Social Group Computing: Advancing the Understanding of Human Behavior and Society through the Lens of Online Social Groups

by

Jason Shuo Zhang

M.S., University of Colorado Boulder, 2018

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Computer Science

2020

This thesis entitled:
Social Group Computing: Advancing the Understanding of Human Behavior and Society through the Lens of Online Social Groups
written by Jason Shuo Zhang
has been approved for the Department of Computer Science

Prof. Qin Lv

Prof. Chenhao Tan

Date

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Zhang, Jason Shuo (Ph.D., Computer Science)

Social Group Computing: Advancing the Understanding of Human Behavior and Society through the Lens of Online Social Groups

Thesis directed by Prof. Qin Lv

My doctoral research aims to advance the understanding of human behavior and society by studying the dynamics of online social groups. Through collecting user behavior data from different communities on social media platforms (e.g., Reddit and Foursquare), designing a social dining mobile application OutWithFriendz, and assembling datasets from various offline contexts (e.g., FiveThirtyEight and Forbes), I study different types of intergroup and intragroup interactions. In intergroup interactions, I examine how polarization and tribalism are reflected in online NBA fan communities. Our findings suggest that encouraging NBA fans of different teams to communicate with each other may not help them get along. In intragroup interactions, I use the newly-designed OutWithFriendz mobile application to collect hundreds of real-world social events. Using this dataset, I make several interesting observations to understand how a group of people organize a social event and vote for their preferred meeting times and venues. With the insights gained from these observations, I build a novel recommendation system (GEVR) that provides venue recommendations for newly created group events. The results in my doctoral research emphasize the connection between online and offline worlds. They also inform the design of online platforms for providing more effective and healthy interactions.

For future research, I will keep focusing on understanding human behavior through the lens of online social groups. On the one hand, I will characterize social issues, such as extremism and misinformation, reflected in online social groups, and explore ways to mitigate their effects. On the other hand, I will explore the potential of machines and AI for providing intelligent services for these groups. The ultimate goal is to design a new online discussion environment that motivates positive interactions between people with different ideologies and backgrounds.

Dedication

To all the people who have built castles and let us come in and have fun.

Acknowledgements

The research projects that have gone into this dissertation have been thoroughly pleasurable. This enjoyment is largely a result of the interactions that I have had with my supervisor, advisors, colleagues, and people who have participated in these studies.

I am indebted to my advisor and friend, Qin Lv, for accepting me into her research group, which has been a game-changer for my career. I feel very privileged to have worked with her; she has devoted so much to my Ph.D. study and research. She has infused me with an enthusiasm for research and an open mind for different cultures and opinions. Without her supervision, I would not have attempted and completed these challenging projects.

Another great fortune during my Ph.D. has been being co-advised by Chenhao Tan. I will not forget the enjoyable and illuminating discussions with him at each stage of my Ph.D. career. From Chenhao, I learned not only the technical skills of computer science, but also a dedication and detail-driven spirit for research.

In addition, this journey would not have been fruitful without my wonderful committee members, collaborators, and colleagues–Richard Han, Shivakant Mishra, Scott Holman, Khaled Alanezi, Mike Gartrell, Brian Keegan, Tamara Lehman, Eric Rozner, Nolen Scaife, Hoang Truong, and Tam Vu. I also appreciate all my lab mates–Vivian Lai, Yawen Zhang, Wanshan Yang, Lei Tian, Yichen Wang, Tao Ruan, Chao-Chun Hsu, and Han Liu for lighting up this venture.

Last but not least, I humbly extend my gratitude to my parents Xiaoning Feng and Gangyue Zhang, who have supported me and continue to support me throughout my entire life. The freedom, encouragement, and assistance they offer are the reasons that I started this journey.

Contents

Tables

Table

Figures

Figure

Chapter 1

Introduction

Alone we can do so little; together we can do so much.

— Helen Keller

An important aspect that motivates this thesis research is the emergence of online social groups. A social group, by definition, is an aggregation of individuals that might form around a shared value, interest, or goal. A growing body of research has shown that people are hardwired to be social. We, as human beings, want to be in a group and interact in groups.

Thirty years ago, most social groups were constructed locally. This can be seen as a group of colleagues meeting at a barbeque party, members of a religious organization meeting at a local church, or sports fans gather together at a stadium to cheer on their favorite teams. These social groups are usually subject to geographic constraints.

With the development of the Internet and technology, the ability to create online groups and share experiences has been exponentially greater. Now we can easily connect with like-minded individuals on a variety of technology platforms to accomplish our goals (Figure 1.1). Take Facebook as an example; there are over 1.4 billion people that use the Facebook Group feature every month.

In many ways, these online groups serve similar purposes that offline groups provide, but break the barrier of time, space, and scale that limit offline interactions. They provide members with opportunities for information sharing and learning. For lonely people, they provide companionship and social support. For active users, these online groups are also great places for entertainment

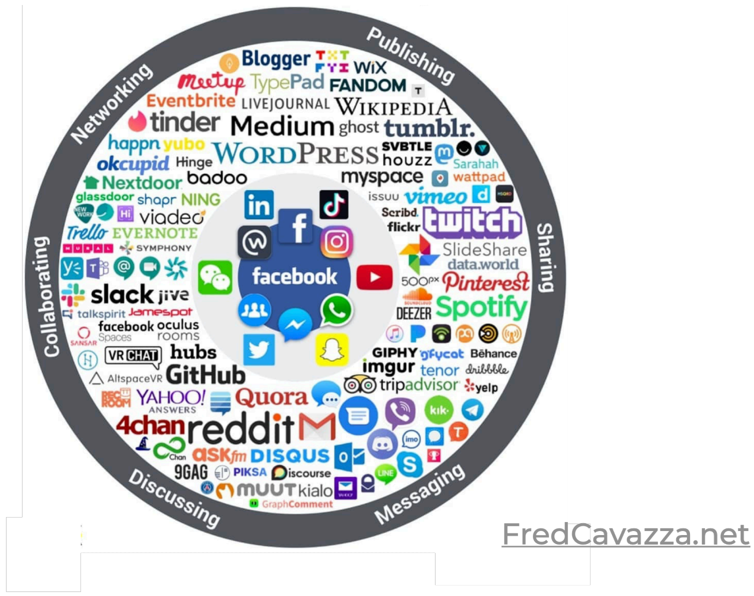


Figure 1.1: Social media landscape in 2019.

purposes. These collective actions also produce a huge benefit for society by creating encyclopedia pages, product reviews, and crowdfunding events.

Because of that, early researchers may have seen the Internet as having the potential to bring people together by bridging cultural, racial, and national divides among different groups. However, the process of bringing people with different backgrounds together has also led to different social issues, from tensions and passive-aggressiveness to harassment and cyberbullying. Moreover, these social groups can also create echo-chambers, which are dominated by like-minded individuals. These echo-chambers can exacerbate polarization and misinformation problems on the Internet.

The goal of this thesis research is to advance the understanding of human behavior and society by studying the dynamics of online social groups. Towards this direction, we collect large-scale datasets from multimodal data resources. For research problems that cannot be answered by existing or available datasets, we also design social systems to collect novel data from scratch. Using all this data, we design quantitative observational studies to generate actionable insights about a variety of social behaviors. Additionally, we connect these observations with theories in social science to better explain them. The ultimate goal is to use all of these insights to improve the design of social systems for supporting more productive and healthy conversations on the Internet. In light of these opportunities and challenges, we will now summarize the most important topics that this thesis will focus on.

1.1 How are Online Sports Communities Shaped by the Offline Context?

Many online communities do not only exist in the virtual world but are also deeply embedded in the offline context and attract people with similar interests. Professional sports provides an interesting case because these online fan communities, in a way, only exist as a result of offline sports teams and games. Fascinated by the strong connection of the offline NBA season and fan behavior in online communities, we construct a large-scale dataset from Reddit, which combines 1.5M posts and 43M comments in online NBA fan communities with offline statistics that document franchise value, team performance, star players, etc. We study interactions of team performance and online fan behavior at the game- and season-level. Our hierarchical regression analyses reveal a series of interesting behavior patterns that are consistent across multiple NBA seasons. For example, the fans of the top teams tend to be more active when their supported team is losing, while the fans of the bottom teams tend to engage more in discussions when their team is winning. Besides, team performance also drives what fans talk about in these online communities. Better teams discuss "season prospects" and "championships" more, and weak teams spend a large amount of time envisioning the "future."

1.2 A Computational Framework to Characterize Tribalism and Polarization on the Internet

In recent years, there are growing concerns about tribalism and polarization in world politics. On popular social media sites, such as Twitter and Reddit, liberals and conservatives tend to cluster according to their political beliefs and rarely talk with each other. To reduce prejudice and build bridges between people with different ideologies, intergroup contact has long been considered a simple yet potentially powerful strategy in the offline setting by psychologists. In practice, there are social media platforms testing this strategy of exposing people to opposing groups/views to reduce polarization.

However, it remains unclear if blindly implementing this offline strategy online can achieve these goals. In this study, we leverage 2.1 million posts and 61 million comments from NBA-related communities on Reddit and design a computational framework to study how interactions between fans of different teams relate with their language usage. We choose sports context as our testbed as sports fans can be really tribal and treat rival team fans almost as enemies, just like people with different political ideologies. My results show that mixing fans of opposing teams together may not moderate their behavior, and it may backfire. The more these fans interact, the more negative their comments are even when they discuss with fans of the home team.

The findings provide implications for thinking about encouraging intergroup contact on the Internet. To bring people together and reduce polarization, just exposing them to opposing groups/views may not work. People who choose to have intergroup contact can be argumentative, which lead to toxic interactions, making people more extreme after the exposure. To improve the design of social media platforms for healthy conversations between groups, we need to think about how we can encourage open attitudes and calm conversations.

1.3 Understanding Group Event Organization

The wide adoption of smartphones and mobile applications has brought significant changes to not only how individuals behave in the real world, but also how groups of users interact with each other when organizing social group events. There is a rich history of research concentrating on individual user behavior analysis. However, social interactions among group members are often ignored. Understanding how users make event decisions as a group online and identifying the contributing factors can offer important insights for social group studies and a more effective system design for group event scheduling.

To address this topic, we design OutWithFriendz, a mobile system that enables a group of people to decide when and where to meet through consensus polls. OutWithFriendz provides an online virtual platform where group members exchange thoughts and share preferences.

With the OutWithFriendz mobile system, we collect more than 200 legitimate events. Our analysis results disclose a series of deciding factors that correlate with the success of a group event (measured by attendance rate). These factors include voting patterns, user mobility, individual preference, and host preference. Interestingly, we find that early voters tend to vote for a wide variety of date/venue options, while late voters report limited availability. Even though the late voters still prefer venues near their frequented places, they would sometimes hide their real preferences and go with the ones that are widely accepted by other group members. These observations align with the famous groupthink phenomenon in psychology: when people negotiate in a group, they will set aside their personal beliefs and adopt the opinion of the rest of the team.

1.4 GEVR: An Event Venue Recommendation System for Groups

After identifying the factors that would affect the group decision-making process, we design a Group Event Venue Recommendation (GEVR) system that incorporates user mobility and context information to provide venue recommendations for newly created group events. This problem is more challenging than recommending items to individual users, as the group members may have diverse spatial and temporal preferences. Groups are also dynamic, as group memberships may change fluidly from event to event.

To tackle the complexity of group recommendation, we propose a novel group location cluster prediction model, which dynamically assigns different group decision strategies to groups based on the groups members' social strength.

The performance of GEVR is uniquely evaluated with over 500 real-world group events collected using the OutWithFriendz mobile system. GEVR can provide an 80% accuracy in predicting which location cluster the group will meet at, outperforming all the baselines by a significant margin. More importantly, our evaluation results indicate that one straightforward group decision strategy, e.g., average satisfaction, does not apply to all groups. Groups with strong social relationships usually apply different strategies compared to the ones with weak social relationships.

Chapter 2

Background and Related Work

In this chapter, we cover prior research studying the evolution of online communities, sports fan behavior, interactions between groups, and group recommendation.

2.1 Online Communities

The proliferation of online communities has enabled a rich body of research in understanding group formation and community dynamics [7, 154, 90, 97]. Most relevant to our work are studies that investigate how external factors affect user behavior in online communities [136, 171, 156, 214]. Palen and Anderson [136] provide an overview of studies on social media behavior in response to natural disasters and point out limits of social media data for addressing emergency management. Romero et al. [156] find that communication networks between traders "turtle" up during shocks in stock price and reveal relations between social network structure and collective behavior. Other offline events studies include the dynamics of breaking news [92, 91, 101], celebrity death [93, 56], and Black Lives Matter [192, 177]. This literature illustrates that online communities do not only exist in the virtual world. They are usually deeply embedded in the offline context in our daily life.

Another relevant line of work examines user engagement in multiple communities and in particular, user loyalty [183, 221, 73]. Hamilton et al. [73] operationalize loyalty in the context of multi-community engagement and consider users loyal to a community if they consistently prefer the community over all others. They show that loyal users employ language that signals collective identity and their loyalty can be predicted from their first interactions.

Reddit has attracted significant interest from researchers in the past few years due to its growing importance. Many aspects and properties of Reddit have been extensively studied, including user and subreddit lifecycle in online platforms [183, 131], hate speech [24, 25, 160], interaction and conflict between subreddits [99, 182], and its relationship with other web sources [193, 131]. Studies have also explored the impacts of certain Reddit evolutions and policy changes on user behaviors. Notable events include pre-default subreddit [106] and Reddit unrest [131, 24, 118].

Our work examines a special set of online communities that derived from professional sports teams. As a result, regular sports games and team performance are central for understanding these communities and user loyalty in these communities. Different from prior studies, we focus on the impact of team performance on user behavior in online fan communities.

2.2 Sports Fan Behavior

As it is crucial for a sports team to foster a healthy and strong fan base, extensive studies in sports management have studied fan behavior. Researchers have studied factors that affect purchasing behavior of sports fans [168, 199, 190], including psychometric properties and fan motivation. A few studies also build predictive models of fan loyalty [15, 213]. Bee and Havitz [15] suggest that fan attraction, involvement, psychological commitment, and resistance can be predictors of fan behavioral loyalty. Dolton and MacKerron [43] estimate that the happiness that fans feel when their team wins is outweighted by the sadness that strikes when their team loses by a factor or two. Yoshida et al. [213] build regression models based on attitudinal processes to predict behavioral loyalty. The potential influence of mobile technology on sports spectators is also examined from different angles [189, 111, 83]. Torrez Riley [189] describes survey results that suggest the current usage of mobile technology among college sports fans. The work by Ludvigsen and Veerasawmy [111] examines the potential of interactive technologies for active spectating at sporting events. Most relevant to our work are studies related to fan identification [22, 48, 49, 79, 82, 89, 176, 180, 198] and we have discussed them to formulate our hypotheses. These studies usually employ qualitative methods through interviews or small-scale surveys.

It is worth noting that fan behavior can differ depending on the environment. Cottingham [34] demonstrates the difference in emotional energy between fans in sports bars and those attending the game in the stadium. In our work, we focus on online communities, which are an increasingly important platform for sports fans. These online fan communities also allow us to study team performance and fan behavior at a much larger scale than all existing studies.

2.3 Language as a Lens of Human Behavior

The proliferation of textual content online has inspired a vast body of literature to understand the language in online communication and its relationship with individual attributes. Prior research in CSCW and related communities has investigated how language can reflect properties of individuals [187, 129, 40, 138, 162, 147, 183]. For instance, Toma and Hancock [187] show that linguistic emotions correlate with deception in online dating profiles; De Choudhury et al. [40] uses linguistics style features to show that mothers with post-partum depression are more likely to use first-person singular pronouns and swear words; Naaman et al. [129] conduct a quantitative analysis of message content from over 350 Twitter users to characterize the type of messages posted on the platform and broadly classify users as self-broadcasters and informers. In general, users' demographic information and personality can also be predicted based on linguistic features extracted from textual social media data [122, 123, 115, 61].

Recently, negative language use has been examined in the context of antisocial behavior in online communities [30, 16, 163, 24, 29, 39, 28]. Conceptually related is a prior study on the effects of community feedback on user behavior, which reveals that negative feedback can lead to future antisocial behavior [28]. Cheng et al. [30] further design an experiment that shows negative mood expressed from textual content increases the likelihood of trolling in online platforms. A supervised learning model proposed by Blackburn and Kwak [16] indicates that negative sentiments are useful for predicting toxic players in online games. In this work, we compare the differences in language usage between intergroup and single-group members, with a focus on expressions of emotions.

2.4 Tribalism, Echo Chambers, and Polarization

A battery of studies in social sciences has shown that human behavior is shaped by our need to belong to a group and by our proclivity to hate rival groups [144, 181, 33]. Such behavior has been documented in a wide variety of contexts. In the political context, for instance, recent studies find that "liberal group" and "conservative group" on social media not only rarely talk to each other, but also use different hashtags and links to various websites within their tweets [116, 169, 69, 66, 141]. Another commonly studied context is brand communities in the marketing literature [35, 128, 78, 13, 70]. For example, Hickman and Ward [78] show that in-groups are strongly motivated to develop negative views of out-groups and engage in "trash talking" about out-groups. In-group members will also gain pleasure at the misfortune of rival brands and their users.

In the sports context, the team sports literature focus on the negative consequences of rivalry, such as negative explicit and implicit attitudes towards the opposing team [33, 103], *schadenfreude* [74, 32], and even riot [71]. These negative perceptions may even transfer to the sponsors of the rival team: Dalakas and Levin [37] explore the negative sponsorship effects and find that sponsors of disliked NASCAR drivers are viewed less positively than sponsors of liked drivers. Similarly, Olson [134] finds that brands faced a steep decline in sales among Manchester City fans when they announced the sponsorship of the soccer club Manchester United, a fierce rival of Manchester City. By examining the intergroup emotions of fans of the Boston Red Sox and New York Yankees, Lehr et al. [103] show that pleasure from a powerful rival's losses can outstrip that from gains of the supported team. Given the competitive nature of professional sports and the importance of emotions in fan behavior, we believe that professional sports provide exciting opportunities for understanding polarization.

2.5 Intergroup Contact

Intergroup contact has long been considered as an effective strategy to reduce prejudice between groups [47]. For instance, a seminal work by Pettigrew [143] shows that intergroup contact relates to reduced prejudice towards immigrants based on self-reported surveys in France, Great Britain, the Netherlands, and West Germany. Wright et al. [209] find correlational evidence that people who knew that an in-group member had an out-group friend had less negative intergroup attitudes. They also experimentally demonstrate that providing this information induces more positive attitudes. Abbott and Cameron [2] examine young people's assertive bystander intentions in an intergroup (immigrant) name-calling situation and find that greater intergroup contact is related to higher levels of empathy, higher levels of cultural openness, and reduced intergroup bias. From the perspective of language usage online, a field experiment designed by White et al. [202] demonstrates that Muslim and Christian high-school students who have structured Internet intergroup interactions tend to use more affective and positive emotion words, and less anger and sadness words. Kim and Wojcieszak [98] test online contact with two distinct out-groups, undocumented immigrants and gay people. They find that direct online contact improves attitudes towards both out-groups through positive and negative emotions, whereas extended online contact reduces negative emotions and improve attitudes towards undocumented immigrants.

However, recent studies on the "backfire" effect suggest that exposure to opposing groups in online platforms can exacerbate political polarization [10, 133, 9, 102]. [1] For instance, Bail et al. [10] introduce intergroup contact by following a Twitter bot that aggregates tweets of opinion leaders from the opposing political ideology and find that Republicans who follow a liberal Twitter bot become substantially more conservative. Lee et al. [102] use panel data collected in South Korea to investigate the effects of social media usage on changing the political view. They highlight the role of social media in activating political participation and pushing users toward ideological poles.

A possible way to reconcile such differences in prior literature is to review the mechanisms that

[1] Wood and Porter [208] show that backfire in Nyhan and Reifler [133] is stubbornly difficult to reproduce, which further demonstrates the varying results in recent studies.

contribute to the positive effects of intergroup contact: (1) enhancing knowledge about other groups, (2) reducing anxiety when facing opposing groups, and (3) increasing empathy and perspective-taking [144, 145, 47, 45, 143, 146, 209]. Depending on the motivations to engage in intergroup contact and the actual activities during the contact, intergroup contact in online platforms may not necessarily achieve these goals. We aim to conduct a large-scale observational study to understand the differences between intergroup and single-group members in their original affiliated group, and also provide some insights on the nature of intergroup contact in /r/NBA.

It is useful to point out that there is little work on intergroup contact in the CSCW community. In the meanwhile, several recent studies provide a characterization of intergroup conflict. Kumar et al. [99] examine cases of intergroup conflict across 36,000 communities on Reddit where users of one community are mobilized by negative sentiment to comment in another community and show that less than 1% communities start 74% conflicts. At the community level, by constructing a conflict network between subreddits, Datta and Adar [38] find that larger subreddits are more likely to be involved in conflicts with a large number of subreddits, and the main "targets" change over time. However, intergroup contact is different from intergroup conflict as it may help improve intergroup attitudes and reduce intergroup tensions and conflicts. Approaching this topic from a CSCW lens raises additional questions about how socio-technical design decisions can influence the outcomes reported in traditional offline settings.

2.6 Group Behavior Analysis

Group Behavior is a hot research topic in recent years. Prior work has shown that people behave significantly different when they are within a group compare to when they are on their own. The work by Jayarajah **et al.** conducted a large-scale group study involving more than 6000 users location traces at a university campus over a four-month period [88]. By analyzing the dataset, they found that individuals vs groups exhibit significant difference in terms of three aspects: mobility pattern, the responsiveness to phone calls and messages, and mobile application usage. More specifically, they found that while in a group

(1) individuals tend to spend more time at places and are less likely to move to another place;

(2) people's tendency to respond/initiate phone call/messages drops sharply; and

(3) individuals check their phone applications more often, but restrict themselves to shorter usage duration.

Jamil **et al.** detect group formation and dynamics at a large religious event: the Hajj pilgrimage in 2014 [86]. Their analysis reveals that group variations and community structures have different identities based on demographic characteristics. Clear bias is observed in sub-communities based on the gender and age group of the pilgrims. Their work also illustrates key issues in the overall management existing in large-scale event organization and suggests possible measures to tackle these issues accordingly.

There are also studies concentrating on studying group collocated interactions using mobile devices [52, 110]. Fischer **et al.** study ways in which participants organize the management of mobile notifications with their collocated teammates. It reveals that notification management within groups routinely feature ignoring, content sharing, negotiating concurrent activity and filtering [52].

2.7 Group-based Social Influence

People's decisions within a group are certainly influenced by other group members. A wide range of studies focus on studying the impact of social influence in user's group behavior. Romero **et al.** conducted an analysis using the Doodle Poll dataset [157]. In this paper, they found that information cascade effects take place during event scheduling. And early respondents will have a larger influence on the outcome of a poll than people who come later. They further simulate possible interventions to reduce this bias and provide a few suggestions for optimizing the success of polls in a social network. Goyal **et al.** studies how to model social influence propagation in social networks by forming an action log [63]. Li **et al.** incorporates friend and foe relationships into a model to provide higher accuracy [105]. These two works treated every group member equally, which means every node carries the same weight in the social relation graph. The article in [222] went a step further. They illustrate that social influence of one user on another user can differ significantly from one group to another (eg. a technology group vs. a movies group). To utilize this group-specific information, they propose a new social propagation network to model people's social influences that are specific to the event's group organizers.

2.8 Group Recommendation Systems

The problem of making recommendations to a group of users has been investigated in a number of works [58, 26, 148, 215, 216, 217]. Gartrell **et al.** propose a group recommendation model that utilizes social and content interests of group members to recommend movies for groups of users [58]. The work by Cheney **et al.** presents a large-scale study of television viewing habits. They provide an analysis of how the viewing patterns shift across various group contexts and discussed the impact of these findings on the performance of a group TV program recommendation system [26]. Quintarelli **et al.** further models users' ability to direct a group's decision by using contextual influence and aggregating it to recommend TV programs for groups of users [148]. All of these works can be summarized as recommending items for groups of users, where user mobility is not much of a concern. Our problem regarding group event venue recommendation is more challenging as users' mobility patterns can play a significant role in group decision making. Before the actual event gathering, group members may be distributed in different geographical areas. Thus, recommending a venue that will optimize group satisfaction is more challenging and requires delicate data analysis and system design.

Chapter 3

Characterizing Online NBA Fan Communities

3.1 Introduction

Our [The Los Angeles Lakers'] collective success having forged some kind of unity in this huge and normally fragmented metropolis, it cuts across cultural and class lines.

— Kareem Abdul-Jabbar

Professional sports not only involve competitions among athletes, but also attract fans to attend the games, watch broadcasts, and consume related products [201]. For instance, the 2017 final game of the National Basketball Association (NBA) attracted 20 million TV viewers [173]; a 30-second commercial during the Super Bowl cost around 4.5 million dollars in 2015 and these commercials have become an integral part of American culture [205].[1] Fans of sports teams can be very emotionally invested and treat fans of rival teams almost as enemies, which can even lead to violence [155].

Such excitement towards professional sports extends to online communities. A notable example is /r/NBA on Reddit, which attracts over a million subscribers and has become one of the most active subreddits on Reddit, a popular community-driven website [153]. A sports writer has suggested that online fan communities are gradually replacing the need for sports blogs and even larger media outlets altogether [62]. The growth of online fan communities thus provides exciting

[1] Most influential Super Bowl commercials: `http://time.com/4653281/super-bowl-ads-commercials-most-influential-time/`.

opportunities for studying fan behavior in professional sports at a large scale.

It is important to recognize that fan behavior is driven by sports events, including sports games, player transfers between teams, and even a comment from a team manager. The dynamic nature of sports games indicates that discussions in online fan communities may echo the development in games, analogous to the waves of excitement in a stadium. Therefore, our goal in this paper is to characterize **online** fan communities in the context of **offline** games and team performance.

To do that, we build a large-scale dataset of online fan communities from Reddit with 479K users, 1.5M posts, and 43M comments, as well as statistics that document offline games and team performance.[2] We choose Reddit as a testbed because 1) Reddit has explicit communities for every NBA team, which allows us to compare the differences between winning teams and losing teams; and 2) Reddit is driven by fan communities, e.g., the ranking of posts is determined by upvotes and downvotes of community members. In comparison, team officials can have a great impact on a team's official Twitter account and Facebook page.

Organization and highlights. We first provide an overview of the NBA fan communities on Reddit as well as necessary background knowledge regarding the NBA games (Section 6.1). We demonstrate the seasonal patterns in online fan communities and how they align with the NBA season in the offline world. We further characterize the discussions in these NBA fan communities using topic modeling.

We investigate three research questions in the rest of the paper. First, we study the relation between team performance in a game and this game's associated fan activity in online fan communities. We are able to identify game threads that are posted to facilitate discussions during NBA games. These game threads allow us to examine the short-term impact of team performance on fan behavior. We demonstrate intriguing contrasts between top teams and bottom teams: user activity increases when top teams lose and bottom teams win. Furthermore, close games with small point differences are associated with higher user activity levels.

Second, we examine how team performance influences fan loyalty in online communities

[2] The dataset is available at `http://jasondarkblue.com/papers/CSCW2018NBADataset_README.txt`.

beyond a single game. It is important for professional teams to acquire and maintain a strong fan base that provides consistent support and consumes team-related products. Understanding fan loyalty is thus a central research question in the literature of sports management [49, 176, 213, 48]. For instance, "bandwagon fan" refers to a person who follows the tide and supports teams with recent success. Top teams may have lower fan loyalty due to the existence of many bandwagon fans. Our results validate this hypothesis by using user retention to measure fan loyalty. We also find that a team's fan loyalty is correlated with the team's improvement over a season and with the average age of the roster.

Third, we turn to the content in online fan communities to understand the impact of team performance on the topics of discussion. Prior studies show that a strong fan base can minimize the effect of a team's short-term (poor) performance on its long-term success [180, 164]. To foster fan identification in teams with poor performance, fans may shift the focus from current failure to future success and "*trust the process*"[3] [48, 22, 89]. Discussions in online fan communities enable quantitative analysis of such a hypothesis. We show that fans of the top teams are more likely to discuss *"season prospects,"* whiles fans of the bottom teams are more likely to discuss *"future."* Here *"future"* refers to the assets that a team has, including talented young players, draft picks, and salary space, which can potentially prepare the team for future success in the following seasons.

We offer concluding discussions in Section 4.7. Our work develops the first step towards studying fan behavior in professional sports using online fan communities and provides implications for online communities and sports management. For online communities, our results highlight the importance of understanding online behavior in the offline context. Such offline context can influence the topics of discussion, the activity patterns, and users' decisions to stay or leave. For sports management, our work reveals strategies for developing a strong fan base such as shifting the topics of discussion and leveraging unexpected wins and potential future success.

[3] A mantra that reflects Philadelphia 76ers' identity [151] 76ers went through a streak of losing seasons to get top talents in draft-lottery and rebuild the team.

Table 3.1: Dataset Statistics. There are in total 30 teams in the NBA league. #users refers to the number of unique users who posted/commented in the subreddit.

	/r/NBA	Average of team subreddits (std)
#users	400K	13K (8K)
#posts	847K	24K (16K)
#comments	33M	328K (282K)

3.2 An Overview of NBA Fan Communities on Reddit

Our main dataset is derived from NBA-related communities on Reddit, a popular website organized by communities where users can submit posts and make comments. A community on Reddit is referred to as a *subreddit*. We use community and subreddit interchangeably in this paper. We first introduce the history of NBA-related subreddits and then provide an overview of activity levels and discussions in these subreddits.

3.2.1 NBA-related Subreddits

On Reddit, the league subreddit /r/NBA is for NBA fans to discuss anything that happened in the entire league, ranging from a game to gossip related to a player. There are 30 teams in total in the NBA league, and each team's subreddit is for fans to discuss team-specific topics. Each subreddit has multiple moderators to make sure posts are relevant to the subreddit's theme. We collected posts and comments in these 31 subreddits (/r/NBA + 30 NBA team subreddits) from pushshift.io [12] from the beginning of each subreddit until October 2017. [4] The overall descriptive statistics of our dataset are shown in Table 6.1.

A brief history of the NBA-related subreddits on Reddit. NBA-related subreddits have thrived since January 2008, when Reddit released a new policy to allow users to create their own subreddit. The Lakers' and the Celtics' subreddits were created by fans in 2008, and they are the first two NBA teams to have their team subreddits. These two teams are also widely

[4] A small amount of data is missing due to scraping errors and other unknown reasons with this dataset [57]. We checked the sensitivity of our results to missing posts with a dataset provided by J.Hessel; our results in this paper do not change after accounting for them.

acknowledged as the most successful franchises in the history of the NBA league [68]. It is also worth noting that these two teams' subreddits were created even before /r/NBA, which was created at the end of 2008. For the remaining 28 teams, 14 of their subreddits were created by users in 2010, and the other 14 were created in 2011. Moreover, three teams' subreddit names have changed. The Pelicans' subreddit changed their subreddit name from /r/Hornets to /r/Nolapelicans and the Hornets' subreddit from /r/Charlottebobcats to /r/Charlottehornets because these two teams changed their official team names. Additionally, the Rockets' subreddit shortened its name from /r/Houstonrockets to /r/Rockets. To rebuild each team's complete subreddit history, we combined posts and comments in these three teams' old and new subreddits. Figure 3.1a presents the number of users that post and comment in each team subreddit.

3.2.2 NBA Season Structure Reflected in Reddit Activity

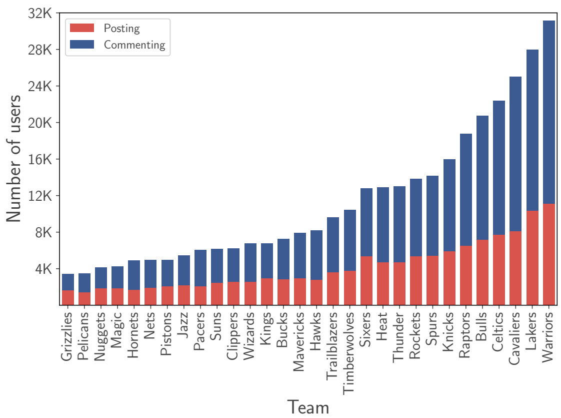
As discussed in the introduction, fan behavior in NBA-related subreddits is influenced by offline events. In particular, the NBA runs in seasons, and seasonal patterns are reflected on Reddit. To show that, we start with a brief introduction of the NBA. 30 teams in the NBA are divided into two conferences (East and West). In each season, teams play with each other to compete for the final championship. The following three time segments make up a complete NBA season: [5]

- **Off season**: from July to mid October. There are no games in this period.[6] Every team is allowed to draft young players, sign free agents, and trade players with other teams. The bottom teams in the last season get the top positions when drafting young players, which hopefully leads to a long-term balance between teams in the league. The goal of the off season for each team is to improve its overall competitiveness for the coming season.

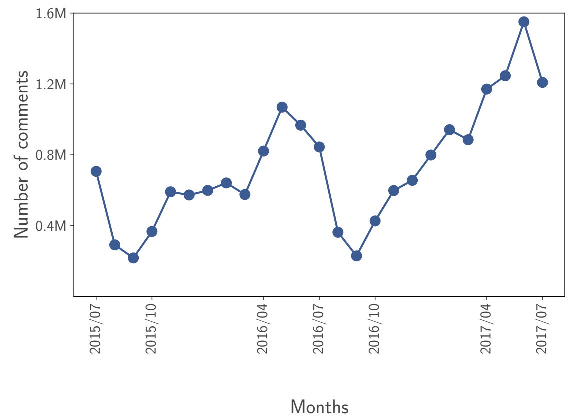
- **Regular season**: from late October to middle of April. Regular season games occur in

[5] Please see the NBA's official description for more details [6].

[6] There are Summer League games and preseason games played in this period, but the results don't count in season record.



(a) Number of users in team subreddits.



(b) User activity in /r/NBA by month.

Figure 3.1: Figure 3.1a shows the number of users that post and comment in team subreddits. /r/Warriors, /r/Lakers, /r/Cavaliers are the top three subreddits in both posting and commenting. The average number of users across all teams is 4,166 for posting and 11,320 for commenting. Figure 3.1b shows user activity level in /r/NBA by month. During the off season (July-mid October), user activity decreases sharply, as no games are played during this period. Then in the regular season (late October to next March), user activity increases steadily. The activity of /r/NBA peaks in May and June, as the championship games happen in these two months.

this period. Every team has 82 games scheduled during this time, 41 home games and 41 away games. A team's regular season record is used for playoff qualification and seeding.

- **Playoff season**: from the end of regular season to June. 16 teams (the top 8 from the Western conference and the top 8 from the Eastern conference) play knockouts in each conference and compete for the conference championship. The champion of the Western Conference and the Eastern Conference play the final games to win the final championship.

Given the structure, a complete NBA season spans two calendar years. In this paper, for simplicity and clarity, we refer to a specific season by the calendar year when it ends. For instance, the official 2016-2017 NBA season is referred to as **the 2017 season** throughout the paper.

User activity in NBA-related subreddits is driven by the structure of the NBA season. As an example, Figure 3.1b shows user activity in /r/NBA by month in the 2016 and 2017 season. From July to September, user activity decreases sharply as there are no games during this period. Then from October to the next March, the number of comments increases steadily as the regular season unfolds. According to the NBA rules, every game in the regular season carries the same weight for playoff qualification, the games in October should be equally important as the games in March. However, fans are much more active on Reddit as it gets closer to the end of the regular season because they deem these games "more critical." This may be explained by the "deadline pressure" phenomenon in psychology [4]. This circumstance has also been observed in other sports. For instance, Paton and Cooke [140] illustrate that the attendance of domestic cricket leagues in England and Wales is much higher in the later segment of the season than the earlier segment. Hogan et al. [80] find that the possibility of the home team reaching the knock-out stage had a significant positive impact on attendance in the European Rugby Cup. We also find that user activity drops a little bit in April in both the 2016 and 2017 season. One possible explanation is that as the regular season is ending, fans of the bottom teams that clearly cannot make the playoffs reduce their activity during this period. After that, the volume of comments increases dramatically during the playoff games. The activity of /r/NBA peaks in May and June, when the conference

Table 3.2: The top five topics by LDA using all the comments in /r/NBA. The top ten weighted words are presented for each topic. In preprocessing, all team and player names are removed. The remaining words are converted to lower case and lemmatized before training the LDA model.

LDA topic	top words	average topic weight
"personal opinion"	opinion, fact, reason, agree, understand, medium, argument, talking, making, decision	0.083
"game strategy"	defense, offense, defender, defensive, shooting, offensive, shoot, open, guard, post	0.082
"season prospects"	final, playoff, series, won, championship, beat, winning, west, east, title	0.078
"future"	pick, trade, star, top, chance, young, future, move, round, potential	0.075
"game stats"	top, number, league, stats, mvp, average, career, assist, put, shooting	0.075

championship and final championship games happen.

3.2.3 Topic analysis

To understand what fans are generally talking about in NBA-related subreddits, we use Latent Dirichlet Allocation (LDA) [17], a widely used topic modeling method, to analyze user comments. We treat each comment as a document and use all the comments in /r/NBA to train a LDA model with the Stanford Topic Modeling Toolbox [54]. We choose the number of topics based on perplexity scores [195]. The perplexity score drops significantly when the topic number increases from 5 to 15, but does not change much from 15 to 50, all within 1370-1380 range. Therefore, we use 15 topics in this paper. Table 3.2 shows the top five topics with the greatest average topic weight and the top ten weighted words in each topic. Two authors, who are NBA fans and active users on /r/NBA, manually assigned a label to each of the five most frequent topics based on the top words in each topic. Each label in Table 3.2 summarizes the topic's gist, and the five labels are *"personal opinion," "game strategy," "season prospects," "future,"* and *"game stats."* We describe our preprocessing procedure and present the other ten topics in Section 3.7.1.

3.3 Research Questions and Hypotheses

We study three research questions to understand how team performance affects fan behavior in online fan communities. The first one is concerned with team performance in a single game and that game's associated user activity, while the other two questions are about team performance in a season and community properties (fan loyalty and the topics of discussion).

3.3.1 Team Performance and Game-level Activity

An important feature of NBA-related subreddits is to support game related discussion. In practice, each game has a game thread in the home-team subreddit, the away-team subreddit, and the overall /r/NBA. Team performance in each game can have a short-term impact on fans' behavior. For instance, Leung et al. [104] show that losing games has a negative impact on the contributions to the corresponding team's Wikipedia page, but winning games does not have a significant effect. However, it remains an open question how team performance in games relate to user activity in *online sports fan communities*.

Previous studies find that fans react differently to top teams than to bottom teams based on interviews and surveys [48, 180, 213]. In particular, Doyle et al. [48] find that fans of teams with an overwhelming loss to win ratio can be insensitive to losses through interviews. In contrast, fans that support top teams may be used to winning. The hype created by the media and other fans elevates the expectation in the fan community. As a result, losing can surprise fans of the top teams and lead to a heated discussion. Therefore, we formulate our first hypothesis as follows:

H1: In subreddits of the top teams, fans are more active on losing days; in subreddits of the bottom teams, fans are more active on winning days.

3.3.2 Team Performance and Fan Loyalty

Researchers in sports management show that a team's recent success does not necessarily lead to a loyal fan base [22, 15, 176]. For instance, "bandwagon fan" refers to individuals who become fans of a team simply because of their recent success. These fans tend to have a weak attachment to the team and are ready to switch to a different team when the team starts to perform poorly. On the contrary, in the bottom teams, active fans that stay during adversity are probably loyal due to their deep attachment to the team [48, 22]. They are able to endure current stumbles and treat them as a necessary process for future success. Our second hypothesis explores the relation between team performance and fan loyalty:

H2: Top team subreddits have lower fan loyalty and bottom team subreddits have higher fan loyalty.

3.3.3 Team Performance and Topics of Discussion

In addition to whether fans stay loyal, our final question examines what fans talk about in an online fan community. As a popular sports quote says, "*Winning isn't everything, it's the only thing.*"[7] A possible hypothesis is that that the discussion concentrates on winning and team success. However, we recognize the diversity across teams depending on team performance. Several studies find that fans of teams with poor performance may shift the focus from current failure to future success: staying optimistic can help fans endure adversity and maintain a positive group identity [48, 22, 89]. This is in clear contrast with the focus on winning the final championship of the top teams [22]. As a result, we formulate our third hypothesis as follows:

H3: The topics of discussion in team subreddits vary depending on team performance. Top team subreddits focus more on *"season prospects"*, while bottom team subreddits focus more on *"future"*.

3.4 Method

In this section, we first provide an overview of independent variables and then discuss dependent variables and formulate linear models to test our hypothesis in each research question.

[7] Usually attributed to UCLA football coach Henry Russel Sanders.

3.4.1 Independent Variables

To understand how team performance affects fan behavior in online fan communities, we need to control for factors such as a team's market value and average player age. We collect statistics of the NBA teams from the following websites: Fivethirtyeight, [8] Basketball-Reference, [9] Forbes. [10] and Wikipedia. [11] We standardize all independent variables for linear regression models. Table 3.3 provides a full list of all variables used in this paper. In addition to control variables that capture the differences between seasons and months, the variables can be grouped in three categories: performance, game information, and team information.

3.4.1.1 Performance

Since our research questions include both team performance in a single game and team performance over a season, we consider performance variables both for a game and for a season. First, to measure a team's game performance, we simply use whether this team wins or loses. Second, to measure a team's performance over a season, we use elo ratings of the NBA teams. The elo rating system was originally invented as a chess rating system for calculating the relative skill levels of players. The popular forecasting website FiveThirtyEight developed an elo rating system to measure the skill levels of different NBA teams [166]. These elo ratings are used to predict game results on FiveThirtyEight and are well received by major sports media, such as ESPN [12] and CBS Sports. [13] The FiveThirtyEight elo ratings satisfy the following properties:

- A team's elo rating is represented by a number that increases or decreases depending on the outcome of a game. After a game, the winning team takes elo points from the losing one, so the system is zero-sum.
- The number of elo points exchanged after a game depends on the elo rating difference

[8] `http://fivethirtyeight.com/`.
[9] `https://www.basketball-reference.com/`.
[10] `https://www.forbes.com/`.
[11] `https://www.wikipedia.org/`.
[12] `http://www.espn.com/`.
[13] `https://www.cbssports.com/`.

Table 3.3: List of variables and their corresponding definitions and sources. Measurements of team performance are in bold.

