A Computational Analysis on the Role of Social Relationships in Online Communication and Information Diffusion

by

Minje Choi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in the University of Michigan
2023

Doctoral Committee:

Assistant Professor David Jurgens, Co-Chair
Associate Professor Daniel M. Romero, Co-Chair
Assistant Professor Lu Wang
Assistant Professor Justine Zhang

Minje Choi
minje@umich.edu
ORCID iD: 0009-0003-1125-9894

ACKNOWLEDGEMENTS

The focus of my Ph.D. studies has been on understanding how various social relationships are formed and maintained, shaping our lives and society overall. Upon reflection on the past five years of my Ph.D., I realize that the social relationships I built and maintained during the process have been immensely helpful in overcoming hardships and being able to pursue my degree. I would like to dedicate this section to mentioning those to whom I would like to express my sincere gratitude.

First and foremost, I would like to thank my two wonderful advisors David Jurgens and Daniel Romero, for being the most supportive in every way I could possibly think of. I was able to obtain so much high- and low-level knowledge on how to do meaningful and truthful research in the domain of computational social sciences and natural language processing, and greatly appreciate their patience in waiting for me to learn the process and be able to further develop. I am so much grateful for their honesty and sincerity during our conversations and how they were always supportive of the decisions I made throughout my Ph.D., both in and out of research. I am eager to push myself further after graduation and try hard to match the high level of academic standards that were practiced by both of them.

I also would like to mention other faculty and staff members who have helped me throughout this journey. I thank Ceren Budak for her thoughtful advice during the completion of my first project. I also thank professors Abigail Jacobs, Yesim Orhun, and Yan Chen, who have provided me with high-quality advice as committee members during my doctoral milestones. I thank professors Lu Wang and Justine Zhang, both of whom I am glad to have on my doctoral dissertation committee, and appreciate the thoughts and suggestions they have given so far. Finally, I thank Jinseok Kim for his honest and encouraging advice during our occasional dinner talks.

I am honored to have been part of the UMSI community, where I have met many great friends and mentors. First I would like to thank everyone in my cohort, in particular Zhuofeng Wu, Deahan Yu, Jane Im, Woosuk Seo, and all of whom I consider my support group in helping me navigate through the early stages of PhD life. I also thank Julia Mendelsohn, Soyoung Lee, Jiazhao Li, Joshua Ashkinaze, Siqi Wu, Anmol Panda, Ashwin Rajadesingan, Sam Carton, Lia Bozarth, Xinyan Zhao, Sungjin Nam, Sagar Kumar, Patrick Park, and other members of UMSI for providing me with a sense of community membership and belongingness.

I had the privilege of belonging to not one but two groups consisting of intellectually motivated and supportive students. My thanks go to all our current and past lab members: Hao Peng, Ed

Platt, Danaja Maldeniya, Chris Quarles, Aparna Ananthasubramaniam, Yulin Yu, David Gamba, Stacey Xiang, Kenan Alkiek, Hong Chen, Lavinia Dunagan, Ben Litterer, Jiaxin Pei, Leo Raabe, Agrima Seth, and Jason Yan. I have learned so much from them throughout all the discussions and meetings we had.

During my Ph.D., I was lucky enough to do internships at a number of different places and meet several exciting teams, which helped me broaden my perspective immensely. I thank everyone at the Social Dynamics Team at Nokia Bell Labs for providing me with a great internship experience, especially to Luca Maria Aiello who was a wonderful mentor both in research and in person. I thank the Risk Modeling Team at Visa Research for the opportunity to engage in industry-aligned ML research, especially my mentors Javid Ebrahimi and Xin Dai who both have been very helpful throughout the process. I would also like to thank the Computational Social Science Team at Snap Research and loved the time there doing cutting-edge research on social networks, shouting out, especially to my mentor Francesco Barbieri who was always there to help. Needless to say, I thank all the interns I was able to meet throughout the internships and do hope our paths cross together at future conferences or research opportunities.

I would like to thank all of my friends in Michigan who will be part of my fond memories during my time at Ann Arbor. I was also proud to be part of the organizing committee for the Korean Student Association - Graduate, and particularly thank Minhyung Ahn, whom I will always remember as a caring leader. I would also like to mention Alex Hongrak Kim and his family for their warm welcome and ongoing support during the last five years, and also Samuel Park for his kind words and wisdom. I thank Youngsun Wi for the unwavering support during the majority of my Ph.D. which helped me overcome numerous obstacles and remain positive. I also thank Sungbok Shin, Jaehoon Lee, Younwoo Ko, Jisoo Lim, Jinhyuk Lee, Myungho Lee, Sanghyuk Yoon, and all of my dear friends outside of Michigan whose presence helped me overcome struggles and maintain faith throughout the journey.

Last but not least, I thank my brother Min-hyuk and my parents for always being at my side and for rooting me at all times.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

ABSTRACT

Social relationships play a crucial role in shaping daily conversations and information sharing within social networks, both in person and on online platforms like Twitter and Facebook. These platforms have become immensely popular for accessing a wide range of information. While previous studies have contributed to understanding the properties of social ties, less attention has been devoted to directly identifying the characteristics of individual social relationships and their influence on dyadic interactions in online social networks.

In this dissertation I present three computational studies to identify and analyze the key characteristics of social relationships within large online social networks. These studies seeks to shed light on how social relationships impact interactions and information diffusion. The first study approaches relationships through the lens of social dimensions, such as conflict or trust, wherein a dyad exhibits varying levels of dimension strength. The findings indicate that the strength of inferred dimensions accurately represents the nature of social relationships in Twitter ties. Additionally, these inferred dimensions can reflect community-level outcomes, such as the stability of organizations or well-being indices. The second study proposes a novel method for identifying different types of interpersonal relationships using a combination of text- and network-based features. Linguistic and diurnal communication patterns are found to differ significantly among various types of relationships, and that it is possible to build accurate classifier models for inferring categories of social relationships based on communication on Twitter. Moreover, incorporating information about these relationships enhances the performance of retweet prediction models. Building upon the relationship classification model developed in Study 2, the third and final study investigates the responses of dyads of users facing unexpected life-shock events. Interestingly, the research uncovers relationship-specific reactions to different types of shocks, providing valuable insights into how social ties are influenced during challenging times. The findings from the three computational studies provide a comprehensive understanding of the dynamics of social relationships in the digital age.

CHAPTER 1

Introduction

1.1 Motivation

Humans have always been social beings by nature. Throughout history, humans have demonstrated their innate social nature by adapting to various environments and overcoming obstacles through interaction within larger social networks [Braun and Plog, 1982]. Association with a social network structure provides numerous benefits, including risk reduction through extended access to resources [Rautman, 1993], faster communication channels for information sharing [Yang and Counts, 2010], and the provision of emotional support and stability [Wellman and Wortley, 1990]. Evidence of strategic social networks can be observed as far back as the Paleolithic ages [Sikora et al., 2017], suggesting a long history of social network formation that has been crucial to the survival and flourishing of civilizations, societies, and cultures.

One distinctive characteristic of social networks, as opposed to other forms of networks, is that actors establish connections in the form of *social relationships* at both dyadic and group levels. Social relationships are determined and defined by social [Massen et al., 2010], biological [Golombok et al., 1999], and psychological [Baumeister and Leary, 1995] factors, and individuals form various relationships with others at different stages in their lives. As a result, the ties formed within social networks exist in relationships of various forms. Relationships play a key role in understanding human behavior in a social context, as the levels of emotions or content exchanged in daily interactions are largely expected to differ by the type of relationship [Bradac, 1983].

In recent decades, advances in communication technology have reshaped the sizes and structures of social networks at unprecedented rates. Most notably, social media platforms such as Twitter and Facebook have had a major impact on how social networks are managed and maintained. These platforms enable users to keep in touch and share information with their social network contacts through communication means such as text, images, and videos. Through various recommendation algorithms, users can also expand their network sizes by befriending or following other user accounts that share similar interests or demographics, whether it be a friend's friend or a total stranger online.

At a large scale, such interactions contribute to the spread of public opinions [Gorodnichenko et al., 2021], political campaigns [Ramaciotti Morales and Cointet, 2021], collective emotions [Kramer et al., 2014], shaping our society through the process. This leads to interesting potential research directions on how social relationships in online platforms play a role in shaping the dissemination of content and information in online social networks.

Social networking platforms contain characteristics of both social and informational networks [Arnaboldi et al., 2013, Kwak et al., 2010], making them an efficient means for forming and maintaining ties with others for various purposes. To gain a clearer understanding of how information diffusion occurs in these social networks via interpersonal communication and information-sharing behaviors, several studies have focused on changes caused by properties of dyads. Studies in this direction have led to valuable findings, for instance, dyads (i) have a greater tendency to share less common hashtags which may indicate community belonging [Romero et al., 2013b], (ii) are more likely to share content than with strangers [Quercia et al., 2012, Bakshy et al., 2009], and (iii) frequently engage through actions such as mentioning each other or sharing posts [Jones et al., 2013]. Other studies have looked into user interactions with a focus on influential users such as celebrities or politicians who possess a large follower base. These users gain widespread influence by specializing in narrow topics [Cha et al., 2010] or posting messages with strong sentiments [Dang-Xuan et al., 2013]. Yet, it still remains largely unknown how this dyadic behavior of interactions and information diffusion is governed by the social relationship itself. In other words, a large volume of existing research focuses on explaining network interactions through *network properties*, shedding less light on how the *social* factors connected to social relationships play a role in network behavior.

Being able to understand the role that social relationships play in online social networks can enrich our understanding on multiple domains of studies, leading to a vast array of potential contributions. In the domain of network science where it is important to understand how information diffuses within networks, the additional knowledge of social relationships and their properties can help create better models for simulating the flow of information, as factors such as the speed or likelihood of sharing information may vary by relationship type [Haas and Sherman, 1982, Keijsers et al., 2010]. This topic is of importance in the social psychology and mental health domain as well, as online social networks have proven to be effective means of social support through support networks [Huh et al., 2016]. As the type of support that one can provide differs by social relationship type [Wellman and Wortley, 1990], correctly modeling the expected levels and types of social support in online social networks can contribute to providing more accurate and timely support measures for distressed individuals via their online support networks. Finally, identifying different social relationships in online social networks can also benefit our understanding of how societies change over time. Being able to identify properties of social relationships opens up questions such as which types of social exchange are prevalent in current versus past societies, and how this is reflected in the composition of identifiable social relationships from online social network interactions. Simply put, the added knowledge of social relationships can lead to several research opportunities across a large range of social science domains.

1.2 Challenges

Understanding the effects of social relationships on the diffusion of online social networks, however, is associated with several practical challenges. The first and foremost issue is that it is difficult to obtain ground-truth information on the relationship types of dyads in online social networks such as Twitter. Most platforms do not explicitly require users to provide information on the relationship type when establishing a connection. One exception would be Facebook where users may provide relationship information such as "married to" or "father of" [Backstrom and Kleinberg, 2014], but this feature does not generalize to other platforms and thus cannot be used in those such as Twitter, which is a major platform where interpersonal relationships exist. While studies such as Tay et al. [2018] use linguistic heuristics to identify romantic ties from Twitter, this method is not comprehensive across various types of relationships. For online platforms that promote anonymity instead of tie formation such as Reddit, it is even harder to obtain meaningful information of the relationship between two users who are interacting through conversations, where an alternative would be to instead identify properties of relationship attributes exchanged through conversations. However, this direction has not been explored thoroughly as of today.

Another challenge is the discrepancy between online and offline relationships. Online platforms such as Twitter and Facebook are actively being used as platforms for growing and maintaining social networks. While there is substantial variance by platform type, several studies suggest a partial overlap between the contacts in social networking sites and offline networks [Subrahmanyam et al., 2008, Reich et al., 2012, Ozenc and Farnham, 2011, Farnham and Churchill, 2011], indicating that usually a substantial amount of one's offline social ties also exist within online social networks. Interactions made in online social networks can strengthen the perceived relationship even outside of the platform [Subrahmanyam et al., 2008, Reich et al., 2012], and so one can assume that the interactions and communication made in online social networks are also important for maintaining social relationships.

However, this should not lead to the conclusion that the interactions among relationship types would be identical between online and offline environments. One reason would be platform differences which exist for maintaining different aspects of social relationships. The decision of choosing which platform to communicate with whom can differ due to factors such as feature availability or generational popularity [Nouwens et al., 2017, Archambault and Grudin, 2012a]. Function wise, platforms that enable one-on-one conversations help strengthen existing social ties [Kim et al., 2007]. In contrast, communication on platforms such as Twitter or Facebook are broadcasted to one's entire follower network by default, differing greatly from one-on-one conversations. This creates a sense of "context collapse" [Marwick and Boyd, 2011b], where information is visible and exposed to everyone. Such public messages may restrict the content one would want to deliver to the receiver, especially to close relationships. Context collapse can be both intentional and unintentional, leading to desired or undesired effects [Davis and Jurgenson, 2014].

The broadcasting nature of online social networks also leads to the formation of new types of relationships that are often unique to the platform. One particular case would be parasocial relationships [Hoffner, 2008]. These generally consist of follower-followee relationships which are formed by following celebrities or media characters who are active on the platforms [Marwick and Boyd, 2011a]. Often the communication structure is one-sided, where the followees construct their posts and messages as a form of impression management intended to exert social presence at scale [Kim and Song, 2016, Chen, 2016].

The final challenge is the nature of social relationships, which can be dynamic and multi-faceted. Theoretic models such as social penetration theory [Altman and Taylor, 1973] or uncertainty reduction theory [Berger and Calabrese, 1974] suggest that relationships can develop through frequent interactions and information disclosure, sometimes introducing additional facets into the relationship such as friendships turning into romantic relationships [Guerrero and Mongeau, 2008]. Likewise, lack of interaction or rise of conflicts can revert intimate relationships back to their original state, or even terminate the relationship altogether. Therefore, studies on interpersonal relationships should be aware of measuring the perceived relationship at the time. This could lead to challenges when attempting to pinpoint the relationship between two users given a specific timeframe.

1.3 Proposed Approach

This leaves a largely unexplored space for studying social relationships in online social networks, which is the topic of this dissertation. In particular, I will focus on tackling the first challenge by devising computational methods for inferring properties of social relationships from the interaction data in online social networks. The development of such models allows one to infer labels of social relationships to the dyads in a social network at an unprecedented scale, enabling large-scale analyses of interactions in online social networks. Combined with large datasets of social networks such as Reddit and Twitter, I will answer research questions such as the effect of social relationship type on responses to events such as shocks, or whether the prevalence of certain social relationships is indicative of desirable or undesirable statuses of our society.

1.4 Contributions of Studies

I propose three studies that aim to understand different aspects of social relationships in online social networks.

1.4.1 Social dimensions extracted from dyadic conversations can increase the understanding of our society

In the first study, I focus on measuring the strengths of social dimensions from conversations that are representative of dyadic interactions. Based on the framework of Deri et al. [2018] which proposes ten types of social dimensions for describing human relationships, I train finetuned RoBERTa classifiers using a dataset of conversations comprising Reddit, movie scripts, and email threads which are labeled by crowdsourced workers. I show that all ten social dimensions can be predicted purely from *conversations*, and that the combination of the predicted dimensions suggests both the types of *relationships* people entertain and the types of real-world *communities* they shape. Further, by applying the classifiers for measuring the dimensions of 160M messages written by geo-referenced Reddit users, 290k emails from the Enron corpus, and 300k lines of dialogue from movie scripts, I show that the presence of the ten dimensions in the language is indicative of the types of communities people shape. For example, some of the dimensions are predictive of societal outcomes in US States, such as education, wealth, and suicide rates.

1.4.2 Interpersonal communication in Twitter conditioned on social relationships

In the second study, I aim to provide labels to dyadic interactions from a different perspective by assigning interpersonal category labels instead of measuring dimension strength. Individuals in these networks are largely organized around social structures such as work, neighborhood, or families [Feld, 1981, 1982], forming *interpersonal relationships*, such as friendships, kinship, and romantic partnerships. These interpersonal relationship types can influence communication and behavior in the network. To study the role of interpersonal relationships in shaping communication and interactions in online social networks, I first collect a massive dataset of interactions between 9.6 million Twitter user dyads with labeled relationships which are obtained using self-declared relationship phrases. I then conduct an extensive analysis of linguistic, topical, network, diurnal characteristics, and social dimensions from the previous study, across dyads of the identified relationship categories. Next, I introduce a neural network model for classifying five relationship types from linguistic and network features, achieving an F1 of 0.70, which substantially improves upon a strong classifier baseline (0.55) and random guess (0.20). Finally, I demonstrate the potential of relationship information for enhancing the ability to predict information diffusion, where I show that compared to a baseline model, adding relationship types for a retweet prediction task can improve the F1 score by 1.4% for tweets that do not contain URLs and 2.0% for tweets that do.

1.4.3 Relationship-specific responses to unexpected life shocks

In the third study, I propose an application of applying the relationship classifier model to studying relationship-specific differences in network behaviors. Specifically, I focus on the case of dyadic activations in response to shocks caused by unexpected life events. Exposure to events such as the sudden death of a close person or job loss can cause adverse effects on one's mental [Burton et al., 2006] and financial status [Atkinson et al., 1986]. In this study, I conduct a large computational analysis of responses to shocks in online social networks to test how interpersonal relationships engage. I introduce a new dataset of over 13,000 Twitter users who posted shock events along with their interactions with others, each labeled with their inferred relationships to the shocked user. Using causal inference methods, I approximate the effect of experiencing and posting shock events on receiving responses from Twitter users, and how these activation levels differ both in magnitude and significance depending on both the type of relationship and the type of shock. To understand the interaction between shock event types, relationships, and responding behavior, I analyze how tie strength and structural embeddedness influence the users in different relationships to reply to a shock, both strongly recognized network properties for determining interactions in social networks. Finally, we identify relationship-specific differences in the content of shock responses by measuring topic shift via a topic model.

1.5 Order of Chapters

The next chapters are arranged as follows: In Chapter 2, I will deliver a literature review of prior work that covers studies on (1) social relationships and their effects on communication and interactions, and (2) prior studies that have aimed to quantify and infer relationships using data-driven approaches. In Chapter 3, I will cover the first study, which measures social dimensions from text and networks. In Chapter 4, I will cover the second study, which infers interpersonal relationships from interaction features and examines their likeliness to offline relationships. In Chapter 5, I will show a case study of how knowledge of relationships can help understand information diffusion of networks during shock events. In Chapter 6, I will conclude by discussing the findings and addressing limitations while discussing future directions.

CHAPTER 2

Related Work

In this chapter, I will provide a literature review of two related areas to the dissertation: (1) social relationships and their effect on interpersonal communication and information, and (2) computational methods to infer properties of relationships using data-driven approaches.

2.1 Defining Social Relationships

An important first step is to correctly understand the concept of *social relationships*, *social relations*, or *interpersonal relationships*, which can be used interchangeably. Broadly defined, social relationships are social structures that consist of contact bonds among different types of groups such as friends, colleagues, relatives, and neighbors [De Belvis et al., 2008]. The distinctions between relationships are made on a person's *focus* or *foci*, a "social, psychological, legal, or physical entity around which joint activities are organized" [Feld, 1981]. People have their activities organized around foci such as workplaces [Lewicki et al., 1996], families [Umberson, 1987], or social hangouts [Oldenburg, 1999], which require them to participate and behave according to different social roles [Farnham and Churchill, 2011, Ozenc and Farnham, 2011]. When an understanding of the relationship type is established between the two parties, the conversations and interactions between the two can be shaped according to the mutually perceived relationship.

Relationships can be formed through various processes. Some relationships are established through nature and biological processes such as families [Golombok et al., 1999], while others are formed by selection depending on factors such as perceived similarity [Leenders, 2013] or attractiveness [Feingold, 1988]. Group affiliation can also be a major reason for the formation of certain relationships, especially organizational relationships in professional settings [Sluss and Ashforth, 2007, Sias and Cahill, 1998].

Regardless of how they are formed, a sufficient and ongoing amount of interaction between the two parties is required to maintain and further develop a relationship. Often, the amount of information and emotion exchanged between individuals is correlated with how deep their relationships are.

Sociologists have established theoretical models such as the uncertainty reduction theory [Berger and Calabrese, 1974] or the social penetration theory [Altman and Taylor, 1973], which propose that people engage in a series of mutual interactions to develop relationships and decide whether to make future commitments, sharing information of deeper, more intimate topics along the process. Such interactions cost resources such as time, money, and attention, setting a limit to the number of strong relationships one can handle at a time [Dunbar, 1993, Dunbar et al., 2015]. While it is true that relationships can indeed exist even without direct interaction between the two actors such as relationships within organizations where two unrelated people are under the same corporate umbrella, I will consider those cases as out of the scope of this dissertation for the focus being on interpersonal communication and interaction among relationships, thus requiring at least some amount of interaction between the two parties.

2.2 The Role of Social Relationships in Social Interactions

How social relationships play a role in shaping various types of social interactions has been widely studied throughout various fields of research. Here, I will briefly introduce studies from three angles: (1) how the language and emotion shared between speakers differ by relationship, (2) how social relationships shape the behavior of sharing information, and (3) the connection between social relationships and tie strength.

2.2.1 Language, Emotions, and Support

The practice of language is a social process and thus is heavily related to the social surroundings around oneself. In that sense, different types of relationships such as friends, acquaintances, and lovers can play a role in how one's language is shaped [Bradac, 1983]. One example would be how individuals acquire and adapt language at different stages of life. For instance, the language of children at a young age is largely influenced by their family, but are later more influenced by their close friends as they reach school years [Schieffelin and Ochs, 1986, Dickinson and Tabors, 2001, Hartup, 1989, Eckert, 1998]. Other studies also confirm that linguistic convergence is more likely to happen in close relationships such as close friends [Kovacs and Kleinbaum, 2020] or couples [Brinberg and Ram, 2021].

A more common phenomenon would be that the language people use in conversations to express their thoughts and emotions would largely differ depending on whom they are talking to. Sometimes, specific words or phrases would be reserved only for a particular relationship, one example being expressions of intimacy among romantic partners [Cook, 1994]. Another example would be language patterns that occur from status or power differences [Danescu-Niculescu-Mizil and Lee, 2011, Danescu-Niculescu-Mizil et al., 2012], clearly showing how characteristics of the relationship shape language and the emotions that are transferred through conversations. Such discrepancies arise from communicative and social norms which are agreed upon by both participants in a particular relationship, often associated with factors such as mutual benefit, power, and social distance [Fiske, 1992].

The differences in social relationships can also affect the type and effectiveness of emotions exchanged from the connection. One key dimension of how relationships play distinctive roles is in the levels of social support they provide. Wellman and Wortley [1990] study how social relationships such as friends or family are associated with social support of different types: *emotional aid*, *small and large services*, *financial aid*, and *companionship*. They discover that networks both consist of densely knit immediate kin who provide broad levels of support and segments of relationships such as friends or workmates who provide specialized support. Additionally, Agneessens et al. [2006] find that people expect more companionship from friends and more instrumental support from immediate kin. On the contrary, distant ties such as acquaintances can help access new information [Granovetter, 1973] and even help improve emotional well-being [Sandstrom and Dunn, 2014]. Overall, the findings suggest at relationships are accountable for the heterogeneity in the levels of linguistic and emotional exchange among individuals in a social network.

2.2.2 Information Sharing

People belong to several groups which are unique with their own interests and purposes [Feld, 1981, Bradac, 1983]. Each group, and the relationships formed within those groups, share knowledge of only partial aspects of the individuals who engage in a relationship [Ozenc and Farnham, 2011]. This structure results in heterogeneity of the information that gets shared and processed by each type of relationship [Ranganath et al., 2015, Haas and Sherman, 1982, Goldsmith and Baxter, 1996], which creates a concept of relationship types as a gateway of deciding which information will be passed on through a dyad.

Individuals may also strategically control the type of information to a specific type of relationship for various purposes. Examples would include not disclosing information to parents [Keijsers et al., 2010, Frijns et al., 2005, Finkenauer et al., 2002], work relationships [Frampton and Child, 2013], or romantic partners [Aldeis and Afifi, 2015]. Disclosure of information unintended to particular relationships can cause discomfort [Cobb and Kohno, 2017].

Finally, the type of relationship can affect the amount of information to be accepted when delivered by another individual. Relationships maintained over a long time tend to have a greater level of trust, thus increasing the possibility of a piece of information becoming accepted [Vanneste et al., 2014]. This can at times be problematic since the additional trust can result in believing incorrect information, potentially leading to harmful consequences. For example, the goodwill and trust put towards family members can in fact make individuals more likely to believe misinformation when it has come from family members [Chadwick et al., 2023, Waddell and Moss, 2023, Dubois et al., 2020].

2.2.3 Tie Strength

A concept in social networks that has significant relevance to social relationships is *tie strength*, the closeness of two nodes that share a tie. In the seminal paper "The Strength of Weak Ties", Granovetter [1973] defines tie strength as "a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie". Using this definition of tie strength, Granovetter further explores how weak ties, which do not have as frequent contact as strong ties, can serve as *bridges* that provide access to otherwise inaccessible information.

Tie strength has been actively studied, and especially during the past decade, several studies have contributed to understanding tie strength in online social media platforms, particularly how to infer it and its downstream effects on communication and information diffusion. Early studies have shown that behavioral and linguistic features from Facebook and Twitter can be used to show which ties are likely to provide assistance such as loaning money, helping with job search, or bringing friends to a new site [Gilbert and Karahalios, 2009, Gilbert, 2012]. Xiang et al. [2010] show that inferred tie strengths from network features in LinkedIn correlate with other metrics such as profile similarity and profile view counts. Other studies have looked into features such as network motifs [Rotabi et al., 2017] and user location [McGee et al., 2011, 2013] to detect strong ties. Recently, large-scale studies on online social networks have tested Granovetter's hypothesis on weak ties, confirming its usefulness in obtaining novel information [Bakshy et al., 2012, Rajkumar et al., 2022].

There is an overlap between the concepts of tie strength and social relationships since theories of relationship formation such as the social penetration theory rely on increased interactions for relationships to progress [Altman and Taylor, 1973]. Therefore, certain relationships such as romantic partners which require a large amount of effort to develop have greater tie strength than colleagues or acquaintances. Nevertheless, one should also be aware that in fact, ties with equal strength can result in a wide variety of relationship types [Marsden and Campbell, 1984, White, 2008, Chowdhary and Bandyopadhyay, 2015]. Again, social relationships are defined by which social foci the two individuals are expected to be at [Feld, 1981], so as long as they co-exist within that foci, there can be great heterogeneity among different individuals with whom the ego shares the same relationship type with. Therefore, social relationships should be separated from tie strength in terms of what additional information of social networks it can bring.

2.3 Inferring Properties of Social Relationships

So far, we have seen that the interactions of each relationship possess unique characteristics which are often distinguishable to a large extent. Coupled with advances in machine learning methodologies that enable the crafting of features based on multimodal data inputs such as text, network, and images, inferring properties of social relationships from interaction data has often been introduced as an end goal task in the applied machine learning domain.

2.3.1 Categories

A common approach is to design datasets and models that can be used for inferring relationship properties from dyadic interactions. Several recent studies have adopted natural language processing methods to build classifier models that infer categories of relationship types from scripts of movies and TV shows [Tigunova et al., 2021, Jia et al., 2021, Yu et al., 2020, Qiu et al., 2021, Jiang et al., 2022] or from text messages [Welch et al., 2019]. Other studies have focused on the network structure of dyads and their interactions to train classifiers that utilize network features from datasets such as mobile phone networks [Min et al., 2013, Reinhardt et al., 2015, Yu et al., 2014] or online social networks [Backstrom and Kleinberg, 2014, Tay et al., 2018].

While less relevant to the scope of this dissertation, relationship prediction from images or videos has also emerged as a popular task in the computer vision domain. In this task setting, an image or video stream containing at least two individuals involved in a particular situation is provided as input data, where the objective is to infer the relationship category based on both (1) demographics of each individual and (2) an understanding of the presented situation. Relationships are labeled from datasets such as TV show snippets [Gao et al., 2021], which are often combined with additional data domains such as text scripts or speech data. Several image- and graph-based deep learning methods have shown strong performances in such tasks [Li et al., 2017b, Wang et al., 2018, Li et al., 2020, Goel et al., 2019, Liu et al., 2019a].

2.3.2 Dimensions

An alternative approach of directly classifying the relationship of a dyad is to instead identify several social dimensions which can be used to distinguish between different types of social relationships [Wish et al., 1976, Fiske, 1992, Reis et al., 2000]. This perspective has led to the development of several theoretical models for determining the properties of social relationships. One example is the relational models theory (RMT) proposed by Fiske [1992] which suggests all human interactions can be described in terms of four types of motivations or "relational models": *communal sharing*, *equality matching*, *authority ranking*, and *market pricing*. Similarly, Foa and Foa [1980] propose relationship categories based on social exchange: *love*, *status*, *information*, *services*, *money*, and *goods*. Several survey methods have been developed for measuring social relationships between individuals, such as the Interview Schedule for Social Interaction (ISSI) [Henderson et al., 1980], the Interview Measure of Social Relationships (IMSR) [Brugha et al., 1987], and Evaluation of Social Systems (EVOS) Scale [Aguilar-Raab et al., 2015]. These surveys include a list of questions to measure dimensions such as level of attachment, social integration, sense of reliable alliance, and ability to get help and guidance during troublesome events. From a computational aspect, a number of studies have aimed to create models that can infer the strength of such dimensions from conversation text data [Rashid and Blanco, 2017, 2018] or online social network structures [Gilbert and Karahalios, 2009, Gilbert, 2012].

2.3.3 Unsupervised Clusters

The last approach in identifying relationships uses unsupervised machine learning methods to identify clusters of dyads with similar properties. A few studies have adopted vector embedding methods to model dyads of fictional characters [Iyyer et al., 2016] or online social network users [Yang et al., 2020] and aggregate the embeddings into clusters, which can be explained to some extent using explanation methods such as topic models. This method does not suffer from the issue of having to provide ground-truth labels for each dyad but lacks interpretability.

CHAPTER 3

Study 1: Ten Social Dimensions of Relationships

3.1 Introduction

Research in the social sciences dedicated considerable efforts to draw systematic categorizations of the fundamental sociological dimensions that describe human relationships [Fiske, 1992, Wellman and Wortley, 1990, Bicchieri, 2005, Spencer and Pahl, 2018]. This was partly motivated by the necessity to model relationships beyond tie strength [DeDeo, 2013, Aiello et al., 2014, Aiello, 2018], as ties with equal strength may result in a wide variety of relationship types [Marsden and Campbell, 1984, White, 2008, Chowdhary and Bandyopadhyay, 2015]. Recently, such extensive literary production was surveyed by Deri et al. [2018], who compiled an extensive review of decades' worth of findings in sociology and social psychology to identify *ten dimensions* that have been widely used as ways to categorize relationships: *knowledge*, *power*, *status*, *trust*, *support*, *romance*, *similarity*, *identity*, *fun*, and *conflict* (description in Table 3.1). Although these categories are not meant to cover exhaustively all possible social experiences, Deri et al. provided empirical evidence that most people are able to characterize the nature of their relationships using these ten concepts only. Through a small crowdsourcing experiment, they asked people to spell out keywords that described their social connections (Table 3.1) and found that all of them fitted into the ten dimensions.

