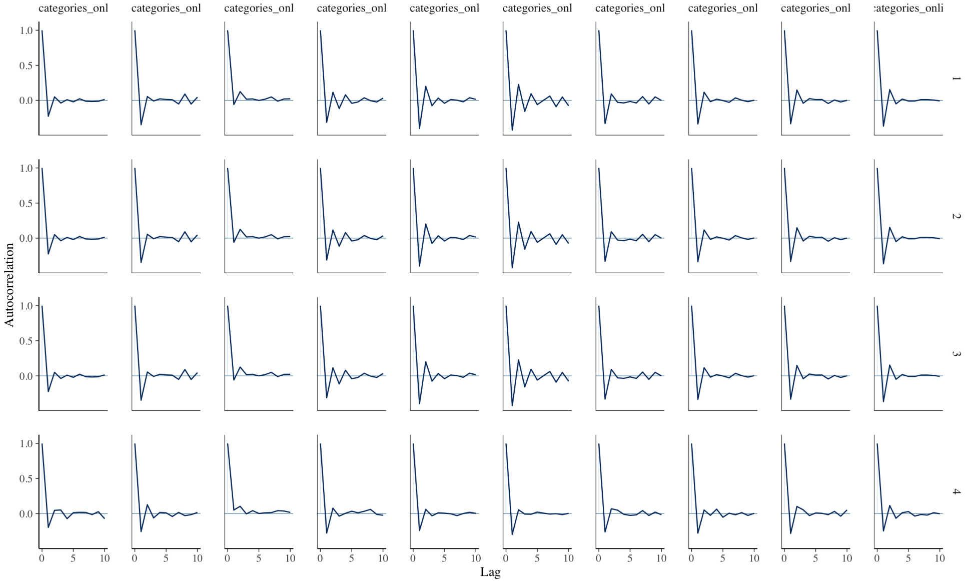
(a) $\beta_{DC} 1 - 10$



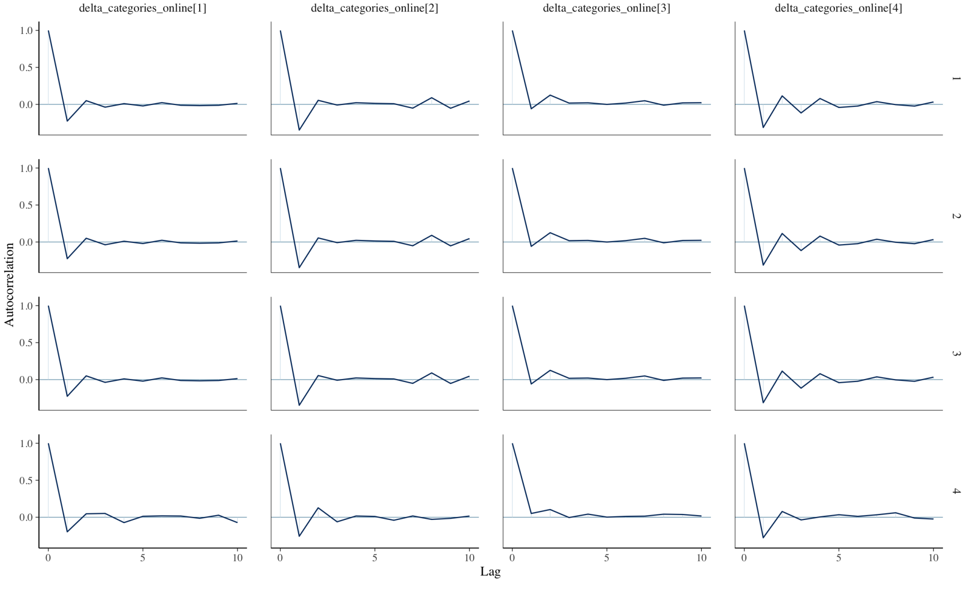
(b) $\beta_{DC} 11 - 20$



(c) $\beta_{DC}21-30$



(d) $\beta_{DC}31-34$



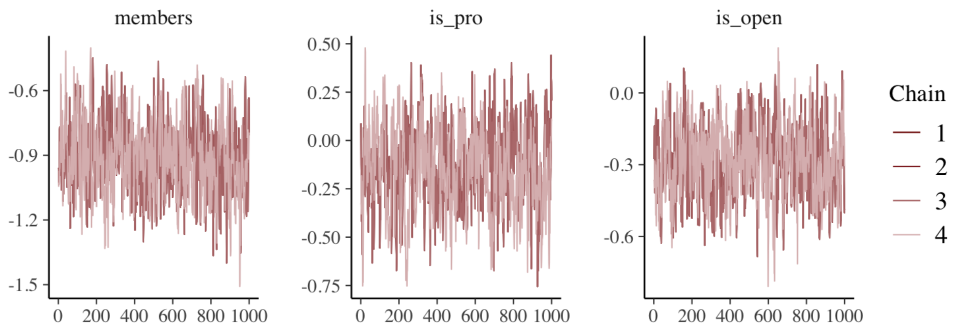
A.5.3 Traceplots

MCMC traceplots show the sampled values of a parameter over the iterations and across chains. These plots help to judge how quickly the MCMC procedure converges to the parameter values, and whether the sampler fails to explore certain areas of the parameters' space.

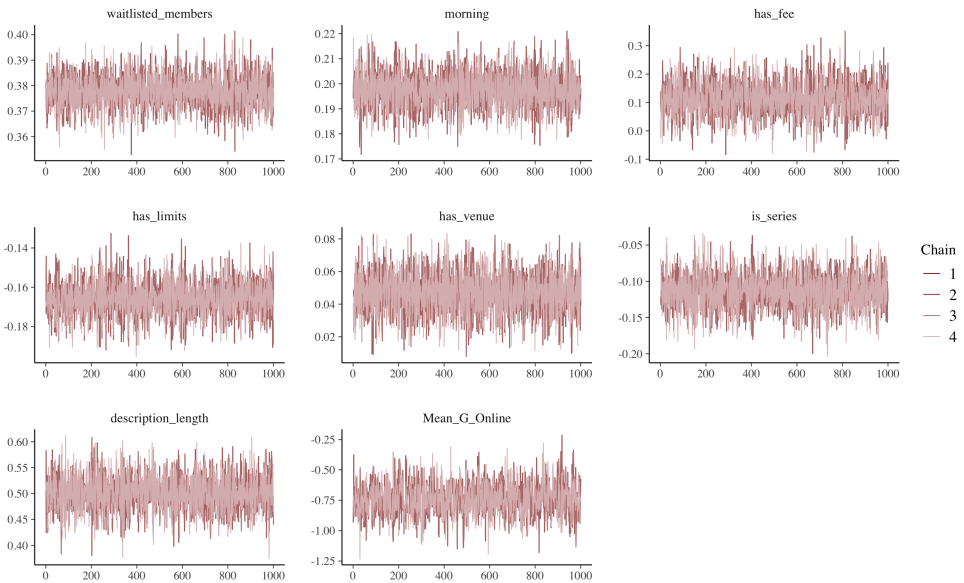
Figures A.10a to A.10d show the traceplots for the parameter estimates referring to group, event, and category variables.

Figure A.10: MCMC Traceplots

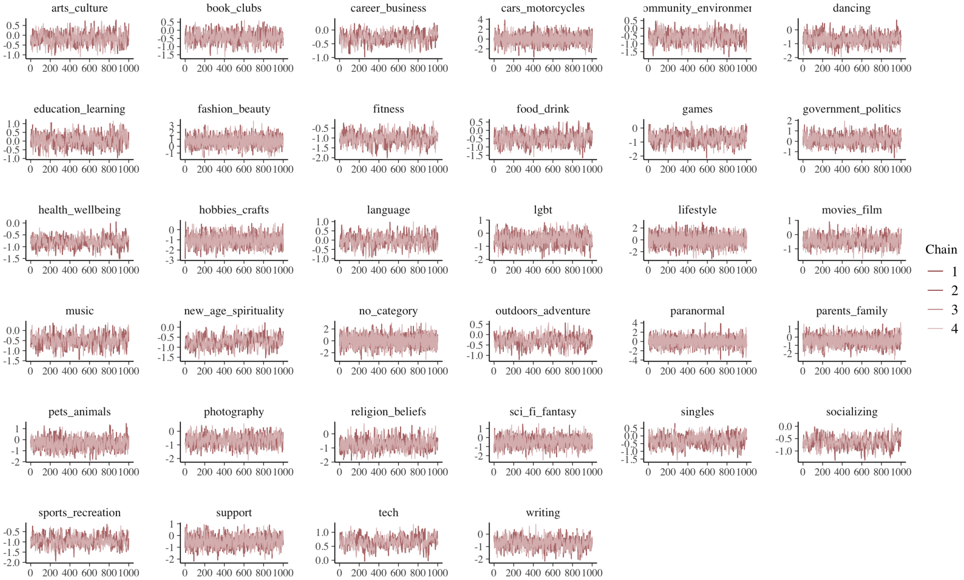
(a) Group-Level Variables ($\gamma_{1,2,3}$)



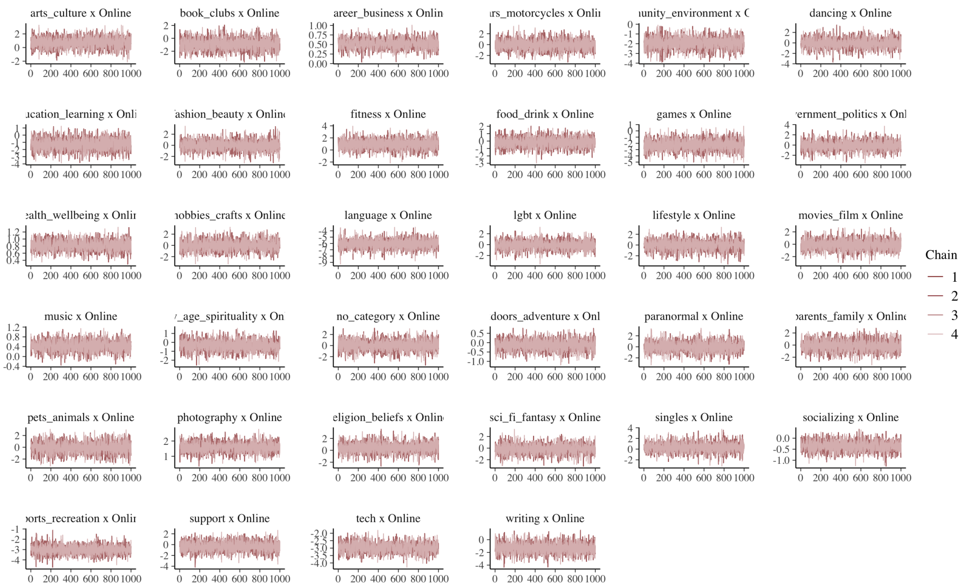
(b) Event-Level Variables (β_3)



(c) Baseline Categories Variables (β_c)



(d) Digitization $\times$ Categories Variables (β_{DC})



A.6 Additional SCM Figures and Tables

Figure A.11: Posterior Density of β_{DC} Estimates – Weighted by Number of Groups per Category

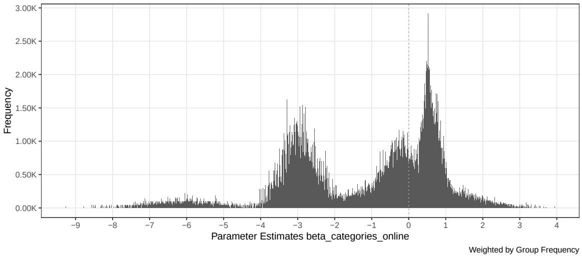


Figure A.12: Posterior Density of β_{DC} Estimates – Weighted by Number of Observations per Category

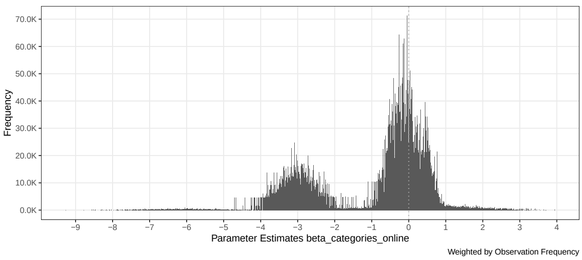


Table A.3: Posterior Parameters for Category × Digitization Probabilities – only considering categories which hosted any digitized event.

Variable	Mean	SE	SD	2.5% CI	50% CI	97.5% CI	ESS	$\hat{R}$
Digitization × Interest Category (β_{DC})								
Photography	1.60	0.00	0.36	0.99	1.60	2.20	5552	1
Health Wellbeing	0.82	0.00	0.15	0.56	0.81	1.10	3439	1
Career Business	0.52	0.00	0.13	0.31	0.52	0.75	2895	1
Music	0.43	0.00	0.23	0.04	0.43	0.82	5240	1
Outdoors Adventure	-0.14	0.00	0.30	-0.63	-0.14	0.37	5787	1
Food Drink	-0.19	0.01	0.70	-1.40	-0.18	0.94	7558	1
Socializing	-0.35	0.00	0.25	-0.78	-0.34	0.06	3405	1
Tech	-3.00	0.01	0.34	-3.50	-3.00	-2.40	3111	1

Note: Posterior distributions estimated using a Hamiltonian Monte Carlo algorithm. Posterior statistics calculated over 4 chains, 1000 iterations per chain. Specification estimated: $u_{ie} = D'_e * C'_g \beta_{DC} + C'_g \beta_c + X'_e \beta_e + X'_{ieg} \mu + \zeta_m + \tau_e + \eta_g + \epsilon_{ie}$; $Y_{ie} = 1$ if $u_{ie} > 0 > L$; $Y_{ie} 0$ if $L < u_{ie} \leq 0$; $Y_{ie} = -1$ if $u_{ie} \leq L$. Priors: $\eta_g \sim N(X'_g \gamma_g, 1)$; $\beta_{DC}, \beta_c, \beta_e, \mu, \gamma, \zeta_m, \tau_e \overset{\text{iid}}{\sim} N(0,1)$; $L \sim N^{-}(0,1)$. Estimated censoring threshold parameter L: mean -0.20, SD 0.00.

Table A.4: Posterior Parameter Estimates for Event, Group, and Member Characteristics (Utility Scale)

Variable	Mean	SE	SD	2.5% CI	50% CI	97.5% CI	ESS	$\hat{R}$
Group Characteristics (γ_g)								
Members	-0.91	0.01	0.16	-1.20	-0.91	-0.63	863	1
Is Pro	-0.14	0.01	0.20	-0.49	-0.14	0.20	546	1
Is Open	-0.28	0.01	0.15	-0.53	-0.28	-0.04	693	1
Event Characteristics (β_e)								
Avg. Digitization in Group	-0.74	0.00	0.14	-0.96	-0.73	-0.51	2220	1
Cap on RSVPs	-0.17	0.00	0.01	-0.18	-0.17	-0.15	8677	1
Recurring Event	-0.11	0.00	0.03	-0.16	-0.11	-0.07	7185	1
Venue is Listed	0.05	0.00	0.01	0.03	0.05	0.07	7108	1
Event Fee Charged	0.12	0.00	0.07	0.01	0.12	0.22	5238	1
Morning Event	0.20	0.00	0.01	0.18	0.20	0.21	7028	1
Members in Waitlist	0.38	0.00	0.01	0.37	0.38	0.39	6068	1
Event Description Length	0.50	0.00	0.04	0.44	0.50	0.56	8366	1
Member-Event Characteristics (μ_{ie})								
N. Positive RSVPs	-0.18	0.00	0.00	-0.18	-0.18	-0.17	7597	1
Tenure	0.04	0.00	0.01	0.03	0.04	0.06	9884	1
Time of Response	0.33	0.00	0.06	0.22	0.33	0.43	8075	1
Member-Group Characteristics (μ_{ig})								
Share Co-Attendees	-0.93	0.00	0.04	-0.99	-0.93	-0.87	9015	1
Past Events Exposure	-0.02	0.00	0.01	-0.04	-0.02	-0.00	10541	1
N.Co-Attendees	0.54	0.00	0.00	0.53	0.54	0.54	10351	1
Avg. Response Rate	1.10	0.00	0.02	1.10	1.10	1.20	6222	1

Note: Posterior distributions estimated using a Hamiltonian Monte Carlo algorithm. Posterior statistics calculated over 4 chains, 1000 iterations per chain. Specification estimated: $u_{ie} = D'_e * C'_g \beta_{DC} + C'_g \beta_c + X'_e \beta_e + X'_{ieg} \mu + \zeta_m + \tau_e + \eta_g + \epsilon_{ie}$; $Y_{ie} = 1$ if $u_{ie} > 0 > L$; $Y_{ie} 0$ if $L < u_{ie} \leq 0$; $Y_{ie} = -1$ if $u_{ie} \leq L$. Priors: $\eta_g \sim N(X'_g \gamma_g, 1)$; $\beta_{DC}, \beta_c, \beta_e, \mu, \gamma, \zeta_m, \tau_e \overset{\text{iid}}{\sim} N(0, 1)$; $L \sim N^{-}(0, 1)$. Estimated censoring threshold parameter L: mean -0.20, SD 0.00.