Variable	Definition	Source
Performance		
winning	Win or lose a game.	FiveThirtyEight
season elo	A team's elo rating at the end of a season.	FiveThirtyEight
season elo difference	A team's elo rating difference between the end of a season and its last season.	FiveThirtyEight
month elo	A team's elo rating at the end of that month.	FiveThirtyEight
month elo difference	A team's elo rating difference between the end of a month and its last month.	FiveThirtyEight
Game information		
team elo	A team's elo rating before the game.	FiveThirtyEight
opponent elo	The opponent's elo rating before the game.	FiveThirtyEight
point difference	Absolute point difference of the game.	FiveThirtyEight
rivalry or not	If the opponent team is a rivalry.	Wikipedia
top team	If a team is with the five highest elo ratings at the end of a season.	FiveThirtyEight
bottom team	If a team is with the five lowest elo ratings at the end of a season.	FiveThirtyEight
Team information		
market value	Transfer fee estimated to buy a team on the market.	Forbes
average age	The average age of the roster.	Basketball-Reference
#star players	The number of players selected to play the NBA All-Star Game.	Basketball-Reference
#unique users	The number of users that made at least one post/comment in the team's subreddit.	N/A
offense	The average points scored per game.	Basketball-Reference
defense	The average points allowed per game.	Basketball-Reference
turnovers	The average turnovers per game.	Basketball-Reference
Temporal information		
season	A categorical variable to indicate the season.	N/A (control variable)
month	A categorical variable to indicate the month.	N/A (control variable)

between two teams prior to the game, final basketball points, and home-court advantage. Teams gain more elo points for unexpected wins, great basketball point differences, and winning away games.

- The long-term average elo rating of all the teams is 1500.

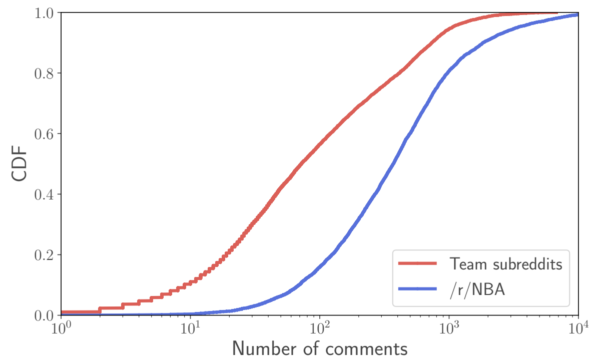
To measure team performance, we use a team's elo rating at the end of a season as well as the elo difference between the end of this season and last season. A high elo rating at the end indicates an absolute sense of strong performance; a great elo rating difference suggests that a team has been improving. In addition to studying how team performance over a season affects fan loyalty, we also include team performance over a month to check the robustness of the results.

3.4.1.2 Game Information

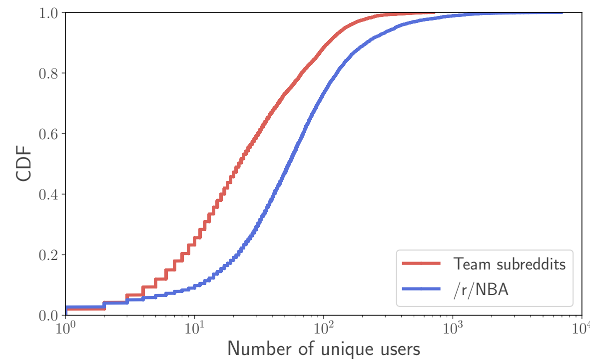
To test **H1**, we need to use the interaction between game performance and top (bottom) team so that we can capture whether a top team loses or a bottom team wins. We define **top team** as teams with the highest five elo ratings at the end of a season and **bottom team** as teams with the lowest five elo ratings at the end of a season. We also include the following variables to characterize a single game: 1) Two team's elo ratings, which can partly measure the importance of a game; 2) (Basketball) point difference, which captures how close a game is; 3) Rivalry game: which indicates known rivalry relations in the NBA, such as the Lakers and the Celtics. We collect all pairs of the NBA rivalry teams from Wikipedia [206].

3.4.1.3 Team Information

To capture team properties, we include a team's market value, average age of players, and the number of star players. Market value estimates the value of a team on the current market. There are three key factors that impact a team's market value, including market size, recent performance and history [41]. We collect market values of all NBA teams from Forbes. We scrape the average age of players on the roster from Basketball-Reference, which computes the average age of players at



(a) The CDF of #comments in game threads.



(b) The CDF of #unique users in game threads.

Figure 3.2: Figure 3.2a shows the cumulative distribution of the number of comments in all game threads in both /r/NBA and team subreddits. Figure 3.2b shows the cumulative distribution of the number of unique users in all game threads in both /r/NBA and team subreddits.

the start of Feb 1st of that season. The website chooses to calculate average age on Feb 1st because it is near the player trade deadline [6], and every team has a relatively stable roster at that time. The number of star players measures the number of players selected to play in the NBA All-Star Game [204] of that season and this information is collected from Basketball-Reference. We further include variables that characterize a team's playing style: 1) Offense: the average points scored per game; 2) Defense: the average points allowed per game; 3) Turnovers: the average number of turnovers per game. Teams' playing style information by season is collected from Basketball-Reference.

3.4.2 Analysis for H1

Online fan communities provide a platform for fans to discuss sports games in real time and make the game watching experience interactive with other people on the Internet. Accordingly, every team subreddit posts a "Game Thread" before the start of a game. Fans are encouraged to make comments related to a game in its game thread. Figure 3.2 shows the cumulative distributions of the number of comments and the number of unique users. A game thread can accumulate hundreds or thousands of comments. The number of comments is usually significantly higher

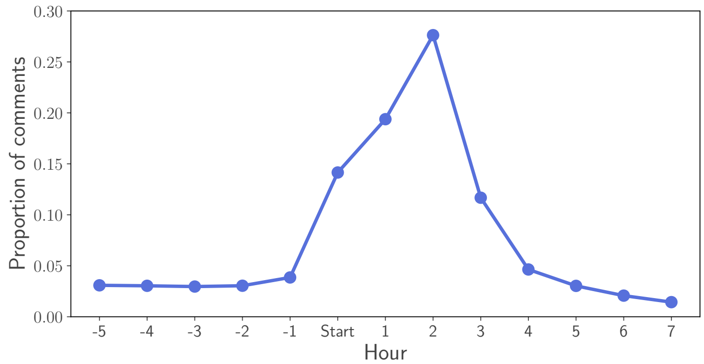


Figure 3.3: The average proportion of comments made in each team subreddit by hour on the game day during the 2017 season (normalized based on game's starting hour). Error bars represent standard errors and are too small to see in the figure. Comment activity increases and peaks at the second hour after the game starts, as a typical NBA game takes around 2.5 hours.

during game time than other time periods. Figure 3.3 shows the average proportion of comments made in each team subreddit by hour on the game day of the 2017 season (normalized based on games' starting hour). The number of comments peaked around the game time.

We use the number of comments in game threads to capture the fan activity level for a game. Most game threads used titles that are similar to this format: "[Game Thread]: team 1 @ team 2".[14] We detected 8,596 game threads in team subreddits and 6,277 game threads in /r/NBA based on regular expression matching. Since NBA-related subreddits allow any fan to create game threads, titles of game threads do not follow the same pattern, especially in the earlier times of team subreddits. A detailed explanation and sanity check is presented in Section 3.7.2.

Hierarchical regression analysis was used to analyze the effect of team performance in a single

[14] If more than one game thread is created for the same game, only the first one is kept, and the others are deleted by the moderator.

game on fan activity. Our full linear regression model to test **H1** is shown below:

$$
\begin{aligned}
\#comments\ in\ game\ thread \sim & \beta_0 + \beta_s\, season + \beta_m\, month + \beta_t\, top\ team + \beta_b\, bottom\ team \\
& + \beta_1\, winning + \beta_2\, top\ team\ winning + \beta_3\, top\ team\ losing \\
& + \beta_4\, bottom\ team\ winning + \beta_5\, bottom\ team\ losing \\
& + \beta_6\, team\ elo + \beta_7\, opponent\ elo + \beta_8\, rivalry\ or\ not + \beta_9\, point\ difference \\
& + \beta_{10}\, market\ value + \beta_{11}\, average\ age + \beta_{12}\, \#star\ players + \beta_{13}\, \#unique\ users \\
& + \beta_{14}\, offense + \beta_{15}\, defense + \beta_{16}\, turnovers.
\end{aligned}
\tag{3.1}
$$

To test our hypothesis in team subreddits, all the variables in Equation 3.1 are included. Unlike game threads in team subreddits, game threads in /r/NBA involve two teams and the following variables are ill-defined: "winning," "offense," "defense," and "turnovers." Therefore, these variables are removed when testing our hypothesis on game threads in /r/NBA.

3.4.3 Analysis for H2

Fan loyalty refers to people displaying recurring behavior and a strong positive attitude towards a team [49]. To examine the relationship between team performance and fan loyalty in team subreddits, we first define active users as those that post or comment in a team subreddit during a time period. We then define two measurements of fan loyalty: *seasonly user retention* and *monthly user retention*. Seasonly user retention refers to the proportion of users that remain active in season $s + 1$ among all users that are active in season s. Monthly user retention refers to the proportion of users that remain active in month $m+1$ among all users that are active in month m.

The full linear regression models to test **H2** are shown below:

$$
\begin{aligned}
\textit{seasonly user retention} \sim & \beta_0 + \beta_s\, \textit{season} \\
& + \beta_1\, \textit{season elo} + \beta_2\, \textit{season elo difference} \\
& + \beta_3\, \textit{market value} + \beta_4\, \textit{average age} + \beta_5\, \textit{\#star players} + \beta_6\, \textit{\#unique players} \\
& + \beta_7\, \textit{offense} + \beta_8\, \textit{defense} + \beta_9\, \textit{turnovers}.
\end{aligned}
\tag{3.2}
$$

$$
\begin{aligned}
\textit{monthly user retention} \sim & \beta_0 + \beta_s\, \textit{season} + \beta_m\, \textit{month} \\
& + \beta_1\, \textit{month elo} + \beta_2\, \textit{month elo difference} \\
& + \beta_3\, \textit{market value} + \beta_4\, \textit{average age} + \beta_5\, \textit{\#star players} + \beta_6\, \textit{\#unique players} \\
& + \beta_7\, \textit{offense} + \beta_8\, \textit{defense} + \beta_9\, \textit{turnovers}.
\end{aligned}
\tag{3.3}
$$

3.4.4 Analysis for H3

Among the five topics listed in Table 3.2, *"season prospects"* and *"future"* topics are closely related to our hypotheses about fans talking about winning and framing the future. By applying the trained LDA model to comments in each team subreddit, we are able to estimate the average topic distribution of each team subreddit by season. Our full linear regression model to test **H3** is shown below:

$$
\begin{aligned}
\textit{topic weight} \sim & \beta_0 + \beta_s\, \textit{season} \\
& + \beta_1\, \textit{season elo} + \beta_2\, \textit{season elo difference} \\
& + \beta_3\, \textit{market value} + \beta_4\, \textit{average age} + \beta_5\, \textit{\#star players} + \beta_6\, \textit{\#unique players} \\
& + \beta_7\, \textit{offense} + \beta_8\, \textit{defense} + \beta_9\, \textit{turnovers},
\end{aligned}
\tag{3.4}
$$

where *topic weight* can be the average topic weight of either *"season prospects"* or *"future."*

3.5 Results

Based on the above variables, our results from hierarchical regression analyses linear regression models by and large validate our hypotheses. Furthermore, we find that the average age of players on the roster consistently plays an important role in fan behavior, while it is not the case for market value and the number of star players.

3.5.1 How does Team Performance Affect Game-level Activity? (H1)

Consistent with **H1**, regression results show that the top team losing and the bottom team winning correlate with higher levels of fan activity in both team subreddits and /r/NBA. Table 3.4 presents the results of our hierarchical regression analyses. The R^2 value is 0.40 for team subreddits and 0.63 for /r/NBA, suggesting that our linear variables can reasonably recover fan activity in game threads. Overall, fans are more active when their team wins in team subreddits (remember that the notion of one's team does not hold in /r/NBA). The interaction with the top team and the bottom team show that surprise can stimulate fan activity: both the top team losing and the bottom team winning have significantly positive coefficients. To put this into context, in the 2017 season, the average winning percentage of the top five teams is 69%. Fans of the top teams may get used to their teams winning games, in which case losing becomes a surprise. On the contrary, the average winning percentage of the bottom five teams is 31%. It is invigorating for these fans to watch their team winning. The extra excitement can stimulate more comments in the game threads in both team subreddits and /r/NBA. In comparison, when top teams win or bottom teams lose, fans are less active, evidenced by the negative coefficient in team subreddits (not as statistically significant in /r/NBA).

To further illustrate this contrast, Figure 3.4 shows the average number of comments on winning, losing, and non-game days for the top three and the bottom three teams in the 2017 and 2016 regular season. Consistent patterns arise: 1) In all top and bottom teams, the average

Table 3.4: Hierarchical regression analyses for game-level activity in team subreddits and /r/NBA. Month is also added as a control variable for each model. **Throughout this paper, the number of stars indicate p-values, ***: $p < 0.001$, **: $p < 0.01$ *: $p < 0.05$.** We report p-values without the Bonferroni correction in all the regression tables. In Section 3.7.3, we report F-test results with the null hypothesis that adding team performance variables does not provide a significantly better fit and reject the null hypothesis after the Bonferroni correction.

Variable	Team subreddits Reg. 1	Team subreddits Reg. 2	Team subreddits Reg. 3	/r/NBA Reg. 1	/r/NBA Reg. 2	/r/NBA Reg. 3
Control: season						
2014	0.011∗∗∗	0.012∗∗∗	0.013∗∗∗	0.007∗∗∗	0.007∗∗∗	0.005∗∗∗
2015	0.032∗∗∗	0.032∗∗∗	0.045∗∗∗	0.010∗∗∗	0.010∗∗∗	0.006∗∗∗
2016	0.046∗∗∗	0.046∗∗∗	0.062∗∗∗	0.014∗∗∗	0.015∗∗∗	0.012∗∗∗
2017	0.067∗∗∗	0.068∗∗∗	0.081∗∗∗	0.018∗∗∗	0.019∗∗∗	0.016∗∗∗
Control: top/bottom team						
top team		0.012∗∗∗	0.012∗∗∗		0.012∗∗∗	0.020∗∗∗
bottom team		−0.012∗∗∗	−0.007∗∗∗		−0.009∗∗∗	−0.007∗∗
Performance						
winning			0.003∗∗			–
top team winning			−0.006∗∗∗			−0.018∗∗∗
top team losing			0.006∗∗∗			0.018∗∗∗
bottom team winning			0.007∗∗∗			0.006∗∗∗
bottom team losing			−0.011∗∗∗			−0.003∗
Game information						
team elo			0.083∗∗∗			0.091∗∗∗
opponent elo			0.070∗∗∗			0.111∗∗∗
rivalry or not			0.010∗∗∗			0.008∗∗∗
point difference			−0.017∗∗∗			−0.010∗∗∗
Team information						
market value			0.051∗∗∗			0.013∗∗∗
average age			−0.067∗∗∗			−0.015∗
#star players			0.040∗∗∗			0.020∗∗∗
#unique users			0.058∗∗∗			0.017∗∗∗
offense			0.023∗∗			–
defense			−0.012∗∗			–
turnovers			0.084∗∗∗			–
intercept	−0.010∗∗	−0.011∗∗	0.085∗∗∗	0.001	−0.004∗∗	−0.156∗∗∗
Adjusted R^2	0.236	0.286	0.440	0.302	0.338	0.644
Intraclass Correlation (Season) [207]	0.087	–	–	0.021	–	–

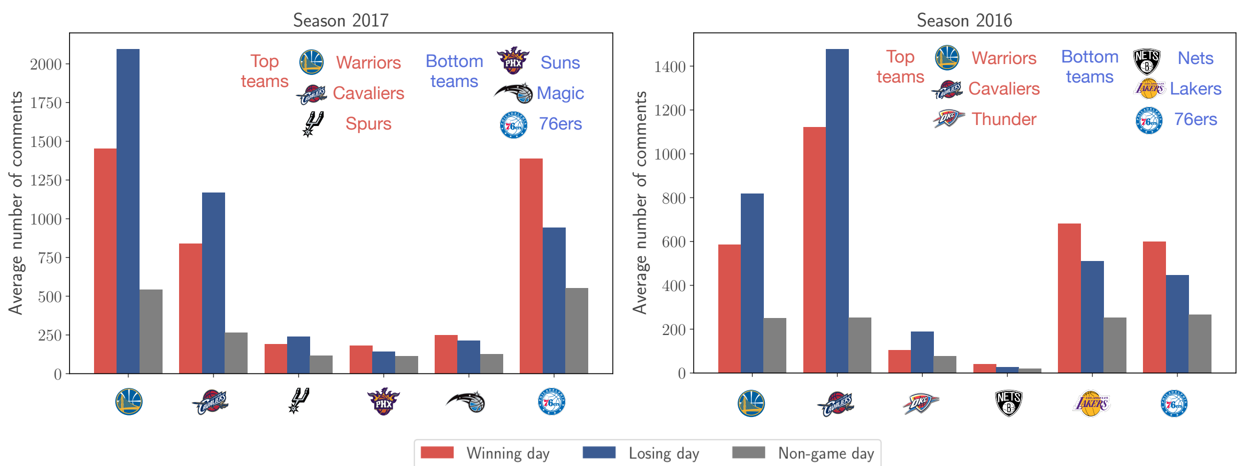


Figure 3.4: The average number of comments on winning, losing and non-game days for the top three and the bottom three teams in the 2017 (left) and 2016 (right) regular season. In all the top and bottom teams, the average number of comments on game days is significantly higher than non-game days. In all top teams, the average number of comments on losing days is higher than winning days, while bottom teams show the opposite trend.

number of comments on game days is significantly higher than non-game days; 2) In all top teams, the average number of comments on losing days is higher than winning days, but bottom teams show exactly the opposite trend. Our results differ from that of Leung et al. [104], which finds that unexpected winning does not have a significant impact on Wikipedia page edits. One of the primary differences between our method and theirs is that they did not specifically control the effect of top/bottom team. It may also be explained by the fact that Wikipedia page edits do not capture the behavior of most fans and are much more sparse than comments in online fan communities. Online fan communities provide rich behavioral data for understanding how team performance affects fan behavior. The number of fans involved in our dataset is much higher than that in their Wikipedia dataset. A comparison between fans' behavior on Reddit and Wikipedia could be an interesting direction for future research. In addition, game information and team information also serve as important factors. Among variables about game information, point difference is negatively correlated with game-level user activity, as the game intensity tends to be higher when the point difference is small (a close game). Better teams (with higher elo ratings) playing against better teams or rivalry teams correlates with higher user activity levels. As for team information, a team's market value, the number of unique users, and the number of star players are positively correlated with the number of comments, since these two factors are closely related to the number of fans. Younger teams with more average points scored and less points allowed per game stimulate more discussion in team subreddits.

3.5.2 How does Team Performance Relate to Fan Loyalty in Team Subreddits? (H2)

Our findings confirm **H2**, that top teams tend to have lower fan loyalty and bottom teams tend to have higher fan loyalty, measured by both seasonly user retention and monthly user retention. Table 3.5 shows the hierarchical regression results. In both regression analyses, elo rating, which measures a team's absolute performance, has a statistically significant negative impact on user retention rate. The coefficient of elo rating also has the greatest absolute value among all vari-

Table 3.5: Hierarchical regression analyses for seasonly user retention rate and monthly user retention rate in team subreddits. Month is also added as a control variable for the monthly user retention analysis. **#unique users** is counted every season for the seasonly user retention analysis and every month for the monthly user retention analysis. For both dependent variables, team's overall performance has a negative coefficient while short-term performance and market value has a positive coefficient.

Variable	Seasonly user retention		Monthly user retention	
	Reg. 1	Reg. 2	Reg. 1	Reg. 2
Control: season				
2014	0.237∗∗∗	0.237∗∗∗	0.086∗∗∗	0.055∗∗∗
2015	0.184∗∗∗	0.187∗∗∗	0.126∗∗∗	0.126∗∗∗
2016	0.088∗∗∗	0.116∗∗∗	0.130∗∗∗	0.139∗∗∗
2017	0.073∗∗∗	0.090∗∗∗	0.137∗∗∗	0.141∗∗∗
Performance				
season elo		−0.370∗∗		–
season elo difference		0.229∗∗∗		–
month elo		–		−0.170∗∗
month elo difference		–		0.032∗∗
Team information				
market value		0.068∗		0.051∗∗
average age		−0.105∗		−0.021
#star players		−0.038		0.041
#unique users		0.181∗∗∗		0.111∗∗∗
offense		−0.168		0.004
defense		−0.077		−0.053
turnovers		−0.037		−0.041
intercept	0.583∗∗∗	0.629∗∗∗	0.478∗∗∗	0.460∗∗∗
Adjusted R^2	0.286	0.503	0.155	0.232
Intraclass Correlation (Season) [207]	0.396	–	0.029	–

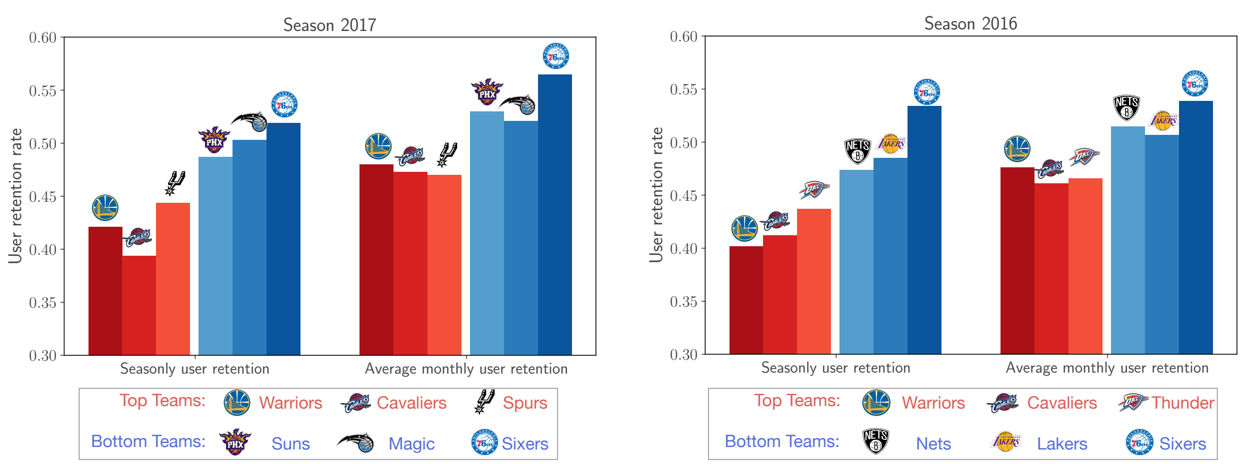


Figure 3.5: Seasonly user retention rate and average monthly user retention rate of the top three and bottom three teams in the 2017 (left) and 2016 (right) season. Bottom teams consistently have higher user retention than top teams.

ables (except intercept). Meanwhile, improved performance reflected by elo difference positively correlates with user retention.

Figure 3.5 presents the seasonly user retention rate and average monthly user retention rate of the top 3 and bottom 3 teams in the 2017 (left) and 2016 (right) season. It is consistent that in these two seasons, bottom teams have higher user retention rate than top teams, both seasonly and monthly. This may be explained by the famous "bandwagon" phenomenon in professional sports [197]: Fans may "jump on the bandwagon" by starting to follow the current top teams, which provides a short cut to achievement and success for them. In comparison, terrible team performance can serve as a loyalty filter. After a period of poor performance, only die-hard fans stay active and optimistic in the team subreddits. Our results echo the finding by Hirt et al. [79]: after developing strong allegiances with a sports team, fans find it difficult to disassociate from the team, even when the team is unsuccessful. It is worth noting that the low fan loyalty of the top teams cannot simply be explained by the fact that they tend to have more fans. In fact, teams with higher market value and more unique users (more fans) tend to have a higher user retention rate, partly because their success depends on a healthy and strong fan community.

Similar to game-level activity, fans are more loyal to younger teams, at least in seasonly user retention (the coefficient is also negative for monthly user retention and $p-$value is 0.07). Surprisingly, according to our hierarchical regression results, a team's number of star players and playing style (offense, defense, and turnovers) have no significant impact on user retention.

3.5.3 How does Team Performance Affect Topics of Discussion in Team Subreddits? (H3)

Our final question is concerned with the relation between team performance and topics of discussion in online fan communities. Our results validate **H3**, that better teams have more discussions on *"season prospects"* and worse teams tend to discuss *"future."* Table 3.6 presents the results of hierarchical regression analyses OLS linear regression on *"future"* topic weight and *"season prospects"* topic weight computed with our LDA model. In both regressions, only team performance

Table 3.6: Hierarchical regression analyses for *"season prospects"* topic weight and *"future"* topic weight in team subreddits. Team performance has positive correlation with *"season prospects"* topic and negative correlation with *"future"* topic.

Variable	*"season prospects"* Reg. 1	*"season prospects"* Reg. 2	*"future"* Reg. 1	*"future"* Reg. 2
Control: season				
2014	0.096∗∗∗	0.056∗∗	0.059∗	0.137∗∗∗
2015	0.067∗∗	0.049∗	0.054∗	0.129∗∗∗
2016	0.069∗∗	0.053∗	0.109∗∗∗	0.175∗∗∗
2017	0.061∗	0.048∗	0.065∗	0.160∗∗∗
Performance				
season elo		0.410∗∗∗		−0.415∗∗
season elo difference		−0.130		−0.099
Team information				
market value		−0.065		0.051
average age		0.149∗∗∗		−0.189∗∗∗
#star players		0.018		−0.187∗∗
#unique users		0.071		−0.104
offense		−0.036		0.037
defense		−0.120		−0.115
turnovers		0.059		−0.092
intercept	0.398∗∗∗	0.286∗∗∗	0.478∗∗∗	0.814∗∗∗
Adjusted R^2	0.013	0.578	0.002	0.619
Intraclass Correlation (Season) [207]	0.059	–	0.003	–

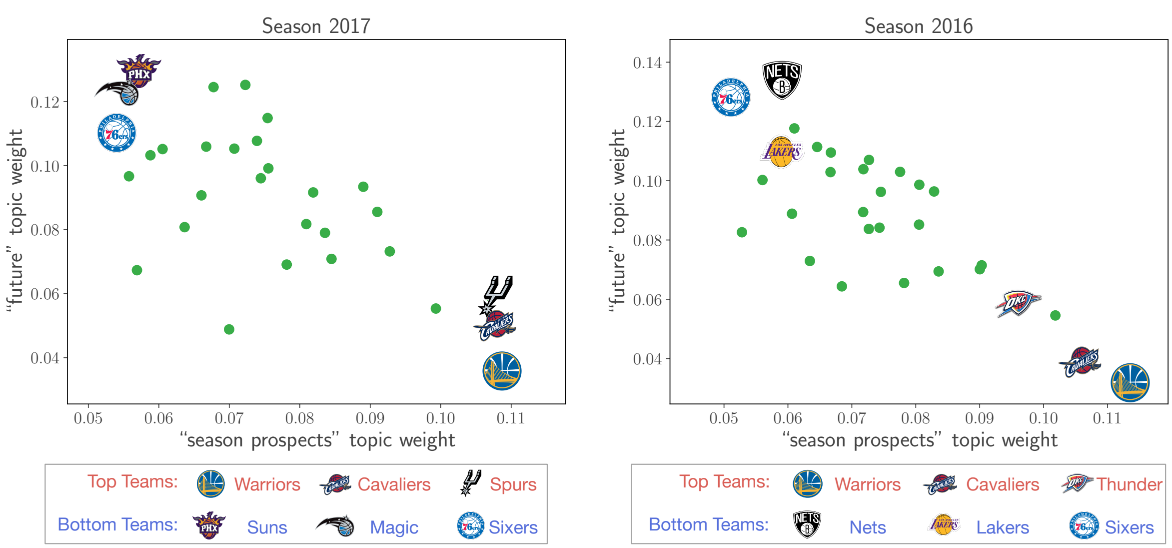


Figure 3.6: Scatterplot of *"season prospects"* topic weight and *"future"* topic weight in all the team subreddits in the 2017 (left) and 2016 (right) season. The top three teams and bottom three teams are represented by team logos instead of points. Teams are ranked by elo rating at the end of each season. Fans of the top teams tend to discuss much more *"season prospects"* topics (lower right corner) and fans of the bottom teams tend to discuss much more *"future"* topics (upper left corner).

(season elo) and average age have statistically significant coefficients. Both team performance and average age are positively correlated with *"season prospects"* and negatively correlated with *"future"*. Moreover, the number of star players has a negative correlation with *"future"* but has no significant effect on *"season prospects."* Despite having only two or three variables (except control variables and intercept) with significant coefficients, both regression analyses achieve strong correlation with R^2 above 0.57. Note that the improvement in team performance (season elo difference) does not have a significant effect.

As an example, Figure 3.6 further shows topic weights of *"future"* and *"season prospects"* for all the teams in the 2017 and 2016 season. The top 3 teams and bottom 3 teams in each season are highlighted using team logos. The top teams are consistently in the lower right corner (high *"season prospects"*, low *"future"*), while the bottom teams are in the upper left corner (low *"season prospects"*, high *"future"*). Our results echo the finding in Doyle et al. [48]: framing the future is an important strategy for fans of teams with poor performance to maintain a positive identity in the absence of success.

The effect of average age reflects the promise that young talents hold for NBA teams. Although it takes time for talented rookies that just come out of college to develop physical and mental strength to compete in the NBA, fans can see great potential in them and remain positive about their team's future, despite the team's short-term poor performance. In contrast, veteran players are expected to bring immediate benefits to the team and compete for playoff positions and even championships. For example, Rothstein [159] lists a number of veteran players who either took a pay cut or accepted a smaller role in top teams to chase a championship ring at the end of their career.

A team's playing style, including offense points, defense points, and turnovers, doesn't seem to influence the topic weights of these two topics. We also run regression for the other three top topics in Table 3.2 and present the results in Section 3.7.4. Team performance plays a limited role for the other three topics, while average age is consistently significant for all three discussion topics.

3.6 Concluding Discussion

In this work, we provide the first large-scale characterization of online fan communities of the NBA teams. We build a unique dataset that combines user behavior in NBA-related subreddits and statistics of team performance. We demonstrate how team performance affects fan behavior both at the game level and at the season level. Fans are more active when top teams lose and bottom teams win, which suggests that in addition to simply winning or losing, surprise plays an important role in driving fan activity. Furthermore, a team's strong performance doesn't necessarily make the fan community more loyal. It may attract "bandwagon fans" and result in a low user retention rate. We find that the bottom teams generally have higher user retention rate than the top teams. Finally, fans of the top teams and bottom teams focus on different topics of discussion. Fans of the top teams talk more about season records, playoff seeds, and winning the championship, while fans of the bottom teams spend more time framing the future to compensate for the lack of recent success.

Limitations. One key limitation of our work is the representativeness of our dataset. First, although our study uses a dataset that spans five years, our period coincides with the rapid growth of the entire Reddit community. We use *season* and *month* to try our best to account for temporal differences, but our sample could still be based upon fans with a mindset of growth. Second, although Goldschein [62] suggests that /r/NBA is now playing an important role among fans, the NBA fan communities on Reddit may not be representative of the Internet and the whole offline population.

Another limitation of our work lies in our measurement. For game-level activity, we only consider the number of comments in the game threads. This measurement provides a nice way to make sure that the comments are about the game, but we may have missed related comments in other threads. We do not consider other aspects of the comments such as sentiment and passion. In addition, our fan loyalty metric is entirely based on user retention. A user who posts on a team subreddit certainly supports the team to a different extent from those who do not. Our metric may fail to capture lurkers who silently support their teams. Finally, our topics of discussion are derived from topic modeling, an unsupervised approach. Supervised approaches could provide more accurate identification of topics, although the deduction approach would limit us to a specific set of topics independent of the dataset.

Implications for online communities. First, our work clearly demonstrates that online communities do not only exist in the virtual world; they are usually embedded in the offline context and attract people with similar offline interests. It is an important research question to understand to what extent and how online communities relate to offline contexts as well as what fraction of online communities are entirely virtual. Professional sports provide an interesting case, because these online fan communities, in a way, only exist as a result of the offline sports teams and games. Such connections highlight the necessity to combine multiple data sources to understand how fans' usage of social media correlates with the on-going events of the topic of their interests. Our study has the potential to serve as a window into the relationship between online social behavior and offline professional sports. We show that subreddit activity has significant correlations with game results and team properties. Exploring the factors that motivate users of interest-based communities to communicate with social media is also an important and rich area for future research. For example, a promising future direction is to study the reasons behind fans departing a team subreddit. Possible reasons include being disappointed by the team performance or playing style, favorite players being traded, and being attacked by other fans in the team subreddit or /r/NBA.

Second, our results show that teams with strong performance correlate with low fan loyalty. These results relate to the multi-community perspective in online community research [183, 221, 73, 224]. One future direction is to examine where fans migrate to and whether fans leave the NBA or the Reddit altogether, and more importantly, what factors determine such migration decisions.

Third, our findings reveal strategies for the design of sports-related online platforms. Our results clearly demonstrate that teams in under-performing periods are more likely to develop a more loyal fan base that discusses more about their team's *"future."* Recognizing these loyal fans and acknowledging their contributions within the fan community can be critical for facilitating attraction and retention of these fans. For example, team subreddits' moderators may reward a unique flair to the users who have been active in the community for a long time, especially during the difficult times.

Implications for sports management. Our findings suggest that winning is not everything. In fact, unexpected losses can stimulate fan activity. The increase of fan activity does not necessarily happen in a good way. For example, the fans of the Cavaliers, which won the Eastern Championship of the 2017 season, started to discuss firing the team's head coach Tyronn Lue after losing three of the first six games in the following season. Managers may try to understand the role of expectation in fan behavior and guide the increased activity and attention towards improving the team and building a strong fan base.

We also find that the average age of the roster consistently plays an important role in fan behavior: younger teams tend to bring more fan activity on game days and develop a more loyal fan base that discusses about *"future."* These results contribute to existing literature on the effect of age in sports management. Timmerman [186] finds that the average age is positively correlated with team performance, while the age diversity is negatively correlated (in other words, veterans improve team performance but are not necessarily compatible with young players). The tradeoff between veteran players and young talents requires more research from the perspective of both team performance and fan engagement.

Finally, it is crucial for teams to maintain a strong fan base that can support them during unsuccessful times because it is difficult for sports teams to sustain winning for a long time. This is especially true in the NBA since the draft lottery mechanism is designed to give bottom teams opportunities to improve and compete. Consistent with Doyle et al. [48], we find that framing the *"future"* can be an important strategy for teams with poor performance to maintain a positive group identity. The absence of success can be a great opportunity to develop a deep attachment with loyal fans. Prior studies show that certain fan group would like to persevere with their supported team through almost anything, including years of defeat, to recognize themselves as die-hard fans. By doing this, they feel that they would reap more affective significance among the fan community when the team becomes successful in the future [197, 82]. It is important for managers to recognize these loyal fans and create ways to acknowledge and leverage their positions within the fan community. For instance, teams may host "Open Day" and invite these loyal fans to visit facilities and interact with star players and coaching staff. Hosting Ask Me Anything (AMA) [203] interviews is another strategy to engage with online fan communities.

3.7 Appendix

3.7.1 Topic Modeling and Additional Topics

The following pre-processing procedures are used to clean data before training topic models:

- Converting all the words to lower case.
- Removing all the HTML links in the comments.
- Removing all the player names and nicknames, such as Lebron, Kobe.
- Removing all the team names, such as Lakers, Celtics.
- Removing common stopwords.
- Lemmatizing all words.

The remaining 10 topics generated after training are shown in Table 3.7.

3.7.2 Game Thread Matching

The percentage of game threads detected among all games by season is shown in Table 3.8. The percentages of game threads detected in /r/NBA are high in all seasons, with an average of 95%. The percentage of game threads detected in team subreddits are high in the 2016 and 2017 seasons, all above 80%. Significant amount of game threads are missing for the 2013 and 2014 seasons.

Table 3.7: The remaining 10 extracted topics by LDA using all comments in /r/NBA. The top ten weighted words are presented for each topic.

LDA topic	**top words**	**average topic weight**
Topic 00	big, level, work, hard, league, talent, basketball, long, skill, playing	0.071
Topic 01	fucking, dude, day, kid, life, friend, face, talking, bitch, court	0.068
Topic 02	injury, minute, played, playing, half, quarter, end, night, ago, start	0.068
Topic 03	watch, watching, basketball, feel, fun, damn, hope, god, fucking, honestly, suck	0.067
Topic 04	call, foul, ref, throw, free, called, hand, rule, hit, foot	0.065
Topic 05	contract, money, deal, cap, million, sign, free, pay, max, salary	0.063
Topic 06	post, comment, thread, edit, read, friend, face, talking, bitch, court	0.055
Topic 07	sport, basketball, school, city, jersey, high, black, world, college, white	0.054
Topic 08	coach, bench, starting, role, system, fit, coaching, front, starter, minute	0.054
Topic 09	prime, greatest, goat, time, career, all, star, history, seasons, era	0.043

Table 3.8: Percentage of game threads detected from team subreddits and /r/NBA by season. For each game, it is supposed to have one game thread in home-team subreddit, one game thread in away-team subreddit and one game thread in /r/NBA. The detected game thread is high for /r/NBA. For team subreddits, certain percentage of Game Thread are missing because: 1) Team subreddits only create game threads for important games when the community is relatively small; 2) Some game threads do not follow standard format.

	Team subreddits					/r/NBA				
year	2013	2014	2015	2016	2017	2013	2014	2015	2016	2017
#game threads detected	503	1176	2131	2362	2424	1106	1230	1311	1314	1305
#games	1314	1319	1311	1316	1309	1314	1319	1311	1316	1309
Percentage	19%	45%	81%	90%	93%	84%	93%	100%	100%	100%

We further investigate the reasons why game threads are not detected in team subreddits. We randomly sampled 50 games in season 2014. If home-team game thread and away-team game thread exist for these games, 100 game threads should have been detected in total. We manually checked them in our dataset and found that 40 game threads are successfully detected, 54 game threads were not created and 6 game threads had special title formats. Based on this limited sample, our detection rate is in fact 87% (40/46). There are two major reasons to explain missing game threads: 1) Some team subreddits were relatively small and fans only created game threads for important games (e.g., games against rivalry teams and strong teams, or games that are critical for playoff spots); 2) Some Game Threads do not follow the standard format. For example, a game thread in /r/Timberwolves is titled "Last regular season game, boys travel to Houston!".

3.7.3 F-tests with Bonferroni correction

Table 3.9 summarizes the results of F-tests where the null hypothesis is that adding team performance variables does not provide a significantly better fit. All the p-values are less than 0.0001 after the Bonferroni correction, so we reject the null hypothesis. In all the regressions in Table 3.9, adding team performance variables provides a significantly better fit.

Table 3.9: F-test results. P-values are all less than 0.0001 after the Bonferroni correction.

	H1		H2		H3	
	Team subreddits	/r/NBA	Seasonly	Monthly	*"season prospects"*	*"future"*
F-value	26.9	51.6	15.2	12.3	16.3	20.1
p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

3.7.4 Additional Linear Regression Models for Topic Weights

Table 3.10 presents the results of OLS linear regression models for average topic weights of top five topics other than *"future"* and *"season prospects,"* i.e., *"personal opinion," "game strategy,"* and *"game stats."*

Table 3.10: Hierarchical regression analyses for the other top 5 topics: *"personal opinion," "game strategy,"* and *"game stats."*

variables	*"personal opinion"* Reg. 1	*"personal opinion"* Reg. 2	*"game strategy"* Reg. 1	*"game strategy"* Reg. 2	*"game stats"* Reg. 1	*"game stats"* Reg. 2
Control: season						
2014	0.110∗∗∗	0.125∗∗∗	0.079∗∗∗	0.082∗∗∗	0.059∗∗∗	0.059∗∗∗
2015	0.146∗∗∗	0.196∗∗∗	0.096∗∗∗	0.095∗∗∗	0.083∗∗∗	0.082∗∗∗
2016	0.063∗∗	0.107∗∗∗	0.101∗∗∗	0.090∗∗∗	0.050∗∗∗	0.088∗∗∗
2017	0.152∗∗∗	0.211∗∗∗	0.089∗∗∗	0.081∗∗∗	0.104∗∗∗	0.103∗∗∗
Performance						
season elo		−0.380∗∗		−0.126		−0.099
season elo difference		−0.125∗		−0.024		0.044
Team information						
market value		0.080		0.015		0.014
average age		−0.121∗		−0.069∗		−0.053∗
#star players		0.018		−0.068		−0.043
#unique users		−0.105		0.068		0.146∗∗∗
offense points		0.158		0.112		−0.100
defense points		−0.213∗		0.038		0.053
turnovers		−0.120		0.032		−0.079
intercept		0.781∗∗∗		0.414∗∗∗	0.043	0.414∗∗∗
Adjusted R^2	0.048	0.208	0.008	0.251	0.025	0.334

Chapter 4

Intergroup Contact in Online NBA Fan Communities

4.1 Introduction

Driven by the growing concerns of tribalism and polarization in world politics [85, 31, 179], it is increasingly important to understand intergroup contact as a straightforward yet potentially powerful strategy to reduce prejudice between groups. Intergroup contact refers to interactions between members of different groups, and groups can be defined using a variety of factors, including political ideology, place of origin, and race. A key hypothesis is that members with intergroup contact (henceforth “intergroup members”) behave differently, e.g., by showing sympathy towards other groups and voicing different opinions in their affiliated group [144, 145, 47, 45, 143, 146, 10, 133, 9, 102, 60]. However, prior studies have observed different effects of intergroup contact. For instance, self-reported surveys show that intergroup contact relates to reduced prejudice towards immigrants in European countries [143], while a recent experimental study finds that exposure to opposing groups on Twitter can exacerbate political polarization [10]. Although self-reported surveys and experimental studies have been the main methods for studying the effect of intergroup contact [133, 9, 102, 21, 18, 65, 135], we believe that observational study allows researchers to characterize intergroup contact in the wild and provide valuable complementary evidence in diverse contexts. Indeed, with the emergence of online groups, it has become possible to observe intergroup contact and individual behavior at a massive scale for substantial periods. Our goal in this study is to investigate how individuals that choose to engage in intergroup contact behave differently from others without intergroup contact in their original affiliated group in online platforms.

We leverage the existing structure of NBA-related discussion forums on Reddit to identify the group affiliations of users and intergroup contact in the context of professional sports, a novel domain different from politics. We choose online fan groups of professional sports teams as a testbed for the following reasons. First, professional sports play a significant role in modern life [201, 23, 72]. People in the United States spent more than 31 billion hours watching sports games in 2015 [132], and the attendance of the 2017-2018 National Basketball Association (NBA) season reached 22 million [130]. Second, professional sports teams are unambiguously competitive in nature. Similar to other common contexts for studies on intergroup contact (e.g., political ideology), fans of sports teams can treat fans of opposing teams as enemies and sometimes even engage in violence [155, 55]. Moreover, sports fans tend to think that the media and supporters from opposing teams are likely to have unfair opinions against their favored teams, just like people with different ideologies [44].