By combining these ten fundamental blocks in opportune proportions, one can draw an accurate, explainable, and intuitive description of the nature of most relationships, as perceived by the people involved. However, although the ten dimensions provide a useful way to conceptualize relationships, it is not clear to what extent these concepts are expressed through language and what role they have in shaping observable dynamics of social interactions. The growing availability of online records of conversational traces provides an opportunity to mine linguistic patterns for markers of their presence. Past research in Web Mining and Natural Language Processing (NLP) studied aspects pertaining some of the dimensions we deal with in this work [Danescu-Niculescu-Mizil et al., 2012, Ma et al., 2017], with special attention to concepts at the extremes of the spectrum of sentiment such as conflict [Kumar et al., 2018] or empathy [Morelli et al., 2017, Polignano et al., 2017] and

Dim.	Description	Keywords	References
Knowledge	Exchange of ideas or information; learning, teaching	teaching, intelligence, competent, expertise, know-how, insight	[Fiske et al., 2007, Levin and Cross, 2004]
Power	Having power over the behavior and outcomes of another	command, control, dominance, authority, pretentious, decisions	[French et al., 1959, French Jr, 1956, Blau, 2017]
Status	Conferring status, appreciation, gratitude, or admiration upon another	admiration, appreciation, praise, thankful, respect, honor	[Blau, 2017]
Trust	Will of relying on the actions or judgments of another	trustworthy, honest, reliable, dependability, loyalty, faith	[Luhmann, 2018, Zaheer et al., 1998]
Support	Giving emotional or practical aid and companionship	friendly, caring, cordial, sympathy, companionship, encouragement	[Baumeister and Leary, 1995, Fiske et al., 2007, Vaux, 1988]
Romance	Intimacy among people with a sentimental or sexual relationship	love, sexual, intimacy, partnership, affection, emotional, couple	[Buss and Schmitt, 1993, Buss, 2006, Emlen and Oring, 1977]
Similarity	Shared interests, motivations or outlooks	alike, compatible, equal, congenial, affinity, agreement	[McPherson et al., 2001, Jackson et al., 2008]
Identity	Shared sense of belonging to the same community or group	community, united, identity, cohesive, integrated	[Tajfel, 2010, Oakes et al., 1994, Cantor and Mischel, 1979]
Fun	Experiencing leisure, laughter, and joy	funny, humor, playful, comedy, cheer, enjoy, entertaining	[Radcliffe-Brown, 1940, Billig, 2005, Argyle, 2013]
Conflict	Contrast or diverging views	hatred, mistrust, tense, disappointing, betrayal, hostile	[Berlyne, 1960, Tajfel et al., 1979]

Table 3.1: The ten social dimensions of relationships studied by decades of research in the social sciences. The keywords are the most popular terms used by people to describe those dimensions, according to Deri at al. [Deri et al., 2018]'s survey.

support [Wang and Jurgens, 2018, Yang et al., 2019]. The operationalization of some of these concepts proved useful to improve the accuracy of prediction tasks [Buntain and Golbeck, 2014, Wang et al., 2016a, Mitra and Gilbert, 2014, Wen et al., 2020].

So far, little work has been conducted to explore all the ten dimensions systematically and jointly in relation to the use of language. In this study, we show that all ten social dimensions can be predicted purely from *conversations*, and that the combination of the predicted dimensions suggests both the types of *relationships* people entertain and the types of real-world *communities* they shape. Specifically, we made three main contributions:

- We collected conversation records from various sources (§3.2), and we labeled them according to the ten dimensions using crowdsourcing. We obtained annotations for a total of ~9k texts and ~5k Twitter relationships (§3.3.1), and found that all dimensions are abundantly expressed in everyday language (§3.4.1).

- Using the collected data, we train multiple classifiers to predict the 10 dimensions purely from text (§3.3.2). Some dimensions are harder to predict because of their more complex lexical variations. Deep learning classifiers are more capable of handling such complexity, yielding an average AUC of 0.85 across the dimensions and a maximum AUC of 0.98 (§3.4.2). The model shows a good level of robustness when tested on unseen data sources.

- We find that the combination of the dimensions predicted from two individuals' conversations on Twitter predicts their type of social relationships (§3.4.3). Further, by applying our framework to 160M messages written by geo-referenced Reddit users, 290k emails from the Enron corpus, and 300k lines of dialogue from movie scripts, we show that the presence of the ten dimensions in the language is indicative of the types of communities people shape (§3.4.4). For example, some of the dimensions are predictive of societal outcomes in US States, such as education, wealth, and suicide rates (§3.4.5).

3.2 Data collection

To test our method on a diverse range of data, we extracted information about conversations and relationships from four sources.

3.2.1 Reddit comments

Reddit is a public discussion website, is one of the most accessed websites in the World and mostly popular in the United States where half of its user traffic is generated. Reddit is structured in 140k+ independent *subreddits* dedicated to a broad range of topics [Medvedev et al., 2019]. Users can post a new *submission* to a subreddit and write *comments* to existing submissions. A dataset containing the vast majority of the submissions and comments published on Reddit since 2007 is publicly available [Baumgartner et al., 2020]. We gathered the data for the year 2017, which is nearly complete, according to recent estimates [Gaffney and Matias, 2018]. In total, we collected 96,212,869 submissions and 886,886,260 comments from 13,874,369 users.

To match Reddit discussions with census data (§3.4.5), we focused our analysis on users whom we could geo-reference at the level of US States. Reddit does not provide explicit information about user location, yet it is possible to get reliable location estimates with simple heuristics. Following the approach by Balsamo et al. [Balsamo et al., 2019], we first selected 2,844 subreddits related to cities or states in the United States. From each of those, we listed the users who posted at least 5 submissions or comments. From the resulting set of users, we removed those who contributed to subreddits in multiple states. This resulted in 967,942 users who are likely to be located in one of the 50 US states. The number of users per state ranges from 1,042 (South Dakota) to 75,548 (California). In 2017, these users posted 9,553,410 submissions and 148,114,859 comments overall.

We used this data to conduct a spatial analysis of the use of language (§3.4.5) and we sample from it to build our training set (§3.3.1).

3.2.2 Enron emails

Enron Corporation was an American company founded in 1985 that went bankrupt in 2001, when its systematic practices of accounting fraud were exposed to the public. After the scandal and the resulting investigation, The Enron Email Dataset [Klimt and Yang, 2004] was released to the public and became a popular resource for research in network science and Natural Language Processing [Coffee Jr, 2001, Klimt and Yang, 2004, Diesner and Carley, 2005, Peterson et al., 2011]. Messages include the full text and the email header. By filtering on the "from:" and "to:" fields, we obtained a corpus of 287,098 messages exchanged among 9,706 employees between year 2000 and 2001. In this study, we use a sample of annotated Enron emails to test our classifier's performance (§3.4.3), and we look at its entirety to conduct a descriptive study (§3.4.4).

3.2.3 Movie dialogs

Scripted movie dialogs are fictional yet plausible representations of conversations that span a wide spectrum of human emotions and relationship types. The Cornell Movie-Dialogs Corpus [Danescu-Niculescu-Mizil and Lee, 2011] is one of the most comprehensive open collections of movie scripts, containing 304,713 utterances exchanged between 10,292 pairs of characters from 617 movies. Past research used it to investigate the relationship between language and social interaction dynamics [Danescu-Niculescu-Mizil et al., 2012]. We used it to test our classification system (§3.4.3), and for conducting a qualitative analysis of its output (§3.4.4).

3.2.4 Twitter relationships

Tinghy.org is a website that hosts a series of "gamified" psychological tests. Launched in 2018, it was conceived by Deri et al. [2018] as a platform to collect data about how social media users perceive their online relationships in terms of the 10-dimensional model of relationships. In one of these games, users log in with their Twitter account and they are sequentially presented with 10 of their Twitter followees. The selection of contacts is biased towards the strongest ties with the player. This is done using a validated linear regression model (see Table 1 in [Gilbert, 2012]) that estimates tie strength through a number of factors that can be calculated from the data exposed by the public Twitter API (e.g., time elapsed since last interaction). The player picks one to three dimensions over the 10 available to describe their relationship with each of the friends displayed (Figure 3.1). With the explicit user consent, interaction data is gathered through the Twitter API.

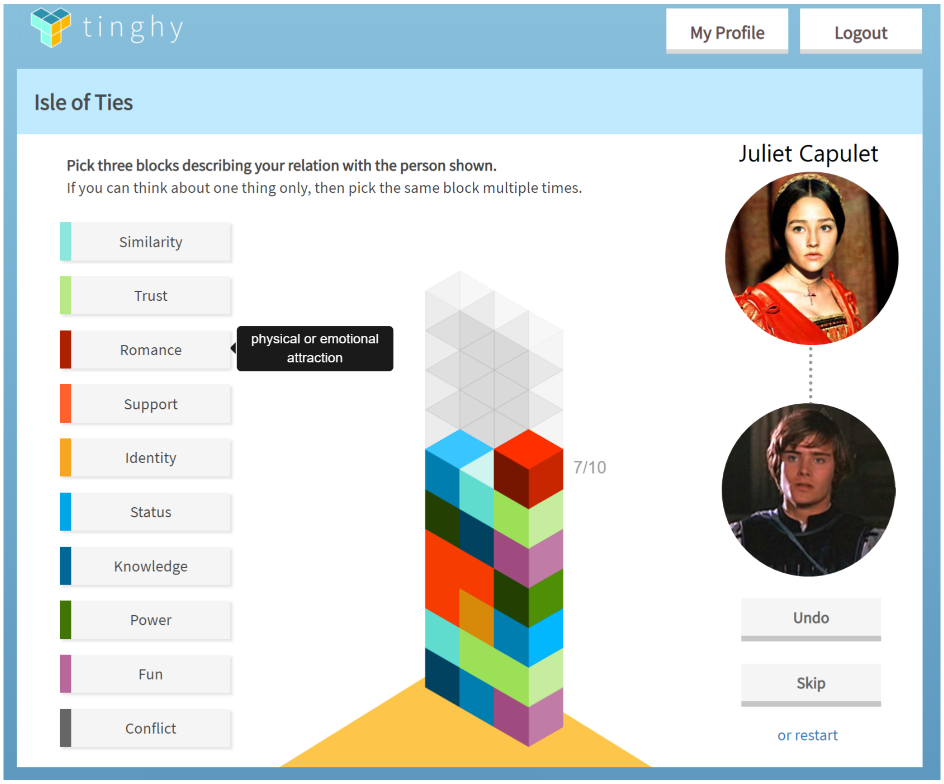


Figure 3.1: Anonymized screenshot of the Tinghy game. The player (bottom profile picture) is presented with 10 Twitter friends, one at the time (top profile picture) and is asked to describe their relationship by picking 1 to 3 dimensions from the menu on the left. By doing so, new blocks are added to the "friendship wall" in the middle. The dimensions are explained to the player with short text snippets.

For every player-friend pair (u, v), the dataset contains *i)* a list of up to three dimensions picked by u, sorted by order of selection; *ii)* the list of all tweets in which u mentions (or replies to) v or viceversa; and *iii)* the list of u's tweets that were retweeted by v or viceversa. To date, 684 people played the game, providing labels for 5,217 social ties between a total of 3,777 unique individuals (the data was recorded even when players quit the game before completion). These ties exchanged 9,960 mentions, 31,100 replies and 8,619 retweets overall. We restricted our study to English tweets that account for 1,772 relationships between 1,406 unique individuals for a total of 8,870 mentions, 19,254 replies and 5,050 retweets.

Unlike the ground-truth labels for the other datsets, which are at sentence-level (§3.3.1), the annotations coming from this game are provided at relationship-level. This allowed us to test the extent to which one could predict the dominant social dimension of a relationship from conversations (§3.4.3).

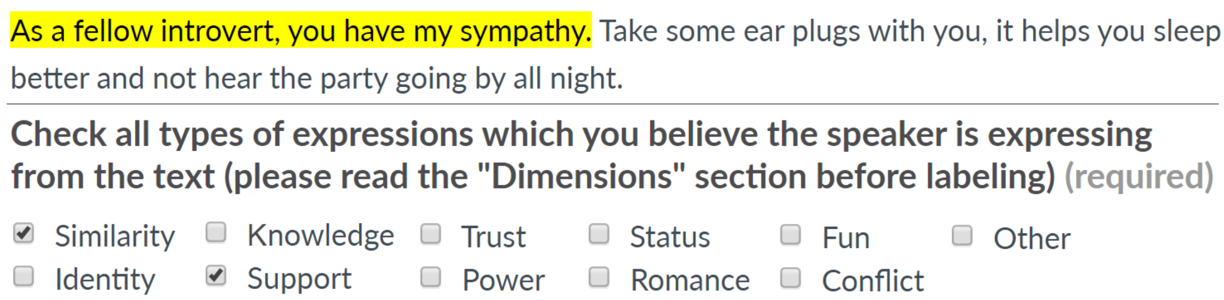


Figure 3.2: Example of the crowdsourcing task. The highlighted sentence conveys a combination of social support and similarity.

3.3 Methodology

We adopted a supervised approach to extract the ten social dimensions from text. We crowdsourced a dataset of conversational texts annotated with the 10 dimensions (§3.3.1), and we used it to train multiple classifiers (§3.3.2).

3.3.1 Crowdsourcing

To annotate text, we followed the same procedure for Reddit comments, movie dialogs, and Enron emails. For each data source, we split all texts into sentences, and retain only the sentences that contain at least one 1^{st} or 2^{nd} person pronoun. This filtering step is meant to bias the selection in favor of phrases that follow a conversational structure. We then selected a random sample of sentences with length between 6 and 20 words, to avoid statements that are too complex to assess or too short to be informative. For each sentence, we also kept the preceding and following sentences from the same text, if any. The addition of neighboring sentences is helpful for the annotators—albeit not strictly necessary—to make better sense of the context around the sentence.

Each resulting *passage*, composed by the target sentence highlighted with color and surrounded by the neighboring phrases, is presented to crowdworkers for annotation. We asked them to read the whole passage and to select the dimensions that they believe the highlighted sentence conveys, among the 10 provided (Figure 3.2). Annotators were encouraged to select multiple dimensions when they felt that more than one applied. A special label *"other"* was provided in case the annotator was uncertain or no available option seemed pertinent. Each sentence was annotated by three people.

Before starting the task, annotators read the definitions of all the 10 dimensions, which were extended versions of the statements in Table 3.1. For example, social support was described as: *"Expressions that suggest the offer of any type of emotional or practical aid, which might come in different forms, including: sympathy, compassion, empathy, companionship, offering to help."* Definitions were accompanied by 3 to 5 examples (e.g., for social support: *"I am so sorry for your* *loss."*). Instructions were accessible at any time during the task, for quick reference.

As a quality-control mechanism, we inserted *test sentences* both at the beginning of each task and at random positions in the task. These consists of variations of the examples provided in the instructions, for which the correct dimension is known. The test sentences served two purposes. First, whenever an annotator provided a wrong answer to a test sentence, the correct answer was shown, so that they could learn from their mistakes. Second, annotators who failed to assign correct labels to 40% of the test sentences or more were banned from the task, and their answers were discarded. Through small-scale preliminary tests, we empirically observed that 40% was a good threshold to filter out misbehaving users.

We deployed the task on the crowdsourcing platform "Figure Eight". We opened the participation only to people residing in five English-speaking countries (United States, United Kingdom, Ireland, Canada, Australia) and who belong to the platform's top-tier expert contributors. We set the price for each annotation task to 0.05$, which amounts to a 9$ hourly wage considering an average time of 20 seconds spent on each sentence. We collected labels for 7,855 sentences from Reddit posts, 400 from movie lines, and 436 from Enron emails, which were provided by 934 annotators who labeled 28 sentences each on average. Workers spent 23s per sentence on average (σ = 35s). The reported level of satisfaction after the task was 4.0 out of 5, on average.

3.3.2 Classification

3.3.2.1 Classifiers

We experiment with four classification frameworks: a traditional ensemble classifier, a simple metric based on distance between words in an embedding space, and two deep-learning models.

Xgboost. An ensemble of decision trees with gradient boosting [Chen and Guestrin, 2016]. It is well-suited to small datasets, makes it easy to interpret the contribution of individual features, and is able to ignore any vacuous features that may be present to prevent overfitting. Xgboost has proven to be the best performing classifier among competitors in popular challenges. We train Xgboost using the features defined in §3.3.2.2, computed at sentence-level. We performed grid search to tune its learning rate and the maximum depth of its trees. In a binary classification task, Xgboost outputs a confidence score in [0,1] that captures the likelihood of the sample belonging to the positive class.

Embedding distance. Word embeddings are dense vector representations of words that capture the linguistic context in which words occur in a corpus. Such representations are generally learned by training neural network models on large text corpora to predict the occurrence of words from their local lexical context. Each word is associated with a point in the embedding space such that words that share common contexts are close to one another. Many embedding techniques have been developed recently [Li and Yang, 2018], and several pre-trained models are readily available. GloVe [Pennington et al., 2014] embeddings with 300 dimensions, trained on the Common Crawl corpus (42B tokens) performed best in the tasks we addressed. In addition to considering a word's local context, GloVe uses also global co-occurrence statistics across the whole text corpus.

We leveraged the properties of the embedding space to implement a simple measure of distance between a sentence and each of the 10 conversational dimensions. We first computed a sentence-level embedding vector $\mathbf{g}_s$ by averaging the embedding vectors of all the words in a sentence s:

$$\mathbf{g}_s = \frac{1}{len(s)} \cdot \sum_{w \in s} \mathbf{g}_w, \tag{3.1}$$

where $\mathbf{g}_w$ is the GloVe vector of word w. We used the same formula to compute an embedding vector $\mathbf{g}_d$ for the words representative of each dimension d, as listed in Table 3.1. We then computed the Euclidean distance between the two resulting vectors: $d(\mathbf{g}_s, \mathbf{g}_d)$. This method yields a single measure that does not offer a natural threshold for binary classification, yet one that can rank sentences by their 'relevance' to a dimension.

LSTM. Long Short-Term Memory (LSTM) [Hochreiter and Schmidhuber, 1997] is a type of recurrent neural network (RNN) particularly suited to process data that is structured in temporal or logical sequences. LSTMs have demonstrated to achieve excellent results in timeseries forecasting [Lipton et al., 2015, Greff et al., 2016] as well as in NLP tasks [Sundermeyer et al., 2012]. LSTM accepts fixed-size inputs; in our experiments, we fed it with a 300-dimension GloVe vector of a word, one word at a time for all the words in a sentence. Each new word updates the model's status by producing a new hidden-state vector. Following the standard approach, we applied a linear transformation to reduce the last hidden vector into one scalar value, and we apply a sigmoid function to transform it into a continuous value between 0 and 1, which indicates the probability of belonging to the positive class. We experimented with a simple LSTM model with no attention, short-cut connections, or other additions. We performed grid search to tune its hyperparameters (learning rate and number of epochs).

BERT. Transformers [Vaswani et al., 2017] are models designed to handle ordered sequences of data by relying on attention mechanisms rather than on recurrence. As opposed to directional models like LSTM, which read the input sequentially, transformers parse an entire sequence of words at once, thus allowing the model to learn the context of a word based on all of its surroundings (left and right context). BERT (Bidirectional Encoder Representations from Transformers) is a language representation model based on Transformers and pre-trained on a 3.3B word corpus from BooksCorpus and Wikipedia [Devlin et al., 2018]. It has been adapted to solve several NLP tasks, achieving state-of-the-art results. We used a pretrained BERT-Base Cased model. Following the

Feature family	Feature names	# features
Linguistic style	politeness [Brown et al., 1987, Danescu-Niculescu-Mizil et al., 2013]; hedging terms [Fu et al., 2017]; morality-related words [Haidt and Graham, 2007]; integrative complexity [Robertson et al., 2019]; syntactic markers [Tchokni et al., 2014]: word elongations, use of capital words, #question marks, #exclamation marks, #ellipsis	50
Readability & complexity	#words; avg. length of words; avg. syllable per word; entropy of words [Tan et al., 2016]; readability indices [Jurafsky, 2000]: Kincaid, ARI, Coleman-Liau, Flesch Reading Ease, Gunning-Fog index, SMOG index, Dale Challenge index	12
Linguistic lexicons	LIWC [Pennebaker et al., 2001]; Empath [Fast et al., 2016]	175
Sentiment	VADER [Hutto and Gilbert, 2014]; Hatesonar [Davidson et al., 2017]	6
Word distribution	n-grams [Jurafsky, 2000]	100

Table 3.2: Interpretable linguistic features for classification

original specifications [Devlin et al., 2018], we fine-tune it to perform binary classification by adding a classification layer on top of the Transformer output, which results into a 2-dimensional output vector representing the two output classes. Last, we apply a softmax transformation to get a single score in [0,1] that reflects the likelihood of the input belonging to the positive class. We performed grid search to tune its learning rate and the number of epochs.

3.3.2.2 Interpretable features

To train the Xgboost model, we extracted a total of 343 features, partitioned in five families (Table 3.2). We picked these sets of features because they have been successfully used to solve a variety of NLP tasks, they are intuitively interpretable, and they cover several facets of language use. Here we summarize them shortly and we refer the reader to the original publications for the detailed formulations. The first family of features captures aspects of *linguistic style*: the use of formulas of politeness [Danescu-Niculescu-Mizil et al., 2013] and complex argumentation [Fu et al., 2017, Robertson et al., 2019]; the presence of words that appeal to morality [Haidt and Graham, 2007]; and the use of a number of simple syntactic markers [Tchokni et al., 2014]. The second one comprises a measures of *readability and writing complexity*, ranging from simple counts to more sophisticated indices [Jurafsky, 2000]. The third one includes LIWC [Pennebaker et al., 2001] and Empath [Fast et al., 2016], two widely used *linguistic lexicons* that map words into linguistic, psychological, and topical categories. The fourth one captures the spectrum of *sentiment* with VADER [Hutto and Gilbert, 2014], a rule-based tool to measure positive/negative emotions in short text, and Hatesonar [Davidson et al., 2017], a tool to detect offensive language. Last, to capture the *distribution of words*, we counted a sentence's unigrams and bigrams. To reduce the sparsity of the *n*-gram space, we considered only those that occur 10 times or more in the training set and we filtered them using log-odd Dirichlet priors to further narrow the set to those *n*-grams that are highly discriminative [Monroe et al., 2008]. Specifically, we kept only the top 100 n-grams ranked by $\xi = \log p(w|w \in P_d) - \log p(w)$, where $p(w)$ is the probability of a n-gram w occurring in the full corpus, and $p(w|w \in P_d)$ is the probability of occurring in the sentences of the positive set for the target dimension (P_d).

3.3.2.3 Task definition

Given a sentence s and a social dimension $d \in D = \{d_1, \ldots, d_{10}\}$, our task was to determine whether s conveys d. Rather than training one multi-class classifier, we treated each dimension independently and trained multiple binary classifiers. This choice was motivated by the non-exclusive nature of the ten dimensions [Deri et al., 2018]: a sentence may convey any pair (or subsets) of dimensions at once—which we confirmed in our results (§3.4.4).

Given a dimension d, we included in its set of positive samples P_d all the sentences that were labeled with d by two annotators or more, and we put all the sentences never labeled with d in the set of negative samples N_d. In each round of a 10-fold cross-validation, we randomly split each set in 80% for training, 10% for tuning, and 10% for testing. Since $|P_d| < |N_d| \forall d$, we performed random oversampling [Ling and Li, 1998] to balance the classes. Specifically, within each training, tuning, and testing split, we added multiple copies of positive samples picked at random until the size of the two classes got balanced. Compared to other oversampling techniques [Chawla et al., 2002, He et al., 2008], random oversampling does not generate synthetic data points, which might end up exhibiting unrealistic features. Its application is equivalent to giving higher importance to positive samples: classifying a positive instance correctly yields a performance gain that is proportional to the number of replicas (or an equally great loss if misclassified).

We measured performance with the average "Area Under the ROC Curve" across all folds—AUC, in short. AUC measures the ability of the model to correctly rank positive and negative samples by confidence score, independent of any fixed decision threshold. Because the data is balanced, the expected value of AUC for a random classification is 0.5.

3.4 Results

3.4.1 Conversations

Most agreement scores are well-defined for sets of items judged by all raters. We compute an inter-annotator agreement score on the set of test sentences which have been rated by all annotators. On this set, the Fuzzy Kappa agreement score [Kirilenko and Stepchenkova, 2016]—an extension of Cohen's Kappa that contemplates the possibility of an instance being placed in multiple categories [McHugh, 2012]—is 0.45, which indicates moderate agreement. On the full set, no consensus was reached

Data	total#	#Dimensions 0	1	2	3+
All	8,691	41%	53%	5%	1%
Reddit	7,855	43%	54%	3%	0%
Movies	400	10%	59%	24%	7%
Enron	436	22%	59%	14%	5%

Table 3.3: Fraction of messages labeled with n numbers of dimensions from the annotators

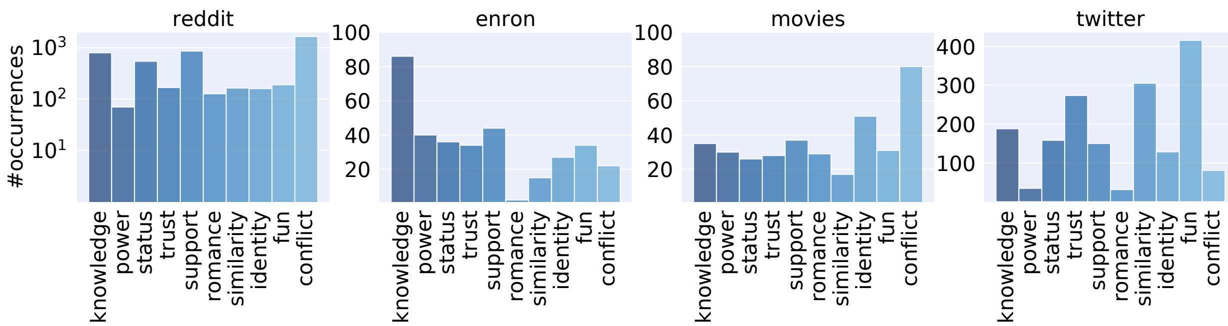


Figure 3.3: Distributions of labels across datasets.

on 41% of the sentences, which were assigned no dimension. Some agreement is reached for the remaining 59%: 53% were assigned exactly one dimension, 5% two, and 1% three or more. Source-specific proportions are listed in Table 3.3. Despite the selection of sentences was performed at random, almost 60% of those from Reddit carry a social value that could be linked to the 10 dimensions. In movie scripts, this fraction raises to 90%, which is expected considering that the narrative structure of movies compresses dense information about character relationships in a limited number of lines. Next, we focused on those sentences on which annotators reached some consensus, and used the remaining ones only as negative examples for training. In §3.5, we discuss the nature of the sentences for which no consensus was reached.

Verbal expressions do not represent all dimensions in equal measure, and the relative proportions vary considerably across data sources (Figure 3.3). In Reddit, conflict is predominant, followed by support, knowledge, and status. This is in line with previous work that showed that Reddit communities are often aimed at providing social support [De Choudhury and De, 2014, Cunha et al., 2016, De Choudhury et al., 2016], but they are also prone to fall prey of misbehaving users [Cheng et al., 2017, Kumar et al., 2018]. In Enron, the relative abundance of knowledge-exchange messages reflects the nature of goal-oriented communication in corporations; unsurprisingly, romance is non-existent. Lines from movie scripts exhibit high level of conflict and identity, likely due to how fictional story arcs pivot around overcoming interpersonal challenges [Field, 2005], often

	Xgboost										
	Linguistic	Hate	Readability	VADER	Empath	LIWC	Ngrams	All	**Embedding**	**LSTM**	**BERT**
Knowledge	0.61	0.6	0.65	0.66	0.7	0.77	0.69	0.76	0.53	**0.82**	**0.82**
Power	0.54	0.56	0.57	0.58	0.68	<0.5	0.58	0.54	0.53	**0.82**	0.74
Status	0.67	0.58	0.61	0.78	0.78	0.79	0.78	0.82	0.78	**0.86**	0.85
Trust	0.7	<0.5	0.61	0.76	0.72	0.75	0.76	**0.80**	0.72	0.77	0.73
Support	0.62	0.55	0.64	0.69	0.75	0.78	0.69	0.79	0.66	0.83	**0.85**
Romance	0.85	0.53	0.77	0.82	0.97	0.93	0.82	0.96	0.78	**0.98**	0.93
Similarity	0.5	0.53	0.55	0.62	0.63	0.6	0.62	0.63	0.64	0.80	**0.82**
Identity	<0.5	<0.5	0.57	0.50	0.55	0.67	0.62	0.59	0.66	**0.75**	0.62
Fun	<0.5	0.62	0.71	0.76	0.86	0.86	0.65	0.95	0.83	0.94	**0.98**
Conflict	0.57	0.57	0.64	0.79	0.75	0.81	0.61	0.84	0.66	0.86	**0.91**

Table 3.4: Performance of different models on each dimension for the Reddit dataset (average AUC over 10-fold cross validation). Top performances are highlighted in bold.

instantiated by cohesive factions opposing each other [Wolfenstein, 2002]. For Twitter relationships, the dominant dimensions are fun, similarity, trust, and knowledge, which reflect partly the bias of the data collection towards strong ties, and partly the nature of Twitter as a community of interest in which like-minded people exchange information [Kwak et al., 2010, Conover et al., 2011].

3.4.2 Classifying conversations

Prediction results are summarized in Table 3.4. Among all the prediction models, the embedding similarity performed worst. LSTM and BERT reached comparable performances, yielding top scores on 5 dimensions each, with a tie on *knowledge*; their performance gap is minor in most dimensions, with peak performances ranging from 0.75 to 0.98. AUC generally drops when using the Xgboost model, even when relying on all available features. Xgboost obtained the best performance on *trust* only, and by a small margin.

Across classifiers, results suggest that some dimensions are easier to predict than others. For example, simple lexicons for sentiment analysis reach AUC scores exceeding 0.85 for *fun* and *romance*. To check the link between performance and size of training data, we plot the AUC against the number of positive samples for each dimension (Figure 3.4, left—LSTM only, for brevity). The AUC increases linearly with the dataset size (R^2 = 0.37) except when considering two outliers: *romance* and *fun*, which are associated with good performances despite the scarcity of their training data. We hypothesize that this discrepancy is due to the diverse nature of verbal expressions:

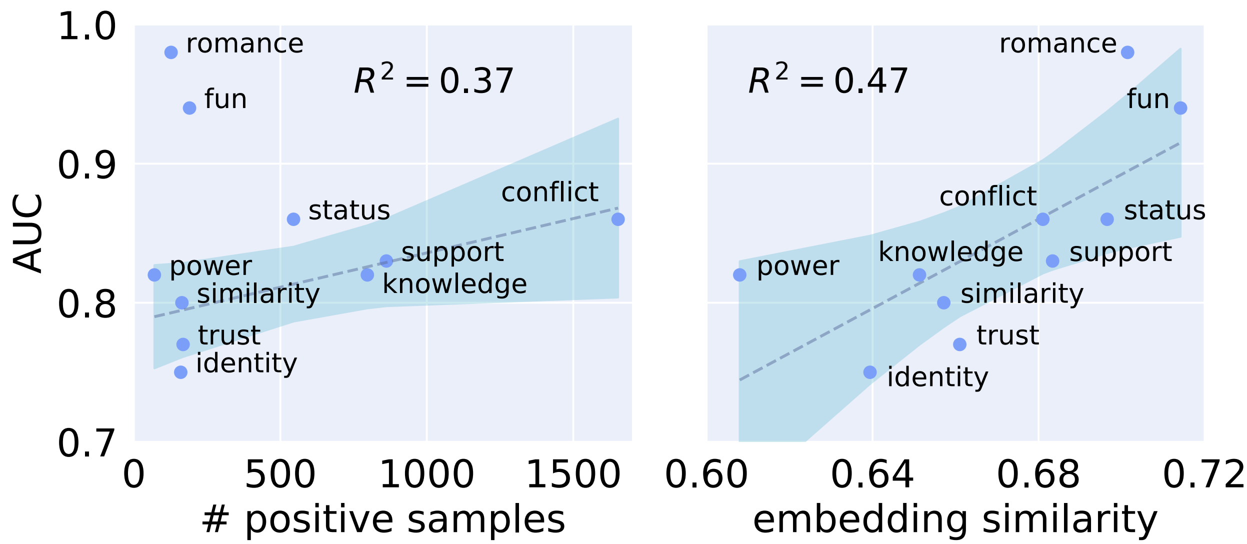


Figure 3.4: AUC increases with the size of the training data (left) and with the lexical homogeneity of the expressions used to express a dimension, estimated with average similarity in the embedding space (right).

the more limited the language variations used to express a dimension, the easier to predict those variations. To verify it, we computed the sentence-level embedding vectors (using Formula 3.1) for all sentences in the sets of positive samples $P_d, \forall d$. We then measure the average cosine similarity between 100k random pairs of sentences within the same set P_d, which gives an estimate on how semantically close the verbal expressions in each dimensions are. We find a significant linear relationship ($R^2 = 0.47$) between average embedding similarity and AUC (Figure 3.4, right). As expected, *romance* and *fun* are the ones with highest similarity. This trend holds for all classifiers but it is particularly pronounced for Xgboost and for the simple embedding similarity baseline.

