Understanding and Reshaping Social Networks with Advanced Computational Techniques

by

Yuan Yuan

Submitted to the Institute for Data, Systems, and Society
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Social and Engineering Systems and Statistics
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September 2021

Author .....
Institute for Data, Systems, and Society
July 22, 2021

Certified by .....
Alex 'Sandy' Pentland
Toshiba Professor of Media Arts and Sciences, MIT
Thesis Supervisor

Certified by .....
Dean Eckles
Associate Professor of Marketing, MIT
Thesis Committee Member

Certified by .....
Sinan Aral
David Austin Professor of Management, MIT
Thesis Committee Member

Certified by .....
Esteban Moro
Associate Professor, Universidad Carlos III de Madrid and MIT
Thesis Committee Member

Accepted by .....
Fotini Christia
Chair, Social and Engineering Systems Program

Understanding and Reshaping Social Networks with Advanced Computational Techniques

by

Yuan Yuan

Submitted to the Institute for Data, Systems, and Society
on July 22, 2021, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Social and Engineering Systems and Statistics

Abstract

Social networks are powerful in modeling interdependence among individuals. Recently, the availability of large-scale social network data and advances in computational tools have facilitated the rapid development in social network research. However, a few important aspects of social networks have been understudied, and advanced computational tools may not directly help social scientists draw scientific knowledge. My thesis thus aims to move towards applying and developing computational tools that help investigate important questions on social networks.

The first component of my thesis focuses on understanding social interactions and networks, which offers implications for reshaping social networks to improve social cohesion. Specifically, I examine the formation and dynamics of social networks, with a focus on social exchange and "long ties". Utilizing large-scale social network data and computational tools, I first discuss benefits of the social exchange with dissimilar people in social networks; and then I proceed to study dynamic social networks and focus on long ties, or the social ties that bridge different communities in dynamic networks. Methodologically, I develop a novel interdisciplinary approach that combines game theory and machine learning techniques.

Second, I study what features on online platforms may improve social interactions and reshape social networks. To do so, I utilize large-scale data of online social media and provide two examples in the field. The first example is the identification of social contagion of online gift giving. This study examines how receiving a gift would promote the person to pay forward the gift, and also discusses how this social contagion can promote social interactions and tight social bonds. The other example is to examine how the designs of peer effects and prosociality on online social platforms encourage users' offline fitness behavior. Methodologically, both studies involve advanced causal inference and machine learning techniques to test the main hypotheses.

Moreover, I develop computational tools that analyze social network data. In the final component of my thesis, I introduce an algorithm for controlled experiments in social networks. This algorithm detects heterogeneous spillover effects – how the

treatment assignments received by one's network neighbors affect a person's behavior – in the data of networked experiments. This interdisciplinary algorithm combines approaches in causal inference, machine learning, and network science.

Thesis Supervisor: Alex 'Sandy' Pentland
Title: Toshiba Professor of Media Arts and Sciences, MIT

Thesis Committee Member: Dean Eckles
Title: Associate Professor of Marketing, MIT

Thesis Committee Member: Sinan Aral
Title: David Austin Professor of Management, MIT

Thesis Committee Member: Esteban Moro
Title: Associate Professor, Universidad Carlos III de Madrid and MIT

Acknowledgments

My thesis discusses social networks and prosocial behavior. So I deeply understand how the completion of my PhD would be difficult without the intellectual and emotional support of my mentors, collaborators, friends, and family.

First, I would like to thank my primary advisor Prof. Alex 'Sandy' Pentland. I am more than grateful that he enrolled me in the Human Dynamics Group five years ago, and then I started my PhD journey. He has taught me a lot about what makes it an important research question and always encourages me to make every effort when we find an interesting question. He gave his students plenty of research freedom. I have explored several different directions during my PhD, and I enjoyed this process. I believe that because of this, I am able to grow to be an independent researcher.

I also want to thank Prof. Dean Eckles, my co-supervisor who has provided me with a lot of advice in the last three years. He guided me to the interesting research area – causal inference and experimental design. He is very sharp in finding my blindpoints during our discussion. Everytime I interact with him, I learn something new. His pursuit in research rigor has deeply impacted me. I hope that in the future, I can polish my data analysis skills as he does.

Moreover, I would like to thank my committee members, Prof. Sinan Aral and Prof. Esteban Moro. Sinan's class on Applied Network Theory is one of the most beneficial classes for me at MIT. I am very grateful for his support throughout my PhD. As he has conducted several highly impactful studies on social networks, I am very fortunate to have his advice on my research.

Esteban is one of the people who provided me a lot of help and guidance at the beginning of my PhD. He offered me many great suggestions on topics in network science, especially in my first year. He is also the person at Human Dynamics Group who would always be happy to answer our technical questions.

In addition to my committee members, I would also love to thank my co-authors who helped to develop my PhD thesis – Ahmad Alabdulkareem, Kristen Altenburger, Tracy Liu, Ding Lyu, Farshad Kooti, Christos Nicolaides, Chenhao Tan, and Jie Tang.

Without them, the studies mentioned in this thesis would not be completed with the current quality. My critical thinking and writing skills have been polished together with the polishment of our papers. They are not just helping me with the papers, but also provided me with much guidance about academia and research.

Moreover, I am fortunate to have many good mentors. Although they have not co-authored any papers with me or our collaboration is still in progress, they helped me generously throughout my PhD. Among them, I hope to give special thanks to Xiaowen Dong, Moshe Hoffman, Shan Huang, and Erez Yoeli. They generously offer a lot of support and mentorship for my research and career development. I don't know if I can reciprocate, but I guess I will try to pay it forward.

I would love to thank my Human Dynamics family. Scholars and peers at Human Dynamics have conducted a variety of interesting and important research and have provided me with a lot of emotional support. I would like to thank visiting professors YY Ahn and Desheng Zhang for their great advice on my research. Very importantly, my peer PhD students are the ones who I would always share my joys and concerns with. Thank you – Yan Leng, Eaman Jahani, Michiel Bakker, Tara Sowrirajan, Abdullah Almaatouq, Abdulrahman Alotaibi, Dhaval Adjodah, Dan Calacci, Zivvy Epstein, Abhi Dubey, Morgan Frank, P. M. Krafft, Oren Lederman, Isabella Loaiza-Saa, Alejandro Noriega Campero, and every other person I met at HD! I also want to thank Stephen Buckey and Teri Hagen for their excellent administrative support.

I would love to thank my IDSS family. I greatly appreciate the efforts made by the IDSS leadership, especially the efforts made by Munther Dahleh, Ali Jadbabaie, Alberto Abadie, John Tsitsiklis, and Fotini Christia. I was fortunate to get into the first cohort of the SES program, where, I was exposed to people from diverse backgrounds, and to a variety of hardcore courses that will benefit me in the long term. I want to give special thanks to our academic administrators, especially Beth Milnes, who warmly provided me and fellow students with numerous help and support. I would like to thank Kim Strampel and Laura Dorson. Also importantly, I have also made great friends at IDSS. I particularly enjoyed my days dining out and playing board games with Chin-Chia Hsu, Yan Jin, Hanwei Li, Minghao Qiu, Yi Sun, Qi Yang,

and Jinglong Zhao these years. Other members of IDSS, such as Kiran Garimella, Amin Rahimian, Ian Schneider, Rui Sun, Amir Tohidi Kalorazi, Manxi Wu, and Yunzong Xu, have also provided me with much support.

I am also fortunate to know peers and friends from Sloan. I would like to thank Christie Ko for her support. I have learned a lot from members at IDE. Here I hope to give shout-outs to Mohsen Mosleh and Jeremy Yang.

I would also love to thank my old and new friends. I enjoy playing board games with my friends, especially during the COVID-19 pandemic, your company makes the days more fun. MIT and the great Boston area are excellent places to make new friends. Because of conferences and other academic events, I am also able to make friends with many scholars. I would love to thank you too for sharing academic and life experiences. Among them, special thanks are sent to Shengjia Zhao!

Finally – love you, Mom and Dad!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Individuals are interdependent. Our social identities depend on who we interact with [237]. Our opinions, actions, emotions can be deeply influenced by our peers [69, 67, 96, 97, 47, 56, 21, 19]. We are inclined to make friends with those who share similar characteristics [177]. The social groups that we belong to provide us with support, information of interest, among other benefits, or pitfalls [28].

Individuals and their interdependence can be modeled by social networks [110, 53]. A social network consists of a set of individuals, who serve as nodes on the graph, and the interactions or relationships between those individuals, who serve as edges [34, 135]. Social networks provide researchers with a powerful, bottom-top angle to view and analyze our societies. On the micro level, we can investigate how the local network structure of a node (or an individual) predicts individual decisions or influences their behaviors [163, 82, 247]. On the macro level, social networks provide a bottom-up approach to examine the global structure of a society and the diffusion of information or behaviors on this network [81, 198].

Social networks have been studied by multiple disciplines, ranging from physics [258, 35], statistics [98, 226], computer science [163, 197], to economics [135], and other social sciences [69, 73]. From my perspective, two main reasons have contributed to the surge in social network research. The first reason is the availability of large-scale

data [158]. Due to the proliferation of mobile and digital technologies, various social network datasets are becoming available to researchers. Examples include mobile communication networks, email networks, instant messaging networks, and social media networks. Importantly, in recent years, some social network platforms have made it possible for users to have various types of virtual interactions. The data of those online behaviors offer opportunities to investigate important human behaviors with large-scale datasets that have not been well studied due to the data availability issue.

Second, the innovation of analytical and computational tools also promote the proliferation of social network studies. The studies of complex networks provide powerful ways to explain important phenomena in the real-world complex network dynamics by simple and tractable analytical models, such as the small-world phenomenon and the power law degree distribution [258, 35]. In addition, microeconomic theory and econometrics have been employed to model the individual rational decisions of choosing neighbors and actions on the network [135]. Furthermore, the recent rapid development of machine learning techniques on graphs, such as graphical models [235], embedding techniques [197, 112], and graph neural networks[144, 253], have been utilized to model the complex patterns on large-scale real-world networks. All these approaches have facilitated the analysis of massive social networks.

However, challenges remain in the studies of social networks from both question driven and methods driven perspectives. On the question-driven side, first, while there has been extensive literature on the naturally occurring patterns on social networks, such as homophily (the phenomenon whereby people tend to interact with others [177]), triad closure or local clustering [101], and group formation or community detection [28], much fewer studies focus on the interactions between individuals and communities who are distant on social networks. Those interactions between socially distant or dissimilar individuals are equally or sometimes even more important, as they may serve as the key channels to spread novel information and connect different communities [110]. Moreover, they may help us understand how to help combat societal problems such as political polarization [71], racism and xenophobia [13], and

misinformation [76, 254].

Second, as online social media designs various novel features that facilitate people's daily life, as researchers, we need to work with those online platforms to understand how to promote positive social interactions and reduce the negative ones. While social platforms are providing us with much convenience for remote interactions, they may unintentionally promote negative interactions such as the formation of echo chambers, hate, and polarization. It would be interesting to illustrate how social network platforms may design features that promote the social good. In my thesis, two examples on online social platforms are used to offer implications for improving the ecosystem of online social platforms, or more specifically for promoting prosocial behavior.

From a methodological perspective, social scientists are faced with challenges when directly applying machine learning for social science research, especially when social scientists aim to effectively understanding causal mechanisms. Since the goal of most machine learning algorithms is to maximize the prediction performance, direct application of machine learning techniques may not help us fully understand the mechanisms of the dynamics of our societies. This "explanation-prediction" trade-off is a challenge in computational social science studies [120]. For social science research, it is even more challenging to understand how to draw useful knowledge by applying the "black-box-like" algorithms. Moreover, machine learning algorithms face challenges in deriving causal relationships from big data, which is the core problem in many social science studies [196]. In particular, in social networks, correlation-based data analysis may sometimes offer misleading conclusions without considering causality [70, 18].

1.2 Problem Summary

My thesis contributes to the social network literature in terms of both drawing social scientific knowledge and improving computational approaches. From Chapter 2 to Chapter 6, I discuss my explorations in further understanding social networks and improving computational methods in social networks.

Chapter 2 introduces a framework for modeling social network formation. It discusses an approach that jointly accounts for homophily and social exchange, which has been studied mostly separately in the prior literature. Methodologically, it proposes a novel approach to incorporate embedding techniques [197, 112, 253] in strategic network formation models [135]. This chapter is based on the paper "*An interpretable Approach for Social Network Formation Among Heterogeneous Agents*," which is joint work with Ahmad Alabdulkareem and Alex 'Sandy' Pentland, and has been published at Nature Communications [268].

Chapter 3 further discusses the dynamics of social networks, with a focus on long ties (or long-range ties), which are the social ties that bridge different communities in social networks [57, 56, 194]. The results of this and last chapters imply the necessity of having social ties that bring beneficial social exchange to stabilize social ties and maintain the structure of our social networks. Methodologically, it is an extension of the approach in Chapter 2 to temporal social networks by combining strategic network formation and network embedding. This chapter is based on an ongoing project titled "*Investigating and Modeling the Dynamics of Long Ties*," which is joint work with Ding Lyu, Lin Wang, Xiaofan Wang, and Alex 'Sandy' Pentland.