Chapter B

Appendix to Chapter 3

B.1 Variables Summaries

Table B.1: Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Treatment Variable				
Loss (Binary)	0.47	0.49	0	1
Community Activity				
Daily Contributions	25.02	40.07	1	2018
Daily Contributions (Adjusted)	1.09	1.05	0.009	97
Core-Periphery Activity				
Core Daily Contributions	7.10	13.36	0	484
Periphery Daily Contributions	11.93	14.93	0	1593
Daily Contributions from Newly Activated Members	5.28	11.5	0	516
Social Network Metrics				
Degree Centrality	7.08	10.78	0	193
Local Clustering Coefficient	0.55	0.44	0	1
N. Cohesive Blocks	18.25	23.53	1	160
Control Variables				
Opening Point Spread	4.73	8.22	-31	44
First Half of Season	0.47	0.5	0	1
Weekday v. Weekend	0.57	0.49	0	1
Top 25 AP Rank	0.14	0.35	0	1
Win Streak	1.34	2.08	0	22
Loss Streak	0.99	1.64	0	17

B.2 Additional Tables

B.2.1 Community Activity

Table B.2: Effect of Negative Events on Daily Contributions per Subreddit (Log Scale)

	Dependent variable: Daily Contributions (Log Scale)		
	(1)	(2)	(3)
Loss × Post-Game	−0.383	−0.379	−0.405
	(0.245)	(0.246)	(0.256)
Seasonality Controls	No	No	Yes
Team Popularity Controls	No	No	Yes
Predicted Point Spreads Control	No	No	Yes
Subreddit-month FE	No	Yes	Yes
Week-year FE	No	Yes	Yes
R^2	0.281	0.282	0.289
Num. obs.	297059	297059	297059

Robust standard errors clustered at the month-year level in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.
Estimating equation: $Y_{ct} = \beta_T T_t + \beta_D D_c + \delta D_c * T_t + \gamma X_c + \eta_{ct} + U_{ct}$. All specifications include subreddit-month and week-year fixed effects.
DV: New daily subreddit contributions (log scale). Treatment: Loss by focal team. Seasonality controls: first-half of season binary, weekend binary, number of cumulative losses in the season until game date. Team popularity controls: AP top-25 ranking binary. Predicted point spreads control: categorical point spread indicators – predicted draw, predicted close, clear predicted outcome.

Table B.3: Effect of Negative Events on Adjusted Daily Contributions per Subreddit (Log Scale)

	Dependent variable: Log(Adjusted Daily Contributions)		
	(1)	(2)	(3)
Loss × Post-Game	-0.0261^{***}	-0.0259^{***}	-0.0258^{***}
	(0.013)	(0.013)	(0.013)
Seasonality Controls	No	No	Yes
Team Popularity Controls	No	No	Yes
Predicted Point Spreads Control	No	No	Yes
Subreddit-month FE	No	Yes	Yes
Week-year FE	No	Yes	Yes
Adj. R^2	0.068	0.069	0.069
Num. obs.	297059	297059	297059

Robust standard errors clustered at the month-year level in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.
Estimating equation: $Y_{ct} = \beta_T T_t + \beta_D D_c + \delta D_c * T_t + \gamma X_c + \eta_{ct} + U_{ct}$. All specifications include subreddit-month and week-year fixed effects.
DV: New daily subreddit contributions divided by pre-game average contributions within subreddit. Treatment: Loss by focal team. Seasonality controls: first-half of season binary, weekend binary, number of cumulative losses in the season until game date. Team popularity controls: AP top-25 ranking binary. Predicted point spreads control: categorical point spread indicators – predicted draw, predicted close, clear predicted outcome.

B.2.2 Community Expectations

Table B.4: Negative Events and Community Activity – Disconfirmed vs. Confirmed Outcomes, 5-point Threshold

	Dependent Variable: Daily Contributions (Log 1+x)	
	Disconfirmed, ±5 points	Confirmed ±5 points
	(1)	(2)
Loss × Post-Game	-0.079^{***}	-0.061^{***}
	(0.019)	(0.017)
Loss	0.007	-0.006
	(0.017)	(0.013)
Post-Game Period	1.280^{***}	1.262^{***}
	(0.011)	(0.014)
Seasonality Controls	Yes	Yes
Team Popularity Controls	Yes	Yes
Predicted Point Spreads	No	No
Subreddit-month FE	Yes	Yes
Week-year FE	Yes	Yes
R^2	0.512	0.562
Num. obs.	93886	76697

Robust standard errors clustered at the month-year level in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$.
Estimating equation: $Y_{ct} = \beta_T T_t + \beta_D D_c + \delta D_c \times T_t + \gamma X_{ct} + \eta_{ct} + \tau_t + U_{ct}$. All specifications include subreddit-month and week-year fixed effects.
DV: New daily subreddit contributions (log 1+x scale). (1) Games for which predictions were disconfirmed within 5 points; (2) Games for which predictions were confirmed within 5 points. Treatment: Loss by focal team. Seasonality controls: first-half of season binary, weekend binary, number of cumulative losses in the season until game date. Team popularity controls: AP top-25 ranking binary. Predicted point spreads control: categorical point spread indicators – predicted draw, predicted close, clear predicted outcome.

Table B.5: Negative Events and Community Activity – Disconfirmed vs. Confirmed Outcomes

	Dependent Variable: Daily Contributions (Log 1+x)			
	Disconfirmed, ± 3 points (1)	Disconfirmed, ± 5 points (2)	Confirmed ± 3 points (3)	Confirmed ± 5 points (4)
Loss $\times$ Post-Game	-0.027	$-0.037^{\dagger}$	-0.018	-0.024
	(0.018)	(0.022)	(0.019)	(0.022)
Loss	-0.042^{**}	-0.011	-0.114^{***}	-0.073^{***}
	(0.016)	(0.020)	(0.014)	(0.017)
Post-Game Period	2.422^{***}	2.396^{***}	2.448^{***}	2.473^{***}
	(0.011)	(0.013)	(0.016)	(0.019)
Seasonality Controls	Yes	Yes	Yes	Yes
Team Popularity Controls	Yes	Yes	Yes	Yes
Predicted Point Spreads	Yes	No	Yes	No
Subreddit-month FE	Yes	Yes	Yes	Yes
Week-year FE	Yes	Yes	Yes	Yes
R^2	0.788	0.786	0.787	0.787
Num. obs.	118475	93886	100652	76697

Robust standard errors clustered at the month-year level in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$.
Estimating equation: $Y_{ct} = \beta_T T_t + \beta_D D_c + \delta D_c \times T_t + \gamma X_{ct} + \eta_{ct} + \tau_t + U_{ct}$. All specifications include subreddit-month and week-year fixed effects.
DV: New daily subreddit contributions (log 1+x scale). Games for which predictions were: (1) disconfirmed within 3 points; (2) disconfirmed within 5 points; (3) confirmed within 3 points; (4) confirmed within 5 points. Treatment: Loss by focal team. Seasonality controls: first-half of season binary, weekend binary, number of cumulative losses in the season until game date. Team popularity controls: AP top-25 ranking binary. Predicted point spreads control: categorical point spread indicators – predicted draw, predicted close, clear predicted outcome.

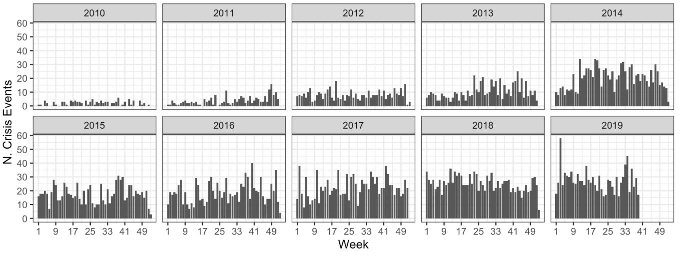
Chapter C

Appendix to Chapter 4

C.1 Brand Crisis Data

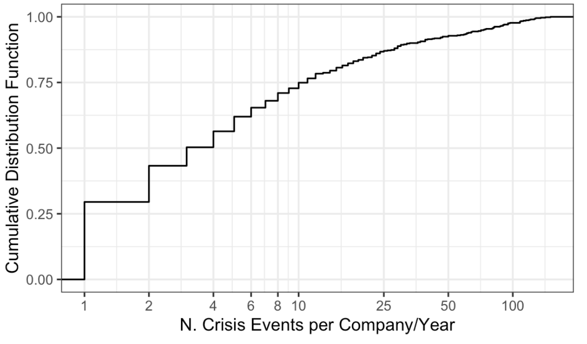
The available data include information on 7805 brand crises occurred between January 2010 and September 2019 (Figure C.1). Every year, most companies only face 1 crisis event. They represent about 29% of the observations in the estimation panels. The median number of crises per company and year is 4, and the average is 13.2 (Figure C.2).

Figure C.1: Number of Crisis Events by Week and Year



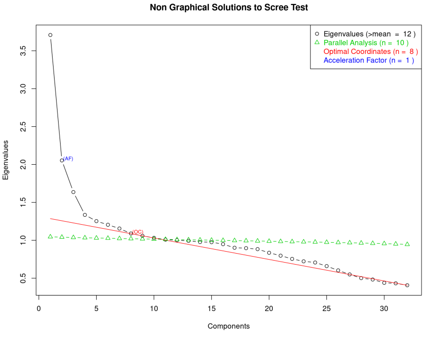
In the heterogeneity analyses, I aimed at measuring the differential impact of different types of crisis issues on the volume and structure of consumer discussions in the online brand communities. To better understand how the 32 original issue types correlate and co-occur with one another, I performed a maximum-likelihood (ML) exploratory factor analysis on the matrix of crisis dates, issues, and communities involved. The objective of the factor analysis was to reduce the dimensionality of

Figure C.2: Number of Crisis Events per Company and Year - Cumulative Distribution



the issue types, such that issues typically mentioned together in the RepRisk dataset would load on the same underlying factor. Figure C.3 shows the results of the ML exploratory factor analysis. The scree plot suggests that the last largest drop in eigenvalues is at 4 factors.

Figure C.3: Scree Plot for the Maximum-Likelihood Factor Analysis, Suggesting a 4 Factor Solution



I ran the factor analysis on the issue type indicators based on how frequently they are triggered together by a single crisis event. Table C.1 shows the issue type indicators, the predicted factor loadings, the factor corresponding to the highest loading, and a description of a possible crisis construct underlying each factor. Note that, while the scree plot solution suggests 4 factors, the last two crisis items ("Other issues" and "Not specified") do not load predominantly on any of the 4 factors.

Therefore, these issues can be considered as a fifth factor (named "Other"). As only 3 events fall under the fifth category, I do not take that category further into consideration.

Table C.1: Crisis Issue Types Factor Analysis - Solution with 4 Factors

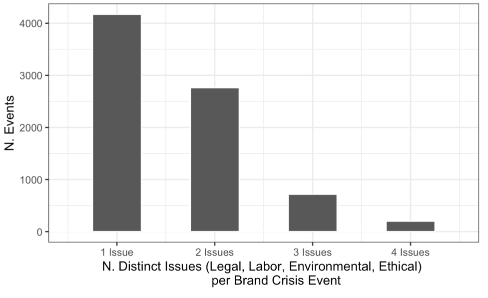
	Factor Loadings						
Crisis Issue Type	1	2	3	4	Factor	Factor Description	Impact on Consumers
Violation Of International Standards	0.24	0.02	0.07	0.16	1	Labor Issues	Indirect
Human Rights Abuses, Corporate Complicity	0.21	-0.02	-0.06	0.13	1	Labor Issues	Direct
Child Labor	0.54	0.18	-0.04	0.11	1	Labor Issues	Indirect
Forced Labor	0.47	0	0.02	0.09	1	Labor Issues	Direct
Freedom Of Association, Collective Bargaining	0.39	0	0.05	0.05	1	Labor Issues	Direct
Supply Chain Issues	0.61	0.25	-0.12	0.02	1	Labor Issues	Indirect
Executive Compensation Issues	0.01	-0.03	-0.07	-0.02	1	Labor Issues	Indirect
Occupational Health, Safety Issues	0.63	0.24	-0.03	-0.04	1	Labor Issues	Direct
Discrimination In Employment	0.13	0	0	-0.06	1	Labor Issues	Direct
Poor Employment Conditions	0.62	0.08	-0.04	-0.07	1	Labor Issues	Direct
Impacts On Landscapes, Ecosystems, Biodiversity	0.11	0.68	0.05	0.29	2	Local/Environmental Impact	Indirect
Impacts On Communities	0.07	0.58	0.05	0.29	2	Local/Environmental Impact	Direct
Local Pollution	0.13	0.71	0.05	0.12	2	Local/Environmental Impact	Indirect
Overuse, Wasting Of Resources	0.06	0.27	0	0.07	2	Local/Environmental Impact	Indirect
Products Health, Environmental Issues	0.02	0.05	-0.13	0.01	2	Local/Environmental Impact	Direct
Waste Issues	0.08	0.64	-0.03	-0.02	2	Local/Environmental Impact	Indirect
Other Environmental Issues	0.03	0	-0.02	-0.02	2	Local/Environmental Impact	Indirect
Tax Evasion	-0.03	0.02	0.25	0.06	3	Legal/Financial Issues	Indirect
Fraud	0.01	0.04	0.22	0.06	3	Legal/Financial Issues	Indirect
Tax Optimization	-0.09	-0.05	0.07	0.03	3	Legal/Financial Issues	Indirect
Anti Competitive Practices	-0.15	-0.04	0.38	-0.18	3	Legal/Financial Issues	Indirect
Violation Of National Legislation	0	0.01	0.91	-0.24	3	Legal/Financial Issues	Indirect
Climate Change, GHG Emissions, Global Pollution	0.05	0.21	0.03	0.43	4	Ethical Issues	Direct
Controversial Products/Services	-0.03	0.07	0.05	0.37	4	Ethical Issues	Direct
Misleading Communication	0.16	0.14	0.07	0.29	4	Ethical Issues	Indirect
Animal Mistreatment	0.02	0.05	0.01	0.2	4	Ethical Issues	Indirect
Corruption, Bribery, Extortion, Laundering	0.09	0.09	0.14	0.18	4	Ethical Issues	Direct
Local Participation Issues	0.08	0.17	0.06	0.18	4	Ethical Issues	Direct
Other Social Issues	0	0.05	0.01	0.1	4	Ethical Issues	Direct
Social Discrimination	-0.06	-0.04	-0.05	0.09	4	Ethical Issues	Indirect
Other Issues	-0.01	0	-0.02	0	5	Other	Indirect
Not Specified	0.01	0	0.01	-0.02	5	Other	Indirect

Using 4-factor solution for issue types, Figure C.4a describes the frequency with which multiple issues get triggered by a single crisis event. Notice that no event triggers all issues at the same time. Figure C.4b describes the distribution of crisis events by issue type.

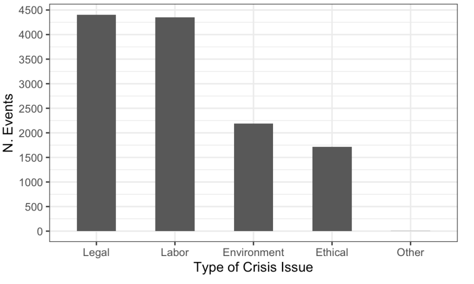
In addition to the 4-factor solution, I also classify the 32 types of crisis into "direct" versus "indirect" impact. A brand crisis has direct impact if it has the potential to affect customers directly (for example, product-harm crises that can physically damage consumers, or crises that involve forced labor and poor employment conditions). A brand crisis with indirect impact implies potential indirect harm for

Figure C.4: Distributions of Issue Types

(a) Number of Issues per Crisis Event



(b) Frequency of Issue Types (A Single Brand Crisis Can Involve Multiple Issues)



consumers (for example, a brand engaging in violation of national laws, issues with executive compensation, or crises involving waste of natural resources). Table C.1 summarises the classification of each issue type into direct versus indirect impact on the consumers. 3613 (46.3%) crisis events trigger issues that have both direct and indirect impact on the final consumer. 2318 (29.7%) events trigger issues with indirect impact, and 1875 (24%) events have a direct impact on the final consumer.

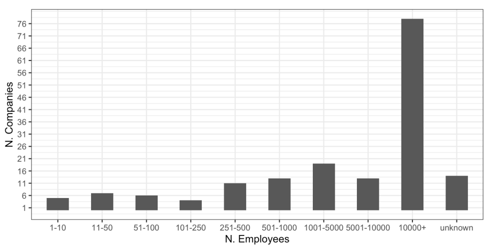
C.2 Company Data

I measure several important variables to describe the 154 companies included in the study. These characteristics are crucial to determine and address any sample selection concerns on observable company features, since company-level characteristics could determine whether a brand crisis receives media coverage at all (Backhaus & Fischer, 2016; Stäbler & Fischer, 2020).