Figure 4.1 illustrates our framework. There are 30 teams in the NBA, and every team has its discussion forum (henceforth **team subreddit**) on Reddit, a place where fans of the corresponding team congregate and discuss news, games, and any other topics that are relevant to the team. The low-access barrier on the Internet also enables users to communicate easily with fans from opposing teams. In fact, /r/NBA is dedicated to interactions between fans of all NBA teams for any discussion related to the NBA. Contrary to each team's "echo chamber," which is dominated by fans that support the same team, /r/NBA represents an open and diverse environment where intergroup contact occurs. We can thus identify intergroup members and single-group members based on whether they have any activity in the intergroup setting (/r/NBA).

As posting comments is a major activity in online platforms such as Reddit, analyzing the language used in online platforms provides an opportunity for capturing individuals' attitudes and emotions [40, 138, 162, 147]. In particular, there has been growing concerns about hate speech and negative language in online communities [39, 29, 30, 163, 24]. We thus focus on characterizing the differences in language usage between intergroup and single-group members in **their affiliated team subreddit** (the intragroup setting), e.g., whether intergroup members swear more than single-group members in their affiliated team subreddit. As a result, we would be able to capture

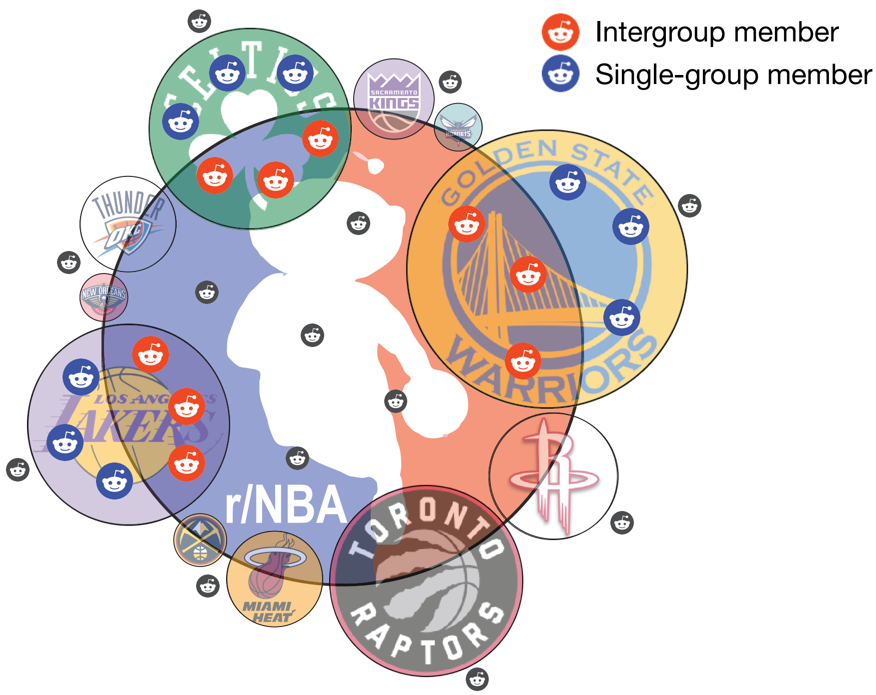


Figure 4.1: Illustration of NBA-related discussion forums (also known as subreddits) on Reddit. We identify group affiliation (i.e., whether a person is a fan of an NBA team) and intergroup contact based on the existing structure of NBA-related subreddits on Reddit. Each team has its team subreddit. Here, we present 11 of the 30 NBA teams (with corresponding team logos) to cover subreddits of different sizes. The central /r/NBA logo represents /r/NBA, where intergroup contact happens. The radius of each team logo is proportional to the number of subscribers in the corresponding subreddit. Users in each team logo represent fans of a team based on their indicated support in NBA-related subreddits. Red icons refer to intergroup members who have engaged in intergroup contact and are thus also in the /r/NBA logo, while blue icons refer to single-group users without such behavior. We only put red and blue icons in the three largest team subreddits due to space limitations, but every team subreddit has these two categories of users (see Figure 4.4 for the number of intergroup and single-group members of each team). Note that not all users who participated in these discussion forums qualify as a fan of an NBA team (grey icons).

behavioral differences reflected in language use between intergroup and single-group members in the intragroup setting. Note that in this work, we do not claim that intergroup contact causes such differences due to endogenous factors that may lead to individuals choosing to engage in intergroup contact in the wild (i.e., individuals who choose to engage in intergroup contact in practice may be inherently different from those who do not).

Organization and highlights. We first introduce our dataset and provide an overview of the framework for identifying group affiliations and intergroup contact in Section 4.2. With intergroup members and single-group members identified, We then investigate two research questions in the rest of the paper (methods in Section 4.3 and results in Section 4.4):

RQ1: *How do members with intergroup contact differ from those without such contact in intragroup language usage in NBA fan groups?*

RQ2: *How do different levels of intergroup contact relate to intragroup language usage?*

For **RQ1**, we first apply matching techniques to make sure the intergroup and single-group members are comparable. We then analyze the behavioral differences between intergroup and single-group members by examining language usage of their comments in their affiliated team subreddit. We demonstrate intriguing contrasts between them: intergroup members tend to use more negative and swear words, and generate more hate speech comments compared to single-group members in their affiliated team subreddit.

For **RQ2**, we are able to quantify different levels of intergroup contact for each intergroup member based on the frequency of intergroup contact. Interestingly, we find varying mechanisms of how different levels of intergroup contact relate to intragroup behavioral differences. For instance, although intergroup contact mostly monotonically relates to differences in language usage, the trends are not necessarily linear. Such varying mechanisms provide complementary evidence to the seemingly conflicting results on intergroup contact in recent studies.

To explore the potential reasons behind the clear behavioral differences in language usage between intergroup and single-group members, we further compare the language usage of intergroup members between the intragroup setting (affiliated team subreddit) and the intergroup setting

(/r/NBA) in Section 4.5. This setup naturally controls for the subject because we compare the same person across two different environments. We find that intergroup members are even more negative and more likely to swear in the intergroup setting. Such negative intergroup contact may partly explain the observed differences in intragroup language usage.

Our work highlights the fact that individuals selectively choose to have intergroup contact in the wild, and in turn interact with people without intergroup contact in their original group. We further demonstrate a variety of ways in which intergroup contact levels can moderate intragroup behavior. These observations may reconcile recent conflicting results with respect to intergroup contact. Our findings indicate that observational studies can provide important complementary evidence to experimental studies on this topic because interventions can hardly result in deep and regular contacts. We offer discussions in Section 6.7 and conclude our work in Section 4.7.

4.2 Professional sports as a testbed

We focus on the professional sports context derived from NBA-related discussion forums (/r/NBA and 30 team subreddits) on Reddit, an active community-driven platform where users can submit posts and make comments. These user-created discussion forums are also called "subreddits". Each subreddit has multiple moderators to make sure that posts are relevant to the subreddit's theme. Over the years, basketball fans all over the world have flocked to /r/NBA, the site's professional basketball subreddit, to discuss games in progress, seek meaning in the latest trade rumors, and debate the legality of calls by the referees. In fact, /r/NBA has become the largest single-sport subreddit with more than 1.9M subscribers [153] and one of the most active subreddits on Reddit [152]. NBA-related subreddits represent an ideal testbed for understanding how intergroup contact relates to intragroup behavior because their structure allows us to identify many users' team affiliation. Moreover, /r/NBA is the place for all basketball fans to congregate, where intergroup contact between fans of different teams happens.

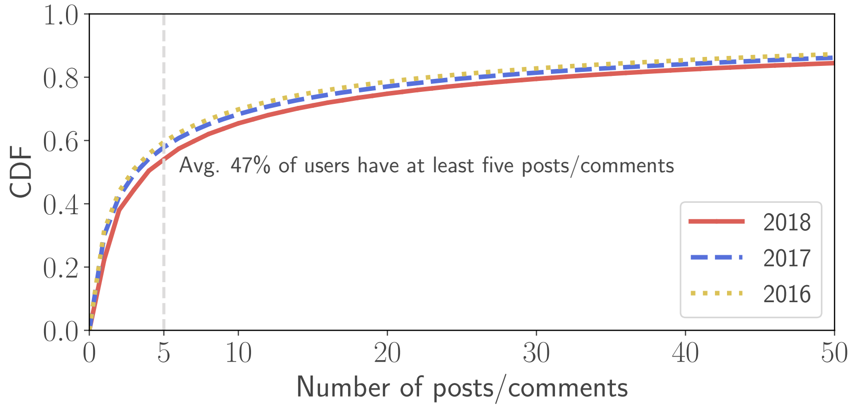


Figure 4.2: The distribution of the number of activities made by users in NBA-related subreddits in the 2018, 2017, and 2016 seasons.

4.2.1 Dataset and NBA Seasons

We obtain 2.1M posts and 61M comments in NBA-related subreddits from `https://pushshift.io` [12]. As pointed out in Zhang et al. [219], offline NBA seasons are reflected in user behavior in these NBA-related subreddits. We organize our dataset according to the timeline of NBA seasons and focus on the most recent three seasons, i.e., from July 2015 to June 2018. For simplicity and clarity, we refer to a specific season by the calendar year when it ends. For instance, the official 2017-2018 NBA season is referred to as *the 2018 season* or *2018* in this paper.

4.2.2 Identifying Team Affiliation and Intergroup Contact

To identify the team affiliation of users in a season, we first define active users in NBA-related subreddits in a season as those who have at least five activities, where an activity refers to either submitting a post or making a comment. Figure 4.2 shows the distribution of the number of activities by a user in NBA-related subreddits. These active users contribute over 95% of all the activities in NBA-related subreddits.

We identify the team affiliation of active users based on where their activities occur and by using a special mechanism on Reddit, known as flair. Flair appears as an icon next to the username in posts and comments. Every comment can have at most one flair. Before April 2018, flairs are

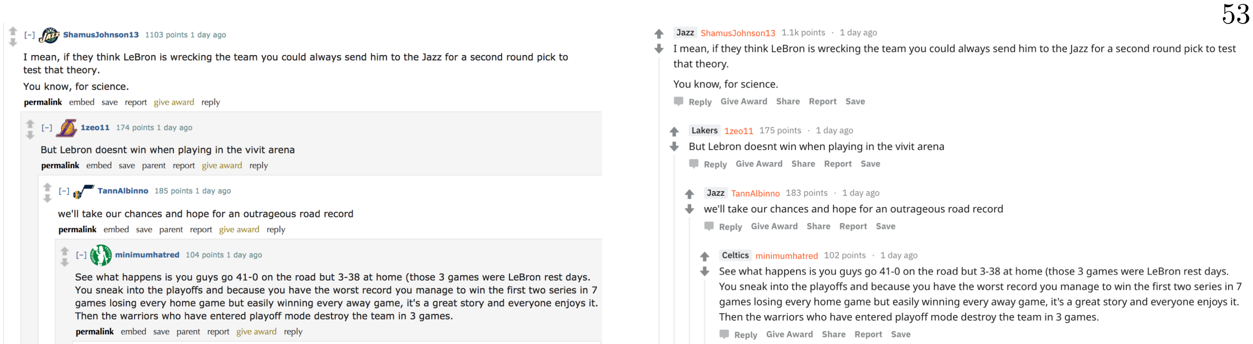


Figure 4.3: An example of the flair usage in /r/NBA before and after the design change. Before the design change, flairs are represented by team logos while after the change, flairs are represented by the team name.

represented by team logos. After that, Reddit adopted a new design to the entire platform, and the flairs are represented by team names in /r/NBA. An example is shown in Figure 4.3. In /r/NBA, fans can use flairs to indicate support of a team. ~80% of the comments/posts in our /r/NBA dataset have been made with flairs even though flairs are optional. We use all the flairs that fans used in /r/NBA for the inference of their team affiliation [1] .

We view posting/commenting in a team subreddit and using a team's flair in /r/NBA as an indication of support towards that team. An active user is defined as a fan of a team if the user indicates support only for that team, and such support sustains over all activities in an entire NBA season. In other words, all activities of a fan indicate support towards his/her affiliated team. It follows that not every active user in NBA-related subreddits is identified as a fan of some team.

We further determine whether a fan of a team is exposed to intergroup contact based on his/her (lack of) behavior in /r/NBA, which we refer to as intergroup status. To summarize, we categorize fans of a team into the following two categories:

- **Intergroup**: Fans of a team who posted in both the affiliated team subreddit and /r/NBA in the season.
- **Single-group**: Fans of a team who had no activity in /r/NBA throughout the season.

[1] In our /r/NBA dataset, every comment's JSON format has the "author_flair_css_class" key, and the corresponding value represents a unique flair this comment uses. The value is a string with the team's name and the flair id. For example, the flair values of the Log Angeles Lakers are phrased as "Lakers1", "Lakers2", "Lakers3", and "Lakers4".

Table 4.1: The number of intergroup and single-group members in the 2018, 2017, and 2016 seasons.

	2018	2017	2016
Single-group	6,023	5,941	4,843
Intergroup	28,296	24,528	20,467

Table 6.1 shows the number of members in each category. Since our study is concerned with intragroup behavior, i.e., behavior in the affiliated team subreddit, we view these intergroup and single-group fans as intergroup and single-group members of the affiliated team and study their behavior in the affiliated team subreddit.

Figure 4.4 presents the number of intergroup and single-group members in all 30 team subreddits in the 2018 season (see Figure 4.14 for the numbers in the 2017 and 2016 seasons). In every team subreddit, there are many more intergroup members than single-group members. Our definitions are based on user behavior in a single NBA season, and the label of a user can change across seasons. However, a single-group member rarely becomes an intergroup member in the next season in our dataset (6.0% of single-group members become intergroup members from 2016 to 2017, and 8.5% of single-group members become intergroup members from 2017 to 2018). This also confirms the tendency of single-group members to avoid intergroup contact.

We use both posting and commenting behavior to identify fans' team affiliation, but we focus on analyzing comments in the rest of the paper since the posts are usually much longer and more formal, and are thus not comparable to comments.

4.3 Methods

We study the behavioral differences between intergroup and single-group members in their affiliated team subreddit by examining the expression of emotions in their comments in team subreddits for two reasons: (1) emotion is a central theme in understanding sports fans and their opinions; (2) textual content constitutes the main observed behavior in NBA-related discussion forums [219, 103, 180].

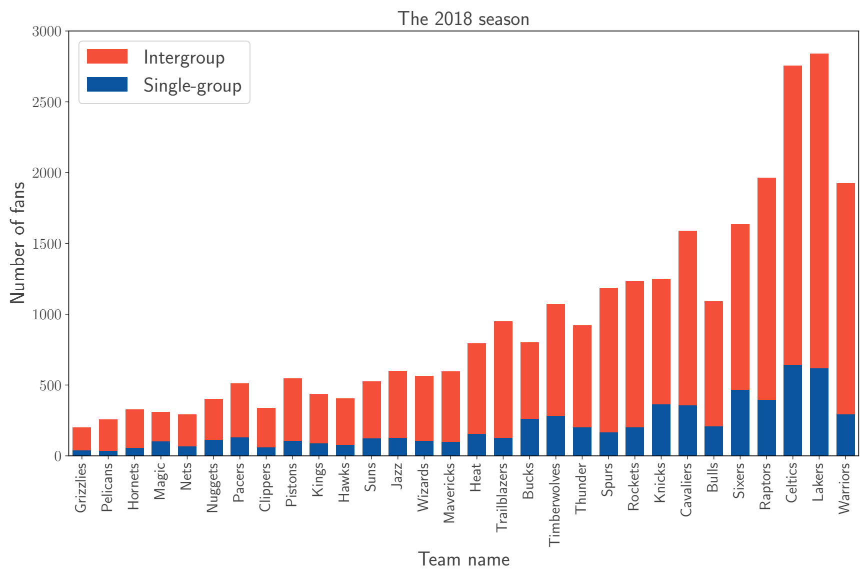


Figure 4.4: The number of intergroup and single-group members affiliated with each NBA team in the 2018 season. We rank 30 team subreddits by the number of subscribers each team has by the end of the 2018 season. (see Figure 4.14 for the data statistics in the 2017 and 2016 seasons).

4.3.1 RQ1: Intragroup Behavior Differences

Matched intergroup members. A naïve way to compare these two categories of users is to directly examine all users in each category. However, such an approach does not take into account other important confounding factors, such as how active a member is in the group. We thus seek to ensure that intergroup and single-group members are a priori balanced on any observable features in the affiliated team subreddit, which indicates similar loyalty to the team. To achieve this, we adopt matching techniques: for each single-group member, we match him/her with the most similar unmatched intergroup member from the same affiliated team, where similarity is based on all the observed features. Due to the observational nature, whether a member has intergroup contact or not is not randomly assigned. In other words, our study reflects the behavioral differences between those who engage in intergroup contact and those who do not.

Following prior studies on factors associated with fan behavior in online sports communities [219, 113, 167, 104], we consider the following observable feature set for matching: (1) the number of comments in the affiliated team subreddit, (2) the average time gap between comments, (3) the average length of comments, (4) the proportion of comments in the playoff season, and (5) the proportion of comments in the game threads (these game threads are created for discussions during a game). All the comments examined here are in the members' affiliated team subreddit. We collect all of these feature values for each season. The similarities between fans are estimated using the nearest neighbor matching technique [178]. Min-max normalization is applied to each feature before feeding it into the matching model so that no single feature dominates the matching. We do not include the feedback (upvotes/downvotes) that members received from the NBA subreddits for matching because it can be endogenous with the language used in the comments (e.g., comments with hate speeches may not get many upvotes).

To evaluate the outcome of our matching procedure, for each observable feature, we check distributional differences between the treatment group (intergroup members) and the control group (single-group members). We compare their empirical cumulative distributions before and after matching using the Mann-Whitney U test [112]. The results of the 2018, 2017, and 2016 seasons are summarized in Figure 4.15. A small p-value here indicates that there exists a significant difference between the treatment group and the control group. Prior to matching, the p-value for each feature is close to 0, implying that the distributions do differ between groups. After matching, we find no difference between the treatment group and the matched control group for any observable feature at the 5% significance level ($\alpha = 0.05$) in all three seasons, indicating that the data is balanced across all the covariates after matching.

Language usage analysis. The proportion of emotional words (i.e., positive emotions, negative emotions, and swear words) in members' comments are analyzed using the Linguistic Inquiry and Word Count software (LIWC [142]), a word frequency-based text analysis tool (see Table 4.3 for examples of emotional words detected using this software). The hate speech comments are identified using an automated hate speech detection model [39]. It is a multi-class classifier that can reliably separate hate speech from other offensive language (see Table 4.4 for examples of hate speech comments detected using this detection model). According to Davidson et al. [39], the model achieved an overall precision of 0.91, recall of 0.90, and F1 score of 0.90 on detecting hate speech tweets.

Fightin-Words model. To identify a list of distinguishing keywords that are over-used by inter-group or single-group members, we apply the Fightin-Words algorithm [126] to compare the word frequencies to the background frequencies found in the other fan group's corpora using the informative Dirichlet prior model. This method estimates the log-odds ratio of each word w between two corpora α and β given the frequencies obtained from the background corpus $\mathcal{D}$. Then the log-odds ratio $\delta_w^{(\alpha\text{-}\beta)}$ for word w can be estimated as:

$$\delta_w^{(\alpha\text{-}\beta)} = log\frac{c_w^\alpha + c_w^\mathcal{D}}{c^\alpha + c^\mathcal{D} - c_w^\alpha + c_w^\mathcal{D}} - log\frac{c_w^\beta + c_w^\mathcal{D}}{c^\beta + c^\mathcal{D} - c_w^\beta + c_w^\mathcal{D}}, \tag{4.1}$$

where c_w^α and c_w^β are the counts of word w in corpora α and β, c^α and c^β are the counts of all words in corpora α and β, $c_w^\mathcal{D}$ is the count of word w in the background corpus $\mathcal{D}$, and $c^\mathcal{D}$ is the count of all words in corpus $\mathcal{D}$. The Fightin-Words algorithm also provides an estimation for the variance of the log-odds ratio,

$$\sigma^2(\delta_w^{(\alpha\text{-}\beta)}) \backsim \frac{1}{c_w^{\alpha} + c_w^{\mathcal{D}}} + \frac{1}{c_w^{\beta} + c_w^{\mathcal{D}}}, \tag{4.2}$$

and the corresponding z-score can be calculated as follows:

$$Z = \frac{\delta_w^{(\alpha\text{-}\beta)}}{\sqrt{\sigma^2(\delta_w^{(\alpha\text{-}\beta)})}}. \tag{4.3}$$

The Fightin-Words model is known to outperform other traditional methods in detecting word usage differences between corpora, such as PMI (pointwise mutual information) [114] and TF-IDF [161], by not over-emphasizing fluctuations of rare words [126]. We use the comments made by intergroup members as the background corpus for single-group members and vice versa to identify differences in language usage by each fan group [2] . We rank each word by averaging its z-scores calculated by the Fightin-Words model across all 30 teams. A higher positive z-score indicates this word is over-used by single-group members, and a higher negative z-score means this word is over-used by intergroup members.

4.3.2 RQ2: Different Levels of Intergroup Contact

Matched intergroup members with different levels of intergroup contact. We define a user's level of intergroup contact based on the fraction of comments in the intergroup setting (/r/NBA). The fraction is calculated as the proportion of the number of comments the user made in the /r/NBA versus the total number of comments in NBA-related subreddits. Specifically, for each single-group member, we again apply the nearest neighbor matching technique to find five closest intergroup members in the same affiliated team and assign a label of 1, 2, 3, 4, or 5 to them based on their fraction of comments in /r/NBA in a complete NBA season. Different from pairing intergroup members and single-group members before, we do it with replacement because there are not enough intergroup members to conduct this matching uniquely. As such, an intergroup member can be matched to multiple single-group members. We compare the empirical cumulative distributions before and after matching for each level using the Mann-Whitney U test [35]. The

[2] Part of our code is borrowed from Jack Hessel's Fightin-Words model implementation [77].

results of the 2018, 2017, and 2016 seasons are presented in Figure 4.16, 4.17, and 4.18, respectively. We also assign a label of 0 to single-group members. A larger label indicates a higher level of intergroup contact that the member has in /r/NBA. We aggregate intergroup members at each level across all 30 team subreddits to compare their intragroup behavior. Note that the number of members at each level is the same, but some intergroup members may be counted more than once.

Regression analyses of the relationship between different levels of intergroup contact and language usage. To understand the relationship between members' intergroup contact level and language usage, we also conduct OLS regression analyses after the above matching procedure. The independent variables considered in the regression model are the same set of features used in matching intergroup and single-group members (Section 4.3.1). We standardize all independent variables before feeding into the regression model. Our full linear regression model to test each language usage pattern is shown below:

$$\begin{aligned} \textit{Proportion of language usage} \sim & \beta_0 + \beta_1 \textit{number of comments} + \beta_2 \textit{average comment hours gap} \\ & + \beta_3 \textit{average comment length} + \beta_4 \textit{proportion of playoff comments} \\ & + \beta_5 \textit{proportion of game thread comments} \\ & + \beta_6 \textit{fraction} + \beta_7 \textit{fraction} \times \textit{level1} + \beta_8 \textit{fraction} \times \textit{level2} \\ & + \beta_9 \textit{fraction} \times \textit{level3} + \beta_{10} \textit{fraction} \times \textit{level4} + \beta_{11} \textit{fraction} \times \textit{level5} \end{aligned}$$

The fraction in the linear regression model refers to the proportion of the number of comments the user made in the intergroup setting (/r/NBA) versus the total number of comments in NBA-related subreddits. All the control variables for matching intergroup members and single-group members are included. There are repeated measures in our regression model as an intergroup member can be matched to more than one single-group member. The average number of times for an intergroup member to be matched is 1.77 (excluding the intergroup members who never get matched). Among the intergroup members who are matched more than once, the average variance of their intergroup contact levels in different matches is 0.27. The small variance shows the consistency of our matching technique.

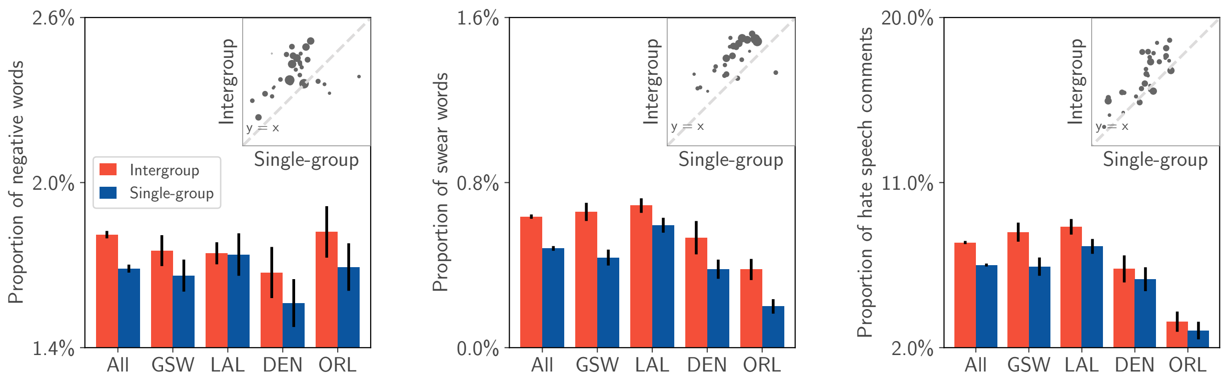
4.4 Results

In this section, we examine intragroup language differences between intergroup and single-group members (RQ1). We further discuss how different levels of intergroup contact relate to intragroup behavior (RQ2).

4.4.1 RQ1: Intragroup Language Differences

Figure 4.5 compares negative language usage between matched intergroup and single-group members. Intergroup members tend to use more negative language than single-group members, which is indicated by the use of more negative words (two-tailed t-test, $t = 6.23$, $p < 0.001$, 95% CI=0.08% to 0.16%; 26 out of 30 teams, two-tailed binomial test $p < 0.001$) and swear words (two-tailed t-test, $t = 3.51$, $p < 0.001$, 95% CI=0.02% to 0.08%; 28 out of 30 teams, two-tailed binomial test $p < 0.001$) based on lexicon analysis. We further compute the proportion of hate speech with an automated hate speech and offensive language detection model [39]. It is consistent that intergroup members also generate more hate speech (two-tailed t-test, $t = 10.44$, $p < 0.001$, 95% CI=1.00% to 1.46%; 26 out of 30 teams, two-tailed binomial test $p < 0.001$). These results indicate that intergroup members are more emotionally charged in their intragroup behavior compared with the single-group members and are somewhat different from the hypothesis that intergroup contact enhances empathy and perspective thinking [11, 145, 175]. Our results are consistent when excluding the NBA playoffs (see Section 4.8.1). We also compare positive language usage between intergroup members and single-group members and do not find a consistent trend at the 5% significance level (see Figure 4.21).

To further understand the difference between intergroup and single-group members in language usage, we identify a list of distinguishing words that are more likely to be used by intergroup



(a) Proportion of negative words. (b) Proportion of swear words. (c) Proportion of hate speech comments.

Figure 4.5: The comparison of language usage between intergroup and single-group members in the 2018 season. Intergroup members use more negative words (Figure 4.5a; two-tailed t-test, $t = 6.23$, $p < 0.001$, 95% CI=0.08% to 0.16%; 26 out of 30 teams, two-tailed binomial test $p < 0.001$) and swear words (Figure 4.5b; two-tailed t-test, $t = 3.51$, $p < 0.001$, 95% CI=0.02% to 0.08%; 28 out of 30 teams, two-tailed binomial test $p < 0.001$), and generate more hate speech comments (Figure 4.5c; two-tailed t-test, $t = 10.44$, $p < 0.001$, 95% CI=1.00% to 1.46%; 26 out of 30 teams, two-tailed binomial test $p < 0.001$). "All" is based on concatenating the samples from all 30 NBA team subreddits, and we also show the top two and bottom two teams ranked by the number of subscribers that have at least 100 single-group members. We further show the scatter plot of all 30 teams in the top right to illustrate that the findings are robust across teams (the size of the dot is proportional to the number of subscribers). Error bars represent standard errors. The results are consistent in the 2017 and 2016 seasons (see Figure 4.19 and Figure 4.20).

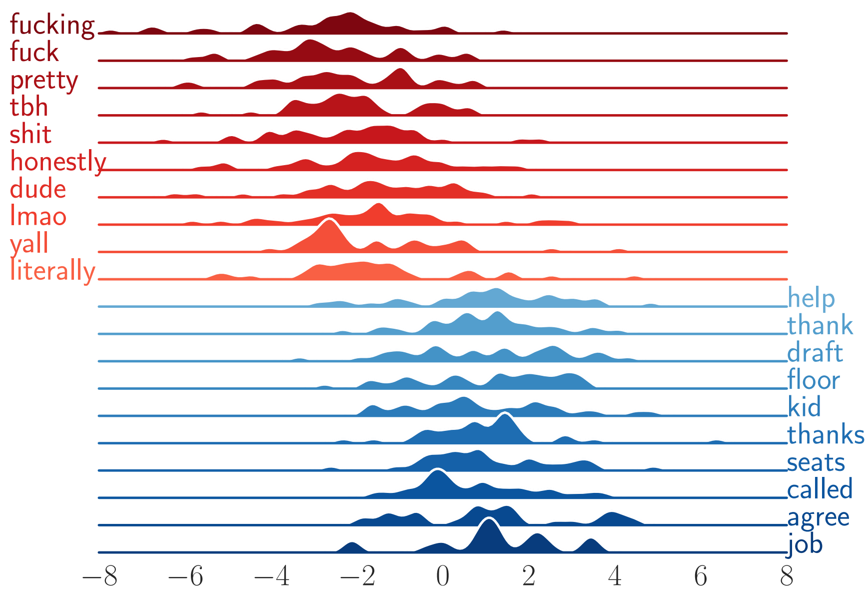


Figure 4.6: The top-10 over-represented words used by intergroup (red) and single-group (blue) members in the 2018 season. For each word, we show the distribution of the z-scores for all 30 teams calculated by the Fightin-Words algorithm [126].

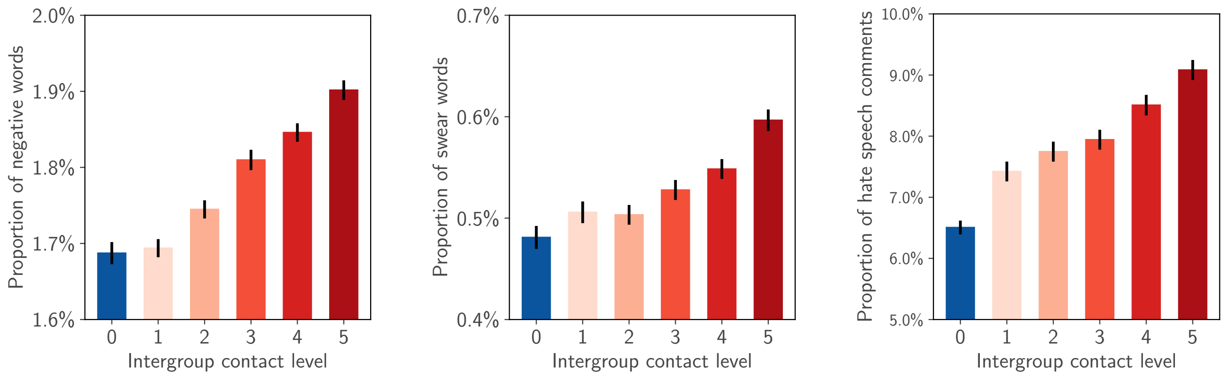
or by single-group members, using the Fightin-Words algorithm with the informative Dirichlet prior model [126]. Figure 4.6 lists the top-10 over-represented words used by intergroup and single-group members in the 2018 season. We rank each word by its average z-score calculated by the Fightin-Words algorithm across all 30 teams. A positive z-score indicates that this word is over-used by single-group members, while a negative z-score suggests that this word is over-used by intergroup members. Our results show that single-group members are more friendly and calm when commenting in the affiliated team subreddit and use more polite words, such as "agree", "thanks", and "help". Also, "seats" suggest that some single-group members are local fans, as they frequently discuss information about attending live games. In comparison, intergroup members use more swear words and talk more about the referees (likely complaining).

4.4.2 RQ2: Different Levels of Intergroup Contact

In addition to identifying the intragroup behavioral differences, our observational study allows us to quantify different levels of intergroup contact, which can be difficult to operationalize in experimental studies. Here, we examine the mechanisms of how increased levels of intergroup contact relate to differences in intragroup behavior.

Figure 4.7 shows language usage differences between members with different intergroup contact levels. Members of higher intergroup contact levels are generally more negative in language usage: They tend to use more negative words (mean = 1.69%, 1.69%, 1.76%, 1.80%, 1.85%, and 1.90%, respectively for labels from 0 to 6) and swear words (mean = 0.48%, 0.50%, 0.50%, 0.53%, 0.55%, and 0.59%, respectively for labels from 0 to 6), and generate more hate speech comments (mean = 6.51%, 7.61%, 7.72%, 7.77%, 8.43%, and 9.15%, respectively for labels from 0 to 6) in the affiliated team subreddit. However, the trends are not necessarily linear. For instance, intergroup members at level 1 do not show significant differences from single-group members in negative word usage, while intergroup members at level 5 present a significant jump from previous levels in negative words, swear words, and the use of hate speech.

Table 4.2 shows the results of regression analyses. The fraction of intergroup contact has



(a) Proportion of negative words. (b) Proportion of swear words. (c) Proportion of hate speech comments.

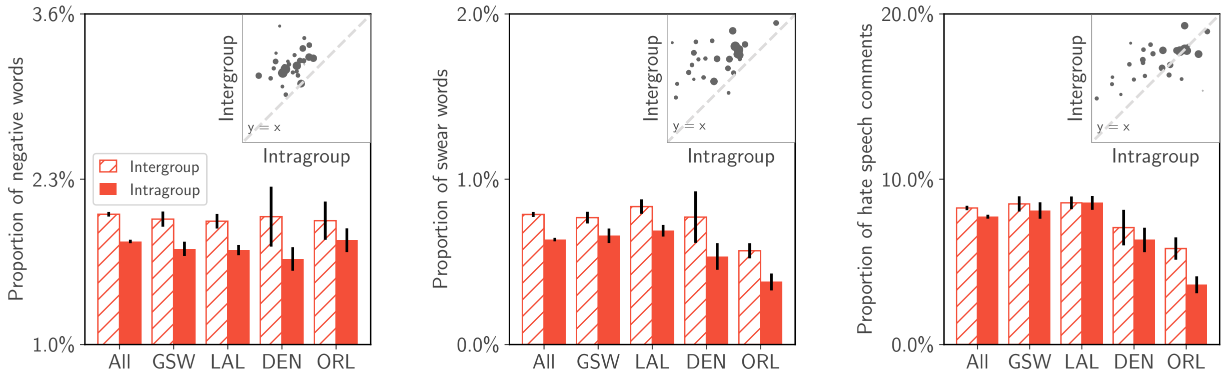
Figure 4.7: Intragroup language usage differences of members with different intergroup contact levels in the 2018 season. x-axis represents intergroup levels determined by the fraction of comments in /r/NBA. We observe a consistent monotonic pattern in the proportion of negative words (mean = 1.69%, 1.69%, 1.74%, 1.81%, 1.85%, and 1.90%, respectively for labels from 0 to 6;), swear words (mean = 0.48%, 0.51%, 0.50%, 0.53%, 0.55%, and 0.60%, respectively for labels from 0 to 6), and hate speech comments (mean = 6.51%, 7.42%, 7.75%, 7.94%, 8.51%, and 9.08%, respectively for labels from 0 to 6). The monotonic trend is consistent in the 2017 and 2016 season (see Figure 4.22 and Figure 4.23). Error bars represent standard errors.

a statistically significant positive coefficient in regressions for the proportion of negative words, swear words, and hate speech comments. Moreover, the coefficients for some interaction terms with levels are also statistically significant (e.g., level 5, β_{11}, is statistically significant in regressions for the proportion of negative words, swear words, and hate speech comments), indicating that nonlinear corrections are required. Note that the BIC score is consistently better by incorporating the interaction terms, although adjusted R^2 remains the same due to the fact that this is a very challenging regression task.

4.5 Intragroup Behavior vs. Intergroup Behavior of the Same User

Given the clear intragroup behavioral differences in language usage between intergroup and single-group members, we end our study by exploring the potential reasons behind them. We study the differences in language usage of the same user in his/her affiliated team subreddit vs. in /r/NBA. We compare the same person in two different contexts and naturally control for most of the confounding factors, which is also connected with the personality vs. situation debate [95].

Figure 4.8 shows that intergroup members use even more negative language in the intergroup setting, as they use more negative words (two-tailed t-test, $t = 9.39$, $p < 0.001$, 95% CI=0.17% to 0.26%; 30 out of 30 teams, two-tailed binomial test $p = 0.001$) and swear words (two-tailed t-test, $t = 13.69$, $p < 0.001$, 95% CI=0.22% to 0.29%; 29 out of 30 teams, two-tailed binomial test $p < 0.001$), and generate more hate speech comments than in the intragroup setting (two-tailed t-test, $t = 2.97$, $p = 0.003$, 95% CI=0.18% to 0.88%; 23 out of 30 teams, two-tailed binomial test $p = 0.005$). This indicates that fans are more hostile when facing fans from other teams than from the same team. This observation is robust after controlling for topics of discussion by only considering game threads (see Section 4.8.3 and Figure 4.13). Our observation suggests that although intergroup members are more emotional than single-group members in the affiliated subreddit, they are not as "outrageous" as they are in the intergroup setting. In comparison, when going to the intergroup setting and confronting fans from other team groups, they tend to have more negative interactions and troll each other.



(a) Proportion of negative words. (b) Proportion of swear words. (c) Proportion of hate speech comments.

Figure 4.8: Intergroup members use more negative language in the intergroup setting than in the intragroup setting in the 2018 season. Here we only consider the matched intergroup members in Figure 4.5 (i.e., the solid red bars in this figure are identical to the red bars in Figure 4.5). They use more negative words (two-tailed t-test, $t = 9.39$, $p < 0.001$, 95% CI=0.17% to 0.26%; 30 out of 30 teams, two-tailed binomial test $p = 0.001$) and swear words (two-tailed t-test, $t = 13.69$, $p < 0.001$, 95% CI=0.22% to 0.29%; 29 out of 30 teams, two-tailed binomial test $p < 0.001$), and generate more hate speech comments (two-tailed t-test, $t = 2.97$, $p = 0.003$, 95% CI=0.18% to 0.88%; 23 out of 30 teams, two-tailed binomial test $p = 0.005$). "All" is based on concatenating the samples from all 30 NBA team subreddits, and we also show the top two and bottom two teams ranked by the number of subscribers that have at least 100 single-group members. We further show the scatter plot of all 30 teams in the top right to illustrate that the findings are robust across teams (the size of the dot is proportional to the number of subscribers). Error bars represent standard errors. The results are consistent in the 2017 and 2016 seasons (see Figure 4.24 and Figure 4.25).

These observations may provide explanations for the characteristics of intergroup fans in intragroup behavior. Prior studies suggest that negative intergroup contact is more influential in shaping people's attitudes and may curb the contact's ability to reduce prejudice [64, 137, 172]. The emotionally charged intergroup contact from the intergroup setting may connect to intergroup fans' more sentimental attitudes in their affiliated team subreddit. It requires further research to establish the causal link here, but the fact that we are able to observe these contrasts demonstrates the importance of such observational studies based on real interactions over substantial time periods.

4.6 Discussion

Although most previous studies have focused on the role of intergroup contact in changing attitudes of individual members, our study highlights the fact that users selectively become intergroup members, and intergroup and single-group members in turn interact with each other in their affiliated group. Such interaction can potentially influence members' language usage and shape the entire group. Moreover, we demonstrate a variety of ways in which intergroup contact levels can moderate intragroup behavior. This indicates that observational studies can provide important complementary evidence to experimental studies on this topic because interventions can hardly result in deep and regular contact. Novel methodologies are required to further bridge the gap between observational studies and experimental studies.

Could social media be driving polarization? Twitter, Reddit, and Facebook have become important platforms for political discussions as well as misinformation [194, 67]. Service providers are designing new features that would actively expose people to opposing views. For example, Twitter recently experimented with new algorithms that would promote alternative viewpoints in Twitter's timeline to address misinformation and reduce the effect of echo chambers [158]. However, the proposed solution may increase polarization. Unlike decades of offline experiments which mostly indicate intimate contact between members of rival groups across an extended period can produce positive effects, the results in Bail et al. [10] and our paper suggest that encountering views from opposing groups online may make them even more wedded to their own views. There are several possible explanations of this contrast by examining the possible mechanisms that intergroup contact affects individual attitudes. First, the comments created on social media are usually brief. These short messages without enough context may not enhance knowledge about opposing groups. Several studies suggest that people interpret short text-based messages inconsistently, which creates significant potential for miscommunication [210, 121, 94]. Second, the discussion structure may facilitate the spread of negative interaction. Cheng et al. [30] examine the evolution of discussions on CNN.com and show that existing trolling comments in a discussion thread significantly increase the likelihood of future trolling comments. The spread of negativity will increase rather than reduce people's anxiety levels when facing opposing groups. Third, the anonymous, spontaneous nature of communications on social media may not be conducive to cultivating empathy. In an experiment designed to examine the relationship between the presence of mobile devices and the quality of social interactions, results show that participants who have conversations in the absence of mobile devices report high levels of empathetic concern [124]. In summary, intergroup contact may lead to diverging outcomes depending on the environment and the nature of the contact. Further research is required to examine these possibilities and understand how social and technical design decisions can influence the outcomes.

Can we design better online discussion forums for different groups? The findings in this work indicate that social platforms designers should consider strategies to shape intergroup contact online. As hinted above, it is insufficient to recommend users to follow members of opposing groups or opposing views. Better design strategies need to be experimented for encouraging civil and extended intergroup contact. It would also be useful to take into account how different levels of intergroup contact may moderate individual opinions differently. Content moderation can be a promising area for future studies in the context of intergroup contact [96]. For instance, Matias [119] shows the displaying community rules can prevent harassment, but how to reduce negative intergroup contact remains an open question. Similarly, a powerful way of spreading online information is through social consensus cues and online endorsement (e.g., upvotes, likes). However, promoting content with the highest popularity can sometimes be problematic. Earlier research suggests that tweets with more sentiment-laden words are likely to be favorited or retweeted, and politicians may intentionally use this strategy to maximize impacts on Twitter [19, 184, 185]. Our study also finds that intergroup members receive better feedback from their affiliated team subreddit even though they use more negative language (Figure 4.12). This type of behavior can generate negative reactions from opposing groups and push the whole discussion to cycle towards more emotionally-laden and potentially polarizing content. It is thus important to develop comment ranking systems that are cognizant of intergroup contact and prioritize constructive interactions.