We conclude that, although Xgboost yields decent performances in some cases, its effectiveness suffers from the higher lexical variety of expressions in some dimensions (e.g., *power* or *identity*) more than that of deep learning models. Nevertheless, the nature of the Xgboost framework allows us to study the importance of its interpretable features in predicting different dimensions, thus providing a human-readable indication of whether the content of verbal exchanges in the labeled data matches theoretical expectations. We measure each feature's effect size using Cohen's d, and report only those with $d > 0.4$, which corresponds to a substantial effect size [Cohen, 2013]. Table 3.5 shows the important features organized into each feature category. The features that emerge echo the theoretical definition of the ten dimensions (Table 3.1). Naturally, sentiment is an important feature for most. Pleasant interactions express positive sentiment, *knowledge* and *power* tend to be neutral, and *conflict* carries negative sentiment. Furthermore: *knowledge* is associated with complex writing; *romance*, *support*, and *trust* with a sense of empathy and attachment; *power* with work-related topics and with words conveying authority; *fun* with words of play and celebration;

Dimension	Top features
Knowledge	Readability (ARI, Kincaid, Gunning Fog Index, avg. words per sentence); VADER (neutral); Style (hedging)
Power	Liwc (power, work); Vader(neutral); Empath (order, business, power)
Status	Liwc (affect, posemo); Vader (positive); Empath (giving, optimism, politeness)
Trust	Liwc (posemo, affect); Vader (positive); Empath (friends, help, trust); Style (empathy words)
Support	Liwc (posemo); Vader (positive); Empath (optimism, help, giving); Ngram (“thank you”); Style (empathy words)
Romance	Empath (affiliation, affection, friends, sexual, wedding, optimism); Style (empathy words); Liwc (affiliation, bio, social, drives, ppron, posemo) Vader (positive); Ngram (“love”)
Similarity	Liwc (compare); Empath (appearance); Ngram (“like”); Style (integration words)
Identity	Liwc (religion); Hatesonar (hatespeech); Empath (sexual)
Fun	Empath (celebration, childish, children, fun, leisure, party, ridicule, toy, vacation, youth, optimism); Liwc (affect, posemo); Vader (positive); Style (“!”)
Conflict	Vader (negative); Liwc (anger, negate, swear, negemo); Readability (Dale Challenge Index); Empath (hate, swearing terms); Hatesonar (offensive language)

Table 3.5: Important feature groups per dimension in the Xgboost classifier (Cohen’s $d > 0.4$)

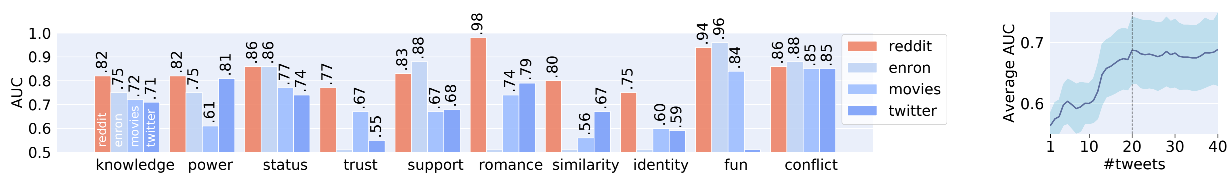


Figure 3.5: Left: AUC of LSTM models trained on the Reddit data and tested on the other datasets. Right: growth of AUC in the classification of Twitter relationships as the number of messages exchanged between the two users increases.

similarity with verbal formulas of comparison.

For simplicity, in the remainder of the paper we report only results for LSTM, which is computationally simpler and faster than BERT, and achieved similar results.

3.4.3 Classifying relationships

To test the adaptability of our model to different domains, we trained dimension-specific LSTM classifiers on all the available Reddit data and tested them on the corpora from Enron and movie scripts. Results are summarized in Figure 3.5 (left).

In Enron, the performance did not drop when detecting *status*, *support*, *fun* and *conflict*, whereas *knowledge* and *power* suffered a loss within 0.1. The AUC dropped when detecting utterances of *similarity* and *identity*, which both rarely appear in our labeled Enron sample. The model adapted to a lesser extent to movie scripts, arguably because the composition of scripted text is intrinsically different from user-generated text in blog posts or emails. Still, we recorded limited or no AUC loss for four dimensions out of ten (*knowledge*, *status*, *fun*, and *conflict*). As we shall see in our qualitative analysis (§3.4.4), even the lowest-performing classifiers dimensions returned meaningful results when applied to larger data sources and only high-confidence sentences were kept.

Last, we used the data collected from the Tinghy game to address an even more challenging task: predicting *relationship-level* labels from conversations. For every pair of Twitter users u, v, we considered only the first dimension that u picked in the game; the first association that comes to mind is likely to be the most relevant and important, according to several models of human attention [Broadbent, 1957, Fleming and Koman, 1998, Cutrell and Guan, 2007]. We leave a multi-dimensional analysis of relationships to future work. We ran our classifier on the text of each mention, reply and retweet between the two users, disregarding the directionality of interaction. We estimated a relationship-level label by picking the most frequent dimension across all the messages.

We observed that the average AUC across dimensions grows with the volume of messages exchanged between the users. After a minimum of 20 messages, the performance reaches a plateau (Figure 3.5, right). Therefore, we limited the prediction only to pairs of users who were involved in at least 20 interactions. In this setting, the prediction worked best (Figure 3.5, left) for *conflict* and *status* ($AUC > 0.8$), and for *power*, *support*, and *romance* ($AUC > 0.7$).

Overall, models that predict *conflict*, *status*, and *knowledge* were the most robust across sources. Predictions suffered limited losses for about half of the dimensions in each dataset, which is remarkable given the limited size of training data. Finally, with the predictions on Twitter relationships, we produced evidence that the model could learn the perceived nature of a social tie from the conversations that flow over it.

3.4.4 Qualifying conversations and relationships

We provide a qualitative assessment of the output of our tool on the Enron emails and on the movie scripts.

3.4.4.1 The fall of Enron

The ability of identifying a rather comprehensive set of dimensions from conversational text enables us to interpret social phenomena with broader nuances compared to traditional tools like sentiment analysis. Both the longitudinal nature of the Enron dataset and the well-documented stages of the company's downfall make it possible to test whether exogenous events impact the presence of certain social dimensions in people's exchanges and relationships.

We ran our ten LSTM models $M_d, d \in \{1 \dots 10\}$ on every email, and marked a text t with dimension d if the maximum confidence score for dimension d across all its sentences is higher than 0.95, namely $max(\{M_d(s), \forall s \in t\}) > 0.95$. In other words, a text conveys a dimension if at least one of its sentences is predicted with high confidence to express that dimension. For all the emails sent during a calendar week t, we calculated the ratio $f_d(t)$ between emails carrying dimension d

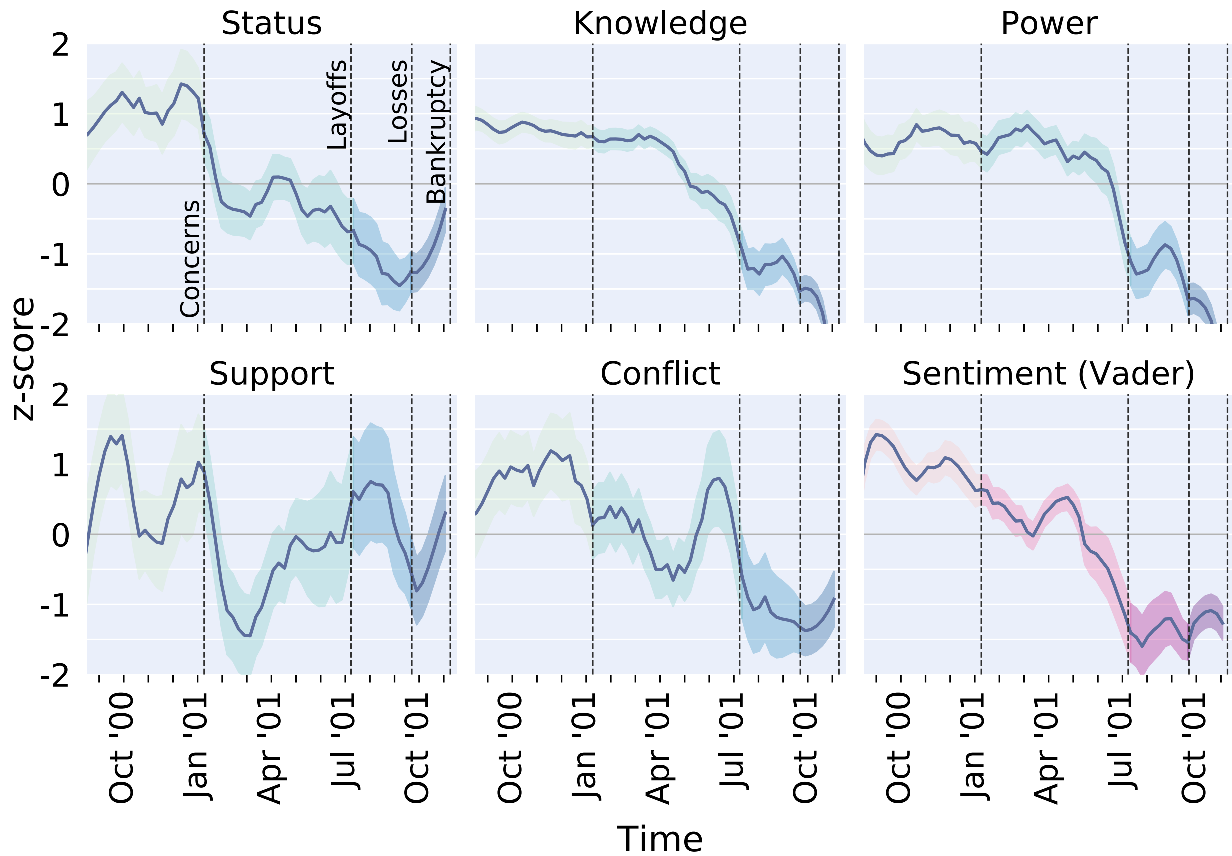


Figure 3.6: How the presence of five social dimensions in Enron employees emails changes over time, compared to a sentiment analysis baseline. Status giving, knowledge transfer, and the power-based exchanges plummet after the first financial concerns. After massive layoffs, the remaining employees give support to each other.

and the total numbers of emails sent. Finally, we transformed these fractions into z-scores to make the values comparable across dimensions:

$$zscore_d(t) = \frac{f_d(t) - \mu_d}{\sigma_d} \tag{3.2}$$

where μ_d and σ_d are the average and standard deviation of f_d across all weeks.

Figure 3.6 shows the trends of the dimensions over time. We excluded from the analysis those dimensions that did not perform well in the cross-domain adaptation of our models (Figure 3.5). For the sake of comparison, we report also the z-score of the sentiment score calculated with VADER. All plots are marked with four significant events in Enron's history: *i)* the beginning of widespread concerns about the financial stability of the company; *ii)* the first round of layoffs; *iii)* the start of financial losses; *iv)* the declaration of bankruptcy. The picture traced by sentiment analysis marks an overall, steady downward trend that reaches its lowest level by the time financial losses were made official. The conversational dimensions, on the other hand, reveal a richer picture that matches the known stages of the company's downfall [McLean and Elkind, 2013]. First, as the initial concerns sparked, the exchange of status and support plummeted: panic started to spread and employees stopped celebrating their achievements, thanking each other, and offering comfort. About three months later, the frequency of knowledge exchange dropped sharply: as concerns grew, employees spent less time in dealing with their everyday duties. A few weeks before the layoffs, as it became clear that many employees would have been made redundant, conflict exploded and the power structure collapsed—fewer orders were given to the angry crowd of employees who were made aware of the impeding jobs cuts. In the aftermath of the layoffs, those who managed to stay in the company gave support to each other for a few weeks before the imminent crack.

3.4.4.2 Movies

Movie dialogues present dense and relatable narratives. Often the story and background of characters is laid out to the audience, which makes it easy to interpret their interactions. This motivated us to manually inspect some lines extracted by our machine learning tool. We ran our models on all lines from the movie script corpus, sorted them by confidence scores, and reported the top three for every dimension.

3.4.5 Predicting community outcomes

We saw that the 10 dimensions can be captured from conversations between pairs of people and reflect their relationships. We then tested whether the presence of those dimensions in conversations is associated with real-world outcomes at community-level. We expect to find such a connection because language is more than a mere communication medium. The words we use effectively reflect and change the reality around us [Green, 1996], and the words that are used collectively by a community reveal the social processes associated to its thriving or decline. Since our Reddit data comprises of messages written by users that are geo-referenced at US State-level (§3.2.1), we conducted a geographical analysis to study the relationship between the presence of the 10 dimensions and socio-economic outcomes. We set out to test three hypotheses:

H1*: Knowledge and education.* People with higher degrees have higher language proficiency [Graham, 1987] and are more likely to access and contribute to technical content online [Glott et al., 2010, Thackeray et al., 2013]. We hypothesize that US States with higher exchanges of knowledge are associated with higher education levels.

H2*: Knowledge and wealth.* Social networks in which knowledge is exchanged create innovation and technological advancements, which result into economic growth [Florida et al., 2005, Bettencourt et al., 2007]. We hypothesize that US States with higher exchanges of knowledge are also associated with higher per-capita income.

H3*: Trust, support, and suicides*. People affected by depression, especially those who have suicidal thoughts, do not tend to trust their peers [Gilchrist and Sullivan, 2006, Shilubane et al., 2012, Cigularov et al., 2008], and seek social support in different contexts, often online [De Choudhury et al., 2016]. We therefore expect to find high levels of social support and reduced level of trust in States with high suicide rates.

To verify these three hypotheses, we downloaded the 2017 American Community Survey statistics from the United States Census Bureau. The survey reports, for each State, the median household income and the proportion of residents with bachelor's degree or higher as a proxy for education levels. From the US Center of Disease Control, we downloaded the State-level suicide death rate calculated from the residents' death certificates.

We ran our classifiers on every sentence of all the ~160M posts and comments published by the ~1M of Reddit users for which we estimated their State of residence. Similar to the analysis of Enron emails, we marked each text with dimensions d whenever the confidence of model M_d exceeded the threshold of 0.95 for at least one sentence in the text. Last, we estimated the prevalence of a dimension d in a State as the number of posts labeled with d normalized by the total number of posts in that State.

We ran a linear regression to estimate each of the census indicators from the State-level prevalence of the 10 dimensions. As a control factor, we added population density, which is associated with a number of socio-economic outcomes [Bettencourt, 2013]. Overall, our hypotheses were confirmed (Table 3.6). *Knowledge* is the strongest significant predictor of education levels and income. Presence of *support* and absence of *trust* are the two most important predictors of suicide rates. As expected, population density alone is a good proxy for all the outcomes (urban areas are richer and more educated, with fewer cases of suicide). Yet, adding the conversational dimensions to the density-only baseline yields an absolute R^2_{adj} increase between 0.25 to 0.52; with all the factors combined, all R^2_{adj} exceed 0.7. Figure 3.7 displays the linear relationship between the outcome variables and the strongest predictors in the three regressions.

A few other significant predictors emerge beyond what we hypothesized. States with higher education exhibit lower levels of *conflict*. This is consistent with studies that found that hate speech is fueled by low education levels [Gagliardone et al., 2015]. Wealth is associated with a reduced number of expressions that point out similarities between points of view, which might be a sign of communities that are structurally and culturally diverse [Cummings, 2004, Lee and Nathan, 2010]. Suicide rates are higher in States with fewer expressions of *identity*, in line with previous studies that found an association between lack of sense of belonging and risk of depression-related suicides among young people [Proctor and Groze, 1994].

	Education		Income		Suicides	
	β	SE	β	SE	β	SE
intercept	.111	.009	.233	.099	.228	.109
Knowledge	**.554*****	.172	**1.140*****	.192	.219	.211
Power	.187	.159	-.209	.177	.004	.195
Status	-.217	.199	.150	.222	.054	.244
Trust	.309	.205	-.050	.223	**-.768*****	.251
Support	.278	.238	.134	.099	**1.103*****	.291
Romance	-.247	.118	-.182	.133	-.044	.145
Similarity	-.496	.191	**-.597*****	.214	-.113	.234
Identity	**.224***	.126	-.053	.141	**-.333****	.154
Fun	.191	.000	-.127	.169	.027	.185
Conflict	**-.300****	.115	-.211	.127	**.280***	.141
Pop. density	**.433*****	.080	**.731*****	.090	**-.614*****	.098
R^2_{adj}	**.782** (+.522)		**0.774** (+.334)		**.707** (+.253)	
Durbin-Watson	2.202		2.134		2.390	

Table 3.6: Linear regressions that predict real-world outcomes (education, income, suicide rate) at US-State level from the presence of the 10 dimensions in the conversations among Reddit users residing in those States. Population density is added as a control variable. Absolute R^2_{adj} increments of the full models over the density-only models are reported in parenthesis.

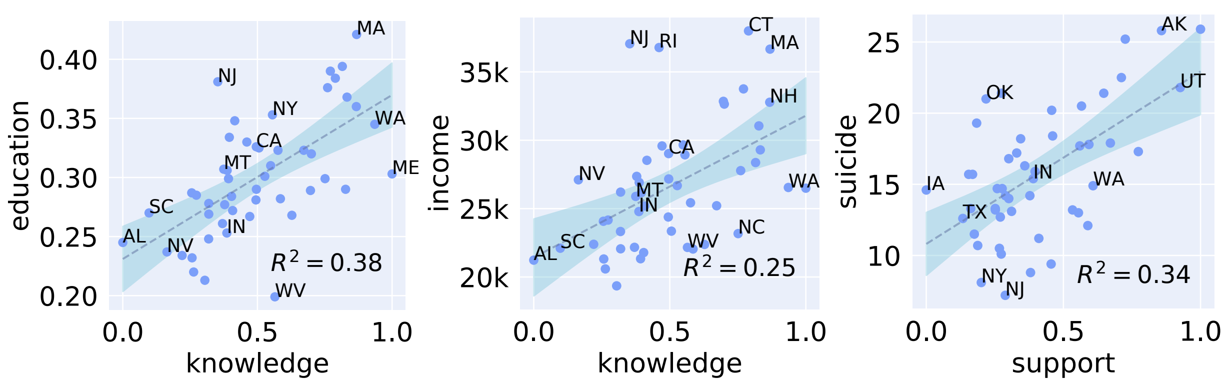


Figure 3.7: Linear relationships between each US-State outcome variable (education, income, suicide rate) and its most predictive social dimension (min-max normalized). Plots are annotated with a few representative US States.

3.5 Discussion

3.5.1 Results and implications

Starting from a unified theory that identifies the fundamental building blocks of social interactions, we collected data to associate these building blocks with verbal expressions, and we trained a deep-learning classifier to detect such expressions from potentially any text. Our tests obtained high prediction performances, showed that our tool correctly qualified the coexistence of different social dimensions in individual sentences and ascertained that the presence of certain dimensions is predictive of real-worlds outcomes.

From the theoretical standpoint, our work contributes to the understanding of how some of the fundamental sociological elements that define human relationships are reflected in the use of language. In particular, we discovered that all the 10 dimensions are represented abundantly in everyday conversations (albeit not equally), and that the way they are expressed can be learned even from a small number of examples. In practice, the data we collected and the classifiers we built could contribute to creating new text analytics tools for social networking sites. In particular, we believe that the dynamics of a number of processes mediated by social networks (including diffusion, polarization, link creation) could be re-interpreted with our application of the 10 dimensional model to conversation networks. To aid this process, we made our code and crowdsourced data available[1] and encourage researchers to experiment with it, while considering the limitations we cover next.

3.5.2 Limitations

Our approach has limitations that future work will need to address.

Data biases. The data sources we used suffer from a number of biases. Our classifiers are trained on a restricted datasets from a single source (Reddit), made of texts posted by US residents, and labeled by annotators from English-speaking countries. As a result, some dimensions were underrepresented in the labeled data. A larger data collection with reduced socio-demographic, cultural, and linguistic biases is in order. We focused on phrases containing 1^{st} or 2^{nd} person pronouns and considered online conversations only; we did not test our tool on conversations happening offline.

Models. Our models do not take into account important aspects of social interactions. First, they do not account for directionality. For example, a sentence classified as *support* could either contain expressions of social support that the speaker is giving to others as well as the acknowledgment that others have provided support to the speaker. Second, we performed training focusing only on the sentences labeled by annotators, and not on the surrounding context. As a result, our models might

[1] https://social-dynamics.net/projects/social_dimensions

fail to grasp the broader context around a phrase, which, for example, resulted in their inability to detect sarcasm.

Exhaustiveness of the 10 dimensions. The theoretical model we operationalized is not meant to exhaustively map all the possible elements that define social interactions. Yet, the 10 dimensions summarize key concepts that have been extensively studied over decades in social and psychological sciences. Therefore, our analysis is comprehensive in that it includes the most frequent dynamics of interpersonal exchange. However, one might wonder why roughly 40% of text samples could not be clearly labeled with any dimensions by the annotators (§3.4.1). To investigate this aspect further, we manually inspected a sample of those instances. We found that, except a few instances of spam-like messages and false negatives, most sentences contained personal opinions on a matter (e.g., *"My concern with this scenario is that she assumes that you would be into it."*) or trivia (e.g., *"My chinchilla attacks the vacuum the same way your rabbit attacks the broom"*). These are, to some extent, soft expressions of knowledge exchange or social support. In short, not all conversations convey a meaningful and clearly identifiable social meaning; a good part of it is generic chatter. Although we did not find any striking evidence that would point towards a need to revise or expand the underlying theoretical model, we still believe that further investigation across multiple datasets and scenarios is required. In conclusion, the ten dimensions might not be orthogonal and exhaustive representations of conversational language, yet we found that they express a very high descriptive power.

CHAPTER 4

Interpersonal Relationships in Twitter

4.1 Introduction

Dyadic relationships between individuals are a fundamental characteristic of online social networks such as Twitter. For relationships, concepts such as tie strength [Granovetter, 1973], signs [Leskovec et al., 2010a,b], and direction [Foster et al., 2010] have been used to study communicative behaviors such as reciprocity [Cheng et al., 2011], topic diffusion [Romero et al., 2013b], and echo chambers [Colleoni et al., 2014]. Individuals in these networks are largely organized around social structures such as work, neighborhood, or families [Feld, 1981, 1982], forming *interpersonal relationships*, such as friendships, kinship, and romantic partnerships. These interpersonal relationship types can influence communication and behavior in the network—e.g., consider what information might be shared between friends versus with a parent. Knowing and inferring relationship types in a social network can have several implications, such as directing messages to the appropriate social audience [Ranganath et al., 2015], improving information diffusion models, and detecting social communities [Tang et al., 2012]. However, due to a lack of data availability, interpersonal relationships have rarely been considered for these tasks in social network research.

In this paper, we aim to close this gap by inferring interpersonal relationship types from dyadic interactions in online social networks. Indeed, several studies have tried to classify relationships in domains such as phone call logs [Min et al., 2013], chatroom conversations [Tuulos and Tirri, 2004], and conversation transcripts across messaging platforms [Welch et al., 2019]. While these studies show predicting relationships is possible for private exchanges or niche topic-based communities, general social media lead to additional challenges due to the substantially higher diversity in content and relationship types. Also, while there has been prior work directly aiming to predict relationships from Twitter data [Adali et al., 2012], the predicted categories are data-driven clusters that do not directly correspond to known social relationship types. Yet, as we will show, interpersonal communication still contains linguistic signals that reveal social relationships, enabling accurate prediction.

Category	Examples
Social	best friend, neighbor, roommate
Romance	dating partner, spouse, fiancé
Family	parent, child, aunt
Organizational	manager, colleague, pastor
Parasocial	idol, fan, hero

Table 4.1: Examples of relationship types per category

The contributions of this study are as follows. First, using a massive dataset of interactions between 9.6 million Twitter user dyads with labeled relationships, we conduct an extensive analysis of linguistic, topical, network, and diurnal characteristics across relationship categories. We show that relationships on Twitter follow existing theories of interpersonal relationships and reveal complex social dynamics. Second, we introduce a neural network model for classifying five relationship types from linguistic and network features, achieving an F1 of 0.70, which substantially improves upon a strong classifier baseline (0.55) and random guess (0.20). Finally, we show that knowing the type of relationship improves performance on the challenging task of predicting whether one user will retweet another's message, improving the F1 by 1.4% for tweets that do not contain URLs and 2.0% for tweets that do, and highlighting the benefit of modeling the interaction between relationship and content. A pretrained version of our model is publicly available[1].

4.2 Interpersonal Relationships

Interpersonal relationships between Twitter users can be broadly grouped into five categories: *social*, *romance*, *family*, *organizational*, and *parasocial*. These categories, based on prior theory from communication studies and sociology, cover the social relationships studied in both offline [Knapp et al., 1980, Feld, 1982] and online settings [Ozenc and Farnham, 2011]. Examples for each category are included in Table 4.1.

Social Peer relationships and friendships are often the most common relationship in one's social network [Gorrese and Ruggieri, 2012]. Characteristics include high levels of reciprocity [Hartup and Stevens, 1999], a wide range of shared topics [Hays, 1984] and homophily [Rivas, 2009]. Strong ties include close friends who provide emotional support [Richey and Richey, 1980], while weaker ties such as acquaintances or neighbors can help build connections and obtain information [Granovetter, 1983]. Several studies on online social networks have focused on the interactions of social relationships [Ellison et al., 2013, Lee, 2009, Burke and Kraut, 2014].

[1]https://github.com/minjechoi/relationships

Romance Romantic relationships are central to adult life, leading to opportunities for intimacy and support [Hartup et al., 1999]. They exist in various stages such as dating, engaged and being married [Stafford and Canary, 1991, Knapp et al., 1980], and can develop into the formation of new families. These relationships are often considered the closest ties. Previous work introduced methods to classify romantic relationships in online social networks based on their network properties [Backstrom and Kleinberg, 2014] and conversation content [Tay et al., 2018].

Family Family relationships are essential for building personalities and receiving social support. Though maintained throughout lifetime, their importance may decline and are partially replaced by social and romantic ties during young adulthood [Shulman, 1975, David-Barrett et al., 2016]. This is reflected in contact frequency, diversity of activities and influence strength, which are lower than romantic relationships and similar to friend relationships [Berscheid et al., 1989]. Topics shared between family relationships in online social networks typically include advice giving and household issues [Burke et al., 2013].

Organizational Relationships are also formed as individuals join organizations and are assigned roles within them [Sluss and Ashforth, 2007, Marwell and Hage, 1970]. Organizational relationships are a mixture of personal and role relationships [Bridge and Baxter, 1992]. This dual status leads to a stronger notion of a community or group identity [Klein and D'Aunno, 1986] and a lesser sense of trust and solidarity compared to friend relationships [Myers and Johnson, 2004]. Information exchange and politeness are expected in conversations [Sias et al., 2012].

Parasocial The final relationship category is highly asymmetrical, consisting of celebrity-fan relationships [Garimella et al., 2017], involving high levels of affection from one side, resembling friendship or romantic relationships [Kehrberg, 2015]. Parasocial relationships are especially important to study in social networks such as Twitter, as influential figures with millions of followers can influence which topics go "viral" [Suh et al., 2010, Stever and Lawson, 2013].

4.3 Extracting Relationships

In order to construct a ground truth set of dyads with labeled relationships, we use self-reported relationships between users where a user declared their relationships to another user in a tweet. A similar strategy has been used for extracting social roles from tweets [Bergsma and Van Durme, 2013, Beller et al., 2014]. We now describe the full procedure.

We begin with a 10% sample of all tweets posted between 2012 and 2019 and remove all non-English tweets using `pycld2`. We search for all instances of the phrase 'my **REL** @username',

Category	Dyads	DM	PM	RT
Social	6.6M	81M	23.9M	47M
Romance	2.3M	36M	11M	20M
Family	324K	3.4M	945K	1.7M
Organizational	92K	419K	316K	470K
Parasocial	360K	4.1M	3M	4.1M
Total	9.6M	125M	39M	74M

Table 4.2: Statistics of dyad pairs and tweet interactions in the final dataset for directed mentions (DM), public mentions (PM), and retweets (RT).

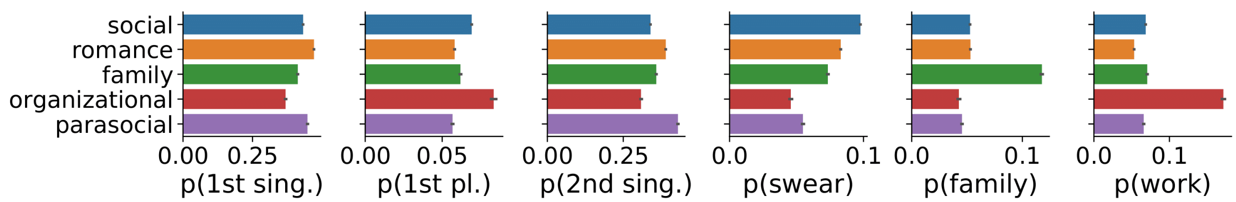


Figure 4.1: Probability of containing a LIWC-category word in a directed mention toward a specific relationship type. Romance and parasocial relationships express high levels of self-disclosure by using more singular pronouns, while organizational relationships use more plural pronouns to show collective identity. Swearing is most common among social and least common within organizational relationships, possibly due to differences in social distance. Work- and family-related words are associated with the respective relationship categories. Here and throughout the paper, error bars denote bootstrapped 95% confidence intervals.

with **REL** being any string of up to three words. Through this search, we capture public relationship declarations such as "My dear husband @username...". Since this will also capture many phrases that do not correspond to a relationship, our next goal will be to filter out such instances.

First, all phrases occurring <1,000 times in the dataset are removed, as we observed that most of the low frequency terms do not correspond to relationships. This process leaves 1,298 phrases that potentially map to a specific relationship. Next, every phrase was assigned to one of the five relationship categories or labeled as invalid through a two-step annotation process by the authors. For each phrase, annotators were shown 50 phrases and asked to choose which categories (up to two) the phrase belonged to, if any. To aid the annotation task, for each phrase, annotators were given five relationship-signaling tweets that used that phrase. Inter-annotator agreement score was measured by averaging the pairwise Fuzzy Kappa score [Kirilenko and Stepchenkova, 2016], which allows for multiple categories selected per item. Annotators obtained a κ=0.69, indicating high agreement. Given the high agreement, the remaining phrases were equally distributed across annotators without overlap.

After annotation, phrases assigned either to zero or more than one relationship category were discarded (see Supplemental Material 1 for details)[2]. Ultimately, 508 phrases were assigned to a single relationship category. These phrases were used to label 9,672,541 relationships between 10,410,262 users. Tweets were then collected for all 10.4M users from our 10% sample from 2012–2019, totalling 238M tweets.

Three types of tweets are used in our analyses, which represent different types of interactions between users. (1) **Directed mentions**: Tweets where a user directs a message to a specific user by adding the username at the beginning of the tweet, typically for starting a conversation or replying to another user. While the mentioned user is notified that they were mentioned, this tweet does not appear on the posting user's timeline. (2) **Public mentions**: Tweets visible for the public audience where a username is mentioned in the middle of a tweet. Public mentions are typically used to refer to other users, but not necessarily to have a conversation. (3) **Retweets**: Instances where a user is broadcasting a tweet posted by another user. The number of dyads, directed mentions, public mentions, and retweets for each relationship category are shown in Table 4.2.

4.4 Behavioral and Structural Differences in Relationships

To test the quality of extracted relationships, we test communicative and network patterns in each type, validating our data using predictions from known trends for specific relationships [Burke et al., 2013, Ellison et al., 2007].

4.4.1 Linguistic Preferences

Linguistic style and content reflect how an individual perceives another [Bell, 1984]. For instance, usage of pronouns reveal levels of self-disclosure [Choudhury and De, 2014, Wang et al., 2016b], and swearing terms indicate a closer social distance between the speakers [Feldman et al., 2017]. Comparing the use of these words by relationship category can reveal how open each relationship types are in Twitter conversations. Using lexicons from LIWC [Pennebaker et al., 2015], we calculate the probability of a directed mention containing one of a specific set of words: 1st person singular and plural pronouns, 2nd person singular pronouns, and swearing terms. Assuming that there exist topics central to a single type of relationship such as work-related topics, we also include the LIWC categories for work- and family-related words. The results are displayed in Figure 4.1.