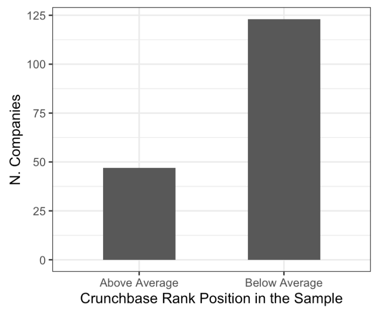
Figure C.5 shows that most companies in the estimation panels have a large employee base (10000 or more employees, Figure C.5a), and score below the in-sample average Crunchbase rank (Figure C.5b).

Figure C.5: Distributions of Company Characteristics

(a) Number of Companies by Employee Count



(b) Number of Companies by Rank



The most popular product category in the estimation panels is "Software and Computer Services", which includes companies and brands like Microsoft, Dropbox, and Android (see Table C.2 for the full list of brands and companies included in this study by main product category). Table C.2 also shows the most likely type of sector (Business to Business and/or Business to Consumer) for each of the companies included in the estimation sample.

Figure C.6: Distribution of Product Categories

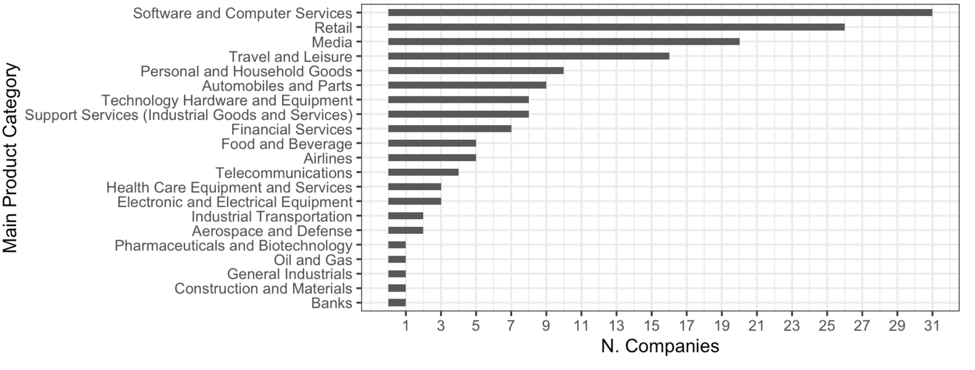


Table C.2: List of Companies by Main Product Category

Product Category	Company	B2C	B2B
Airlines	Alaska Air Group Inc	1	0
Airlines	Alaska Airlines Inc	1	0
Airlines	American Airlines Inc and Group Inc	1	0
Airlines	Delta Air Lines Inc	1	0
Automobiles	Audi of America LLC	1	0
Automobiles	Audi UK	1	0
Automobiles	General Motors Co, LLC and Ltd	1	1
Automobiles	Lexus	1	0
Automobiles	McLaren Group	0	1
Automobiles	Tesla Inc	1	0
Automobiles	Volkswagen UK	1	0
Banks	Bank of America Corp (BOA)	1	1
Construction and Materials	Eagle Inc	0	1
Electronic/Electrical Equipment	Alienware Corp	1	0
Electronic/Electrical Equipment	Dell Inc	1	1
Electronic/Electrical Equipment	Garmin Ltd	1	1
Financial Services	American Express Co and Bank Intl	1	0
Financial Services	Blackstone Group LP	0	1
Financial Services	Blockchain	1	0
Financial Services	HSBC Holdings PLC (HSBC)	1	0
Financial Services	Kickstarter Inc	1	0
Financial Services	Knights of Columbus Inc	1	0
Food and Beverage	Domino's Pizza Inc	1	0
Food and Beverage	General Mills Inc	1	0
Food and Beverage	Guinness Ltd	1	0
Food and Beverage	Nestle Holdings Inc	1	0
Food and Beverage	Nestle UK	1	0
General Industrials	Mauser Corp	0	1
Health Care	Centene Corp	0	1
Health Care	Cigna Corp (Cigna)	1	1
Health Care	Humana Inc	1	0
Industrial Transportation	DoorDash Inc	1	0
Industrial Transportation	FedEx Corp	1	0
Media	ABC Cable Networks Group Inc	0	1
Media	AdMob Inc	0	1
Media	American Broadcasting Company	0	1
Media	Archie Comic Publications Inc	1	1
Media	BBC Worldwide Ltd	0	1
Media	Bloomberg LP	1	1
Media	Comcast Corp	1	0
Media	ESPN Inc	1	1
Media	Facebook Inc	1	0
Media	Facebook UK Ltd	1	0
Media	Flickr Inc	1	0
Media	Gawker Media LLC	0	1
Media	Graham Holdings Co	0	1
Media	Home Box Office Inc (HBO)	1	0
Media	Hulu LLC	1	0
Media	Instagram Inc	1	0
Media	Last.fm Ltd	1	0
Media	LinkedIn Corp	1	0
Media	WhatsApp Inc	1	0
Media	Zynga Inc	1	0
Oil and Gas	Apache Corp	1	1
Personal/Household Goods	Beats Electronics LLC	1	0
Personal/Household Goods	Blizzard Entertainment Inc	1	0
Personal/Household Goods	Bose Corp (Bose)	1	0
Personal/Household Goods	Converse Inc	1	0
Personal/Household Goods	Dyson Ltd	1	0
Personal/Household Goods	Epic Games Inc	1	0
Personal/Household Goods	Fitbit Inc	1	0
Personal/Household Goods	Hasbro Inc	1	0
Personal/Household Goods	Jawbone Inc	1	0
Personal/Household Goods	Keurig Inc	1	0
Pharma and Biotechnology	23andMe Inc	1	0
Retail	7-Eleven Inc and Hawaii Inc	1	0
Retail	Amazon Web Services Inc (AWS)	1	0
Retail	Amazon.com Inc (Amazon)	1	0
Retail	Audible Inc	1	0
Retail	Best Buy Co Inc	1	0
Retail	Blue Apron Inc	1	0
Retail	Davidson College	1	0
Retail	Dollar Tree Inc	1	0
Retail	eBay Inc	1	1
Retail	Emory University	1	0
Retail	Etsy Inc	1	0
Retail	Forever 21 Inc and Retail	1	0
Retail	Georgetown University	1	0
Retail	Grinnell College	1	0
Retail	Groupon Inc	1	0
Retail	Hamilton College	1	0
Retail	John Lewis PLC	1	0
Retail	Kroger Co; The	1	0
Retail	Kwik Trip Inc	1	0
Retail	Macy's Inc	1	0
Retail	Marshalls Inc	1	0
Retail	Menards Inc	1	0
Retail	Nordstrom Inc	1	0
Retail	Northeastern University	1	0
Software/Computer	Adobe Systems Inc (Adobe)	1	0
Software/Computer	Alteryx Inc	0	1
Software/Computer	Android Inc	1	0
Software/Computer	AOL Inc	1	0
Software/Computer	Autodesk Inc	1	1
Software/Computer	BitTorrent Inc	1	0
Software/Computer	Bittrex Inc	1	0
Software/Computer	Booking Holdings Inc	1	0
Software/Computer	Coinbase Inc	1	0
Software/Computer	Dropbox Inc	1	0
Software/Computer	Duolingo Inc	1	0
Software/Computer	Epic Systems Corp	1	1
Software/Computer	Flipboard Inc	1	1
Software/Computer	Fortinet Inc	0	1
Software/Computer	GitHub Inc	1	0
Software/Computer	Google LLC (Google) and UK Ltd	1	1
Software/Computer	Grindr LLC	1	0
Software/Computer	LastPass	1	0
Software/Computer	Lyft Inc	1	0
Software/Computer	Magic Leap Inc	0	1
Software/Computer	Microsoft Corp	1	1
Software/Computer	MongoDB Inc	0	1
Software/Computer	Mozilla Corp	1	0
Software/Computer	Netcore Solutions LLC	0	1
Software/Computer	Quora Inc	1	0
Software/Computer	Reddit Inc	1	0
Software/Computer	Roblox Corp	1	0
Software/Computer	Snapchat Inc	1	0
Software/Computer	Tinder Inc	1	0
Software/Computer	Vimeo LLC	1	0
Support/Industrial Goods	Accenture Ltd and LLP	0	1
Support/Industrial Goods	Boston Consulting Group Inc (BCG)	0	1
Support/Industrial Goods	Carfax Inc	1	0
Support/Industrial Goods	Deloitte LLP	0	1
Support/Industrial Goods	Frostburg State University	1	0
Support/Industrial Goods	KPMG LLP (UK and USA)	0	1
Technology	Advanced Micro Devices Inc (AMD)	1	1
Technology	Apple Inc (Apple)	1	0
Technology	Apple UK	1	0
Technology	Dell Inc	1	1
Technology	F5 Networks Inc	0	1
Technology	Intel Corp	1	1
Technology	Motorola Inc	1	1
Technology	Motorola Solutions Inc	1	1
Telecommunications	AT&T Communications Inc	1	1
Telecommunications	AT&T Corp and AT&T Inc	1	1
Telecommunications	Avaya Inc	0	1
Travel and Leisure	Airbnb Inc	1	0
Travel and Leisure	Cedar Fair LP	1	1
Travel and Leisure	Chili's Inc	1	0
Travel and Leisure	Cleveland Cavaliers	1	1
Travel and Leisure	Costa Ltd	1	1
Travel and Leisure	Dallas Mavericks Inc	1	1
Travel and Leisure	Del Taco LLC	1	1
Travel and Leisure	Everton Football Club	1	1
Travel and Leisure	Indianapolis Colts Inc	1	1
Travel and Leisure	Kentucky Fried Chicken	1	0
Travel and Leisure	Los Angeles Lakers	1	1
Travel and Leisure	Major League Baseball (MLB)	1	1
Travel and Leisure	McDonald's Corp	1	0
Travel and Leisure	Miami Dolphins Ltd Inc	1	1
Travel and Leisure	Taco Bell Corp	1	0
Travel and Leisure	TripAdvisor Inc	1	0

In Table C.3, I map each product category into the most likely type of sector (Business to Business or Business to Consumer), and to a larger "macro-category" that can best describe the main product category for each company.

Table C.3: List of Product Categories

Main Product Category	(Most Likely) Sector Type	MacroCategory	N. Companies	N. Crises
Aerospace and Defense	B2B	Services	2	64
Airlines	B2C	Travel	5	201
Automobiles and Parts	B2C	Travel	9	762
Banks	B2C	Finance	1	443
Construction and Materials	B2B	Services	1	3
Electronic and Electrical Equipment	B2C	Tech	3	111
Financial Services	B2C	Finance	7	111
Food and Beverage	B2C	Consumer Goods	5	100
General Industrials	B2B	Services	1	1
Health Care Equipment and Services	B2C	Health	3	23
Industrial Transportation	B2B	Travel	2	72
Media	B2C	Information	20	845
Oil and Gas	B2C	Services	1	71
Personal and Household Goods	B2C	Consumer Goods	10	62
Pharmaceuticals and Biotechnology	B2C	Health	1	6
Retail	B2C	Consumer Goods	26	1237
Software and Computer Services	B2C	Tech	31	1565
Support Services (Industrial Goods and Services)	B2B	Services	8	64
Technology Hardware and Equipment	B2C	Tech	8	1269
Telecommunications	B2C	Information	4	255
Travel and Leisure	B2C	Travel	16	695

To understand the extent to which similar companies engage in similar corporate misbehavior, Table C.4 shows the average levels of crisis severity, reach, and novelty by company size. The table suggests that larger or smaller companies sizes do not noticeably engage in more or less severe corporate misbehavior. On the other hand, it appears that larger companies engage in brand crises more frequently than smaller ones: companies with more than 5,000 employees score the lowest average score for novelty in media coverage.

Table C.4: Average Crisis Characteristics by Company Size.

N. Employees	Avg. Severity	Avg. Reach	Avg. Novelty
1-10	1.40	1.58	1.84
11-50	1.00	1.88	2.00
51-100	1.30	2.36	1.85
101-250	1.11	2.03	2.00
251-500	1.02	2.20	1.58
501-1000	1.01	2.30	1.80
1001-5000	1.11	2.04	1.57
5001-10000	1.31	1.84	1.81
10000+	1.23	2.13	1.41
unknown	1.16	2.39	1.45

Note: Severity levels as reported by the RepRisk SGC dataset: 1 = low, 2= medium, 3 = high; News outlets' reach levels as reported by the RepRisk SGC dataset: 1 = low reach, 2= medium reach, 3 = high reach; Novelty levels as reported by the RepRisk SGC dataset: 1 = not first company offense, 2 = first company offense.

C.3 Community Data

The estimation panels include data on 300 brand communities on Reddit.com. Table C.5 shows that most communities are organized around companies in the sectors "software and computer services", "retail", and "media". This statistics reflects the fact that most companies in the sample operate in these product categories.

Table C.5: Number of Communities per Product Category

Product Category	N. Communities
Software and Computer Services	77
Retail	39
Media	36
Technology Hardware and Equipment	35
Travel and Leisure	27
Personal and Household Goods	23
Automobiles and Parts	13
Financial Services	13
Support Services (Industrial Goods and Services)	8
Electronic and Electrical Equipment	4
Food and Beverage	4
Health Care Equipment and Services	4
Industrial Transportation	4
Airlines	3
Pharmaceuticals and Biotechnology	3
Telecommunications	2
Aerospace and Defense	1
Banks	1
Construction and Materials	1
General Industrials	1
Oil and Gas	1

Figure C.7 shows that on average, there is a change in the levels of consumer activity recorded in the communities around and after the date of a brand crisis. In particular, on average, community activity decreases about a week before the event is covered by the news – suggesting a lag between the event unfolding and the news coverage – sharply increases for the 4 weeks following the event with a peak at 3 weeks, and then declines around weeks 5 and 6.

In terms of social network metrics, Figure C.8 and Table C.6 show that all the metrics related to speed and ease of information spread, on average, increase in the weeks following the news of a brand crisis. Figure C.8 clarifies that the average weekly clustering coefficient and degree centrality follow the pattern of average weekly activity – a decrease at week -1, and an increase throughout weeks $0-4$.

Figure C.7: Weekly Community Contributions – controlling for community and week f.e.

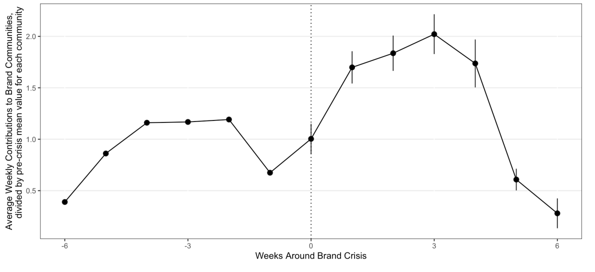
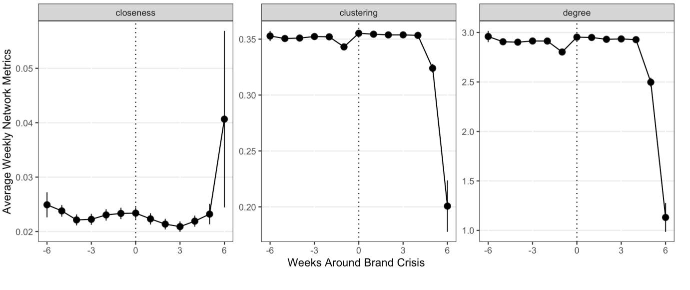


Table C.6: Average Network Metrics Before and After News of Brand Crisis.

	Pre-Crisis		Post-Crisis	
Metrics	Mean	SD	Mean	SD
Clustering	0.354	0.218	0.357	0.217
Degree	2.927	2.668	2.950	2.680
Closeness	0.022	0.104	0.021	0.101

Figure C.8: Weekly Community Network Metrics – controlling for community and week f.e.