Limitations. Our findings are subject to the following limitations. First, the causal relationship between intergroup contact and negative language usage is not entirely clear. Due to the nature of our observational study, whether a member has intergroup contact is not randomly assigned. Though we match users based on a series of activity features, an important confounding factor could be that people who seek intergroup contact are inherently different from those who do not.

Second, our definition of intergroup contact entails that we focus on relatively active users. Thus, we cannot observe indirect intergroup contact, such as browsing /r/NBA. Prior studies have shown that indirect contact, such as imagining oneself interacting with an out-group member and observing an in-group member interacting with an out-group member [191, 46], may also shape human behavior. It also follows that intergroup members have more activities on NBA-related discussion forums as a whole than single-group members. We want to note that the nature of intergroup contact is that given the same amount of time in life, individuals with intergroup contact put more effort into intergroup contact than those without such contact.

Third, we use a coarse proxy to consider any users who have posted in our intergroup setting (/r/NBA) as intergroup members, and study the language differences in the intragroup setting at the user level instead of at the dyad level. However, some comments created by the fans in /r/NBA may be replies to the fans who are from the same team or do not have a team affiliation. More in-depth characterization of different types of discussions happened in the intergroup setting is required to further understand the differences observed in this study.

Fourth, the observations made in this study are limited to Reddit NBA fan groups. The sports context might be a strong case for understanding intergroup relations, as all the teams are created to compete with each other for the final championship. The expression of hostile attitudes towards opposing sides are culturally acceptable and even encouraged [33]. We should expect less negative intergroup contact between groups that do not contend for the same resources (e.g., music fans of different musicians may not have conflicts with each other at all). However, politics, especially in a polarized bipartisan situation, share common properties with the sports context. Examining the generalization of our results in other contexts is a promising avenue for future work.

Finally, the negative language observed in our study may not necessarily bring negative effects to the community. Prior studies suggest the main reason people use swearing words on the online platform is to express some strong emotions, such as anger and frustration [196, 42, 30]. Heath et al. [75] examine users' emotional selection in memes when emotion is manipulated and observe that people prefer the version of the story that produced the highest levels of disgust and evoke strong sentiment. Jay [87] further argues that only when cursing occurs in the form of insults toward others, such as name-calling, harassment, and hate speech, it becomes harmful. In addition, earlier literature suggests that the reason people use swearing words on the online platform may relate to Internet humor, such as jokes and memes. Posting humorous content on the Internet has the potential to engage other users in art activities that are closely connected to their lives and receive online endorsements [212, 165]. Attempting to be funny could be another reason that intergroup members adopt a more negative language style than single-group members. However, as pointed out by Lockyer and Pickering [108], a significant proportion of Internet humor has offensive, sexism, and racism content, and its consequences are often overlooked.

4.7 Conclusion

In this paper, by applying our computational framework to NBA-related discussion forums on Reddit, we identify clear language differences between intergroup and single-group members in their affiliated group (the intragroup setting). We find that in the affiliated team subreddit, intergroup members tend to use more negative and swear words, and generate more hate speech comments compared with single-group members. Moreover, we quantify different levels of intergroup contact for each intergroup member based on the fraction of their comments in the intergroup setting (/r/NBA). Interestingly, the level of intergroup contact can relate to differences in language usage in different ways, though the relationship is mostly monotonic. To further shed light on the behavior of intergroup members, we also compare the language usage of intergroup members between the intragroup setting and the intergroup setting. This setup naturally controls for the subject because we compare the same person across two different environments. We observe that intergroup members are even more negative and more likely to swear in the intergroup setting. As intergroup contact in online platforms becomes increasingly common and can play an important role in opinion formation, our work demonstrates how observational studies can provide complementary evidence to experimental studies on this topic.

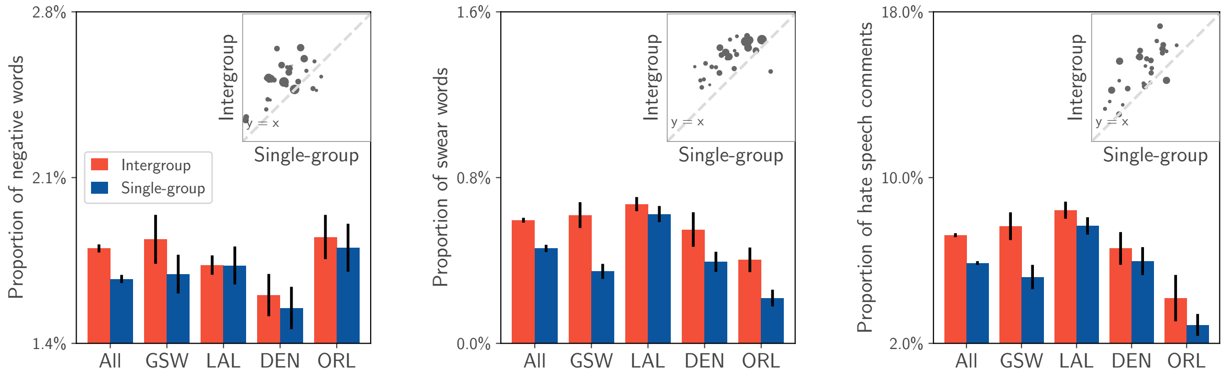
4.8 Appendix

4.8.1 The language differences between intergroup members and single-group members are consistent when we exclude the NBA playoffs

In Section 4.4.1, we find that intergroup members use more negative and swear words and generate more hate speech comments compared to single-group members in the intragroup setting. It is possible that comments posted during the playoff season play a significant factor, as the activity in /r/NBA peaks in the playoff season [219]. Figure 4.9, 4.10, and 4.11 show that our results are consistent in all three seasons when we exclude the playoff season.

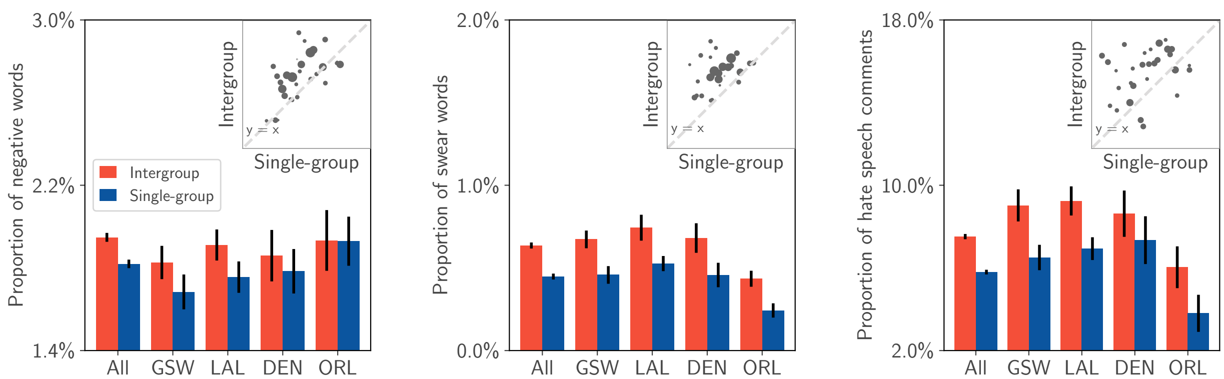
4.8.2 Intergroup members receive better feedback than single-group members

Figure 4.12 shows that intergroup members receive better feedback in the intragroup setting than single-group members in all three seasons (two-tailed t-test, $t = 15.68$, $p < 0.001$, 95% CI=0.048 to 0.062%, 29 out of 30 teams, two-tailed binomial test $p < 0.001$ for the 2018 season; two-tailed t-test, $t = 12.56$, $p < 0.001$, 95% CI=0.043 to 0.058%, 30 out of 30 teams, two-tailed binomial



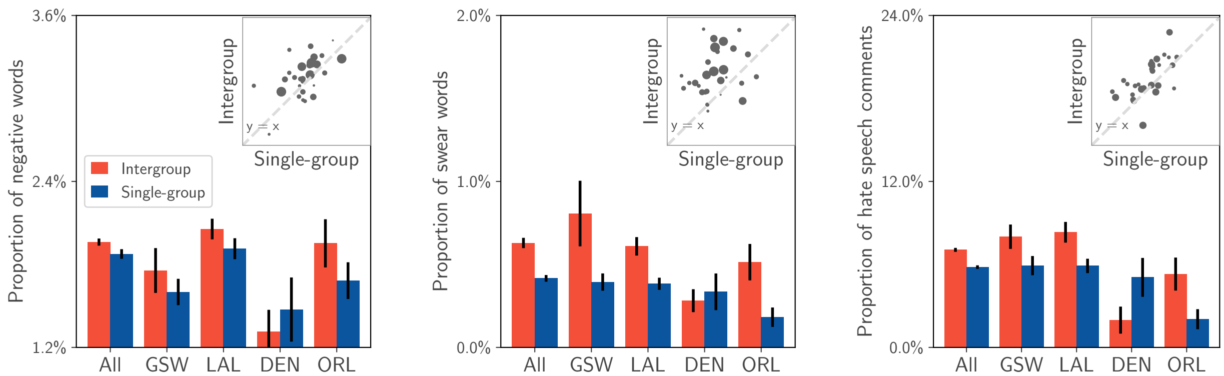
(a) Proportion of negative words. (b) Proportion of swear words. (c) Proportion of hate speech comments.

Figure 4.9: The comparison of language usage between intergroup and single-group members in the 2018 season when excluding the NBA playoffs. Intergroup members use more negative words (two-tailed t-test, t = 4.31, $p < 0.001$, 95% CI=0.07% to 0.19%; 24 out of 30 teams, two-tailed binomial test $p = 0.001$) and swear words (two-tailed t-test, t = 3.43, $p < 0.001$, 95% CI=0.04% to 0.14%; 28 out of 30 teams, two-tailed binomial test $p < 0.001$) and generate more hate speech comments (two-tailed t-test, t = 10.05, $p < 0.001$, 95% CI=1.38% to 2.05%; 24 out of 30 teams, two-tailed binomial test $p = 0.001$). Error bars represent standard errors.



(a) Proportion of negative words. (b) Proportion of swear words. (c) Proportion of hate speech comments.

Figure 4.10: The comparison of language usage between intergroup and single-group members in the 2017 season when excluding the NBA playoffs. Intergroup members use more negative words (two-tailed t-test, t = 4.31, $p < 0.001$, 95% CI=0.07% to 0.19%; 24 out of 30 teams, two-tailed binomial test $p = 0.001$) and swear words (two-tailed t-test, t = 3.43, $p = 0.002$, 95% CI=0.04% to 0.14%; 28 out of 30 teams, two-tailed binomial test $p < 0.001$) and generate more hate speech comments (two-tailed t-test, t = 10.05, $p < 0.001$, 95% CI=1.38% to 2.05%; 23 out of 30 teams, two-tailed binomial test $p < 0.005$). Error bars represent standard errors.



(a) Proportion of negative words.

(b) Proportion of swear words.

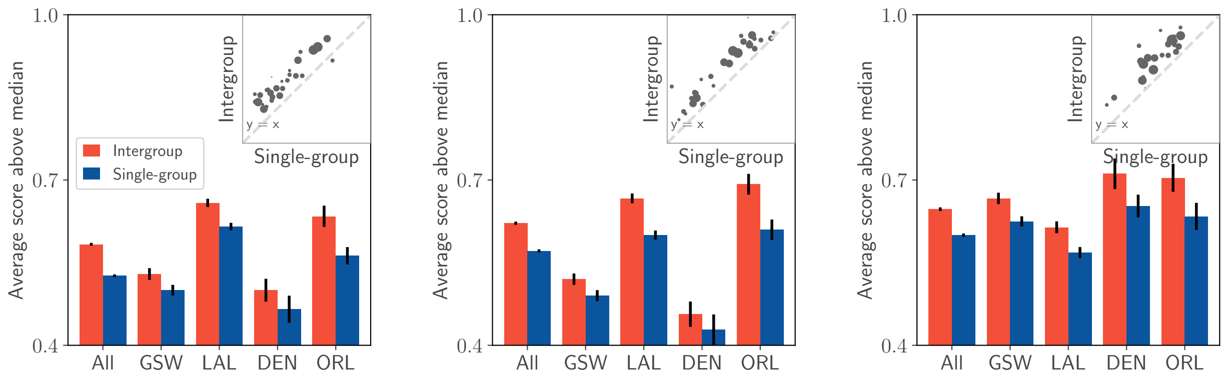
(c) Proportion of hate speech comments.

Figure 4.11: The comparison of language usage between intergroup and single-group members in the 2016 season when excluding the NBA playoffs. Intergroup members use more negative words (two-tailed t-test, $t = 1.99$, $p = 0.04$, 95% CI=0.01% to 0.17%; 21 out of 30 teams, two-tailed binomial test $p = 0.04$) and swear words (two-tailed t-test, $t = 3.06$, $p = 0.002$, 95% CI=0.04% to 0.18%; 24 out of 30 teams, two-tailed binomial test $p = 0.001$) and generate more hate speech comments (two-tailed t-test, $t = 7.95$, $p < 0.001$, 95% CI=0.92% to 1.53%; 25 out of 30 teams, two-tailed binomial test $p < 0.001$). Error bars represent standard errors.

test $p < 0.001$ for the 2017 season; two-tailed t-test, $t = 12.43$, $p < 0.001$, 95% CI=0.046 to 0.064%, 30 out of 30 teams, two-tailed binomial test $p < 0.001$ for the 2016 season). Comment feedback is defined by whether the comment score (#upvotes- #downvotes) is above the median score of that team subreddit in that month, which accounts for the differences across subreddits. This observation suggests that using negative language is likely to draw attention in the corresponding team subreddits.

4.8.3 Intergroup members are more emotional in the intergroup setting than in the intragroup setting when controlling for the discussion topic

In Section 4.4.2, we find that intergroup members use more negative language in the intergroup setting than in the intragroup setting. This difference may occur due to the fact that more heated topics are discussed in /r/NBA than in team subreddits. To control for this factor, we further limit our comparison to the game threads in both settings. Game threads are important components of NBA-related team subreddits to facilitate game-related discussions during NBA games. In practice, each game has a game thread in the home-team subreddit, the away team subreddit, and the overall /r/NBA. Figure 4.13 shows the language usage difference of intergroup members in the game threads of the intergroup and intragroup setting. Only members who made comments in the game threads of both settings are included in this analysis (2118 members for the 2018 season, 1495 members for the 2017 season, and 1289 members for the 2016 season). In all three seasons, it is consistent that intergroup members use more negative words in the game threads of the intergroup setting than of the intragroup setting. We do not compare the language usage patterns per team in this analysis as there are teams with less than 20 members.

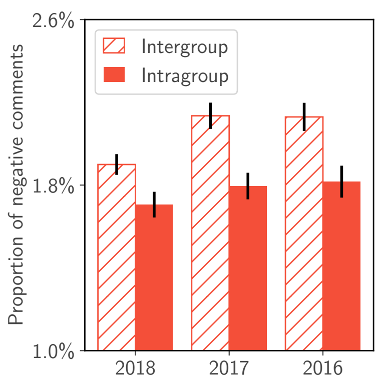


(a) The 2018 season.

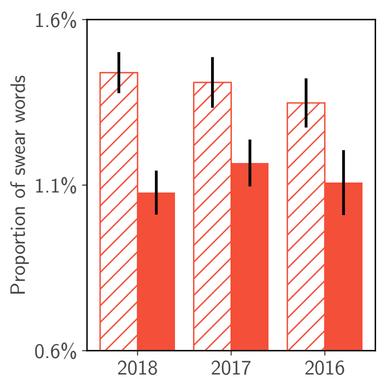
(b) The 2017 season.

(c) The 2016 season.

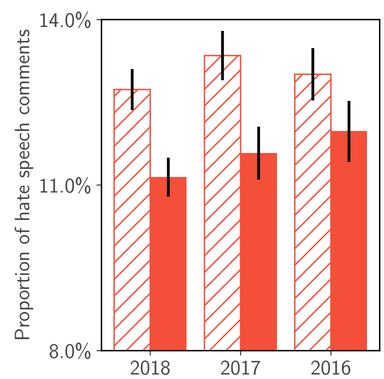
Figure 4.12: The comparison of feedback received from affiliated team subreddit between intergroup and single-group members in the 2018, 2017, and 2016 seasons. Intergroup members receive better feedback than single-group members in all three seasons (two-tailed t-test, $t = 15.68$, $p < 0.001$, 95% CI=0.048 to 0.062%, 29 out of 30 teams, two-tailed binomial test $p < 0.001$ for the 2018 season; two-tailed t-test, $t = 12.56$, $p < 0.001$, 95% CI=0.043 to 0.058%, 30 out of 30 teams, two-tailed binomial test $p < 0.001$ for the 2017 season; two-tailed t-test, $t = 12.43$, $p < 0.001$, 95% CI=0.046 to 0.064%, 30 out of 30 teams, two-tailed binomial test $p < 0.001$ for the 2016 season). Error bars represent standard errors.



(a) Proportion of negative words.



(b) Proportion of swear words.



(c) Proportion of hate speech comments.

Figure 4.13: The observation that intergroup members are more negative in the intergroup setting than in the intragroup setting is robust after controlling for topics of discussion by only considering game threads. Intergroup members use more negative words (two-tailed t-test, $t = 2.43$, $p = 0.015$, 95% CI=0.04% to 0.35% for the 2018 season; two-tailed t-test, $t = 3.76$, $p < 0.001$, 95% CI=0.16% to 0.52% for the 2017 season; two-tailed t-test, $t = 3.04$, $p = 0.002$, 95% CI=0.11% to 0.51% for the 2016 season) and swear words (two-tailed t-test, $t = 3.99$, $p < 0.001$, 95% CI=0.18% to 0.54% for the 2018 season; two-tailed t-test, $t = 2.34$, $p = 0.023$, 95% CI=0.04% to 0.32% for the 2017 season; two-tailed t-test, $t = 1.96$, $p = 0.048$, 95% CI=0.00% to 0.48% for the 2016 season) and generate more hate speech comments (two-tailed t-test, $t = 3.10$, $p = 0.002$, 95% CI=0.06% to 2.59% for the 2018 season; two-tailed t-test, $t = 2.70$, $p = 0.007$, 95% CI=0.05% to 3.05% for the 2017 season; two-tailed t-test, $t = 1.42$, $p = 0.155$, 95% CI=0.00% to 2.46% for the 2016 season) in the game threads of the intergroup setting than that of the intragroup setting. We do not compare the language usage patterns per team in this analysis, as there are teams with less than 20 members after this control. Error bars represent standard errors.

Table 4.2: Regression analyses for the proportion of negative words, swear words, and hate speech comments. The variables used for matching intergroup and single-group members are also included for control. For each of the analyses, the fraction of intergroup contact has a positive coefficient. The number of stars indicates p-values, ***: $p < 0.001$, **: $p < 0.01$ *: $p < 0.05$.

	Prop. of negative words		Prop. of swear words		Prop. of hate speech	
Variable	Reg. 1	Reg. 2	Reg. 1	Reg. 2	Reg. 1	Reg. 2
Control						
number of comments	0.003*	0.003*	0.023***	0.022***	0.002*	0.002*
average comment hours gap	−0.004***	0.004***	−0.017***	−0.017***	−0.003***	−0.003***
average comment length	−0.028***	−0.028***	−0.000	−0.000	−0.042***	−0.042***
Prop. of playoff comments	0.006***	0.006***	0.019***	0.018***	0.004***	0.004***
Prop. of game thread comments	0.053***	0.053***	0.090***	0.089***	0.028***	0.028***
Fraction						
fraction	0.007***	0.005***	0.027***	0.019***	0.004***	0.003***
Levels						
fraction × level1		−0.002		−0.001		0.001
fraction × level2		0.001		−0.001		−0.001
fraction × level3		0.002**		0.003		0.000
fraction × level4		0.002**		0.005*		0.001
fraction × level5		0.002***		0.010***		0.002***
intercept	0.052***	0.053***	0.040***	0.042***	0.028***	0.028***
Adjusted R^2	0.172	0.172	0.033	0.033	0.138	0.138
BIC	-147675	-147641	-45838	-45803	-164859	-164835

Table 4.3: A sample of positive, negative and swear words in the Linguistic Inquiry and Word Count dictionary (LIWC [142]). Words ending with "*" match any string with the same prefix.

Positive	credit*, graced, attract*, graceful*, terrific*, bonus*, affection*, humour*, delicious*, love, openness, sweetheart*, bless*, bold*, madly, fine, friend*, hurra*, ready, trust*, secur*, won, improving, fiesta*, dynam*, toleran*, sunniest, optimal*, helpful*, neat*, enthus*, joking, favour*, giving, agreeab*, easiness, supportive*, frees*, graces, gentler
Negative	ignor*, aggravat*, unattractive, scary, attack*, offend*, grief, fright*, domina*, unfriendly, violat*, grave*, nast*, suck, shock*, sucker*, impatien*, wept, heartless*, shake*, battl*, moron*, vanity, aggress*, masochis*, unsure*, screw*, lost, losing, mocker*, envie*, sadness, nag*, timid*, afraid, hateful*, turmoil, agoniz*, obnoxious*, pain
Swear	prick*, dyke*, tit, cock, dicks, butt, bloody, dick, sob, asshole*, pussy*, screw*, suck, wanker*, mofo, fucks, shit*, bastard*, arse, butts, darn, sucked, jeez, nigger*, fucker*, arses, ass, hell, crappy, dang, motherf*, dumb*, heck, crap, tits, queer*, bitch*, sonofa*, titty, fuckin*

Table 4.4: Examples of hate speech comments detected with the automated detection model [39].

Kevin Sorbo's Hercules was such a pussy magnet
Because I want losers like you to fuck off
Same reason KAT ass rapes our team everytime we play.
Holy shit Sabonis is stuntin' like his daddy right now with these passes
That's like me saying you're a dumbass because your team is currently shit. Sorry pal, don't talk about basketball until you make the playoffs.
Melo gets NO calls ever - I've always said he should bitch more
No defense, no rebounding. Same old shit. Embarrassing
Fuck the Celtics!
now since your dumb ass sees that it doesnt make difference like ive been saying what is your excuse?
Tyler Ennis you bum. Worst player in the league.

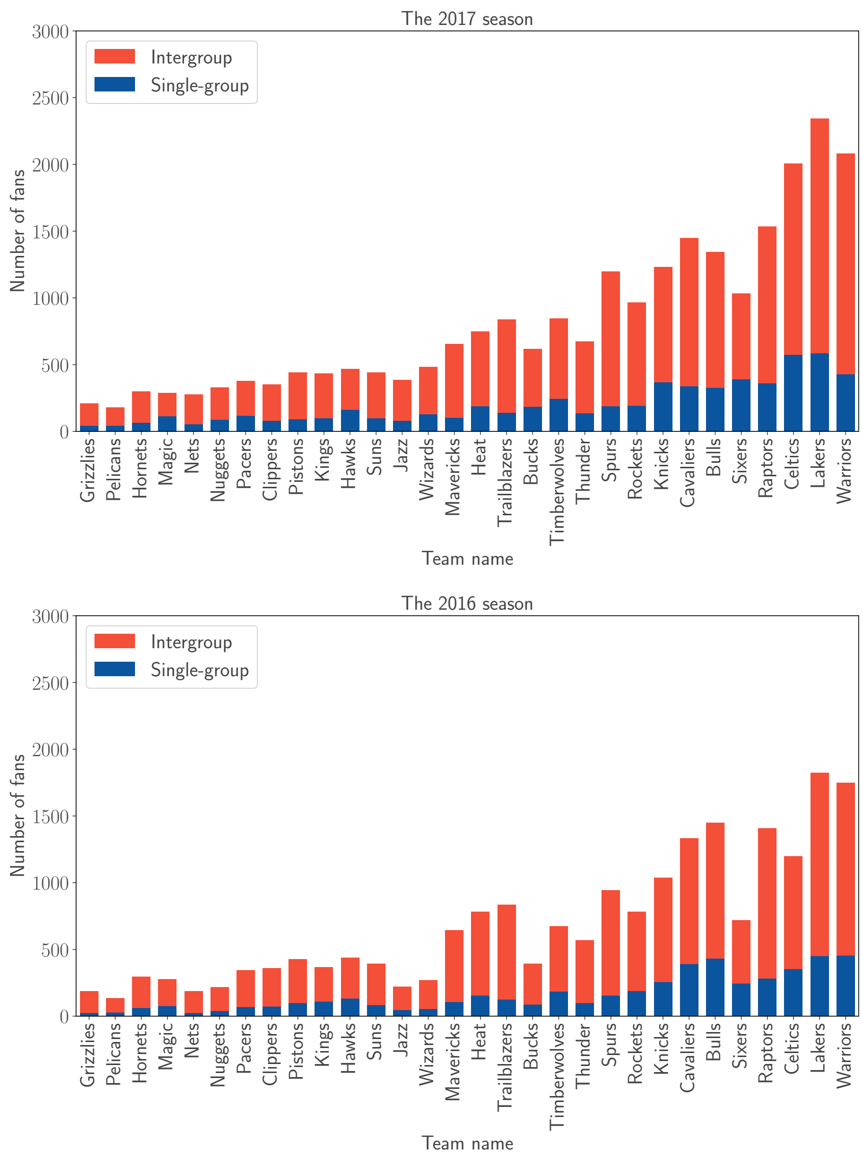
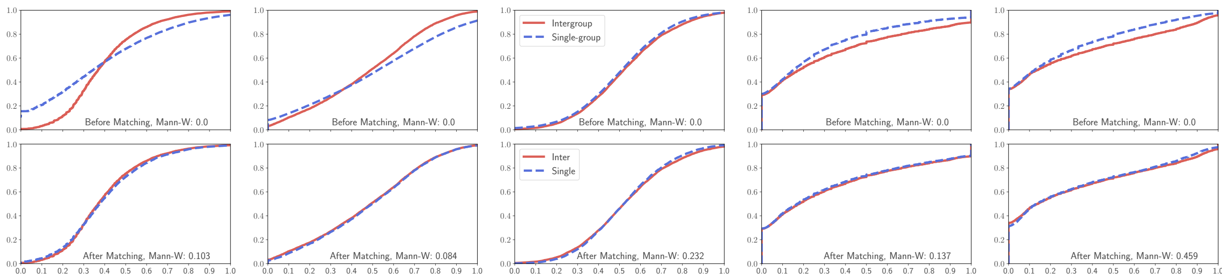
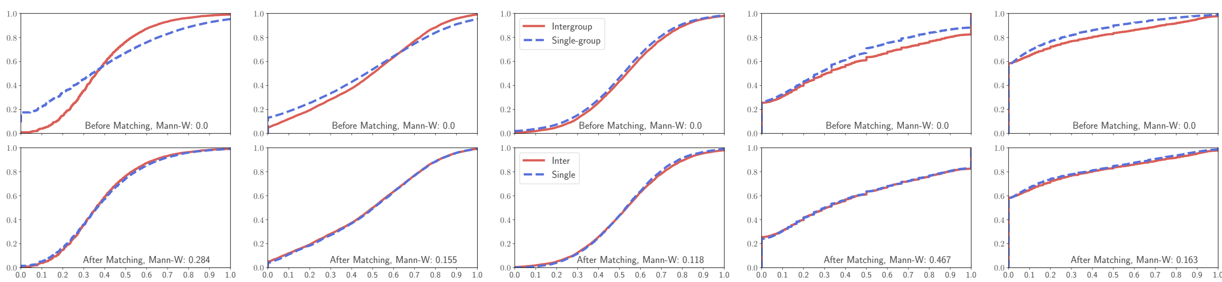


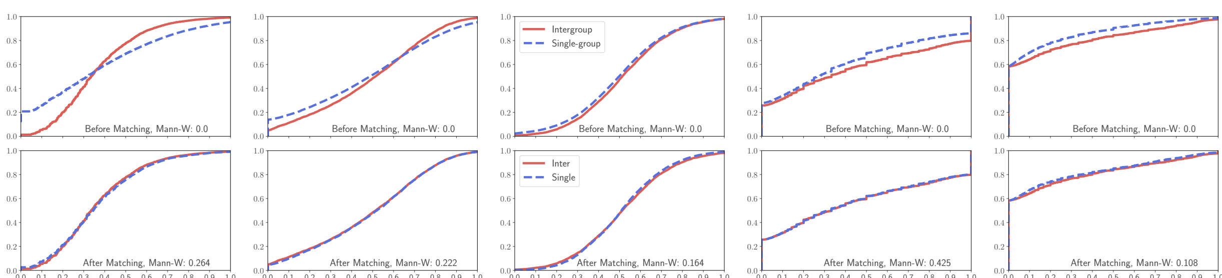
Figure 4.14: The number of intergroup and single-group members affiliated with each NBA team in the 2017 and 2016 seasons. We rank 30 team subreddits by the number of subscribers each team has by the end of the 2018 season.



(a) Season 2018

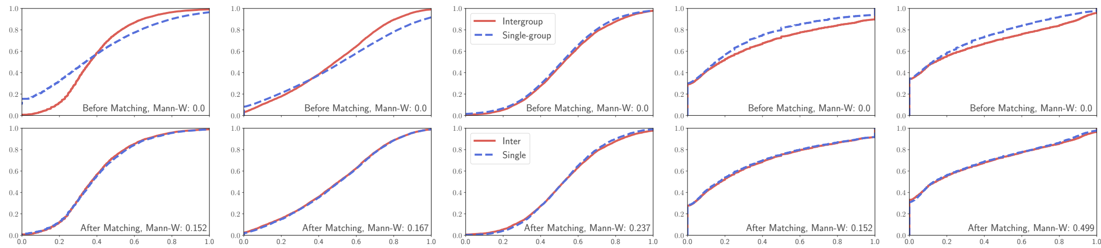


(b) Season 2017

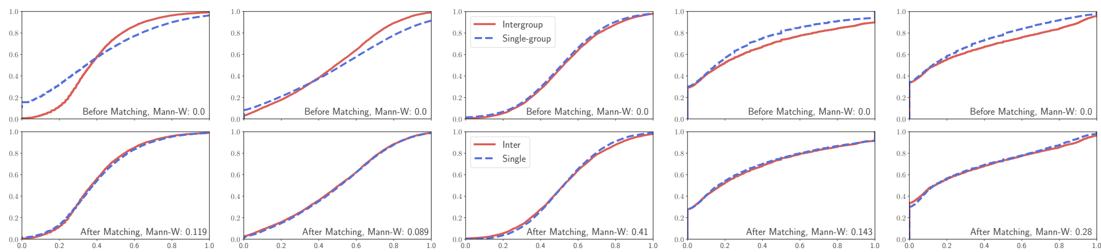


(c) Season 2016

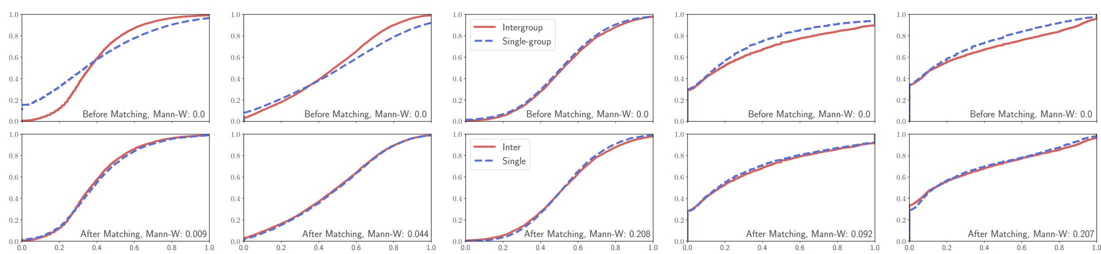
Figure 4.15: Empirical cumulative distribution of each activity feature before and after the matching technique in the 2018, 2017, and 2016 seasons. The activity features from left to right are the number of comments, the average hour gap between comments, the average comment length, the proportion of playoff comments, and the proportion of game thread comments. The corresponding p-values of the Mann-Whitney tests are also reported. Recall that a small p-value indicates that there is dependence between the treatment and control groups (relative frequencies are different). Prior to matching, each p-value is very close to 0.0. After the matching, at the 0.05 significance level ($\alpha = 0.05$), we find no dependence on the group label for any activity feature observed in the matched dataset. These trends are consistent in all three seasons.



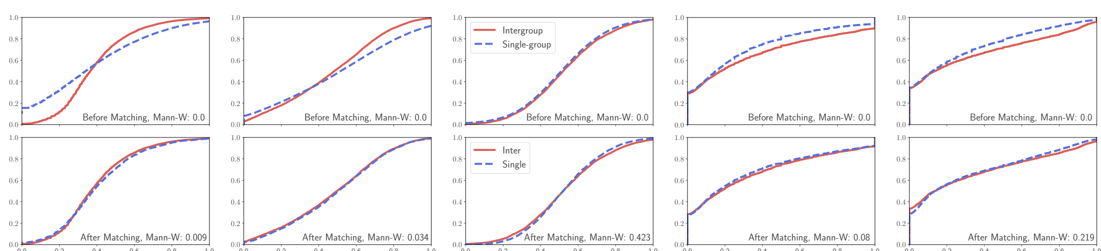
(a) Matching results for level 1.



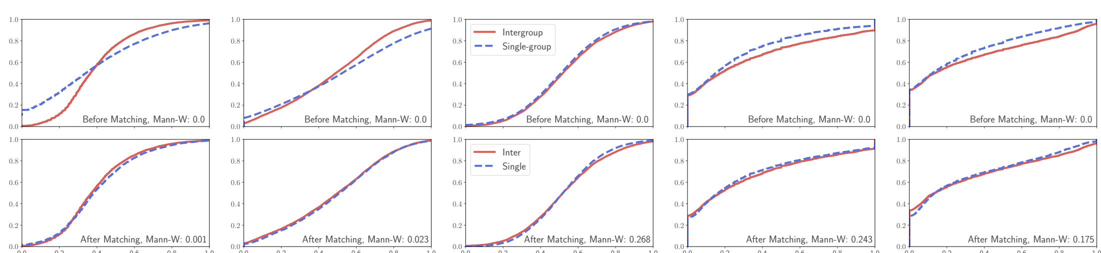
(b) Matching results for level 2.



(c) Matching results for level 3.

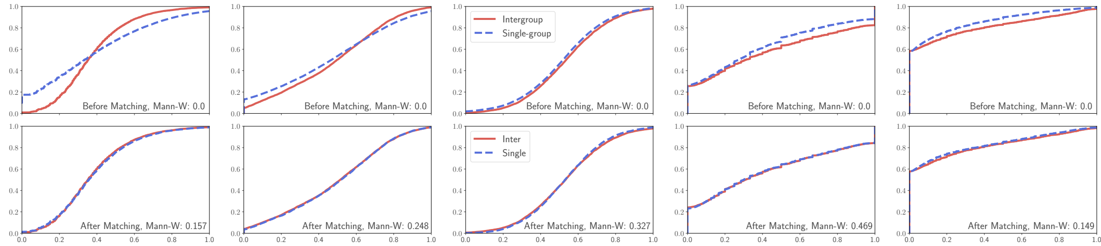


(d) Matching results for level 4.

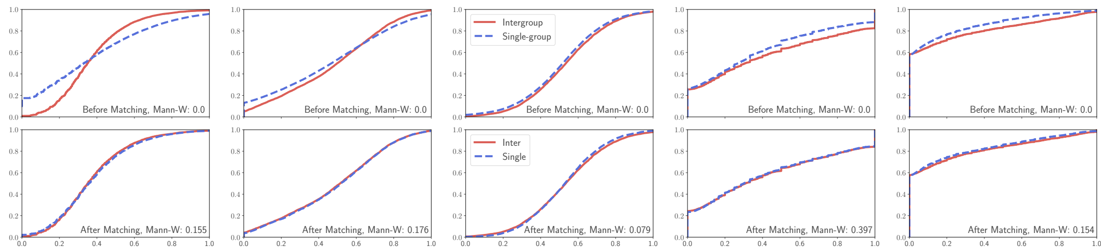


(e) Matching results for level 5.

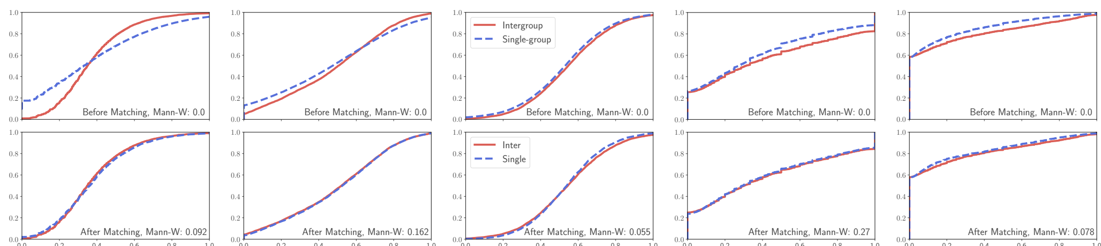
Figure 4.16: Empirical cumulative distribution of each activity feature before and after the matching technique from level 1 to level 5 in the 2018 season. The activity features from left to right are the number of comments, the average hour gap between comments, the average comment length, the proportion of playoff comments, and the proportion of game thread comments.



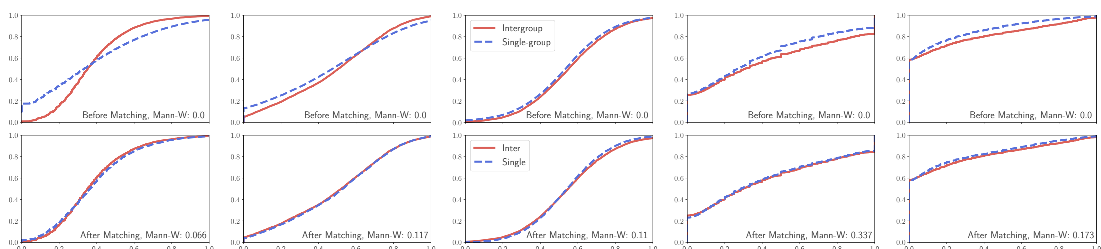
(a) Matching results at level 1.



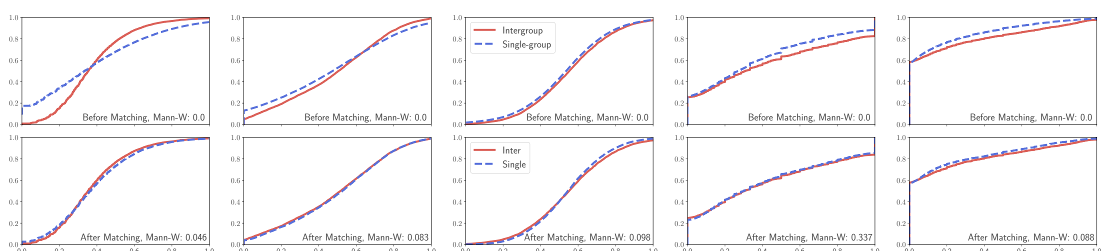
(b) Matching results at level 2.



(c) Matching results at level 3.

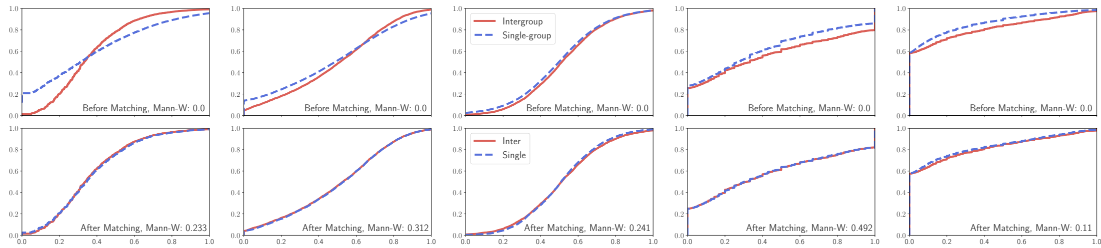


(d) Matching results at level 4.

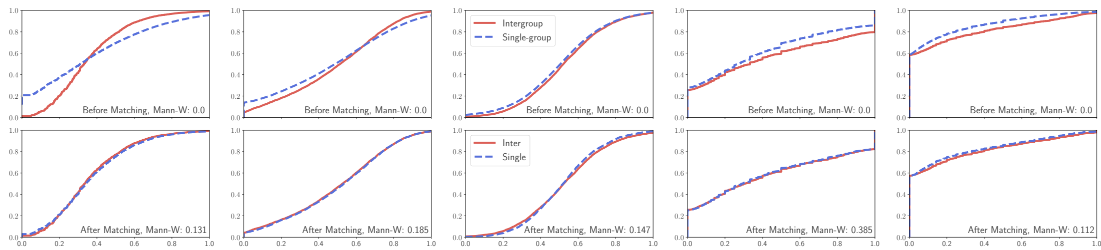


(e) Matching results at level 5.

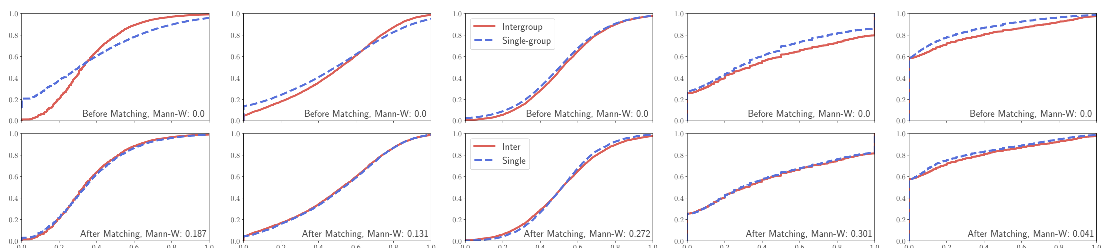
Figure 4.17: Empirical cumulative distribution of each activity feature before and after the matching technique from level 1 to level 5 in the 2017 season. The activity features from left to right are the number of comments, the average hour gap between comments, the average comment length, the proportion of playoff comments, and the proportion of game thread comments.



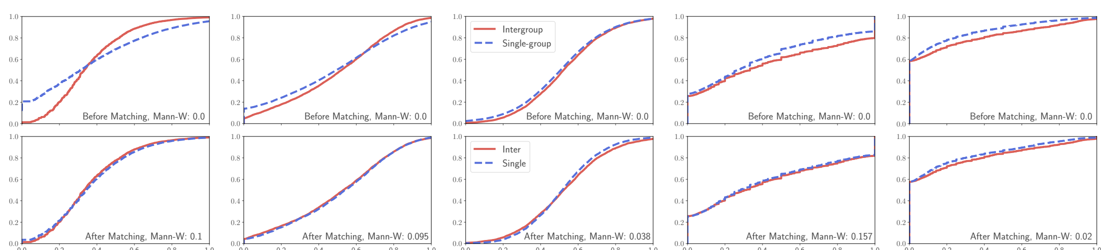
(a) Matching results at level 1.