After the two chapters that focus on modeling social network structures, Chapter 4 and Chapter 5.1 provide two examples of how online social platforms can improve well-being and promote positive social interactions in online social platforms. Chapter 4 discusses the social contagion of gift giving (precisely red packets [271]) on a major social network platform in China, which further contributes to the literature of social contagion. The results of this chapter also highlights how online gift giving can promote social interactions and solidarity in online social networks. From a methodological perspective, it applies a structural causal model [196] to a large-scale observational dataset as a neat natural experiment. This is joint work with Tracy Xiao Liu, Chenhao Tan, Qian Chen, Alex 'Sandy' Pentland, and Jie Tang, and is based on a paper under review and revision titled "*Gift Contagion in Online Groups*" [270].

Chapter 5.1 provides a comparison between the effect of social contagion (or peer

pressure) versus prosocial incentives (i.e., the incentives from strangers). This chapter focuses on their effect on people's daily exercise patterns. This chapter does not only quantify the peer effect (social contagion) of exercises through online social network platforms, but also compares it with the effectiveness of prosocial incentives by the same platform. This chapter involves multiple modern causal inference approaches, including a large scale field experiment, instrumental variable design, matching design, and a regression discontinuity design, to estimate the effects and investigate the mechanisms. This is joint work with Christos Nicolaides, Alex 'Sandy' Pentland, and Dean Eckles, is based on the working paper "*Promoting Physical Activity through Prosocial Incentives on Mobile Platforms*."

Chapter 6 is a methodological paper, which proposes an approach that combines machine learning and network science to help mitigate the network interference [248, 22] effect when analyzing large-scale experimental data in networks. It is also connected to theories in social networks such as structural diversity [247] and complex contagion [57]. This is joint work with Kristen M. Altenburger and Farshad Kooti, and is based on the paper "*Causal Network Motifs: Identifying Heterogeneous Spillover Effects in A/B Tests*" appeared in the ACM Web Conference 2021 [269].

Finally, Chapter 7, the Conclusions Chapter, summarizes my thesis and proposes future directions that I would like to explore in network science and computational social science [158].

Chapter 2

Modeling social exchange and homophily with embedding techniques

2.1 Background

Previous work on modeling social network formation has typically employed game theory or agent-based modeling [258, 33, 136, 223, 189, 185, 135]. These studies typically propose simple and tractable micro-level rules for link formation mechanisms and show that these rules have implications for known macro-level properties. Several studies in statistics and econometrics have also used game theory to model empirical networks [178, 68, 59], but they typically have been focused on estimating and identifying the effects of special interest such as racial segregation. To date, these models have not been capable of accounting for the effects of broad heterogeneity among individuals and, therefore, a lack predictive power for link formation in complex, real-world networks.

Studies on network embedding techniques [197, 238, 112, 144] could partially fill this gap in the network formation literature because these techniques consider node heterogeneity and show predictability of both link formation and individual charac-

teristics. Network embedding techniques are aimed at representing each node with a fixed-length vector learned from social network observation data. The agents in a network may be so diverse that representing all their characteristics would require very high dimensionality for the vectors. The philosophy of network embedding is aimed at reducing the dimensionality by mapping all the characteristics of agents onto a low-dimensional latent space. Each dimension in the latent space, therefore, usually does not correspond to a separate attribute of the agents. The latent space representation of the network provides considerable potential for measuring the heterogeneity among agents. However, because network embedding methods are designed for data representation and compression rather than for explaining network formation, they do not attempt to capture micro, inter-agent effects such as social status or macro effects such as social segregation and thus do not provide social science explanations for the link formation.

There are few previous papers which have attempted to account for network formation by the heterogeneity of agent without losing micro-level interpretability. A study on ecological networks by McKane and Drossel utilized a similar approach with ours, wherein agents were represented by a small number of attributes among a large attribute pool [175]. However, this work did not directly estimate the latent variables for networks of agents. More abstractly, our method is also reminiscent of mixed membership stochastic blockmodels where agents respectively follow a probability distribution of membership within several communities [8]. However, probabilistic membership models typically do not seek to explain the economic and sociological mechanism and dynamics of network formation. We hope to extend these previous works to the estimation of agent characteristics and network link formation using observed interaction data. In addition, we want to incorporate a more complex but interpretable inter-agent exchange utility function, by modeling both exchange benefits and coordination costs arising from the differences between agents.

Furthermore, an important question rarely studied in literature is the trade-off between coordination costs and exchange benefits. On the one hand, the coordination between dissimilar people incurs higher costs than between similar people [137],

a relationship which encourages homophily, i.e., the tendency to interact more with people who have shared characteristics. On the other hand, the rationale of exchanging endowments comes from welfare economics: agents have different endowments and their preferences drive them to interact and exchange endowments [171]. The exchange nature therefore encourages heterophily, i.e., the tendency to interact with dissimilar individuals [205]. Empirical studies have found that heterophily exists in various scenarios [138, 141], and that complimentary heterophily between two agents may bring more mutual benefits than homophily [12]. However, most prior studies of social network formation consider either only coordination costs and homophily [68, 44, 75] or only social exchange and heterophily [74, 100, 147], rather than the integrated consideration of exchange and coordination as we do in this study. The trade-off between exchange benefits and coordination costs is also reminiscent of the identity-diversity balance in organizational performance literature [257, 125].

Inspired by the network embedding techniques, we develop a social network formation model and representation learning method for heterogeneous agents that seek to retain the interpretability. We maintain the inter-agent micro-structure characteristic of most agent-based models and have the macro-level structures that are the focus of sociology. In our model, agents are characterized by vectors, called their endowment vectors; they maximize their utility by having link formation driven by comparing their own endowment vectors with those of others. Importantly, we take an economic view of human networks which considers link formation to be driven by the trade-off between the benefit of exchanges [191] among individuals with different endowments against the coordination costs due to differences in individual endowments [177]. We apply optimization methods to ascertain the endowment vectors of all agents using empirical observation of network interactions. The effectiveness of this method is validated by predicting link formation and individual characteristics. Subsequently, the agent-based models that are derived from empirical data are evaluated in terms of their micro- and macro-level behavior, compared with the behavior of human networks. Abstractly, we model link formation as a reaction-diffusion system, a framework found in many biological systems.

2.2 Methods

Problem setup

Let $\mathcal{I} = \{1, 2, ..., N\}$ be a group of N and potentially connected agents indexed by i (or j, l). Let K be the dimensionality of endowments that drives the formation of the social network of the group, indexed by k. Each agent has a latent endowment vector $\mathbf{w_i} = (w_{i1}, ..., w_{iK})^T$, with each dimension indicating an aspect of the individual's attributes. Let $\mathbf{W} = (\mathbf{w}_1, ..., \mathbf{w}_N)^T$. We observe all edges among the N agents. Let D be a set of $N \times N$ adjacency matrices among agents in all periods. D_{ij} is binary ($\{0, 1\}$). $D_{ij} = 1$ if there is an edge from i to j, and $D_{ij} = 0$ otherwise. For the convenience of showing pairwise stability, the study is restricted to undirected graphs, i.e. $D_{ij} = D_{ji}$.

Agents make rational choices by comparing their endowment vectors with potential friends. Agents maximize their utility functions ($U_i : 2^{\mathcal{I}/\{i\}} \to \mathbb{R}$ for each i) dependent on the differences between their endowment vectors and all possible candidates (all other agents). U_i is also parameterized by $\mathbf{W}$, $\mathbf{b}$, and $\mathbf{c}$. $\Delta u_i(j)$ is the marginal utility that j brings to i. We therefore predict D_{ij} by $\Delta u_i(j)$.

Data description

- Andorra. We collected the nationwide call detail records in Andorra from July 2015 to June 2016. Utilizing the country code, we filtered out all non-citizens, leaving 32,829 citizens with at least one call interactions with another. If the (i, j) had at least one effective call (duration greater than zero seconds), we set $D_{ij} = D_{ji} = 1$; otherwise $D_{ij} = D_{ji} = 0$. This process results in $513,931$ links. To demonstrate the effectiveness of the learned endowments, we also extracted three characteristics of individuals: phone type, frequent city, and Internet usage. The phone type was identified by the type allocation code, and we classified each type into Apple, Samsung and others (the distribution of three types is balanced). For each phone number, we employed the last phone type that we observed. Note that type phone is strongly correlated with important

individual characteristics such as income. The most frequent city was identified by the cell tower id. We classified each phone number by the location that it shows up most frequently throughout the year, this location is thus likely the work location of the individual (some individuals' work location may be their home). Internet usage was computed by the total duration of cellular data. In the prediction task, we classified Internet usage into high (more than median) and low (less or equal than median). Details of the datasets, such as statistics of individual characteristics and network degree distribution, are shown in Description in detail is reported in Appendix.

- Movie. To highlight the exchange effects, we examine a specific type of social network, director-cast movie collaboration network, where a node represents either a movie director or an actor/actress, and an edge between a director i and an actor/actress j represents a collaboration between i and j. $D_{ij} = D_{ji} = 1$ means that i and j collaborated at least once; 0 otherwise. Note that the social network is close to a bipartite graph where nodes are partitioned into directors and cast (some people have both cast and director experience). We extracted 3,493 movies throughout 2000 to 2016, and retained individuals with at least five movies within this period, resulting in 160 directors and 2,628 cast members, and 10,399 director-cast pairs. To validate the effectiveness of the learned endowments, we extracted two individual characteristics: occupation and gender. For occupation, we labeled an individual as a director if she functioned as a director in more than a half of the movies in which she engaged; cast otherwise. For gender, we collected 1,840 males and 761 females and 186 unlabeled.

- Synthetic. We manually establish a network of 2,500 agents. Agents are indexed by (x, y) $(i = 50x + y)$, $0 \leq x \leq 49$, $0 \leq y \leq 49$, $x, y \in \mathbb{N}$. Each agent therefore resides at a unique location on the 50×50 grid, and the agent has a probability of 0.5 to be either type A (e.g., a buyer) or type B (e.g., a seller). Buyers (sellers) are exploring sellers (buyers) in their neighborhood with Manhattan distance ≤ 3. The network is therefore a bipartite graph where buyers and

sellers exchange goods and money. This data generating process results in 14,453 edges. We predict the type and location (divide the plane into four parts) for all agents.

- Company. A network of employees in a company where edges represent a call and text communication (MobileD in [239]). Each employee is labeled as a manager or a subordinate. In total, we have 420 managers and 1,564 subordinates, with 12,751 edges among them. In this network, managers are mostly connected with managers and subordinates are mostly connected with subordinates. At the same time, subordinates also interact with their respective managers occasionally. We believe that this dataset should show a trade-off between coordination and exchange; for example, managers and subordinates have exchange effects, and they have lower coordination costs to interact with the same type.

- Trade. We use the 2014 international trade data provided by the United Nations Statistical Division (UN Comtrade Database: [`https://comtrade.un.org/`]), specifically the cleaned version provided by the BACI team using their own methodology of harmonization [104]. We created a network of countries, where an edge indicates that the trade value between two countries is greater than one billion dollars (for both directions). This process resulted in 100 countries with at least one link, and 703 undirected edges among them. We predict the GDP, economic complexity index (ECI) [116], and the countries' continents for this dataset.

Details in learning

For computational simplicity and better fitting performance (see Appendix), we split the dimensions into "beneficial dimensions" and "costly dimensions". In equation (2.5), every dimension (say the k-th) can contribute to both benefits and costs if both b_k and c_k are greater than zero. However, it is not difficult to see that if we constrain some dimensions to have zero-valued beneficial scaling parameters ($b_k = 0$) or costly

scaling parameters ($c_k = 0$), the dimensionality of the model (K) will increase but the capacity of data fitting will not change. During the learning process, a connected pair (i, j) may result in either an increase in the difference on some beneficial dimension (with $b_k > 0$) or a decrease in the difference on some costly dimension (with $c_k > 0$) between their endowment vectors. Empirically, if both b_k and c_k are positive, these two conflicting effects (to increase or to decrease the utility on the same dimension) would hinder an effective convergence (shown in Appendix); we conjecture that this is because we are optimizing a non-linear non-convex loss function. Therefore, we separate the K dimension into K_{bnf} "beneficial dimensions" and K_{cst} "costly dimensions" ($K_{\text{bnf}} + K_{\text{cst}} = K$). By comparing the performances of link fitting for different K_{bnf} and K_{cst}, we select the optimal K^*_{bnf} and K^*_{cst}, and consequently K^*. For simplicity, we let $b_k = 0$, for $k > K_{\text{bnf}}$; and $c_k = 0$, for $k \leq K_{\text{bnf}}$, and $\boldsymbol{\theta} = (b_1, b_2, ..., b_{K_{\text{bnf}}}, c_{K_{\text{bnf}}+1}, c_{K_{\text{bnf}}+2}, ..., c_K)$. In Appendix, we show empirically that the performances of link fitting and node classifications are worse when we do not split dimensions into beneficial and costly dimensions; and that even when we do not split dimensions, the learning algorithm will lead most dimensions to be either "beneficial" or "costly", i.e. either b_k or c_k is very close to zero. More details, including how to select λ_{fp} and λ_{reg} for different datasets, can be found in Appendix.