Communication patterns match prior expectations, as illustrated in three trends. First, conversations in organizational relationships focus on collective identity, as shown in the highest probability of 1st-personal plurals, but lowest in 1st-person singular, as individuals in these relationships associate each other in the context of a larger collective entity [Klein and D'Aunno, 1986]. Second,

[2] Please refer below for the supplemental material

parasocial relationships use lesser 1st-person plural pronouns but more 1st- and 2nd-person singular pronouns, making their behavior similar to that of romantic relationships. This result is consistent with previous findings on behaviors in parasocial relationships that resemble love and affection due to the intense focus on the higher status individual [Tukachinsky, 2010]. However, parasocial conversations also contain substantially fewer swear words compared to romantic relationships, reflecting higher perceived cost of social norm violation due to the relationships larger social distance [Fägersten, 2012]. Also, consistent with previous findings stating the positive relationship between profanity and social distance [Feldman et al., 2017], swear words appear most commonly for social relationships, followed by romance and family relationships. Finally, the figure also reveals that work- and family-related words match folk expectations: organizational and family relationships are the most likely to use their respective LIWC categories, underscoring the topical differences between relationships. The most frequent five words for each LIWC category (shown in Table 2 in Supplemental Material) confirm that these topical-relationship interactions are not primarily driven by a single word in a category.

4.4.2 Topical Diversity

Do some relationships talk about more diverse topics than others? Social penetration theory predicts that the variety of topics shared in conversations should increase as relationships further develop [Altman and Taylor, 1973]. We test this prediction using a topic model to analyze the diversity in communication. Following prior work [Quercia et al., 2012], we measure topical diversity using the entropy of the message's distribution of topics from a trained LDA model. A 100-topic LDA model is fit using Mallet [McCallum, 2002] on a sample of 100K dyads balanced across the five categories and using five tweets per dyad to control for differences in communication frequency. To measure diversity we calculate entropy over the mean topic distribution per dyad, then aggregate by category.

The observed topical diversity (Figure 4.2) matches predictions from social penetration theory, with more diverse topics (higher entropy) seen in relationship categories that are more likely to contain deeper relationships with stronger ties and to have developed further such as romance, social, and family. In contrast, organizational relationships are less likely to communicate on topics outside their common ground [Marwell and Hage, 1970]. These results were consistent over several runs with topic models trained on 20 or 50 topics.

4.4.3 Network and Communication Properties

Given their different functions, relationships are expected to differ in their network and communication proprieties. We test for these differences by examining the labeled dyads within the larger social

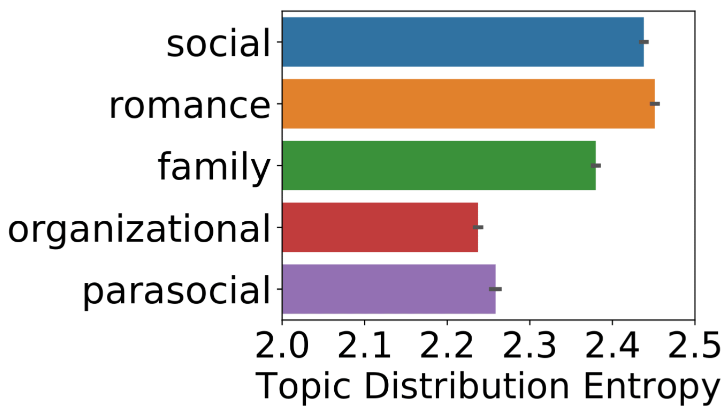


Figure 4.2: The average entropy of topic distributions obtained from directed mention tweets. The entropy is significantly higher for social and romance relationships, which shows these relationships contain more topics in their conversations.

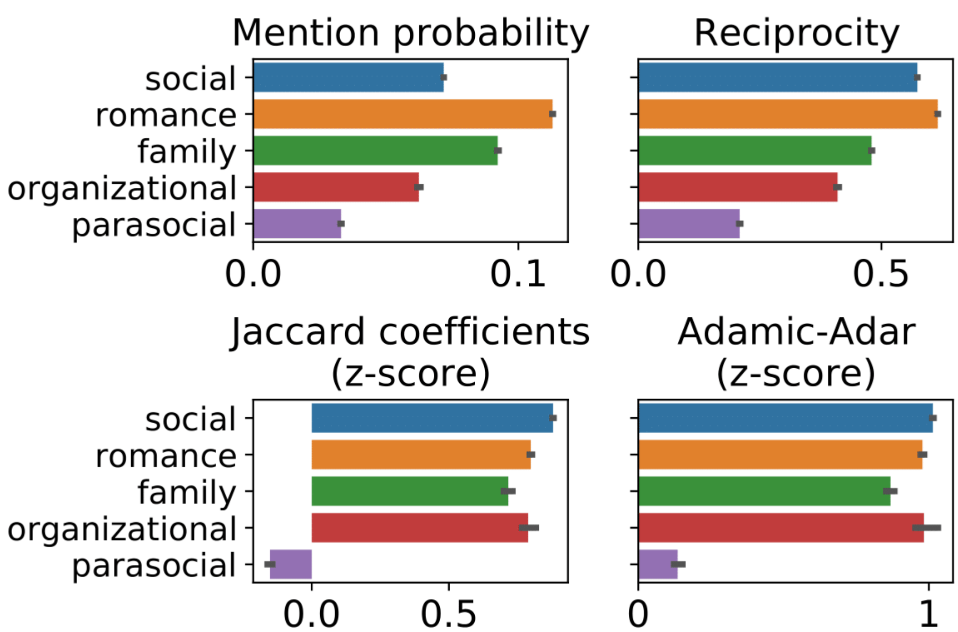


Figure 4.3: Network and communication features. Jaccard and Adamic-Adar scores are lowest for parasocial relationships, indicating a low similarity in neighbors of a dyad. Romance has both the highest mention probability and reciprocity, signalling the strongest level of mutual communication.

network constructed from our entire 10% Twitter sample. Here, two users have an edge if they both mention each other at least once, which results in a network with ~1.1B edges. The dyads with labeled relationships represent a small fraction of this comprehensive social network, as the majority of dyads have not declared a relationship. All the network properties for the dyads in our study are measured according to their users' statistics in this larger network. Using this network and directed tweets between two people, we consider two aspects of a relationship: (1) communication frequencies and (2) the local network structure around a relationship.

4.4.3.1 A comparison of communication frequencies across relationships

Communication frequencies are measured using (a) the probability of a user tweeting to the other in a relationship, relative to all others in their ego network, and (b) the reciprocity in communication, measured as the ratio of tweets between two users, scaled to [0,1] where 0 indicates only one person tweets another and 1 is equal communication. We denote $\Gamma(u)$ as the set of neighbors of user u, and $\mathbf{m}_{u\rightarrow w}$ as the number of times user u mentions another user w. The probability of mentioning a specific user out of all possible neighbors is obtained as

$$\frac{\mathbf{m}_{u\rightarrow v}}{\sum_{w\in\Gamma(u)} \mathbf{m}_{u\rightarrow w}}.$$

We also compute the reciprocity between two users as the fraction of communications each user has made, denoted as

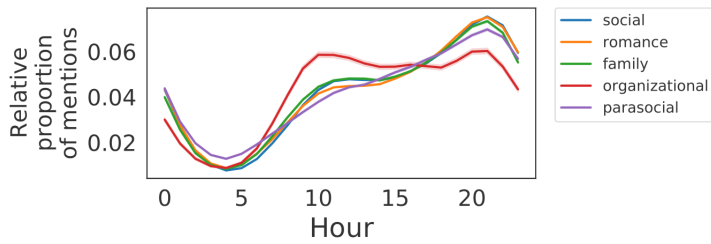
$$2\times\frac{\mathbf{min}\left(\mathbf{m}_{u\rightarrow v},\mathbf{m}_{v\rightarrow u}\right)}{\mathbf{m}_{u\rightarrow v}+\mathbf{m}_{v\rightarrow u}}.$$

A score of 1.0 means a fully reciprocal dyad with both users communicating equally, and 0 a fully imbalanced dyad where only one mentions the other. To ensure that a relationship is valid, we only calculate reciprocity using dyads where each user has made at least one interaction with the other.

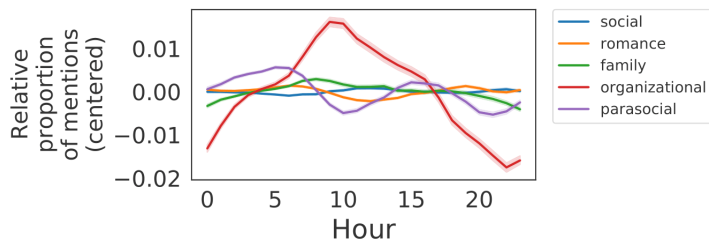
Communication frequencies, which are presented in the first row of Figure 4.3, exhibit clear differences across categories. Individuals prioritize communication within romantic relationships, consistent with prior work [Burton-Chellew and Dunbar, 2015], and have the highest reciprocity. High reciprocity implies two people have similar social status [Verbrugge, 1983]; this behavior is seen most in categories likely to be between peers: romantic and social relationships. Reciprocity levels follow expectations for differences in social status within each category, with the highest-distance parasocial relationship having lowest reciprocity.

Reciprocity relates to status difference, where high reciprocity in contact frequency implies two users belonging to similar social statuses [Verbrugge, 1983]. While relationships of social and romance categories are genuinely considered as equal in status, other categories contain non-reciprocal relationships such as parent-child relationships or manager-subordinate relationships, which can explain the lower scores for family and organizational categories. Reciprocity is lowest again in parasocial relationships, showing there exist large status differences between celebrity-fan relationships.

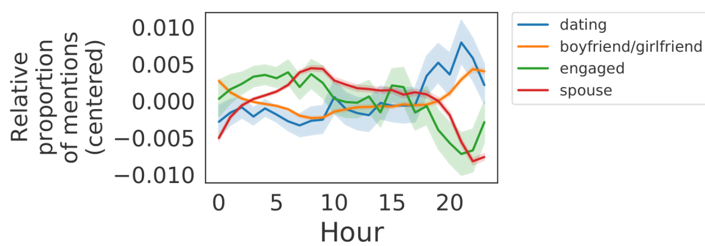
By observing mention probability or the likelihood to get mentioned instead of all neighboring users, we can see that romance has the highest level of relative importance by far, while, surprisingly,



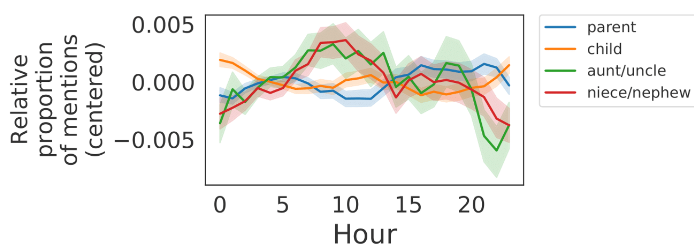
(a) Raw frequency, category-wise



(b) Normalized frequency, category-wise



(c) Specific to *romance*



(d) Specific to *family*

Figure 4.4: A comparison of mention frequency across hours of day reveal striking difference in temporal dynamics between relationship categories (a,b) and subcategories (c,d) where (b), (c) and (d) are centered relative to the mean temporal distribution across all relationship categories: (a) The un-centered communication frequency among categories (b) the centered communication frequency among categories (c) The centered communication frequency for four *Romance* subcategories (d) The centered communication frequency for four *Family* subcategories. Shaded regions show 95% bootstrapped confidence intervals.

social relationships drop to the level of other types. This result is consistent with findings showing that individuals prioritize communicating with their romantic partners over other relationships [Roberts and Dunbar, 2011].

4.4.3.2 A comparison of network properties across relationships

We also consider two types of network properties: (a) the Jaccard Index of the two users' friends and (b) Adamic-Adar index [Adamic and Adar, 2003], both frequently used for measuring the likelihood of an edge between two nodes. To allow for direct comparisons among dyads, we use the z-normalized score for each metric instead of the raw score as $zscore(u) = \frac{x-\mu}{\sigma}$, where x is the raw score, μ and σ are the mean and standard deviation computed from the neighboring dyads of u other than v. Supplemental Material §3 provides a longer explanation of how these values are computed.

The patterns in network structure (Figure 4.3 bottom) also match expectations. First, parasocial relationships exhibit very low Jaccard index and Adar-Adamic scores. This is expected as celebrities are embedded in very different social structures from that of their fans and do not have many connections in common. Second, the *family* category has significantly lower Jaccard index and Adar-Adamic score than the social, romance, and organizational categories. This is likely due to two reasons: First, unlike social, romance, and organizational relationships, family relationships do not depend on network structure. Indeed, a family relationship is established regardless of social proximity. However, social, romance, and organizational highly depend on social proximity as these relationships are established through social mechanisms such as friends introducing their friends to each other (i.e. triadic closure) [Kossinets and Watts, 2006]. Second, family ties tend to be well embedded within family networks (e.g. siblings may have other common family connections), but also tend to be much smaller than other relationships such as social and organizational. Due to these differences in volume, family ties are overall less embedded.

4.4.4 Diurnal Communication Patterns

Individuals manage their communications differently according to social relationships or "identities" such as *social*, *work* and *family* [Ozenc and Farnham, 2011, Min et al., 2013]. By observing diurnal Twitter usage patterns through the timestamps of tweet messages [Golder and Macy, 2011], we show the existence of both between- and within-category communication differences across relationship types.

Using our massive volume of communications, we compute a distribution representing the fraction of messages exchanged for each relationship dyad during each hour of a day. With the number of mentions from user u to user v as $\mathbf{m}_{u\rightarrow w}$, we bin each mentioning tweet according to the hour of the day it was created. We then define $t_{u\rightarrow w}(i)$ as the fraction of mentions produced in the i-th hour so that $\sum_{i=0}^{23} t_{u \to w}(i) = 1$. We restrict our analysis to tweets where a local timezone of the tweeting user is provided along with its global timestamp and convert the tweet to its local time. Also, we only consider cases where the sending user has made at least 5 activities for better smoothing of the distribution.

After we compute the diurnal communication distributions for each dyad, we can aggregate across different relationship categories or subcategories to obtain category-wise diurnal distributions. We provide a comparison of the diurnal distributions aggregated across different categories, shown in Figure 4.4(a). While all categories share the same pattern of substantially lower communication during dusk and peaks around evening, similar to previous work Golder and Macy [2011], we can observe slight differences around daytime. To further examine such differences, we center the distributions by subtracting each with a global mean

$$t_{global}(i) = \frac{1}{|S|} \sum_{j=1}^{|S|} t_{(u \to v)_j}(i), \left\{(u \to v)_j \in S\right\}$$

with S as the set of all dyads across all categories, or in Figures 4.4(c) and (d), the different subcategories we consider. The centered result (Figure 4.4(b)) shows a notably higher communication rate for the *Organization* category during work hours (9-16) which drops afterwards, possibly due to moving away from a work activities and towards friends and family chatter [Farnham and Churchill, 2011].

Due to our data scale, we can examine communication patterns *within* the same relationship category. The trends for the *Romance* category, shown in Figure 4.4(c), reveal that even within romance relationships, diurnal patterns can be partitioned into early ("dating" and "boyfriend/girlfriend") versus stable ("engaged" and "spouse") stages, where the former communicates more during late hours. One possible reason is that the latter group may consist of more married couples that share the same physical space during the evening and have fewer reasons to communicate through Twitter. Within family relationships (Figure 4.4(d)), we show that aunt/uncle - niece/nephew communication is more intense during the day compared to parent-child communication. We conjecture that this tendency reflects the lesser degree of perceived closeness between extended families as opposed to direct kin, restricting communication during late hours.

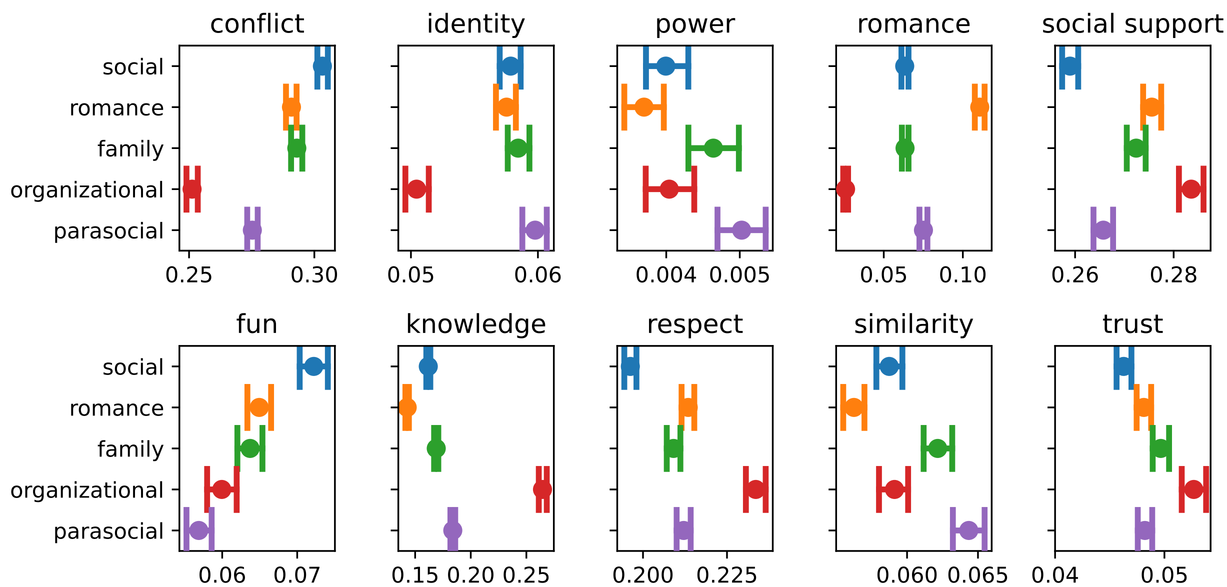
4.4.5 Levels of Social Dimensions

Previously in Study 1, we examined the possibility of identifying different properties of social relationships by identifying and measuring ten different types of social dimensions. By measuring the levels of social dimensions exchanged through conversations in Twitter, one can examine whether the prevalence of dimensions differ by interpersonal relationship categories in Twitter as well.

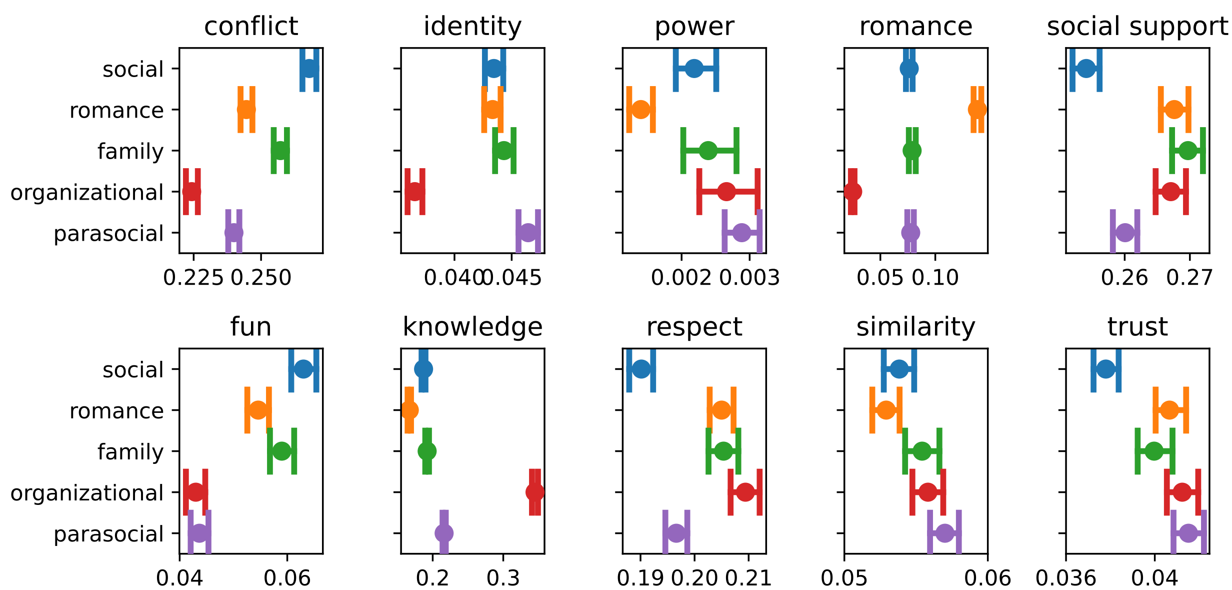
For each of the five categories we identified all dyads that had at least five total interactions between the two users in a dyad. Next, we sampled 10,000 dyads per each category and aggregated all direct- and public-mention tweets exchanged between each dyad. We then ran every single tweet with each of the ten BERT classifiers from Study 1 to measure dimension strength at an individual tweet level. This results in ten different probabilistic scores between 0 and 1 for each tweet interaction, with each score corresponding to a single dimension.

Figure 4.5 displays differences in the average dimensional strength per relationship category for each of the ten dimensions. In Figure 4.5(a) which contains the scores measured from direct mention and reply tweets, we can observe heterogeneity across all of the dimensions, where the *organizational* category stands out across several dimensions. For instance, messages exchanged between organizational ties contain the greatest amount of social support, knowledge transfer, respect and trust, while containing the least amount of conflict, identity-related language, and romance-related language. These results partially reflect the higher standards of professionalism required for communication in workplace environments, where greater levels of formality and informativeness are typically expected [Roberts, 2010]. We also observe results consistent to our understanding of the social relationships, such as romance-related language being highest for *romantic* relationships and fun-related language appearing most in the *social* category. Figure 4.5(b) presents the results measured on public mention tweets, where we observe similar patterns in general to Figure 4.5(a) with slight variations. Overall, a measurement of social dimensions on the social relationship categories confirms our belief that the extracted social relationship categories do contain strong signals of variation that align with language that aligns with our expected behaviors of those specific relationship types.

Do dyads in the same relationship category get clustered in terms of similarity in dimensional strength? Considering each dyad's 10-dimensional vector as a representation, we run a t-SNE model with a perplexity of 30 to plot the distances between all dyads in a 2-D space, examining whether dyads of the same relationship type are clustered together (Figure 4.6). Our results are somewhat mixed. While all relationship categories are generally dispersed throughout the 2-D space, we do observe a strong prevalence of dyads in the organizational category in the left quartile of the figure, as well as a small concentrated region of dyads in the parasocial category which can also be observed in the leftmost part of the figure. For social, romance, and family ties, we do not observe such concentrated regions and instead can find them intertwined across most regions. One potential interpretation of these results is that compared to social, romance, and family categories where the amount of dimensional strength exchanged in messages can vary greatly, in organizational or parasocial relationships there is less variability and flexibility. Dyads in the latter two relationships are more likely to result in exchanging conversations only related to a limited number of topics compared to the former three categories where one might feel more comfortable sharing various



(a) Direct mentions & replies



(b) Public mentions

Figure 4.5: A comparison of social dimension strength on the tweets exchanged between different types of social relationships when measured using the ten social dimensions from Study 1. Across several dimensions (e.g. conflict, romance, fun, knowledge) we can observe results that align with our understanding of social relationships.

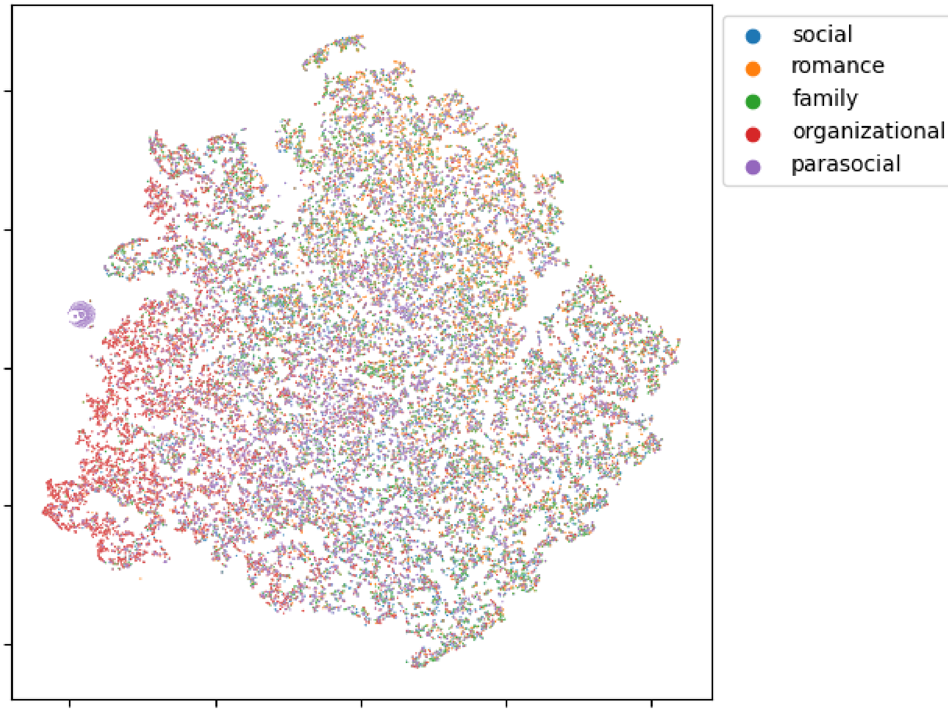


Figure 4.6: A t-SNE visualization created using the 10-dimensional average scores for each dyad. Clustered regions can be observed for both organizational and parasocial categories on the left region of the figure, but not for the other three categories which are evenly dispersed throughout the figure.

aspects of his/her life [Farnham and Churchill, 2011].

4.5 Relationship Classification

Given that relationship types differ in linguistic preferences, topic diversity, and network properties, we now test whether they can be accurately classified from these features.

4.5.1 Task and Experimental Setup

We classify a dyad into one of the five relationship categories on the basis of its behavioral and communication features. The prediction task is conducted with both balanced and imbalanced datasets. **Balanced set:** Training data uses 200K dyads per category, randomly upsampling organizational relationships which had fewer samples. A random sample of 2K dyads were used for validation data. The test set contained 17,522 dyads per category, where all classes were downsampled to match the least common class, organizational relationships. **Imbalanced set:** 2M dyads are randomly selected and split by a 8:1:1 ratio into training, testing and validation partitions. In both settings, users were constrained to be in only one partition. We ensure that at least one user of the dyad performs at least five interactions, in order to sufficiently represent a dyad's interactions. To control for differences in communication frequency, we restrict the data to at most 15 tweets from each user in a dyad, keeping up to 5 tweets for each type of communication: directed tweets, retweets, and public mentions. To avoid data leakage, all tweets that contained the phrase used for labeling the relationship type were removed prior to this process.

4.5.2 Proposed Models

To capture information from the different types of communication, we introduce a new deep learning model that performs multi-level encoding to represent these texts. Our base architecture builds upon the model of Huang and Carley [2019] using the parameters of the RoBERTa pre-trained language model [Liu et al., 2019b]. Each tweet and bio text for a dyad are encoded as constant-length vectors by encoding each with RoBERTa and mean pooling the output layer's word piece embeddings for the content. These pooled content encodings are fed into a separate 6-layer Transformer network [Vaswani et al., 2017]; this second level of encoding summarizes the different sources of text information through its attention mechanism. The output layer of this second-level model is mean pooled as a representation of all communication.

The final dyad representation concatenates the multi-level encoding with (i) character-level embeddings for the username created using 1-dimensional convolutional filters [Kim, 2014] and (ii) the four network statistics described in Section 4.4. This representation is fed through two linear layers using ReLU activation with dropout before a softmax is applied to classify the dyad. RoBERTa models were first pre-trained using 3M training set tweets; then the full classification model was trained end-to-end. Supplemental Material 4.1 provides details of all hyperparameters and training procedures.

Baselines We introduce two baselines: the first is a random guess for all samples, and the stronger second baseline combines text and network features and uses them to train an XGBoost [Chen and Guestrin, 2016] classifier. Uni-, bi-, and tri-grams with more than 10K frequency are used as features. We also add features for frequencies in category of the LIWC and Empath [Fast et al., 2016] lexicons. Network features are identical to the neural model. Details are included in Supplemental Material §4.2.

4.5.3 Results and Findings

Our model can accurately recognize different relationships, attaining a macro F1 of 0.70 on the balanced dataset (Table 4.3), indicating that relationship categories are identifiable from their

	Model	Soc.	Rom.	Fam.	Org.	Para.	F1
Bal.	*Random*	0.20	0.20	0.20	0.20	0.20	0.20
	GBT Model	0.45	0.57	0.55	0.64	0.55	0.55
	Our Model	**0.60**	**0.69**	**0.69**	**0.79**	**0.72**	**0.70**
Imbal.	*Random*	0.62	0.30	0.05	0.02	0.10	0.20
	Maj. Class	0.76	0.00	0.00	0.00	0.00	0.15
	GBT Model	0.80	0.52	0.33	0.28	0.28	0.44
	Our model	**0.84**	**0.68**	**0.51**	**0.50**	**0.38**	**0.58**

Table 4.3: Performance comparison for different settings. The F1 score is used to measure the performance in all cases. The first five columns show the F1 scores measured only from samples whose ground truth label belongs to each category as a binary classification task. The last column is the combined Macro F1 score computed as a multi-task classification task.

network and communication patterns. This performance substantially improves upon that of the XGBoost baseline and random chance. In the balanced setting, the model is most accurate at predicting organizational and parasocial and least at social relationships. We attribute this difference to the intra-class diversity; while organizational and parasocial relationships typically have narrowly-exhibited behavior (e.g., low topic diversity for organizational in Figure 4.2), social relationships can take many forms, e.g., friends, neighbors. This diversity likely makes the class harder to distinguish.

In the imbalanced setting that reflects the natural distribution of classes, our model offers an even larger performance improvement over baselines. Here, most dyads have a social relationship (76%; cf. Table 4.3), yet the model is still able to reliably identify all classes. This result highlights the applicability of the model in real-world settings, which is crucial for studying communication dynamics.

Separate ablation studies were performed to test the information from each type of communication (direct, public mention, or retweet) and for the addition of user and network features. Among communication types, public mentions provided the most information (highest performance) for predicting relationship types (0.56 in balanced setting). We speculate that a user has to include context about the mentioned user and their relationship when mentioning them in a public tweet, to provide an explanation to the audience. This information is not required in directed mentions since the expected audience is only the two users. The addition of user features and network features to text features also significantly improves performance, most notably for parasocial relationships, where adding user profiles and network features boosts the F1 score by 0.13. The significantly lower Jaccard coefficients and Adamic-Adar scores of dyads in the parasocial category (Figure 4.3) likely makes it easier to identify the relationship type, when incorporated as features. Table with full results is in Supplemental Material (Table 3).

4.5.4 Testing the Validity of Classifier Models

As a further validation of the relationship classification model, we test whether the relationships inferred by our model mirror the behavioral properties seen in the labeled relationships. A random sample of 1M dyads is collected where one user has made at least five interactions, mirroring our classifier setup (§4.5.1). We then collect mention, reply, and retweet activities made between both users of the dyad, and apply the classifiers on these new users. As temporal information is not used in the classifier but shows clear differences by relationship (Figure 4.4), we test whether the communication patterns in these inferred relationships are similar. The resulting time series from inferred users were highly correlated with those of the labeled data correlations, ranging from 0.928 to 0.947, shown in detail in Supplemental Table 5 and visualized in Supplemental Figure 1. This result indicates the model infers relationships that have highly similar behavior to the labeled data in practice (despite not being trained on these features), e.g., dyads in the organizational category focusing more of their communication during daytime and dropping in volume after work hours for both labeled and inferred data.

As a second test of validity, we examine the distribution of inferred relationships in the random sample. The resulting distribution differs substantially from that of the labeled data, as expected: Social 23%, Romance 23%, Family 14%, Organizational 5%, and Parasocial 36% (cf. Table 4.2). In the random sample we observe more Parasocial, which aligns with earlier expectations that Twitter is largely a mass media platform rather than a social network [Kwak et al., 2010]. However, unlike expectations from earlier work, we observe a significant uptick in stronger social relationships: Social and Romance together account for ~46% of the random sample. We attribute this result, in part, to the requirement that dyads in the random sample must have at least five directed tweets, which likely increases the presence of stronger ties who are more likely to talk more. Our results point to the social nature of Twitter and the need for future work to examine how all relationships are manifested on Twitter—not just those that communicate—to establish to what degree the platform now serves as a social network.

4.6 Retweet Prediction with Relationships

Retweeting is central to information spread on Twitter. Given its significance, several studies have introduced approaches to model and predict this behavior. Factors identified include: content-based features (e.g., hashtags, URLs [Bakshy et al., 2011]), network features (e.g., tie strength [Yuan et al., 2016]), and user popularity [Hong et al., 2011]. Despite all these efforts, retweet prediction remains

Model	Soc.	Rom.	Fam.	Org.	Para.
Baseline	0.61	0.63	0.60	0.64	0.61
With relationships	0.64	0.65	0.63	0.64	0.62

Table 4.4: Classwise F1 performance comparison of retweet prediction task on a balanced dataset. The presented order is Social, Romance, Family, Organizational, and Parasocial.

a notoriously hard problem [Martin et al., 2016]. Our findings in Section 4.5 suggest one potential way to improve past efforts. The fact that relationships can be predicted based on tweet conversations indicates a promising connection between the topic of a tweet and the type of relationship that would have interest in it. This leads us to hypothesize that a user's probability of retweeting depends on the interaction between the tweet's content and the relationship to the tweet's author. As such, we test whether incorporating the relationship type between two users u and v can improve the prediction accuracy for whether user v will retweet a particular tweet by u.