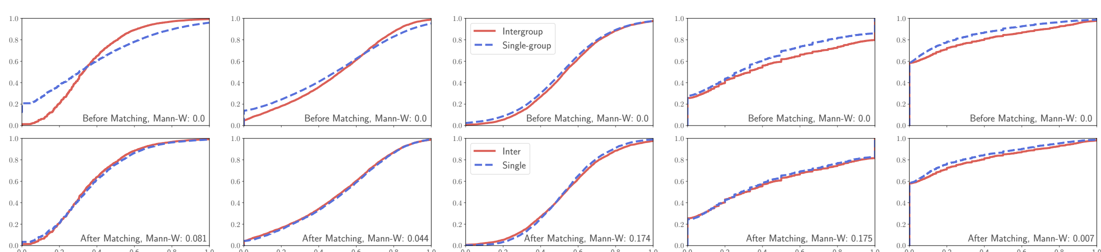
(b) Matching results at level 2.



(c) Matching results at level 3.

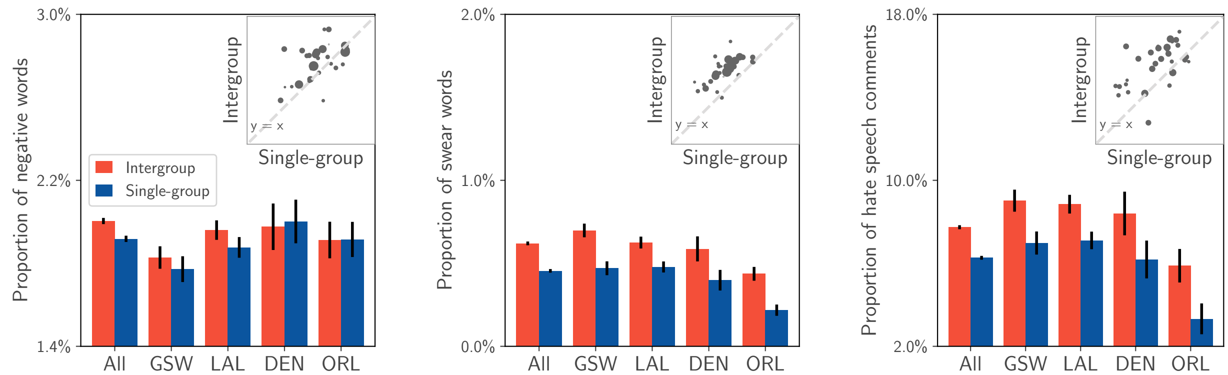


(d) Matching results at level 4.



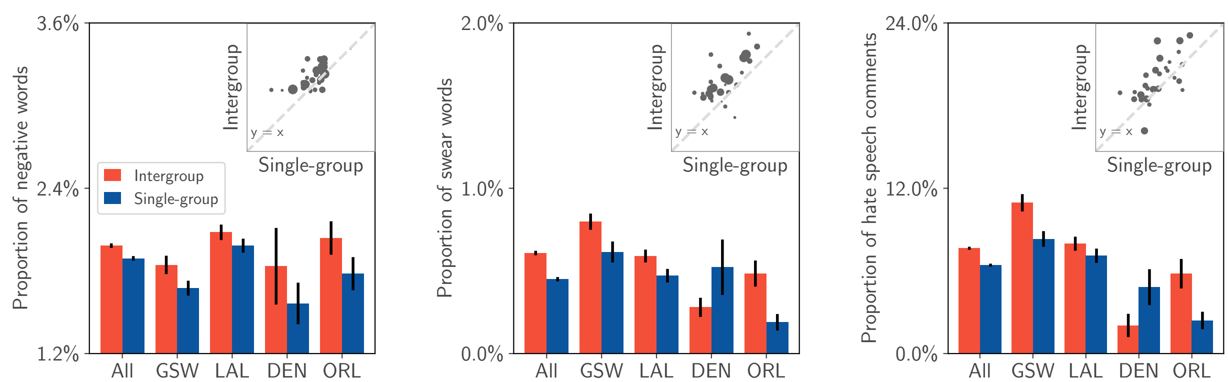
(e) Matching results at level 5.

Figure 4.18: Empirical cumulative distribution of each activity feature before and after the matching technique from level 1 to level 5 in the 2016 season. The activity features from left to right are the number of comments, the average hour gap between comments, the average comment length, the proportion of playoff comments, and the proportion of game thread comments.



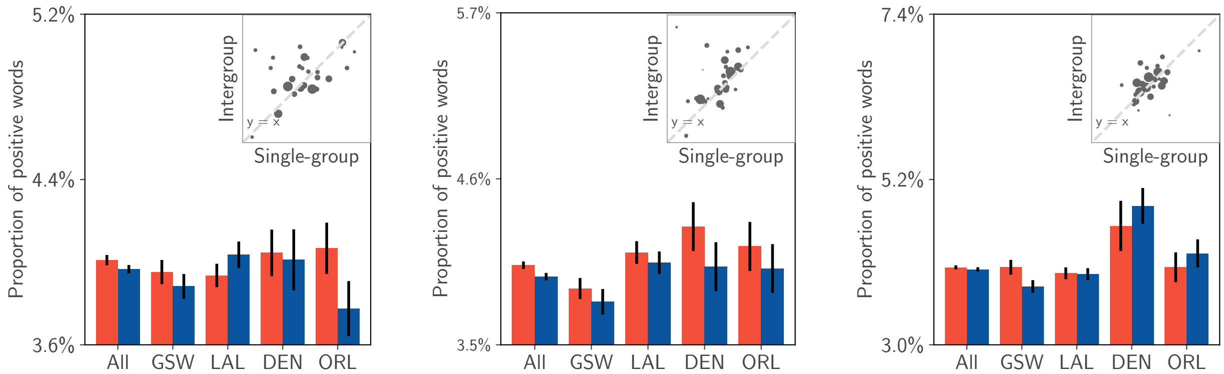
(a) Proportion of negative words. (b) Proportion of swear words. (c) Proportion of hate speech comments.

Figure 4.19: Comparison of language usage between intergroup and single-group members in the 2017 season. Intergroup members use more negative words (two-tailed t-test, $t = 4.04$, $p < 0.001$, 95% CI=0.04% to 0.13%; 24 out of 30 teams, two-tailed binomial test $p = 0.001$) and swear words (two-tailed t-test, $t = 4.17$, $p < 0.001$, 95% CI=0.03% to 0.10%; 29 out of 30 teams, two-tailed binomial test $p < 0.001$) and generate more hate speech comments (two-tailed t-test, $t = 11.01$, $p < 0.001$, 95% CI=1.21% to 1.74%; 24 out of 30 teams, two-tailed binomial test $p = 0.001$). Error bars represent standard errors.



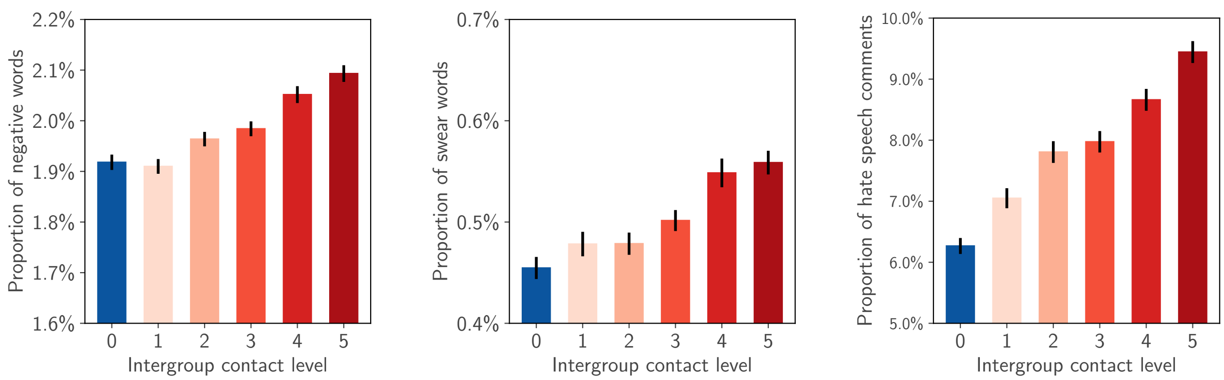
(a) Proportion of negative words. (b) Proportion of swear words. (c) Proportion of hate speech comments.

Figure 4.20: Comparison of language usage between intergroup and single-group members in the 2016 season. Intergroup members use more negative words (two-tailed t-test, $t = 3.93$, $p < 0.001$, 95% CI=0.05% to 0.14%; 23 out of 30 teams, two-tailed binomial test $p = 0.005$) and swear words (two-tailed t-test, $t = 3.10$, $p = 0.002$, 95% CI=0.02% to 0.09%; 28 out of 30 teams, two-tailed binomial test $p < 0.001$) and generate more hate speech comments (two-tailed t-test, $t = 7.95$, $p < 0.001$, 95% CI=0.92% to 1.53%; 25 out of 30 teams, two-tailed binomial test $p < 0.001$). Error bars represent standard errors.



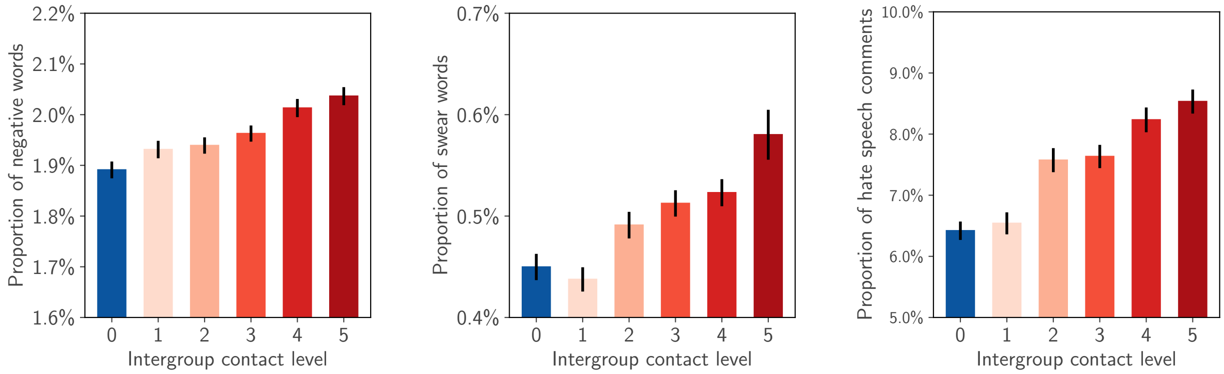
(a) The 2018 season. (b) The 2017 season. (c) The 2016 season.

Figure 4.21: The comparison of positive language usage between intergroup and single-group members in the 2018, 2017, and 2016 seasons. We find no consistent trend at the 5% significance level ($\alpha = 0.05$) (two-tailed t-test, $t = 1.36$, $p = 0.174$, 95% CI=0.019% to 0.107%, 16 out of 30 teams, two-tailed binomial test $p = 0.856$ for the 2018 season; two-tailed t-test, $t = 2.15$, $p = 0.03$, 95% CI=0.000% to 0.144%, 16 out of 30 teams, two-tailed binomial test $p = 0.856$ for the 2017 season; two-tailed t-test, $t = 0.66$, $p = 0.508$, 95% CI=-0.050% to 0.102%, 14 out of 30 teams, two-tailed binomial test $p = 0.856$ for the 2016 season). Error bars represent standard errors.



(a) Proportion of negative words. (b) Proportion of swear words. (c) Proportion of hate speech comments.

Figure 4.22: Intragroup language usage differences of members with different intergroup contact levels in the 2017 season. x-axis represents intergroup levels determined by the fraction of comments in /r/NBA. We observe a consistent monotonic pattern in the proportion of negative words (mean = 1.92%, 1.91%, 1.96%, 1.98%, 2.05%, and 2.09%, respectively for labels from 0 to 6), swear words (mean = 0.45%, 0.48%, 0.48%, 0.50%, 0.55%, and 0.56%, respectively for labels from 0 to 6), and hate speech comments (mean = 6.27%, 7.05%, 7.81%, 8.97%, 8.67%, and 9.44%, respectively for labels from 0 to 6).

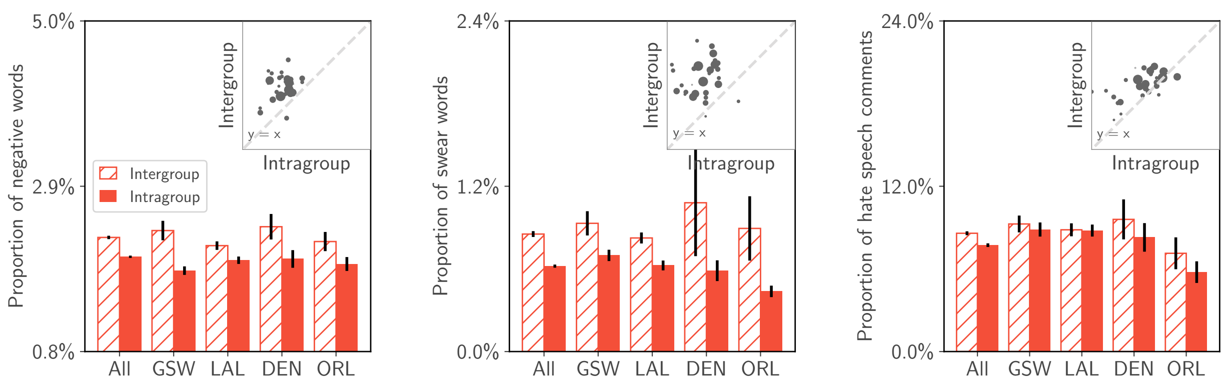


(a) Proportion of negative words.

(b) Proportion of swear words.

(c) Proportion of hate speech comments.

Figure 4.23: Intragroup language usage differences of members with different intergroup contact levels in the 2016 season. x-axis represents intergroup levels determined by the number of comments in /r/NBA. We observe a consistent monotonic pattern in the proportion of negative words (mean = 1.89%, 1.93%, 1.94%, 1.96%, 2.01%, and 2.04%, respectively for labels from 0 to 6), swear words (mean = 0.45%, 0.44%, 0.49%, 0.51%, 0.52%, and 0.58%, respectively for labels from 0 to 6), and hate speech comments (mean = 6.42%, 6.54%, 7.57%, 7.63%, 8.23%, and 8.53%, respectively for labels from 0 to 6).

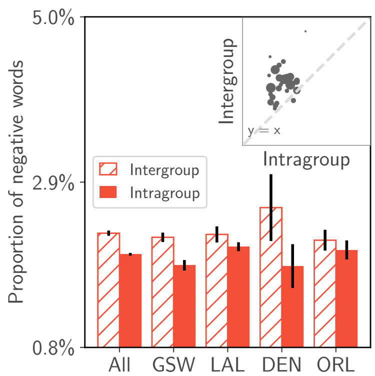


(a) Proportion of negative words.

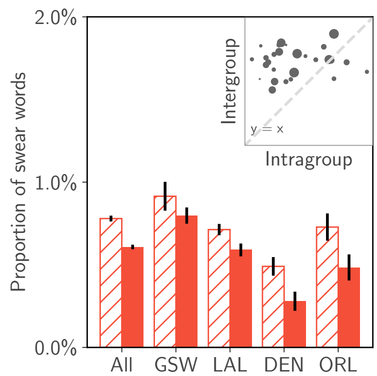
(b) Proportion of swear words.

(c) Proportion of hate speech comments.

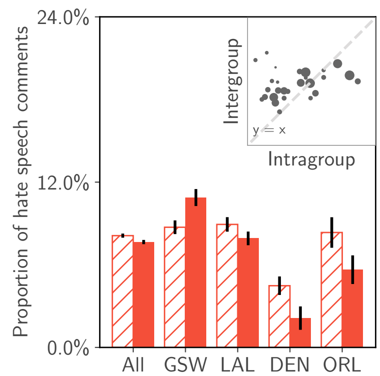
Figure 4.24: Intergroup members use more negative language in the intergroup setting than in the intragroup setting in the 2017 season. They use more negative words (two-tailed t-test, $t = 9, 36$, $p < 0.001$, 95% CI=0.19% to 0.30%; 29 out of 30 teams, two-tailed binomial test $p < 0.001$) and swear words (two-tailed t-test, $t = 13.80$, $p < 0.001$, 95% CI=0.29% to 0.38%; 28 out of 30 teams, two-tailed binomial test $p < 0.001$) and generate more hate speech comments (two-tailed t-test, 4.43, $p < 0.001$, 95% CI=0.48% to 1.25%; 26 out of 30 teams, two-tailed binomial test $p = 0.005$). Error bars represent standard errors.



(a) Proportion of negative words.



(b) Proportion of swear words.



(c) Proportion of hate speech comments.

Figure 4.25: Intergroup members use more negative language in the intergroup setting than in the intragroup setting in the 2016 season. They use more negative words (two-tailed t-test, $t = 7.21$, $p < 0.001$, 95% CI=0.20% to 0.34%; 28 out of 30 teams, two-tailed binomial test $p < 0.001$) and swear words (two-tailed t-test, $t = 12.22$, $p < 0.001$, 95% CI=0.23% to 0.32%; 27 out of 30 teams, two-tailed binomial test $p < 0.001$) and generate more hate speech comments (two-tailed t-test, $t = 2.17$, $p = 0.030$, 95% CI=0.04% to 0.89%; 22 out of 30 teams, two-tailed binomial test $p = 0.016$). Error bars represent standard errors.

Chapter 5

Understanding Group Event Scheduling via the OutWithFriendz Mobile Application

To address the limitations of existing services for group event organization, we have developed OutWithFriendz, a mobile application that enables groups of people to decide together through a voting process the date/time the group would like to meet, as well as the location where they would like to meet. OutWithFriendz is implemented as a client-server architecture that is comprised of both iOS and Android based clients that communicate with a server implemented as a Java Web application.

The main elements of our OutWithFriendz mobile application are shown in Figure 5.1. To start using it, a user may create a new invitation acting as a host. During this process, she can specify the details of this invitation, including a title, a list of suggested dates, a list of suggested locations and invited participants. After this host submits a new invitation to the server, all invited participants receive it and can view the preferences of their invitees. They can then suggest more dates, locations, or vote for their preferred options and comment on the invitation. After the voting process has ended, the host then decides on the final meeting location and time, which are then sent to all participants. In this example (Figure 5.1), the host suggested four locations and three date/time options. After the suggesting and voting process, she selected Cafe Mexicali and Friday, 03-17-2017, 18:00 as final decisions, which received the most votes. Please note that, in our design, the host can make decisions based on the voting results, but she does not have to always obey them. We did find few hosts in our field study whose final decisions were different from the options

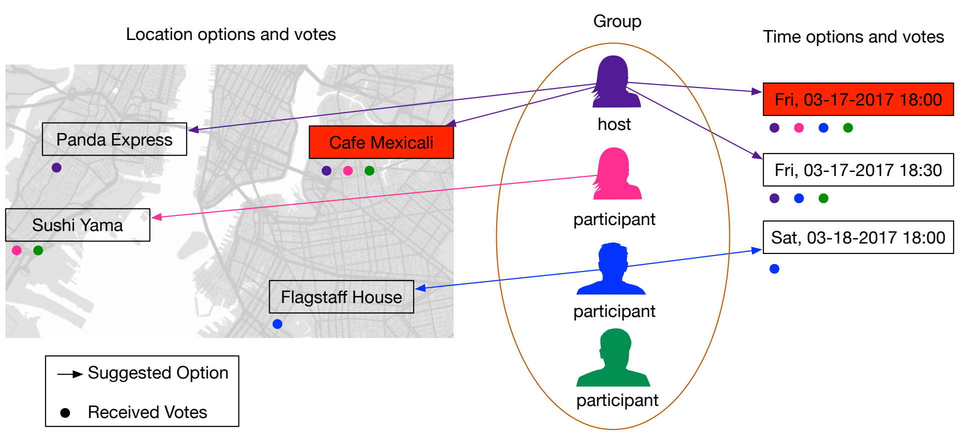


Figure 5.1: An illustration of the key elements in our OutWithFriendz system. The colored arrows (dots) indicate which user suggested (voted) for a meeting time or location, and the red boxes indicate the final decisions.

that received the most votes. This scenario will be discussed later.

Introducing our newly designed OutWithFriendz mobile application, which embeds group decision making into the voting process, raises new questions: How do the mobile app users collaborate to organize their group events? What are the major factors that will impact group decisions? How is the voting behavior processed? And how do users improve event attendance rate?

Our mobile application also collects user mobility-related data. The app posts GPS user location traces to the server. Users may opt out from providing their location traces, although most users did not disable location tracking for the entire duration of their participation in the study. This user mobility data provides a great opportunity to derive such input factors as spread, movement, mobility, and to investigate their impact on group event scheduling.

5.1 System Design

In this section, we describe the design, architecture, and implementation of the OutWithFriendz system. We also present a walk-through example to illustrate the user workflow of our app for group event scheduling.

5.1.1 System Architecture

In order to understand group user behavior at scale, we designed the OutWithFriendz mobile client for both the iPhone and Android platforms. The mobile client communicates with a remote server, which is implemented as a Java Web application using the Spring Roo application framework [170]. We decided to use Spring Roo because it allowed us to quickly prototype and iterate the design of the OutWithFriendz app. All required functionalities to the client are exposed through the server's REST APIs. We use MongoDB to store and manage all data on the server [125]. MongoDB is a widely used NoSQL solution that provides high performance, low complexity, and more importantly, native support of geospatial data and queries. To push notifications between server and clients, GCM services are used to handle all aspects of message queueing and delivery to client applications running on both mobile platforms. We also call upon the Google Maps API

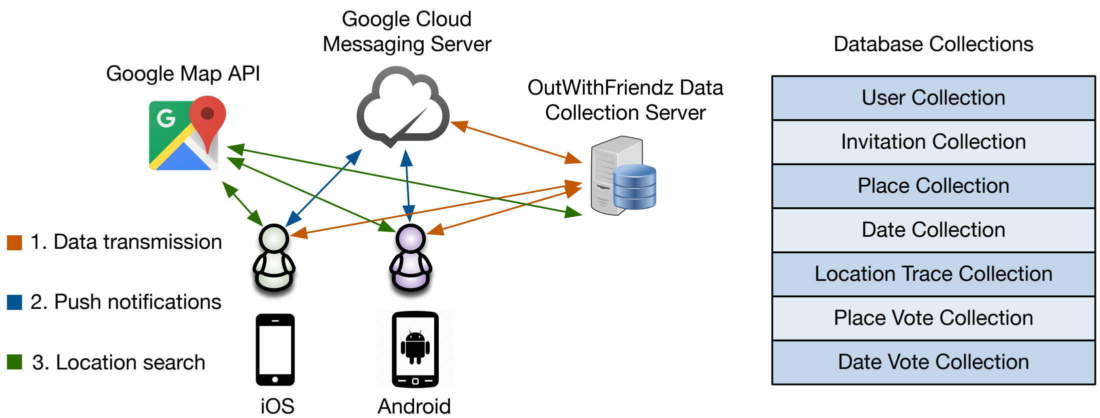
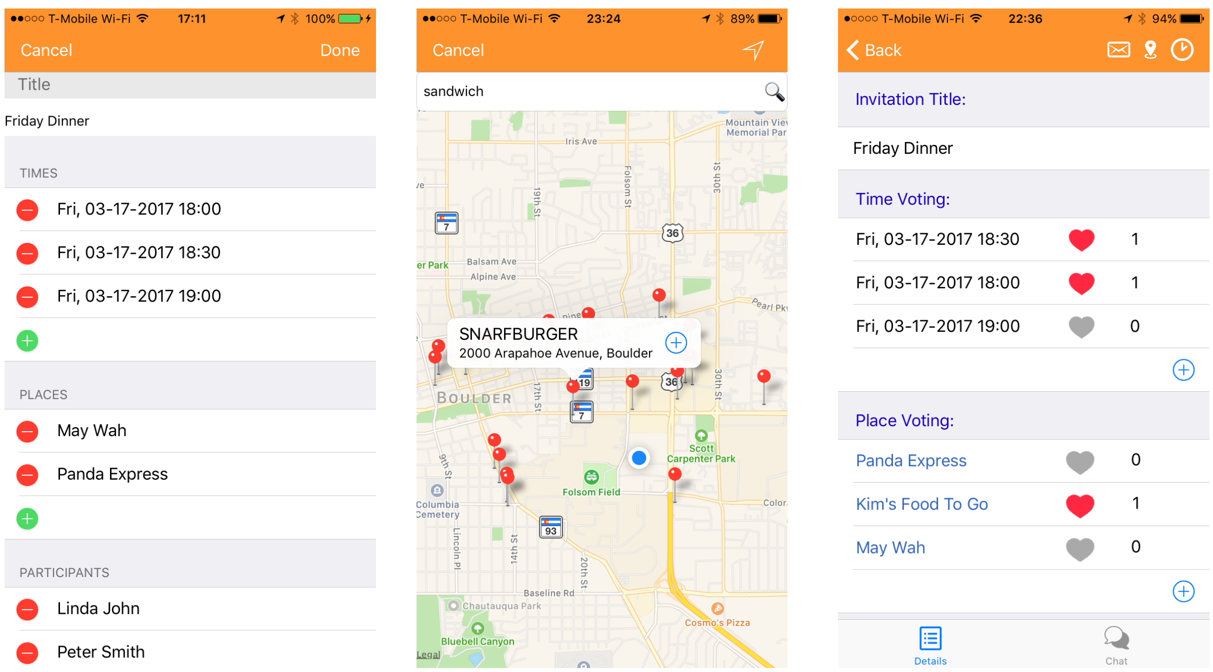


Figure 5.2: The architecture of the OutWithFriendz system.

to retrieve location search results. Figure 5.2 shows the overall architecture of our OutWithFriendz system.

5.1.2 UI Design Challenges

In order to enable a natural group decision-making workflow, we iteratively improved the UI and workflow of the OutWithFriends app based on feedback collected from user studies. We started with an initial usage survey before releasing the app to the market. During our survey, we hired seven students on campus who have different academic backgrounds. They formed three groups to use our app and provided useful feedbacks for improving UI design. For example, these users suggested (1) adding a chat board to allow group members to discuss their opinions; (2) allowing users to edit the location title and provide detailed information for each location; (3) allowing users to link suggested locations with the Google Places application; (4) pushing notifications if an invitation is created or modified; and (5) replacing text buttons with interactive icon buttons. Implementing this functionality helped us improve our app to better support the real-life group event scheduling process. We also added and altered application functionality to improve usability, based on the many user suggestions received during application usage. At the beginning of the study, we focused on dining events only. Later, we came to realize that the users would also like to



(a) Create invitation. (b) Add location. (c) Suggest and vote for preferences.

Figure 5.3: Main workflow of the OutWithFriendz mobile application.

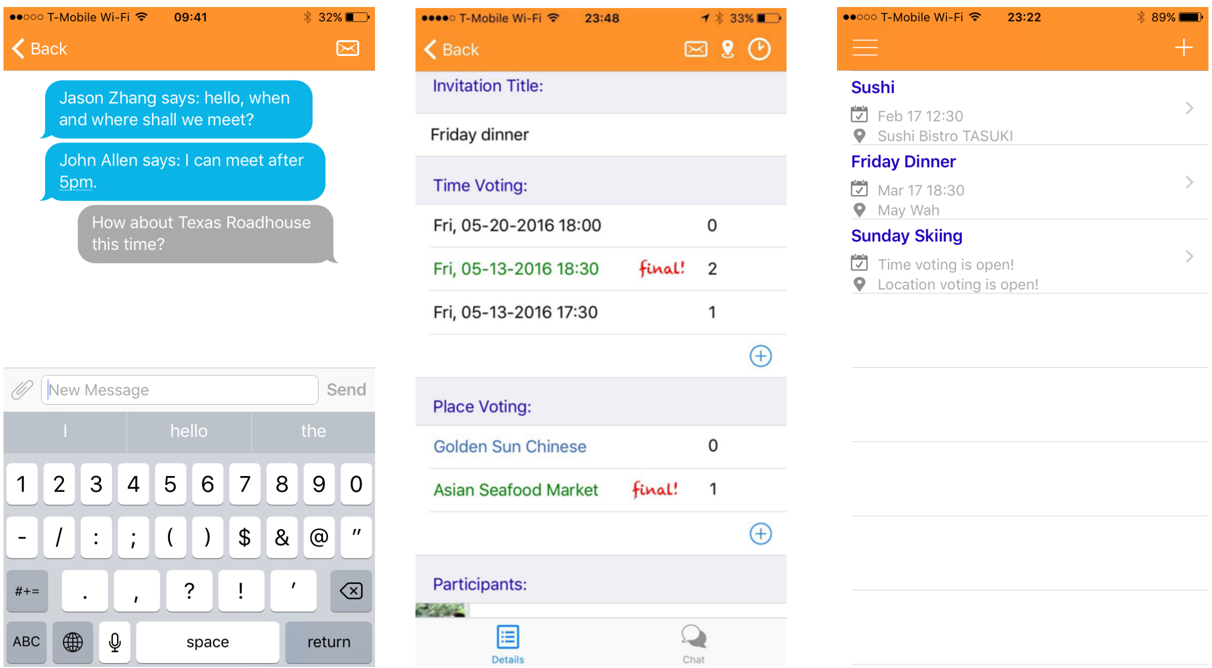
use the OutWithFriends app for generic group gatherings, such as going for a hike or watching a movie. To support this functionality, we shifted from integrating with the Foursquare API to the more suitable Google Places API. We also changed the workflow for the voting process to make it more flexible. Initially, invitation participants were required to decide on the meeting time before starting the voting process for the location. However, our users preferred to perform time voting and location voting concurrently, which is more flexible. Next, we describe the main workflow of our app.

5.1.3 A Walk-through Example

To better understand the workflow of our mobile application, we provide a walk-through example of how the main functions are used for group event scheduling.

A host invites two friends to meet for dinner. In this use case, we describe the actions a user would take to invite some friends to meet for dinner. Here we call this user the host. When creating a new event invitation, the host will go to the window shown in Figure 5.3a and perform the following steps: (1) create a title for the invitation, such as Friday Dinner; (2) specify one or more of the possible dates and times for the invitation; (3) suggest meeting locations using Google Map Services, as shown in Figure 5.3b; and (4) add one or more friends that want to be included as participants in the invitation. Finally, when the host is satisfied with the invitation settings, she taps the "send invitation" icon to send the invitation to all selected participants. She can also start voting for her own preferences right after the new invitation shows up on her screen.

A user receives an invitation to meet several friends for dinner. First, the user receives a notification from the OutWithFriendz application indicating that she has received a new invitation. The user can express her preferences by voting on one or more possible options for meeting dates and locations, as shown in Figure 5.3c. One important feature of our system is that the user may also add new proposed dates/time or locations to the invitation. Once the user has added a new option, it will be automatically made visible to all other participants. Users are also allowed to change their suggestions and votes throughout the voting process. In the "Chat" tab shown in



(a) Chat screen. (b) Invitation finalization screen. (c) Invitation list screen.

Figure 5.4: Main functions of the OutWithFriendz mobile application.

Figure 5.4a, the user is also able to send text messages to other group members for discussion and better coordination of the scheduling process.

Host finalizes the invitation based on voting results. The voting process continues until the host decides to finalize the meeting time and location. Only the host is permitted to finalize this decision, which is shown in Figure 5.4b. After the host has finalized the invitation, each participant receives a notification regarding this action. To support unforeseen changes, the host could still update the final decision after it is finalized. Each user's main screen will show a list of invitations that she has participated in, as shown in Figure 5.4c.

5.2 Data Collection, Methodology, and General Characteristics

We first describe the dataset we collected using our OutWithFriendz system and the methodology we used throughout the paper, then conduct data distribution analysis to understand the key characteristics of our dataset.

5.2.1 Data Collection

We deployed our OutWithFriendz mobile application on the Google Play and Apple Store marketplaces. To collect enough data for group dynamics analysis, we posted advertisements on Microworkers [120] and Craigslist [36] for participants. For each legitimate completed invitation, we paid the host of a group 20 dollars, with the provisions that (1) the host and participants must live in the US; (2) the host should invite at least two other friends to the invitation using our app; (3) the group must demonstrate a full voting process; (4) the host must finalize the meeting time and location for the invitation; (5) each participant would open their location services on their smartphone during the study and allow us to track their mobility traces; and (6) at least half of the group members attended the finalized event[1] .

[1] We added this requirement to prevent workers from creating fake invitations and making dishonest money. It would be interesting to analyze the low-attendance events, which we plan to investigate when we have more users and events.

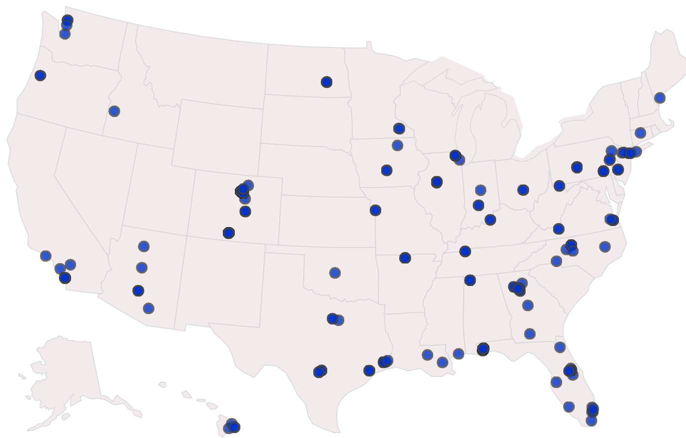


Figure 5.5: The geographic distribution of all finalized locations across the US.

5.2.2 Dataset and User Demographics

From these two job post websites, we collected 246 legitimate invitations over a 5-month period from 432 users. In addition, 71 students on our campus used the app without getting any payment, which contributed another 76 legitimate invitations. The whole data collection period spanned from January 2016 to May 2017. In total, 503 distinct users of our OutWithFriendz application were identified, generating 322 legitimate invitations. Figure 6.1a shows the distribution of all suggested locations recorded in our server across the US. It indicates that our users are widespread throughout 34 different states and 81 cities in the country.

To better understand the demographics of our users, we have conducted an anonymized survey using Google Forms. We contacted all users through their Facebook accounts to complete the survey. In total, 294 users responded and completed the survey, representing 58.4% of all users. The results are shown in Figure 5.6. We can see that 60.8% of the users who responded to the survey are female and 48.5% are self-employed. The ratio of self-employed users is high because many users from crowdsourcing websites are freelancers or housewives who have more free time to complete assignments and earn extra money. Students, including some from our campus and some from the Microworker website, account for 27.5% of our participants. In addition, most (83.0%) of

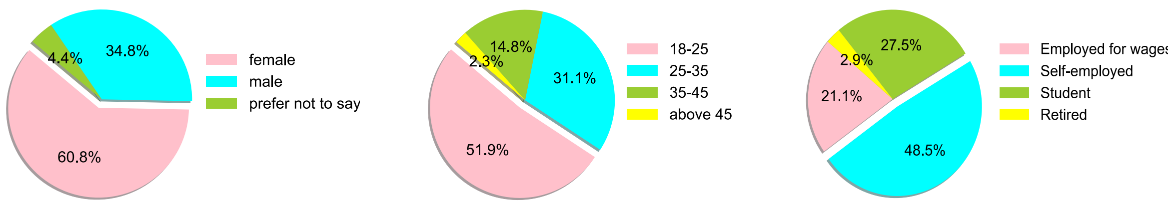


Figure 5.6: User demographic information of OutWithFriendz Users (294 participants). (L) Gender distribution. (M) Age distribution. (R) Profession distribution

our users are young, aged 18-35.

In addition to collecting data about the event organization process, we also collected user mobility-related data. The OutWithFriendz app posts GPS user location traces to the server either every 5 minutes if the app is running in the background or every 30 seconds if the app is running in the foreground. Before participation, all of our users were required to provide informed IRB consent. They would turn on the location services on their smartphone during the test, so we were able to collect the data. All location traces are anonymized and permission to use this anonymized data is provided when installing the app. The total amount of user traces data collected was about 1.1 GB.

5.2.3 Methodology

We define some concepts and notations that will be used throughout the paper. In the following analysis, we only use completed invitations in our dataset. A "group decision" refers to the information submitted by the host after an event has occurred, including the final group consensus rating.

In addition, for each event e, we define T_e as the set of suggested meeting times and L_e as the set of suggested locations. For each participant i and option o of event e, we let $V(i,o)$ be an indicator function that indicates whether i voted for option o:

$$V(i,o) = \begin{cases} 1 & \text{Participant } i \text{ voted for option } o \\ 0 & \text{Otherwise.} \end{cases} \tag{5.1}$$

Then we define a user's available time and location options as

$$\text{user } i\text{'s time availability for event } e = \frac{1}{|T_e|} \sum_{o=1}^{T_e} V(i,o), \tag{5.2}$$

$$\text{user } i\text{'s location availability for event } e = \frac{1}{|L_e|} \sum_{o=1}^{L_e} V(i,o). \tag{5.3}$$

Using an individual user's location trace data, we are able to analyze statistical properties of individual mobility. One way to consider a movement is to calculate the distance between two consecutive location trace points in our dataset. This will result in the detection of many very short movements, such as from one office to another in the same building. However, due to the location services limitation in today's mobile phones, these short movements cannot be traced precisely. For instance, when users are indoors, mobile phones may record false location changes where there was no movement, or miss actual location changes that are small. To eliminate these very short and uncertain movements and extract long movements, we implement an algorithm introduced by Ye et al [211], which was originally designed for GPS data. Assume that each individual's location trace points detected by mobiles devices are ordered by timestamp $L = \{l_1, l_2, l_3, ..., l_n\}$. We identify two types of movements. *Type 1* refers to the short movements of a user within a building. In *Type 2*, the user will travel from one area to another with a significant travel distance larger then r, for some period of time. In our experiments, r is set to 0.12 miles (200 meters) and the period threshold is set to 30 minutes, as suggested by [211]. To extract all the *Type 2* movements and eliminate *Type 1* movements, we iteratively seek spatial regions where the user remains for more than 30 minutes and all the tracked points within this spatial region lie within 0.12 miles. Then the location points in this spatial region are fused together by calculating the centroid of these points. The centroid point is considered as a **stationary point** for the spatial region.

5.2.4 Group Size Distribution

Figure 5.7 left summaries the distribution of group size in our OutWithFriendz dataset. Here we define group size as the number of participants who finally stayed in the invitation. We do not

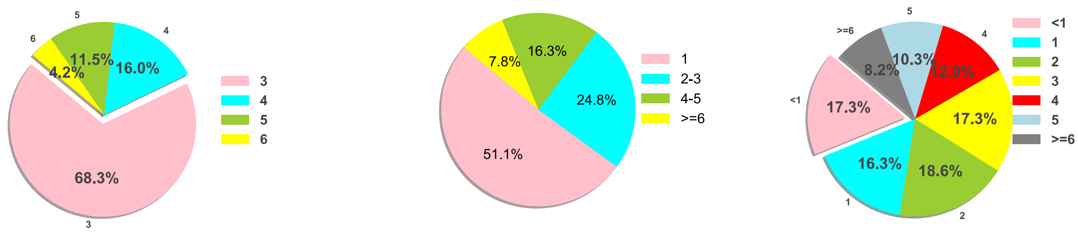


Figure 5.7: The distribution of group size (left), event count by host (middle) and number of days to make final decision (right).

count users who were removed from the invitation, either by themselves or by the host, because they did not participate in the whole scheduling process and their votes were not shown after the removal. We observe that most of the groups in our study are small, with the large majority being groups of three. We were pleased to see a significant fraction of groups with five (11.5%) and six members (4.2%) who were able to use the app concurrently. Our work focused primarily on obtaining data for groups of three or more, which we feel represent many typical social group interactions of interest to us. As a result, we did not focus on examining pairwise groups in this study. The figure's trend lines suggest that if we had opened up our study to pairwise groups, then our data would have been overwhelmingly skewed toward pairwise groups. However, now that we have obtained substantial initial data for larger groups, we plan to also explore the behavior of pairwise groups in our future works.

5.2.5 Distribution of Event Count by Host

Figure 5.7 (middle) shows the distribution of event number created by hosts. Here we only count completed events. During our study, 141 unique users have hosted at least one event, and 72 (51.1%) hosts have created exactly one event each. Please note that each user is allowed to create and join multiple invitations in this study. Moreover, 11 groups (7 from paid users and 4 from students in our university) have used our mobile app frequently, with more than six legitimate invitations per group in our dataset.

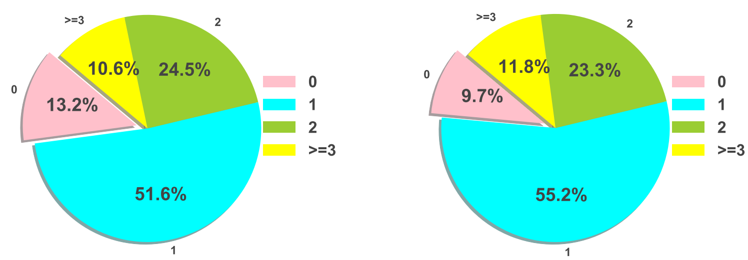


Figure 5.8: The distribution of number of votes by a single user for event time (left) and place (right).

5.2.6 Distribution of Days to Make Final Decision

We are interested in the duration that it took for event organizers to make their final decisions. As shown by the right figure in Figure 5.7, the number of days to make the final decision is somewhat evenly distributed, and there is no dominant duration in this distribution. This is a bit surprising or counter-intuitive, since we expected that there may be a more pronounced duration of decision-making within the first couple of days of creating an invitation. However, there are also a substantial fraction of events that took four or more days to decide (about 30%), indicating that a large fraction of hosts are taking a long time to decide. This may be affected by the type of events and the amount of lead time. For daily meals, users can make a decision within thirty minutes while for some weekend activities, they will start planning it at the beginning of the week.

5.2.7 Voting Distribution of Individual Users

Voting distribution is based on the number of votes made per individual user. The distributions for time and place voting are shown in Figure 5.8. The majority of users will vote for one option as far as the event time. Similarly, the majority of users will vote for one option in terms of the place voting. In both processes, around 10% didn't vote and 10% voted for more than 2 options. This voting behavior is analyzed further in later sections.

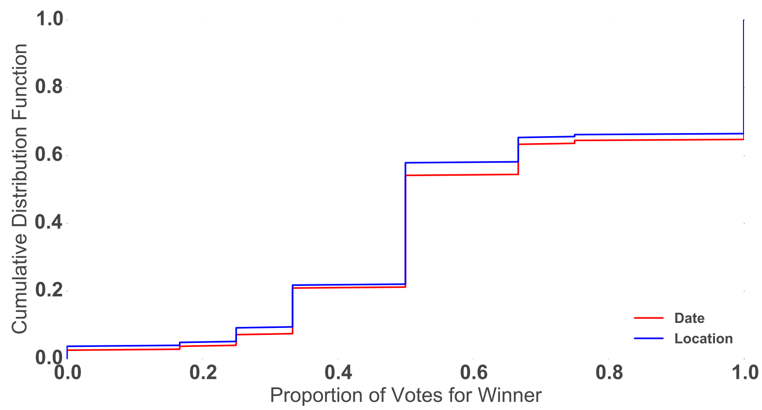


Figure 5.9: The proportion of users who voted for final event time (red) and location (blue).