C.4 Estimation Sample: Summary Statistics

Table C.7: Summary Statistics

Variable	N.Obs.	Mean	SD	Min.	Max.
Weekly Contributions per Subreddit	429387	356.829	400.514	0	1508
Crisis Occurrence Indicator ($\in \{0,1\}$)	429387	0.504	0.500	0	1
Crisis Characteristics					
High Severity ($\in \{0,1\}$)	429387	0.012	0.110	0	1
High Reach ($\in \{0,1\}$)	429387	0.408	0.491	0	1
Novelty ($\in \{1, 2\}$)	429387	1.557	0.497	1	2
N. Issues per Crisis	429387	2.737	2.106	1	21
N. Countries Affected per Crisis	429387	2.041	3.787	1	80
Crisis Impact and Type ($\in \{0, 1\}$)					
Direct	429387	0.672	0.469	0	1
Indirect	429387	0.796	0.403	0	1
Legal	429387	0.610	0.488	0	1
Labor	429387	0.593	0.491	0	1
Environment	429387	0.226	0.418	0	1
Ethical	429387	0.173	0.378	0	1
Other	429387	0.001	0.023	0	1
Company Characteristics					
Received Funding ($\in \{0, 1\}$)	429387	0.890	0.313	0	1
Crunchbase Rank (In-Sample Normalized)	429387	-0.002	1.019	-0.178	22.226
N. Crises per Company	429387	597.873	277.461	1	929
Company Type ($\in \{0, 1\}$)					
B2C	429387	0.904	0.295	0	1
B2B	429387	0.271	0.444	0	1
Product Categories ($\in \{0, 1\}$)					
Consumer Goods	429387	0.165	0.371	0	1
Health	429387	0.001	0.032	0	1
Information	429387	0.071	0.257	0	1
Services	429387	0.004	0.065	0	1
Tech	429387	0.665	0.472	0	1
Travel	429387	0.078	0.268	0	1
Finance	429387	0.015	0.120	0	1

C.5 User-Generated Content

Table C.8: Brand Crises and User-Generated Content. Including crisis, company, and product category controls, and company-month and week-of-month fixed effects.

	Weekly Average % of Words per Contribution					
Outcome Variable	Brand Crisis Estimate	Robust Std.Err.	T-value	P-value	R^2	N. Obs.
Weekly Average Word Count (Log1p)						
Wordcount	0.13542	0.01134	11.94195	p < .001	0.88324	428968
Positive and Negative Emotion						
Positive Emotion	-0.00002	0.00005	-0.40650	p = 0.684	0.09329	428968
Negative Emotion	0.00004	0.00001	2.79040	p = 0.005	0.29355	428968
Conflict	0.00003	0.00001	2.95787	p = 0.003	0.20172	428968
In-Group v. Out-Group Expressions						
We Pronoun	0.00003	0.00001	1.99403	p = 0.046	0.30980	428968
Third Person Pronouns	0.00007	0.00003	2.53687	p = 0.011	0.58095	428968
Cognitive Processes						
Cognitive Processes	0.00143	0.00016	8.93875	p < .001	0.88273	428968
Time Orientation						
Future Focus	0.00012	0.00003	4.51203	p < .001	0.65542	428968
Present Focus	0.00046	0.00007	7.00242	p < .001	0.83773	428968
Past Focus	0.00042	0.00005	7.77473	p < .001	0.75177	428968

Note: the coefficients are estimated with a multivariate OLS model. Robust standard errors clustered at the level of the product category and week are in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Specification tested: $Y_{bt} = \mathbb{I}(t > T_s)_{bt}\delta + X_{bt}\beta_1 + X_b\beta_2 + \gamma_{b,m(t))} + \zeta_t + \varepsilon_{bt}$. *Treatment variable* $\mathbb{I}(t > T_s)_{bt}$: Brand crisis occurrence indicator. *Outcome variables* Y_{bt}: Average weekly share of words (%) per contribution. *Crisis control variables* X_{bt}: crisis severity, news novelty, media reach, number of countries affected, number of issues raised by the crisis, type of crisis issue. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset. *Product category*: main product macro-category in which the company operates. *Fixed effects* $\gamma_{b,m(t)}, \zeta_t$: company-month and week of month fixed effects.

Table C.9: Brand Crises and User-Generated Content by Member Type. Including crisis, company, and product category controls, and company-month and week-of-month fixed effects.

	Weekly Average % of Words per Contribution				
Outcome	Brand Crisis	H-Type	Brand Crisis × H-Type	R^2	N. Obs.
Word Count (Count, Log1p)					
Wordcount	-1.2137, p < .001	0.51363, p < .001	0.43211, p < .001	0.83389	857936
Positive and Negative Emotion					
Positive Emotion	-0.00073, p < .001	-5e-04, p < .001	0.00026, p = 0.077	0.10148	857936
Negative Emotion	-6e-05, p = 0.179	-3e-05, p = 0.552	-2e-04, p = 0.004	0.13936	857936
Conflict	1e-05, p = 0.727	0.00012, p < .001	-2e-05, p = 0.657	0.12972	857936
In-Group v. Out-Group Expressions					
We Pronoun	-0.00033, p < .001	0.00026, p < .001	0.00021, p < .001	0.21179	857936
Third Person Pronouns	-0.00036, p < .001	0.00025, p < .001	6e-05, p = 0.516	0.48599	857936
Cognitive Processes					
Cognitive Processes	-0.00885, p < .001	0.00211, p < .001	0.0037, p < .001	0.81998	857936
Time Orientation					
Future Focus	-9e-05, p = 0.216	0.00098, p < .001	-0.00031, p = 0.001	0.51447	857936
Present Focus	-0.00343, p < .001	-0.00044, p < .001	0.00181, p < .001	0.74683	857936
Past Focus	-0.00392, p < .001	-0.00207, p < .001	0.0012, p < .001	0.64439	857936

Note: the coefficients are estimated with a multivariate OLS model. Robust standard errors clustered at the level of the product category and week are in parenthesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$. Specification tested: $Y_{ibt} = \mathbb{I}(i = \mathrm{H})_{ibt}\beta_1 + \mathbb{I}(t > T_s)_{bt}\beta_2 + \mathbb{I}(t > T_s)_{bt} \times \mathbb{I}(i = \mathrm{H})_{ibt}\delta + X_{bt}\beta_3 + X_b\beta_4 + \gamma_{b,m(t))} + \zeta_t + \varepsilon_{ibt}$. *Treatment variable* $\mathbb{I}(t > T_s)_{bt}$: Brand crisis occurrence indicator. *Outcome variable* Y_{bt}: Average weekly share of words (%) per contribution. *Moderator* $\mathbb{I}(i = H)_{ibt}$: Type of member indicator: {1=H-type; 0=L-type}, based on above vs. below-average pre-crisis activity level. *Crisis control variables* X_{bt}: crisis severity, news novelty, media reach, number of countries affected, number of issues raised by the crisis, type of crisis issue. *Company control variables* X_b: Crunchbase rank, indicator for reception of funding, number of crises in the dataset. *Product category*: main product macro-category in which the company operates. *Fixed effects* $\gamma_{b,m(t)}, \zeta_t$: company-month and week of month fixed effects.

References

6Connect. (2020, 09). *Virtual Events vs. Physical Events: The Definitive Pro-Con List.* Retrieved from https://www.6connex.com/blog/virtual-events-vs-physical-events-pro-con-list/

Aaker, J., Fournier, S., & Brasel, S. A. (2004). When good brands do bad. *Journal of Consumer research*, *31*(1), 1–16.

Aggarwal, P. (2004). The effects of brand relationship norms on consumer attitudes and behavior. *Journal of consumer research*, *31*(1), 87–101.

Ahluwalia, R., Burnkrant, R. E., & Unnava, H. R. (2000). Consumer response to negative publicity: The moderating role of commitment. *Journal of marketing research*, *37*(2), 203–214.

Alexa Internet Ranking. (2021). *Reddit.com Competitive Analysis, Marketing Mix and Traffic - Alexa.* Retrieved from https://web.archive.org/web/20210419124707/https://www.alexa.com/siteinfo/reddit.com

Algesheimer, R., Borle, S., Dholakia, U. M., & Singh, S. S. (2010). The impact of customer community participation on customer behaviors: An empirical investigation. *Marketing science*, *29*(4), 756-769.

Ardichvili, A., Vaugh, P., & Wentling, T. (2003). Motivation and barriers to participation in virtual knowledge-sharing communities of practice. *Journal of Knowledge Management*, *7*(1), 64-77.

Armstrong, A., & Hagel, J. (2000). The real value of online communities. *Knowledge and communities*, *74*(3), 85–95.

Assmann, J., Sandner, P., & Ahrens, S. (2009). Users' influence on the success of online communities. In *2009 42nd hawaii international conference on system sciences* (pp. 1–10).

Associated Press. (n.d.). *AP Top 25 Poll.* Retrieved from https://collegebasketball.ap.org/poll/

Athey, S., & Wager, S. (2019). Estimating treatment effects with causal forests: An application. *arXiv preprint arXiv:1902.07409*.

Backhaus, M., & Fischer, M. (2016). Brand damage from product harm and corporate social irresponsibility: how deep and how long. *Marketing Science Institute Working Paper Series*, *16*(133), 1–63.

Bampo, M., Ewing, M. T., Mather, D. R., Stewart, D., & Wallace, M. (2008). The effects of the social structure of digital networks on viral marketing performance. *Information systems research*, *19*(3), 273–290.

Banerjee, A. V., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2018). Changes in social network structure in response to exposure to formal credit markets.

Available at SSRN 3245656.

Barberá, P., Wang, N., Bonneau, R., Jost, J. T., Nagler, J., Tucker, J., & González-Bailón, S. (2015). The critical periphery in the growth of social protests. *PloS one*, *10*(11).

Barger, V., Peltier, J. W., & Schultz, D. E. (2016). Social media and consumer engagement: a review and research agenda. *Journal of Research in Interactive Marketing*, *10*(4), 268-287.

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). The pushshift reddit dataset. In *Proceedings of the international aaai conference on web and social media* (Vol. 14, pp. 830–839).

Berger, J., Sorensen, A. T., & Rasmussen, S. J. (2010). Positive effects of negative publicity: When negative reviews increase sales. *Marketing science*, *29*(5), 815–827.

Bevy, C. (2021). *The 2021 community industry report* (Tech. Rep.). Retrieved from https://cmxhub.com/community-industry-report-2021/

Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social networks,*, *21*(4), 375-395.

Bramoulle, Y., & Kranton, R. (2005). *Strategic experimentation in networks.* (NajEcon Working Paper no. 784828000000000417 from www. najecon. org.)

Brodie, R. J., Ilic, A., Juric, B., & Hollebeek, L. (2013). Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of business research*, *66*(1), 105-114.

Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of interactive marketing*, *21*(3), 2–20.

Burnap, P., Williams, M. L., Sloan, L., Rana, O., Housley, W., Edwards, A., ... Voss, A. (2014). Tweeting the terror: modelling the social media reaction to the woolwich terrorist attack. *Social Network Analysis and Mining*, *4*(1), 206.

Bussgang, J., & Bacon, J. (2020, January). When community becomes your competitive advantage. *Harvard Business Review*.

Card, D., & Dahl, G. (2009). *Family violence and football: The effect of unexpected emotional cues on violent behavior* (Tech. Rep.). National Bureau of Economic Research.

Card, D., & Dahl, G. B. (2011). Family violence and football: The effect of unexpected emotional cues on violent behavior. *The Quarterly Journal of Economics*, *126*(1), 103-143.

Cheng, S. Y., White, T. B., & Chaplin, L. N. (2012). The effects of self-brand connections on responses to brand failure: A new look at the consumer–brand relationship. *Journal of Consumer Psychology*, *22*(2), 280–288.

Cleeren, K., Van Heerde, H. J., & Dekimpe, M. G. (2013). Rising from the ashes: How brands and categories can overcome product-harm crises. *Journal of Marketing*, *77*(2), 58–77.

Cohn, A., Gesche, T., & Maréchal, M. A. (2018). Honesty in the digital age.

Crunchbase. (2021). *DonBest - Crunchbase Investor Profile Investments.* Retrieved from https://www.crunchbase.com/organization/donbest

Crunchbase Rank. (2019, 02). *What Is Crunchbase Rank and Trend Score?* Retrieved from https://about.crunchbase.com/blog/crunchbase-rank-trend-score/

Dawar, N., & Pillutla, M. M. (2000). Impact of product-harm crises on brand equity: The moderating role of consumer expectations. *Journal of marketing research*, *37*(2), 215–226.

De Matos, C. A., Henrique, J. L., & Alberto Vargas Rossi, C. (2007). Service recovery paradox: a meta-analysis. *Journal of service research*, *10*(1), 60–77.

De Souza, C. S., & Preece, J. . (2004). A framework for analyzing and understanding online communities. *Interacting with computers*, *16*(3), 579-610.

Dessart, L., Veloutsou, C., & Morgan-Thomas, A. (2015). Consumer engagement in online brand communities: a social media perspective. *Journal of Product and Brand Management*, *24*(1).

De Valck, K., Van Bruggen, G., & Wierenga, B. (2009). Virtual communities: A marketing perspective. *Decision Support Systems*, *47*(3), 185–203.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PloS one*, *6*(12), e26752.

Dondio, P., & Usher, J. (2017). Analysing the behaviour of online investors in times of geopolitical distress: a case study on war stocks. In *Proceedings of the international conference on web intelligence* (pp. 275–283).

Dover, Y., & Kelman, G. (2018). Emergence of online communities: Empirical evidence and theory. *PloS one*, *13*(11).

Dutta-Bergman, M. J. (2005). The antecedents of community-oriented internet use: Community participation and community satisfaction. *Journal of Computer-Mediated Communication*, *11*(1), 97–113.

Easley, D., Kleinberg, J., et al. (2010). *Networks, crowds, and markets* (Vol. 8). Cambridge university press Cambridge.

Eismann, K., Posegga, O., & Fischbach, K. (2016). Collective behaviour, social media, and disasters: A systematic literature review.

El-Sayed, A. M., Seemann, L., Scarborough, P., & Galea, S. (2013). Are network-based interventions a useful antiobesity strategy? an application of simulation models for causal inference in epidemiology. *American journal of epidemiology*, *178*(2), 287–295.

Faraj, S., Kudaravalli, S., & Wasko, M. (2015). Leading collaboration in online communities. *MIS quarterly*, *39*(2), 393–412.

Farzan, R., Lu, D., & Lin, Y. R. (2016, July). *What happens offline stays offline?: examining sustainability of a hybrid social system* (Vol. Proceedings of the 27th ACM Conference on Hypertext and Social Media (pp. 261-266)).

Festinger, L. (1957). *A theory of cognitive dissonance* (Vol. 2). Stanford university press.

Fischer, D. (2018). Social networking sites in the aftermath of a crisis-the enabling role for self-organization. In *Proceedings of the 51st hawaii international conference on system sciences.*

Fletcher, L., & Aliaj, O. (2021). Reddit led market turmoil hits big quant hedge funds. *Financial Times*.

Forrest, R., & Kearns, A. (2001). Social cohesion, social capital and the neighbourhood. *Urban studies*, *38*(12), 2125–2143.

Forsyth, D. R., & Burnette, J. (2010). Group processes.

Fournier, S., & Lee, L. (2009). Getting brand communities right. *Harvard business review*, *87*(4).

Granovetter, M. (2018). Economic action and social structure: The problem of embeddedness. In *The sociology of economic life* (pp. 22–45). Routledge.

Grégoire, Y., Tripp, T. M., & Legoux, R. (2009). When customer love turns into lasting hate: The effects of relationship strength and time on customer revenge and avoidance. *Journal of marketing*, *73*(6), 18–32.

Harmon-Jones, E., & Mills, J. (2019). An introduction to cognitive dissonance theory and an overview of current perspectives on the theory.

Heilweil, R., & Ghaffary, S. (2021, 01). *How Trump supporters planned the Capitol riot online.* Retrieved from https://www.vox.com/recode/22221285/trump-online-capitol-riot-far-right-parler-twitter-facebook

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *Journal of interactive marketing*, *18*(1), 38–52.

Hsu, L., & Lawrence, B. (2016). The role of social media and brand equity during a product recall crisis: A shareholder value perspective. *International journal of research in Marketing*, *33*(1), 59–77.

Huffaker, D. (2010). Dimensions of leadership and social influence in online communities. *Human Communication Research*, *36*(4), 593–617.

Hunt, K., Bristol, T., & Bashaw, R. (1999). A conceptual approach to classifying sports fans. *Journal of Services Marketing*, *13*(6), 439–452.

Hurlbert, J. S., Haines, V. A., & Beggs, J. J. (2000). Core networks and tie activation: What kinds of routine networks allocate resources in nonroutine situations? *American Sociological Review*, 598–618.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.

Iyer, S., Cheng, J., Brown, N., & Wang, X. (2020). When does trust in online social groups grow? In *Proceedings of the international aaai conference on web and social media* (Vol. 14, pp. 283–293).

Jalili, M., & Perc, M. (2017). Information cascades in complex networks. *Journal of Complex Networks*, *5*(5), 665–693.

Jasper, J. M., & Poulsen, J. D. (1995). Recruiting strangers and friends: Moral shocks and social networks in animal rights and anti-nuclear protests. *Social problems*, *42*(4), 493–512.

Jenson, J. (1998). *Mapping social cohesion: The state of canadian research.* Canadian policy research networks Ottawa.

Jung, K., & Park, H. W. (2014). Citizens' social media use and homeland security information policy: Some evidences from twitter users during the 2013 north korea nuclear test. *Government Information Quarterly*, *31*(4), 563–573.