4.6.1 Dataset

We first select the same number of dyads per relationship category to balance across different relationship types. For each dyad we collect all retweets that occurred between the two users, but remove instances where user mentions occurred in the original tweet. While being mentioned in a tweet is a strong motivator for a retweet to occur [Jenders et al., 2013], here our goal is to understand the interplay between the content of the tweet and the relationship properties of the dyad and thus remove tweets with mentions.

We focus on a balanced prediction task where for each positive retweet that occurred between a dyad, we assign one tweet produced by the same user around the same time which did not get retweeted by the other user. As a result, 50,000 positive and negative tweets were sampled per category and split into training, validation, and test sets at a ratio of 8:1:1.

4.6.2 Models for Retweet Prediction

The models used for the prediction task are also based on a pretrained RoBERTa model for text classification.

Baseline model encodes the tweet text using RoBERTa and preserves the embedding of the first position corresponding to the [CLS] token, which is a common practice in BERT-based classification tasks [Devlin et al., 2018]. Features for the baseline model are the contents of the tweet and the number of followers (log-scaled) and the existence of a URL, both frequently used features in retweet prediction [Petrovic et al., 2011, Suh et al., 2010]. The [CLS] embedding is passed through a linear

	without URLs			with URLs		
Model	Pre.	Rec.	F1	Pre.	Rec.	F1
Baseline	0.58	0.69	0.63	0.53	0.85	0.65
With relationships	0.58	0.71	0.64	0.53	0.87	0.66

Table 4.5: Model performance (Precision, Recall, F1) at predicting retweets of messages with or without URLs. The addition of relationship types leads to an increase in the F1 score by boosting recall.

layer and is concatenated with sparse features: the log-scaled number of followers and existence of a URL. This concatenated embedding is passed through two additional layers to be transformed into a single scalar value, where we apply a sigmoid function to convert into a score between 0 and 1, with 1 indicating a potential retweet.

Relationship-aware Model extends our baseline model using a number of representations for relationship information. We first use a direct encoding of the relationship category into a 256-dimensional vector which is trained along with other parameters. The textual information of the declared phrase associated with the relationship type (e.g., "my *best friend*") is also encoded using a character-level CNN with 1-d convolution [Kim, 2014] then max-pooled, resulting in another representation vector of size 768. This number is obtained by using 256 convolutional filters each for kernels with sizes of 3,4, and 5. Finally, we add the relationship category as one-hot features in addition to the other sparse features, and concatenate them with the dense embeddings.

We set d to 768 and the learning rate to 1e-6. All models are trained with batch sizes of 16 and for a maximum of 10 epochs. We select the model with the highest validation F1 score, which is computed for every 5000 steps.

4.6.3 Results

Adding relationship information improves performance for retweet prediction, as shown in Table 4.5, which is known to be a difficult task [Martin et al., 2016]. Incorporating relationship types provides a 1% performance increase in F1 score due to an increase in recall by 2%. Here, we show separate performances for test data tweets with and without URLs; tweets with URLs are likely to have different retweet dynamics on the basis of the content in the URL (e.g., retweeting a linked news story versus a personal message) and the different social uses of retweets [boyd et al., 2010].

Analyzing performance improvement by relationship type reveals a more complex picture of improvement, shown in Figure 4.7. For tweets containing URLs, the addition of relationship information consistently improves performance, while we observe the largest performance gains

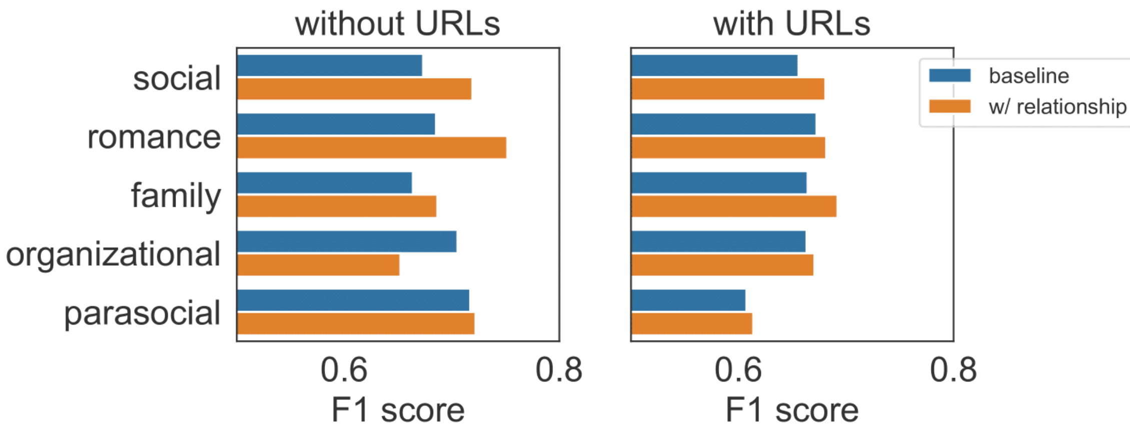


Figure 4.7: A comparison of baseline vs. relationship-infused models for retweet prediction on tweets with and without URLs. In both settings, the addition of a relationship type improves the predictive performance.

for tweets without URLs. In particular, in tweets without URLs, the model sees increases of 2.2%, 3.1% and 2.3% for social, romance, and family categories respectively, signalling that the model can use this social information to decide where a person in that relationship is likely to retweet based on the content. However, the model performs worse for predicting retweets from organizational relations (2.8% decrease) which lowers the overall performance reported in Table 4.5. In the case of tweets containing URLs, F1 scores increase across all categories, with social (3.8%) and family (4.2%) categories benefiting from the largest gains.

We observe that the larger gains seen in social, romance, and family relationships are due to increased recall (see Table 4 in Supplemental Material). This increase suggests that individuals embedded in social, romance, and family relationships retweet content that is less likely to be retweeted normally (e.g., mundane personal events) because of the nature of the relationship, which the relationship-aware model is able to use to correctly identify the content will be retweeted. This result is further evidence of the interaction between communication patterns and relationship types.

4.7 Summary of Work

Not all ties are equal: friends, family, and lovers all have different social, linguistic, and temporal behaviors—yet, social network studies have typically limited themselves to networks with edges that encode only the existence of a relationship, but ignore the *type* of that relationship. Using a dataset consisting of the interactions between 9.6M dyads on Twitter with known relationship types, we introduce a new approach that explicitly models interpersonal relationship types in social networks. We make the following three contributions towards understanding relationships in networks. First, we show that the linguistic, topical, network, and diurnal properties in online communication between different relationship types match predictions from theory and observational studies. Second, we demonstrate that relationship types can be accurately predicted using text and network features combined with state-of-the-art deep learning models. Third, we show that knowing relationship types improves performance when predicting retweets—demonstrating the benefits of predicting relationships at scale. The addition of relationship type significantly improves the recall of retweets for social, family, and romance relationships, which are considered more personal.

Our proposed approach, combined with the consistency of our results with existing literature on social relationships, further demonstrates the value in studying social media networks to further understand the differences in communicative behaviour across interpersonal relationships. Furthermore, as evident from the performance of our relationship classifier and the improvement on retweet prediction, our work enables new types of analyses that benefit from large-scale relationship-aware networks such as modeling network evolution, information diffusion dynamics, and community structure. Overall, our work provides a stepping stone towards incorporating relationship types in several research questions in social and network sciences.

4.8 Ethics and Limitations

Ethical Considerations This research was performed on only public data, in accordance with the Twitter Terms of Service. However, users do not necessarily expect their data to be collected for research purposes nor to be disclosed [Fiesler and Proferes, 2018]. Furthermore, the potentially-sensitive nature of interpersonal relationships being revealed necessitates that additional privacy steps must be taken and that benefits must outweigh harms. To mitigate risk to individuals, we report only aggregate information and focus on broad effects, avoiding any focus on marginalized groups or sensitive relationships [Townsend and Wallace, 2016]. Further, data and models will only be shared upon confirmation of ethical principles of use. Counterbalancing these risks, this study offers substantial benefit to our understanding of social processes and how relationships influence what we share and what we hear about. As demonstrated in Section 4.6, understanding these relationships can improve the algorithms that individuals come into contact with regularly, such as social and content recommendation systems.

Limitations Data used in this study relies on self-reported, public declarations, which may not occur across all dyads in Twitter. The willingness to declare such relationships is likely indicative of a stronger tie between users. Therefore, our work may not reflect the behaviors of people in relationships less likely to be declared. Such non-declarations can be due to a variety of reasons such as increased desire for privacy, weaker association of that relationship type, or even the potential social stigma around declaring the relationship. Further, our data and model depend on observing communication between two users; as not all users with a meaningful social relationship also communicate on Twitter (e.g., spouses on Twitter who talk offline), our model is unable to identify such relationship. While the results of our work are largely in line with prior expectations from sociology and psychology, future work is needed to understand what biases, if any, stem from using only self-reported relationships and how these relationships fit within the broader space of relationships that exist between users on Twitter. Finally, our validation efforts are at the population level and future work could perform additional validation on specific relationships through crowdsourced labeling of both the self-declared and the inferred relationships obtained through our model.

Our analysis is based on a 10% sample of Twitter data and while consisting of billions of tweets, this sampling by nature omits tweets between users that would affect the inference of network and communication statistics. Specifically, while the measured using the best-available data and follow past scholarship that used similar datasets to infer network properties [e.g., Bliss et al., 2014, Pierri et al., 2020], our results likely underestimate the presence of edges and rates of communication frequency due to sampling.

The current work makes two simplifying assumptions about a dyad's relationships: (i) the relationship is of only one type and (ii) does not change over time. In practice, relationships evolve over time and categories such as social and organizational can indeed overlap. Our simplifying assumptions allow us to perform these initial studies at large scale. However, future work could relax these constraints with sufficient longitudinal data or with additional self-reported data.

4.9 Related Work

Due to the difficulty of collecting ground truth data, only a handful of studies have examined real social relationships of different categories on social media. Min et al. [2013] conduct a survey on 40 participants to obtain their SMS data and categorize their contacts into *family*, *work*, and *social*. The authors show that it is possible to infer these relationship types using features such as the geographical similarity between two users and their contact patterns. More recently, Welch et al. [2019] collected private data for the ego network of one user and 104 alters where their exact relationship was known; the combination of text and behavioral data (such as message duration) was used to predict relationship properties. Other studies have predicted the existence of a specific relationship from interpersonal interactions, the most common being a romantic relationship. Backstrom and Kleinberg [2014] show that dispersion—a metric which the authors introduce for measuring how well-connected mutual neighbors are—is a better predictor of romantic relationships than embeddedness within the Facebook network. Similarly, Tay et al. [2018] introduce a model that compares the similarity of messages shared between two people on Twitter to predict whether they are in a romantic relationship. Our study goes substantially beyond these studies by simultaneously testing a comprehensive set of relationship types and examining orders of magnitude more data for each relationship type.

Rather than study categories of relationships, some studies instead measure attributes of the relationship themselves, e.g., their relative status. Adali et al. [2012] use social and behavioral features derived from Twitter activities to distinguish between relationship groups defined by their word usage and a clustering algorithm. Gilbert and Karahalios [2009] and Gilbert [2012] use interaction-based features provided by the Facebook platform to predict the tie strength of user dyads in Facebook and Twitter. Rashid and Blanco [2017] and Rashid and Blanco [2018] use a conversation dataset from the TV series *Friends* to label the relationships between the main characters with properties such as "equal-hierarchical" or "pleasure-task oriented", then show that these properties can be predicted using text features. Choi et al. [2020] predict the prevalence of specific social dimensions such as trust or power differences using deep learning classifiers on Twitter interactions between two users. These studies offer a complementary view of relationships and our introduction of a new large-scale dataset with relationship categories opens up new future work for testing how these attributes align or differ across categories.

Social networking platforms serve multiple roles as both social and informational networks Arnaboldi et al. [2013], and as such individuals may form ties with others for different purposes. Several works have examined how different properties of the dyads reflect their information sharing behavior. In particular, studies have focused on the characteristics of interactions between close users, showing that such dyads (i) have a greater tendency to share less common hashtags which may indicate community belonging [Romero et al., 2013b], (ii) are more likely to share content than with strangers [Quercia et al., 2012, Bakshy et al., 2009], and (iii) frequently engage through actions such as mentioning each other or sharing posts [Jones et al., 2013]. Other work focuses on user interactions with influential users such as celebrities or politicians who possess a large follower base. These users gain widespread influence by specializing on narrow topics [Cha et al., 2010] or posting messages with strong sentiments [Dang-Xuan et al., 2013]. While our study also aims to identify communication and interaction differences between different types of user relationships, we go beyond the widely studied themes, such as close ties or influential users, to study how specific types of relationships interact with information sharing, showing differences in which content is communicated across specific types (e.g., Figure 4.1) and that knowing the relationship type aid in predicting which content is retweeted.

CHAPTER 5

Relationship-specific Interactions toward Life Shocks in Twitter

5.1 Introduction

Not everything in life comes with preparation: People may experience sudden events such as the death of a loved one or a sudden job loss. Exposure to such events can cause adverse effects on one's mental Burton et al. [2006] and financial status Atkinson et al. [1986]. To mitigate such harmful effects that arise from experiencing unexpected life events or *shocks*, people often ask for help through their own accessible social network, which is made of ties belonging to different categories or relationships such as friends, family, or workplace relationships. The different relationships are known to provide varying levels and types of support, which contribute to the overall support network of an individual Vaux and Harrison [1985]. Here, we study behavioral differences caused by relationship types by comparing responses to shock events.

In online settings, shocks like a sudden change in a user's status can cause behavioral changes from other users in response Oh and LaRose [2016]. Prior studies have shown online shocks are associated with a variety of network-level reactions such as changes in centralization levels Zhang et al. [2017] and increased communication among closer ties Hobbs and Burke [2017], Romero et al. [2016]. Furthermore, users experiencing shock events may choose to publicly disclose it on social media, a behavior which has been increasingly studied by social researchers Haimson et al. [2018], Andalibi [2019]. However, these studies have largely assumed that all ties are equal rather than modeling the potential interaction between different relationships and shocks, and little is yet known of the relationship-specific response behaviors towards these disclosures. We pursue this open question, asking whether relationship-specific shock behavior in online platforms mirrors that of the offline world, or whether structural and normative constraints of online social networks may cause individuals to interact differently with their relationships when compared to offline settings.

In this study, we conduct a large computational analysis of responses to shocks in online social networks to test how interpersonal relationships engage. We introduce a new dataset of over 13,000

Twitter users who posted shock events along with their interactions with others, each labeled with their inferred relationships to the shocked user (Section 5.3). Using causal inference methods, we approximate the effect of experiencing and posting shock events on receiving responses from Twitter users, and how these activation levels differ both in magnitude and significance depending on both the type of relationship and the type of shock (Section 5.4). To understand the interaction between shock event types, relationships and responding behavior, we analyze how tie strength and structural embeddedness influence the users in different relationships to reply to a shock, both strongly recognized network properties for determining interactions in social networks (Section 5.5). Finally, we identify relationship-specific differences in the content of shock responses by measuring topic shift via a topic model (Section 5.6).

Our contributions are as follows. First, we demonstrate a method for identifying and extracting instances of shock experiences from Twitter posts using active learning and result in a corresponding dataset of ~13K users experiencing shocks and 179,563 users making a total of 110,540 replies to them. This data is augmented by labeling the relationship between shocked and replying users and adding a corresponding control set of users, matched for aspects such as demographics, location, and activity level. Second, through a large-scale quasi-causal analysis, we demonstrate how relationship types determine levels of responsiveness and topic shift in the responses to shock tweets that are posted online. Our findings partially align with existing theories on social relationships in offline settings; however, we also discover contrasting results that may relate to differences between online and offline settings. For instance, we observe that romance and family relationships generally respond less to shocks than social relationships, which may be due to the existence of other communication channels preferred over Twitter's public space. Third, we show that tie strength and network embeddedness each have different effects in predicting responsiveness specific to the relationship type and shock. These results point to the existence of unique social dynamics for each relationship type and suggest how individuals and their supporters can better mobilize their social networks in times of unexpected distress.

5.2 Shocks, Relationships, and Networks

Separate strands of research have examined shocks and social relationships in times of stress, creating expectations for different social ties that might interact in these events. Here, we outline the major work in each area to motivate specific research questions pursued in our study.

5.2.1 Shocks and Engagement in Social Networks

A shock is defined as an unforeseen event capable of disrupting an individual, a group, or a social network Jackson and Dutton [1988]. Prior studies have looked at how exposure to exogenous shocks such as community censorship Zhang et al. [2017], sudden price changes Romero et al. [2016] or disasters Corbo et al. [2016] affected the network's communication behavior such as contact frequency and clustering tendency.

A particular category of shocks widely studied across various academic fields is that of individuals experiencing unexpected life events during their life courses and how these events affect their social networks. Events such as the death of a family member or unexpected pregnancy can increase intra-family strain and harm well-being levels Lavee et al. [1987], which can also develop into health or depression issues Kendler et al. [1999]. Individuals are also challenged by their ability to make discrete decisions and have difficulty maintaining economic stability Shirani and Henwood [2011]. As a means of overcoming such issues, individuals may turn to their social network connections, who in turn offer informational and emotional support De Choudhury and Kıcıman [2017]. The process of support-seeking and caregiving is reciprocal in that the more stressful an event is, the more support a person seeks, which returns greater support from others Collins and Feeney [2000]. The social support provided by one's network is known to reduce stress levels and obtain adequate resources, a concept known as the *buffering hypothesis* Cohen and Wills [1985]. The importance of social support in stressful life events and the buffering hypothesis has been extensively studied in several domains, including but not limited to health studies Mitchell et al. [2014], psychology Jackson [1992], and family studies Szkody and McKinney [2019].

The social support that one may receive or provide during shocks may not be identical. Not only the nature of the shock event but also factors such as social status Smith et al. [2012] and gender Liebler and Sandefur [2002] determine the type or impact of the support given. A distinct characteristic of social networks that we will focus on is the type of *interpersonal relationship* between the support provider and receiver. It is known that a person's support system consists of various types of relationships such as spouses, immediate family, close friends, and social acquaintances Vaux and Harrison [1985], which can provide different types of support. Friends and neighbors within close proximity are capable of providing more instantaneous and instrumental support than family members, but families in turn provide more stable support that is unaffected by temporal factors Wellman and Wortley [1990]. Work-related stress can be alleviated through support from friends and workplace relationships rather than other relationships Henderson and Argyle [1985]. Overall, different relationships are capable of delivering different types and levels of support depending on the stress event.

5.2.2 Shock Responses in Online Social Networks

Researchers have increasingly been turning towards online social networks such as Twitter and Facebook for studying interactions during shock events. The formation and structure of online social networks mirror those formed offline Dunbar et al. [2015], and the abundance, as well as the accessibility of interaction data among users has made it a popular research subject. People may post support-seeking messages on networking services visible to others to receive social support Oh and LaRose [2016]. As a response to the messages, one's neighboring users may choose to send messages of support, which can lead to increased levels of well-being Burke and Kraut [2016]. Previous studies have examined the roles of users in online communities for providing support on cases such as medical issues Huh et al. [2016] or suicidal ideation De Choudhury and Kıcıman [2017]. In the context of unexpected life events, some studies have looked at changes in the structures of online social networks following events such as the death of a close friend Hobbs and Burke [2017], unemployment Gee et al. [2017], or breakups Garimella et al. [2014]. While these studies examine events similar to those considered in our work, they do not provide comparisons of behavioral differences among relationship types.

5.2.3 Research Questions

Our research identifies shock-response behaviors that occur in online social networks, with a focus on the interpersonal relationships among the users. Given different social expectations for certain relationships (e.g., friends vs. co-workers), we hypothesize that relationship type provides different degrees of support following shock events and that this behavior which has been observed in offline studies can also be found in online social networks, even if the type of support and relationships who provide it vary from offline to online. Therefore, we formulate our first research question to investigate whether the activation levels in response to a shock differ by shock, and more importantly, whether there exist differences among relationships.

RQ 1 *Does the likelihood to respond to a certain type of shock differ by relationship type?*

The closeness between two individuals is a well-known factor for determining whether support will be provided upon experiencing a shock Collins and Feeney [2000]. However, it is not well understood whether closeness is a determinant of support across different types of shocks and relationships. In addition, closeness can be measured in different ways, such as through communication frequency or structural embeddedness in a network. While these two measures are correlated, they measure fundamentally different aspects of closeness, and it is possible to experience cases such as ties that have high communication frequency but without any common connections Park et al. [2018]. Thus, we investigate the role of closeness, both tie strength and

embeddedness, in the likelihood of response across relationship and shock types.
RQ 2 *Does the degree to which tie strength and structural embeddedness affect responsiveness to shocks differ by relationship type?*

Finally, we examine the shift in the context of communication between a dyad in response to a shock. Online social networks are both platforms for obtaining and sharing information and also for maintaining contact with existing social ties Kwak et al. [2010]. While a user's social network is thus expected to contain messages of both social and informational content, exposure to a shock may rapidly shift the composition of the topics surrounding her. Specifically, the social buffering hypothesis suggests neighboring users will offer some form of support to the shocked user which comes at the expense of their original roles of sharing other types of information. We expect that during a shock, neighboring users will modify the topics of the messages that they share with the shocked user to provide support. Our third research question investigates how this shift occurs differently for each relationship and for each shock.
RQ 3 *What are the topics that increase or decrease for each relationship type following a shock?*

5.3 Dataset

To study how social behavior varies after shock events, we create and introduce a new dataset of Twitter users undergoing specific types of personal events and the social interactions they have after. Following, we describe how the shock events are identified and how the dataset is created to enable a pseudo-causal analysis of the effects of a shock.

5.3.1 Identifying Shock Events

Since our study focuses on the changes in a person's online social network that can be caused by exposure to shock events, we introduce a new data collection procedure to capture a large and accurate sample of shock instances and the network activity around the time of the shock. We begin by identifying four types of well-studied life events as shocks:

- **Romantic breakups** Romantic relationships provide a strong source of attachment that deepens over time and leads to strong levels of intimacy Hartup et al. [1999]. Ending a relationship through a breakup can lead to heightened levels of depression and anxiety Sprecher et al. [1998]. As a replacement, individuals undergoing breakups may turn towards other close members of their social networks, friends, and family members to compensate for the lost relationship Moller et al. [2003].

- **Exposure to crime** Being the victim of a crime is an uncontrollable life event that can cause negative effects on one's emotional, physical, and financial status Cutrona and Russell [1990]. Social support from others can help the shocked individual mitigate negative psychological effects caused by the incident and obtain guidance so that they can go through appropriate measures and solve potential issues Mason and Benson [1996].

- **Death of a close person** The death of a close person such as a friend or family member is, without doubt, a very stressful event that leads to strong negative emotions such as loneliness, and depression Burton et al. [2006]. Social support from others can help reduce such levels of loneliness, and people close to the bereaved can form new connections as a means of coping with grief Hobbs and Burke [2017].

- **Unexpected job loss** Unemployment can lead to a decline in one's time structure, social contacts, and activity levels Jahoda [1981]. The impact of job loss increases depending on financial situations and attachment to the job LeanaI and Feldman [1990] and the loss of friends and colleagues, harming one's social support network Morris and Irwin [1992]. Social networks can enhance an individual's mobility by providing information and various resources Podolny and Baron [1997].

While all four categories of shocks are known to cause high levels of distress and thus have the afflicted call for social support, the social groups or relationships that would respond may actually differ by shock ty pe. For instance, a colleague from work would respond differently to one's breakup versus one's sudden unemployment and choose to provide different levels of support. Additionally, being inflicted by these shock events is known to cause disruptions in one's online social network connections [Garimella et al., 2014, Hobbs and Burke, 2017, Burke and Kraut, 2013, Deal et al., 2020]. We can thus expect to identify varying degrees of interactions by relationship type, which differs by shock type.

Our approach, described next, involves a series of active learning-based filtering that combines regular expressions and deep learning text classifiers to extract a dataset of tweets describing a shock event, denoted as *shock tweets*.

To be considered as a shock tweet, we set a number of requirements that each tweet needs to satisfy:

- **Topic relevance**: The event described in each tweet should fall into one of the categories defined earlier as shock events: *romantic breakups*, *crime*, *death of a close person*, or *unexpected job loss*.

- **Recency**: The event described in each tweet should be about an event that has happened "recently". We preserve tweets that contain phrases describing that this event happened

Shock type	Regex filtering	Active learning	Covariate matching
Breakup	18,277	3,249	1,191
Crime	83,246	4,743	1,762
Death	69,707	22,870	9,456
Job loss	7,186	2,171	1,160
Total	178,416	33,033	13,569

Table 5.1: The number of tweets retained after each step.

"recently" or within a week. This helps remove tweets posting shock events that happened in the past and have already been resolved.

- **Self-centeredness**: The event described in each tweet should be about an event that happened to the author. This prevents capturing tweets about events that happened to others, which we do not consider as personal life events. An exception is the case of shocks from the 'death' category, where we only use tweets that describe the death events of a close person (e.g., a friend or family member).

- **General tweets**: We limit our scope to *general tweets* or tweets addressed to all followers of a user account[1]. This differs from replies, which although publicly visible, are more targeted towards the original tweet's author rather than the entire followers of a user and will thus be processed differently by others. We also remove all retweets or quotes as we are only interested in the messages directly generated by the shocked user.

We construct a set of regular expressions[2] to match these requirements. For each shock, we list all possible ways to address a shock event (e.g., "passed away", "died", "passing of"), a recent event (e.g., "last week", "yesterday", "this morning"), and a close person in the case of death shocks. Our source data is comprised of a 10% percent sample of tweets produced between January 2019 and June 2020 obtained through the Twitter Decahose API. We remove reply tweets, retweets, and replies with comments to another tweet. We then apply regular expressions to filter out tweets that do not satisfy our conditions on topic relevance, recency, and self-centeredness, resulting in an initial set of 178,416 candidate shock tweets (Table 1).

Even after applying the regular expressions, we observe that the precision of the filtered tweets is low (Table 2). As a next step, we improve the precision of our dataset by providing labels for the tweets and training a text classifier model through manual annotation. To ensure annotation quality, we randomly select 50 tweets per shock type and have three annotators with sufficient background knowledge to determine whether each tweet is a shock tweet or not. We then measure the Krippendorff's α Hayes and Krippendorff [2007] to measure inter-annotator agreement, where

[1]https://help.twitter.com/en/using-twitter/types-of-tweets
[2]https://github.com/minjechoi/relationships-shocks

Shock type	Precision of regular expressions
Breakup	0.147
Crime	0.048
Death	0.218
Job loss	0.202

Table 5.2: The precision of valid shock tweets after filtering on regular expressions only.

we achieve high agreement scores of 0.875 (breakup), 0.894 (crime), 0.891 (death) and 0.869 (job loss). Once high levels of agreement are ensured, we then train multiple rounds of text classifiers accompanied by augmenting annotated samples for each round, which is a form of active learning Settles [2009]. For each shock type an annotator provides labels for 1,000 tweets from our regex-filtered set of tweets, which we divide into a train/test/validation split of 400/400/200. This dataset is then used to train a classifier that predicts whether a given tweet is a shock tweet for that respective shock type. We use the pre-trained cased BERTweet Nguyen et al. [2020] model in Pytorch 1.8. For each model, we train for 50 epochs and save the model with the highest F1 score on the validation set, which is used to compute the F1 score on the test set. We use Adam Kingma and Ba [2015] with a learning rate of 1e-8 and 100 warmup steps.

Once our initial classifier is trained, we improve performance through rounds of further labeling and training with additional samples. We use the initially trained classifier to infer the probability values of being a shock tweet, a continuous score ranging between 0 and 1, for all unlabeled tweets. We then select 200 tweets of which the inferred score is closest to 0.5, the decision boundary, and additionally annotate these selected tweets, which are added to the training set to train a new classifier model using identical hyperparameter settings. We repeat this process for five rounds, after which the performance plateaus (Figure 5.1). We then select the best-performing classifiers for each shock type, which produce F1 scores of 0.88 (breakup), 0.85 (crime), 0.92 (death), and 0.91 (job loss). Using these models we infer the remaining unlabeled tweets and preserve the samples whose inferred score is higher than 0.5. We combine these samples with the positively labeled samples to use as our set of 33,033 shock tweets. The number of tweets preserved through each step is shown in Table 1.

5.3.2 Propensity Score Matching

In order to identify what causal effects the shock event has on social ties' behaviors, we need to control for counterfactual settings of users who would have experienced a shock but did not, which exist in our initial observational data. For each shock tweet, we assume that the author of the tweet

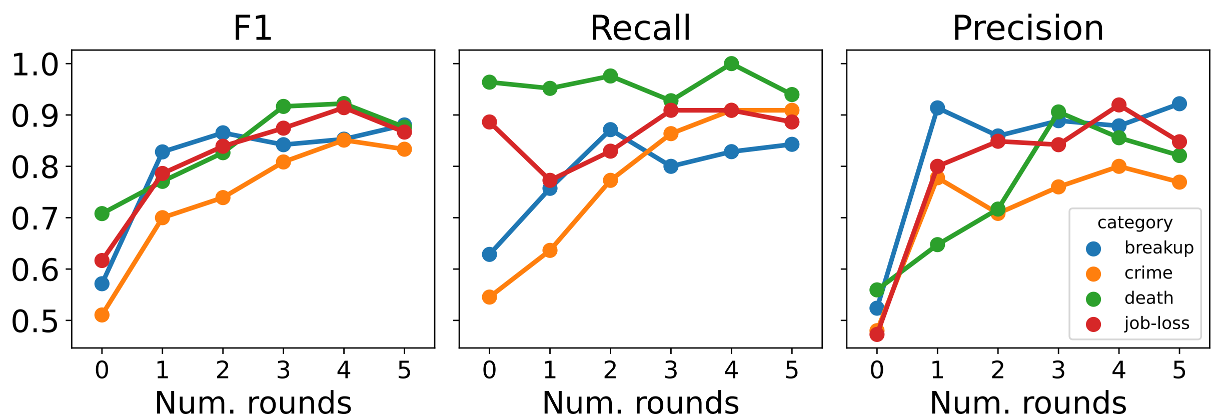


Figure 5.1: Comparison of metrics on the test set after performing additional rounds of active learning.

is a treated user, and so for users who posted multiple shock tweets (0.39% of the treated users), we only considered the earliest instance as a valid shock tweet. By comparing our treated users against a control group with similar covariates, we can ensure robustness in measuring the effect of self-disclosing a shock tweet and removing potential confounding effects. Therefore, we adopt a causal inference approach and use propensity score matching (PSM) to create a control set of users who have similar covariates as those undergoing a shock event Imbens and Rubin [2015]. As a matched user may change one's behavior over time, we match control users experiencing a shock with another user according to their behavior on a specific date.

Selecting Covariates for Matching

The first step of PSM requires identifying adequate covariates for each user to be later used in the matching task. These covariates should be relevant to whether a user posts a shock tweet or not or are likely to undergo the shock event itself. For every active user in our dataset, we consider the following covariates:

- **User demographics**: As a proxy for actual demographic information of a Twitter user, we obtain inferred gender and age using M3 model Wang et al. [2019]. The M3 model returns a distribution over age categories and a continuous score of gender performance[3].
- **Twitter account properties**: We use the number of followers, friends (followers), and activities of each user as covariates. As these numbers change slightly over time, we use the earliest available measure in our dataset.

[3]For age, we drop one category due to collinearity.