5.2.8 Distribution of the Proportion of the Votes

We also analyze the proportion of users who voted for event time or location in the final decision. The distribution is shown in Figure 5.9. More than 70% of the final decisions for both time and location received majority votes to become the final choice. This is understandable since groups tend to agree on the majority votes. For the remaining 30%, we observe some very interesting behavior. In these polls, the final decisions did not receive the majority of the votes. In fact, in a small fraction of cases, there is a non-zero proportion of polls in which the final decision received no votes. In these cases, the group host, who is the only one with the power to finalize the event time and location, decided to override the majority voting results, either by personal fiat or possibly through a discussion with other group members that caused them to change their minds.

5.2.9 Suggestion Distribution

OutWithFriendz app allows group participants not only to vote for their preferences, but also to suggest new options. Figure 5.10 shows the suggestion distributions for host and participants. Most hosts will suggest 2 or more options for the event. We also observe a small portion of hosts who provide no options of their own, and rely on other group members to provide suggestions. For participants, more than 60% didn't make new suggestions. They just vote for the existing options.

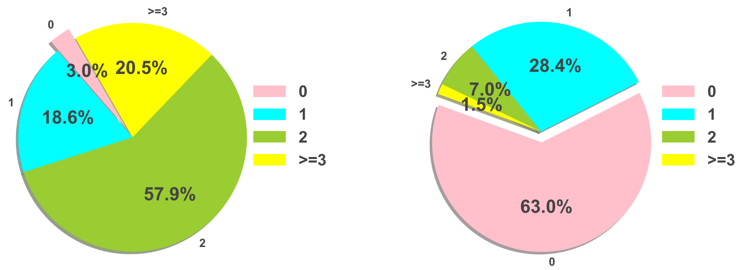


Figure 5.10: The distribution of the number of suggestions made by group host (left) and other participants (right).

Some made one new suggestion while very few of them would make too many new suggestions. We will further compare the influence of group host and participants in our group decision section.

5.2.10 Metro vs Non-metro Areas

Using the location trace data we collected from our users and the U.S. census data, we are able to identify locations frequently visited by our users. The technique we used will be introduced in detail in Section 5. Then we can project each user's home county using the frequently visited locations. According to the latest Rural-Urban Continuum Codes released in May 2013 [1], every county is classified as a non-metro or metro area. In our dataset, 48% the users live in metro areas and 52% live in non-metro areas[2] .

5.2.11 Weather Factor

An important external factor that can influence event organization relates to weather. Here we examine the impact of rain and temperature on our dataset. For this analysis, weather and temperature information for each event was scraped from weathersource API [200] at its location and starting time in our dataset. Note that we can only get hourly weather data from weathersource. If the starting time of one event is 19:35, in the analysis we use 19:00 weather data at the same day crawled from weathersource. Here we decide it is raining if the precipitation of an event's starting

[2] Please note our dataset contains 71 students who lived in Boulder doing this study. If we remove this student population, the proportion of metro and non-metro users would be 39% and 61%, all from crowdsourcing market users.

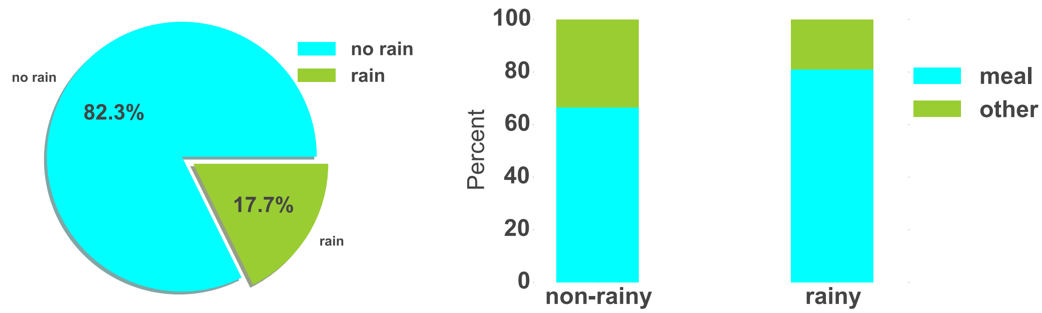


Figure 5.11: (L) The distribution of events on rainy days vs. non-rainy days; (R) The distribution of meal events and other events on non-rainy days and rainy days.

time is above 0. On snowy days, usually the precipitation will also be above 0, which we classify as rainy in our analysis.

Figure 6.5 shows the distribution of events that happened in rainy weather or not. 82.3% of events organized in our app occur in non-rainy weather. There are two reasonable explanations: (1) In many places around the country non-rainy days happen more often than rainy days; (2) Bad weather would have negative influence on real event attendance. Looking deeper, bad weather appears to affect the types of events that are organized. In OutWithFriendz, any place can be added to the Google Map as an option for voting, and need not be confined to a restaurant only. For example, people have used the app to organize events such as outdoor hiking and going to the movies. We divide all events into two category types: meal events and other events. Meal events refer to people hanging out for lunch or dinner, which is the majority event type in our dataset. Other events include activities that are not primarily dining, e.g. sporting and entertainment events. In our study, we found that bad weather would have less impact on meal events compare with other types of events. Figure 6.5 show that 66.7% of the events belong to meal events on non-rainy weather while this number goes up to 81.1% on rainy days.

Table 5.1: The correlation of user mobility and voting availability

	Pearson correlation	p-value
The correlation of user mobility and date voting availability.	0.276	7.12e-05
The correlation of user mobility and location voting availability.	0.281	2.92e-06

5.3 Group Decision Analysis

The analysis in this section examines the impact of a number of factors on group decision. We divide them into four categories: (1) impact of user mobility, (2) impact of individual preference, (3) impact of host preference, and (4) impact of voting behavior.

5.3.1 Impact of User Mobility

We now examine the impact of user mobility on group behavior in OutWithFriendz. Here we define user mobility as the total travel distance traveled by a user in the 48-hour period preceding an invitation. Our assumption before was that users who traveled longer distances will be more exhausted, and thus less likely to have significant voting availability. However, our analysis refutes this conjecture:

Observation 1 Users with higher mobility are more active in attending social events.

We use the Pearson correlation coefficient [100] to calculate the relationship between user mobility and voting availability. Table 5.1 shows that the correlation of user mobility with both date and location voting availability is positive, and the results are significant ($p < 0.001$). These results indicate that highly mobile users are more available for event attendance.

Observation 2 Group mobility has a positive correlation with an area's urban density.

Given the spatial regions that are detected, we are interested in investigating whether there exists any pattern between a group's mobility and an area's urban density. Our hypothesis is that groups living in metro areas have higher mobility than groups living in non-metro areas, since metro group members may be more spread out in big cities and generate longer travel distances. To perform

Table 5.2: The correlation of group mobility and area's urban density.

	Pearson Correlation	p-value
Population density	0.1834	0.013
Housing unites	0.1572	0.018

this analysis, we downloaded the 2016 U.S. area development degree data from the U.S. Census Bureau [20]. Here we use population density and number of housing units to calculate the urban density of an area. For simplicity, we only consider the location of each group event. It is possible that group members live in a city but traveled to a rural area for the event. But this is rare in our dataset. Table 5.2 shows the relationship between group's total travel distance and the corresponding county's population density and housing units. The Pearson correlation coefficient for these two parameters are positive with p-values that are smaller than 0.05.

5.3.2 Impact of Individual Preference

To discover underlying factors that may lead users to vote for specific event options, we first focus our analysis on individual users. A social event is typically characterized by two major factors: event time and location. Using the OutWithFriendz dataset we have collected, we first analyze the travel distance between event suggested locations and each participant's closest location cluster, with the requirement that this cluster must contain a point with a timestamp that occurs within 2 hours before or after the finalized time for the invitation. The suggested location options are further divided into two categories: the location options with votes and location options without votes. Based on the results, we make the observation:

Observation 3 Most users would like to vote for event locations near their frequented locations.

Figure 5.12 shows the cumulative distribution of travel distances among locations voted for and not voted for by each invitation participant. The average travel distance for voted locations is 4.19 miles while for non-voted locations is 7.53 miles. A Wilcoxon test found this to be a significant difference ($z = -4.57, p < 0.001$), which indicates users have clear preference to attend events near

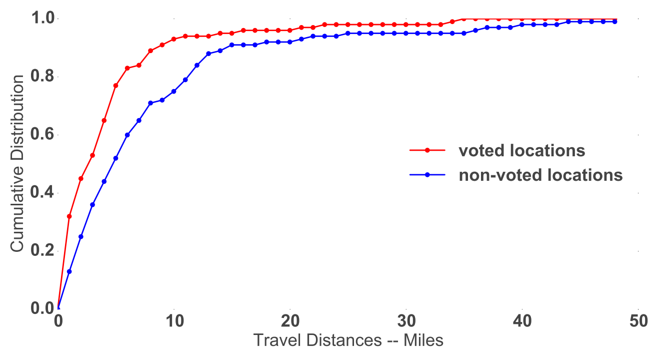


Figure 5.12: The cumulative distribution of travel distances among voted locations and non-voted locations for each participant.

their frequented places. This is reasonable in daily life. For example, we would intuitively expect that users would prefer to go to dinner at restaurants that are close to their office or home.

Observation 4 People like to attend social events after work on weekdays, while on weekends, events are distributed relatively evenly.

Additionally, we are also interested in investigating individual user's temporal preference. Our hypothesis is that participants are more likely to attend events after work. Figure 5.13 depicts the suggested event times on weekdays and weekends. It is clear that in weekdays there is a high spike around 6pm. While in comparison, event times are distributed more evenly throughout the day on weekends.

5.3.3 Impact of Host Preference

In our OutWithFriendz system, the host has more authority than other participants. The host can not only decide who to invite, but also finalizes the event time and location. This suggests that the host will have more influence on the group decision-making process. In our dataset, we have several significant observations about host behavior.

Observation 5 The final meeting location is closer to a host's frequented place than other participants.

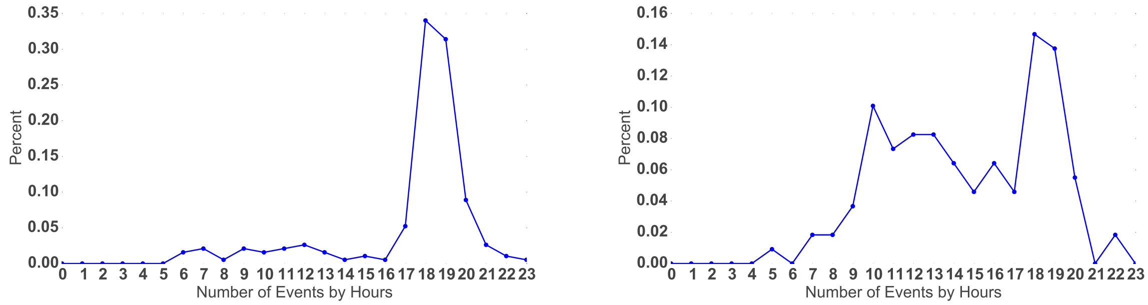


Figure 5.13: The distribution of events by hours on weekday (left) and weekend (right).

It's not surprising that event host would show some "selfishness" when making the final decision. We calculated that the average distance between the final location and host's closest frequented place is 5.23 miles. While the same metric for common participants is 6.75 miles, 29% longer than host, a significant difference according to a Wilcoxon test ($z = -3.38, p < 0.001$).

Observation 6 The probability that the final event date and location is voted by the host is significantly higher than that for other group participants. For events in which the host did not choose his/her voting option as the final decision, the main reason is to respect the majority voting results.

Table 5.3 shows the probability that final event option is voted for by the host and by another participant. It is clear that the final option is much more likely to have been voted for by the host than by other group members, with a probability of 0.71 vs 0.36 for the final event date ($z = -13.22, p < 0.001$). and 0.72 vs 0.34 ($z = -11.87, p < 0.001$) for the final event location. We also observe that among all the invitations in which the final event time was not the host's voting option, 95.2% coincided with the majority voting results. The percentage is 94.4% for the final

Table 5.3: The probability of final event option voted by host and participant

	Probability
Final event date voted by host	0.71
Final event date voted by participant	0.36
Final event location voted by host	0.72
Final event location voted by participant	0.34

Table 5.4: The correlation between whether host comply voting results and event attendance rate

	Pearson Correlation	p-value
Whether host comply location voting results and event attendance rate	0.48	$< 10^{-10}$
Whether host comply date voting results and event attendance rate	0.47	$< 10^{-10}$

event location. This indicates that, although hosts have a higher impact on making decisions, they still highly respect other group members' opinions.

Observation 7 The host choosing not to use the consensus voting result as the final decision would have negative influence on the event attendance rate.

In our OutWithFriendz application, the host can select a final decision that is contrary to the voting results. According to our user study, there are two main reasons for this behavior: (1) The option that received most votes is not suitable for the event host; (2) the users discussed through using the app's chat function and some members changed their minds but did not update their votes. In our OutWithFriendz dataset, 7.3% of final dates and 9.2% of final locations are contrary to voting results. We calculated the Pearson correlation between whether the host complies with the consensus opinion and the corresponding event attendance rate. The results are shown in Table 5.4. The positive correlation is significant here for both location voting and date voting. These results confirm that for event organization, hosts that don't comply with voting results have negative impact on the attraction of participants.

5.3.4 Impact of Voting Process

Voting is one of the most innovative aspects of our OutWithFriendz system. In contrast to traditional online event organization services, such as Meetup and Douban Events, where the meeting location and time is decided only by group host when the invitation is created, OutWithFriendz allows all group members to express their preferences through suggestions and votes. After all invitees have responded to the poll, the group host is able to find a mutually agreeable location

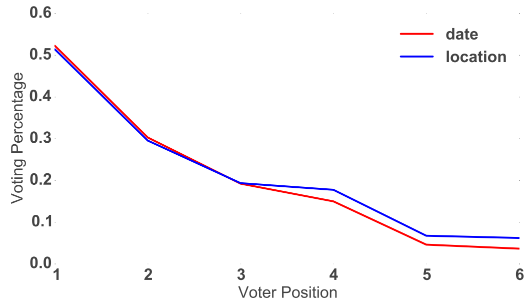


Figure 5.14: The relationship between average availability and voter position.

and time that usually accommodates most of the group members. Tracking the group's voting process using our system offers a great opportunity to study group decision making behavior.

Observation 8 Early voters tend to vote for a wide variety of options, while later voters are more likely to report limited availability.

Figure 5.14 shows the relationship between group members' average availability and their "voting position". A user's availability is defined by Equations 5.2 and 5.3. Voter position refers to the temporal index of casting votes within the scope of an invitation. The host's position is 1, the first voter's position is 2, and so on. There is a clear decrease of availability as voter position increases, and this result is consistent for both location voting and time voting.

Observation 9 Late voters tend to vote for options that align with existing voting results and are mutually agreeable.

Due to the fact that new voters can observe other voters' responses, these early responses would easily affect future voting behaviors. In our dataset, we find that later voters are more willing to vote for options that coincide with existing voting results, which makes it easier for the host to find common mutually agreeable options. Here we define the voting coincidence by cosine similarity:

$$Coincidence = Cosine(\vec{v}, \vec{e}) \tag{5.4}$$

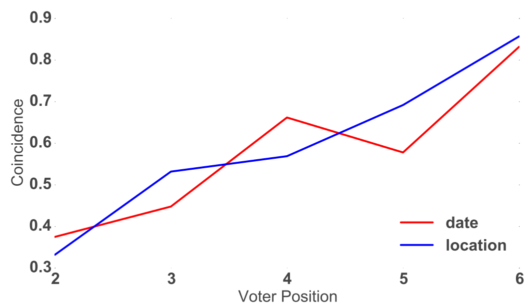


Figure 5.15: The relationship between average voting coincidence and voter position.

where $\vec{v}$ refers to the new voter's voting vector, and $\vec{e}$ refers to the existing voting vector. For example, if there are four date options in invitation i, and they receive 1, 3, 1, and 0 votes respectively, then the $\vec{e}$ is $[1, 3, 1, 0]$. If a new voter v votes for second, and fourth option, then $\vec{v}$ is [0, 1, 0, 1]. The coincidence here is the cosine similarity between $\vec{v}$ and $\vec{e}$, which is 0.640. Figure 5.15 shows the relationship between average voting coincidence and voter position. It is clear that there is a positive relationship between voter position and coincidence in both date voting and location voting. Later voters will try to consider their options in light of the whole group's voting behavior. Sometimes, these later voters may vote for less convenient options in order to make the host's life easier.

5.4 Discussion

In this section, we summarize the key results of our analysis and provide insights into how these results can be utilized for better group event planning experiences. The results cover the impact of user mobility, individual preference, and the host preference on the group event scheduling process. We also discuss the impact of user behavior during the voting process.

User Mobility. The analysis in Section 5.3.1 showed that users with higher mobility are more likely to be active participants in group events. They are more active in voting for proposed event location and event time. There are two reasonable explanations for this phenomenon:

- Previous studies have shown that users who travel by car, bus, and foot in daily life differ substantially in their value of time, in both revealed-preference and stated-preference surveys [107, 50]. In our OutWithFriendz dataset, the users who travel long distances may travel by car. This increases their likelihood of attending events far away from their frequented spots.

- Users who have higher mobility are more likely to be active event attendees. They are used to meeting with friends after school or work, which results in longer travel distances. Conversely, office workers who sit at their desks during the day have little mobility detected, but may still be tired after work and less likely to travel.

One way to use this observation for better event planning would be to recommend more diverse locations and dates for highly mobile users, as they tend to be more willing to explore new options. On the other hand, users with less mobility should not be overwhelmed with a large number of choices. In addition to recommending event locations and time, this observation can be used for forming groups by matching users with similar mobility levels in the same group, which can lead to smoother event planning experiences.

Individual Preference. The analysis in Section 5.3.2 revealed typical patterns related to individual preferences for event time and locations. First, with regard to event location, users tend to arrange events at nearby locations to avoid traveling long distances. Second, with regard to event time, on weekdays, users want to schedule events after their working hours, while on weekends users show more flexibility. These observations are worth considering for smarter group event planning. For event locations, the application could suggest places such that the mean travel distance for the group members is minimized, so as to provide a reasonable compromise for the whole group. Likewise, suggested event time should occur outside of typical working hours of the group members.

Host Preference. The results in Section 5.3.3 show that group event planning is heavily influenced by the host who creates the invitation. From the analysis in this section, it is evident that the final event location is on average closer to the host's frequented locations than that of other group members. We also see that the final event locations and dates are more likely to be the options voted by the host. However, the influence of the host can also lead to negative outcomes: when the host chooses not to follow the group's consensus, event attendance is reduced. These effects point to the need to carefully consider the preferences of the host, and how these preferences align with the preferences of the group, when providing recommendations to event participants. Effective communication mechanisms between the host and the participants should also be provided.

Voting Analysis. The analysis in Section 5.3.4 shows that the votes cast by early voters are very likely to affect late voters. Late voters tend to vote for fewer options, and these options tend to match those that have already received votes from early voters. There are several possible explanations for this observation:

- People who came to the poll later may be busier than early voters, and had a smaller time window before the actual event time, thus their availability is more limited compared with early voters.
- The polls in the OutWithFriendz application are all open polls, which means later voters can see the current voting results. Their votes may not be able to change the current status significantly because every voter can only vote once for a given option.
- Late voters will vote only for agreeable options that help the host to more easily finalize decisions.

This phenomenon can be used to improve the event planning experience. For example, we could encourage users to vote early, so that their votes will carry more weight. We could also hide existing voting results, so as to prevent existing votes from biasing later voters, and facilitate the voting process by providing voting recommendations to users based on their historical voting patterns.

Chapter 6

GEVR: An Event Venue Recommendation System for Groups of Mobile Users

The implementation of the OutWithFriendz mobile system provides a great opportunity for tackling the problem of recommending event venues for groups of mobile users. Online venue recommendation services, such as Foursquare, Google Maps, and Yelp, demonstrate the importance placed on recommending venues to an individual mobile user, e.g., what coffee shop or restaurant should that person visit next. However, such venue recommendation does not address the larger challenge of recommending venues to a group of mobile users who are trying to choose a place to meet, e.g., friends and colleagues trying to decide on a restaurant for lunch or dinner. In this case, the members of the group may have a wide range of behaviors and preferences and be scattered across many locations.

We design a service that we expect will become increasingly important as people get more and more digitally connected and organize their social gatherings via mobile technologies. While effective recommendation to individual users is already a challenging task, developing event venue recommendations for groups of mobile users is significantly more challenging due to a number of reasons. Groups of users often have a wide variety of behaviors and preferences that must be resolved to achieve the best recommendations for the group. In addition, different members of a group may be at varied locations, so mobility considerations should be taken into account when recommending a place to meet. Groups are also dynamic rather than static, and so group membership may change fluidly from event to event, unlike individual recommendation for the same target user.

Prior studies on group-based recommendation have addressed the issue of resolving differing preferences of individual members of the group in order to provide the best recommendation [58, 26, 148]. However, the role that location and mobility of different group members may play in influencing the group's final choice has not been addressed in these articles. Some scant prior research in group event venue recommendation [14, 139] has been confined only to in-lab surveys and has not studied the real-world dynamics of groups of mobile users, nor its impacts on group decision-making. We believe this is an exciting area ripe for exploration by the ubiquitous computing research community. Due to the explosion of people scheduling group events online using smart devices, we expect mobile and context-aware computing to be used extensively to reduce the friction inherent in coordination and assist groups in making decisions.

As the first step towards in direction, we present GEVR, a novel Group Event Venue Recommendation system that tackles the recommendation challenge by incorporating mobility patterns and social relations among group members. To tackle the complexity of group recommendation, GEVR splits the recommendation process into three separate steps:

(1) **Group location cluster detection:** Using individual group members' historical location traces, GEVR first identifies location clusters where the group members have colocated in the past.

(2) **Group event location cluster prediction:** After identifying these group location clusters, GEVR uses a newly designed, social-based group location cluster prediction model to estimate which location cluster the group is more likely to meet at.

(3) **Event venue recommendation:** Finally, GEVR aggregates restaurants near group members' frequented locations using the Foursquare Venue Recommendation API [53], and reranks them based on the prediction results calculated by Step (2) to build a final top-K recommendation list.

Evaluation results show that our prediction model achieves over 80% accuracy for predicting a final meeting location cluster. Furthermore, through the combination of the group location clusters

Table 6.1: Dataset Statistics. The users in our dataset are from two sources. Campus users are solicited from our university and volunteered to use the mobile app. Crowdsourcing users are recruited from Microworkers and Craigslist. Our users are all over the US, across 40 states and 117 cities.

	Total	Campus	Crowdsourcing
Number of users	625	75	550
Number of events	502	151	351

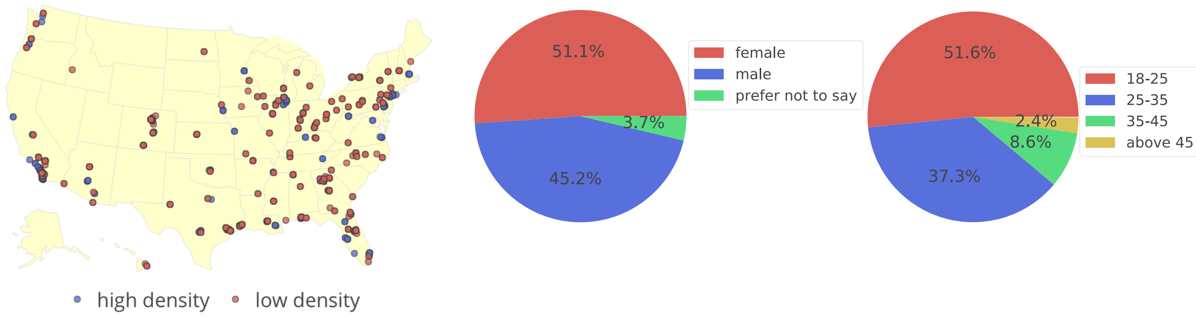
identified, cluster prediction results and Foursquare Venue Recommendation API, GEVR achieves promising recommendation performance and outperforms the baseline models by a significant margin.

6.1 Dataset

The dataset we use in this work is collected using the OutWithFriendz mobile application, which is introduced in Chapter 5. The whole dataset contains group event data collected from January 2017 to December 2017. The experimental protocol is reviewed and conducted under IRB Protocol #12-0008.

Since our GEVR system utilizes users' location traces to provide recommendation for groups of users, to ensure sufficient data coverage, we remove group events that have members with less than three days of location trace data before the invitation was created. Also, we remove group events in which the final meeting venue is not a restaurant. [1] The basic statistics of our dataset after the filtering are shown in Table 6.1. The geographic distribution of all suggested venues across the US is projected in Figure 6.1a. All the venues are further divided into low density and high density areas based on the population density of the venue location. Here, the population density information is collected from the 2016 US area development degree data from the US Census Bureau [20]. Our users are widespread over the US, covering 40 states and 117 cities. After the user study, we conduct an anonymized user demographic questionnaire and 341 study participants

[1] Groups can organize any type of events using OutWithFriendz, including going to the gym, watching a movie at the theater, or participating in outdoor activities. In this project, we focus on dining events, as it is the dominating type in the dataset: 81% of the events are dining events.



(a) All the suggested venues across the US.

(b) User demographics.

Figure 6.1: Figure 6.1a shows the geographic distribution of all the suggested venues across the US. Blue marks represent places of high-density areas and red ones represent low-density areas. Figure 6.1b shows the demographics (gender and age) of 341 users in the dataset who fill out the anonymized user demographic questionnaire.

completed it. Figure 6.1b shows these users' demographic information. We also observe that all of the groups in our study are either small or medium sized, with group sizes of 3–6. Among them, 64% of the groups have three members, 19% of the groups have four members, 12% of the groups have five members, and the remaining 5% have six members.

6.2 Framework Overview

Given all the information collected from the mobile app, to build an effective event venue recommendation system for groups, we tackle the challenges in three steps:

Step 1: Group location cluster detection. In this step, our objective is to gather location trace points of group members into location clusters. The detected location clusters should be able to represent group members' frequented locations. The details of group location cluster detection will be introduced in Section 6.3. The venues are searched and selected using the integrated Google Maps API.

Step 2: Group location cluster prediction. Using the group location clusters detected from Step 1, our goal of this step is to predict which location clusters the group is likely to meet at. More specifically, we model the probability of each detected location cluster being selected as the final meeting location cluster of the group. The design of our prediction model will be explained in Section 6.4.

Step 3: Event venue recommendation for groups of mobile users. In this step, our goal is to recommend a list of restaurants for each group event. We use the Foursquare Venue Recommendation API to create a candidate restaurant pool near every detected location cluster generated by Step 1, and re-rank the candidate pool using the prediction results calculated by Step 2 to build a final top-N restaurant recommendation list. The construction of our group event venue recommendation framework will be described in Section 6.5.

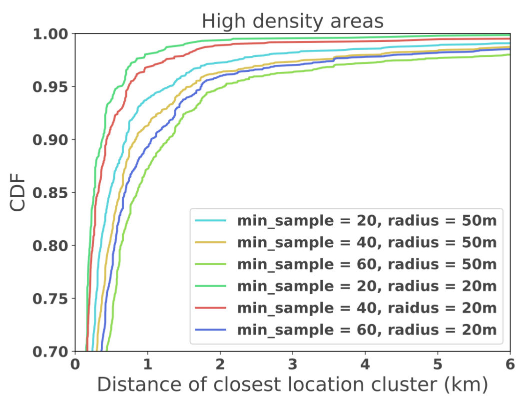
To make an effective event venue recommendation for groups of mobile users, the ultimate goal is to reduce and prioritize the venue recommendation list. The three steps gradually build towards this goal. We first use Step 1 to detect group members' frequented geographic areas and concentrate our recommendation on these areas instead of the whole city. After that, by modeling group members' location preferences and aggregating them using the group decision strategy in Step 2, we estimate the likelihood of each location cluster that the group will decide to meet at. Finally, in Step 3, we re-rank the candidate venues in concentrated areas by the estimated likelihood of each location cluster.

6.3 Group Location Cluster Analysis

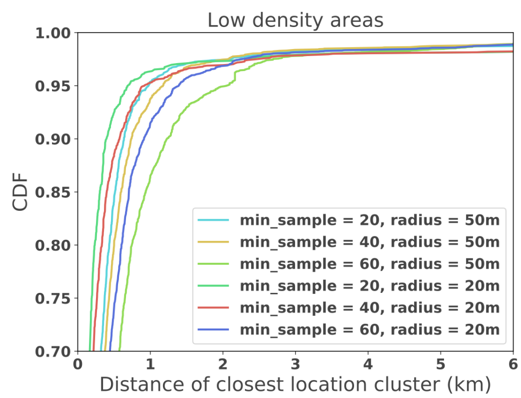
Given a group of multiple mobile users and their location traces, our first goal is to identify **group location clusters**, i.e., concentrated regions that group members have visited. Intuitively, a location cluster may represent home, work, or a district, say downtown, near a mall, or a zoned business area where there are many restaurants. Due to the fact that user location data recorded by mobile applications usually contains errors of a few feet, it is impractical to use GPS points in location traces directly for modeling. A common approach in user mobility studies is to cluster nearby geolocation points into location clusters [5, 81, 27] and use them instead of location points when predicting human movements. Inspired by these papers, to generate group location clusters, we first combine all group members' temporally ordered location trace points and apply the widely-used DBSCAN [51] location clustering algorithm to detect group location clusters, which represent significant places that the group members visit frequently. More specifically, every location point is represented by its geographical coordinate pair (latitude and longitude), and fed directly into the DBSCAN clustering algorithm, which identifies dense neighborhoods and gathers location points that are very close to each other into the same cluster to represent a location. After the clustering procedure, the ideal situation is that every location point detected from group members' location traces is assigned to a specific location cluster; nearby location points would belong to the same location cluster, and each cluster is relatively separated from each other on the map. We further represent each group location cluster using the centroid of all the location trace points that belong to that cluster. To ensure that the DBSCAN algorithm is able to find sensible and meaningful location clusters using our dataset, and the group's final meeting venue falls within one of these clusters, it is important to analyze and select appropriate values for the parameters.

The main parameters of DBSCAN are *min_sample* and *radius*. *Min_sample* is the minimum number of points required to form a dense region, and *radius* is the maximum distance allowed between two samples for them to be considered as the same neighborhood. To determine a reasonable combination of *min_sample* and *radius* parameter values, we tried different combinations and examined whether groups' final event venues fall within one of the group location clusters detected. Each group location cluster is represented by the centroid point of all location trace points that belong to that cluster. Figure 6.2 shows the cumulative distribution function (CDF) of distances between a group event's final venue and the closest group location cluster using different combinations of *min_sample* and *radius* values. For comparison purpose, all groups are also divided equally into low-density and high-density areas based on the population density of their centroid points. The population density information we use is collected from the 2016 US area development degree data from the US Census Bureau [20].

We find that in both high density and low density areas, when setting $min\_sample = 40, radius = 20\ meters$, more than 90% of the final event venues fall into the closest clusters



(a) Distribution of distances in high density areas.



(b) Distribution of distances in low density areas.

Figure 6.2: Figure 6.2a and Figure 6.2b show the cumulative distribution function of distances between each group event's final venue and the closest group location cluster for groups in high density and low density areas, respectively. As can been seen in the figures, when setting $min\_sample = 40, radius = 20meters$, more than 90% of group events' final venues fall within the closest group location cluster (within 500 meters).

(within 500 meters). When setting $min\_sample = 20, radius = 20\ meters$, locations such as highway transitions may fall within the same cluster. The other combinations fail to cover a significant proportion of group event venues in any of the group location clusters. We further confirm that by choosing $min\_sample = 40, radius = 20\ meters$, the DBSCAN algorithm is able to find sensible clusters in groups' location traces based on a visual inspection of these clusters plotted on a map. These location clusters appear to correspond to locations frequently visited by our participants, such as work, school, home, etc. Based on this analysis, we set the min_sample value as 40 and the $radius$ value as 20 meters in our group location cluster detection task.

It is also worth noting that in roughly 10% of the group events, the groups' final meeting venues did not fall into any of the group clusters detected. One reasonable explanation is that some group members may have been to this location some time ago, but not recently. For example, users may have been to Costco two weeks ago, which is not covered by more recent location traces. It is also possible that groups of users decided to explore a new venue that they had never tried before. We plan to investigate this problem in more detail in the future when we collect more data.

6.4 Social-based Group Location Cluster Prediction Model

Given the group location clusters determined in the previous section, our goal in this section is to design a statistical model to predict which location cluster the group is going to meet at. Motivated by prior work in group recommendation [58, 26], we divide this problem into two steps. First, we analyze strong predictors that will impact group member's meeting decisions and construct an innovative *"Edge-RWR+CART"* model to incorporate these identified features for predicting individual location preference. After that, we aggregate group members' location preferences to predict group decisions. An earlier study also suggests that the group decision strategy varies depending on the strength of social connections among group members [58]. Inspired by this observation, in the context of group event venue recommendation, we propose a novel *"social-based"* group location prediction model, which adaptively applies different group decision strategies to groups with different social relationship strength to aggregate group members' individual location

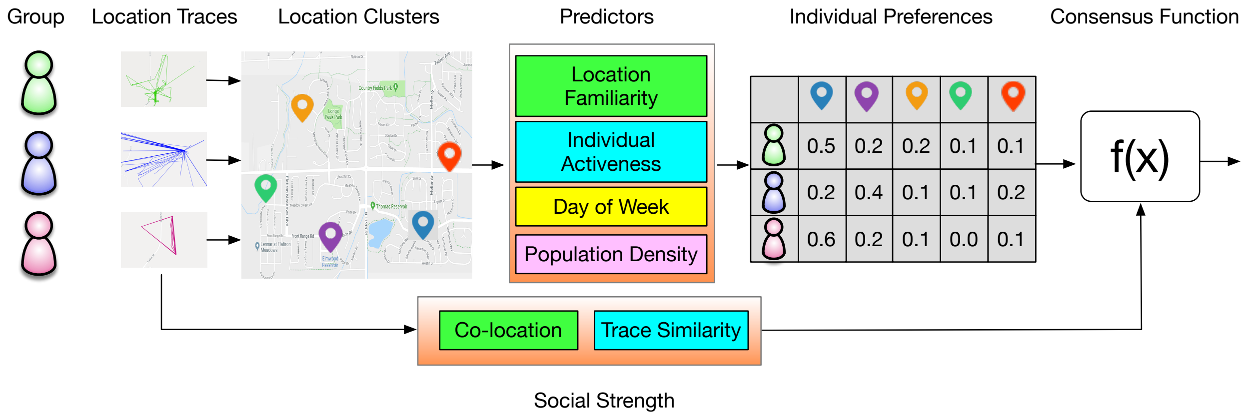


Figure 6.3: The overall architecture of our group location prediction model. With the detected group location clusters, we model each group member's individual location cluster preference based on his/her location traces. Meanwhile, the location traces are also used for estimating the group's social relationship strength. After that, a newly designed consensus function that we term a *"social-based"* group decision strategy is used to aggregate individual members' location cluster preferences and predict the group's final decision.

preferences.

6.4.1 Overall Architecture of the Model

The overall architecture of our prediction model is illustrated in Figure 6.3. As shown in the figure, individual group members' location traces are first merged to generate the group location clusters shown on the map. We then determine each individual's likelihood of visiting each location cluster, as shown in the table on the right side of the figure. This is accomplished by performing an analysis of the key factors influencing each individual's likelihood of visiting a location cluster, including location familiarity, individual activeness, day of the week, and population density. After determining the individual likelihoods of visiting a location cluster, we then measure the social strength of the group in order to combine the individual location preferences into a prediction of the group's likelihood of visiting each location cluster. This is depicted as the group consensus function $f(x)$ in the figure, along with the two social strength factors that we will use, co-location and trace similarity.

6.4.2 Social Relationship Strength

The social relationships among group members play a significant role in group decision making. Consider the following scenario: Tony and Paul are in the same group voting on a location for dinner on Friday night, and Paul does not like spicy food. If Tony knows Paul very well, he may not suggest or vote for Thai restaurants that are famous for spicy cuisine, even though he himself loves spicy food. A strong social relationship with Paul helps Tony to propose better meeting options. Here we estimate a group's social relationship strength using two factors:

- *Average number of daily meetings detected between any two group members using their location traces*: The more often two users meet in real life, the more likely that they know each other well. We use the following criteria to define a meeting in our analysis: Two members must be approximately at the same location (within 20 meters) for at least five minutes. Please note that we use GPS location points, instead of location clusters, as a more accurate estimate of group members' meetings can be obtained using this data. For example, two friends can coincidentally appear at the same farmers market for more than five minutes but not see each other, and thus this should not be counted as a meeting.

- *Average similarity of group location cluster familiarity between every pair of group members*: Two users with more similar location cluster familiarity tend to know each other better. Here we model the group location cluster familiarity of user u by vector $\mathbf{loc}(u)$, where the ith element of $\mathbf{loc}(u)$ represents the number of location trace points in the ith location cluster detected in our dataset. Furthermore, $\mathbf{loc}(u)$ is normalized per user.

Formally, we first measure the social relationship between two group members u_i and u_j as

$$\text{social}(u_i, u_j) = w_1 M(u_i, u_j) + (1 - w_1) \cdot Cosine(\mathbf{loc}(u_i), \mathbf{loc}(u_j)), \tag{6.1}$$

where $M(u_i, u_j)$ is the normalized average number of meetings per day detected between group members u_i and u_j, and $\mathbf{loc}(u_i)$ is the vector representing $u_i's$ location familiarity for each group location cluster. w_1 balances the relative importance of co-occurrence and location trace similarity.

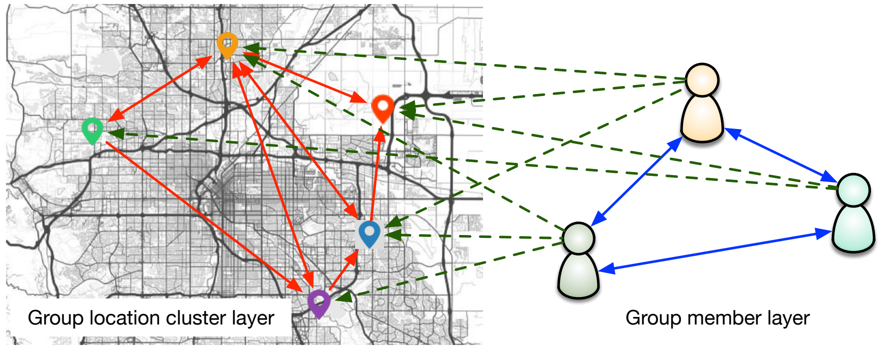


Figure 6.4: An illustration of a group location cluster-group member graph. There are two layers in the graph: the group location cluster layer and the group member layer. The nodes in the location layer represent group location clusters, while nodes in the group layer represent group members. There are three types of edges in this graph: the user-cluster edge (green), cluster-cluster edge (red) and user-user edge (blue).

Normally user location co-occurrence is a much stronger signal than location trace similarity in indicating social relationship strength. Hence w_1 is set to 0.8 in our experiments based on the observations from our dataset. The similarity of two group members' location cluster familiarity vectors is measured by Cosine similarity:

$$Cosine(\mathbf{loc}(u_i), \mathbf{loc}(u_j)) = \frac{\mathbf{loc}(u_i) \cdot \mathbf{loc}(u_j)}{||\mathbf{loc}(u_i)|| \times ||\mathbf{loc}(u_j)||}. \tag{6.2}$$

Given the social relationship strength computed between pairs of users, a group's social relationship strength is defined as the average of all the pairwise social relationships among the group members:

$$\text{Social}(G) = \frac{2\sum_{u_i, u_j \in G} \text{social}(u_i, u_j)}{|G|(|G|-1)}, \tag{6.3}$$

where $|G|$ represents the group size.

6.4.3 Location Familiarity Modeling for Individuals

Next, we focus on what factors may influence an individual to visit a particular location cluster. Figure 6.3 listed four such factors that we considered, and in this section we focus first on location familiarity as a predictor, followed by the three other factors considered in the next subsection. The location familiarity of individuals within a group may vary from user to user. To model users' location familiarity more accurately, we leverage transitions between group location clusters and the social relationship between group members. Intuitively, if a user's close friend is very familiar with one location cluster, it is highly likely that this user has been there or at least heard about it before, even though a visit to this location was not detected in location trace data. Taking this into consideration, we construct a two-layer graph that effectively combines group members' social relationships and location trace information. The structure of the graph is shown in Figure 6.4. There are two layers in the graph: the group location cluster layer and the group member layer. The nodes in the group location cluster layer represent group location clusters, and the nodes in the group member layer represent group members. There are three types of edges in this graph: the user-cluster edge (green), the cluster-cluster edge (red), and the user-user edge (blue).

Edge-specified Random Walk with Restart (Edge-RWR). To estimate a group member's familiarity to a group location cluster, we propose a novel Edge-specified Random Walk with Restart (Edge-RWR) algorithm to calculate the relevance score between a group member and a group location cluster. Unlike the traditional RWR method that treats every edge equally, we define the transition probability of each edge follows.

The *user-cluster* edge describes the probability that a user will visit a location cluster. Instead of simply using the number of visits, we apply the widely used idea of TF-IDF [150], where higher values indicate that the user is more interested in this location cluster and this interest is significant among other group members. The same approach is utilized when estimating cluster-cluster and user-user edges' probabilities. Specifically, given m users and n location clusters of a group, the user-cluster subgraph $\mathbf{G_{UC}}$ is represented by a $m \times n$ adjacency matrix $\mathbf{M_{UC}}$, where $\mathbf{M_{UC}} = \{p(c_j|u_i)\}, 0 \leq i < m, 0 \leq i < n$, and $p(c_j|u_i)$ is estimated by:

$$p(c_j|u_i) = \alpha_1 \frac{Avg(u_i, c_j)}{Avg(c_j)} + (1 - \alpha_1)\frac{1}{m}, \tag{6.4}$$

where $Avg(u_i, c_j)$ refers to the average number of visits per day made by user u_i to cluster c_j. $Avg(c_j)$ refers to the average number of total visits to cluster c_j per day made by all the users in the group. And $1-\alpha_1$ is the probability that the walker will restart from a randomly selected user node.