2.3 Results

2.3.1 A game theoretical model

Endowment is a well-known and useful concept in microeconomic theory [171], for example, fundamental theorems of welfare economics are based on endowment vectors. In our model, an endowment vector represents all of the features (assets, abilities, capacities, qualities, etc.) that each agent possesses, and are treated as fixed, invariant characteristics of the agent. We do not consider the situation where endowments are dynamic in this study. Note that since we limit the dimensionality of endowment vectors, similar to network embedding algorithms (see Methods), each dimension

does not necessarily have a specific meaning, but may be a combination of many attributes of an individual.

Agents establish social ties according to the comparison between their endowments. If we assume that there are K dimensions of endowments in a society, each agent has a K-dimensional endowment vector $\mathbf{w}$. Note that dimensions may be mutually correlated; for example, in the Karate club, leaders and followers have high values in their respective dimensions, and these two dimensions should be negatively correlated. We constrain the first and second moments of each dimension ($\|\mathbf{W}_{:k}\|$) to be zero and one respectively for computational simplicity.

We assume the utility function of agent i is only determined by agent i's neighbors' endowment vectors. We define the utility function $U_i : 2^{\mathcal{I}/\{i\}} \to \mathbb{R}$ for all i, as equation (2.1). The argument S is the potential neighbors, denoting an arbitrary subset of all agents except i herself, i.e., $\mathcal{I}/\{i\}$. Each agent i selects her neighbor set S by maximizing her utility function U_i. U_i is composed of two terms, the benefits of exchange (F_i) and the costs of coordination (G_i):

$$U_i(S; \mathbf{W}, \mathbf{b}, \mathbf{c}) = \underbrace{F_i(S; \mathbf{W}, \mathbf{b})}_{\text{benefits of exchange}} - \underbrace{G_i(S; \mathbf{W}, \mathbf{c})}_{\text{costs of coordination}}, \qquad \forall S \subset \mathcal{I}/\{i\}. \tag{2.1}$$

When S_i^* is the optimal neighbor set for i, we define the marginal utility that j brings to i as:

$$\Delta u_i(j) = \begin{cases} U_i(S_i^*; \mathbf{W}, \mathbf{b}) - U_i(S_i^*/\{j\}; \mathbf{W}, \mathbf{b}), & \text{if } j \in S_i^*; \\ U_i(S_i^* \cup \{j\}; \mathbf{W}, \mathbf{b}) - U_i(S_i^*; \mathbf{W}, \mathbf{b}), & \text{if } j \notin S_i^*. \end{cases} \tag{2.2}$$

In this study, we are focused on a specific form for F_i and G_i and, consequently, for U_i. For the costs of coordination, agent i's cost incurred by agent j is measured by the difference between $\mathbf{w_j}$ and $\mathbf{w_i}$.

$$G_i(S; \mathbf{W}, \mathbf{c}) = \sum_{i \in S} g(\mathbf{w_j}, \mathbf{w_i}, \mathbf{c}) = \sum_{i \in S} \|\mathbf{c} \circ (\mathbf{w_j} - \mathbf{w_i})\|_2. \tag{2.3}$$

"$\circ$" denotes element-wise multiplication. $\|x\|_2$ denotes ℓ_2 norm. Note that the costs are symmetric, i.e. $\|\mathbf{c} \circ (\mathbf{w_i} - \mathbf{w_j})\|_2 = \|\mathbf{c} \circ (\mathbf{w_j} - \mathbf{w_i})\|_2$. The *costly scaling parameter*, c_k, measures the importance of k-th dimensions on the costs. A higher c_k will amplify the difference between i and j's endowment vectors on the k-th dimension ($w_{jk} - w_{ik}$). This term encourages homophily: dissimilar pairs have to suffer from high coordination costs before forming a link.

For F_i, we propose the following form:

$$F_i(S_i^*; \mathbf{W}, \mathbf{b}) = \sum_{j \in S_i^*} \sum_{k=1}^{K} b_k \max(w_{jk} - w_{ik}, 0) \tag{2.4}$$

Intuitively, $w_{jk} - w_{ik}$ measures the "advantage" of agent j on the k-th dimension over agent i. Since we do not want negative benefits, we consider the benefit on the k-th dimension is zero if $w_{jk} - w_{ik} < 0$. In deep learning, $\max(x, 0)$ is called the "ReLU" function. TensorFlow [1], a machine learning programming library, provides methods to optimize functions that contain ReLU functions. Similar to c_k, the beneficial scaling parameter b_k measures how beneficial the k-th dimension is. This term indicates that if a person is high in several dimensions, she could bring more benefits to others. Therefore, other people are inclined to link to her. However, she does not necessarily reciprocate every link because, for example, when she is high in every dimension, she will not benefit from others in any dimension. Note that we do not consider comparative advantages in this study. In addition, this term encourages heterophily: people whose expertise are complimentary have high potential benefits for link formation. Therefore, in this specific form, we have

$$\Delta u_i(j) = \sum_{k=1}^{K} b_k \max(w_{jk} - w_{ik}, 0) - \|\mathbf{c} \circ (\mathbf{w_j} - \mathbf{w_i})\|_2 \tag{2.5}$$

There are of course many other variations for the functional form (equation (2.1)). For example, we can let F_i non-separable in terms of the neighbor set S, e.g., $F_i(S) = \frac{1}{|S|} \sum_{j \in S_i^*} \sum_{k=1}^{K} b_k \max(w_{jk} - w_{ik}, 0)$. The intuition is that when one agent has many neighbors, the benefit brought by each neighbor decreases; Do et al. provide

a good example of a decreasing marginal utility [77]. However, this functional form indicates that $\Delta u_i(j)$ depends on the neighbor set S, which leads to a time-consuming combinatorial optimization in the learning process; specifically, when the learning algorithm chooses S_i^*, it may need $\mathcal{O}(N2^N)$ computations of the utility functions, which is computationally infeasible for even a small-scale network. This is beyond the scope of this study. We can also change G_i into other norms, such as ℓ_1 norm, or change F_i into a smoother version of $\max(x, 0)$, but these changes do not significantly affect the results in the later sections, as shown in Appendix. Therefore we concentrate on this specific form in later sections (equation (2.5)).

In network game theory, pairwise stability [135] refers to the situation where no increased marginal utility can be brought to both of any unconnected pairs, and no increased marginal utility can be brought to any agents who want to drop their neighbors. Following the definition, we derive the conditions when pairwise stability in undirected networks is satisfied. The proof is straightforward and can be found in Appendix.

Proposition 1 *An undirected network ($\mathcal{G} = (\mathcal{V}, \mathcal{E})$) implied by neighbor sets S_i^*, $i = 1, 2, ..., N$ is pairwise stable, if the following conditions are satisfied:*

1. *if $j \in S_i^*$, then $i \in S_j^*$;*
2. *$\forall j \in S_i^*$, $\Delta u_i(j) \geq 0$;*
3. *$\forall j \notin S_i^*$, $\min\left(\Delta u_i(j), \Delta u_j(i)\right) < 0$.*

2.3.2 Learning endowments

We have established a model for social network formation with many parameters and latent variables. Before we examine the proprieties of the model, we have to assign values for the unknown variables, including the endowment vectors ($\mathbf{W}$), and scaling parameters ($\mathbf{b}$ and $\mathbf{c}$). To equip our model with the capability of fitting real-world networks, we learn the endowment vectors using the observations of real-world networks, by assuming real-world networks are at or close to pairwise stability.

Let $\mathcal{L}(\mathbf{b}, \mathbf{c}, \mathbf{W}|D)$ be the loss function that we want to minimize. The definition of $\mathcal{L}(\mathbf{b}, \mathbf{c}, \mathbf{W}|D)$ is reported in Appendix. Then we solve the optimization problem in equation (2.6).

$$
\begin{aligned}
\text{Minimize}_{\ \mathbf{b},\mathbf{c},\mathbf{W}} :& \mathcal{L}(\mathbf{b}, \mathbf{c}, \mathbf{W}|D) \\
\text{Subject to: } & b_k \geq 0,\ \forall k = 1, 2, ...K \\
& c_k \geq 0,\ \forall k = 1, 2, ...K \\
& \frac{\sum_{i=1}^{N} w_{ik}}{N} = 0,\ \forall k = 1, 2, ...K \\
& ||\mathbf{W_{\cdot k}}||_2^2 = N,\ \forall k = 1, 2, ...K
\end{aligned}
\tag{2.6}
$$

The constraints that b_k and c_k should not be less than 0 are required by the properties of our model. The constraint for the mean of each dimension is to limit the number of equivalent solutions, so that the optimizer could typically find a better solution. The constraint of $\mathbf{W_{\cdot k}}$ is to guarantee that the standard deviation of each dimension is approximately 1, so that the values of $\mathbf{b}$ and $\mathbf{c}$ are comparable across dimensions.

Since $\mathcal{L}(\mathbf{b}, \mathbf{c}, \mathbf{W}|D)$ is non-linear and non-convex (dimensions are interchangeable) with respect to $(\mathbf{b}, \mathbf{c}, \mathbf{W})$, we have to approximate the global optimum by a local optimum. By employing Adam optimizer (an improved stochastic gradient descent method) [143], we are able to learn the local optimum of $\mathcal{L}(\mathbf{b}, \mathbf{c}, \mathbf{W}|D)$; Adam optimizer is good at deriving good local optima when solving non-linear and non-convex problems. To obtain a solution that approximates the global optimum, we start from many randomly selected initial points and then analyze the results of the multiple runs to find the parameters that generate the smallest loss and therefore the best link fitting performance. Technical details, including the definition of $\mathcal{L}$ and methods that assist learning, are presented in Appendix.

2.3.3 Validation of learning

Here we show that we have learned meaningful endowment vectors from empirical networks. In particular, we first use a toy example–Zachary's karate club network [272] to illustrate the learning results. We then validate the effectiveness of our model and learning method by showing their performance at fitting link formation and predicting individual characteristics for a variety of large-scale social networks: a synthetic network where two types of people exchange, a Trade network among countries, a movie collaboration network, a Company communication network, and the Andorra network, which is a nationwide mobile phone network (see Description in Methods).

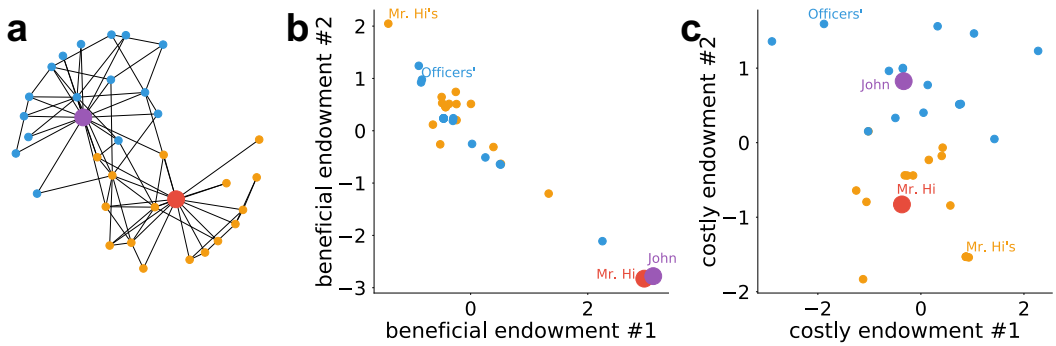


Figure 2-1: Illustration of the learned endowments for karate club network. **a** The network structure of karate club network. Mr. Hi, the instructor, is marked red, and other people in his faction are marked orange. John, the student leader or officer, is marked purple, and other people in his faction are marked blue. **b** The first two dimensions of the learned endowment vector for each individual; these two dimensions are only related to exchange benefits so we call them "beneficial endowments". **c** The last two dimensions of the learned endowment vector for each individual; these two dimensions are only related to coordination cost so we call them "costly endowments".

We start with a toy example to illustrate both the rationale of the present model and the effectiveness of learning performance. Because of a conflict between an instructor (Mr. Hi) and a student officier (John), the social network of Zachary's karate club is polarized into two factions (see Fig. 2-1, Panel a). We set $K = 4$ and the first two dimensions as "beneficial endowments" and the last two dimensions "costly endowments" (see Methods) because it is more convenient for visualization if the numbers of

beneficial and costly dimensions are both even. Note that $K = 4$ is not necessarily the optimal dimensionality and here we did not add regularization term (see Appendix) for this result; however we also show in Appendix that $K = 4$ is a reasonable (almost optimal) selection.