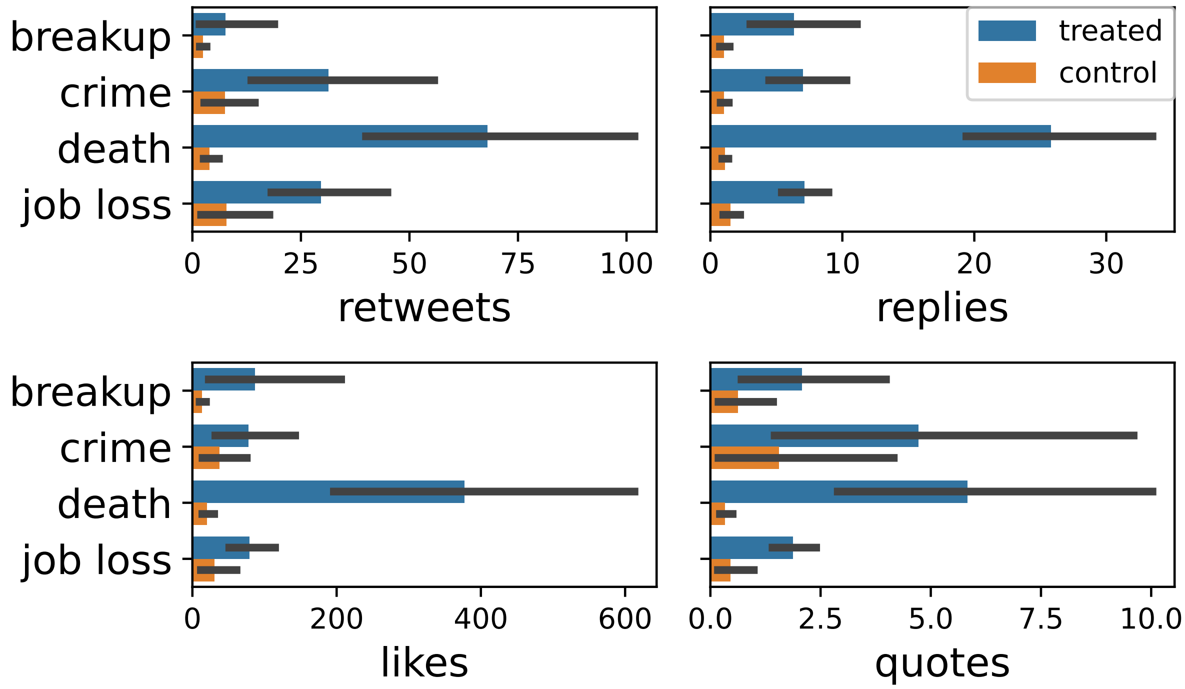


Figure 5.2: Comparison of response metrics between shock tweets and control tweets. Shock tweets receive a substantially larger amount of retweets, replies, likes, and quotes compared to the control tweets.

- **Social Condition**: We obtain inferred geolocations of Twitter users from the total variation geoinference method Compton et al. [2014] and the user's corresponding US census tract. We use the American Community Survey 2018 data for each tract to associate the user with the following covariates as proxies of the users' social conditions: racial distribution, ratio of income to poverty, education level, marriage rate, divorce rate, industry group composition, Gini index, unemployment ratio, and health insurance cover rate.

- **User activity levels**: To account for variation in the attention a user receives and their relative activity levels over time, we obtain the number of tweets (replies, mentions, retweets, and quotes) up to seven days before a particular date and aggregate them by tweet type.

For the initial set of control users, we consider all users active during the time range of our initial Twitter dataset (Jan. 2019 to Jun. 2020). We then preserve the users for which we can identify all covariates, which leaves 5,035,811 candidate users. We identify the same covariates for treated users, which reduces the number of valid treated users to 13,569 (Table 1).

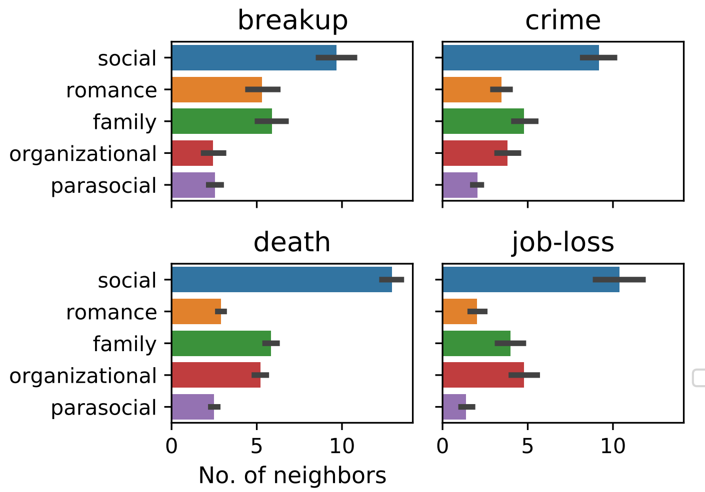


Figure 5.3: The average number of neighbors of each predicted relationship type for shock type.

5.3.2.1 Propensity Matching Users

Instead of performing propensity matching at a user level, we perform matching at the user-date level. In other words, the same user exists through several different samples, as the user's activity covariates differ by each week. We include a matching candidate for a control user for their behavior each day they were active, referred to as a *user-date pair*; this process allows closely matching the behavior of the treated user. Following standard practices of PSM, we estimate propensities using a logistic regression model where the objective is to predict whether a user-date pair is a shock instance based on the associated user-date covariates Eckles and Bakshy [2021]. The fitted model is then used to infer the propensity scores in [0, 1] for every instance using the same covariates. PSM assumes the model learns to predict the likelihood of a shock from the provided covariates, and thus similar propensity scores will indicate similar covariates. Following Eckles and Bakshy [2021], we sort the propensity scores for each user-date pair and divide them into n strata of equal sizes, with the root of the number of treated samples as n. Note that n is different for each shock type. Once both treated and control user-date pairs are sorted into the bins, we match each treated user with five randomly sampled control users whose behavior was recorded on the same day as the treated user, similar to the 1:5 ratio used in Maldeniya et al. [2020].

To verify the quality of our matching process, we measure the Cohen's d effect size for each covariate by comparing the distribution obtained from the treated users with that of the matched users. For all covariates, we obtain an effect size of lesser than 0.269, which corresponds to the group status of shocked or matched user-date pairs accounting for lesser than 1% of the variance Imbens and Rubin [2015].

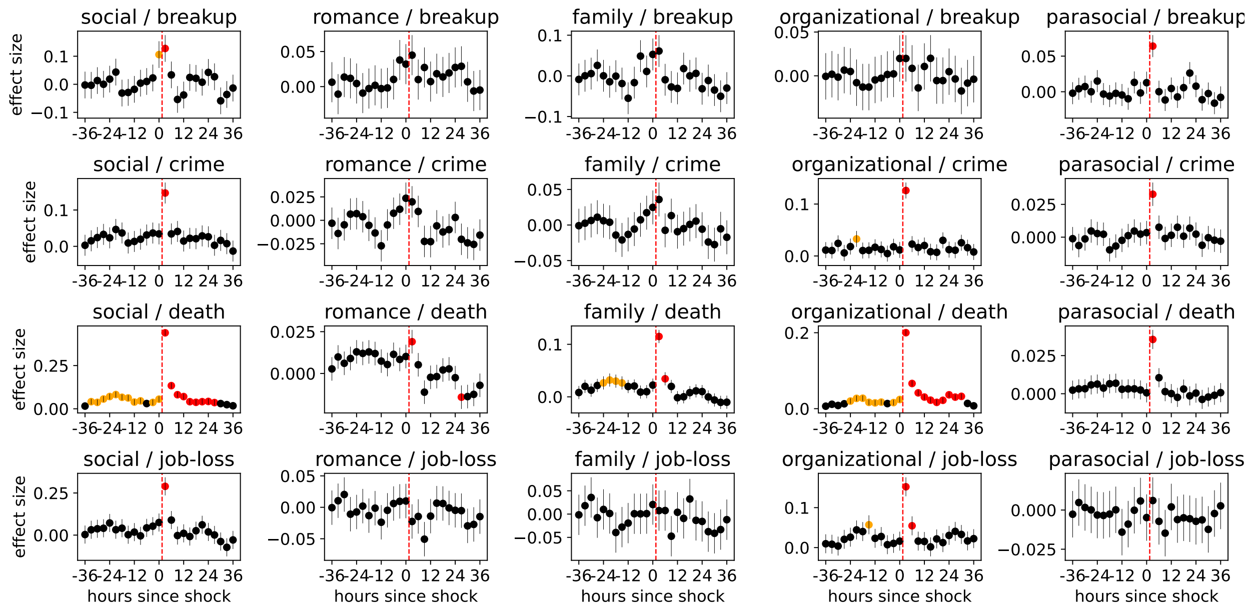


Figure 5.4: Changes in the volume of replies to shocked users across 3-hour blocks, measured using diff-in-diff relative to control users. The x-axis is the number of hours relative to the shock (e.g., the first hour of the shock corresponds to x=1). Statistically significant values (Bonferroni corrected) are colored in orange if before the shock and red if after the shock. The red dashed line indicates the time of posting the shock tweet.

5.3.3 Identifying Relationships

Social relationships can take on many distinct types, from general categories like family and friends to more specific types like step-parent or direct-report Wellman and Wortley [1990]. Here, we adopt the model of Choi et al. [2021] that infers five relationship categories from Twitter interactions: *social*, *romance*, *family*, *organizational*, and *parasocial*. These high-level categories capture the majority of relationship types expected to be observed on Twitter and, critically, have different social expectations in their behavior.

For the matched treated and control users, we identify all dyads that exchanged at least three interactions from the beginning of the time frame of our dataset to seven days before the shock event. Each tie is then classified using its text content, network features, and user description. Our selection to drop all interactions made after seven days prior to the shock tweet is to ensure that the relationship prediction is completely independent of interactions related to the shock itself. We compare our predictions with the results of another setting where we include the interactions around the time of the shock, and find that the predicted relationships of the two settings match 78% (breakup), 80% (crime), 83% (death), and 86% (job loss) of the time.

While the Choi et al. model requires at least 5 interactions, we relax the minimum amount of required interactions to three tweets as our time frame is much shorter (1.5 years vs. 7 years), leading to a smaller number of classifiable dyads. Only using tweets up to one week prior to the shock event avoids potential confounds from inferring relationships based on interactions during the shock event itself. To test the robustness of this setting, we identify dyads with 5 or more interactions, then create two feature sets of all dyads - one using only three interactions per dyad, and the other using all available interactions. We observe that the predicted relationship categories for the two settings match in 83% of the predicted dyads, largely exceeding a random baseline of 20% and showing that the predictions are robust in the number of interactions.

The distributions of the relationship types for treated and control users of each shock type are shown in Figure 5.3. Across all shock types, the social relationship type is most common, consistent with Choi et al. [2021]. We also observe that the size of the average number of ties per relationship type is similar across shocks.

5.4 Relationships and Shock Response

The shocked users receive far more engagement from their peers (Figure 5.2). Across all four types of shocks, the treated users received more retweets, replies, likes, and quotes of their tweets than those in the propensity-matched control users, suggesting that their social network is aware and responsive to these events while using their own networks to increase visibility and utilize social capital [boyd et al., 2010]. Does this responsiveness vary based on the relationship the other user, referred to as an *alter*, has with the shocked user and the nature of the shock? Here, we test RQ1 using a pseudo-causal difference-in-difference to test the effect of another person communicating with the person experiencing a shock.

5.4.1 Experiment Setup

For each user-date pair, we collect all tweet activities involving the user within ±36 hours of the timestamp of the corresponding shock tweet. This timeframe represents one full day (24 hours) before and after the event, plus 12 hours to account for time zone differences. We group the tweets into 3-hour bins, where we count the activities conditioning on (1) the direction (the shocked user responding to an alter user vs. an alter user responding to the shocked user), (2) the relationship type (five categories), and (3) the shock type (four categories). For each time series in the 80 conditions, we construct a difference-in-difference model with lead and lag variables. This model is summarized as follows:

$$y_{d,t,s} = \alpha + D_d + T_t + S_s + \pi_t P_t + \epsilon_t$$

where $y_{d,t,s}$ indicates the number of replies a user in strata s received treatment at day d with t hours since the posting of the shock tweet, divided by the number of neighbors of the corresponding relationship type. d is the date of the tweet (e.g., 2019-01-05) represented as a category variable, t is a categorical variable for 3-hour intervals from -36 to +36, and s is the corresponding strata of the user-date pair which was used for stratifying based on propensity scores. D_d and T_t capture temporal trends while S_s captures differences between strata. P is the treatment interaction of the relative hour from a shock tweet, which allows π to capture not only post-treatment effects but also any pre-treatment effects around the shock. The effect sizes correspond to changes in the attention of the shocked user's neighbors.

5.4.2 Results

Social ties strongly engaged with individuals experiencing a shock but engagement varied significantly by the type of relationship and shock (Figure 5.4). We highlight three findings.

Romance and family ties are not as responsive online Studies on stress management through social interactions focus on the role of "close" relationships which consist of family members or romance partners Collins and Feeney [2000]. In contrast, we observe that in online social settings, the largest effect sizes on response rates come from social and organizational relationships. Indeed, the effect sizes of romance and family relationships are insignificant for all shocks apart from death, which is also much lower than those of social and organizational ties. Interestingly, replies from individuals with romantic ties *drop* shortly after experiencing a death shock, an effect not observed in any other case. One possible explanation is that, unlike social or organizational ties where online social networks serve as the main channel for communication, romance or family ties would be able to communicate through other offline means such as phone calls, text messages, or in-person conversations Burke and Kraut [2014]; the gravity of a death shock could potentially decrease the alter's online communication, leading to the observed drop.

Responses to shocks can be relationship-specific Based on the nature of the shock, individuals in certain relationships were more likely to respond—or not respond at all. This trend can be observed in the case of organizational ties, where in contrast to its high responsiveness for crime, death, and job loss shocks, individuals with organizational relationships had no significant increase in responding to breakup shocks. Similarly, parasocial relationships show significant levels of responsiveness in all shock categories except for job loss. These trends likely reflect normative boundaries in which content people share, which differ by relationship type; as a result, alters may feel uncomfortable showing interest in events that are considered out of their social boundaries Collins and Feeney [2000].

Pre-shock effects for death shocks For death shocks, a small but significant pre-shock effect can be seen before the actual shock. By manually examining a portion of the pre-shock activity, we were able to discover that this effect was partly due to announcements that their close ones were in a critical state, such as someone getting injured and being transported to the ER or entering a critical condition, which are stressful situations themselves. These foreboding tweets can also be seen as request signals for support Oh and LaRose [2016] along with the shock tweet, which is returned by increased attention and support from others.

5.5 Impact of Closeness on Shock Response

While our findings showed relationship-specific trends that differ for each shock type, these properties may differ even for the same relationship, depending on other network properties of the dyad such as tie strength and embeddedness. Here, we formulate a task of predicting whether an interaction occurs within a dyad when one user posts a shock tweet, and investigate how tie strength and embeddedness play different roles across relationship and shock types.

5.5.1 Experiment Setup

We examine how responsiveness can be predicted with respect to tie strength and structural embeddedness, two well-known proxies of closeness in a social network. We construct an undirected network using 10% of all Twitter mention activities during a time frame of three months before the shock occurrence, where an edge is formed if a user mentions the other. We compute (1) tie strength, measured as the number of mentions an alter made to a shock user during that period, and (2) structural embeddedness, measured through Jaccard similarity which is the size of mutual neighbors of two users divided by the size of the union of their neighbors. For a dyad of the shocked user i and alter j with a known relationship, we assign a label of 'responded' (1) or 'not responded' (0) to $y_{i,j}$ based on whether the alter user responded to the shocked user within 36 hours of the shock tweet. We then formulate a logistic regression task with the following formula:

$$y_{i,j} = \alpha + t_{i,j} + s_{i,j} + X_i + X_j$$

where the number of mentions towards the shocked user $t_{i,j}$ represents tie strength and the Jaccard similarity $s_{i,j}$ represents structural embeddedness. The number of followers, friends, posts, in-degree and out-degree of each user (vectors X_i and X_j) are also included as independent variables which we control for.

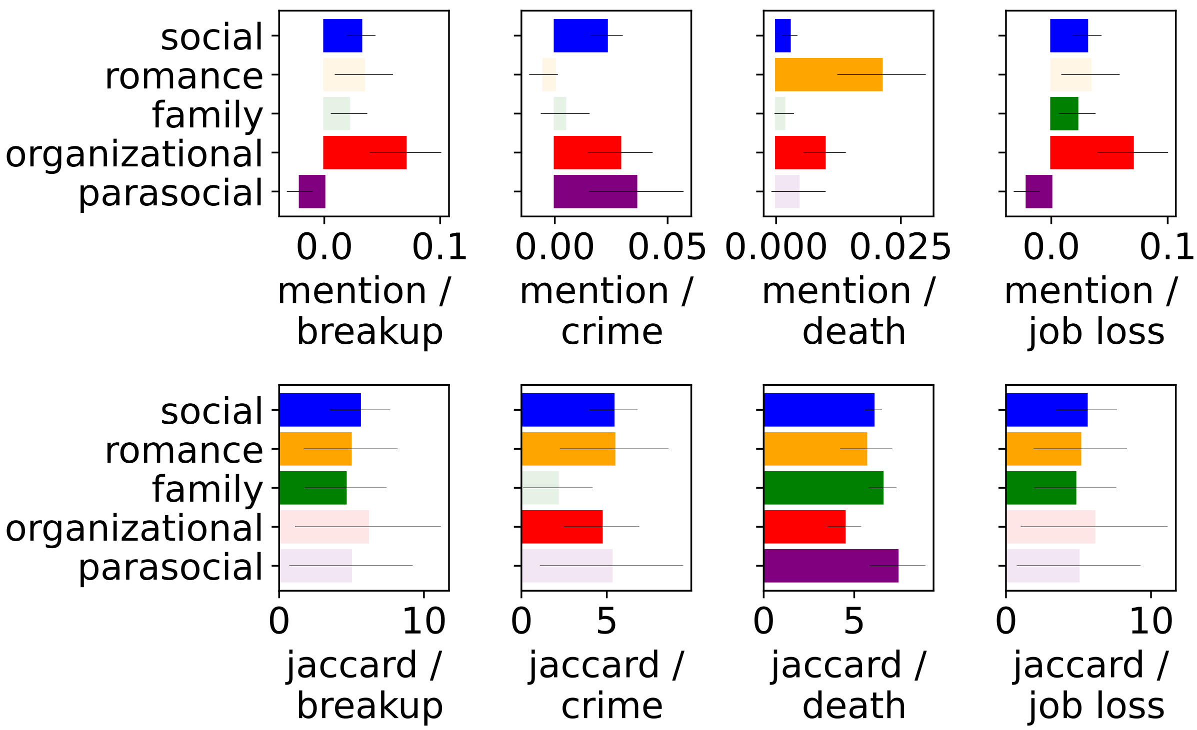


Figure 5.5: Coefficient sizes obtained from a logistic regression of response on mention frequency (tie strength) and Jaccard similarity (structural embeddedness). Values with solid colors indicate statistically significant (Bonferroni-corrected) coefficient values.

5.5.2 Results

We display the coefficients of the network properties for each relationship-shock pair and compare them across relationship types for each shock in Figure 5.5. When significant, close ties are almost always likely to have a positive effect on response prediction, which is expected given the relationship between closeness and social support. Comparing relationship and shock type provides us with a clearer picture. For example, in family relationships, mention frequency is insignificant for most shocks while the Jaccard similarity is almost always significant. This shows that social support from family members depends more on social embeddedness rather than communication frequency. The situation reverses for organizational ties, where communication frequency is more important than embeddedness across shocks. In addition, response during death events depends much more on embeddedness than the frequency of communication. In summary, we generally confirm the important role of closeness in the likelihood that a social connection provides support during shocks, but at the same time, we find it is highly dependent on the type of closeness, shock, and relationship.

5.6 Topic Shifts in Shock Responses

While the response to another person's shock is likely to be of social support, the topics discussed take into account the information shared between the relationship and the underlying social background. Our final question is whether users adjust their topics when responding to shocks and if this level of adjustment differs by relationship.

5.6.1 Experiment Setup

By considering each tweet message as a document and running a probabilistic topic model Blei et al. [2003] across all documents, we can aggregate the output topic distributions for each document across shock type and relationship type. We can compare the topic distributions to interpret which relationship type is more likely to be associated with a particular topic. We aggregate the text content of the replies a treated or control user received during the first 36 hours since posting a shock tweet and run an LDA topic model MALLET McCallum [2002] on 10, 20, 50, and 100 topics. Recent work [Hoyle et al., 2021] have shown that evaluation metrics such as coherence scores may not align well with human judgments on topic quality, and thus we manually examine the topic model results to select the topic size that contains the most diversity and least number of overlapping topics, leading to a size of 50. As the topics produced by topic models are unsupervised, we manually label each topic after inspecting its most probable words; topic labels were obtained for 35 of the 50 topics; the remaining 15 did not have clear organization and were not used further in our study. The

Topic no.	Keywords	Annotated topic
1	hair wear wearing black blue clothes	appearance
2	head face back hand hands eyes	body
3	family loss love hear prayers condolences	condolence
4	people it's don't that's i'm make	conversations
5	it's i'm don't people good that's	conversations
6	day great happy hope good birthday	days
7	hope glad i'm man hear stay	emotional support
8	child mom baby kids age mother	family
9	money work buy pay make free	finance
10	food eat chicken cheese eating pizza	food
11	drink water ice tea milk coffee	food
12	game play games playing played switch	games
13	balls demon harder image hill silent	games
14	moon pokemon shiny sun trade frame	games
15	brain test order blood levels cancer	health
16	gay ship gender canon signs dark	identity
17	lol shit i'm don't ass lmao	informal language
18	english language spanish speak words	language
19	movie watch show good episode character	media
20	song music album listen songs love	music
21	cats dogs pet yang ship they're	pets
22	adopt home pledge foster rescue save	pets
23	trump case court rule report issues	politics
24	black white women men people racist	social issues
25	trump vote party government state win	politics
26	love cute good amazing beautiful nice	positive emotions
27	god jesus soul church rest lord	religion
28	school high class year kids teacher	school
29	twitter tweet post account lol send	social media
30	red sox mlb fans white starts	sports
31	team year game win season play	sports
32	match champion wrestling wwe heel turn	sports
33	data apps desktop storage celica folders	technology
34	time i'm back lol day years	time
35	side city car big south back	traffic

Table 5.3: Top keywords and annotated topics from resulting topic model

most probable words of each topic and its label are shown in Table 5.3. We observe a wide variety of topics, reflecting the range of information exchanged in Twitter conversations.

Once the topic distribution for each tweet is obtained, we examine whether posting a shock significantly increases or decreases the usage of specific topics seen in the replies from each relationship type. For each topic, relationship type, and shock type, we obtain two probability distributions from (1) the probability of that topic in replies toward shocked users and (2) the topic's probabilities in replies to control users. We then test whether the two distributions have an equal mean using the student's t-test. This results in a total of 1,000 comparison t-tests as there are 50 topics, 5 relationships, and 4 shocks. To account for the multiple comparisons, we apply Bonferroni correction by defining an effect as significant if and only if the p-value of the t-test is less than 0.05 divided by 1,000, imposing a strong threshold for significance.

Shock	Social	Romance	Family	Organizational	Parasocial
Breakup	conversations ↑				informal language ↑
Crime	condolence ↑				
	emotional support ↑	emotional support ↑	emotional support ↑	emotional support ↑	emotional support ↑
Death	appearance ↓				
	body ↓			body ↓	
	condolence ↑	condolence ↑	condolence ↑	condolence ↑	condolence ↑
	conversations ↓			conversations ↓	
	emotional support ↑		emotional support ↑	emotional support ↑	
	finance ↓			finance ↓	
	food ↓			food ↓	
	games ↓		games ↓		
	informal language ↓		informal language ↓	informal language ↓	
	media ↓		media ↓	media ↓	
	music ↓				
	pets ↓		pets ↑		
	positive emotions ↓				
			religion ↑		
	social media ↓		social media ↓		
	sports ↓				
	time ↓			time ↓	
	traffic ↓				
Job loss	condolence ↑			condolence ↑	
	emotional support ↑			emotional support ↑	

Table 5.4: A comparison of topics that experienced significant increases or decreases following a shock. The arrow direction indicates whether the topic usage has increased or decreased after a shock when compared with the control group.

5.6.2 Results

We find 66 out of 1,000 possible cases across the combination of all shocks, relationships, and topics in which the change in usage is significant, shown in Table 5.4. Following, we highlight three trends by shock type.

Breakups For breakup we observe two cases of significant increase in topic usage, coming from social (conversations; T4) and parasocial (informal language; T17) relationships. These topics are both situated in the 'conversations' category which indicates an increased amount of casual conversation-based communication. When experiencing and posting a breakup, people considered in a social or parasocial relationship would likely engage more in conversations, either to discuss the event or to divert the attention away Moller et al. [2003]. This shock type is also intriguing in that there is no apparent increase in condolence (T3) or emotional support (T7), which is observed in all other shocks. Support following breakups is shown not explicitly through words of condolence but instead in a form of increased interactions and conversations.

Crime & job loss All relationship types responded to crime shocks by increasing emotional support (T7), and also condolence (T3) in social relationships. For job loss, we observe an increase in condolence and emotional support only from the social and organizational categories, consistent with our earlier findings from Section 5.4.2.

Death Responses following a death shock show the most diverse changes in topic composition for all relationships. For romance and parasocial ties, we observe significant changes only in condolence (T3). Other relationships display more dynamic topic shifts. For social, family, and organizational relationships, along with the increase in condolence and emotional support (T7) we observe decreased usage of several topics ranging from appearance (appearance, body) to entertainment (games, media, sports). As for decreasing topics, we can notice that (1) relationships such as social ties discuss a wide variety of topics in pre-shock communication, and (2) in the case of death shocks these relationships are discouraged to transmit information other than messages of condolence and direct emotional support. We also find evidence of relationship-specific social support, seen in the case of topics related to pets (T22) and religion (T27) being increasingly used in family relationships.

In general, our results show that once a user experiences a shock, their neighbors may choose to significantly reduce the likelihood of conveying topics related to everyday events such as sports or politics, and shift towards expressing condolences and support. At the same time, there are also relationship-specific topics that are exchanged, such as topics related to pets or religion from family members.

5.7 Discussion

General discussion and implications One consistent theme throughout this study is that relationship types have their own functionalities, leading to different levels of engagement and communication following shocks. Our findings highlight the existence of latent social processes within the interactions of online social networks, which can be revealed only by explicitly modeling relationship types. Even well-established social dimensions such as tie strength and structural embeddedness which are known to contribute to responsiveness contribute differently by relationship. Our work suggests incorporating interpersonal relationships along with other network properties to gain a richer understanding of the dynamics within social networks.

Our study shows that existing theories on social relationships and social support do not always fit in the context of online social networks. Previous studies have emphasized the role of "close" relationships such as family members and romantic partners as major providers of social support. However, our results show that, in online settings, social and organizational ties are the most responsive and supportive, and also that closeness does not always lead to higher responsiveness. This discrepancy may result from several reasons, such as romance or family ties having other communicative means to provide support instead of Twitter Burke and Kraut [2014] or a disinclination to exchange personal conversations on public social networking sites.

Our work can help identify which ties in online networks are most supportive during shock experiences. For individuals experiencing stressful life events, it is often challenging to disclose their problems in public online spaces due to issues such as impression management concerns Oh and LaRose [2016]. The response behavior seen in our findings can help indicate the users in specific relationships that are likely to be receptive to outreach after different shock events. Our findings can be used as a basis for providing topic- and relationship-specific support to shocked users by recommending the most suitable ties for different situations.

Limitations Our work only examines shocks in an online setting, leaving open the question of how these trends hold in offline settings for the same events. Prior work has studied these offline behaviors separately Kendler et al. [1999], yet the challenge of growing these studies to our scale is likely prohibitive. Future mixed-method work is needed to fruitfully identify how individuals leverage their offline *and* online networks during shock events.

Similarly, we only examine public responses to life events. As evidenced by the large number of public shock event announcements and the many responses and expressions of support we found in our data, we believe this is an important phenomenon to study. We do acknowledge that considering private messages can lead to different findings. For instance, studying the pattern and content of private messages or how interactions differ from previous years where private messaging was unavailable could lead to interesting findings. However, we leave those to future studies.

Our study focused on short-term interactions following a shock. Although life shocks are often serious events, our results suggest the impact on one's online network was very brief, with the shocked user returning to normal activity within a week. However, the event may still lead to the long-term impact, such as the rewiring of network structure or changing ties, which longer-scale studies may identify.

Finally, the scope of our relationships is limited to five categories based on the classifier of Choi et al. [2021]. These categories may not capture the entirety of social relationships in online social networks. Further, while the best available model for our setting, the classifier itself may exhibit bias that affects our results. Future work on relationship inference could expand the scope of our analysis by identifying additional relationships or more specific subtypes of relationships to capture fine-grained trends.

In unexpected distressing circumstances, a helping hand or a kind word can make all the difference. Our results show that when individuals undergo unexpected shocks, others do reach out online but who specifically reaches out depends on the nature of their relationship and its interaction with the type of shock. Through a large pseudo-causal study of individuals experiencing four types of shocks paired with a precisely matched control set, we demonstrate that social relationships vary significantly in which types of shocks they engage with and even in the context of how they communicate to the shocked individual. Further, we show that while higher tie strength and social embeddedness both increase the likelihood of whether a person will respond in the case of most shocks, these trends are highly dependent on the relationship and shock event type. The insights provide further evidence for the importance of explicitly modeling social relationships in studies on social network dynamics. We release our code, trained classifiers, and data at *https://github.com/minjechoi/relationships-shocks*.

CHAPTER 6

Conclusion

In the final chapter, I discuss the findings of the three proposed studies and their implications. Furthermore, I address future research directions where computationally modeling properties of social relationships can lead to answering unsolved research questions.

6.1 Research contributions and findings

Despite often being considered an essential component of social networks and social interactions, social relationships have remained difficult to quantify and model in online social networks. The three studies proposed in the previous chapters shed new light on how to infer such knowledge from online conversations using natural language data, and furthermore how to use these inferred labels as a lens to improve our understanding of the human behaviors of communications in online social networks, especially in response to dire situations such as life shock events.

The first study contributes to the understanding of how some of the fundamental sociological elements that define human relationships are reflected in the use of language from a theoretical standpoint. In particular, I show that all 10 dimensions are represented abundantly in everyday conversations and that the way they are expressed can be learned even from a small number of examples. In practice, the data I collected and the classifiers I built can contribute to creating new text analytics tools for social networking platforms. The use cases I propose in the study are examples of such tools being applied, where I show that the strengths of social dimensions within communities are correlated with societal outcomes such as education levels or suicide rates. Building upon these findings, I believe that the dynamics of a number of processes mediated by social networks (including diffusion, polarization, and link creation) could be re-interpreted with the application of the 10-dimensional model to conversation networks.

The second study starts from the assumption that not all ties are equal: friends, family, and lovers all have different social, linguistic, and temporal behaviors—yet, social network studies have typically limited themselves to networks with edges that encode only the existence of a relationship, but ignore the *type* of that relationship. Using a dataset consisting of the interactions between 9.6M dyads on Twitter with known relationship types, I introduced a new approach that explicitly models interpersonal relationship types in social networks. I show that the linguistic, topical, network and diurnal properties in online communication between different relationship types match predictions from theory and observational studies. Next, I demonstrate that relationship types can be accurately predicted using text and network features combined with state-of-the-art deep learning models. Last, I show that knowing relationship types improves performance when predicting retweets—demonstrating the benefits of predicting relationships at scale. The addition of relationship type significantly improves the recall of retweets for social, family, and romance relationships, which are considered more personal. The proposed approach, combined with the consistency of our results with existing literature on social relationships, further demonstrates the value of studying social media networks to further understand the differences in communicative behavior across interpersonal relationships. Furthermore, as evident from the performance of my relationship classifier and the improvement in retweet prediction, my work enables new types of analyses that benefit from large-scale relationship-aware networks such as modeling network evolution, information diffusion dynamics, and community structure.

In the third study, I show that when individuals undergo unexpected shocks, others do reach out online but who specifically reaches out depends on the nature of their relationship and its interaction with the type of shock. Through a large pseudo-causal study of individuals experiencing four types of shocks paired with a precisely matched control set, we demonstrate that social relationships vary significantly in which types of shocks they engage with and even in the context of how they communicate to the shocked individual. Further, I show that while higher tie strength and social embeddedness both increase the likelihood of whether a person will respond in the case of most shocks, these trends are highly dependent on the relationship and shock event type. Finally, I show evidence of topics that are unique to specific relationship types when providing support to someone in need. The insights provide further evidence for the importance of explicitly modeling social relationships in studies on social network dynamics.