The *cluster-cluster* edges measure the transition probabilities of users going from one location cluster to another. Given n detected location clusters, the cluster-cluster subgraph $\mathbf{G_{CC}}$ is represented by a $n \times n$ adjacency matrix $\mathbf{M_{CC}}$, where $\mathbf{M_{CC}} = \{p(c_j|c_i)\}, 0 \leq i < n, 0 \leq j < n$, and $p(c_j|c_i)$ is estimated by:

$$p(c_j|c_i) = \alpha_2 \frac{Avg(c_i, c_j)}{Avg(c_j)} + (1-\alpha_2)\frac{1}{n}, \tag{6.5}$$

where $Avg(u_i, c_j)$ refers to the average number of transitions per day from cluster c_i to cluster c_j. $Avg(c_j)$ refers to the average number of total transitions to cluster c_j per day from all of the group location clusters. And $1-\alpha_2$ is the probability that the walker will restart from a randomly selected location cluster node.

The *user-user* edges represent how well two users know each other. We use $social(u_i, u_j)$, introduced in Equation 6.1, to model the social relationship strength between two users, u_i and u_j. Given m group members, the user-user subgraph $\mathbf{G_{UU}}$ is represented by a $m \times m$ adjacency matrix $\mathbf{M_{UU}}$, where $\mathbf{M_{UU}} = \{p(u_j|u_i)\}, 0 \leq i < m, 0 \leq j < m$, and $p(u_j|u_i)$ is estimated by:

$$p(u_j|u_i) = \alpha_3 social(u_i, u_j) + (1-\alpha_3)\frac{1}{m}, \tag{6.6}$$

where $1-\alpha_3$ is the probability that the walker will restart from a randomly selected user node.

ALGORITHM 1: Edge-specified Random Walk with Restart (Edge-RWR)

Input: Network $G = G_{UC} + G_{CC} + G_{UU}$; starting node u_i; restart probability α_1, α_2, α_3;
Output: Stationary group location cluster vector w_C for a random walk starting at u_i;

1: Let all the nodes be initialized to 0, except for a 1 for the u_i node;
2: Let w_U denote the weights of the column vector of users, and w_C denote the weights of the column vector of location clusters;
3: While ($\mathbf{U}$ and $\mathbf{C}$ have not converged):
4: $\quad {w_C}^{k+1} = \alpha_1 \mathbf{M_{UC}} \cdot {w_U}^k + \alpha_2 \mathbf{M_{CC}} \cdot {w_C}^k + (1-\alpha_2)e_C$;
5: $\quad {w_U}^{k+1} = (1-\alpha_1)e_U + \alpha_3 \mathbf{M_{UU}} \cdot {w_U}^k + (1-\alpha_3)e_U$;
6: Output group location cluster vector w_C.

Time Complexity. Algorithm 1 presents the algorithm for finding the stationary vector of the random walk from a single starting user node. The time complexity of *"Edge-RWR"* is $O(k \cdot |V|^2)$, where k is the number of iterations required for convergence, and $|V|$ is the total number of nodes. According to [188], the ratio of the first two eigenvalues of the transition matrix specifies the rate of convergence to the stationary point.

After constructing the group location cluster-group member graph, *"Edge-RWR"* is then performed to compute the location familiarity between a user and a location cluster. The random walker will start from one node, and travel following the probability assigned to each edge to select the next transition. In addition, in order to avoid stoppage of the random walk, we assume that the random walker will in some cases jump back and restart at the same node. At each transition, the walker would either restart at one node (with probability $1 - \beta$), or move to a neighboring node (with probability β). In the two-layer graph that we constructed, the walker will start from a group member node. For each transition, the walker will either jump to an adjacent node, or jump back to the same starting user node. We set the parameter $\alpha_1 = \alpha_2 = \alpha_3 = 0.85$, which has also been widely used in prior literature. The intuition is that if a group location cluster node can be easily reached from a group member node, the user has a higher familiarity with this location cluster. Formally, the location familiarity of an individual group member u for location cluster c is modeled as

$$\text{loc\_familiarity}(u, c) = \text{PageRank}_u(c), \tag{6.7}$$

where $PageRank_u(c)$ is the PageRank score of group location cluster c for user u after running the Edge-RWR algorithm starting from user u.

6.4.4 Contextual Predictors

User Mobility. User mobility can significantly impact daily activities. In a preference survey, Elgar **et al.** find that users with different levels of mobility differ substantially in how they value time [50]. Previous insights also suggest that users with high mobility (longer daily travel distance) tend to be more active in attending group events, and an individual's mobility has a positive correlation with her time and location availability [218]. There are two reasonable explanations for this phenomenon:

- As researchers have found, people with different modes of transportation, such as walking, taking a bus, riding a bike, and driving, exhibit significant difference in how they value time [50, 223]. Users with high mobility may travel by car or public transportation. Highly mobile users are more likely to attend group events that are far away from their current locations.

- Zhang **et al.** also found that users with high mobility are more likely to be active event attendees, as they are used to frequent meetings with friends after school or work, which generates longer travel distances [218].

As such, we model the influence of user mobility as follows: a user's group event attendance is proportional to the user's mobility. Here we define user mobility as the total travel distance in the 48-hour period preceding the creation of the event. More specifically, we estimate user mobility by summing up the geographical distance between every two consecutive location points detected from the user's location trace data in this time period.

Day of the Week. Using the detected group location clusters, we can identify a user's daily movement pattern among these location clusters. We find in our data analysis that users' movement patterns can be quite different on weekdays compared with weekends. The average inter-cluster movement distance is 5.3 km on weekends and 4.8 km on weekdays. This pattern seems reasonable, since intuitively people are more willing to travel longer distances for activities on weekends. And the average movement number is 5.2 on weekdays and 5.4 on weekends.

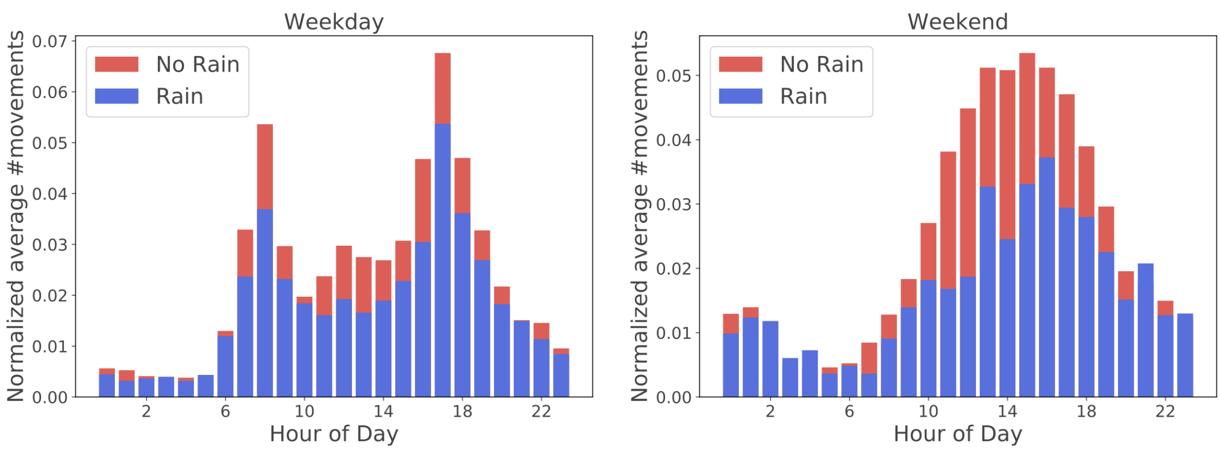
We also find that weather has a significantly different impact on weekdays compared with weekends. We control geographical confounding factors relating to weather by only using a partial dataset collected at our university. When obtaining the weather information during the time of a given movement made by a user, we round the time to the nearest hour. For example, if the starting time of the movement is at 18:15, we collect the weather data at 18:00 that day using the Weathersource API [200]. For the hours with positive precipitation values (including snow), we consider it to be rain. The normalized average movements by hours under different weather conditions on weekdays and weekends are shown in Figure 6.5. These results demonstrate that bad weather causes a sharp decrease in daily movements on both weekdays and weekends. We also find that the movement difference between rain and no-rain hours is larger on weekends, indicating that weather has a higher impact on weekend activities compared with weekdays.

In summary, users' movement patterns are considerably different on weekdays and weekends, which also impact users' group meeting preferences. Therefore, it would be beneficial to consider the day of the week information in our prediction model.

Population Density. As shown in Figure 6.1a, our users are spread all over the US, across 40 states and 117 cities. Population density can play a significant role in modeling group members' location preferences. Previous research has found that geography and social relationships are inextricably intertwined. People living in rural and urban places exhibit significant differences in using social technologies, such as Facebook and Myspace [59, 8]. These findings inspire us to study the effect of population density on group event scheduling. Therefore, we also incorporate the population density information into our prediction model.

6.4.5 Individual Location Preference Prediction

Taking all of the features we have discussed thus far into consideration, we integrate all of the contextual features into a Classification and Regression Tree (CART) model [109]. We predict a user's location preference with four types of features: *location familiarity*, *user mobility*, *day of the week*, and *population density*. Location familiarity is calculated using Equation 6.7 introduced in Section 6.4.3. User mobility models a user's average travel distance on a daily basis. For users with high mobility, traveling longer distances can be easier compared with low-mobility users, due to the fact that high-mobility users generally travel by car or public transportation, and they are accustomed to longer-distance travel. Our analysis further illustrates that users' movement patterns vary significantly on weekdays versus weekends, and in high population density areas versus low population density areas; thus, we also add these two features into our regression model. We also evaluated other popular regression methods, such as Linear Regression, Support Vector



(a) Number of movements by hour on weekdays.

(b) Number of movements by hour on weekends.

Figure 6.5: Figure 6.5a and 6.5b show the normalized average number movements on weekdays and weekends. Bad weather causes a sharp decrease of daily movements on both weekdays and weekends, and the impact is larger on weekends compared with weekdays.

Regression, and Gradient Boosting Regression. CART outperforms all of these regression models for our application. We present detailed results in Section 6.6.

One major advantage of our individual location preference prediction model is its flexibility in leveraging new features. Higher-level mobility features, which are related to users' group event attendance activity, can be easily added to our current model. For example, [174] and [76] have found that a user's mode of transportation, including car, bus, train, walking, and bike, can be accurately inferred using GPS location traces. As future work, we plan to investigate how users' modes of transportation can impact group event attendance.

6.4.6 Social-based Group Location Prediction Model

Thus far, we have constructed a prediction model that integrates all the features we have identified in an intelligent way to model an individual's location preferences. Now, we propose a new *"social-based"* group location prediction model to aggregate the predicted location preferences for individuals and predict the final location cluster where the group will meet.

Given every group member's location preferences, how does the group arrive at a final decision of where to meet? In previous group recommendation work, three main strategies for aggregating individual preferences have been widely explored [84, 58, 26]: *"least misery," "average satisfaction,"* and *"maximum satisfaction."*

- *Least misery* minimizes the dissatisfaction of the least satisfied member: $\min_{u \in G} \mathbf{p}(u, c)$.
- *Average satisfaction* is a straightforward strategy that assumes every group member carries the same weight and computes the average satisfaction: $\frac{1}{|G|} \sum_{u \in G} \mathbf{p}(u, c)$.
- *Maximum satisfaction* maximizes the enjoyment of the most satisfied group member: $\max_{u \in G} \mathbf{p}(u, c)$.

Earlier studies have also argued that groups are diverse and none of these group decision strategies are dominant across all groups [117, 58]. Gartrell **et al.** found that the social relationship strength of a group plays a significant role in the group decision-making process. When the relationship strength is high, the final decision tends to reflect the maximum satisfaction strategy. When the relationship strength is weak, the final decision tends to reflect average satisfaction or least misery [58]. The intuition here is clear: If the group members don't know each other, the main guidance for making the group decision is to avoid dissatisfying the other group members. If the group has high social relationship strength, some people who are knowledgeable about the topic may lead the discussion and others tend to follow their suggestions. Taking these factors into consideration, we propose the following *"social-based"* group consensus function:

$$\mathbf{P}(G,c) = \begin{cases} min_{u \in G}\mathbf{p}(u,c) & \text{if} \quad \text{Social}(G) < \beta \qquad \text{(least misery)} \\ \frac{1}{|G|}\sum_{u \in G}\mathbf{p}(u,c) & \text{if} \quad \beta \leq \text{Social}(G) \leq \alpha \quad \text{(average satisfaction)} \\ max_{u \in G}\mathbf{p}(u,c) & \text{if} \quad \alpha < \text{Social(G)} \qquad \text{(maximum satisfaction)}, \end{cases} \tag{6.8}$$

where Social(G) is defined in Equation 6.3, and parameters α and β are thresholds for the group's social relationship strength. We choose $\alpha = 0.6$ and $\beta = 0.2$ in our evaluation based on a grid search targeting the best performance for predicting groups' final meeting location clusters using our dataset. We will present the details of our experiments in Section 6.6.

6.5 Venue Recommendation Framework

In this section, we present our proposed group event venue recommendation framework. The intuition behind our framework is to re-rank group's nearby venues, returned by Foursquare's API, based on the group location cluster prediction results calculated by Equation 6.8 to provide event venue recommendation for groups. The main flow of the framework is shown in Figure 6.6. Aiming to recommend event venues for groups of mobile users in real-life, our framework is composed of three parts:

(1) The group location clusters generated by grouping location traces' points described in Section 6.3;

(2) The *"social-based"* group location prediction model introduced in Section 6.4;

(3) The venue recommendation system provided by Foursquare's API [53].

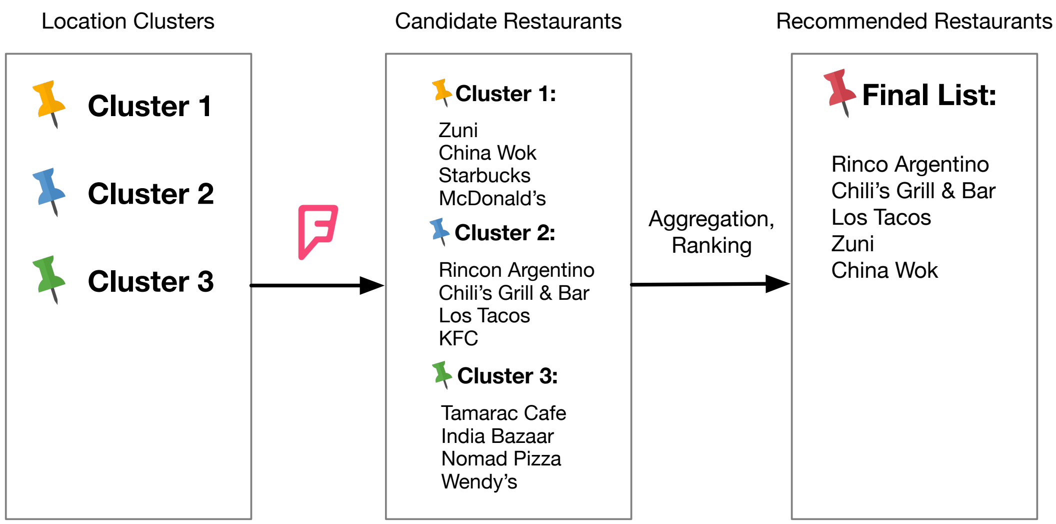


Figure 6.6: The main flow of the group event venue recommendation framework: For each detected group location cluster, we call the Foursquare Venue Recommendation API to get a list of nearby restaurants. We aggregate the restaurants of each cluster to generate a candidate restaurant pool. Prediction results from the *"social-based"* group location cluster prediction model are then used to rank the restaurant pool and generate the final Top-K restaurant list.

Given a combination of location coordinates, a radius, and a limit (maximum number of results to return), the Foursquare API will return a list of recommended venues near the specified location within the given radius. In our setting, given a group G and its group location clusters $\{c_1, c_2, \cdots, c_n\}$, we call the Foursquare API to return a list of m recommended venues $\{v_{i1}, v_{i2}, \cdots, v_{im}\}$ for each cluster c_i. Here $\{v_{i1}, v_{i2}, \cdots, v_{im}\}$ follows the order returned by the Foursquare API. All recommended venues are combined to generate a candidate venue pool. Two factors are leveraged to re-rank the venue pool: the venue rank in the returned list provided by the Foursquare API and probabilities of corresponding location clusters being selected as the final one calculated by our *"social-based"* group location prediction model. More specifically, the rank score of a venue v_{ij} in the final recommendation list is defined as

$$\text{rankscore}(v_{ij}) = \frac{m-(j-1)}{m} \cdot \mathbf{P}(G, c_i), \tag{6.9}$$

where $\frac{m-(j-1)}{m}$ models the rank weight returned by the Foursquare API and $\mathbf{P}(G, c_i)$ is the probability of location cluster c_i being selected as the final location cluster computed by Equation 6.8. The final venue recommendation list created by our framework is ranked by rankscore(v_{ij}) among all the venues in the candidate venue pool.

6.6 Evaluation

In this section, we evaluate the effectiveness of GEVR's proposed individual location preference prediction model, social-based group location cluster prediction model, and group event venue recommendation framework.

6.6.1 Performance of Predicting Individual Location Preference

We first evaluate the performance of predicting individual location preference. Specifically, we evaluate "whether a group member will vote for a suggested venue or not," and the answer is a binary label with 1 representing *yes* and 0 representing *no*. The evaluation metrics we use are AUC (the area under the ROC curve) and F1 score. Since our model quantifies a group member's location cluster preference by outputting a probability value, AUC can find the best threshold for the binary classification. And F1 score is computed to evaluate whether the model provides comparatively robust performance. We use a stratified 5-fold cross validation with respect to the users. All the features are standardized.

As introduced in Section 6.4, our prediction model utilizes four main types of features: location familiarity, user activeness, day of the week, and population density. By comparing the performance of the popular supervised models shown in Table 6.2, we find that *"Edge-RWR+CART"* achieves the best results. The other supervised regression models include Linear Regression (LR), Support Vector Regression (SVR), and Gradient Boosting Regression (GR). *"Edge-RWR+CART"* achieves the best overall performance in estimating users' location cluster preference (AUC = 0.851, F1 = 0.831) and outperforms all the other supervised models by a clear margin (beating the second best one *"Edge-RWR+GR"* by 8.9% in AUC and 7.1% in F1 Score). More importantly, our newly proposed *"Edge-RWR"* beats the traditional *"RWR"* consistently when combined with each supervised learning model that we tested with an average improvement of 13%, indicating that our edge-specified network is superior in estimating the location cluster preference between a group member and a location cluster.

6.6.2 Performance of Predicting Where Groups Meet

The next question is whether we can accurately predict where a group meets by using our newly designed *"social-based"* group decision strategy. We formulate the problem as a **multi-class prediction task**: Given all the group location clusters, can we accurately predict which location cluster the group will decide to meet at. Note that the design of the mobile application allows a host to override the voting results by selecting a final venue that is not necessarily the winning one (based on group members' votes). This may happen when group members discussed where to hang out and made an agreement to disregard the poll results. In our dataset, the host decided to override the majority voting results in roughly 20% of the group events. Therefore, we could measure the effectiveness as either predicting the most popular location cluster according to the poll

Table 6.2: Performance for predicting individual location preference. Standard error values are shown in parentheses. *"Edge-RWR+CART"* outperforms all the other supervised models by a clear margin. Moreover, our newly proposed *"Edge-RWR"* beats the traditional *"RWR"* consistently when combined with each supervised learning model that we tested, indicating that our edge-specified network is superior in estimating the familiarity between a group member and a location cluster.

Model	**AUC**	**F1 Score**
RWR+LR	0.702 (0.020)	0.675 (0.021)
Edge-RWR+LR	0.705 (0.017)	0.669 (0.022)
RWR+SVR	0.745 (0.010)	0.722 (0.008)
Edge-RWR+SVR	0.763 (0.011)	0.757 (0.009)
RWR+GR	0.788 (0.015)	0.747 (0.013)
Edge-RWR+GR	0.795 (0.011)	0.766 (0.008)
RWR+CART	0.838 (0.011)	0.812 (0.009)
Edge-RWR+CART	**0.851 (0.011)**	**0.831 (0.010)**

Table 6.3: Accuracy of predicting the winning location cluster and final location cluster. Individual location preferences are estimated using *"Edge-RWR+CART"* as it achieves the highest performance in modeling individual location preference. Standard error values are shown in parentheses. Our *"social-based"* group location prediction model achieves better prediction accuracy than all the other comparative models, with an average improvement of 12%.

Model	Winning location cluster	Final location cluster
Least misery [148]	0.689 (0.019)	0.668 (0.018)
Average satisfaction [58, 26, 148]	0.771 (0.016)	0.753 (0.017)
Most pleasure [26, 148]	0.760 (0.019)	0.739 (0.022)
Logistic regression-based [26]	0.767 (0.016)	0.741 (0.019)
Rule-based [58]	0.686 (0.022)	0.657 (0.024)
Expertise-based [148]	0.758 (0.018)	0.729 (0.016)
Social-based	**0.826 (0.013)**	**0.801 (0.013)**

results, i.e., the "winning location cluster," or predicting the final location cluster actually visited by the group including overrides, namely the "final location cluster." To evaluate the performance of our model, we report the prediction accuracy of both the winning location cluster and the final location cluster. We compare our *"social-based"* group location prediction model with the three commonly used group decision models introduced in Section 6.4.6, and three other state-of-the-art group recommendation models that have been shown to be effective in their specific group settings in prior literature:

- *Logsitic regression-based* [26]: A logistic regression-based approach that determines the probability of a group view (p_G) from the individual probabilities:

$$\log \frac{p_G}{1-p_G} = \alpha_0 + \alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + ... + \alpha_n p_n. \tag{6.10}$$

- *Rule-based* [58]: A rule-based group consensus framework that constructs associative classification rules by mining the training dataset.

- *Expertise-based* [148]: Estimating group decision by averaging the individual preference with weighted expertise. The expertise is defined as the number of times this user participated before.

The results of predicting the winning location cluster and the final location cluster are shown in Table 6.3. As can be observed in the table, our newly proposed social-based group decision strategy achieves the best performance in both predicting winning location cluster (Accuracy=0.826) and final location cluster (Accuracy=0.801). It is also worth noting that our *"social-based"* group location prediction model achieves better prediction accuracy than all other three widely-used single group decision strategies, with an average improvement of 12%. This echoes the insights illustrated in [58, 117]: groups are diverse and none of the single group decision strategies are dominant across all groups. The accuracy achieved by the *"logistic regression-based"* approach also shows that a linear combination of individual preferences without considering the social relationship strength among group members does not perform well. Moreover, our *"social-based"* model performs much better than the *"expertise-based"* model. One possible explanation is that our dataset currently does not have enough events created by the same group. As a result, the expertise of group members cannot be accurately modeled in such cold-start scenarios. We would like to explore if the *"expertise-based"* model will perform better when we have more group events data. We save this topic for future study.

Performance for groups of different sizes. To better understand the impact of group size on recommendation performance, we computed the prediction accuracy for different group sizes using our *"social-based"* group location cluster prediction model. Our model achieves an average accuracy of 0.818 for groups with three members (0.826 for the winning location cluster and 0.810 for the final location cluster), 0.807 for groups with four members (0.833 for the winning location cluster and 0.781 for the final location cluster), and 0.806 for groups with 5–6 members (0.823 for the winning location cluster and 0.788 for the final location cluster). These results show that our model performs reasonably well for groups with sizes between 3–6. However, the trend or the change in prediction performance is not statistically significant. As described in Section 6.1, only 17% of the groups in our dataset have more than four members. This would be an interesting problem to explore when we collect more data for larger groups.

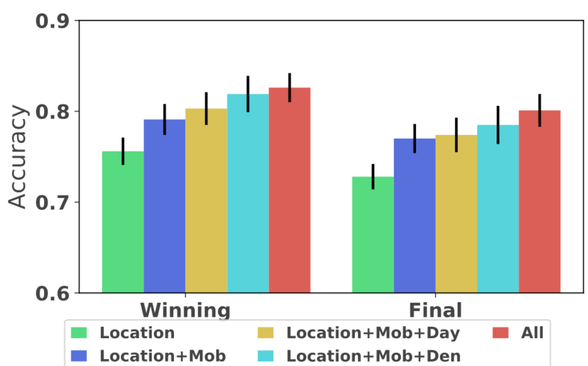
Feature analysis. Our *"social-based"* prediction model utilizes four types of features: Location familiarity (Location), user mobility (Mob), day of the week (Day), and population density

(Den). To understand the effectiveness of these features, we evaluate the prediction performance using different feature combinations. As shown in Figure 6.7a, using location familiarity alone achieves reasonable performance, and adding user mobility further increases the accuracy (from 0.756 to 0.791 for predicting the winning location cluster, and from 0.728 to 0.770 for predicting the final location cluster). One possible explanation is that users with high mobility may travel by car or use public transportation. Reaching far-away places are thus easier for them. It is also likely that users with high mobility are inherently active event attendees, as they are used to frequent meetings with friends after school or work [218]. Additionally, both day of the week and population density features contribute to the overall improvement.

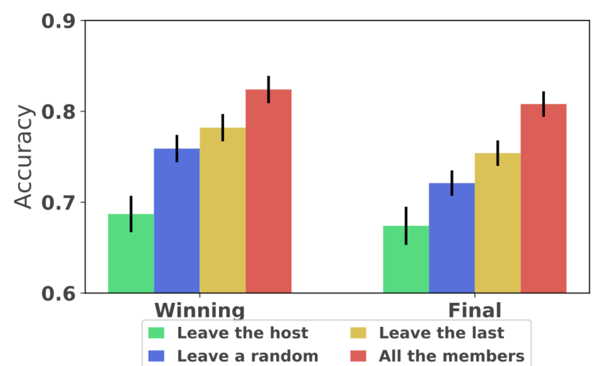
Performance on Groups with Cold-Start Users. As we discussed in previous sections, it is common for groups to have newly joined members without (sufficient) historical information. To evaluate our *"social-based"* group location prediction model in such cold-start scenarios, we set up two experiments:

- *Leave-one-out*: Assuming one group member is a new user of the app, we do not have any historical information about him/her.
- *Effect of data sparsity*: We further study how our model deals with the cold-start problem by using only a proportion of each user's data (20%, 40%, 60%, 80%, and 100%).

Figure 6.7b shows the performance of the "leave-one-out" experiment. We simulate three "leave-one-out" situations by leaving (a) the group host, (b) a random group member (except the host and the last joined group member), or (c) the last joined group member as the new user with no historical information available. As we can see in Figure 6.7b, the group host's location information has the largest impact on the overall prediction performance when predicting both the "winning location cluster" and the "final location cluster," while the last joined group member has the least impact. This finding echoes an observation made in [218]: "The group host will have more influence on the group decision-making process. The late coming voters tend to vote for options that align with existing voting results, thus, having smaller impacts on the final decisions." More importantly,

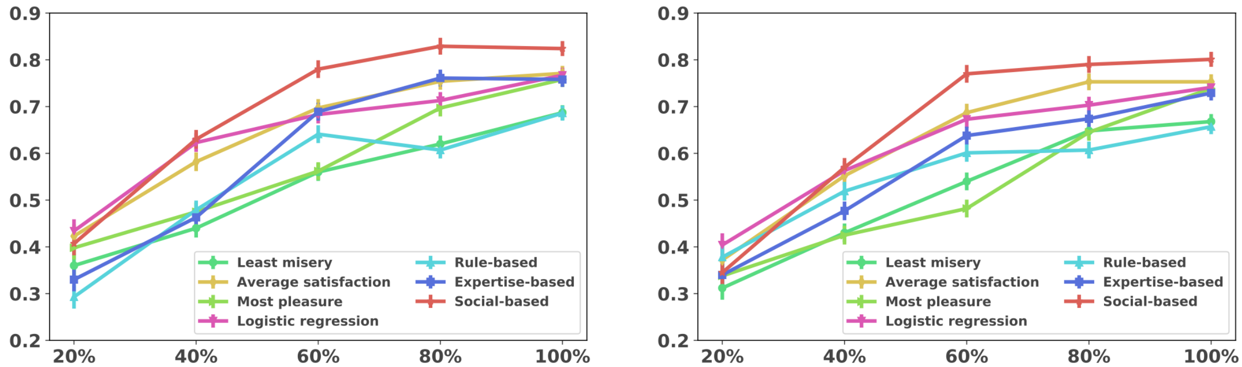


(a) Feature analysis.



(b) Cold-start analysis.

Figure 6.7: Feature analysis (Figure 6.7a) and cold-start analysis (Figure 6.7b). In Figure 6.7a, using the location familiarity and user mobility features already provide reliable performance. Both day of the week and population density features also contribute to the overall improvement. In Figure 6.7b, the group host's location information has the most significant impact on the overall prediction performance when predicting both the "winning location cluster" and "final location cluster," while the last joined group member has the least impact.



(a) Accuracy of predicting the winning location cluster.

(b) Accuracy of predicting the final location cluster.

Figure 6.8: Accuracy of predicting the winning location cluster (Figure 6.8a) and the final location cluster (Figure 6.8b) using all the models in Table 6.3 with different proportions of users' location trace data. Our proposed *"social-based"* model consistently outperforms all the other models when the proportion of data is over 40%.

our *"social-based"* group location prediction model achieves promising results with cold-start users in the group, with an average accuracy of 0.768 when assuming the last group member' data is unknown and an average accuracy of 0.740 when leaving a random group member's data out.

We further study how our proposed model deals with the data sparsity problem by using different proportions of the users' location traces and compare it with all the other models shown in Table 6.3. The results are shown in Figure 6.8. We gradually increase the proportion of users' location traces used for testing the performance from the first 20%, first 40%, till 100% (temporally ordered). As can be seen in the figure, our proposed *"social-based"* model consistently outperforms all the other models when the proportion is above 40%. It does not perform very well when only 20% of the location traces are available for prediction. It is likely that the social relationship strength among group members cannot be precisely estimated with very limited data, which affects the overall prediction accuracy.

6.6.3 Performance of Event Venue Recommendation for Groups of Mobile Users

To evaluate whether our proposed GEVR system is effective for providing event venue recommendations for groups of mobile users, we compare it with the following baselines:

- *Foursquare*: Get 50 recommended restaurants provided by the Foursquare Venue Recommendation API by searching the city name where the group is located at and retain the restaurants' Foursquare rankings without any modification.
- *Most popular*: Same as the Foursquare model but re-ranks the restaurants by Foursquare number of check-ins (popularity).
- *Highest rating*: Same as the Foursquare model, but re-ranks the restaurants by Foursquare user review star rating. If there is a tie, the restaurant with more check-ins is ranked higher.
- *Equal-weighted*: For each detected group location cluster, getting 50 recommended restaurants by searching the nearby area within 0.5 km using the Foursquare Venue Recommendation API and rank them using Equation 6.9, assuming every location cluster is weighted equally. If there is a tie, restaurants with more check-ins are ranked higher.
- *Least misery* [148]: Using the *"least misery"* group location prediction model to recommend venues.
- *Average satisfaction* [58, 26, 148]: Using the *"average satisfaction"* group location prediction model to recommend venues.
- *Most pleasure* [26, 148]: Using the *"most pleasure"* group location prediction model to recommend venues.
- *Logistic regression-based* [26]: Using the *"logistic regression-based"* group location prediction model to recommend venues.
- *Rule-based* [58]: Using the *"rule-based"* group location prediction model to recommend venues.

- *Expertise-based* [148]: Using the *"expertise-based"* group location prediction model to recommend venues.

Here, getting 50 recommended restaurants means setting the field $limit = 50$ as the number of results to return when using the Foursquare API. We choose 50 because it is the maximum number allowed. The returned list could be shorter than 50 if there are not enough restaurants registered nearby.

The three basic baselines *"most popular," "highest rating,"* and *"Foursquare"* are intended to demonstrate the recommendation performance without leveraging the detected group location clusters. The *"equal-weighted"* baseline uses group location clusters information, but assumes no prediction results are provided, so all detected group location clusters are equally weighted. The next six baselines combine individual location preference with different group decision strategies, aiming to determine the feasibility of our social-based group decision strategy for event venue recommendation. We use a widely used metric, **hit rate**, to evaluate the recommendation performance of these models. For each group, we recommend top-N (N=5,10,15,20 in our experiments). Hit rate is defined as the proportion of groups' top-N recommendation lists that includes the final venue decided by the group in our dataset. For example, suppose there are 100 group events; then we would generate 100 top-N lists. For a given event, if the final venue chosen by the participants is included in the top-N recommendation list generated for that event, then we consider that the recommendation to be a success. If out of 100 events, 30 events' final venues are included in their corresponding top-N lists, then the hit rate (final venue) is 0.3.

Results of hitting the final group meeting venue for each of these baseline models, as well as our GEVR, are shown in Figure 6.9, top subfigure. We can see that our baselines gradually build towards the performance of our final recommendation system GEVR, indicating that each of the components (**location clusters**, **individual location preference model**, and **social-based group decision strategy**) is a meaningful contribution to the event venue recommendation for groups of mobile users.

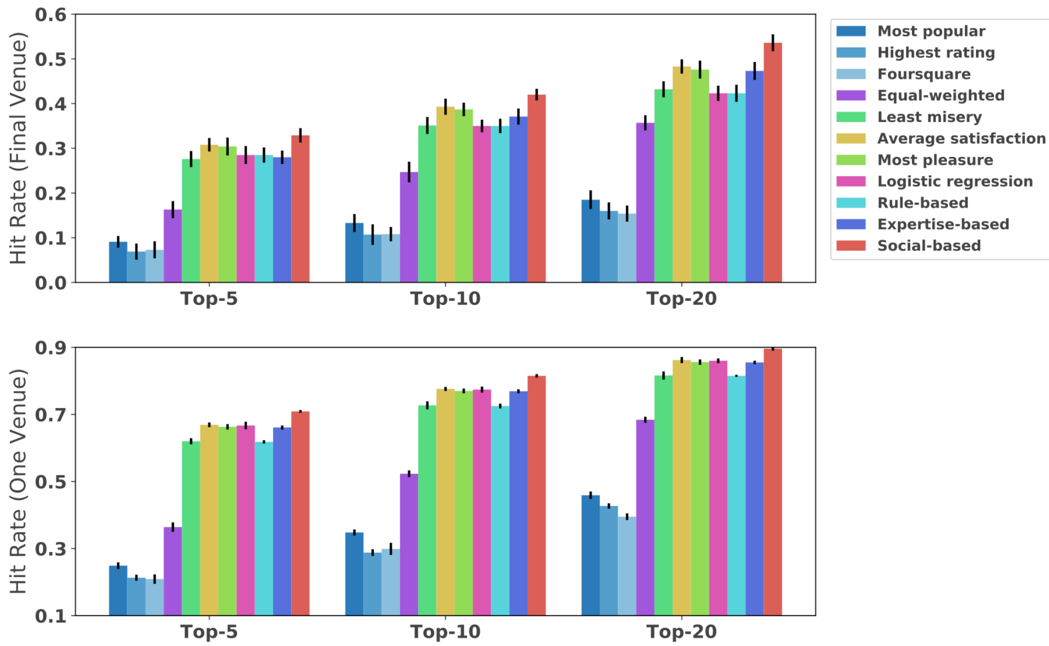


Figure 6.9: Recommendation performance of different models on hitting a group's final venue (top subfigure) and hitting at least one of the group's suggested venues (bottom subfigure). Our *"social-based"* group location prediction model outperforms all the other comparative models consistently in both cases.

The three basic baselines are clearly ineffective, achieving an average hit rate for Top 5, Top 10, and Top 20, with only 0.08, 0.12, and 0.17, respectively. In comparison, the *"equal-weighted"* model, which considers group location cluster information, achieves better performance. But it is not as effective as the remaining ones that model the group decision-making process to help re-rank the recommendation list. Our *"social-based"* group location prediction model outperforms the other six comparative models with an average improvement of 14% for Top-5, 15% for Top-10, and 19% for Top-20. This fits our expectation, as the social-based group decision strategy provides the most promising accuracy in predicting the groups' final location cluster, as shown in Table 6.3.

We further calculate the success rate of hitting at least one of the group's suggested venues, i.e., venues that they voted on, using all these models. This is a different and broader metric than the prior hit rate (final venue), so we define this as hit rate (one venue). For example, suppose for a given event that the group members have suggested three venues that they are interested in visiting. Then if any one of those three appear in the top-N list, we consider that recommendation to be a success. If out of 100 group events, at least one of the voted-upon venues appear in a top-N list for 30 of those events, then the hit rate (one venue) is 0.3. The results of this metric are shown in Figure 6.2, bottom subfigure. As can be seen, the relative recommendation performance of the different methods follows a similar pattern as the previous metric. The overall absolute value of this hit rate (one venue) metric is higher than the final venue hit rate because there is a higher chance of any one of the suggested venues appearing in a top-N list compared with just the final venue appearing.

6.7 Discussion

In this work, we present GEVR, the first event venue recommendation system for groups of mobile users. The system tackles the recommendation challenges by splitting the problem into three separate steps: (1) detecting group location clusters using group members' location traces; (2) predicting where the group will meet; and (3) aggregating and re-ranking venues in nearby areas, crawled through the Foursquare Venue Recommendation API, to create a final venue recommendation list based on the prediction results. When predicting a group's gathering decision, we first model individual group members' location preferences based on four types of features: location familiarity, user activeness, day of the week, and population density. We then design a novel social-based group location prediction model to aggregate individual preferences and predict the group's final decisions based on the group's social relationship strength.

To evaluate the performance of GEVR, we have collected a unique dataset from a group event scheduling mobile application, OutWithFriendz. In total, 625 users participated in this study and created 502 group events. Our users are widespread across the US, covering 40 states and 117 cities. Evaluation results show that GEVR can provide over 80% accuracy for predicting which location cluster the group will meet at, and our newly designed social-based group location prediction model outperforms all the other state-of-the-art group location prediction models with an average improvement of 16% on hit rate.

One of the significant implications of this work is that it is essential to integrate location considerations into an event venue recommendation framework for groups of mobile users. As demonstrated clearly in Figure 6.9, approaches that do not consider group members' location behavior in the recommendation process performed much worse than our GEVR approach. We expect that the research community will build upon this finding, and future solutions for group recommendation in other contexts will also incorporate mobility into their recommendation strategies.

Similarly, our work clearly demonstrates that methods for individual venue recommendation such as *"most Popular," "highest rating,"* and *"Foursquare"* are considerably less effective when applied to group venue recommendation than algorithms that incorporate knowledge of group characteristics. Group properties, such as mobility spread, membership fluidity, and diversity of interests, introduce a higher level of complexity that impacts the final venue chosen by the group. Unlike individual recommendation, which mostly relies on personal preferences, negotiation and coordination are necessary for group members to reach a final agreement. Understanding the group negotiation and coordination process is critical for understanding group event scheduling behaviors and providing effective group event venue recommendation. Our work also indicates that one simple group decision strategy, e.g., average satisfaction, cannot apply to all the groups. Groups with strong social relationships usually apply different strategies compared with groups with weak social relationships.

Our results are uniquely validated based on a large number of real-world group venue events that have been quite challenging to obtain. For privacy reasons, we are unable to release the dataset of group events, because the location traces are difficult to fully anonymize. However, we can lower the barrier to entry for future researchers and have obtained permission from the authors of the software used in this study to release the applications as open source. This will enable other researchers to obtain valuable ground truth data to study the fine-grained behavior of groups of mobile users, and thus drive the growth in the research community for developing practically validated recommendation solutions for groups of mobile users.

Our work builds upon commercial location-based social services (LBSNs) such as Foursquare. This strategy opens up opportunities for more in-depth partnerships with the industry on joint development of novel group recommendation algorithms and systems that can be validated in the real world. This research strengthens the capabilities of academia in approaching the industry for collaborations, such as Google, Apple, and mobile social networks like Facebook, Instagram, and Snapchat, in hopes of introducing mobile tools that can facilitate useful group event coordination and venue recommendation.

There are a variety of strong use cases for our GEVR in real life. First, online group event organization services could use GEVR to find venues in their databases that are highly likely to be suitable for most group members. As previous research has pointed out, for many social groups, how to organize an event and attract more participants continues to be a difficult task. Even experienced organizers still feel stressed when planning a group event [3]. Our GEVR could to some extent reduce the group event host's burden of deciding where to meet. Additionally, although future research is needed to verify this hypothesis, it is possible that our GEVR could help target potential participants who would be interested in attending this group event. With the suggested venues and ongoing voting results of a new group event, we may be able to find group members' friends who fit this proposal perfectly. Validating this hypothesis could also be an exciting topic for future research.

Chapter 7

Future Work

This thesis aims to advance the understanding of human behavior and society through the lens of online social groups. Towards this direction, there are many research topics I would like to explore in the next few years. Here I pick three of the most exciting ones.

7.1 Pathways Towards Extreme Online Spaces

In the digital era, some online spaces have become a heaven for conspiracy theories. For example, /r/The_Donald [1] and 4chan [2] have been known for supporting conspiracy theories, harassment, and trolling. In the recent COVID-19 pandemic [220], wild conspiracy theories and misinformation began sprouting online [127]. Given the emergence of these extreme online spaces, it is critical to understand how they are created and develop. For people who are already radical in real life, a large proportion of them may end up in extreme places in the online world. It can be difficult for service providers to prevent them from doing so. However, other people may first participate in less extreme groups. This experience will expose them to radical ideologies and prepare them to participate in more extreme groups later on. The goal of this project is to design a framework that can detect users who are at risk of joining extreme online spaces and detect social groups that may serve as pathways towards extremism. Hopefully, with such a computational framework, service providers can implement interventions accordingly.

[1] https://www.reddit.com/r/the_donald
[2] https://www.4chan.org/

7.2 Chacterizing Human and AI Interactions in Online Social Groups

Online discussion forums generate enormous volumes of conversation. To effectively interpret the massive content and better serve users, AI can support a radical transformation of experience and management. In practice, service providers have started leveraging different AI tools on their platforms, such as providing the automation of routine tasks, enhancing personalization, and improving moderation. The usage of these AI tools provides a great opportunity to understand human and AI interactions. For example, when the AI did a good job or made a mistake, how would normal users react? When we delegate the moderation task to AI, would normal users accept them better than relying on human moderators? Additionally, how do these attitudes differ across communities? These everyday interactions on the Internet between humans and AI can provide abundant data for understanding human-AI interactions: how can machines and humans collaborate to improve the performance of various human tasks [149]?

7.3 Combating Misinformation on the Internet

The spread of misinformation is an increasing concern for society. In the digital era, there are ecosystems created on popular social media sites by foreign actors and domestic groups that intentionally promote fake news, conspiracy stories, and hoaxes. This issue is a real threat to our democracy, as it can disrupt the public trust of legitimate news sources and undermine our political spectrum. I am passionate about examining ways to combat misinformation on the web through multiple angles. The goal of this research is to characterize online communities' resistance to fake news. In the history of online communities, many may have been flooded with fake news. Tracking these communities' dynamics of resistance can provide a more comprehensive picture. For example, if a community's resistance to fake news is getting better and better over time, there can be effective strategies implemented for preventing the spread of misinformation. Contrastingly, if a community's resistance towards fake news is getting worse, there are probably lessons we can learn from this. In summary, a more in-depth look into the history of these online communities, like the user structure and rules being implemented, may teach us lessons in combating misinformation on the entire platform or even the entire Internet.