Panels b and c in Fig. 2-1 plot the values of the learned endowments of individuals in Zachary's karate club. In Panel b, both Mr. Hi and John are high in dimension #1 and low in dimension #2, while the rest are generally low in endowment dimension #1 and high in dimension #2. We interpret this result as the tendency of exchanges between instructors and students: dimension #1 represents the professional skill of karate and leadership in their factions; endowment #2 represents the willingness to learn Karate. As for costly endowments (Panel c), we find that dimension #4 corresponds to the faction to which each individual belongs: Mr. Hi and his followers (orange) have values generally higher than 0 while John and his followers (blue) are generally lower than 0. Dimension #4 can be explained as the individual's identification with the two factions. We interpret cost endowment #3 as other unobserved characteristics that might influence the interactions between individuals, such as the time and frequency to participate in club activities. We also illustrate the learning results for the Trade and Synthetic datasets graphically in Appendix.

Because our goal is to use the learned endowment vectors to further analyze the micro- and macro- patterns of the network, we learn the endowment vectors by using all the information (the links) of the network. Therefore, rather than split the input links into training and test sets, we use all the links as the input. A potential concern is that we might "overfit" the network by using a large K; we partially address this concern by introducing the regularization term $\mathcal{L}_{\text{reg}}$ as mentioned in Appendix. We use $\Delta u_i(j)$ as the predictor and AUC (area under the curve) as a measurement for the fitting performance. AUC trades off between true positive and false positive rates, and serves as a fair measure when there is a strong imbalance between positive and negative samples. By using an approach provided in Appendix, we obtain the optimal dimensionality (K) and the optimal number of beneficial and costly endowments (K_{bnf} and K_{cst}, see their definitions in Methods).

Dataset	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{E}\rvert$	K^*	K^*_{bnf}	K^*_{cst}	AUC
Karate	34	78	3	2	1	98.48%
Country trade	100	703	5	3	2	96.85%
Synthetic	2,500	14,453	4	2	2	99.92%
Company	1,984	12,751	11	4	7	98.70%
Movie	2,788	10,399	7	4	3	96.08%
Andorra	32,829	513,931	15	8	7	94.76%

Table 2.1: Learning results and link fitting performance of learned endowment vectors. $|\mathcal{V}|$ denotes the number of nodes and $|\mathcal{E}|$ denotes the number of edges. K^*, K^*_{bnf} and K^*_{cst} represent the optimal number of dimensions, beneficial dimensions and costly dimensions, respectively.

As shown in Table 2.3.3, our model is able to obtain very good fits to the input networks. For all datasets, the AUC of link fitting is over 94%. Moreover, we demonstrate that for all datasets, it is necessary to incorporate both the benefit and the cost terms into the utility functions (i.e. $K_{\text{bnf}} > 0$ and $K_{\text{cst}} > 0$). This finding highlights the importance of integrating both exchange effects and coordination costs into the link formation mechanisms. Other technical details, including learning curves and the performance on all the dimensions, are presented in Appendix.

Although our goal is not to design a network embedding algorithm that outperforms the state-of-the-art algorithms, it is interesting to examine the model's ability to predict individual characteristics as a network embedding algorithm. If the learned endowments have a decent predictive power for individual characteristics, we can then believe that we have effectively learned the endowment vectors, which can be used for further analysis such as agent-based modeling. We extract characteristics that are not directly relevant to nodes' ego network attributes (see Appendix for a full list). We split the nodes and their learned endowment vectors into training (75%) and test (25%) sets. We use support vector machine (SVM) and k-nearest neighbors algorithm (k-NN) to train the classifiers, and use cross-validation to tune the classifiers' hyperparameters.

As shown in Fig. 2-2, the learned endowment vectors can well predict most individual characteristics by SVM. Note that k-NN has similar results Appendix. This result shows that our model can encapsulate the latent features of the agents. It is

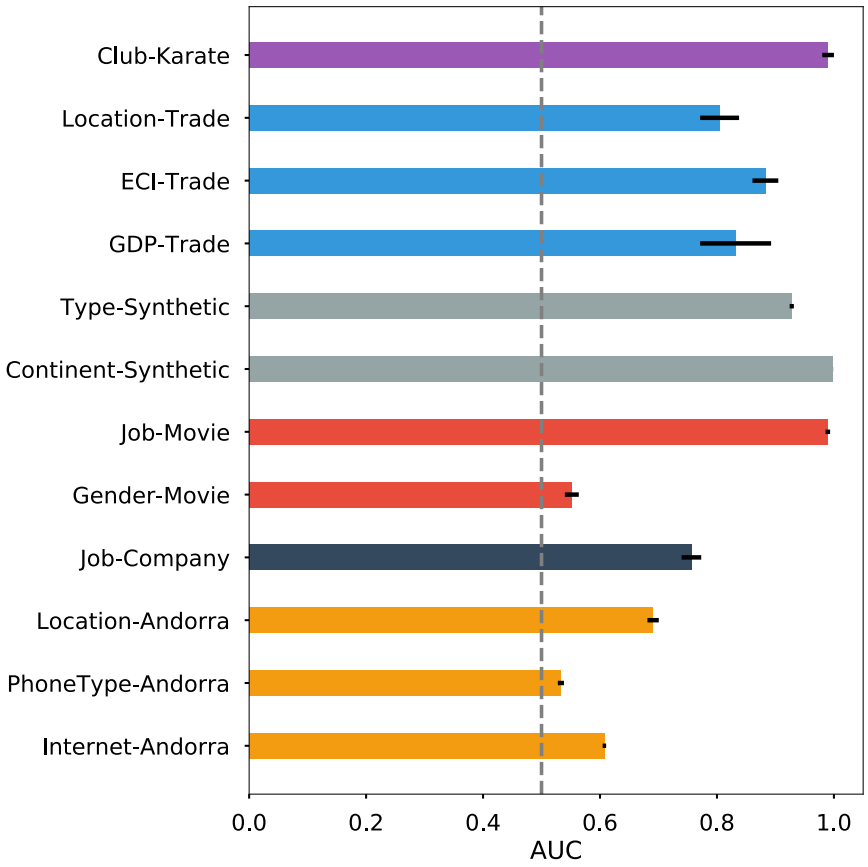


Figure 2-2: Prediction performance of the learned endowments. We use support vector machine as the classifier. The baseline, random guess algorithm, is indicated by the dashed line. Error bars represent the standard errors for the average AUCs in five random splitting of training-test sets.

important to highlight that individual characteristics might not be fully reflected in the network; therefore, neither network embedding algorithms nor the present model can guarantee high AUCs for all prediction tasks. However, the learned endowment vectors in fact contain more information than the presented agent features; therefore, they could predict very different agent characteristics, e.g., preferences of movie genres.

The accuracy at estimating agent characteristics beyond the input data could be because they are important either in coordination costs (e.g. locations) or exchange benefits (e.g. collaboration between cast members and directors). Some characteristics may have both exchange effects and coordination costs: for example, in a

company, subordinates mostly communicate with each other (low coordination costs), but would also interact with their managers occasionally (exchange benefits).

We also compare our results with a network embedding algorithm, DeepWalk [197], with the same number of dimensions and therefore the same degree of freedom (Appendix). Recall that network embedding methods are designed only for dimension reduction; they therefore do not provide economic or sociological insights about the network. Algorithmically, DeepWalk uses an energy function that considers only similarity and not the benefit that can flow from exchanges between agents with very different endowments. Consequently, as might be expected, when our model is compared to DeepWalk, we have better performance if the predicted characteristics are explicitly implied by exchange effects. However, for characteristics explicitly implied by low coordination costs between similar people, the performance of the present model is somewhat lower than that of DeepWalk, probably because DeepWalk considers the similarity between neighbors spanning multiple hops. In sum, the ability to predict agent characteristics shows that our model has learned useful information implicit in the network, and that this implicit information can be used for further agent-based modeling.

2.3.4 Agent-based modeling

We next analyze the properties of the model as an agent-based model. Because of the high degree of freedom of the present model, any manually input distributions of $\mathbf{W}$, $\mathbf{b}$ and $\mathbf{c}$ may appear too arbitrary and do not reflect any real-world situation. We therefore use the learned endowments and parameters as the input to study both micro- and macro- level properties of this model. Our model exhibits many complex and well-known social phenomena, suggesting that these phenomena could be caused by the simple mechanisms of exchange benefits and coordination costs among heterogeneous agents.

On the micro level, an interesting question is how an agent's endowments will affect their ego networks. In particular, we consider two variables for agents based on our model. The first variable is a quantitative measure of social status that we call

"social power"

$$\text{social power}(i) = \mathbf{b} \cdot \mathbf{w_i}. \tag{2.7}$$

Social power means "the potential for social influence" [99], or the potential benefits that one could bring to the other. Recall that b_k measures how beneficial the k-th dimension is. w_{ik} is the i-th agent's value on the k-th dimension. As $b_k \times w_{ik}$ increases, i is more likely to benefit others on the k-th dimension. Therefore, it is sensible to represent an agent's social status by the dot product of $\mathbf{b}$ and $\mathbf{w_i}$. Therefore, the definition of this variable is consistent with the concept, social power. The utility of this social power for social exchange leads naturally to the formation of a network structure that is often described as hierarchical, especially within the surrounding homophilic group.

The second variable is "social exclusion", which measures the extent to which an agent is marginalized [231]:

$$\text{social exclusion} = \|\mathbf{c} \circ \mathbf{w_i}\|_2. \tag{2.8}$$

Recall that we have constrained the means for all dimensions to be 0. If an agent has a large absolute value on some dimension, she is believed to be on the margin of that dimension because a higher cost is needed when she links to another arbitrary person.

We are interested in the correlation between the social power or social exclusion and statistics of their ego networks (i.e., degree and clustering coefficient). The results of the Andorra dataset is presented in Fig. 2-3, and similar results for other datasets are reported in Appendix. We find that "social power" is strongly positively correlated with degree, while "social exclusion" is strongly negatively correlated with degree. This finding is consistent with the implication of the proposed model: people with high (beneficial) endowments can potentially benefit others to a greater degree; people on the margin of the society have fewer opportunities to interact with others. More interestingly, we examine the correlations between social power/exclusion and the clustering coefficients for the nodes. A high clustering coefficient means that

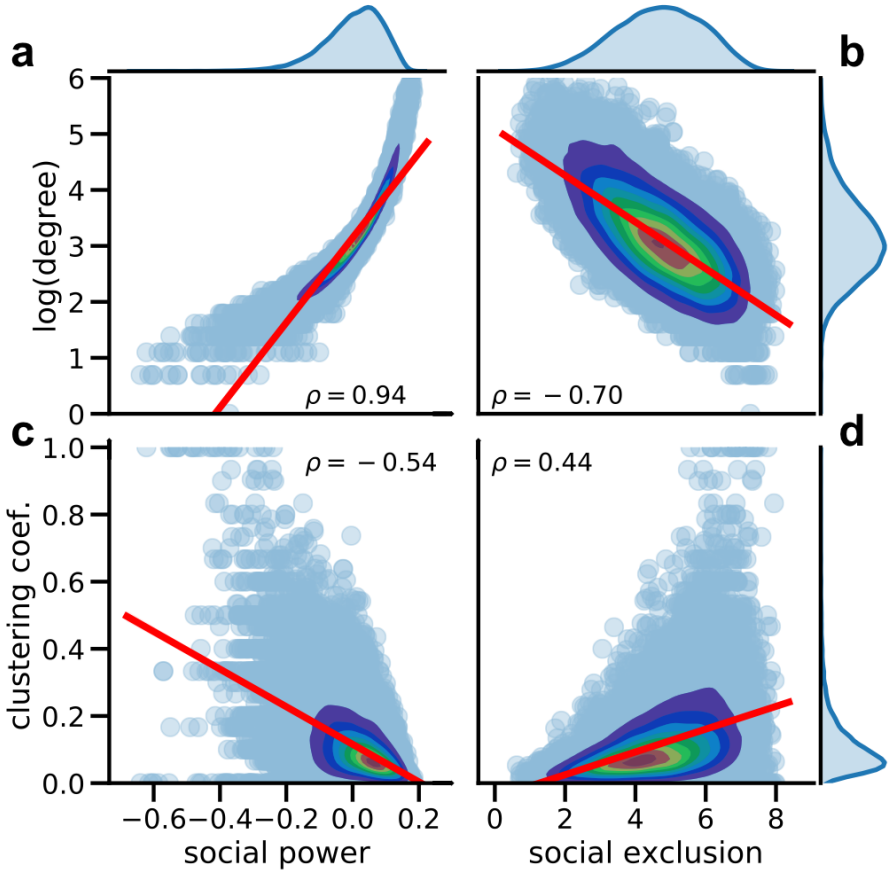


Figure 2-3: Correlations between endowment-related variables and ego networks statistics for the Andorra dataset. **a** Social power versus degree. **b** Social exclusion versus degree. **c** Social power versus clustering coefficient. **d** Social exclusion versus clustering coefficient. Colors represent the density of data points, where red indicates dense areas and purple indicates sparse areas. All data points are plotted as light blue dots. ρ denotes Pearson correlation coefficients. All Pearson correlation coefficients are significant at level $p < 0.001$.

the agent's neighbors are closely connected, and therefore indicates that the agent's neighbors might lack diversity. We find that people have lower clustering coefficients on the network if they have higher social power or lower social exclusion; that is, high status (power) people have more diverse social networks, a well-known and important aspect of human networks.