The interdisciplinary nature of the three studies offers contributions toward multiple research fields. From a network science perspective, my findings provide evidence that social relationships do affect the process of how information is shared throughout online social networks, a venue that had not been previously studied. Specifically, Study 3 showed that the likelihood of providing a supportive message as well as its type differs by relationship type. This suggests that technology can be used to improve social support networks by providing suggestions as to which type of information or individual would be most helpful in a given crisis situation.

Another significance of this dissertation arises from the fact it aims to bridge studies on social networks and relationships with state-of-the-art computational techniques developed by computer science and machine learning researchers. While several effective methods have been developed to infer properties of social relationships (refer to Chapter 2.3), often it is the case that these approaches do not get recognized by social scientists and applied in their studies. One of the aims of this dissertation is to demonstrate the possibility of such applications, which can open up new possibilities for social scientists to explore online social network datasets, which can enable observational studies at a much larger scale.

One of the key findings of the first study is that social relationship dimensions measured collectively at US state level can reflect societal outcomes such as suicide rates and education levels. Future research can further validate this finding, ultimately creating the possibility of using the measurement of community-level interactions and the associated relationship attributes to obtain a quick and accurate reflection of our society, which can be of use especially to policymakers.

6.2 Discussion

6.2.1 Validity of Studying Offline Relationships in Online Settings

One of the challenges described in Chapter 2.4.1 is whether the behaviors of offline relationships can be expected in online platforms, especially given platform purposes and publicized messages. Studies 2 and 3 in the dissertation have been conducted under the assumption that the properties of offline social networks will be also visible in online social networks, following the findings of prior work [Subrahmanyam et al., 2008, Reich et al., 2012, Ozenc and Farnham, 2011, Farnham and Churchill, 2011]. Based on the findings, we can find that interpersonal communication in online social networks leads to an interesting landscape. Factors such as the level of perceived closeness or mutual ties in a relationship are reflected in Twitter usage, as can be found in levels of linguistic patterns, network connections, and diurnal communication patterns (Chapter 4.4). At the same time, the fact that interpersonal communication on Twitter can be publicly viewed results in the suppression of exchanging more personal messages. This is evident from the case of responses to shock events, where romance and family ties, who tend to share a deeper level of emotion [Adams, 1986], are less likely to provide responses on Twitter. A relevant future research question would be to examine platform-level differences in how relationships are exerted in interpersonal communication. Twitter places itself in an interesting position, where a person's offline social network is combined with ties that provide information [Kwak et al., 2010].

6.2.2 Dynamic and Multi-Faceted Social Relationships

The final challenge discussed in the literature review is that relationships may not be static. Social relationships can change over time, and a single relationship can have multiple facets due to increased or decreased levels of communication and emotion sharing [Altman and Taylor, 1973]. During my studies, this trend was most noticed in the data collection process of Study 2, where phrases used to describe dyadic relationships were mapped into relationship categories. A nontrivial number of dyads were associated with multiple different phrases, some that would even exist in different relationship categories (e.g. my best friend (social) / my colleague (organizational)). In the study, however, such dyads were intentionally removed from the dataset and any subsequent analyses. The rationale behind restricting relationships to static, single-faceted ones was to remove any ambiguity in the dataset of relationship-labeled dyads, which was to be used for training relationship classifier models. As a result, users who were associated with more than one relationship phrase were removed.

How can one identify dynamic and multi-faceted social relationships? Based on the methods in the proposed studies, one approach would be to run the relationship classifier on different time points using past interaction data around each time point. A stable relationship would have the same category regardless of time point whereas a dynamic relationship might fluctuate in the categorized relationship type based on when it was inferred. Such results could be further used to examine under which factors would a perceived relationship change over time.

6.2.3 Studying Online Social Relationships in the Post-Twitter Era

The three studies would have been impossible without the existence of publicly available datasets of online social network platforms, namely Twitter and Reddit. In the past decade, researchers have benefitted greatly from the accessibility of social network datasets that have been made publicly available for research purposes. This trend has unfortunately been reversed where company-provided datasets and API access to social media platforms are becoming more restricted. A most prominent change has been observed from Twitter, where changes in leadership have led to the removal of free academic-tier access and increased API pricing, significantly reducing the possibility of collecting data and conducting studies on Twitter [1]. Meanwhile, Reddit is also undergoing policy changes that will require payment for using its API services [2], another step which will hinder data-driven research on online social network platforms. These changes are detrimental to research in online social networks and information diffusion, as it greatly depends on rich datasets containing social network structures and interaction history among users.

[1] https://www.theverge.com/2023/5/31/23739084/twitter-elon-musk-api-policy-chilling-academic-research
[2] https://techcrunch.com/2023/04/18/reddit-will-begin-charging-for-access-to-its-api/

Here I address a number of foreseeable changes in the research space regarding changes in public data availability. One will be an even heightened importance of collaboration with various research groups. The effort to understand social network dynamics has been covered in a large number of fields including psychology, sociology [Dunbar et al., 2015], and public health [Giorgi et al., 2023], where researchers often maintain datasets of social network interactions from a sampled set of participants. A more lucrative option would be to directly collaborate with research teams within social networking companies such as Facebook or LinkedIn that have elevated access to such resources. In fact, a number of influential studies on social network structures and user behavior have been conducted under such collaborative effort [Rajkumar et al., 2022, Chetty et al., 2022a,b]. However, this option remains unavailable for the vast majority of researchers who lack such connections.

Another alternative is to study newly emerging online social networks that may not have such access restrictions. In particular, the advent of alternative platforms to Twitter shed new light on the hope of computational social science research. For instance, the mass migration from Twitter to Mastodon [La Cava et al., 2023] suggests that the latter can replace the need for the former to some extent, where one can hope that the network structures and interaction patterns are also similar. The newly introduced Threads platform from Meta also serves as a possible alternative for conducting social network behavior research. However, this direction requires a sufficient amount of research on understanding how users of these platforms behave and intend to use them.

A new but promising direction is to simulate large-scale studies using large-language models (LLMs). Previously, social scientists have simulated human interaction at scale in social networks using agent-based methods [Marshall and Galea, 2015]. Human interactions are summarized into simple abstractions, where influence is mathematically computed. Recent work has shown that LLMs can be injected with different personas and relationships and then be deployed to virtual environments, where agents interact to create responses that align with how humans would act in social settings [Park et al., 2023]. Of course, at the current stage, it is highly questionable whether these LLMs do contain human knowledge, and whether it is appropriate to learn how humans behave based on their interactions. Nevertheless, the data that LLMs are trained and processed are massive sources of text created by humans, so one can still argue that the outputs created from LLMs represent different humans to some extent. Provided that the right validations and assumptions can be made, simulating relationship-wise human behavior using LLMs can be an interesting direction to pursue in an age marked by the absence of public data.

6.2.4 Social Relationships: Dimensions or Categories?

In this dissertation I approached the task of identifying the properties of social relationships from two different perspectives: by measuring the strength of dimensions of relationships (Study 1) and by directly inferring relationship categories (Study 2). These two approaches measure different concepts of relationships and can provide a complementary understanding of the relationship, as is shown in Chapter 4.4.5. First, measuring the strength of different social dimensions from a single sentence in a conversation can provide indirect information about the social processes that occur within that specific relationship. This is especially ideal for identifying relationships in online platforms where a dyad is not expected to have frequent interactions and there is no additional information to understand other social factors between the two users such as mutual network structure. Second, directly inferring categories of social relationships from user interaction data is a much more challenging task that requires a sufficient amount of interaction history as well as additional network data. While this setting is more restricted due to its requirements, the benefits of this approach are that one can use the inferred relationship categories to relate to theories or hypotheses expected of different social roles and focus groups [Feld, 1981, 1982] which are capable of providing distinctive types of social support and information exchange. Given these use cases, I argue that my dissertation can contribute to providing a framework for flexibly alternating between the two methods of identifying relationship properties depending on the available data within an online platform.

6.2.5 Platform-wise Effects on Relationship-specific Interactions

Another aspect of platform-specific differences is how the design of a platform can lead to the formation and maintenance of different types of social relationships. One example is algorithmic curation such as friend or content recommendation algorithms, which can limit one's resulting social network in terms of demographic [Stoica et al., 2018] and political [Santos et al., 2021, Ramaciotti Morales and Cointet, 2021] diversity. It is often the case that online social networks are built for different purposes, leading to recommendation algorithms optimized towards their respective goals. Platforms such as Snapchat or Facebook focus on finding friends, while LinkedIn specializes in discovering potential work relationships. Meanwhile, Twitter's recommendation algorithm may focus more on suggesting users who share similar interests. People adopt different mindsets depending on the platform they are participating in [Jaidka et al., 2018, Archambault and Grudin, 2012b], and coupled with the heterogeneity in terms of end-goals of each platform, one can expect that the composition of relationships that are prevalent in each platform may differ, even for the same individual.

6.3 Future Directions

The findings of the proposed studies lead to several interesting potential directions. I will conclude by addressing a number of feasible research directions based on the findings of my three studies which can be applicable to various domains.

6.3.1 Comptational Modeling of Relationships in Conversation Settings

Relationships lie at the fundamental level of social factors in language, along with personal properties of the speaker and receiver [Hovy and Yang, 2021]. Compared to the amount of work towards building models as conversational chatbots that incorporate personal attributes [Demasi et al., 2020, Zheng et al., 2020, Song et al., 2020], much less work has been done on generating content for conversations in specific relationships, which is the next level of social factors. The findings of my dissertation along with other recent work [Jurgens et al., 2023] confirm that interpersonal relationships affect how people tune not only their language in online social conversations but also under which contexts they talk. This knowledge can be further applied to build tools that are aware of relationships, with potential applications including generating relationship-aware conversations in stories, scripts, etc.

6.3.2 Interplay of Information Diffusion and Relationship Type

My findings indicate that factors such as social norms and other types of predefined social knowledge among relationships can play a part in determining how and whether information gets transmitted between users in online social networks (Study 3). In other words, we can better model and estimate the diffusion of information in social networks if we can label the properties of the relationships among dyads. Provided that one can successfully identify properties of relationships of every edge in a social network, it then may be possible to include such knowledge in designing studies for understanding the impact of information diffusion within networks. The focus on relationship properties as a determinant of information diffusion can greatly complement existing work in this direction which has looked instead into factors such as tie strength, community membership, and type of information [Romero et al., 2011, 2013a, Cheng et al., 2014, Kupavskii et al., 2012, Cao et al., 2017, Li et al., 2017a].

At an individual dyad level, another promising research direction is to conduct quantitative studies of the impact of information when conditioned on relationship properties. While Study 3 examines the likelihood of sending messages of social support to crisis-affected individuals and does show some evidence of relationship- and shock-specific behavior, this can be further expanded to information transmission other than social support. For instance, an interesting research direction would be to examine the impact of susceptibility toward misinformation when transmitted to various types of relationships. Social relationships are associated with varying levels of trust as can be seen in Study 1, and there can be potential impact if one can discover that the prior level of trust established between the two parties of a relationship may in fact be a significant factor that contributes to the spread of misinformation.

6.3.3 Measuring the Trajectory of Social Relationships in Online Social Networks

As mentioned in Chapter 6.2.2, the aspect of dynamic and multi-faceted relationships has been less prioritized in my studies, yet remains an important component to understand how social relationships behave in online social networks. Future work should place more emphasis on this nature and try to understand how relationships change over time. One possible extension would be to combine a user's social network trajectory data with timestamps of life events that result in status change. Examples of possible life events include relocation to a new area, starting a new job, or changes in romantic relationship status. When given a sufficiently long time frame, the trajectory of relationships affected by these events can provide us with information on how the changes in individuals' perceived relationships are reflected in the language, topics, and interaction frequency of the dyad. One particularly interesting direction would be to identify cases when the inferred category of the relationship itself changes over time, for instance, a coworker relationship transitioning into a friend relationship after one of the users moves to a new company, or a friend relationship transitioning into a romantic relationship over time. One caveat is that Twitter, the platform that was mainly used for conducting the studies presented in this dissertation, may not be the ideal platform to measure such changes due to its public disclosure nature and focus on the diffusion of information content rather than emotional exchanges. In an ideal setting, additional sources of interpersonal communication such as messenger chat data or text messages may help us gain a more comprehensive and nuanced view of how relationships change over time.

BIBLIOGRAPHY

S. Adali, F. Sisenda, and M. Magdon-Ismail. Actions speak as loud as words: Predicting relationships from social behavior data. In *WWW*, 2012.

L. A. Adamic and E. Adar. Friends and neighbors on the web. *Social Networks*, 25(3):211 – 230, 2003.

R. G. Adams. Emotional closeness and physical distance between friends: Implications for elderly women living in age-segregated and age-integrated settings. *The International Journal of Aging and Human Development*, 22(1):55–76, 1986.

F. Agneessens, H. Waege, and J. Lievens. Diversity in social support by role relations: A typology. *Social networks*, 28(4):427–441, 2006.

C. Aguilar-Raab, D. Grevenstein, and J. Schweitzer. Measuring social relationships in different social systems: the construction and validation of the evaluation of social systems (evos) scale. *PLoS One*, 10(7):e0133442, 2015.

L. M. Aiello. Nature of social structures. In R. Alhajj and J. G. Rokne, editors, *Encyclopedia of Social Network Analysis and Mining, 2nd Edition*. Springer, 2018. doi: 10.1007/978-1-4939-7131-2_110180. URL https://doi.org/10.1007/978-1-4939-7131-2_110180.

L. M. Aiello, R. Schifanella, and B. State. Reading the source code of social ties. In *Proceedings of the 2014 ACM conference on Web science*, pages 139–148, 2014.

D. Aldeis and T. D. Afifi. Putative secrets and conflict in romantic relationships over time. *Communication Monographs*, 82(2):224–251, 2015.

I. Altman and D. A. Taylor. *Social penetration: The development of interpersonal relationships.* Holt, Rinehart & Winston, 1973.

N. Andalibi. What happens after disclosing stigmatized experiences on identified social media: Individual, dyadic, and social/network outcomes. In *CHI*, 2019.

A. Archambault and J. Grudin. A longitudinal study of facebook, linkedin, & twitter use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 2741–2750, New York, NY, USA, 2012a. Association for Computing Machinery. ISBN 9781450310154. doi: 10.1145/2207676.2208671. URL https://doi.org/10.1145/2207676.2208671.

A. Archambault and J. Grudin. A longitudinal study of facebook, linkedin, & twitter use. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2741–2750, 2012b.

M. Argyle. *The psychology of happiness*. Routledge, 2013.

V. Arnaboldi, M. Conti, A. Passarella, and R. Dunbar. Dynamics of personal social relationships in online social networks: a study on twitter. In *Proceedings of the first ACM conference on Online social networks*, pages 15–26, 2013.

T. Atkinson, R. Liem, and J. H. Liem. The social costs of unemployment: Implications for social support. *J. Health Soc. Behav*, 1986.

L. Backstrom and J. Kleinberg. Romantic partnerships and the dispersion of social ties: A network analysis of relationship status on facebook. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, page 831–841, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450325400. doi: 10.1145/2531602.2531642. URL https://doi.org/10.1145/2531602.2531642.

E. Bakshy, B. Karrer, and L. A. Adamic. Social influence and the diffusion of user-created content. In *EC*, 2009.

E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: Quantifying influence on twitter. In *WSDM*, 2011.

E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, page 519–528, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450312295. doi: 10.1145/2187836.2187907.

D. Balsamo, P. Bajardi, and A. Panisson. Firsthand opiates abuse on social media: monitoring geospatial patterns of interest through a digital cohort. In *The World Wide Web Conference*, pages 2572–2579, 2019.

R. F. Baumeister and M. R. Leary. The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin*, 117(3):497, 1995.

J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, volume 14, pages 830–839, 2020.

A. Bell. Language style as audience design. *Language in society*, 13(2):145–204, 1984.

C. Beller, R. Knowles, C. Harman, S. Bergsma, M. Mitchell, and B. Van Durme. I'm a belieber: Social roles via self-identification and conceptual attributes. In *ACL*, 2014.

C. R. Berger and R. J. Calabrese. Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. *Human communication research*, 1(2): 99–112, 1974.

S. Bergsma and B. Van Durme. Using conceptual class attributes to characterize social media users. In *ACL*, 2013.

D. E. Berlyne. Conflict, arousal, and curiosity. 1960.

E. Berscheid, M. Snyder, and A. M. Omoto. The relationship closeness inventory: Assessing the closeness of interpersonal relationships. *Journal of Personality and Social Psychology*, 57(5): 792–807, Nov 1989.

L. M. Bettencourt. The origins of scaling in cities. *science*, 340(6139):1438–1441, 2013.

L. M. Bettencourt, J. Lobo, and D. Strumsky. Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size. *Research policy*, 36(1):107–120, 2007.

C. Bicchieri. *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press, 2005.

M. Billig. *Laughter and ridicule: Towards a social critique of humour*. Sage, 2005.

P. Blau. *Exchange and power in social life*. Routledge, 2017.

D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 2003.

C. A. Bliss, M. R. Frank, C. M. Danforth, and P. S. Dodds. An evolutionary algorithm approach to link prediction in dynamic social networks. *Journal of Computational Science*, 5(5):750 – 764, 2014.

d. boyd, S. Golder, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *HICSS*, 2010.

J. J. Bradac. The language of lovers, flovers, and friends: Communicating in social and personal relationships. *Journal of Language and Social Psychology*, 2(2-3-4):141–162, 1983.

D. P. Braun and S. Plog. Evolution of "tribal" social networks: Theory and prehistoric north american evidence. *American Antiquity*, 47(3):504–525, 1982.

K. Bridge and L. A. Baxter. Blended relationships: Friends as work associates. *Western Journal of Communication*, 56(3):200–225, 1992.

M. Brinberg and N. Ram. Do new romantic couples use more similar language over time? evidence from intensive longitudinal text messages. *Journal of Communication*, 71(3):454–477, 2021.

D. E. Broadbent. A mechanical model for human attention and immediate memory. *Psychological review*, 64(3):205, 1957.

P. Brown, S. C. Levinson, and S. C. Levinson. *Politeness: Some universals in language usage*, volume 4. Cambridge university press, 1987.

T. S. Brugha, E. Sturt, B. MacCarthy, J. Potter, T. Wykes, and P. Bebbington. The interview measure of social relationships: the description and evaluation of a survey instrument for assessing personal social resources. *Social Psychiatry*, 22:123–128, 1987.

C. Buntain and J. Golbeck. Identifying social roles in reddit using network structure. In *Proceedings of the 23rd international conference on world wide web*, pages 615–620, 2014.

M. Burke and R. Kraut. Using facebook after losing a job: Differential benefits of strong and weak ties. In *CSCW*, 2013.

M. Burke and R. E. Kraut. Growing closer on facebook: Changes in tie strength through social network site use. In *CHI*, 2014.

M. Burke and R. E. Kraut. The relationship between facebook use and well-being depends on communication type and tie strength. *Journal of Computer-mediated Communication*, 2016.

M. Burke, L. Adamic, and K. Marciniak. Families on facebook. In *ICWSM*, 2013.

A. Burton, W. Haley, and B. Small. Bereavement after caregiving or unexpected death: Effects on elderly spouses. *Aging and Mental Health*, 2006.

M. N. Burton-Chellew and R. I. Dunbar. Romance and reproduction are socially costly. *Evolutionary Behavioral Sciences*, 9(4):229–241, 10 2015.

D. M. Buss. Strategies of human mating. *Psihologijske teme*, 15(2):239–260, 2006.

D. M. Buss and D. P. Schmitt. Sexual strategies theory: an evolutionary perspective on human mating. *Psychological review*, 100(2):204, 1993.

N. Cantor and W. Mischel. Prototypes in person perception. In *Advances in experimental social psychology*, volume 12, pages 3–52. Elsevier, 1979.

Q. Cao, H. Shen, K. Cen, W. Ouyang, and X. Cheng. Deephawkes: Bridging the gap between prediction and understanding of information cascades. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, page 1149–1158, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450349185. doi: 10.1145/3132847.3132973. URL https://doi.org/10.1145/3132847.3132973.

M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *ICWSM*, 2010.

A. Chadwick, N.-A. Hall, and C. Vaccari. Misinformation rules!? could "group rules" reduce misinformation in online personal messaging? *New Media & Society*, 0(0):14614448231172964, 2023. doi: 10.1177/14614448231172964.

N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

C.-P. Chen. Forming digital self and parasocial relationships on youtube. *Journal of Consumer culture*, 16(1):232–254, 2016.

T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

J. Cheng, D. M. Romero, B. Meeder, and J. Kleinberg. Predicting reciprocity in social networks. In *SocialCom*, 2011.

J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec. Can cascades be predicted? In *WWW*, 2014. ISBN 978-1-4503-2744-2.

J. Cheng, M. Bernstein, C. Danescu-Niculescu-Mizil, and J. 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, 2017.

R. Chetty, M. O. Jackson, T. Kuchler, J. Stroebel, N. Hendren, R. B. Fluegge, S. Gong, F. Gonzalez, A. Grondin, M. Jacob, et al. Social capital i: measurement and associations with economic mobility. *Nature*, 608(7921):108–121, 2022a.

R. Chetty, M. O. Jackson, T. Kuchler, J. Stroebel, N. Hendren, R. B. Fluegge, S. Gong, F. Gonzalez, A. Grondin, M. Jacob, et al. Social capital ii: determinants of economic connectedness. *Nature*, 608(7921):122–134, 2022b.

M. Choi, L. M. Aiello, K. Z. Varga, and D. Quercia. Ten social dimensions of conversations and relationships. In *WWW*, 2020.

M. Choi, C. Budak, D. M. Romero, and D. Jurgens. More than meets the tie: Examining the role of interpersonal relationships in social networks. In *ICWSM*, 2021.

M. D. Choudhury and S. De. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *ICWSM*, 2014.

G. Chowdhary and S. Bandyopadhyay. Ties that matter. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 2398–2403. IEEE, 2015.

K. Cigularov, P. Y. Chen, B. W. Thurber, and L. Stallones. What prevents adolescents from seeking help after a suicide education program? *Suicide and Life-Threatening Behavior*, 38(1):74–86, 2008.

C. Cobb and T. Kohno. How public is my private life? privacy in online dating. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, page 1231–1240, Republic and Canton of Geneva, CHE, 2017. International World Wide Web Conferences Steering Committee. ISBN 9781450349130. doi: 10.1145/3038912.3052592. URL https://doi.org/10.1145/3038912.3052592.

J. C. Coffee Jr. Understanding enron: It's about the gatekeepers, stupid. *Bus. Law.*, 57:1403, 2001.

J. Cohen. *Statistical power analysis for the behavioral sciences*. Academic press, 2013.

S. Cohen and T. A. Wills. Stress, social support, and the buffering hypothesis. *Psychological Bulletin*, 1985.

E. Colleoni, A. Rozza, and A. Arvidsson. Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data. *Journal of Communication*, 64(2):317–332, 03 2014.

N. L. Collins and B. C. Feeney. A safe haven: an attachment theory perspective on support seeking and caregiving in intimate relationships. *J. Pers. Soc. Psychol*, 2000.

R. Compton, D. Jurgens, and D. Allen. Geotagging one hundred million twitter accounts with total variation minimization. In *IEEE BigData*, 2014.

M. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, F. Menczer, and A. Flammini. Political polarization on twitter. In *Proceedings of the international aaai conference on web and social media*, volume 5, pages 89–96, 2011.

G. Cook. Repetition and learning by heart: An aspect of intimate discourse, and its implications. 1994.

L. Corbo, R. Corrado, and S. Ferriani. A new order of things: Network mechanisms of field evolution in the aftermath of an exogenous shock. *Organization Studies*, 2016.

J. N. Cummings. Work groups, structural diversity, and knowledge sharing in a global organization. *Management science*, 50(3):352–364, 2004.

T. O. Cunha, I. Weber, H. Haddadi, and G. L. Pappa. The effect of social feedback in a reddit weight loss community. In *Proceedings of the 6th International Conference on Digital Health Conference*, pages 99–103, 2016.

E. Cutrell and Z. Guan. What are you looking for? an eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 407–416, 2007.

C. E. Cutrona and D. W. Russell. *Type of social support and specific stress: Toward a theory of optimal matching*. John Wiley & Sons, 1990.

C. Danescu-Niculescu-Mizil and L. Lee. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. *arXiv preprint arXiv:1106.3077*, 2011.

C. Danescu-Niculescu-Mizil, L. Lee, B. Pang, and J. Kleinberg. Echoes of power: Language effects and power differences in social interaction. In *Proceedings of the 21st international conference on World Wide Web*, pages 699–708, 2012.

C. Danescu-Niculescu-Mizil, M. Sudhof, D. Jurafsky, J. Leskovec, and C. Potts. A computational approach to politeness with application to social factors. *arXiv preprint arXiv:1306.6078*, 2013.

L. Dang-Xuan, S. Stieglitz, J. Wladarsch, and C. Neuberger. An investigation of influentials and the role of sentiment in political communication on twitter during election periods. *Information, communication & society*, 16(5):795–825, 2013.

T. David-Barrett, J. Kertesz, A. Rotkirch, A. Ghosh, K. Bhattacharya, D. Monsivais, and K. Kaski. Communication with family and friends across the life course. *PLOS ONE*, 11(11):1–15, 11 2016.

T. Davidson, D. Warmsley, M. Macy, and I. Weber. Automated hate speech detection and the problem of offensive language. In *Proceedings of the international AAAI conference on web and social media*, volume 11, pages 512–515, 2017.

J. L. Davis and N. Jurgenson. Context collapse: Theorizing context collusions and collisions. *Information, communication & society*, 17(4):476–485, 2014.

A. G. De Belvis, M. Avolio, L. Sicuro, A. Rosano, E. Latini, G. Damiani, and W. Ricciardi. Social relationships and hrql: A cross-sectional survey among older italian adults. *BMC public health*, 8 (1):1–10, 2008.

M. De Choudhury and S. De. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 71–80, 2014.

M. De Choudhury and E. Kıcıman. The language of social support in social media and its effect on suicidal ideation risk. In *ICWSM*, 2017.

M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2098–2110, 2016.

B.-E. Deal, L. S. Martinez, B. H. Spitzberg, and M.-H. Tsou. "i definitely did not report it when i was raped...# webelievechristine# metoo": A content analysis of disclosures of sexual assault on twitter. *Social Media+ Society*, 2020.

S. DeDeo. Collective phenomena and non-finite state computation in a human social system. *PLoS one*, 8(10):e75818, 2013.

O. Demasi, Y. Li, and Z. Yu. A multi-persona chatbot for hotline counselor training. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3623–3636, 2020.

S. Deri, J. Rappaz, L. M. Aiello, and D. Quercia. Coloring in the links: Capturing social ties as they are perceived. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–18, 2018.

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

D. K. Dickinson and P. O. Tabors. *Beginning literacy with language: Young children learning at home and school.* Paul H Brookes Publishing, 2001.

J. Diesner and K. M. Carley. Exploration of communication networks from the enron email corpus. In *SIAM International Conference on Data Mining: Workshop on Link Analysis, Counterterrorism and Security, Newport Beach, CA*, pages 3–14, 2005.

E. Dubois, S. Minaeian, A. Paquet-Labelle, and S. Beaudry. Who to trust on social media: How opinion leaders and seekers avoid disinformation and echo chambers. *Social Media + Society*, 6 (2):2056305120913993, 2020. doi: 10.1177/2056305120913993.

R. I. Dunbar, V. Arnaboldi, M. Conti, and A. Passarella. The structure of online social networks mirrors those in the offline world. *Social Networks*, 2015.

R. I. M. Dunbar. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, 16(4):681–694, 1993.

P. Eckert. Gender and sociolinguistic variation. *Language and gender: A reader*, pages 64–75, 1998.

D. Eckles and E. Bakshy. Bias and high-dimensional adjustment in observational studies of peer effects. *Journal of the American Statistical Association*, 2021.

N. B. Ellison, C. Steinfield, and C. Lampe. The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168, 07 2007.

N. B. Ellison, R. Gray, J. Vitak, C. Lampe, and A. T. Fiore. Calling all facebook friends: Exploring requests for help on facebook. In *ICWSM*, 2013.

S. T. Emlen and L. W. Oring. Ecology, sexual selection, and the evolution of mating systems. *Science*, 197(4300):215–223, 1977.

K. B. Fägersten. *Who's swearing now?: the social aspects of conversational swearing*. Cambridge Scholars Publishing, 2012.

S. D. Farnham and E. F. Churchill. Faceted identity, faceted lives: social and technical issues with being yourself online. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pages 359–368, 2011.

E. Fast, B. Chen, and M. S. Bernstein. Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 4647–4657, 2016.

A. Feingold. Matching for attractiveness in romantic partners and same-sex friends: A meta-analysis and theoretical critique. *Psychological bulletin*, 104(2):226, 1988.

S. L. Feld. The focused organization of social ties. *American journal of sociology*, 86(5):1015–1035, 1981.

S. L. Feld. Social structural determinants of similarity among associates. *American Sociological Review*, 47(6):797–801, 1982.

G. Feldman, H. Lian, M. Kosinski, and D. Stillwell. Frankly, we do give a damn: The relationship between profanity and honesty. *Social Psychological and Personality Science*, 8(7):816–826, 2017.

S. Field. *Screenplay: The foundations of screenwriting*. Delta, 2005.

C. Fiesler and N. Proferes. "participant" perceptions of twitter research ethics. *Social Media + Society*, 4(1), 2018.

C. Finkenauer, R. C. Engels, and W. Meeus. Keeping secrets from parents: Advantages and disadvantages of secrecy in adolescence. *Journal of Youth and Adolescence*, 31:123–136, 2002.

A. P. Fiske. The four elementary forms of sociality: framework for a unified theory of social relations. *Psychological review*, 99(4):689, 1992.

S. T. Fiske, A. J. Cuddy, and P. Glick. Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences*, 11(2):77–83, 2007.

J. Fleming and R. Koman. *Web navigation: designing the user experience*. O'reilly Sebastopol, CA, 1998.

R. L. Florida et al. *Cities and the creative class*. Psychology Press, 2005.

E. B. Foa and U. G. Foa. Resource theory: Interpersonal behavior as exchange. In *Social exchange: Advances in theory and research*, pages 77–94. Springer, 1980.

J. G. Foster, D. V. Foster, P. Grassberger, and M. Paczuski. Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences*, 107(24):10815–10820, 2010.

B. D. Frampton and J. T. Child. Friend or not to friend: Coworker facebook friend requests as an application of communication privacy management theory. *Computers in Human Behavior*, 29 (6):2257–2264, 2013.

J. R. French, B. Raven, and D. Cartwright. The bases of social power. *Classics of organization theory*, 7(311-320):1, 1959.

J. R. French Jr. A formal theory of social power. *Psychological review*, 63(3):181, 1956.

T. Frijns, C. Finkenauer, A. A. Vermulst, and R. C. Engels. Keeping secrets from parents: Longitudinal associations of secrecy in adolescence. *Journal of Youth and Adolescence*, 34: 137–148, 2005.

L. Fu, L. Lee, and C. Danescu-Niculescu-Mizil. When confidence and competence collide: Effects on online decision-making discussions. In *Proceedings of the 26th international conference on world wide web*, pages 1381–1390, 2017.

D. Gaffney and J. N. Matias. Caveat emptor, computational social science: Large-scale missing data in a widely-published reddit corpus. *PloS one*, 13(7):e0200162, 2018.

I. Gagliardone, D. Gal, T. Alves, and G. Martinez. *Countering online hate speech*. Unesco Publishing, 2015.

X. Gao, Y. Zhao, J. Zhang, and L. Cai. Pairwise emotional relationship recognition in drama videos: Dataset and benchmark. In *Proceedings of the 29th ACM International Conference on Multimedia*, MM '21, page 3380–3389, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450386517. doi: 10.1145/3474085.3475493. URL https://doi.org/10.1145/3474085.3475493.

K. Garimella, J. Cohen, and I. Weber. Understanding para social breakups on twitter. In *WebSci*, 2017.

V. R. K. Garimella, I. Weber, and S. Dal Cin. From "i love you babe" to "leave me alone"-romantic relationship breakups on twitter. In *ICSI*, 2014.

L. K. Gee, J. Jones, and M. Burke. Social networks and labor markets: How strong ties relate to job finding on facebook's social network. *Journal of Labor Economics*, 2017.

E. Gilbert. Predicting tie strength in a new medium. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, page 1047–1056, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450310864. doi: 10.1145/2145204.2145360. URL https://doi.org/10.1145/2145204.2145360.

E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, page 211–220, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605582467. doi: 10.1145/1518701.1518736. URL https://doi.org/10.1145/1518701.1518736.