Bibliography

[1] Rural-urban continuum codest. `https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/`, 2013.

[2] Nicola Abbott and Lindsey Cameron. What makes a young assertive bystander? the effect of intergroup contact, empathy, cultural openness, and in-group bias on assertive bystander intervention intentions. Journal of Social Issues, 70(1):167–182, 2014.

[3] Judy Allen. Event planning: The ultimate guide to successful meetings, corporate events, fundraising galas, conferences, conventions, incentives and other special events. John Wiley & Sons, 2008.

[4] Dan Ariely. Predictably irrational. HarperCollins New York, 2008.

[5] Daniel Ashbrook and Thad Starner. Using gps to learn significant locations and predict movement across multiple users. Personal and Ubiquitous Computing, 7(5):275–286, 2003.

[6] National Basketball Association. NBA Official Rules 2017-18. New York, NY, USA, 2017. URL `https://turnernbahangtime.files.wordpress.com/2017/10/2017-18-nba-rule-book.pdf`.

[7] Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. Group Formation in Large Social Networks: Membership, Growth, and Evolution. In Proceedings of KDD, pages 44–54. ACM, 2006.

[8] Lars Backstrom, Eric Sun, and Cameron Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In Proceedings of WWW, pages 61–70, 2010.

[9] Christopher A Bail. Terrified: How anti-Muslim fringe organizations became mainstream. Princeton University Press, 2014.

[10] Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. Exposure to opposing views on social media can increase political polarization. Proceedings of the National Academy of Sciences, 2018.

[11] C Daniel Batson, Marina P Polycarpou, Eddie Harmon-Jones, Heidi J Imhoff, Erin C Mitchener, Lori L Bednar, Tricia R Klein, and Lori Highberger. Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group? Journal of Personality and Social Psychology, 72(1):105, 1997.

[12] Jason Baumgartner. Reddit dataset. `https://files.pushshift.io/reddit/`, 2018. [Online; accessed 15-June-2019].

[13] Daniel J Beal, Janet B Ruscher, and Sherry B Schnake. No benefit of the doubt: Intergroup bias in understanding causal explanation. British Journal of Social Psychology, 40(4):531–543, 2001.

[14] Christoph Beckmann and Tom Gross. Agremo: providing ad-hoc groups with on-demand recommendations on mobile devices. In Proceedings of the 29th Annual European Conference on Cognitive Ergonomics, pages 179–182. ACM, 2011.

[15] Colleen C Bee and Mark E Havitz. Exploring the relationship between involvement, fan attraction, psychological commitment and behavioural loyalty in a sports spectator context. International Journal of Sports Marketing and Sponsorship, 11(2):37–54, 2010.

[16] Jeremy Blackburn and Haewoon Kwak. Stfu noob!: predicting crowdsourced decisions on toxic behavior in online games. In Proceedings of the 23rd International Conference on World Wide Web, pages 877–888. ACM, 2014.

[17] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. Journal of machine Learning research, 3(Jan):993–1022, 2003.

[18] Johanne Boisjoly, Greg J Duncan, Michael Kremer, Dan M Levy, and Jacque Eccles. Empathy or antipathy? the impact of diversity. American Economic Review, 96(5):1890–1905, 2006.

[19] William J Brady, Julian A Wills, Dominic Burkart, John T Jost, and Jay J Van Bavel. An ideological asymmetry in the diffusion of moralized content on social media among political leaders. Journal of Experimental Psychology: General, 2018.

[20] US Census Bureau. Us population density dataset 2010-2016, 2016.

[21] Braz Camargo, Ralph Stinebrickner, and Todd Stinebrickner. Interracial friendships in college. Journal of Labor Economics, 28(4):861–892, 2010.

[22] Richard M Campbell Jr, Damon Aiken, and Aubrey Kent. Beyond birging and corfing: Continuing the exploration of fan behavior. Sport Marketing Quarterly, 13(3), 2004.

[23] Ellis Cashmore. Making sense of sports. Routledge, 2005.

[24] Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. Proceedings of the ACM on Human-Computer Interaction, 1 (CSCW):31, 2017.

[25] Eshwar Chandrasekharan, Mattia Samory, Anirudh Srinivasan, and Eric Gilbert. The bag of communities: identifying abusive behavior online with preexisting internet data. In Proceedings of CHI, pages 3175–3187. ACM, 2017.

[26] Allison JB Chaney, Mike Gartrell, Jake M Hofman, John Guiver, Noam Koenigstein, Pushmeet Kohli, and Ulrich Paquet. A large-scale exploration of group viewing patterns. In Proceedings of TVX, pages 31–38, 2014.

[27] Longbiao Chen, Daqing Zhang, Leye Wang, Dingqi Yang, Xiaojuan Ma, Shijian Li, Zhaohui Wu, Gang Pan, Thi-Mai-Trang Nguyen, and Jérémie Jakubowicz. Dynamic cluster-based over-demand prediction in bike sharing systems. In Proceedings of UbiComp, pages 841–852, 2016.

[28] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. How community feedback shapes user behavior. In Proceedings of the International AAAI Conference on Weblogs and Social Media, 2014.

[29] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. Antisocial behavior in online discussion communities. In Proceedings of International AAAI Conference on Web and Social Media, 2015.

[30] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. Anyone can become a troll: Causes of trolling behavior in online discussions. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, pages 1217–1230. ACM, 2017.

[31] Amy Chua. Political Tribes: Group Instinct and the Fate of Nations. Penguin, 2018.

[32] Mina Cikara and Susan T Fiske. Their pain, our pleasure: stereotype content and schadenfreude. Annals of the New York Academy of Sciences, 1299(1):52–59, 2013.

[33] Mina Cikara, Matthew M Botvinick, and Susan T Fiske. Us versus them: Social identity shapes neural responses to intergroup competition and harm. Psychological Science, 22(3): 306–313, 2011.

[34] Marci D Cottingham. Interaction ritual theory and sports fans: Emotion, symbols, and solidarity. Sociology of Sport Journal, 29(2):168–185, 2012.

[35] Bernard Cova and Stefano Pace. Brand community of convenience products: new forms of customer empowerment–the case "my nutella the community". European Journal of Marketing, 40(9/10):1087–1105, 2006.

[36] Craigslist. Craigslist. `https://www.craigslist.org/`, 2017.

[37] Vassilis Dalakas and Aron M Levin. The balance theory domino: How sponsorships may elicit negative consumer attitudes. ACR North American Advances, 2005.

[38] Srayan Datta and Eytan Adar. Extracting inter-community conflicts in reddit. In Proceedings of the International AAAI Conference on Web and Social Media, volume 13, pages 146–157, 2019.

[39] Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. Automated hate speech detection and the problem of offensive language. In Proceedings of International AAAI Conference on Web and Social Media, pages 512–515, 2017.

[40] Munmun De Choudhury, Scott Counts, Eric J Horvitz, and Aaron Hoff. Characterizing and predicting postpartum depression from shared facebook data. In Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, pages 626–638. ACM, 2014.

[41] Jeff Desjardins and Visual Capitalist. The world's 50 most valuable sports teams. http://www.businessinsider.com/the-worlds-50-most-valuable-sports-teams-2017-8, 2017. [Online; accessed 15-April-2018].

[42] Nicholas Diakopoulos and Mor Naaman. Towards quality discourse in online news comments. In Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work, pages 133–142. ACM, 2011.

[43] Peter Dolton and George MacKerron. Is football a matter of life and death–or is it more important than that? 2018.

[44] Bradley Doucet. Why are sports fans so biased? https://goo.gl/mQZTXQ, 2010. [Online; accessed 15-June-2019].

[45] John F Dovidio, Samuel L Gaertner, and Kerry Kawakami. Intergroup contact: The past, present, and the future. Group Processes & Intergroup Relations, 6(1):5–21, 2003.

[46] John F Dovidio, Anja Eller, and Miles Hewstone. Improving intergroup relations through direct, extended and other forms of indirect contact. Group Processes & Intergroup Relations, 14(2):147–160, 2011.

[47] John F Dovidio, Angelika Love, Fabian MH Schellhaas, and Miles Hewstone. Reducing intergroup bias through intergroup contact: Twenty years of progress and future directions. Group Processes & Intergroup Relations, 20(5):606–620, 2017.

[48] Jason P Doyle, Daniel Lock, Daniel C Funk, Kevin Filo, and Heath McDonald. 'i was there from the start': The identity-maintenance strategies used by fans to combat the threat of losing. Sport Management Review, 20(2):184–197, 2017.

[49] Brendan Dwyer. Divided loyalty? an analysis of fantasy football involvement and fan loyalty to individual national football league (nfl) teams. Journal of Sport Management, 25(5): 445–457, 2011.

[50] Alon Elgar and Shlomo Bekhor. Car-rider segmentation according to riding status and investment in car mobility. Transportation Research Record: Journal of the Transportation Research Board, (1894):109–116, 2004.

[51] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In KDD, volume 96, pages 226–231, 1996.

[52] Joel E Fischer, Stuart Reeves, Stuart Moran, Chris Greenhalgh, Steve Benford, and Stefan Rennick-Egglestone. Understanding mobile notification management in collocated groups. In ECSCW 2013: Proceedings of the 13th European Conference on Computer Supported Cooperative Work, 21-25 September 2013, Paphos, Cyprus, pages 21–44. Springer, 2013.

[53] Foursquare. Foursquare venue recomendation api. https://developer.foursquare.com/docs/api/venues/explore, 2018. [Online; accessed 01-November-2018].

[54] Bleacher Report Adam Fromal. Ranking every nba franchise on historical success. http://bleacherreport.com/articles/2733544-ranking-every-nba-franchise-on-historical-success, 2017. [Online; accessed 15-April-2018].

[55] Steve Frosdick and Peter Marsh. Football Hooliganism. Willan, 2013.

[56] Katie Z. Gach, Casey Fiesler, and Jed R. Brubaker. "control your emotions, potter": An analysis of grief policing on facebook in response to celebrity death. Proc. ACM Hum.-Comput. Interact., 1(CSCW):47:1–47:18, December 2017. ISSN 2573-0142.

[57] Devin Gaffney and J. Nathan Matias. Caveat emptor, computational social science: Large-scale missing data in a widely-published reddit corpus. PLOS ONE, 13(7):1–13, 07 2018.

[58] Mike Gartrell, Xinyu Xing, Qin Lv, Aaron Beach, Richard Han, Shivakant Mishra, and Karim Seada. Enhancing group recommendation by incorporating social relationship interactions. In Proceedings of ACM Group, pages 97–106, 2010.

[59] Eric Gilbert, Karrie Karahalios, and Christian Sandvig. The network in the garden: an empirical analysis of social media in rural life. In Proceedings of CHI, pages 1603–1612, 2008.

[60] Nabeel Gillani, Ann Yuan, Martin Saveski, Soroush Vosoughi, and Deb Roy. Me, my echo chamber, and i: introspection on social media polarization. In Proceedings of the 2018 World Wide Web Conference on World Wide Web, pages 823–831. International World Wide Web Conferences Steering Committee, 2018.

[61] Jennifer Golbeck, Cristina Robles, and Karen Turner. Predicting personality with social media. In CHI'11 Extended Abstracts on Human Factors in Computing Systems, pages 253–262. ACM, 2011.

[62] Eric Goldschein. It's Time to Give /r/nba the Respect it Deserves. `https://www.sportsgrid.com/as-seen-on-tv/media/its-time-to-give-rnba-the-respect-it-deserves/`, 2015. [Online; accessed 15-June-2019].

[63] Amit Goyal, Francesco Bonchi, and Laks VS Lakshmanan. Learning influence probabilities in social networks. In Proceedings of the third ACM international conference on Web Search and Data Mining, pages 241–250. ACM, 2010.

[64] Sylvie Graf, Stefania Paolini, and Mark Rubin. Negative intergroup contact is more influential, but positive intergroup contact is more common: Assessing contact prominence and contact prevalence in five central european countries. European Journal of Social Psychology, 44(6):536–547, 2014.

[65] Donald P Green and Janelle S Wong. Tolerance and the contact hypothesis: A field experiment. The Political Psychology of Democratic Citizenship, pages 1–23, 2009.

[66] Catherine Grevet, Loren G Terveen, and Eric Gilbert. Managing political differences in social media. In Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, pages 1400–1408. ACM, 2014.

[67] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. Fake news on twitter during the 2016 us presidential election. Science, 363(6425):374–378, 2019.

[68] The Stanford Natural Language Processing Group. Stanford topic modeling toolbox. `https://nlp.stanford.edu/software/tmt/tmt-0.4/`, 2010. [Online; accessed 15-April-2018].

[69] Anatoliy Gruzd and Jeffrey Roy. Investigating political polarization on twitter: A canadian perspective. Policy & Internet, 6(1):28–45, 2014.

[70] Pedro Henrique Calais Guerra, Wagner Meira Jr, Claire Cardie, and Robert Kleinberg. A measure of polarization on social media networks based on community boundaries. In Proceedings of International AAAI Conference on Web and Social Media, 2013.

[71] Richard Guilianotti. Football, violence and social identity. Routledge, 2013.

[72] Allen Guttmann. From ritual to record: The nature of modern sports. Columbia University Press, 2004.

[73] William L Hamilton, Justine Zhang, Cristian Danescu-Niculescu-Mizil, Dan Jurafsky, and Jure Leskovec. Loyalty in online communities. In Proceedings of ICWSM, pages 377–386. AAAI, 2017.

[74] Cody T Havard. Glory out of reflected failure: The examination of how rivalry affects sport fans. Sport Management Review, 17(3):243–253, 2014.

[75] Chip Heath, Chris Bell, and Emily Sternberg. Emotional selection in memes: the case of urban legends. Journal of Personality and Social Psychology, 81(6):1028, 2001.

[76] Samuli Hemminki, Petteri Nurmi, and Sasu Tarkoma. Accelerometer-based transportation mode detection on smartphones. In Proceedings of SenSys, page 13, 2013.

[77] Jack Hessel. Fighting words model quick implementation. `https://github.com/jmhessel/FightingWords`, 2016. [Online; accessed 15-June-2019].

[78] Thomas Hickman and James Ward. The dark side of brand community: Inter-group stereotyping, trash talk, and schadenfreude. ACR North American Advances, 2007.

[79] Edward R Hirt, Dolf Zillmann, Grant A Erickson, and Chris Kennedy. Costs and benefits of allegiance: Changes in fans' self-ascribed competencies after team victory versus defeat. Journal of personality and social psychology, 63(5):724, 1992.

[80] Vincent Hogan, Patrick Massey, and Shane Massey. Analysing match attendance in the european rugby cup: Does uncertainty of outcome matter in a multinational tournament? European Sport Management Quarterly, 17(3):312–330, 2017.

[81] Cheng-Kang Hsieh, Hongsuda Tangmunarunkit, Faisal Alquaddoomi, John Jenkins, Jinha Kang, Cameron Ketcham, Brent Longstaff, Joshua Selsky, Betta Dawson, Dallas Swendeman, et al. Lifestreams: A modular sense-making toolset for identifying important patterns from everyday life. In Proceedings of SenSys, number 5, 2013.

[82] Craig G Hyatt and William M Foster. Using identity work theory to understand the de-escalation of fandom: A study of former fans of national hockey league teams. Journal of Sport Management, 29(4):443–460, 2015.

[83] Giulio Jacucci, Antti Oulasvirta, Antti Salovaara, and Risto Sarvas. Supporting the shared experience of spectators through mobile group media. In Proceedings of SIGGROUP, pages 207–216. ACM, 2005.

[84] Anthony Jameson and Barry Smyth. Recommendation to groups. In The Adaptive Web, pages 596–627. Springer, 2007.

[85] Kathleen Hall Jamieson and Joseph N Cappella. Echo chamber: Rush Limbaugh and the conservative media establishment. Oxford University Press, 2008.

[86] Shuja Jamil, Anas Basalamah, Ahmed Lbath, and Moustafa Youssef. Hybrid participatory sensing for analyzing group dynamics in the largest annual religious gathering. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pages 547–558. ACM, 2015.

[87] Timothy Jay. Do offensive words harm people? Psychology, Public Policy, and Law, 15(2): 81, 2009.

[88] Kasthuri Jayarajah, Youngki Lee, Archan Misra, and Rajesh Krishna Balan. Need accurate user behaviour?: pay attention to groups! In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pages 855–866. ACM, 2015.

[89] Ian Jones. A model of serious leisure identification: The case of football fandom. Leisure Studies, 19(4):283–298, 2000.

[90] Sanjay Ram Kairam, Dan J. Wang, and Jure Leskovec. The Life and Death of Online Groups: Predicting Group Growth and Longevity. In Proceedings of WSDM, 2012.

[91] Brian Keegan, Darren Gergle, and Noshir Contractor. Staying in the loop: structure and dynamics of wikipedia's breaking news collaborations. In Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration, page 1. ACM, 2012.

[92] Brian Keegan, Darren Gergle, and Noshir Contractor. Hot off the wiki: Structures and dynamics of wikipedia's coverage of breaking news events. American Behavioral Scientist, 57 (5):595–622, 2013.

[93] Brian C Keegan and Jed R Brubaker. 'is' to'was': Coordination and commemoration in posthumous activity on wikipedia biographies. In Proceedings of CSCW, pages 533–546. ACM, 2015.

[94] Lynne Kelly, James A Keaten, Bonnie Becker, Jodi Cole, Lea Littleford, and Barrett Rothe. "it's the american lifestyle!": An investigation of text messaging by college students. Qualitative Research Reports in Communication, 13(1):1–9, 2012.

[95] Douglas T Kenrick and David C Funder. Profiting from controversy: Lessons from the person-situation debate. American Psychologist, 43(1):23, 1988.

[96] Sara Kiesler, Robert Kraut, Paul Resnick, and Aniket Kittur. Regulating behavior in online communities. Building Successful Online Communities: Evidence-Based Social Design. MIT Press, Cambridge, MA, 2012.

[97] Amy Jo Kim. Community Building on the Web: Secret Strategies for Successful Online Communities. Addison-Wesley Longman Publishing Co., Inc., 1st edition, 2000.

[98] Nuri Kim and Magdalena Wojcieszak. Intergroup contact through online comments: Effects of direct and extended contact on outgroup attitudes. Computers in Human Behavior, 81: 63–72, 2018.

[99] Srijan Kumar, William L Hamilton, Jure Leskovec, and Dan Jurafsky. Community interaction and conflict on the web. In Proceedings of the 2018 World Wide Web Conference on World Wide Web, pages 933–943. International World Wide Web Conferences Steering Committee, 2018.

[100] I Lawrence and Kuei Lin. A concordance correlation coefficient to evaluate reproducibility. Biometrics, pages 255–268, 1989.

[101] Alex Leavitt and Joshua A Clark. Upvoting hurricane sandy: event-based news production processes on a social news site. In Proceedings of CHI, pages 1495–1504. ACM, 2014.

[102] Changjun Lee, Jieun Shin, and Ahreum Hong. Does social media use really make people politically polarized? direct and indirect effects of social media use on political polarization in south korea. Telematics and Informatics, 35(1):245–254, 2018.

[103] Steven A Lehr, Meghan L Ferreira, and Mahzarin R Banaji. When outgroup negativity trumps ingroup positivity: Fans of the boston red sox and new york yankees place greater value on rival losses than own-team gains. Group Processes & Intergroup Relations, 22(1): 26–42, 2019.

[104] Weiwen Leung, Haiyi Zhu, and Joseph A. Konstan. The effect of emotional cues from the nfl on wikipedia contributions. Proc. ACM Hum.-Comput. Interact., 1(CSCW):66:1–66:21, 2017. ISSN 2573-0142.

[105] Yanhua Li, Wei Chen, Yajun Wang, and Zhi-Li Zhang. Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In Proceedings of the sixth ACM international conference on Web Search and Data Mining, pages 657–666. ACM, 2013.

[106] Zhiyuan Lin, Niloufar Salehi, Bowen Yao, Yiqi Chen, and Michael S Bernstein. Better when it was smaller? community content and behavior after massive growth. In Proceedings of ICWSM, pages 132–141. AAAI, 2017.

[107] Rongfang Liu, Ram Pendyala, and Steven Polzin. Assessment of intermodal transfer penalties using stated preference data. Transportation Research Record: Journal of the Transportation Research Board, (1607):74–80, 1997.

[108] Sharon Lockyer and Michael Pickering. Beyond a joke: The limits of humour. Springer, 2005.

[109] Wei-Yin Loh. Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1):14–23, 2011.

[110] Andrés Lucero, Danielle Wilde, Simon Robinson, Joel E Fischer, James Clawson, and Oscar Tomico. Mobile collocated interactions with wearables. In Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, pages 1138–1141. ACM, 2015.

[111] Martin Ludvigsen and Rune Veerasawmy. Designing technology for active spectator experiences at sporting events. In Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction, pages 96–103. ACM, 2010.

[112] Henry B Mann and Donald R Whitney. On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics, pages 50–60, 1947.

[113] Leon Mann. Sports crowds and the collective behavior perspective. Sports, games, and play: Social and psychological viewpoints, pages 299–331, 1989.

[114] Christopher D Manning, Christopher D Manning, and Hinrich Schütze. Foundations of statistical natural language processing. MIT press, 1999.

[115] Dejan Markovikj, Sonja Gievska, Michal Kosinski, and David J Stillwell. Mining facebook data for predictive personality modeling. In Proceedings of the International AAAI Conference on Weblogs and Social Media, 2013.

[116] Lilliana Mason. Uncivil agreement: How politics became our identity. University of Chicago Press, 2018.

[117] Judith Masthoff. Group modeling: Selecting a sequence of television items to suit a group of viewers. In Personalized digital television, pages 93–141. Springer, 2004.

[118] J Nathan Matias. Going dark: Social factors in collective action against platform operators in the reddit blackout. In Proceedings of CHI, pages 1138–1151. ACM, 2016.

[119] J Nathan Matias. Preventing harassment and increasing group participation through social norms in 2,190 online science discussions. Proceedings of the National Academy of Sciences, 116(20):9785–9789, 2019.

[120] Microworkers. Microworkers. `https://ttv.microworkers.com/`, 2017.

[121] Hannah Miller, Daniel Kluver, Jacob Thebault-Spieker, Loren Terveen, and Brent Hecht. Understanding emoji ambiguity in context: The role of text in emoji-related miscommunication. In Proceedings of the International AAAI Conference on Web and Social Media, 2017.

[122] Atsunori Minamikawa and Hiroyuki Yokoyama. Blog tells what kind of personality you have: egogram estimation from japanese weblog. In Proceedings of the ACM 2011 Conference on Computer-Supported Cooperative Work, pages 217–220. ACM, 2011.

[123] Atsunori Minamikawa and Hiroyuki Yokoyama. Personality estimation based on weblog text classification. In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, pages 89–97. Springer, 2011.

[124] Shalini Misra, Lulu Cheng, Jamie Genevie, and Miao Yuan. The iphone effect: the quality of in-person social interactions in the presence of mobile devices. Environment and Behavior, 48(2):275–298, 2016.

[125] MongoDB. Mongodb. `https://www.mongodb.com/`, 2017.

[126] Burt L Monroe, Michael P Colaresi, and Kevin M Quinn. Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict. Political Analysis, 16 (4):372–403, 2008.

[127] Robert T Muller. Covid-19 brings a pandemic of conspiracy theories. https://www.psychologytoday.com/us/blog/talking-about-trauma/202004/covid-19-brings-pandemic-conspiracy-theories, 2020. [Online; accessed 1-June-2020].

[128] Albert M Muniz and Thomas C O'guinn. Brand community. Journal of Consumer Research, 27(4):412–432, 2001.

[129] Mor Naaman, Jeffrey Boase, and Chih-Hui Lai. Is it really about me?: message content in social awareness streams. In Proceedings of the ACM Conference on Computer-Supported Cooperative Work, pages 189–192. ACM, 2010.

[130] NBA.com. NBA breaks all-time attendance record for fourth straight year. http://www.nba.com/article/2018/04/12/nba-breaks-attendance-record-fourth-straight-year, 2018. [Online; accessed 15-June-2019].

[131] Edward Newell, David Jurgens, Haji Mohammad Saleem, Hardik Vala, Jad Sassine, Caitrin Armstrong, and Derek Ruths. User migration in online social networks: A case study on reddit during a period of community unrest. In Proceedings of ICWSM, pages 279–288. AAAI, 2016.

[132] Nielsen.com. The year in sports media report: 2015. https://www.nielsen.com/us/en/insights/reports/2016/the-year-in-sports-media-report-2015.html, 2015. [Online; accessed 15-June-2019].

[133] Brendan Nyhan and Jason Reifler. When corrections fail: The persistence of political misperceptions. Political Behavior, 32(2):303–330, 2010.

[134] Erik L Olson. Are rival team fans a curse for home team sponsors? the moderating effects of fit, oppositional loyalty, and league sponsoring. Marketing Letters, 29(1):115–122, 2018.

[135] Elizabeth Page-Gould, Rodolfo Mendoza-Denton, and Linda R Tropp. With a little help from my cross-group friend: Reducing anxiety in intergroup contexts through cross-group friendship. Journal of Personality and Social Psychology, 95(5):1080, 2008.

[136] Leysia Palen and Kenneth M Anderson. Crisis informatics: New data for extraordinary times. Science, 353(6296):224–225, 2016.

[137] Stefania Paolini, Jake Harwood, and Mark Rubin. Negative intergroup contact makes group memberships salient: Explaining why intergroup conflict endures. Personality and Social Psychology Bulletin, 36(12):1723–1738, 2010.

[138] Gregory Park, H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Michal Kosinski, David J Stillwell, Lyle H Ungar, and Martin EP Seligman. Automatic personality assessment through social media language. Journal of Personality and Social Psychology, 108 (6):934, 2015.

[139] Moon-Hee Park, Han-Saem Park, and Sung-Bae Cho. Restaurant recommendation for group of people in mobile environments using probabilistic multi-criteria decision making. In Asia-Pacific Conference on Computer Human Interaction, pages 114–122. Springer, 2008.

[140] David Paton and Andrew Cooke. Attendance at county cricket: An economic analysis. Journal of Sports Economics, 6(1):24–45, 2005.

[141] Warren Pearce, Kim Holmberg, Iina Hellsten, and Brigitte Nerlich. Climate change on twitter: Topics, communities and conversations about the 2013 ipcc working group 1 report. PloS one, 9(4):e94785, 2014.

[142] James W Pennebaker, Martha E Francis, and Roger J Booth. Linguistic inquiry and word count: Liwc 2007. Technical report, 2007.

[143] Thomas F Pettigrew. Generalized intergroup contact effects on prejudice. Personality and Social Psychology Bulletin, 23(2):173–185, 1997.

[144] Thomas F Pettigrew and Linda R Tropp. A meta-analytic test of intergroup contact theory. Journal of Personality and Social Psychology, 90(5):751, 2006.

[145] Thomas F Pettigrew and Linda R Tropp. How does intergroup contact reduce prejudice? meta-analytic tests of three mediators. European Journal of Social Psychology, 38(6):922–934, 2008.

[146] Thomas F Pettigrew, Linda R Tropp, Ulrich Wagner, and Oliver Christ. Recent advances in intergroup contact theory. International Journal of Intercultural Relations, 35(3):271–280, 2011.

[147] Daniele Quercia, Renaud Lambiotte, David Stillwell, Michal Kosinski, and Jon Crowcroft. The personality of popular facebook users. In Proceedings of the ACM Conference on Computer-Supported Cooperative Work, pages 955–964. ACM, 2012.

[148] Elisa Quintarelli, Emanuele Rabosio, and Letizia Tanca. Recommending new items to ephemeral groups using contextual user influence. In Proceedings of RecSys, pages 285–292, 2016.

[149] Iyad Rahwan, Manuel Cebrian, Nick Obradovich, Josh Bongard, Jean-François Bonnefon, Cynthia Breazeal, Jacob W Crandall, Nicholas A Christakis, Iain D Couzin, Matthew O Jackson, et al. Machine behaviour. Nature, 568(7753):477–486, 2019.

[150] Juan Ramos et al. Using tf-idf to determine word relevance in document queries. In Proceedings of the first instructional conference on machine learning, volume 242, pages 133–142, 2003.

[151] Max Rappaport. The definitive history of 'trust the process'. `http://bleacherreport.com/articles/2729018-the-definitive-history-of-trust-the-process`, 2018. [Online; accessed 15-April-2018].

[152] Reddit. Top reddits. `http://redditmetrics.com/top`, 2018. [Online; accessed 15-June-2019].

[153] Redditlist. Most active subreddits. `http://redditlist.com/all`, 2018. [Online; accessed 15-June-2019].

[154] Yuqing Ren, Robert Kraut, and Sara Kiesler. Applying Common Identity and Bond Theory to Design of Online Communities. Organization Studies, 28(3):377–408, 2007.

[155] Alan Roadburg. Factors precipitating fan violence: A comparison of professional soccer in britain and north america. British Journal of Sociology, pages 265–276, 1980.

[156] Daniel M Romero, Brian Uzzi, and Jon Kleinberg. Social networks under stress. In Proceedings of WWW, pages 9–20. International World Wide Web Conferences Steering Committee, 2016.

[157] Daniel M. Romero, Katharina Reinecke, and Lionel P. Robert, Jr. The influence of early respondents: Information cascade effects in online event scheduling. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, WSDM '17, pages 101–110, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4675-7. doi: 10.1145/3018661.3018725. URL http://doi.acm.org/10.1145/3018661.3018725.

[158] Tony Romm and Dwoskin Elizabeth. Jack dorsey says he's rethinking the core of how twitter works. https://www.washingtonpost.com/technology/2018/08/15/jack-dorsey-says-hes-rethinking-core-how-twitter-works/?utm_term=.ca3c42463c6f, 2018. [Online; accessed 15-June-2019].

[159] Matthew Rothstein. For aging veterans, one last shot at an nba title is all they can think about. https://uproxx.com/dimemag/ring-chasing-richard-jefferson-shane-battier/, 2017. [Online; accessed 15-April-2018].

[160] Haji Mohammad Saleem, Kelly P Dillon, Susan Benesch, and Derek Ruths. A web of hate: Tackling hateful speech in online social spaces. In First Workshop on Text Analytics for Cybersecurity and Online Safety (TA-COS 2016), 2017.

[161] Gerard Salton and Michael J McGill. Introduction to Modern Information Retrieval. McGraw-Hill, Inc., 1986.

[162] H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. Personality, gender, and age in the language of social media: The open-vocabulary approach. PloS one, 8(9):e73791, 2013.

[163] Joseph Seering, Robert Kraut, and Laura Dabbish. Shaping pro and anti-social behavior on Twitch through moderation and example-setting. In Proceedings of Computer-Supported Cooperative Work and Social Computing (CSCW) '17, New York, NY, USA, 2017. ACM. URL http://doi.acm.org/10.1145/2998181.2998277.

[164] Matthew D Shank and Mark R Lyberger. Sports marketing: A strategic perspective. Routledge, 2014.

[165] Limor Shifman. Memes in digital culture. Mit Press, 2014.

[166] Nate Silver and Reuben Fischer-Baum. How we calculate nba elo ratings. https://fivethirtyeight.com/features/how-we-calculate-nba-elo-ratings/, 2018. [Online; accessed 15-April-2018].

[167] Eric Simons. The psychology of why sports fans see their teams as extensions of themselves. https://www.washingtonpost.com/opinions/

the-psychology-of-why-sports-fans-see-their-teams-as-extensions-of-themselves/2015/01/30/521e0464-a816-11e4-a06b-9df2002b86a0_story.html, 2015. [Online; accessed 15-June-2019].

[168] Aaron CT Smith and Bob Stewart. The travelling fan: Understanding the mechanisms of sport fan consumption in a sport tourism setting. Journal of sport & tourism, 12(3-4):155–181, 2007.

[169] Marc A Smith, Lee Rainie, Ben Shneiderman, and Itai Himelboim. Mapping twitter topic networks: From polarized crowds to community clusters. Pew Research Center, 20:1–56, 2014.

[170] Spring Roo. Spring roo framework. https://docs.spring.io/spring-roo/reference/html/intro.html, 2017.

[171] Kate Starbird, Leysia Palen, Amanda L Hughes, and Sarah Vieweg. Chatter on the red: what hazards threat reveals about the social life of microblogged information. In Proceedings of CSCW, pages 241–250. ACM, 2010.

[172] Tobias H Stark, Andreas Flache, and René Veenstra. Generalization of positive and negative attitudes toward individuals to outgroup attitudes. Personality and Social Psychology Bulletin, 39(5):608–622, 2013.

[173] Statista. Average tv viewership of nba finals games in the united states from 2002 to 2017 (in millions). https://www.statista.com/statistics/240377/nba-finals-tv-viewership-in-the-united-states/, 2018. [Online; accessed 01-November-2018].

[174] Leon Stenneth, Ouri Wolfson, Philip S Yu, and Bo Xu. Transportation mode detection using mobile phones and gis information. In Proceedings of SIGSPATIAL, pages 54–63, 2011.

[175] Walter G Stephan and Krystina Finlay. The role of empathy in improving intergroup relations. Journal of Social issues, 55(4):729–743, 1999.

[176] Shawn Stevens and Philip J Rosenberger. The influence of involvement, following sport and fan identification on fan loyalty: An australian perspective. International Journal of Sports Marketing and Sponsorship, 13(3):57–71, 2012.

[177] Leo Graiden Stewart, Ahmer Arif, A. Conrad Nied, Emma S. Spiro, and Kate Starbird. Drawing the lines of contention: Networked frame contests within #blacklivesmatter discourse. Proc. ACM Hum.-Comput. Interact., 1(CSCW):96:1–96:23, December 2017. ISSN 2573-0142.

[178] Elizabeth A Stuart. Matching methods for causal inference: A review and a look forward. Statistical science: a review journal of the Institute of Mathematical Statistics, 25(1):1, 2010.

[179] Cass R Sunstein. Going to extremes: How like minds unite and divide. Oxford University Press, 2009.

[180] William A Sutton, Mark A McDonald, George R Milne, and John Cimperman. Creating and fostering fan identification in professional sports. Sport Marketing Quarterly, 6:15–22, 1997.

[181] Henri Tajfel. Social psychology of intergroup relations. Annual Review of Psychology, 33(1): 1–39, 1982.

[182] Chenhao Tan. Tracing Community Genealogy: How New Communities Emerge from the Old. In Proceedings of ICWSM, pages 395–404. AAAI, 2018.

[183] Chenhao Tan and Lillian Lee. All who wander: On the prevalence and characteristics of multi-community engagement. In Proceedings of the 24th International Conference on World Wide Web, pages 1056–1066. International World Wide Web Conferences Steering Committee, 2015.

[184] Chenhao Tan, Lillian Lee, and Bo Pang. The effect of wording on message propagation: Topic- and author-controlled natural experiments on twitter. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, volume 1, pages 175–185, 2014.

[185] Chenhao Tan, Adrien Friggeri, and Lada Adamic. Lost in propagation? unfolding news cycles from the source. In Proceedings of ICWSM, 2016.

[186] Thomas A Timmerman. Racial diversity, age diversity, interdependence, and team performance. Small Group Research, 31(5):592–606, 2000.

[187] Catalina L Toma and Jeffrey T Hancock. Reading between the lines: linguistic cues to deception in online dating profiles. In Proceedings of the ACM Conference on Computer-Supported Cooperative Work, pages 5–8. ACM, 2010.

[188] Hanghang Tong, Christos Faloutsos, and Jia-yu Pan. Fast random walk with restart and its applications. In Proceedings of ICDM, pages 613–622, 2006.

[189] Jessica Torrez Riley. A look at spectator technology: location-based services and mobile habits of collegiate sports fans. In Proceedings of MobileHCI Companion, pages 41–46. ACM, 2012.

[190] Galen T Trail and Jeffrey D James. The motivation scale for sport consumption: Assessment of the scale's psychometric properties. Journal of sport behavior, 24(1):108, 2001.

[191] Rhiannon N Turner and Richard J Crisp. Imagining intergroup contact reduces implicit prejudice. British Journal of Social Psychology, 49(1):129–142, 2010.

[192] Marlon Twyman, Brian C Keegan, and Aaron Shaw. Black lives matter in wikipedia: Collective memory and collaboration around online social movements. In Proceedings of CSCW, pages 1400–1412. ACM, 2017.

[193] Nicholas Vincent, Isaac Johnson, and Brent Hecht. Examining wikipedia with a broader lens: Quantifying the value of wikipedia's relationships with other large-scale online communities. In Proceedings of CHI, page 566. ACM, 2018.

[194] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. Science, 359(6380):1146–1151, 2018.

[195] Hanna M Wallach, Iain Murray, Ruslan Salakhutdinov, and David Mimno. Evaluation methods for topic models. In Proceedings of ICML, pages 1105–1112, 2009.

[196] Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, and Amit P Sheth. Cursing in english on twitter. In Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing, pages 415–425. ACM, 2014.

[197] Daniel L Wann and Nyla R Branscombe. Die-hard and fair-weather fans: Effects of identification on birging and corfing tendencies. Journal of Sport and Social issues, 14(2):103–117, 1990.

[198] Daniel L Wann, Joel L Royalty, and Al R Rochelle. Using motivation and team identfication to predict sport fans' emotional responses to team performance. Journal of Sport Behavior, 25(2):207, 2002.

[199] Daniel L Wann, Frederick G Grieve, Ryan K Zapalac, and Dale G Pease. Motivational profiles of sport fans of different sports. Sport Marketing Quarterly, 17(1):6, 2008.

[200] weathersource. Weather source api. `https://weathersource.com/products/onpoint-api/`, 2017.

[201] Lawrence A Wenner. Media, sports, and society. sage, 1989.

[202] Fiona A White, Hisham M Abu-Rayya, Ana-Maria Bliuc, and Nicholas Faulkner. Emotion expression and intergroup bias reduction between muslims and christians: Long-term internet contact. Computers in Human Behavior, 53:435–442, 2015.

[203] Wikipedia. /r/IAmA. `https://en.wikipedia.org/wiki//r/IAmA`, 2018. [Online; accessed 15-April-2018].

[204] Wikipedia. NBA All-Star Game — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/w/index.php?title=NBA%20All-Star%20Game&oldid=839843474`, 2018. [Online; accessed 8-July-2018].

[205] Wikipedia. Super bowl commercials. `https://en.wikipedia.org/wiki/Super_Bowl_commercials`, 2018. [Online; accessed 15-April-2018].

[206] Wikipedia. List of national basketball association rivalries. `https://en.wikipedia.org/wiki/List_of_National_Basketball_Association_rivalries`, 2018. [Online; accessed 15-April-2018].

[207] Matthew Wolak. Package 'icc', facilitating estimation of the intraclass correlation coefficient. `https://cran.r-project.org/web/packages/ICC/ICC.pdf`, 2018. [Online; accessed 17-August-2018].

[208] Thomas Wood and Ethan Porter. The elusive backfire effect: Mass attitudes' steadfast factual adherence. Political Behavior, pages 1–29, 2016.

[209] Stephen C Wright, Arthur Aron, Tracy McLaughlin-Volpe, and Stacy A Ropp. The extended contact effect: Knowledge of cross-group friendships and prejudice. Journal of Personality and Social Psychology, 73(1):73, 1997.

[210] Robert Wu, Lora Appel, Dante Morra, Vivian Lo, Simon Kitto, and Sherman Quan. Short message service or disservice: issues with text messaging in a complex medical environment. International Journal of Medical Informatics, 83(4):278–284, 2014.

[211] Yang Ye, Yu Zheng, Yukun Chen, Jianhua Feng, and Xing Xie. Mining individual life pattern based on location history. In Mobile Data Management: Systems, Services and Middleware, 2009. MDM'09. Tenth International Conference on, pages 1–10. IEEE, 2009.

[212] InJeong Yoon. Why is it not just a joke? analysis of internet memes associated with racism and hidden ideology of colorblindness. Journal of Cultural Research in Art Education, 33, 2016.

[213] Masayuki Yoshida, Bob Heere, and Brian Gordon. Predicting behavioral loyalty through community: Why other fans are more important than our own intentions, our satisfaction, and the team itself. Journal of Sport Management, 29(3):318–333, 2015.

[214] Ark Fangzhou Zhang, Danielle Livneh, Ceren Budak, Lionel P Robert Jr, and Daniel M Romero. Shocking the crowd: The effect of censorship shocks on chinese wikipedia. In Proceedings of ICWSM, pages 367–376. AAAI, 2017.

[215] Jaosn Shuo Zhang and Qin Lv. Event Organization 101: Understanding Latent Factors of Event Popularity. In The International AAAI Conference on Web and Social Media, pages 716–719. AAAI, 2017.

[216] Jason Shuo Zhang and Qin Lv. Hybrid egu-based group event participation prediction in event-based social networks. Knowledge-Based Systems, 143:19–29, 2018.

[217] Jason Shuo Zhang and Qin Lv. Understanding event organization at scale in event-based social networks. ACM Trans. Intell. Syst. Technol., 10(2):16:1–16:23, January 2019. ISSN 2157-6904.

[218] Jason Shuo Zhang, Khaled Alanezi, Mike Gartrell, Richard Han, Qin Lv, and Shivakant Mishra. Understanding group event scheduling via the outwithfriendz mobile application. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1 (4):175, 2018.

[219] Jason Shuo Zhang, Chenhao Tan, and Qin Lv. "This is why we play": Characterizing Online Fan Communities of the NBA Teams. Proc. ACM Hum.-Comput. Interact., 2(CSCW), 2018.

[220] Jason Shuo Zhang, Brian C Keegan, Qin Lv, and Chenhao Tan. A tale of two communities: Characterizing reddit response to covid-19 through/r/china_flu and/r/coronavirus. arXiv preprint arXiv:2006.04816, 2020.

[221] Justine Zhang, William L Hamilton, Cristian Danescu-Niculescu-Mizil, Dan Jurafsky, and Jure Leskovec. Community identity and user engagement in a multi-community landscape. In Proceedings of ICWSM, pages 377–386. AAAI, 2017.

[222] Shuo Zhang and Qin Lv. Event organization 101: Understanding latent factors of event popularity. In International AAAI Conference on Web and Social Media, pages 716–719. AAAI, 2017.

[223] Yu Zheng, Like Liu, Longhao Wang, and Xing Xie. Learning transportation mode from raw gps data for geographic applications on the web. In Proceedings of WWW, pages 247–256, 2008.

[224] Haiyi Zhu, Jilin Chen, Tara Matthews, Aditya Pal, Hernan Badenes, and Robert E Kraut. Selecting an Effective Niche: An Ecological View of the Success of Online Communities. In Proceedings of CHI, pages 301–310. ACM, 2014.