The proposed model can also predict the macro-level dynamics of networks. As an illustration, we are focused on the impact the systematic change of cost scaling parameters $\mathbf{c}$ (i.e. reducing $\mathbf{c}$ to $\mathbf{c}' = (1-\alpha)\mathbf{c}$, $\alpha \in [0,1]$) on the macro statistics of

the social network. Decreases in coordination costs are typically caused by advances in information technology (e.g. the Internet) or transportation (e.g. a new railway). We then employ agent-based modeling according to the learned endowment vectors and utility functions to reconstruct the empirical social networks (see Appendix for the approach of reconstruction). Finally, we compute density, average clustering coefficient, average shortest path in the giant component, and interaction diversity (defined as equation (2.9)) where $\mathcal{E}$ represents the set of edges for the network and $\mathbf{c}$ is the value after being reduced. Note that here we do not change the relative ratios among c_k $(1 \leq k \leq K)$; it is therefore sensible to incorporate the $\mathbf{c}$ into equation (2.9) after being normalized by $\|\mathbf{c}\|_2$.

$$\text{interaction diversity} = \frac{1}{|\mathcal{E}|} \sum_{(i,j)\in\mathcal{E}} \frac{\|\mathbf{c} \circ (\mathbf{w_i} - \mathbf{w_j})\|_2}{\|\mathbf{c}\|_2} \tag{2.9}$$

Figure 2-4 shows the impact of reducing $\mathbf{c}$ on the macro statistics of all networks. We find that as the cost scaling parameter $\mathbf{c}$ decreases, the density significantly increases while clustering coefficient does not increase much. This indicates that the decrease in coordination costs (e.g. adoption of the Internet) results in more links, and increases social cohesion or balance [55], i.e., the connectivity between one's neighbors. The decreasing trend of shortest paths between pairs reveals that the decrease of the coordination cost could diminish the power of social hierarchy. The increase of interaction diversity indicates that the decrease of coordination costs leads to greater connections between more dissimilar individuals. These synthetic findings indicate that the coordination costs' reduction, usually caused by technology advances, results in a society with more opportunities to connect to others, especially to dissimilar people; this is also a society with less hierarchy.

2.4 Discussion

Inspired by network embedding methods that represent agents by vectors, this study also applies vector representations for heterogeneous agents, referred to as their "en-

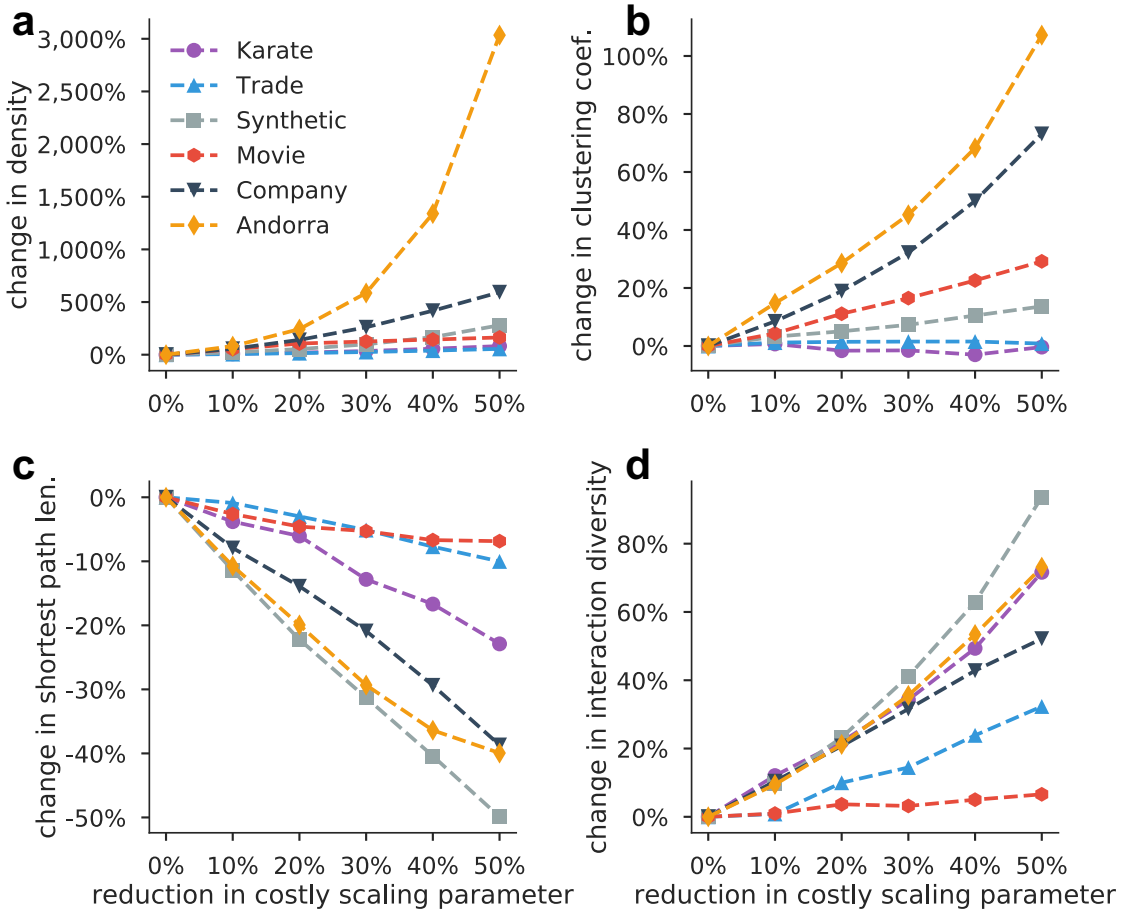


Figure 2-4: Impact of reduction in costly scaling parameters on macro-level network statistics. **a** Density. **b** Clustering coefficient. **c** Average of shortest path lengths. **d** Interaction diversity. X-axis represents the reduction of costly scaling parameters; y-axis represents the dynamics of the macro-level network statistics when costly scaling parameters are reduced.

dowment vectors". Our model is more interpretable than network embedding algorithms because we can economically and sociologically explain the link formation mechanism, by the trade-off between the exchange benefits and coordination costs among agents. We learned the endowment vectors from empirical network data, which can be used to predict a variety of other agent properties, and to demonstrate inter-agent network characteristics such as social status and diversity that are well-known from social science literature.

In particular, we highlight the necessity of trading off between beneficial exchange

effects and coordination costs. Most link formation models use only one or the other. We show that we can effectively learn the representations for agents from empirical networks by optimization methods that incorporate these trade-offs, without explicitly modeling social status, hierarchy, or the dynamics of social networks. This result suggests that many characteristics that are described in the social science literature are due to the trade-off between coordination costs and exchange benefits, rather than being fundamental effects or biases.

There are several interesting future directions based on this work. First, it is intriguing to consider the influence of existing neighbors on the marginal utilities of adding one more neighbor. For instance, the marginal utility of befriending a person should be higher when an ego has 10 friends than when the ego has 100 friends. Incorporating this interaction effect is difficult because this will require combinatorial optimization methods. Second, it is a promising direction to incorporate an indirect effect: the utility of "friends' friends". When we befriend a person, we do not only benefit from this person, but also this person's friends because we obtain useful information from and have small coordination costs with this person's friends. The indirect effect is reminiscent of several network embedding methods, including DeepWalk, which embed nodes on randomly sampled paths to have similar representations. Finally, we may take into account broader interaction effects such as "reputation": when people reach out to an ego, the ego may reciprocate a link even if the link does not directly benefit the ego.

2.5 Appendix

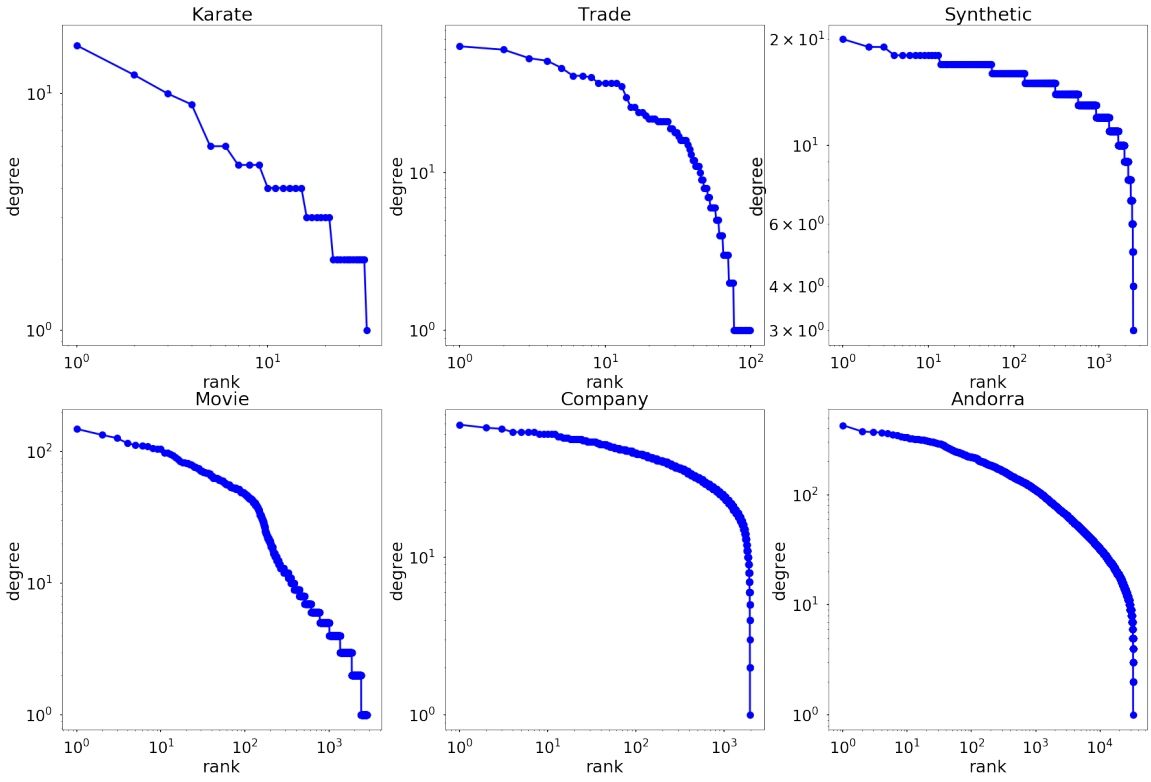


Figure 2-5: Degree distribution for all datasets.

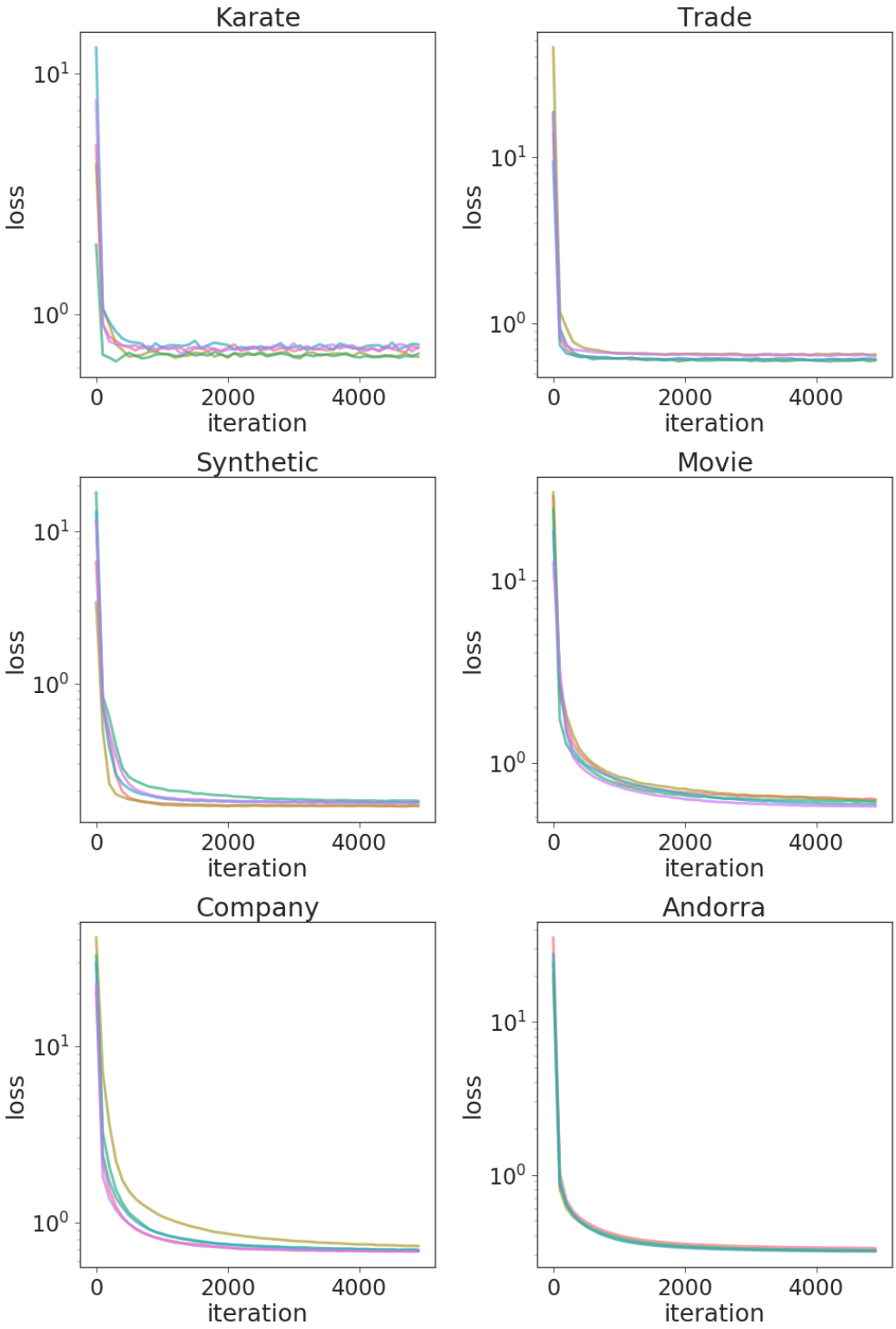


Figure 2-6: Learning curves for losses.