H. Gilchrist and G. Sullivan. Barriers to help-seeking in young people: Community beliefs about youth suicide. *Australian Social Work*, 59(1):73–85, 2006.

S. Giorgi, D. Bellew, D. R. S. Habib, J. Sedoc, C. Smitterberg, A. Devoto, M. Himelein-Wachowiak, and B. Curtis. Lived experience matters: Automatic detection of stigma toward people who use substances on social media. *arXiv preprint arXiv:2302.02064*, 2023.

R. Glott, P. Schmidt, and R. Ghosh. Wikipedia survey–overview of results. *United Nations University: Collaborative Creativity Group*, 8:1158–1178, 2010.

A. Goel, K. T. Ma, and C. Tan. An end-to-end network for generating social relationship graphs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11186–11195, 2019.

S. A. Golder and M. W. Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881, 2011.

D. J. Goldsmith and L. A. Baxter. Constituting relationships in talk: A taxonomy of speech events in social and personal relationships. *Human Communication Research*, 23(1):87–114, 1996.

S. Golombok, C. Murray, P. Brinsden, and H. Abdalla. Social versus biological parenting: family functioning and the socioemotional development of children conceived by egg or sperm donation. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40(4):519–527, 1999.

Y. Gorodnichenko, T. Pham, and O. Talavera. Social media, sentiment and public opinions: Evidence from# brexit and# uselection. *European Economic Review*, 136:103772, 2021.

A. Gorrese and R. Ruggieri. Peer attachment: A meta-analytic review of gender and age differences and associations with parent attachment. *Journal of Youth and Adolescence*, 41(5):650–672, 2012.

J. G. Graham. English language proficiency and the prediction of academic success. *TESOL quarterly*, 21(3):505–521, 1987.

M. Granovetter. The strength of weak ties: A network theory revisited. *Sociological Theory*, 1: 201–233, 1983.

M. S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

G. M. Green. *Pragmatics and natural language understanding*. Psychology Press, 1996.

K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber. Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 28(10):2222–2232, 2016.

L. K. Guerrero and P. A. Mongeau. On becoming "more than friends": The transition from friendship to romantic relationship. *Handbook of relationship initiation*, pages 175–194, 2008.

A. Haas and M. A. Sherman. Conversational topic as a function of role and gender. *Psychological Reports*, 51(2):453–454, 1982.

J. Haidt and J. Graham. When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social justice research*, 20(1):98–116, 2007.

O. L. Haimson, N. Andalibi, M. D. Choudhury, and G. R. Hayes. Relationship breakup disclosures and media ideologies on facebook. *New Media & Society*, 2018.

W. W. Hartup. Social relationships and their developmental significance. *American psychologist*, 44 (2):120, 1989.

W. W. Hartup and N. Stevens. Friendships and adaptation across the life span. *Current Directions in Psychological Science*, 8(3):76–79, 1999.

W. W. Hartup, C. Shantz, N. Eisenberg, R. Emde, L. Hoffman, E. E. Maccoby, F. J. Monks, R. Parke, M. Rutter, and C. Zahn-Waxler. *The development of romantic relationships in adolescence*. Cambridge University Press, 1999.

A. F. Hayes and K. Krippendorff. Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 2007.

R. B. Hays. The development and maintenance of friendship. *Journal of Social and Personal Relationships*, 1(1):75–98, 1984.

H. He, Y. Bai, E. A. Garcia, and S. Li. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, pages 1322–1328. IEEE, 2008.

M. Henderson and M. Argyle. Social support by four categories of work colleagues: Relationships between activities, stress and satisfaction. *Journal of Organizational Behavior*, 1985.

S. Henderson, P. Duncan-Jones, D. Byrne, and R. Scott. Measuring social relationships the interview schedule for social interaction. *Psychological medicine*, 10(4):723–734, 1980.

W. R. Hobbs and M. K. Burke. Connective recovery in social networks after the death of a friend. *Nature Human Behaviour*, 2017.

S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

C. Hoffner. Parasocial and online social relationships. *The handbook of children, media, and development*, pages 309–333, 2008.

L. Hong, O. Dan, and B. D. Davison. Predicting popular messages in twitter. In *WWW*, 2011.

D. Hovy and D. Yang. The importance of modeling social factors of language: Theory and practice. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 588–602, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.49. URL https://aclanthology.org/2021.naacl-main.49.

A. M. Hoyle, P. Goel, A. Hian-Cheong, D. Peskov, J. L. Boyd-Graber, and P. Resnik. Is automated topic model evaluation broken? the incoherence of coherence. In M. Ranzato, A. Beygelzimer, Y. N. Dauphin, P. Liang, and J. W. Vaughan, editors, *NeuRIPS*, 2021.

B. Huang and K. Carley. A hierarchical location prediction neural network for twitter user geolocation. In *EMNLP*, 2019.

J. Huh, B. C. Kwon, S.-H. Kim, S. Lee, J. Choo, J. Kim, M.-J. Choi, and J. S. Yi. Personas in online health communities. *Journal of Biomedical Informatics*, 2016.

C. Hutto and E. Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225, 2014.

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

M. Iyyer, A. Guha, S. Chaturvedi, J. Boyd-Graber, and H. Daumé III. Feuding families and former friends: Unsupervised learning for dynamic fictional relationships. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1534–1544, 2016.

M. O. Jackson et al. *Social and economic networks*, volume 3. Princeton university press Princeton, 2008.

P. B. Jackson. Specifying the buffering hypothesis: Support, strain, and depression. *Soc. Psychol. Q*, 1992.

S. E. Jackson and J. E. Dutton. Discerning threats and opportunities. *Administrative Science Quarterly*, 1988.

M. Jahoda. Work, employment, and unemployment: Values, theories, and approaches in social research. *American Psychologist*, 1981.

K. Jaidka, S. Guntuku, and L. Ungar. Facebook versus twitter: Differences in self-disclosure and trait prediction. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 12, 2018.

M. Jenders, G. Kasneci, and F. Naumann. Analyzing and predicting viral tweets. In *WWW*, 2013.

Q. Jia, H. Huang, and K. Q. Zhu. Ddrel: A new dataset for interpersonal relation classification in dyadic dialogues. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13125–13133, 2021.

Y. Jiang, Y. Xu, Y. Zhan, W. He, Y. Wang, Z. Xi, M. Wang, X. Li, Y. Li, and Y. Yu. The CRECIL corpus: a new dataset for extraction of relations between characters in Chinese multi-party dialogues. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2337–2344, Marseille, France, June 2022. European Language Resources Association. URL https://aclanthology.org/2022.lrec-1.250.

J. J. Jones, J. E. Settle, R. M. Bond, C. J. Fariss, C. Marlow, and J. H. Fowler. Inferring tie strength from online directed behavior. *PLOS ONE*, 8(1):1–6, 01 2013. doi: 10.1371/journal.pone.0052168. URL https://doi.org/10.1371/journal.pone.0052168.

D. Jurafsky. *Speech & language processing*. Pearson Education India, 2000.

D. Jurgens, A. Seth, J. Sargent, A. Aghighi, and M. Geraci. Your spouse needs professional help: Determining the contextual appropriateness of messages through modeling social relationships. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10994–11013, Toronto, Canada, July 2023. Association for Computational Linguistics. URL https://aclanthology.org/2023.acl-long.616.

A. K. Kehrberg. 'i love you, please notice me': the hierarchical rhetoric of twitter fandom. *Celebrity Studies*, 6(1):85–99, 2015.

L. Keijsers, S. J. Branje, T. Frijns, C. Finkenauer, and W. Meeus. Gender differences in keeping secrets from parents in adolescence. *Developmental psychology*, 46(1):293, 2010.

K. S. Kendler, L. M. Karkowski, and C. A. Prescott. Causal relationship between stressful life events and the onset of major depression. *American Journal of Psychiatry*, 1999.

H. Kim, G. J. Kim, H. W. Park, and R. E. Rice. Configurations of relationships in different media: Ftf, email, instant messenger, mobile phone, and sms. *Journal of Computer-Mediated Communication*, 12(4):1183–1207, 2007.

J. Kim and H. Song. Celebrity's self-disclosure on twitter and parasocial relationships: A mediating role of social presence. *Computers in Human Behavior*, 62:570–577, 2016.

Y. Kim. Convolutional neural networks for sentence classification. In *EMNLP*, 2014.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.

A. P. Kirilenko and S. Stepchenkova. Inter-coder agreement in one-to-many classification: fuzzy kappa. *PloS one*, 11(3):e0149787, 2016.

K. J. Klein and T. A. D'Aunno. Psychological sense of community in the workplace. *Journal of Community Psychology*, 14(4):365–377, 1986.

B. Klimt and Y. Yang. The enron corpus: A new dataset for email classification research. In *Machine Learning: ECML 2004: 15th European Conference on Machine Learning, Pisa, Italy, September 20-24, 2004. Proceedings 15*, pages 217–226. Springer, 2004.

M. L. Knapp, D. G. Ellis, and B. A. Williams. Perceptions of communication behavior associated with relationship terms. *Communication Monographs*, 47(4):262–278, 1980.

G. Kossinets and D. J. Watts. Empirical analysis of an evolving social network. *Science*, 311(5757): 88–90, 2006.

B. Kovacs and A. M. Kleinbaum. Language-style similarity and social networks. *Psychological science*, 31(2):202–213, 2020.

A. D. Kramer, J. E. Guillory, and J. T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National academy of Sciences of the United States of America*, 111(24):8788, 2014.

S. Kumar, W. L. Hamilton, J. Leskovec, and D. Jurafsky. Community interaction and conflict on the web. In *Proceedings of the 2018 world wide web conference*, pages 933–943, 2018.

A. Kupavskii, L. Ostroumova, A. Umnov, S. Usachev, P. Serdyukov, G. Gusev, and A. Kustarev. Prediction of retweet cascade size over time. In *CIKM*, 2012.

H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600, 2010.

L. La Cava, L. M. Aiello, and A. Tagarelli. Get out of the nest! drivers of social influence in the# twittermigration to mastodon. *arXiv preprint arXiv:2305.19056*, 2023.

Y. Lavee, H. I. McCubbin, and D. H. Olson. The effect of stressful life events and transitions on family functioning and well-being. *Journal of Marriage and the Family*, 1987.

C. R. LeanaI and D. C. Feldman. Individual responses to job loss: Empirical findings from two field studies. *Human Relations*, 1990.

N. Lee and M. Nathan. Knowledge workers, cultural diversity and innovation: evidence from london. *International journal of knowledge-based development*, 1(1-2):53–78, 2010.

S. J. Lee. Online Communication and Adolescent Social Ties: Who benefits more from Internet use?*. *Journal of Computer-Mediated Communication*, 14(3):509–531, 04 2009.

R. T. A. Leenders. Longitudinal behavior of network structure and actor attributes: modeling interdependence of contagion and selection. In *Evolution of social networks*, pages 165–184. Routledge, 2013.

J. Leskovec, D. Huttenlocher, and J. Kleinberg. Predicting positive and negative links in online social networks. In *WWW*, 2010a.

J. Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. In *CHI*, 2010b.

D. Z. Levin and R. Cross. The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management science*, 50(11):1477–1490, 2004.

R. J. Lewicki, B. B. Bunker, et al. Developing and maintaining trust in work relationships. *Trust in organizations: Frontiers of theory and research*, 114:139, 1996.

C. Li, J. Ma, X. Guo, and Q. Mei. Deepcas: An end-to-end predictor of information cascades. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, page 577–586, Republic and Canton of Geneva, CHE, 2017a. International World Wide Web Conferences Steering Committee. ISBN 9781450349130. doi: 10.1145/3038912.3052643. URL https://doi.org/10.1145/3038912.3052643.

J. Li, Y. Wong, Q. Zhao, and M. S. Kankanhalli. Dual-glance model for deciphering social relationships. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017b.

W. Li, Y. Duan, J. Lu, J. Feng, and J. Zhou. Graph-based social relation reasoning. In *European conference on computer vision*, pages 18–34. Springer, 2020.

Y. Li and T. Yang. Word embedding for understanding natural language: a survey. *Guide to big data applications*, pages 83–104, 2018.

C. A. Liebler and G. D. Sandefur. Gender differences in the exchange of social support with friends, neighbors, and co-workers at midlife. *Social Science Research*, 2002.

C. X. Ling and C. Li. Data mining for direct marketing: Problems and solutions. In *Kdd*, volume 98, pages 73–79, 1998.

Z. C. Lipton, J. Berkowitz, and C. Elkan. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*, 2015.

X. Liu, W. Liu, M. Zhang, J. Chen, L. Gao, C. Yan, and T. Mei. Social relation recognition from videos via multi-scale spatial-temporal reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019a.

Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692, 2019b.

N. Luhmann. *Trust and power*. John Wiley & Sons, 2018.

X. Ma, J. T. Hancock, K. Lim Mingjie, and M. Naaman. Self-disclosure and perceived trustworthiness of airbnb host profiles. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pages 2397–2409, 2017.

D. Maldeniya, C. Budak, L. P. Robert Jr, and D. M. Romero. Herding a deluge of good samaritans: How github projects respond to increased attention. In *WebConf*, 2020.

P. V. Marsden and K. E. Campbell. Measuring tie strength. *Social forces*, 63(2):482–501, 1984.

B. D. Marshall and S. Galea. Formalizing the role of agent-based modeling in causal inference and epidemiology. *American journal of epidemiology*, 181(2):92–99, 2015.

T. Martin, J. M. Hofman, A. Sharma, A. Anderson, and D. J. Watts. Exploring limits to prediction in complex social systems. In *WWW*, 2016.

G. Marwell and J. Hage. The organization of role-relationships: A systematic description. *American Sociological Review*, 35(5):884–900, 1970.

A. Marwick and D. Boyd. To see and be seen: Celebrity practice on twitter. *Convergence*, 17(2): 139–158, 2011a.

A. E. Marwick and D. Boyd. I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society*, 13(1):114–133, 2011b.

K. A. Mason and M. L. Benson. The effect of social support on fraud victims' reporting behavior: A research note. *Justice Quarterly*, 1996.

J. Massen, E. Sterck, and H. de Vos. Close social associations in animals and humans: functions and mechanisms of friendship. *Behaviour*, 147(11):1379–1412, 2010.

A. K. McCallum. Mallet: A machine learning for language toolkit. `http://mallet.cs.umass.edu`, 2002. Accessed: 2023-04-15.

J. McGee, J. A. Caverlee, and Z. Cheng. A geographic study of tie strength in social media. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, page 2333–2336, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450307178. doi: 10.1145/2063576.2063959. URL `https://doi.org/10.1145/2063576.2063959`.

J. McGee, J. Caverlee, and Z. Cheng. Location prediction in social media based on tie strength. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, CIKM '13, page 459–468, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450322638. doi: 10.1145/2505515.2505544. URL https://doi.org/10.1145/2505515.2505544.

M. L. McHugh. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282, 2012.

B. McLean and P. Elkind. *The smartest guys in the room: The amazing rise and scandalous fall of Enron*. Penguin, 2013.

M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

A. N. Medvedev, R. Lambiotte, and J.-C. Delvenne. The anatomy of reddit: An overview of academic research. *Dynamics On and Of Complex Networks III: Machine Learning and Statistical Physics Approaches 10*, pages 183–204, 2019.

J.-K. Min, J. Wiese, J. I. Hong, and J. Zimmerman. Mining smartphone data to classify life-facets of social relationships. In *CSCW*, 2013.

I. Mitchell, L. Evans, T. Rees, and L. Hardy. Stressors, social support, and tests of the buffering hypothesis: Effects on psychological responses of injured athletes. *British Journal of Health Psychology*, 2014.

T. Mitra and E. Gilbert. The language that gets people to give: Phrases that predict success on kickstarter. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 49–61, 2014.

N. P. Moller, R. T. Fouladi, C. J. McCarthy, and K. D. Hatch. Relationship of attachment and social support to college students' adjustment following a relationship breakup. *Journal of Counseling & Development*, 2003.

B. L. Monroe, M. P. Colaresi, and K. M. Quinn. Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16(4):372–403, 2008.

S. A. Morelli, D. C. Ong, R. Makati, M. O. Jackson, and J. Zaki. Empathy and well-being correlate with centrality in different social networks. *Proceedings of the National Academy of Sciences*, 114(37):9843–9847, 2017.

L. Morris and S. Irwin. Unemployment and informal support: dependency, exclusion, or participation? *Work, Employment and Society*, 1992.

S. A. Myers and A. D. Johnson. Perceived solidarity, self-disclosure, and trust in organizational peer relationships. *Communication Research Reports*, 21(1):75–83, 2004.

D. Q. Nguyen, T. Vu, and A. T. Nguyen. Bertweet: A pre-trained language model for english tweets. In *EMNLP*, 2020.

M. Nouwens, C. F. Griggio, and W. E. Mackay. " whatsapp is for family; messenger is for friends" communication places in app ecosystems. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 727–735, 2017.

P. J. Oakes, S. A. Haslam, and J. C. Turner. *Stereotyping and social reality.* Blackwell Publishing, 1994.

H. J. Oh and R. LaRose. Impression management concerns and support-seeking behavior on social network sites. *Computers in Human Behavior*, 2016.

R. Oldenburg. *The great good place: Cafes, coffee shops, bookstores, bars, hair salons, and other hangouts at the heart of a community*. Da Capo Press, 1999.

F. K. Ozenc and S. D. Farnham. Life "modes" in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 561–570, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450302289. doi: 10.1145/1978942.1979022. URL https://doi.org/10.1145/1978942.1979022.

J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein. Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2304.03442*, 2023.

P. S. Park, J. E. Blumenstock, and M. W. Macy. The strength of long-range ties in population-scale social networks. *Science*, 2018.

J. W. Pennebaker, M. E. Francis, and R. J. Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001, 2001.

J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn. The development and psychometric properties of liwc2015. https://repositories.lib.utexas.edu/handle/2152/31333, 2015. Accessed: 2021-04-10.

J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

K. Peterson, M. Hohensee, and F. Xia. Email formality in the workplace: A case study on the enron corpus. In *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, pages 86–95, 2011.

S. Petrovic, M. Osborne, and V. Lavrenko. Rt to win! predicting message propagation in twitter. In *ICWSM*, 2011.

F. Pierri, C. Piccardi, and S. Ceri. Topology comparison of twitter diffusion networks effectively reveals misleading information. *Scientific reports*, 10(1):1–9, 2020.

J. M. Podolny and J. N. Baron. Resources and relationships: Social networks and mobility in the workplace. *American Sociological Review*, 1997.

M. Polignano, P. Basile, G. Rossiello, M. de Gemmis, and G. Semeraro. Learning inclination to empathy from social media footprints. In *Proceedings of the 25th conference on user modeling, adaptation and personalization*, pages 383–384, 2017.

C. D. Proctor and V. K. Groze. Risk factors for suicide among gay, lesbian, and bisexual youths. *Social work*, 39(5):504–513, 1994.

L. Qiu, Y. Liang, Y. Zhao, P. Lu, B. Peng, Z. Yu, Y. N. Wu, and S.-C. Zhu. SocAoG: Incremental graph parsing for social relation inference in dialogues. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 658–670, Online, Aug. 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.54. URL https://aclanthology.org/2021.acl-long.54.

D. Quercia, L. Capra, and J. Crowcroft. The social world of twitter: Topics, geography, and emotions. In *ICWSM*, 2012.

A. R. Radcliffe-Brown. On joking relationships. *Africa*, 13(3):195–210, 1940.

K. Rajkumar, G. Saint-Jacques, I. Bojinov, E. Brynjolfsson, and S. Aral. A causal test of the strength of weak ties. *Science*, 377(6612):1304–1310, 2022.

P. Ramaciotti Morales and J.-P. Cointet. Auditing the effect of social network recommendations on polarization in geometrical ideological spaces. In *Proceedings of the 15th ACM Conference on Recommender Systems*, pages 627–632, 2021.

S. Ranganath, J. Tang, X. Hu, H. Sundaram, and H. Liu. Leveraging social foci for information seeking in social media. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.

F. Rashid and E. Blanco. Dimensions of interpersonal relationships: Corpus and experiments. In *EMNLP*, 2017.

F. Rashid and E. Blanco. Characterizing interactions and relationships between people. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4395–4404, 2018.

A. E. Rautman. Resource variability, risk, and the structure of social networks: An example from the prehistoric southwest. *American antiquity*, 58(3):403–424, 1993.

S. M. Reich, K. Subrahmanyam, and G. Espinoza. Friending, iming, and hanging out face-to-face: overlap in adolescents' online and offline social networks. *Developmental psychology*, 48(2):356, 2012.

D. Reinhardt, F. Engelmann, A. Moerov, and M. Hollick. Show me your phone, i will tell you who your friends are: Analyzing smartphone data to identify social relationships. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, MUM '15, page 75–83, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450336055. doi: 10.1145/2836041.2836048. URL https://doi.org/10.1145/2836041.2836048.

H. T. Reis, W. A. Collins, and E. Berscheid. The relationship context of human behavior and development. *Psychological bulletin*, 126(6):844, 2000.

M. H. Richey and H. W. Richey. The significance of best-friend relationships in adolescence. *Psychology in the Schools*, 17(4):536–540, 1980.

J. Rivas. Friendship selection. *International Journal of Game Theory*, 38(4):521, 2009.

C. Roberts. Language socialization in the workplace. *Annual Review of Applied Linguistics*, 30: 211–227, 2010.

S. G. B. Roberts and R. I. M. Dunbar. Communication in social networks: Effects of kinship, network size, and emotional closeness. *Personal Relationships*, 18(3), 2011.

A. Robertson, L. M. Aiello, and D. Quercia. The language of dialogue is complex. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 428–439, 2019.

D. Romero, C. Tan, and J. Ugander. On the interplay between social and topical structure. In *ICWSM*, 2013a.

D. M. Romero, B. Meeder, and J. Kleinberg. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. In *WWW*, 2011.

D. M. Romero, C. Tan, and J. Ugander. On the interplay between social and topical structure. In *ICWSM*, 2013b.

D. M. Romero, B. Uzzi, and J. Kleinberg. Social networks under stress. In *WWW*, 2016.

R. Rotabi, K. Kamath, J. Kleinberg, and A. Sharma. Detecting strong ties using network motifs. In *Proceedings of the 26th International Conference on World Wide Web Companion*, WWW '17 Companion, page 983–992, Republic and Canton of Geneva, CHE, 2017. International World Wide Web Conferences Steering Committee. ISBN 9781450349147. doi: 10.1145/3041021.3055139. URL https://doi.org/10.1145/3041021.3055139.

G. M. Sandstrom and E. W. Dunn. Social interactions and well-being: The surprising power of weak ties. *Personality and Social Psychology Bulletin*, 40(7):910–922, 2014.

F. P. Santos, Y. Lelkes, and S. A. Levin. Link recommendation algorithms and dynamics of polarization in online social networks. *Proceedings of the National Academy of Sciences*, 118 (50):e2102141118, 2021.

B. B. Schieffelin and E. Ochs. Language socialization. *Annual review of anthropology*, 15(1): 163–191, 1986.

B. Settles. Active learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2009. URL http://digital.library.wisc.edu/1793/60660.

H. N. Shilubane, R. A. Ruiter, A. E. Bos, B. van den Borne, S. James, and P. S. Reddy. Psychosocial determinants of suicide attempts among black south african adolescents: A qualitative analysis. *Journal of Youth Studies*, 15(2):177–189, 2012.

F. Shirani and K. Henwood. Taking one day at a time: Temporal experiences in the context of unexpected life course transitions. *Time & Society*, 2011.

N. Shulman. Life-cycle variations in patterns of close relationships. *Journal of Marriage and Family*, 37(4), 1975.

P. M. Sias and D. J. Cahill. From coworkers to friends: The development of peer friendships in the workplace. *Western Journal of Communication (includes Communication Reports)*, 62(3): 273–299, 1998.

P. M. Sias, E. B. Gallagher, I. Kopaneva, and H. Pedersen. Maintaining workplace friendships: Perceived politeness and predictors of maintenance tactic choice. *Communication Research*, 39 (2):239–268, 2012.

M. Sikora, A. Seguin-Orlando, V. C. Sousa, A. Albrechtsen, T. Korneliussen, A. Ko, S. Rasmussen, I. Dupanloup, P. R. Nigst, M. D. Bosch, et al. Ancient genomes show social and reproductive behavior of early upper paleolithic foragers. *Science*, 358(6363):659–662, 2017.

D. M. Sluss and B. E. Ashforth. Relational identity and identification: Defining ourselves through work relationships. *The Academy of Management Review*, 32(1), 2007.

E. B. Smith, T. Menon, and L. Thompson. Status differences in the cognitive activation of social networks. *Organization Science*, 2012.

H. Song, W.-N. Zhang, J. Hu, and T. Liu. Generating persona consistent dialogues by exploiting natural language inference. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8878–8885, 2020.

L. Spencer and R. Pahl. Rethinking friendship. In *Rethinking Friendship*. Princeton University Press, 2018.

S. Sprecher, D. Felmlee, S. Metts, B. Fehr, and D. Vanni. Factors associated with distress following the breakup of a close relationship. *Journal of Social and Personal Relationships*, 1998.

L. Stafford and D. J. Canary. Maintenance strategies and romantic relationship type, gender and relational characteristics. *Journal of Social and Personal Relationships*, 8(2):217–242, 1991.

G. Stever and K. Lawson. Twitter as a way for celebrities to communicate with fans: Implications for the study of parasocial interaction. *North American Journal of Psychology*, 15(2):339–354, 01 2013.

A.-A. Stoica, C. Riederer, and A. Chaintreau. Algorithmic glass ceiling in social networks: The effects of social recommendations on network diversity. In *Proceedings of the 2018 World Wide Web Conference*, pages 923–932, 2018.

K. Subrahmanyam, S. M. Reich, N. Waechter, and G. Espinoza. Online and offline social networks: Use of social networking sites by emerging adults. *Journal of applied developmental psychology*, 29(6):420–433, 2008.

B. Suh, L. Hong, P. Pirolli, and E. H. Chi. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *SocialCom*, 2010.

M. Sundermeyer, R. Schlüter, and H. Ney. Lstm neural networks for language modeling. In *Thirteenth annual conference of the international speech communication association*, 2012.

E. Szkody and C. McKinney. Stress-buffering effects of social support on depressive problems: Perceived vs. received support and moderation by parental depression. *Journal of Child and Family Studies*, 2019.

H. Tajfel. *Social identity and intergroup relations*, volume 7. Cambridge University Press, 2010.

H. Tajfel, J. C. Turner, W. G. Austin, and S. Worchel. An integrative theory of intergroup conflict. *Organizational identity: A reader*, 56(65):9780203505984–16, 1979.

C. Tan, V. Niculae, C. Danescu-Niculescu-Mizil, and L. Lee. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th international conference on world wide web*, pages 613–624, 2016.

J. Tang, X. Wang, and H. Liu. Integrating social media data for community detection. In M. Atzmueller, A. Chin, D. Helic, and A. Hotho, editors, *MSM*, 2012.

Y. Tay, L. Tuan, and S. Hui. Couplenet: Paying attention to couples with coupled attention for relationship recommendation. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 12, 2018.

S. Tchokni, D. O. Séaghdha, and D. Quercia. Emoticons and phrases: Status symbols in social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, pages 485–494, 2014.

R. Thackeray, B. T. Crookston, J. H. West, et al. Correlates of health-related social media use among adults. *Journal of medical Internet research*, 15(1):e2297, 2013.

A. Tigunova, P. Mirza, A. Yates, and G. Weikum. PRIDE: Predicting Relationships in Conversations. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4636–4650, Online and Punta Cana, Dominican Republic, Nov. 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.380. URL https://aclanthology.org/2021.emnlp-main.380.

L. Townsend and C. Wallace. Social media research: A guide to ethics. The University of Aberdeen, 2016.

R. Tukachinsky. Para-romantic love and para-friendships: Development and assessment of a multiple-parasocial relationships scale. *American Journal of Media Psychology*, 3(1/2):73–94, 2010.

V. H. Tuulos and H. Tirri. Combining topic models and social networks for chat data mining. In *WI*, 2004.

D. Umberson. Family status and health behaviors: Social control as a dimension of social integration. *Journal of health and social behavior*, pages 306–319, 1987.

B. S. Vanneste, P. Puranam, and T. Kretschmer. Trust over time in exchange relationships: Meta-analysis and theory. *Strategic Management Journal*, 35(12):1891–1902, 2014.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

A. Vaux. *Social support: Theory, research, and intervention.* Praeger publishers, 1988.

A. Vaux and D. Harrison. Support network characteristics associated with support satisfaction and perceived support. *American Journal of Community Psychology*, 1985.

L. M. Verbrugge. A research note on adult friendship contact: A dyadic perspective. *Social Forces*, 62(1):78–83, 1983.

T. F. Waddell and C. Moss. Fake news in the family: How family communication patterns and conflict history affect the intent to correct misinformation among family members. *Communication Studies*, 74(3):183–199, 2023.

A. Wang, W. L. Hamilton, and J. Leskovec. Learning linguistic descriptors of user roles in online communities. In *Proceedings of the first workshop on nlp and computational social science*, pages 76–85, 2016a.

Y.-C. Wang, M. Burke, and R. Kraut. Modeling self-disclosure in social networking sites. In *CSCW*, 2016b.

Z. Wang and D. Jurgens. It's going to be okay: Measuring access to support in online communities. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 33–45, 2018.

Z. Wang, T. Chen, J. Ren, W. Yu, H. Cheng, and L. Lin. Deep reasoning with knowledge graph for social relationship understanding. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, IJCAI'18, page 1021–1028. AAAI Press, 2018. ISBN 9780999241127.

Z. Wang, S. Hale, D. I. Adelani, P. Grabowicz, T. Hartman, F. Flöck, and D. Jurgens. Demographic inference and representative population estimates from multilingual social media data. In *WWW*, 2019.

C. Welch, V. Pérez-Rosas, J. K. Kummerfeld, and R. Mihalcea. Look who's talking: Inferring speaker attributes from personal longitudinal dialog. In *CICLing*, 2019.

B. Wellman and S. Wortley. Different strokes from different folks: Community ties and social support. *American journal of Sociology*, 96(3):558–588, 1990.

Q. Wen, P. A. Gloor, A. Fronzetti Colladon, P. Tickoo, and T. Joshi. Finding top performers through email patterns analysis. *Journal of Information Science*, 46(4):508–527, 2020.

H. C. White. Notes on the constituents of social structure. soc. rel. 10-spring'65. *Sociologica*, 2(1): 0–0, 2008.

M. Wish, M. Deutsch, and S. J. Kaplan. Perceived dimensions of interpersonal relations. *Journal of Personality and social Psychology*, 33(4):409, 1976.

M. Wolfenstein. Movie analyses in the study of culture. *Film and nationalism*, pages 68–86, 2002.

R. Xiang, J. Neville, and M. Rogati. Modeling relationship strength in online social networks. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, page 981–990, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781605587998. doi: 10.1145/1772690.1772790. URL https://doi.org/10.1145/1772690.1772790.

C. Yang, J. Zhang, H. Wang, S. Li, M. Kim, M. Walker, Y. Xiao, and J. Han. Relation learning on social networks with multi-modal graph edge variational autoencoders. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 699–707, 2020.

D. Yang, Z. Yao, J. Seering, and R. Kraut. The channel matters: Self-disclosure, reciprocity and social support in online cancer support groups. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–15, 2019.

J. Yang and S. Counts. Predicting the speed, scale, and range of information diffusion in twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 4, pages 355–358, 2010.

D. Yu, K. Sun, C. Cardie, and D. Yu. Dialogue-based relation extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4927–4940, 2020.

M. Yu, W. Si, G. Song, Z. Li, and J. Yen. Who were you talking to-mining interpersonal relationships from cellphone network data. In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, pages 485–490. IEEE, 2014.

N. J. Yuan, Y. Zhong, F. Zhang, X. Xie, C.-Y. Lin, and Y. Rui. Who will reply to/retweet this tweet?: The dynamics of intimacy from online social interactions. In *WSDM*, 2016.

A. Zaheer, B. McEvily, and V. Perrone. Does trust matter? exploring the effects of interorganizational and interpersonal trust on performance. *Organization science*, 9(2):141–159, 1998.

A. F. Zhang, D. Livneh, C. Budak, L. Robert, and D. M. Romero. Shocking the crowd: The effect of censorship shocks on chinese wikipedia. In *ICWSM*, 2017.

Y. Zheng, R. Zhang, M. Huang, and X. Mao. A pre-training based personalized dialogue generation model with persona-sparse data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9693–9700, 2020.