Kalra, A., Li, S., & Zhang, W. (2011). Understanding responses to contradictory information about products. *Marketing Science*, *30*(6), 1098–1114.

Kang, J., Tang, L., & Fiore, A. M. (2014a). Enhancing consumer–brand relationships on restaurant facebook fan pages: Maximizing consumer benefits and increasing active participation. *International Journal of Hospitality Management*, *36*, 145-155.

Kang, J., Tang, L., & Fiore, A. M. (2014b). Enhancing consumer–brand relationships on restaurant facebook fan pages: Maximizing consumer benefits and increasing active participation. *International Journal of Hospitality Management,*, *36*, 145-155.

Khamitov, M., Grégoire, Y., & Suri, A. (2020). A systematic review of brand transgression, service failure recovery and product-harm crisis: integration and guiding insights. *Journal of the Academy of Marketing Science*, *48*(3), 519–542.

Klein, J., & Dawar, N. (2004). Corporate social responsibility and consumers' attributions and brand evaluations in a product–harm crisis. *International Journal* *of research in Marketing*, *21*(3), 203–217.

Koh, J., Kim, Y. G., Butler, B., & Bock, G. W. (2007). Encouraging participation in virtual communities. *Communications of the ACM*, *50*(2), 68-73.

Kozinets, R. V. (1999). E-tribalized marketing?: The strategic implications of virtual communities of consumption. *European management journal*, *17*(3), 252–264.

Kuchmaner, C. A., Wiggins, J., & Grimm, P. E. (2019). The role of network embeddedness and psychological ownership in consumer responses to brand transgressions. *Journal of Interactive Marketing*, *47*, 129–143.

Kumar, M., & Bhalla, S. (2021). Global sports market today: An overview. *International Journal of Physical Education, Sports and Health*, *8*(4).

Li, Y. (2021, 01). *GameStop mania explained: How the Reddit retail trading crowd ran over Wall Street pros.* Retrieved from https://www.cnbc.com/2021/01/27/gamestop-mania-explained-how-the-reddit-retail-trading-crowd-ran-over-wall-street-pros.html

Lin, H. F. (2007). The role of online and offline features in sustaining virtual communities: an empirical study. *Internet Research*, *17*(2), 119-138.

Lin, N. (2002). *Social capital: A theory of social structure and action* (No. 19). Cambridge university press.

Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., ... others (2005). Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, *10*(4), 00–00.

Lu, Y., & Yang, D. (2011). Information exchange in virtual communities under extreme disaster conditions. *Decision Support Systems*, *50*(2), 529–538.

Luo, X. (2009). Quantifying the long-term impact of negative word of mouth on cash flows and stock prices. *Marketing Science*, *28*(1), 148–165.

Lyons, K. (2021, 01). *GameStop stock halts trading after Reddit drama.* Retrieved from https://web.archive.org/web/20210124184432/https://www.theverge.com/2021/1/22/22244900/game-stop-stock-halted-trading-volatility

Ma, A., & Mondragón, R. J. (2015). Rich-cores in networks. *PloS one*, *10*(3), e0119678.

Mahony, D., Nakazawa, M., Funk, D., James, J., & Gladden, J. (2002). Motivational factors influencing the behavior of j. league spectators. *Sport Management Review*, *5*(1), 1–24.

Manchanda, P., Packard, G., & Pattabhiramaiah, A. (2015). Social dollars: The economic impact of customer participation in a firm-sponsoredonline customer community. *Marketing Science*, *34*(3), 367-387.

Marwell, G., & Oliver, P. (1993). *The critical mass in collective action.* Cambridge University Press.

Marwell, G., Oliver, P. E., & Prahl, R. (1988). Social networks and collective action: A theory of the critical mass. iii. *American journal of Sociology*, *94*(3), 502–534.

McGrath, J. E. (1984). *Groups: Interaction and performance* (Vol. 14). Prentice-Hall Englewood Cliffs, NJ.

Meetup. (2020, 03). *Meetup Spins Off From WeWork.* Retrieved from https://www.prnewswire.com/news-releases/meetup-spins-off-from-wework-301031766.html

Millington, R. (2021, 05). The Real Value of Your Brand Community. *Harvard Business Review*. Retrieved from https://hbr.org/2021/05/the-real-value-of-your-brand-community

Mochon, D., Johnson, K., Schwartz, J., & Ariely, D. (2017). What are likes worth? a facebook page field experiment. *Journal of Marketing Research*, *54*(2), 306-317.

Muniz, A. M., & O'Guinn, T. C. (2001). Brand community. *Journal of consumer research*, *27*(4), 412-432.

Nelson, A. (2020, 02). *FIVB impresses in BCW's international federations social media study.* Retrieved from https://www.sportbusiness.com/news/fivb-impresses-in-bcws-international-sports-federations-social-media-ranking-2019/

Park, H., & Cho, H. (2012). Social network online communities: information sources for apparel shopping. *Journal of Consumer Marketing*.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of liwc2015* (Tech. Rep.).

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count.* Mahwah, NJ: Lawerence Erlbaum Associates.

Preece, J. (2000). *Online communities: Designing useability, supporting sociability.* Chichester: John Wiley and Sons.

Preece, J. (2001). Sociability and usability in online communities: Determining and measuring success. *Behaviour & Information Technology*, *20*(5), 347–356.

Qu, Y., Wu, P. F., & Wang, X. (2009). Online community response to major disaster: A study of tianya forum in the 2008 sichuan earthquake. In *2009 42nd hawaii international conference on system sciences* (pp. 1–11).

Racca, P., Casarin, R., Squazzoni, F., & Dondio, P. (2016). Resilience of an online financial community to market uncertainty shocks during the recent financial crisis. *Journal of computational science*, *16*, 190-199.

Rasmussen, J., & Ihlen, Ø. (2017). Risk, crisis, and social media: A systematic review of seven years' research. *Nordicom Review*, *38*(2), 1–17.

Reddit Inc. (2021). *Press - Reddit.* Retrieved from https://web.archive.org/web/20210416025129/https://www.redditinc.com/press/

RepRisk AG. (n.d.). *RepRisk's ESG Platform.* Retrieved from https://www.reprisk.com/solutions#the-esg-risk-platform

Robitzsch, A., & Grund, S. (2021). miceadds: Some additional multiple imputation functions, especially for 'mice' [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=miceadds (R package version 3.11-6)

Roehm, M. L., & Brady, M. K. (2007). Consumer responses to performance failures by high-equity brands. *Journal of Consumer Research*, *34*(4), 537–545.

Romero, D. M., & Kleinberg, J. (2010). The directed closure process in hybrid social-information networks, with an analysis of link formation on twitter. *arXiv preprint arXiv:1003.2469*.

Romero, D. M., Uzzi, B., & Kleinberg, J. (2016, April). *Social networks under stress* (No. 25). International Conference on World Wide Web.

Rooderkerk, R. P., Van Heerde, H. J., & Bijmolt, T. H. (2013). Optimizing retail assortments. *Marketing Science*, *32*(5), 699–715.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Rothaermel, F. T., & Sugiyama, S. (2001). Virtual internet communities and commercial success: individual and community-level theory grounded in the atypical case of timezone. com. *Journal of management*, *27*(3), 297–312.

Schaefers, T., & Schamari, J. (2016). Service recovery via social media: The social influence effects of virtual presence. *Journal of Service Research*, *19*(2), 192–208.

Sharma, P., Sadh, A., Billore, A., & Motiani, M. (2021). Investigating brand community engagement and evangelistic tendencies on social media. *Journal of Product & Brand Management*.

Sherman, B. (2020). *Zoom sees sales boom amid pandemic.* Retrieved from https://www.bbc.com/news/business-52884782

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications.* Toronto; London; New York: Wiley.

Spinks, D. (2020, 03). *Tools you can use to host virtual conferences and events.* Retrieved from https://web.archive.org/web/20200422202751/https://twitter.com/DavidSpinks/status/1235993594705534976

Staats, B. R., Kc, D. S., & Gino, F. (2018). Maintaining beliefs in the face of negative news: The moderating role of experience. *Management Science*, *64*(2), 804–824.

Stäbler, S., & Fischer, M. (2020). When does corporate social irresponsibility become news? evidence from more than 1,000 brand transgressions across five countries. *Journal of Marketing*, *84*(3), 46–67.

Stam, W. (2009). When does community participation enhance the performance of open source software companies? *Research Policy*, *38*(8), 1288-1299.

Stan Development Team. (2021). *Stan Modeling Language Users Guide and Reference Manual, 2.27.* Retrieved from https://mc-stan.org/

Tajfel, H. E. (1978). *Differentiation between social groups: Studies in the social psychology of intergroup relations.* Academic Press.

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., & Gelman, A. (2018). Validating bayesian inference algorithms with simulation-based calibration. *arXiv preprint arXiv:1804.06788*.

The CMO Survey. (2021, 02). *CMO Survey.* Retrieved from https://cmosurvey.org/results/26th-edition-february-2021/

The Community Roundtable. (2021). *State of Community Management 2021.* Retrieved from https://s3.amazonaws.com/SoCM/SOCM_2021.pdf

Thoits, P. A. (1995). Stress, coping, and social support processes: Where are we? what next? *Journal of health and social behavior*, 53–79.

Tibshirani, J., Athey, S., & Wager, S. (2021). grf: Generalized random forests [Computer software manual]. Retrieved from https://github.com/grf-labs/grf (R package version 1.2.0.0)

Torres, M. M., Toral, S. L., Perales, M., & Barrero, F. (2011, June). Analysis of the core team role in open source communities. In *2011 international conference on complex, intelligent, and software intensive systems (pp. 109-114).*.

Tortoriello, M., Reagans, R., & McEvily, B. (2012). Bridging the knowledge gap: The influence of strong ties, network cohesion, and network range on the transfer of knowledge between organizational units. *Organization science*, *23*(4), 1024–1039.

Tsarenko, Y., & Tojib, D. (2015). Consumers' forgiveness after brand transgression: the effect of the firm's corporate social responsibility and response. *Journal of Marketing Management*, *31*(17-18), 1851–1877.

U.S. Census Bureau. (2010). *American community survey.* Retrieved 2 December 2019, from https://is.gd/3wGOwg

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228-1242.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication research*, *23*(1), 3-43.

Wang, X. (2016). People's motivation to participate in social network sites, subsequent behaviours, and situation self-awareness following a crisis: Evidence from the mh370 flight incident. *Australasian Journal of Information Systems*, *20*.

Wang, Y., & Fesenmaier, D. R. (2004). Towards understanding members' general participation in and active contribution to an online travel community. *Tourism management*, *25*(6), 709-722.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, *45*(4), 1191–1207.

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.

Wirtz, J., Den Ambtman, A., Bloemer, J., Horváth, C., Ramaseshan, B., Van De Klundert, J., ... Kandampully, J. (2013). Managing brands and customer engagement in online brand communities. *Journal of service Management*, *24*(3), 223-244.

Yang, J., Anderson, E. T., & Gordon, B. R. (2021). Digitization and flexibility: Evidence from the south korean movie market. *Marketing Science*.

Yoshida, M., Gordon, B., Heere, B., & James, J. D. (2015). Fan community identification: an empirical examination of its outcomes in japanese professional sport. *Sport Marketing Quarterly*, *24*(2), 105-120.

Yuan, D., Lin, Z., Filieri, R., Liu, R., & Zheng, M. (2020). Managing the product-harm crisis in the digital era: The role of consumer online brand community engagement. *Journal of Business Research*, *115*, 38–47.

Zander, A. (2018). *Motives and goals in groups.* Routledge.

Zhang, J. S., Tan, C., & Lv, Q. (2018). "this is why we play": Characterizing online fan communities of the nba teams. In *Proceedings of the acm on human-computer interaction,* (Vol. 2, p. 1-25).

Zhou, T. (2011). Understanding online community user participation: a social influence perspective. *Internet research*, *21*(1), 67-81.

Summary

Shared-interest communities are social groups of people who gather around a common interest. These communities provide people with a centralized source of information about their common interest. They are important hubs of knowledge, social support, socialization, and entertainment for consumers, brands, and institutions alike. For this reason, every day, millions of people resort to their shared-interest communities – both online and in-person – to meet, discuss, solve problems, and even manage disruptive situations of crisis or emergency, such as terrorist attacks, natural or civil disasters, financial instability, as well as product recalls and service failures. Given their importance of consumers, businesses, institutions, and citizens, several streams of literature across business and economics has investigated the antecedents of community participation, success, and resilience. In the course of this investigation, both scientific evidence and industry events demonstrated that the complex environment of institutions, businesses, and technologies, and the daily dynamics of shared-interest communities are inevitably interconnected. However, it is still unclear how the emergence of new technologies and the occurrence of (often disruptive) external events relate to the success and sustainability of shared-interest communities.

With three essays, in this dissertation, I shed light into the dynamics of shared-interest communities under the influence of changing technologies and potentially disruptive external events. In particular, I investigate three main questions: *(i) What is the impact of digitizing community activities on the participation intentions of community members?; (ii) What is the effect of a negative vs. positive shock to the shared purpose of an online community on members' engagement and social cohesion in the affected community?; (iii) What is the effect of a brand crisis on the engagement and social network resilience of consumers in brand communities?*

In Chapter 2, I investigate the first question, and focus on how increasing the extent of digitization of community activities impacts community participation. I address this issue using rich and unique data from the event-based community platform Meetup.com. Employing structural causal models and causal random forests, I find that increasing the extent of activity digitization decreases members' intentions

to attend such events. A counterfactual analysis shows that completely digitizing in-person activities causes an average 2.97% decrease in positive RSVPs. Furthermore, I find that the effect is heterogeneous across communities operating in different interest categories. This chapter contributes literature in marketing and economics studying the effects of digitizing human interactions on people's behavior in social groups. The chapter also informs marketing professionals, community managers, and policy makers, facing the urgent need to evaluate the consequences of digitization in their communities.

In Chapter 3, I turn to the second question, and assess the impact of negative vs positive shocks on the stated purpose of the community on social cohesion in online communities. To address this issue, I exploit quasi-experimental conditions in the empirical context of online sport communities, and I integrate difference-in-difference models with social network analyses. The results show that (i) negative shocks to a community's purpose cause a decrease in activity compared to positive shocks; (ii) the decrease is attributable mostly to members who belong to the "core" of the social networks; (iii) social cohesion is significantly affected by a negative purpose shock. In a series of heterogeneity analyses, I assess whether the disruptions to activity and cohesion can be mitigated by community managers. In particular, I evaluate two managerially relevant tools to address purpose-related shocks: expectations management and content moderation in the affected communities. This chapter supports any community-facing professionals, such as marketing and community managers, in maintaining their community in times of crisis, and in creating more value for their members during advantageous times.

In Chapter 4, I still investigate the effects of disruptive events on community dynamics, but I focus more specifically on the relationship between brand communities and the brand environment. In this chapter, I assess the effect of brand crises on the volume of customer interactions in online brand communities, and the properties of the brand social network correlated with ease and speed of information spread. I use data from 300 brand communities on Reddit.com exposed to different brand crises. The data includes brand crises reported by media outlets between 2010 and 2019.

In a series of difference-in-difference analyses, I find that brand crises (i) increase the volume of consumer discussions in online brand communities, and (ii) affect the patterns of information-sharing in the brand networks. Focusing on consumer types, I show that consumers who were active any time before the crisis effectively *disengage* from their brand communities after a crisis event. This result suggests that the average boost in brand-related activity is attributable to consumers who only become after the crises. Furthermore, I show that the decrease in engagement is mitigated among consumers who had more experience, loyalty, or status within the brand community – although this mitigation seems heterogeneous at least across product categories. In line with this set of evidence, I suggest that brand crises are a serious threat to the integrity of online brand communities, but that consumer loyalty and commitment has the potential to preserve the functioning of brand spaces online in under certain circumstances. The insights from this chapter support businesses and organizations managing online communities in situations of external stress and unexpected reputational threats.

Overall, the findings of this dissertation contribute to the scientific and managerial knowledge about the internal and environmental circumstances that allow shared-interest communities to thrive in a complex world. As human interactions via digitization technologies become the new norm, and as external events prompt dramatic collective action on digital platforms, the findings of this dissertation are both extremely timely and useful for researchers, practitioners, and policy-makers alike.