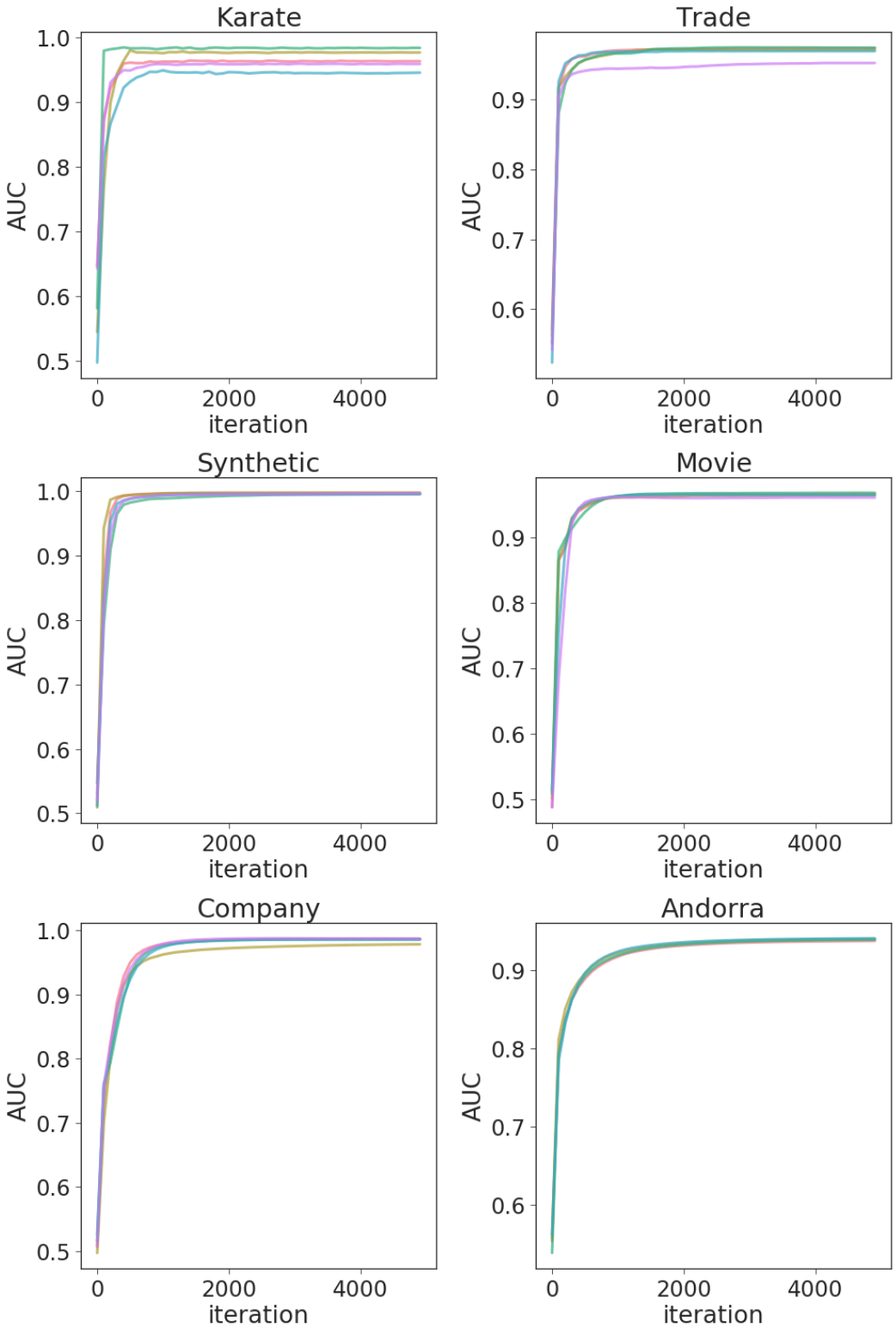


Figure 2-7: Learning curves for AUCs.

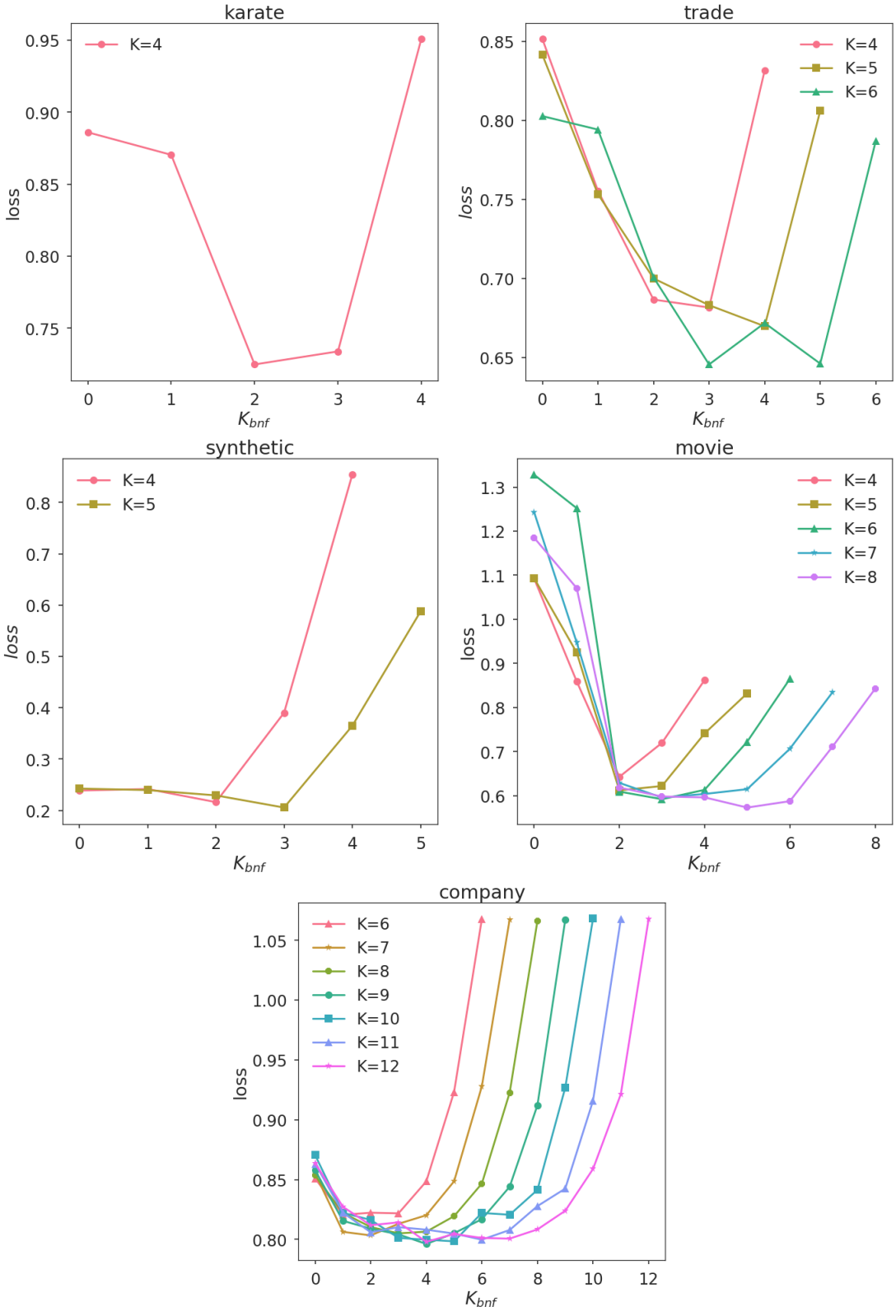
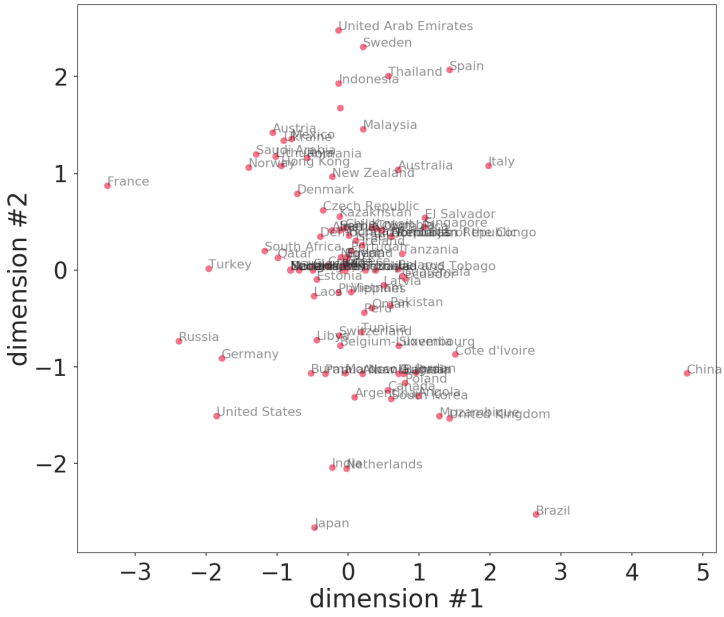
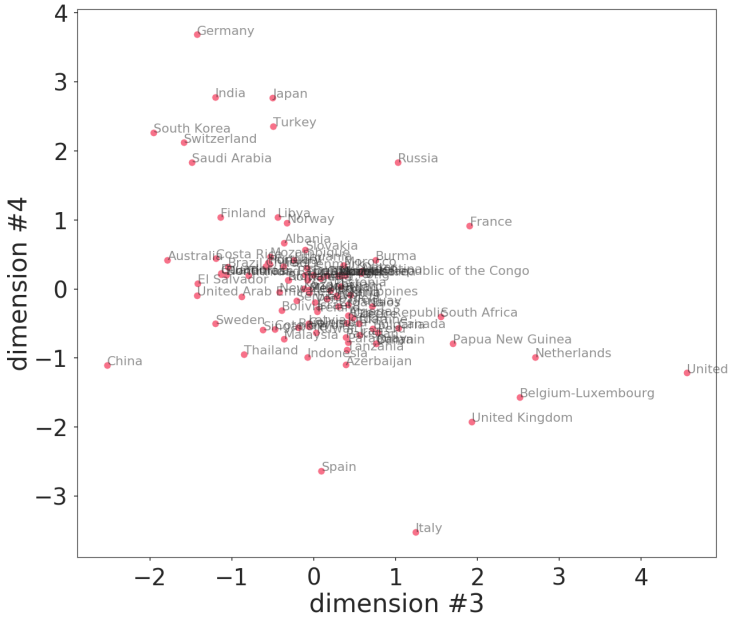
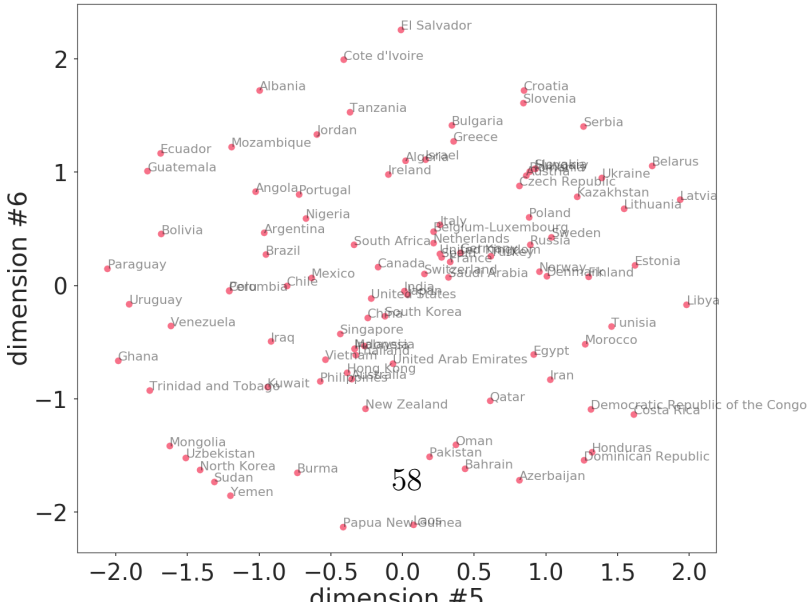


Figure 2-8: Learning losses for K and K_{bnf}.


dimension #2
dimension #1
United Arab Emirates
Sweden
Thailand
Spain
Indonesia
Malaysia
Australia
Italy
France
Denmark
New Zealand
Czech Republic
El Salvador
Tanzania
Turkey
Estonia
Pakistan
Russia
Germany
Cote d'Ivoire
China
United States
United Kingdom
Netherlands
Brazil
Japan


dimension #4
dimension #3
Germany
India
Japan
Turkey
South Korea
Switzerland
Saudi Arabia
Russia
Finland
Norway
France
Burma
South Africa
Papua New Guinea
Netherlands
China
Thailand
Azerbaijan
United States
Belgium-Luxembourg
United Kingdom
Spain
Italy


dimension #6
dimension #5
El Salvador
Cote d'Ivoire
Albania
Croatia
Slovenia
Tanzania
Bulgaria
Serbia
Jordan
Greece
Mozambique
Ecuador
Guatemala
Belarus
Ukraine
Angola
Portugal
Kazakhstan
Latvia
Lithuania
Nigeria
Poland
Bolivia
Argentina
Italy
Belgium-Luxembourg
Netherlands
Sweden
Brazil
South Africa
Russia
Paraguay
Canada
Estonia
Mexico
Chile
Uruguay
Libya
Venezuela
Tunisia
Iraq
Singapore
Morocco
Ghana
Egypt
United Arab Emirates
Iran
Trinidad and Tobago
Kuwait
Qatar
New Zealand
Democratic Republic of the Congo
Costa Rica
Mongolia
Oman
Honduras
Uzbekistan
Pakistan
Dominican Republic
North Korea
Burma
Bahrain
Sudan
Azerbaijan
Yemen
Papua New Guinea

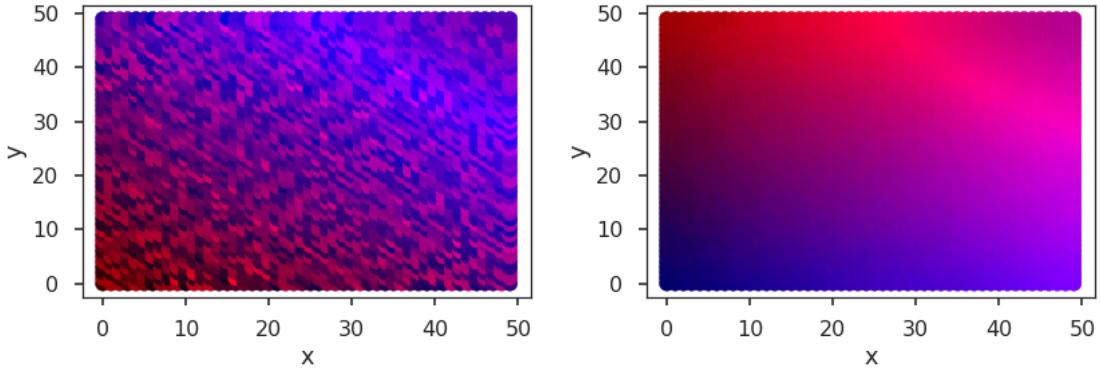


Figure 2-10: Learned endowment vectors for agent i, $i = 50x + y$. The left panel shows the two beneficial endowments, while the right panel shows the two costly endowments. Colors indicate the value on respective dimensions: red for the first and third dimensions and blue for the second and fourth dimensions. Purple indicates high value on both dimensions; black indicates low value on both dimensions.

Table 2.2: Statistics of Individual Characteristics in Andorra CDR

Phone Type	
Samsung	14,609
Apple	9,422
Others	8,798
Location Cluster	
Centre	20,830
St Julia	3,377
Massana	3,703
Arans	264
Encamp	3,641
Soldeu	536
Pas De La Cas	478
Internet usage (hour)	
0%	0.00
25%	5.95
50%	1,076.28
75%	3,088.07
100%	8,746.29

Table 2.3: Prediction performance of the present model and DeepWalk. Standard errors of the means in five classifications are reported in parentheses.

Dataset	**Present-SVM**	**Present-KNN**	**DeepWalk-SVM**	**DeepWalk-KNN**
Club-Karate	0.9900	0.9894	1.0000	1.0000
	(0.0100)	(0.0065)	(0.0000)	(0.0000)
Location-Trade	0.8045	0.8125	0.8597	0.9099
	(0.0333)	(0.0337)	(0.0439)	(0.0215)
ECI-Trade	0.8828	0.9098	0.8750	0.8504
	(0.0221)	(0.0186)	(0.0272)	(0.0257)
GDP-Trade	0.8319	0.7633	0.7355	0.7459
	(0.0607)	(0.0697)	(0.0622)	(0.0680)
Type-Synthetic	0.9279	0.9446	0.5340	0.4470
	(0.0036)	(0.0048)	(0.0081)	(0.0106)
Location-Synthetic	0.9986	0.9908	0.9971	0.9987
	(0.0001)	(0.0016)	(0.0021)	(0.0008)
Job-Company	0.8538	0.8549	0.9491	0.9352
	(0.0098)	(0.0075)	(0.0061)	(0.0040)
Job-Movie	0.9931	0.9854	0.5774	0.4824
	(0.0008)	(0.0045)	(0.0187)	(0.0041)
Gender-Movie	0.5445	0.5266	0.5546	0.5294
	(0.0105)	(0.0006)	(0.0140)	(0.0075)
Location-Andorra	0.6908	0.7141	0.8265	0.8215
	(0.0097)	(0.0064)	(0.0058)	(0.0084)
PhoneType-Andorra	0.5331	0.5592	0.5918	0.5805
	(0.0052)	(0.0019)	(0.0034)	(0.0049)
Internet-Andorra	0.6075	0.5822	0.6493	0.6446
	(0.0030)	(0.0305)	(0.0022)	(0.0070)

Table 2.4: Pearson correlation coefficients between social power or exclusion and degree or clustering coefficients. *: $p < 0.1$; **: $p < 0.01$.

Dataset	Karate	Trade	Synthetic	Movie	Company
social power vs log(degree)	0.713**	-0.179*	0.406**	0.808**	-0.011
social exclusion vs log(degree)	-0.365*	-0.833**	-0.257**	-0.593 **	-0.675**
social power vs clustering coef.	-0.583**	-0.345**	-	-	0.007
social exclusion vs clustering coef.	0.368*	-0.636**	-	-	-0.007

Table 2.5: The optimal AUC in five runs in the "split" and "non-split" conditions.

Condition	**Karate**	**Trade**	**Synthetic**	**Movie**	**Company**	**Andorra**
Split	0.9848	0.9685	0.9992	0.9608	0.9870	0.9476
Non-Split	0.9245	0.9537	0.9957	0.9611	0.9321	0.9318

Table 2.6: Comparison of predictive power of learned endowments on "split" and "non-split" conditions.

Dataset	**Split-SVM**	**Split-KNN**	**Nonsplit-SVM**	**Nonsplit-KNN**
Club-Karate	0.9900	0.9894	0.9089	0.9189
	(0.0100)	(0.0065)	(0.0365)	(0.0308)
Location-Trade	0.8045	0.8125	0.8140	0.8849
	(0.0333)	(0.0337)	(0.0222)	(0.0421)
ECI-Trade	0.8828	0.9098	0.9050	0.9093
	(0.0221)	(0.0186)	(0.0267)	(0.0322)
GDP-Trade	0.8319	0.7633	0.7833	0.7843
	(0.0607)	(0.0697)	(0.0672)	(0.0653)
Type-Synthetic	0.9279	0.9446	0.9346	0.9541
	(0.0036)	(0.0048)	(0.0044)	(0.0017)
Location-Synthetic	0.9986	0.9908	0.9986	0.9890
	(0.0001)	(0.0016)	(0.0001)	(0.0008)
Job-Company	0.8538	0.8549	0.6745	0.6442
	(0.0098)	(0.0075)	(0.0077)	(0.0123)
Job-Movie	0.9931	0.9854	0.9926	0.9864
	(0.0008)	(0.0045)	(0.0010)	(0.0025)
Gender-Movie	0.5445	0.5266	0.5199	0.5770
	(0.0105)	(0.0006)	(0.0109)	(0.0021)
Location-Andorra	0.6908	0.7141	0.5315	0.6215
	(0.0097)	(0.0064)	(0.0336)	(0.0116)
PhoneType-Andorra	0.5331	0.5592	0.4880	0.5463
	(0.0052)	(0.0019)	(0.0100)	(0.0067)
Internet-Andorra	0.6075	0.5822	0.5107	0.5889
	(0.0030)	(0.0305)	(0.0316)	(0.0104)

Table 2.7: Number of beneficial and costly dimensions in the non-split condition.

Dataset	$b_k, c_k = 0$	$b_k = 0,\ c_k > 0$	$b_k = 0,\ c_k < 0$	$b_k, c_k > 0$
Karate	0	1	2	0
Trade	0	2	3	0
Synthetic	0	2	2	0
Company	1	5	5	1
Movie	0	4	2	2
Andorra	0	5	11	0

Pairwise Stability

The proof of the Proposition (1) is derived directly from the definition of pairwise stability. On the one hand, removing an agent in S_i^* leads to a decrease in U_i, i.e.,

Table 2.8: The optimal AUC in five runs for variations of the functional form.

Condition	**Karate**	**Trade**	**Synthetic**	**Movie**	**Company**	**Andorra**
Present model	0.9848	0.9685	0.9992	0.9608	0.9870	0.9476
Smooth	0.9727	0.9675	0.9885	0.9567	0.9943	0.9395
Abs (ℓ_1)	0.9730	0.9722	0.9907	0.9717	0.9923	0.9379

Table 2.9: Comparison of the prediction performances between the present model and variations. Standard errors are reported in parentheses.

Dataset	**Present-SVM**	**Present-KNN**	**Abs-SVM**	**Abs-KNN**	**Smooth-SVM**	**Smooth-KNN**
Club-Karate	0.9900	0.9894	1.0000	0.9900	0.9900	0.9600
	(0.0100)	(0.0065)	(0.0000)	(0.0100)	(0.0100)	(0.0127)
Location-Trade	0.8045	0.8125	0.6939	0.7656	0.7594	0.7636
	(0.0333)	(0.0337)	(0.0313)	(0.0474)	(0.0215)	(0.0236)
ECI-Trade	0.8828	0.9098	0.7789	0.7562	0.8834	0.8399
	(0.0221)	(0.0186)	(0.0306)	(0.0533)	(0.0296)	(0.0362)
GDP-Trade	0.8319	0.7633	0.8768	0.8601	0.8730	0.8633
	(0.0607)	(0.0697)	(0.0512)	(0.0621)	(0.0568)	(0.0461)
Type-Synthetic	0.9279	0.9446	0.8970	0.9200	0.9307	0.9147
	(0.0036)	(0.0048)	(0.0052)	(0.0041)	(0.0037)	(0.0060)
Location-Synthetic	0.9986	0.9908	0.9985	0.9905	0.9514	0.9415
	(0.0001)	(0.0016)	(0.0002)	(0.0012)	(0.0014)	(0.0044)
Job-Company	0.8538	0.8549	0.77589	0.7657	0.8044	0.7814
	(0.0098)	(0.0075)	(0.0158)	(0.0082)	(0.0102)	(0.0071)
Job-Movie	0.9931	0.9854	0.9882	0.9795	0.9861	0.9774
	(0.0008)	(0.0045)	(0.0011)	(0.0048)	(0.0043)	(0.0042)
Gender-Movie	0.5445	0.5266	0.5011	0.5297	0.5043	0.5526
	(0.0105)	(0.0006)	(0.0086)	(0.0057)	(0.0096)	(0.0094)
Location-Andorra	0.6908	0.7141	0.5840	0.6033	0.6205	0.6395
	(0.0097)	(0.0064)	(0.0191)	(0.0089)	(0.0049)	(0.0054)
PhoneType-Andorra	0.5331	0.5592	0.5608	0.5631	0.5456	0.5654
	(0.0052)	(0.0019)	(0.0053)	(0.0061)	(0.0014)	(0.0030)
Internet-Andorra	0.6075	0.5822	0.6096	0.6122	0.5993	0.6176
	(0.0030)	(0.0305)	(0.0073)	(0.0103)	(0.0112)	(0.0053)

$$U_i(S_i^*; \mathbf{W}, \mathbf{b}, \mathbf{c}) \geq U_i(S_i^* / \{j\}; \mathbf{W}, \mathbf{b}, \mathbf{c}), \text{ for } j \in S_i^*. \tag{2.10}$$

In our specific form (equation (5) in main text), we have

$$\Delta u_i(j) \geq 0, \text{ for } j \in S_i^*. \tag{2.11}$$

On the other hand, forming a new link does not increase the utility of both agents at the same time, i.e., if

$$U_i(S_i^*; \mathbf{W}, \mathbf{b}, \mathbf{c}) \leq U_i(S_i^* \cup \{j\}; \mathbf{W}, \mathbf{b}, \mathbf{c}), \text{ for } j \notin S_i^*, \tag{2.12}$$

then

$$U_j(S_j^*; \mathbf{W}, \mathbf{b}, \mathbf{c}) > U_j(S_j^* \cup \{i\}; \mathbf{W}, \mathbf{b}, \mathbf{c}), \text{ for } i \notin S_j^*. \tag{2.13}$$

Equivalently, between $\Delta u_i(j)$ and $\Delta u_j(i)$, at least one should be less than zero. There-

fore, we have

$$\min\left(\Delta u_i(j), \Delta u_j(i)\right) < 0, \text{ for } i \notin S_j^* \text{ and } j \notin S_i^*. \tag{2.14}$$

Data description

Here we describe in detail the datasets that we utilize, with a focus on the individual characteristics to predict. Degree distributions are presented in 2-5.