Samenvatting

Gemeenschappen met gedeelde belangen zijn sociale groepen mensen die bij elkaar komen in het kader van een gemeenschappelijk belang. Deze gemeenschappen voorzien mensen van een gecentraliseerde informatiebron over hun gemeenschappelijk belang. Het zijn belangrijke hubs voor kennis, sociale steun, socialisatie en entertainment voor zowel consumenten als merken en instellingen. Daarom doen miljoenen mensen elke dag een beroep op hun gemeenschappen met gedeelde belangen – zowel online als in persoon – om elkaar te ontmoeten, dingen te bespreken, problemen op te lossen en zelfs ontwrichtende crisissituaties of noodsituaties het hoofd te bieden, zoals terreuraanslagen, civiele of natuurrampen, financiële instabiliteit, maar ook het terugroepen van producten en storingen in de dienstverlening. Gezien het belang ervan voor consumenten, bedrijven, instellingen en burgers, is in verschillende stromingen van de zakelijke en economische literatuur onderzoek gedaan naar de antecedenten van de deelname aan gemeenschappen en het succes en de veerkracht ervan. In de loop van dit onderzoek hebben zowel wetenschappelijk bewijs als gebeurtenissen in de sectoren aangetoond dat de complexe omgeving van instellingen, bedrijven en technologieën, en de dagelijkse dynamiek van gemeenschappen met gedeelde belangen onvermijdelijk met elkaar verbonden zijn. Het is echter nog onduidelijk hoe de opkomst van nieuwe technologieën en het optreden van (vaak ontwrichtende) externe gebeurtenissen zich verhouden tot het succes en de duurzaamheid van gemeenschappen met gedeelde belangen.

Met drie essays werp ik in dit proefschrift licht op de dynamiek van gemeenschappen met gedeelde belangen onder invloed van veranderende technologieën en potentieel ontwrichtende externe gebeurtenissen. Meer bepaald onderzoek ik drie hoofdvragen: (i) Wat is de impact van digitalisering van gemeenschapsactiviteiten op de deelname-intenties van gemeenschapsleden; (ii) Wat is het effect van een negatieve tegenover een positieve schok t.o.v. het gedeelde doel van een online gemeenschap op de betrokkenheid en sociale cohesie van leden in de getroffen gemeenschap; (iii) Wat is het effect van een merkcrisis op de veerkracht van de betrokkenheid en het sociale netwerk van consumenten in merkgemeenschappen?

In hoofdstuk 2 onderzoek ik de eerste vraag en richt ik me op hoe de toenemende mate van digitalisering van gemeenschapsactiviteiten de gemeenschapsdeelname beïnvloedt. Ik onderzoek deze kwestie met behulp van uitgebreide en unieke gegevens van het op evenementen gebaseerde communityplatform Meetup.com. Aan de hand van structurele causale modellen en causale random forests kom ik tot de conclusie dat een toenemende mate van digitalisering van activiteiten de intentie van leden om dergelijke evenementen bij te wonen, vermindert. Een contrafeitelijke analyse toont aan dat het volledig digitaliseren van offline activiteiten de mediaan van positieve RSVP's (reacties op uitnodigingen) met 2.97% vermindert. Bovendien stel ik vast dat het effect heterogeen is in de verschillende gemeenschappen. Dit hoofdstuk levert een bijdrage aan de marketing- en economieliteratuur waarin de effecten worden bestudeerd van digitalisering van menselijke interacties op het gedrag van mensen in sociale groepen. Het hoofdstuk biedt ook informatie aan marketing professionals, community managers en beleidsmakers die geconfronteerd worden met de dringende noodzaak om de gevolgen van digitalisering in hun gemeenschappen te evalueren.

In hoofdstuk 3 ga ik in op de tweede vraag en beoordeel ik de impact van negatieve tegenover positieve schokken t.o.v. het verklaarde doel van de gemeenschap op de sociale cohesie in online gemeenschappen. Om deze vraag te beantwoorden, maak ik gebruik van quasi-experimentele omstandigheden in de empirische context van online sportgemeenschappen en integreer ik difference-in-difference-modellen met socialenetwerkanalyses. De resultaten tonen aan dat (i) negatieve schokken t.o.v. het doel van een gemeenschap leiden tot een afname van de activiteit in vergelijking met positieve schokken; (ii) de afname vooral toe te schrijven is aan leden die tot de "kern" van de sociale netwerken behoren; (iii) de sociale cohesie significant wordt beïnvloed door een negatieve schok t.o.v. het doel. In een reeks heterogeniteitsanalyses ga ik na of de ontwrichtingen van de activiteit en van de cohesie kunnen worden verminderd door gemeenschapsmanagers. In het bijzonder evalueer ik twee managementtechnisch relevante instrumenten om doelgerelateerde schokken aan te pakken: verwachtingsmanagement en contentmoderatie in de getroffen gemeenschappen. Dit hoofdstuk ondersteunt alle professionals die zich tot de gemeenschap richten, zoals marketing en community managers, bij het in stand houden van hun gemeenschap in tijden van crisis, en bij het creëren van meer waarde voor hun leden in gunstige tijden.

In hoofdstuk 4 onderzoek ik nog steeds de effecten van ontwrichtende gebeurtenissen op de dynamiek van gemeenschappen, maar richt ik me vooral op de relatie tussen merkgemeenschappen en de merkomgeving. In dit hoofdstuk evalueer ik het effect van merkcrises op het volume van klantinteracties in online merkgemeenschappen en de eigenschappen van het sociale netwerk van het merk in correlatie met gemak en snelheid van informatieverspreiding. Ik gebruik gegevens van 300 merkgemeenschappen op Reddit.com die aan verschillende merkcrises werden blootgesteld. De gegevens omvatten merkcrises die tussen 2010 en 2019 door mediakanalen zijn gemeld. In een reeks difference-in-difference-analyses kom ik tot de vaststelling dat merkcrises (i) het volume van consumentendiscussies in online merkgemeenschappen doen toenemen; (ii) de patronen van informatieuitwisseling in de merknetwerken beïnvloeden; en (iii) een negatieve impact hebben op het sentiment van door gebruikers aangemaakte inhoud. Als we kijken naar consumententypes, blijkt dat consumenten die vóór de crisis actief waren, zich na een crisis niet meer aansluiten bij de merkgemeenschappen. Uit de resultaten kunnen we afleiden dat de gemiddelde toename van merkgerelateerde activiteit toe te schrijven is aan consumenten die zich aansluiten na de crises. Verder laat ik zien dat de afname in betrokkenheid wordt afgezwakt bij consumenten die meer ervaring, loyaliteit of status binnen de merkgemeenschap hadden – hoewel deze afzwakking heterogeen lijkt te zijn, althans in de verschillende bedrijven en crises. In lijn met dit bewijsmateriaal suggereer ik dat merkcrises een serieuze bedreiging vormen voor de integriteit van online merkgemeenschappen, maar dat consumentenloyaliteit en -betrokkenheid het functioneren van online merkruimten onder bepaalde omstandigheden in stand kunnen houden. De inzichten uit dit hoofdstuk ondersteunen bedrijven en organisaties die online gemeenschappen beheren in situaties van externe stress en onverwachte bedreigingen voor de reputatie.

Over het algemeen dragen de bevindingen van dit proefschrift bij aan de wetenschappelijke en bestuurskundige kennis over de interne en omgevingsomstandigheden die gemeenschappen met gedeelde belangen in staat stellen om te gedijen in een complexe wereld. Nu menselijke interacties via digitaliseringstechnologieën de nieuwe norm worden, en externe gebeurtenissen ingrijpende collectieve actie op digitale platforms uitlokken, zijn de bevindingen van dit proefschrift zowel uiterst actueel als nuttig voor onderzoekers, mensen in de praktijk en beleidsmakers.

About the Author

Martina Pocchiari was born in Potenza, Italy on July 30, 1993. She holds a Bachelor of Science in Economics and Business cum laude from the University of Bologna, and a Master of Science in Marketing Management cum laude from the Rotterdam School of Management, Erasmus University. She pursued a PhD in Empirical Quantitative Marketing under the supervision of Dr. Jason M.T. Roos and Prof. Dr. Ir. Gerrit van Bruggen at the Rotterdam School of Management, Erasmus University. Since the start of her PhD in 2017, Martina studied the factors contributing to the success and resilience of shared-interest communities – a topic that gained unparalleled relevance during the Covid–19 pandemic. Her work has been presented at prestigious international academic conferences, including the INFORMS Marketing Science conference, EMAC, the Journal of Marketing Research conference, and the IJRM conference. Her work has been featured in industry outlets, podcasts and newsletters for community management, and presented at the FOSDEM industry conference in 2021. Martina established successful research collaborations with international coauthors at the Hebrew University; the Wharton School, University of Pennsylvania; and MIT Sloan. She was a visiting research fellow at the Hebrew University, Jerusalem Business School in Spring 2022. Next to her research output, Martina has supervised close to 100 Bachelor students' theses at RSM. In August 2022, Martina will join the National University of Singapore, NUS Business School as an assistant professor of marketing. She will continue to research human interaction in digital platforms and online communities, and the consumption of products and information in increasingly digitized and complex environments.

April, 2022

Martina Pocchiari

Rotterdam School of Management, Erasmus University
Postbus 1738, Room T10-24, 3000 DR Rotterdam, Netherlands

pocchiari@rsm.nl https://sites.google.com/view/martinapocchiari @pccmtn

Research Interests

Causal inference and causal identification, online communities and social influence, social network analysis, machine learning (e.g. NLP algorithms, one-class and multi-class classification problems, ensamble methods for causal identification), behavioral lab/field experiments.

Academic Appointments

2022-	Assistant Professor of Marketing *NUS Business School* *National University of Singapore*

Education

2017-2022	PhD Candidate in **Empirical Quantitative Marketing** (Expected Graduation: June 2022) *Rotterdam School of Management, Erasmus University* Thesis advisors: Gerrit van Bruggen and Jason M.T. Roos
Spring 2022	Visiting Research Fellow, Marketing Department *Hebrew University, Jerusalem School of Business Administration*
	PhD Coursework: *Tinbergen Institute* – Microeconomics, Game Theory, Behavioral Economics, Experimental Economics, Econometrics I, Econometrics II, Applied Microeconometrics, Bayesian Econometrics
	Erasmus Research Institute of Management – Causal identification Summer School, Advanced Statistical Methods, Empirical Research Methodology & Measurement, Current Topics in Marketing Research, Academic Writing, Publishing Strategy, Scientific Integrity

TU Delft Faculty of Electrical Engineering, Mathematics & Computer Science
– Modeling and Data Analysis in Complex Networks

University of Amsterdam - Psychological Networks (Winter School)

2015-2016 MSc **Marketing Management** – Digital Marketing & Analytics Track, Cum Laude
Rotterdam School of Management, Erasmus University
Thesis advisors: Pieter Schoonees and Alina Ferecatu

2012-2015 BSc **Business and Economics**, Cum Laude
University of Bologna
Thesis advisor: Sara Valentini

Working Papers

Drafts available upon request.

Pocchiari, M., Roos, J.M.T. ”The Effect of Digitizing Community Activities on Community Participation: Evidence from Meetup.com”.

Featured in:
Talk About Your Community liveshow. July 30 2021. Link.
A View from the Clocktower – Clocktower Advisors newsletter. July 2 2021. Link.
The Developer Advocates newsletter. June 7 2021. Link.
The Communities Show Podcast. May 13 2021. Link.

Pocchiari, M., Dover, Y. ”The Role of Community Shared Purpose in Online Community Dynamics”.

Featured in:
Talk About Your Community liveshow. January 2022. Link.

Roos, J.M.T., Ferecatu, A., **Pocchiari, M.** ”An Experimental Paradigm for Studying Exposure to Fake News”.

Research in Progress

Pocchiari, M., Yildirim, P., Almaatouq, A. ”How Do Brand Networks Break in Face of a Crisis?”. Manuscript in preparation.

Conference Presentations

2022

"How Do Brand Networks Break in Face of a Crisis?"

INFORMS Marketing Science. June 2022.

"The Role of Community Shared Purpose in Online Community Dynamics"

INFORMS Marketing Science. June 2022 (Presenter: Yaniv Dover).

Industry Talks

Talk About Your Community Show. January 2022.

2021

"The Effect of Digitizing Community Activities on Community Participation: Evidence from Meetup.com"

INFORMS Marketing Science. June 3-5 2021 (Session Chair).

EMAC 2021 - Digital Marketing and Social Media Track. May 25-28 2021.

EMAC Doctoral Colloquium Advanced Marketing Research. May 23-25 2021.

Marketing in Israel Conference. February 3 2021.

"The Effect of External Events on Online Community Dynamics"

Rotterdam School of Management brownbag seminar series. October 21 2021.

Journal of Marketing Research – Mitigation in Marketing Conference. September 20-21 2021 (Presenter: Yaniv Dover).

INFORMS Marketing Science. June 3-5 2021 (Presenter: Yaniv Dover).

EMAC 2021 - Digital Marketing and Social Media Track. May 25-28 2021. (Presenter: Yaniv Dover)

Industry Talks

Talk About Your Community Show. July 30 2021.

The Communities Show Podcast. "Using Science to Research Online Communities". May 12 2021.

Free and Open Source Developers' European Meeting (FOSDEM). "Switching Open Source Communities: How to Stay Authentic to Yourself and Find Hidden Benefit In Your New Role" (Joint talk with Anastasia Raspopina). February 6-7 2021.

2020 (All Virtual)

"How Do Social Networks Break in Face of a Crisis?"

RSM Marketing Brownbag Seminar. November 26, 2020.

"The Effect of External Events on Online Community Resilience"

RSM Marketing PhD day, October 14 2020.

"The Effect of Activity Digitization on Participation Intentions in Brand Communities: Evidence from Meetup.com"

Interactive Marketing Research Conference. October 29 2020.

EMAC Doctoral Colloquium Advanced Marketing Research, September 2-4 2020.

INFORMS Marketing Science Doctoral Consortium. June 17 2020.

INFORMS Marketing Science. June 11-13 2020 (Session chair).

SIMktg - Società Italiana Marketing, Doctoral & Research Colloquium. May 18 2020.

2020 (Canceled due to COVID–19)

"The Effect of Activity Digitization on Participation Intentions in Brand Communities: Evidence from Meetup.com"

Invited for presentation at Customer Journeys in a Digital World conference. Amsterdam, June 21-23 2020.

Accepted for presentation at EMAC Digital Marketing and Social Media track. Budapest, May 27-29 2020.

2019

"The Impact of Community and Activity Characteristics on Active Participation: Evidence from Meetup.com"

INFORMS Marketing Science. Rome, June 19-22 2019 (Session chair).

Marketing Effectiveness Through Customer Journeys and Multichannel Management. Bologna, June 16-18 2019.

EMAC Doctoral Colloquium. Hamburg, May 27-31 2019.

"An Experimental Paradigm for Studying Exposure to Fake News"

Invited seminar at ESADE Business School. Online, May 20 2020. (Presenter: J.M.T. Roos)

INFORMS Marketing Science. Rome, June 19-22 2019. (Presenter: J.M.T. Roos)

EMAC Main Conference. Hamburg, May 27-31 2019. (Presenter: J.M.T. Roos)

Invited Seminars

Ben-Gurion University - Guilford Glazer Faculty of Business & Management, April 2022

Reichman University (formerly IDC Herzliya) - Arison School of Business, March 2022

The Hebrew University Business School, March 2022

McGill University - Desautels Faculty of Management, November 2021 (canceled)

Universitat Pompeu Fabra - Department of Economics and Business, October 2021

ESADE Business School, October 2021

Simon Fraser University - Beedie School of Business, October 2021

Hong Kong University (HKU) Business School, October 2021

Nova School of Business and Economics, September 2021

Tilburg School of Economics and Management, September 2021

University of Notre Dame - Mendoza College of Business, September 2021

IESE Business School, September 2021

April, 2022

Santa Clara University - Leavey School of Business, September 2021
National University of Singapore - NUS Business School, August 2021
EDHEC Business School - February 2021

Recognitions

Aug 2021	AMA-Sheth Foundation Doctoral Consortium Fellow (Indiana University, Bloomington, IN)
2020 & 2021	European Marketing Academy Conference (Online). EMAC Doctoral Colloquium Fellow
2020 & 2021	ISMS Marketing Science. Doctoral Consortium Fellow
May 18, 2020	SIMktg - Società Italiana Marketing. Doctoral Colloquium Fellow
April 22, 2020	ERIM Talent Placement Grant (∼€30,000)
June 19, 2019	ISMS Marketing Science. Doctoral Consortium Fellow (Rome, Italy)
May 2019	European Marketing Academy Conference. Doctoral Colloquium Fellow (Hamburg, Germany)
Sept 2017	Fully funded PhD scholarship in Empirical Quantitative Marketing (ERIM, Erasmus University)
2015-2016	Marketing Management Honours Program Cohort Member (RSM, Erasmus University)

Teaching Interests and Experience

I have taught and assisted a variety of classes in marketing and statistics, from undergraduate to master levels. I am passionate about teaching and flexible about different classes including (but not limited to) digital marketing and analytics, statistics, big data analytics for marketing insights, marketing research, marketing principles and strategy, causal inference and experimental economics.

Instructor:

Mar-July 2022	Research Training and Bachelor Thesis - Analytical Decision-Makers Track *BSc International Business Administration – Rotterdam School of Management, Erasmus University* N=24, average instructor evaluation not yet available.
Jan-June 2021, Jan-June 2020	Research Training and Bachelor Thesis *BSc International Business Administration – Rotterdam School of Management, Erasmus University* N=30 per year, average instructor evaluation: 9.8/10 (2021), 9/10 (2020)

Teaching Assistant:

Sept-Oct 2020	Experimentation and Causal Inference *MSc Business Analytics and Management – Rotterdam School of Management, Erasmus University*
Jan-Feb 2020, Jan-Feb 2019	Lab Sessions, Marketing Analytics *MSc Marketing Management – Rotterdam School of Management, Erasmus University*
Feb-Apr 2016	Marketing FEB11008X *BSc International Economics and Business Economics – Erasmus School of Economics, Erasmus University*

Academic Service

Trainee reviewer at the Journal of Consumer Research.

2020-	Member of the Customer Analytics Lab, Erasmus Centre for Data Analytics
2020-	Writer and contributor for Towards Data Science
2019-	Organizer of PhD feedback sessions for doctoral and post-doc students at the Marketing Department of the Rotterdam School of Management
2017-2019	Board Member Alumni Association - Bachelor Degree Course in Business and Economics, University of Bologna

Methodologies and Tech Skills

Methods	Causal inference with potential outcomes and graphical models, machine learning, analysis of complex networks, Bayesian econometrics, behavioral experiments (lab and field)
Tech	R and Stan, Python, oTree (based on JavaScript, Django and HTML), Qualtrics, C++, LaTeX

Other Working Experience

2016-2017	Marketing Research Consultant - Nielsen, division Pointlogic
2014-2015	Teaching Assistant - Angolo del Sapere srl
2014	Internship - Geocart ltd, a Geocart Group Company
2011-2013	Pre-school/kindergarten teacher (kids aged 3-5)

Personal Interests

I love learning about art history, playing Nintendo games, learning Russian, bouldering (indoor climbing), and running (quite slowly).

April, 2022

I love attending concerts – my favourite music genres are rock 'n' roll, electronic, alternative and new-prog rock, and industrial music (Nine Inch Nails, Porcupine Tree and Steven Wilson, early Elton John, the Mars Volta, Interpol, & more).

The ERIM PhD Series

The ERIM PhD Series contains PhD dissertations in the field of Research in Management defended at Erasmus University Rotterdam and supervised by senior researchers affiliated to the Erasmus Research Institute of Management (ERIM). All dissertations in the ERIM PhD Series are available in full text through the ERIM Electronic Series Portal: http://repub.eur.nl/pub. ERIM is the joint research institute of the Rotterdam School of Management (RSM) and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam (EUR).

Dissertations in the last four years

Ahmadi, S., A motivational perspective to decision-making and behavior in organizations, Promotors: Prof. J.J.P. Jansen & Dr T.J.M. Mom, EPS-2019-477-S&E, https://repub.eur.nl/pub/116727

Akemu, O., Corporate Responses to Social Issues: Essays in Social Entrepreneurship and Corporate Social Responsibility, Promotors: Prof. G.M. Whiteman & Dr S.P. Kennedy, EPS-2017-392-ORG, https://repub.eur.nl/pub/95768

Albuquerque de Sousa, J.A., International stock markets: Essays on the determinants and consequences of financial market development, Promotors: Prof. M.A. van Dijk & Prof. P.A.G. van Bergeijk, EPS-2019-465-F&A, https://repub.eur.nl/pub/115988

Alserda, G.A.G., Choices in Pension Management, Promotors:Prof. S.G. van der Lecq & Dr O.W. Steenbeek, EPS-2017-432-F&A, https://repub.eur.nl/pub/103496

Anantavrasilp, S., Essays on Ownership Structures, Corporate Finance Policies and Financial Reporting Decisions, Promotors: Prof. A. de Jong & Prof. P.G.J. Roosenboom, EPS-2021-516-F&E, https://repub.eur.nl/pub/134947

Arampatzi, E., Subjective Well-Being in Times of Crises: Evidence on the Wider Impact of Economic Crises and Turmoil on Subjective Well-Being, Promotors: Prof. H.R. Commandeur, Prof. F. van Oort & Dr. M.J. Burger, EPS-2018-459-S&E, https://repub.eur.nl/pub/111830

Arslan, A.M., Operational Strategies for On-demand Delivery Services, Promotors: Prof. R.A. Zuidwijk & Dr N.A. H. Agatz, EPS-2019-481-LIS, https://repub.eur.nl/pub/126463

Aydin Gökgöz, Z. Mobile Consumers and Applications: Essays on Mobile Marketing, Promotors: Prof. G.H. van Bruggen & Dr B. Ataman, EPS-2021-519-MKT, whttps://repub.eur.nl/pub/135352

Azadeh, K., Robotized Warehouses: Design and Performance Analysis, Promotors: Prof. dr. ir M.B.M. de Koster & Prof. D. Roy, EPS-2021-515-LIS, https://repub.eur.nl/pub/135208

Avci, E., Surveillance of Complex Auction Markets: A Market Policy Analytics Approach, Promotors: Prof. W. Ketter, Prof. H.W.G.M. van Heck & Prof. D.W. Bunn, EPS-2018-426-LIS, https://repub.eur.nl/pub/106286

Balen, T.H. van, Challenges of Early-Stage Entrepreneurs: The Roles of Vision Communication and Team Membership Change, Promotors: Prof. J.C.M. van den Ende & Dr M. Tarakci, EPS-2019-468-LIS, https://repub.eur.nl/pub/115654

Bansraj, S.C., The Principles of Private Equity: Ownership and Acquisitions, Promotors: Prof. J.T.J Smit & Dr V. Volosovych, EPS-2020-507-F&A, https://repub.eur.nl/pub/132329

Bavato, D., With New Eyes: The recognition of novelty and novel ideas, Promotors: Prof. D.A. Stam & Dr. S. Tasselli, EPS-2020-500-LIS, https://repub.eur.nl/pub/134264

Bernoster, I., Essays at the Intersection of Psychology, Biology, and Entrepreneurship, Promotors: Prof. A.R. Thurik, Prof. I.H.A. Franken & Prof. P.J.F Groenen, EPS-2018-463-S&E, https://repub.eur.nl/pub/113907

Blagoeva, R.R., The Hard Power of Soft Power: A behavioral strategy perspective on how power, reputation, and status affect firms, Promotors: Prof. J.J.P. Jansen & Prof. T.J.M. Mom, EPS-2020-495-S&E, https://repub.eur.nl/pub/127681

Bouman, P., Passengers, Crowding and Complexity: Models for Passenger Oriented Public Transport, Prof. L.G. Kroon, Prof. A. Schöbel & Prof. P.H.M. Vervest, EPS-2017-420-LIS, https://repub.eur.nl/pub/100767

Breugem, T., Crew Planning at Netherlands Railways: Improving Fairness, Attractiveness, and Efficiency, Promotors: Prof. D. Huisman & Dr T.A.B. Dollevoet, EPS-2020-494-LIS, https://repub.eur.nl/pub/124016

Bunderen, L. van, Tug-of-War: Why and when teams get embroiled in power struggles, Promotors: Prof. D.L. van Knippenberg & Dr. L. Greer, EPS-2018-446-ORG, https://repub.eur.nl/pub/105346

Burg, G.J.J. van den, Algorithms for Multiclass Classification and Regularized Regression, Promotors: Prof. P.J.F. Groenen & Dr. A. Alfons, EPS-2018-442-MKT, https://repub.eur.nl/pub/103929

Chammas, G., Portfolio concentration, Promotor: Prof. J. Spronk, EPS-2017-410-F&E, https://repub.eur.nl/pub/94975

Chan, H.Y., Decoding the consumer's brain: Neural representations of consumer experience, Promotors: Prof. A. Smidts & Dr M.A.S. Boksem, EPS-2019-493-MKT, https://repub.eur.nl/pub/124931

Couwenberg, L., Context dependent valuation: A neuroscientific perspective on consumer decision-making, Promotors: Prof. A. Smit, Prof. A.G. Sanfrey & Dr M.A.S. Boksem, EPS-2020-505-MKT, https://repub.eur.nl/pub/129601

Dalmeijer, K., Time Window Assignment in Distribution Networks, Promotors: Prof A.P.M. Wagelmans & Dr R. Spliet, EPS-2019-486-LIS, https://repub.eur.nl/pub/120773

Dennerlein, T. Empowering Leadership and Employees' Achievement Motivations: the Role of Self-Efficacy and Goal Orientations in the Empowering Leadership Process, Promotors: Prof. D.L. van Knippenberg & Dr J. Dietz, EPS-2017-414-ORG, https://repub.eur.nl/pub/98438

Dolgova, E., On Getting Along and Getting Ahead: How Personality Affects Social Network Dynamics, Promotors: Prof. P.P.M.A.R Heugens & Prof. M.C. Schippers, EPS-2019-455-S&E, https://repub.eur.nl/pub/119150

Duijzer, L.E., Mathematical Optimization in Vaccine Allocation, Promotors: Prof. R. Dekker & Dr W.L. van Jaarsveld, EPS-2017-430-LIS, https://repub.eur.nl/pub/101487

Fasaei, H., Changing the Narrative: The Behavioral Effects of Social Evaluations on the Decision Making of Organizations, Promotors: Prof. J.J.P. Jansen, Prof. T.J.M. Mom & Dr. M.P. Tempelaar, EPS-2020-492-S&E, https://repub.eur.nl/pub/129598

Eijlers, E., Emotional Experience and Advertising Effectiveness: on the use of EEG in marketing, Prof. A. Smidts & Prof. M.A.S. Boksem, EPS-2019-487-MKT, https://repub.eur.nl/pub/124053

El Nayal, O.S.A.N., Firms and the State: An Examination of Corporate Political Activity and the Business-Government Interface, Promotor: Prof. J. van Oosterhout & Dr. M. van Essen, EPS-2018-469-S&E, https://repub.eur.nl/pub/114683

Feng, Y., The Effectiveness of Corporate Governance Mechanisms and Leadership Structure: Impacts on strategic change and firm performance, Promotors: Prof. F.A.J. van den Bosch, Prof. H.W. Volberda & Dr J.S. Sidhu, EPS-2017-389-S&E, https://repub.eur.nl/pub/98470

Frick, T.W., The Implications of Advertising Personalization for Firms, Consumer, and Ad Platforms, Promotors: Prof. T. Li & Prof. H.W.G.M. van Heck, EPS-2018-452-LIS, https://repub.eur.nl/pub/110314

Fytraki, A.T., Behavioral Effects in Consumer Evaluations of Recommendation Systems, Promotors: Prof. B.G.C. Dellaert & Prof. T. Li, EPS-2018-427-MKT, https://repub.eur.nl/pub/110457

Gai, J., Contextualized Consumers: Theories and Evidence on Consumer Ethics, Product Recommendations, and Self-Control, Promotors: Prof. S. Puntoni & Prof. S.T.L. Sweldens, EPS-2020-498-MKT, https://repub.eur.nl/pub/127680

Ghazizadeh, P. Empirical Studies on the Role of Financial Information in Asset and Capital Markets, Promotors: Prof. A. de Jong & Prof. E. Peek, EPS-2019-470-F&A, https://repub.eur.nl/pub/114023

Giurge, L., A Test of Time; A temporal and dynamic approach to power and ethics, Promotors: Prof. M.H. van Dijke & Prof. D. De Cremer, EPS-2017-412-ORG, https://repub.eur.nl/pub/98451

Gobena, L., Towards Integrating Antecedents of Voluntary Tax Compliance, Promotors: Prof. M.H. van Dijke & Dr P. Verboon, EPS-2017-436-ORG, https://repub.eur.nl/pub/103276

Groot, W.A., Assessing Asset Pricing Anomalies, Promotors: Prof. M.J.C.M. Verbeek & Prof. J.H. van Binsbergen, EPS-2017-437-F&A, https://repub.eur.nl/pub/103490

Hanselaar, R.M., Raising Capital: On pricing, liquidity and incentives, Promotors: Prof. M.A. van Dijk & Prof. P.G.J. Roosenboom, EPS-2018-429-F&A, https://repub.eur.nl/pub/113274

Harms, J. A., Essays on the Behavioral Economics of Social Preferences and Bounded Rationality, Prof. H.R. Commandeur & Dr K.E.H. Maas, EPS-2018-457-S&E, https://repub.eur.nl/pub/108831

Hartleb, J., Public Transport and Passengers: Optimization Models that Consider Travel Demand, Promotors: Prof. D. Huisman, Prof. M. Friedrich & Dr. M.E. Schmidt, EPS-2021-535-LIS, https://repub.eur.nl/pub/135664

Hendriks, G., Multinational Enterprises and Limits to International Growth: Links between Domestic and Foreign Activities in a Firm's Portfolio, Promotors: Prof. P.P.M.A.R. Heugens & Dr. A.H.L Slangen, EPS-2019-464-S&E, https://repub.eur.nl/pub/114981

Hengelaar, G.A., The Proactive Incumbent: Holy grail or hidden gem? Investigating whether the Dutch electricity sector can overcome the incumbent's curse and lead the sustainability transition, Promotors: Prof. R.J. M. van Tulder & Dr K. Dittrich, EPS-2018-438-ORG, https://repub.eur.nl/pub/102953

Hoogervorst, R., Improving the Scheduling and Rescheduling of Rolling Stock: Solution Methods and Extensions, Promotors: Prof. D. Huisman & Dr T.A.B. Dollevoet, EPS-2021-534-LIS, https://repub.eur.nl/pub/135726

Jacobs, B.J.D., Marketing Analytics for High-Dimensional Assortments, Promotors: Prof. A.C.D. Donkers & Prof. D. Fok, EPS-2017-445-MKT, https://repub.eur.nl/pub/103497

Jia, F., The Value of Happiness in Entrepreneurship, Promotors: Prof. D.L. van Knippenberg & Dr Y. Zhang, EPS-2019-479-ORG, https://repub.eur.nl/pub/115990

Kahlen, M. T., Virtual Power Plants of Electric Vehicles in Sustainable Smart Electricity Markets, Promotors: Prof. W. Ketter & Prof. A. Gupta, EPS-2017-431-LIS, https://repub.eur.nl/pub/100844

Kampen, S. van, The Cross-sectional and Time-series Dynamics of Corporate Finance: Empirical evidence from financially constrained firms, Promotors: Prof. L. Norden & Prof. P.G.J. Roosenboom, EPS-2018-440-F&A, https://repub.eur.nl/pub/105245

Karali, E., Investigating Routines and Dynamic Capabilities for Change and Innovation, Promotors: Prof. H.W. Volberda, Prof. H.R. Commandeur & Dr J.S. Sidhu, EPS-2018-454-S&E, https://repub.eur.nl/pub/106274

Keko. E, Essays on Innovation Generation in Incumbent Firms, Promotors: Prof. S. Stremersch & Dr N.M.A. Camacho, EPS-2017-419-MKT, https://repub.eur.nl/pub/100841

Kerkkamp, R.B.O., Optimisation Models for Supply Chain Coordination under Information Asymmetry, Promotors: Prof. A.P.M. Wagelmans & Dr. W. van den Heuvel, EPS-2018-462-LIS, https://repub.eur.nl/pub/109770

Khattab, J., Make Minorities Great Again: a contribution to workplace equity by identifying and addressing constraints and privileges, Promotors: Prof. D.L. van Knippenberg & Dr A. Nederveen Pieterse, EPS-2017-421-ORG, https://repub.eur.nl/pub/99311

Kim, T. Y., Data-driven Warehouse Management in Global Supply Chains, Promotors: Prof. R. Dekker & Dr C. Heij, EPS-2018-449-LIS, https://repub.eur.nl/pub/109103

Klitsie, E.J., Strategic Renewal in Institutional Contexts: The paradox of embedded agency, Promotors: Prof. H.W. Volberda & Dr. S. Ansari, EPS-2018-444-S&E, https://repub.eur.nl/pub/106275

Koolen, D., Market Risks and Strategies in Power Systems Integrating Renewable Energy, Promotors: Prof. W. Ketter & Prof. R. Huisman, EPS-2019-467-LIS, https://repub.eur.nl/pub/115655

Kong, L. Essays on Financial Coordination, Promotors: Prof. M.J.C.M. Verbeek, Dr. D.G.J. Bongaerts & Dr. M.A. van Achter. EPS-2019-433-F&A,https://repub.eur.nl/pub/114516

Koritarov, V.D., The Integration of Crisis Communication and Regulatory Focus: Deconstructing and Optimizing the Corporate Message, Promotors: Prof. C.B.M. van riel, Dr G.A.J.M. Berens & Prof. P. Desmet, EPS-2021-522-ORG,https://repub.eur.nl/pub/136979

Korman, B., Leader-Subordinate Relations: The Good, the Bad and the Paradoxical, Promotors: S.R. Giessner & Prof. C. Tröster, EPS-2021-511-ORG, https://repub.eur.nl/pub/135365

Kyosev, G.S., Essays on Factor Investing, Promotors: Prof. M.J.C.M. Verbeek & Dr J.J. Huij, EPS-2019-474-F&A, https://repub.eur.nl/pub/116463

Lamballais Tessensohn, T., Optimizing the Performance of Robotic Mobile Fulfillment Systems, Promotors: Prof. M.B.M de Koster, Prof. R. Dekker & Dr D. Roy, EPS-2019-411-LIS, https://repub.eur.nl/pub/116477

Leung, W.L., How Technology Shapes Consumption: Implications for Identity and Judgement, Promotors: Prof. S. Puntoni & Dr G Paolacci, EPS-2019-485-MKT, https://repub.eur.nl/pub/117432

Li, Wei., Competition in the Retail Market of Consumer Packaged Goods, Promotors: Prof. D.Fok & Prof. Ph.H.B.F. Franses, EPS-2021-503-MKT, https://repub.eur.nl/pub/134873

Li, X. Dynamic Decision Making under Supply Chain Competition,
Promotors: Prof. M.B.M de Koster, Prof. R. Dekker & Prof. R. Zuidwijk, EPS-2018-466-LIS, https://repub.eur.nl/pub/114028

Liu, N., Behavioral Biases in Interpersonal Contexts,
Supervisors: Prof. A. Baillon & Prof. H. Bleichrodt, EPS-2017-408-MKT, https://repub.eur.nl/pub/95487

Maas, A.J.J., Organizations and their external context: Impressions across time and space, Promotors: Prof. P.P.M.A.R Heugens & Prof. T.H. Reus, EPS-2019-478-S&E, https://repub.eur.nl/pub/116480

Maira, E., Consumers and Producers, Promotors: Prof. S. Puntoni & Prof. C. Fuchs, EPS-2018-439-MKT, https://repub.eur.nl/pub/104387

Manouchehrabadi, B., Information, Communication and Organizational Behavior, Promotors: Prof. G.W.J. Hendrikse & Dr O.H. Swank, EPS-2020-502-ORG, https://repub.eur.nl/pub/132185

Matawlie, N., Through Mind and Behaviour to Financial Decisions, Promotors: Prof. J.T.J. Smit & Prof. P. Verwijmeren, EPS-2020-501-F&A, https://repub.eur.nl/pub/134265

Mirzaei, M., Advanced Storage and Retrieval Policies in Automated Warehouses, Promotors: Prof. M.B.M. de Koster & Dr N. Zaerpour, EPS-2020-490-LIS, https://repub.eur.nl/pub/125975

Nair, K.P., Strengthening Corporate Leadership Research: The relevance of biological explanations, Promotors: Prof. J. van Oosterhout & Prof. P.P.M.A.R Heugens, EPS-2019-480-S&E, https://repub.eur.nl/pub/120023

Nikulina, A., Interorganizational Governance in Projects: Contracts and collaboration as alignment mechanisms, Promotors: Prof. J.Y.F. Wynstra & Prof. L. Volker, EPS-2021-523- LIS, https://repub.eur.nl/pub/137001

Nullmeier, F.M.E., Effective contracting of uncertain performance outcomes: Allocating responsibility for performance outcomes to align goals across supply chain actors, Promotors: Prof. J.Y.F.Wynstra & Prof. E.M. van Raaij, EPS-2019-484-LIS, https://repub.eur.nl/pub/118723

Okbay, A., Essays on Genetics and the Social Sciences,
Promotors: Prof. A.R. Thurik, Prof. Ph.D. Koellinger & Prof. P.J.F. Groenen, EPS-2017-413-S&E, https://repub.eur.nl/pub/95489

Peng, X., Innovation, Member Sorting, and Evaluation of Agricultural Cooperatives, Promotor: Prof. G.W.J. Hendriks, EPS-2017-409-ORG, https://repub.eur.nl/pub/94976

Petruchenya, A., Essays on Cooperatives: Emergence, Retained Earnings, and Market Shares, Promotors: Prof. G.W.J. Hendriks & Dr Y. Zhang, EPS-2018-447-ORG, https://repub.eur.nl/pub/105243

Plessis, C. du, Influencers: The Role of Social Influence in Marketing, Promotors: Prof. S. Puntoni & Prof. S.T.L.R. Sweldens, EPS-2017-425-MKT, https://repub.eur.nl/pub/103265

Pocock, M., Status Inequalities in Business Exchange Relations in Luxury Markets, Promotors: Prof. C.B.M. van Riel & Dr G.A.J.M. Berens, EPS-2017-346-ORG, https://repub.eur.nl/pub/98647

Polinder, G.J., New Models and Applications for Railway Timetabling, Prof. D. Huisman & Dr. M.E. Schmidt, EPS-2020-514-LIS, https://repub.eur.nl/pub/134600

Pozharliev, R., Social Neuromarketing: The role of social context in measuring advertising effectiveness, Promotors: Prof. W.J.M.I. Verbeke & Prof. J.W. van Strien, EPS-2017-402-MKT, https://repub.eur.nl/pub/95528

Qian, Z., Time-Varying Integration and Portfolio Choices in the European Capital Markets, Promotors: Prof. W.F.C. Verschoor, Prof. R.C.J. Zwinkels & Prof. M.A. Pieterse-Bloem, EPS-2020-488-F&A, https://repub.eur.nl/pub/124984

Ramezan Zadeh, M.T., How Firms Cope with Digital Revolution: Essays on Managerial and Organizational Cognition, Prof. H.W. Volberda & Prof. J.P. Cornelissen, EPS-2021-508-S&E, https://repub.eur.nl/pub/135682

Reh, S.G., A Temporal Perspective on Social Comparisons in Organizations, Promotors: Prof. S.R. Giessner, Prof. N. van Quaquebeke & Dr. C. Troster, EPS-2018-471-ORG, https://repub.eur.nl/pub/114522

Riessen, B. van, Optimal Transportation Plans and Portfolios for Synchromodal Container Networks, Promotors: Prof. R. Dekker & Prof. R.R. Negenborn, EPS-2018-448-LIS, https://repub.eur.nl/pub/105248

Romochkina, I.V., When Interests Collide: Understanding and modeling interests alignment using fair pricing in the context of interorganizational information systems, Promotors: Prof. R.A. Zuidwijk & Prof. P.J. van Baalen, EPS-2020-451-LIS, https://repub.eur.nl/pub/127244

Schie, R. J. G. van, Planning for Retirement: Save More or Retire Later? Promotors: Prof. B. G. C. Dellaert & Prof. A.C.D. Donkers, EOS-2017-415-MKT, https://repub.eur.nl/pub/100846

Schneidmüller, T., Engaging with Emerging Technologies: Socio-cognitive foundations of incumbent response, Promotors: Prof. H. Volberda & Dr S.M. Ansari, EPS-2020-509-S&E, https://repub.eur.nl/pub/131124

Schouten, K.I.M. Semantics-driven Aspect-based Sentiment Analysis, Promotors: Prof. F.M.G. de Jong, Prof. R. Dekker & Dr. F. Frasincar, EPS-2018-453-LIS, https://repub.eur.nl/pub/112161

Sihag, V., The Effectiveness of Organizational Controls: A meta-analytic review and an investigation in NPD outsourcing, Promotors: Prof. J.C.M. van den Ende & Dr S.A. Rijsdijk, EPS-2019-476-LIS, https://repub.eur.nl/pub/115931

Sluga, A., Hour of Judgment: On judgment, decision making, and problem solving under accountability, Promotors: Prof. F.G.H. Hartmann & Dr M.A.S. Boksem, EPS-2021-520-F&A, https://repub.eur.nl/pub/136967

Slob, E., Integrating Genetics into Economics, Promotors: Prof. A.R. Thurik, Prof. P.J.F. Groenen & Dr C.A. Rietveld, EPS-2021-517-S&E, https://repub.eur.nl/pub/135159

Smolka, K.M., Essays on Entrepreneurial Cognition, Institution Building and Industry Emergence, Promotors: P.P.M.A.R. Heugens, & Prof. J.P. Cornelissen, EPS-2019-483-S&E, https://repub.eur.nl/pub/118760

Stirnkorb, S., Changes in the Information Landscape and Capital Market Communication, Promotors: Prof. E. Peek & Prof. M. van Rinsum, EPS-2021-536-F&A, https://repub.eur.nl/pub/136970

Straeter, L.M., Interpersonal Consumer Decision Making,
Promotors: Prof. S.M.J. van Osselaer & Dr I.E. de Hooge, EPS-2017-423-MKT, https://repub.eur.nl/pub/100819

Stuppy, A., Essays on Product Quality, Promotors: Prof. S.M.J. van Osselaer & Dr N.L. Mead. EPS-2018-461-MKT, https://repub.eur.nl/pub/111375

Subaşi, B., Demographic Dissimilarity, Information Access and Individual Performance, Promotors: Prof. D.L. van Knippenberg & Dr W.P. van Ginkel, EPS-2017-422-ORG, https://repub.eur.nl/pub/103495

Suurmond, R., In Pursuit of Supplier Knowledge: Leveraging capabilities and dividing responsibilities in product and service contexts, Promotors: Prof. J.Y.F Wynstra & Prof. J. Dul. EPS-2018-475-LIS,https://repub.eur.nl/pub/115138

Toxopeus, H.S. Financing sustainable innovation: From a principal-agent to a collective action perspective, Promotors: Prof. H.R. Commandeur & Dr. K.E.H. Maas. EPS-2019-458-S&E, https://repub.eur.nl/pub/114018

Turturea, R., Overcoming Resource Constraints: The Role of Creative Resourcing and Equity Crowdfunding in Financing Entrepreneurial Ventures, Promotors: Prof. P.P.M.A.R Heugens, Prof. J.J.P. Jansen & Dr. I. Verheuil, EPS-2019-472-S&E, https://repub.eur.nl/pub/112859

Valboni, R., Building Organizational (Dis-)Abilities: The impact of learning on the performance of mergers and acquisitions, Promotors: Prof. T.H. Reus & Dr A.H.L. Slangen, EPS-2020-407-S&E, https://repub.eur.nl/pub/125226

Vandic, D., Intelligent Information Systems for Web Product Search, Promotors: Prof. U. Kaymak & Dr Frasincar, EPS-2017-405-LIS,https://repub.eur.nl/pub/95490

Verbeek, R.W.M., Essays on Empirical Asset Pricing, Promotors: Prof. M.A. van Dijk & Dr M. Szymanowska, EPS-2017-441-F&A,https://repub.eur.nl/pub/102977

Visser, T.R., Vehicle Routing and Time Slot Management in Online Retailing, Promotors: Prof. A.P.M. Wagelmans & Dr R. Spliet, EPS-2019-482-LIS, https://repub.eur.nl/pub/120772

Vlaming, R. de., Linear Mixed Models in Statistical Genetics, Prof. A.R. Thurik, Prof. P.J.F. Groenen & Prof. Ph.D. Koellinger, EPS-2017-416-S&E, https://repub.eur.nl/pub/100428

Vongswasdi, P., Accelerating Leadership Development: An evidence-based perspective, Promotors: Prof. D. van Dierendonck & Dr H.L. Leroy, EPS-2020-512-ORG, https://repub.eur.nl/pub/134079

Vries, H. de, Evidence-Based Optimization in Humanitarian Logistics, Promotors: Prof. A.P.M. Wagelmans & Prof. J.J. van de Klundert, EPS-2017-435-LIS, https://repub.eur.nl/pub/102771

Wang, R., Corporate Environmentalism in China, Promotors: Prof. P.P.M.A.R Heugens & Dr F. Wijen, EPS-2017-417-S&E, https://repub.eur.nl/pub/99987

Wang, R., Those Who Move Stock Prices, Promotors: Prof. P. Verwijmeren & Prof. S. van Bekkum, EPS-2019-491-F&A, https://repub.eur.nl/pub/129057

Wasesa, M., Agent-based inter-organizational systems in advanced logistics operations, Promotors: Prof. H.W.G.M van Heck, Prof. R.A. Zuidwijk & Dr A. W. Stam, EPS-2017-LIS-424, https://repub.eur.nl/pub/100527

Wessels, C., Flexible Working Practices: How Employees Can Reap the Benefits for Engagement and Performance, Promotors: Prof. H.W.G.M. van Heck, Prof. P.J. van Baalen & Prof. M.C. Schippers, EPS-2017-418-LIS, https://repub.eur.nl/

Wiegmann, P.M., Setting the Stage for Innovation: Balancing Diverse Interests through Standardisation, Promotors: Prof. H.J. de Vries & Prof. K. Blind, EPS-2019-473-LIS, https://repub.eur.nl/pub/114519

Wijaya, H.R., Praise the Lord!: Infusing Values and Emotions into Neo-Institutional Theory, Promotors: Prof. P.P.M.A.R. Heugens & Prof. J.P. Cornelissen, EPS-2019-450-S&E, https://repub.eur.nl/pub/115973

Williams, A.N., Make Our Planet Great Again: A Systems Perspective of Corporate Sustainability, Promotors: Prof. G.M. Whiteman & Dr. S. Kennedy,EPS-2018-456-ORG, https://repub.eur.nl/pub/111032

Witte, C.T., Bloody Business: Multinational investment in an increasingly conflict-afflicted world, Promotors: Prof. H.P.G. Pennings, Prof. H.R. Commandeur & Dr M.J. Burger, EPS-2018-443-S&E, https://repub.eur.nl/pub/104027

Wu, J., A Configural Approach to Understanding Voice Behavior in Teams, Promotors: Prof. D.L. van Knippenberg & Prof. S.R. Giessner, EPS-2020-510-ORG, https://repub.eur.nl/pub/132184

Ye, Q.C., Multi-objective Optimization Methods for Allocation and Prediction, Promotors: Prof. R. Dekker & Dr Y. Zhang, EPS-2019-460-LIS, https://repub.eur.nl/pub/116462

Yuan, Y., The Emergence of Team Creativity: a social network perspective, Promotors: Prof. D. L. van Knippenberg & Dr D. A. Stam, EPS-2017-434-ORG, https://repub.eur.nl/pub/100847

Zhang, Q., Financing and Regulatory Frictions in Mergers and Acquisitions, Promotors: Prof. P.G.J. Roosenboom & Prof. A. de Jong, EPS-2018-428-F&A, https://repub.eur.nl/pub/103871

Zhu, S., Spare Parts Demand Forecasting and Inventory Management: Contributions to Intermittent Demand Forecasting, Installed Base Information and Shutdown Maintenance, Promotors: Prof. R. Dekker & Dr W.L. van Jaarsveld, EPS-2021-538-LIS, https://repub.eur.nl/pub/135684

Zon, M. van, Cost Allocation in Collaborative Transportation, Promotors: Prof. A.P.M. Wagelmans, Dr R. Spliet & Dr W. van den Heuvel, EPS-2021-530-LIS, https://repub.eur.nl/pub/136975

Shared-interest communities – social groups of people gathering around a common interest – represent a centralized source of information, knowledge, social support, and entertainment. Every day, millions of people resort to digital and in-person communities to meet, discuss, solve problems, and even manage disruptive situations, such as natural or civil crises, financial instability, and product recalls. Given the role of shared-interest communities in the lives of consumers, businesses, institutions, and citizens, several streams of literature investigated the antecedents of community participation, success, and resilience. However, it is still unclear how the emergence of new technologies and the occurrence of (often disruptive) external events relate to the success and sustainability of shared-interest communities. In three essays, this dissertation sheds light into the dynamics of shared-interest communities under the influence of changing technologies and external events. Three questions are addressed: (i) What is the impact of digitizing community activities on the participation intentions of community members? (ii) What is the effect of negative vs. positive shocks to a community's purpose on community dynamics?; (iii) What is the effect of a brand crisis on consumer engagement and patterns of information spread in brand communities? This dissertation contributes to the scientific and managerial understanding of the circumstances that allow shared-interest communities to thrive in a complex world. As digitized human interactions become the new norm, and external events prompt dramatic collective action on digital platforms, the findings of this dissertation are both extremely timely and insightful for researchers, practitioners, and policy-makers alike.

ERIM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoekschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are the Rotterdam School of Management (RSM), and the Erasmus School of Economics (ESE). ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focused on the management of the firm in its environment, its intra- and interfirm relations, and its business processes in their interdependent connections.

The objective of ERIM is to carry out first rate research in management, and to offer an advanced doctoral programme in Research in Management. Within ERIM, over three hundred senior researchers and PhD candidates are active in the different research programmes. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.

ERIM

ERIM PhD Series
Research in Management

Erasmus University Rotterdam (EUR)
Erasmus Research Institute of Management
Mandeville (T) Building
Burgemeester Oudlaan 50
3062 PA Rotterdam, The Netherlands

P.O. Box 1738
3000 DR Rotterdam, The Netherlands
T +31 10 408 1182
E info@erim.eur.nl
W www.erim.eur.nl