Description of Andorra Dataset

We begin with a brief illustration of the format of the dataset. For each entry, we have the anonymized phone number which initiates a record (call/text/cellular), the starting time of the record, the duration of the record, the cell tower id of the initiator, the type of the record (call/text/cellular), cell phone carrier (containing country code), the anonymized phone number that receives the call or text, the type allocation code (containing phone type). The statistics of individual characteristics are presented in 2.2. Because we do not have fine-grained information for users, we use the following three variables as the proxy variables for individual characteristics.

- **Phone type**. In Andorra, the first and second largest phone types are Samsung and Apple, respectively. Therefore, we distinguish these two phone types from other phone types. We believe that phone type should be (weakly) correlated to individual income and social status; therefore we employ the phone type as one of the variables to predict.

- **Location**. Andorra Telecom records the cell tower location to which each call connects. We are then able to label each user by the place where she appears most frequently. Andorra Telecom classifies cell towers into seven clusters (as shown in 2.2). Note that the clusters do not strictly correspond to the seven districts in Andorra.

- **Internet usage**. For Internet usage, we show the minimum, the first quartile, median, the third quartile, and the maximum in the dataset. Youths tend to use the Internet more frequently than old people. Thus Internet usage should be a decent proxy for age since we lack sufficient fine-grained characteristics for the individuals.

Description of Movie Dataset

The TMDB is a public dataset[1]. In the dataset, we have 4,803 movies in total. For each movie, participants are classified into either cast or crew. Each individual is identified by a unique ID. We only extract the actors/actresses and directors, and then establish links between the director and the cast, meaning that "the director invited the actor/actress to the collaboration and the actor/actress agreed". If an individual serves as the director in more than half of the movies that she has engaged in, then we label her as "director", otherwise "actor/actress". The gender of each individual is also identified in the dataset.

Description of Company Dataset

The company dataset is also a public dataset (MobileD [2]). We examine this network because it shows the collaboration within a company and provides the labels of individual characteristics (managers or subordinates). In this network, managers' neighbors are mostly managers and subordinates' neighbors are mostly subordinates. Therefore, strong homophily (caused by reduced coordination costs) exists on the "job" dimension in this network but there are also exchanges between these two types of employees. We predict whether an employee is a manager or a subordinate.

Description of Trade Dataset

Because we do not consider the strength of a link, we will have a very densely connected network if we do not filter out some links. We filtered out links with export

[1] `https://www.kaggle.com/tmdb/tmdb-movie-metadata`
[2] `https://aminer.org/socialtieacross`

amounts lower than one billion; countries with high trade volumes would retain more links. We also remove countries without any link because then we have no information to estimate information for these countries. This process yields a network of 100 countries.

We extract three features for the 100 countries (or region) in the network:

1. Continent. We predict whether a country belongs to Africa, America, Asia/Pacific and Europe, and measure the performance by the average AUC of these four predictions. We have 14 African countries, 19 American countries, 33 Asian or Pacific countries, and 33 European countries in this network.

2. GDP. We use the GDP in 2014 as the second variable to predict. We label as 1 the countries with a GDP higher than the median; and otherwise 0.

3. Export complexity index (ECI). ECI measures the economic complexity for a country. For example, Japan has a high ECI because it exports products with high uniqueness. Similar to GDP, we label the countries with an ECI higher than the median as 1; and otherwise 0.

Description of Synthetic Dataset

As mentioned in the main article, we generate the network with two types of agents (e.g., buyers and sellers) who explore exchanges in their neighborhoods. We predict two individual characteristics in this dataset:

1. Type. As mentioned previously, we predict the type of each agent (a binary variable). The probability that the agent i is type A (e.g., buyers) is 0.5, and agents draw their types independently.

2. Location. We divide the 50×50 grid into four divisions: (1) $0 \leq x < 25$ and $0 \leq y < 25$; (2) $0 \leq x < 25$ and $25 \leq y < 50$; (3) $25 \leq x < 50$ and $0 \leq y < 25$; (4) $25 \leq x < 50$ and $25 \leq y < 50$. We predict whether a node belongs to each division, and then take the average AUC to measure the prediction performance for this characteristic.

Learning

Learning Method

Random noise. Since real-world social networks may incorporate random noises, we therefore add an *i.i.d.* random shock for the marginal utility into each pair (i, j). Following the conditions in Proposition (1), for $j \in S_i^*$,

$$\Delta u_i(j) + \epsilon_{ij} \geq 0; \tag{2.15}$$

and for $l \notin S_i^*$,

$$\min\Big(\Delta u_i(j), \Delta u_j(i)\Big) + \epsilon_{ij} < 0. \tag{2.16}$$

For computational simplicity, we assume that the CDF of random shock ϵ_{ij} is a sigmoid function,

$$\phi(x) = Pr[-\epsilon_{ij} \leq x] = \frac{1}{1 + \exp(-x - u_0)}. \tag{2.17}$$

We then establish the loss function. We learn the $\mathbf{W}$, $\mathbf{b}$, $\mathbf{c}$ by minimizing the loss function. u_0 can be considered as a baseline benefit or cost of forming a link. u_0 is learned along with $\mathbf{b}$ and $\mathbf{c}$ (can be perceived as a reserved utility) but is omitted in later equations.

Loss. We next specify the definition of the loss function:

$$\mathcal{L} = \mathcal{L}_{\text{pos}} + \mathcal{L}_{\text{neg}} + \mathcal{L}_{\text{fp}} + \mathcal{L}_{\text{reg}}. \tag{2.18}$$

- $\mathcal{L}_{\text{pos}} = -\frac{\sum_{(i,j)\in\mathcal{E}} \log\Big(\phi\big(|S_i^*|\Delta u_i(j)\big)\Big)}{\sum_{(i,j)\in\mathcal{E}} 1}$. This loss measures how well connected pairs are predicted. $\phi(x)$ is the sigmoid function ($\phi(x) = \frac{1}{\exp(-x-u_0)+1}$). u_0 is an unknown scalar to be learned along with $\mathbf{b}$ and $\mathbf{c}$. Multiplying $|S_i^*|$ is to consider that fact that when an agent has many agents (a large $|S_i^*|$), the marginal benefit from each neighbor ($\Delta u_i(j)$) does not need to be very large. In other words, we require $\Delta u_i(j)$ to be large if i has few neighbors but not necessarily large if

i has many neighbors.

- $\mathcal{L}_{\text{neg}} = -\frac{\sum_{(i,j)\notin\mathcal{E}} \log\left(1-\phi\left(\min\left(\Delta u_i(j), \Delta u_j(i)\right)\right)\right)}{\sum_{(i,j)\notin\mathcal{E}} 1}$. This loss measures how well disconnected pairs are predicted. This term follows the third condition in Proposition 1.

- $\mathcal{L}_{\text{fp}} = \lambda_{\text{fp}} \frac{\sum_{(i,j)\notin\text{FP}} \log\left(1-\phi\left(\min\left(\Delta u_i(j), \Delta u_j(i)\right)\right)\right)}{\sum_{(i,j)\in\mathcal{E}} 1}$. The penalty of existing false positives. FP is the set of pairs that are not in $\mathcal{E}$ but rank in the top $|\mathcal{E}|$ among all $\min(\Delta u_i(j), \Delta u_j(i)), \forall i, j$. Because the number of disconnected pairs is typically much larger than the number of connected pairs, false positives are usually not properly penalized. λ_{fp} scales the weight of this term.

- $\mathcal{L}_{\text{reg}} = \lambda_{\text{reg}}\|\theta\|_1$. Lasso ($\ell_1$) helps to eliminate unnecessary dimensions and to determine the choice of K. Therefore, if we learn a b_k or a c_k close to zero, we can remove this dimension and find an optimal K^*.

Dimensionality selection. We next provide a method to select the optimal dimensionality (K^*) along with (K^*_{bnf} and K^*_{cst}). Algorithm 1 presents the method. The philosophy is that for each K, we enumerate all possible K_{bnf}. "SGD" represents the algorithm for learning $\mathbf{W}$ and $\boldsymbol{\theta}$ via stochastic gradient descent and will be discussed later. If K_{bnf} gives rise to the best AUC for the current K, and the corresponding $\boldsymbol{\theta}$ contains 0, then we terminate the enumeration. After removing the dimensions with zero-valued θ_k (< 0.01 in practice), we obtain the optimal dimensionality and the values of corresponding variables. We stop increasing K if we find a zero-valued θ_k in the optimal run in the best K_{bnf} for this K. A zero-valued θ_k means that the k-th dimension does not contribute to either benefits or costs so we can drop that dimension without impacting the performance of this model. We call such dimensions **"degenerate dimensions"**.

Result: $\mathbf{W}^*, \boldsymbol{\theta}^*, K^*, K^*_{\text{bnf}}, K^*_{\text{cst}}$

Initialize K to a small number;

repeat

  $K = K + 1$;

  $\mathcal{L}^*(K) = +\infty$;

  for $0 \leq K_{bnf} \leq K$ **do**

    $\mathbf{W}(K, K_{\text{bnf}}), \boldsymbol{\theta}(K, K_{\text{bnf}}) = \text{SGD}(\mathcal{L}, K, K_{\text{bnf}})$;

    Let $\mathcal{L}(K, K_{\text{bnf}})$ be the current loss;

    if $\mathcal{L}(K, K_{bnf}) < \mathcal{L}^*(K)$ **then**

      $\mathcal{L}^*(K) = \mathcal{L}(K, K_{\text{bnf}})$;

      $K^*_{\text{bnf}} = K_{\text{bnf}}$;

    end

  end

until $\exists k$ *s.t.* $\theta_k \approx 0$;

$K^* = K - |\{\theta_k \approx 0\}|$;

$K^*_{\text{bnf}} = K^*_{\text{bnf}} - |\{b_k \approx 0\}|$;

$K^*_{\text{cst}} = K^* - K^*_{\text{bnf}}$;

$\mathbf{W}^* = \mathbf{W}(K^*, K^*_{\text{bnf}})$ (after removing the k-th dimension with $\boldsymbol{\theta}_k = 0$);

$\boldsymbol{\theta}^* = \boldsymbol{\theta}(K^*, K^*_{\text{bnf}})$;

Algorithm 1: The procedure of learning endowment vectors and other parameters.

Stochastic gradient descent (SGD) Here we describe the algorithm for SGD given K and K_{bnf}. We use TensorFlow [1] to implement the learning algorithm. Specifically, we use Adam optimizer [143] as the optimizer. We run a large number of iterations (5,000) to guarantee the convergence of learning. Importantly, because the memory cannot typically afford N^2 links as the input for one iteration, we retain all the connected pairs as the input while randomly sampling the number of unconnected pairs. Specifically, we sample ten times the number of connected pairs each iteration (four times for Andorra because of the memory limit); however we still have a consistent estimation for $\mathcal{L}_{\text{neg}}$ in expectation. This technique is called "negative sampling" and

is also used in many machine learning methods. For a typical network with a density of 0.1%, each iteration we sample approximately 1% pairs so every 100 iterations may cover all pairs on expectation. Also note that the stochastic nature of this learning process reduces the likelihood to converge to unwanted local minima. To approximate the optimum, we selected the run with the best AUC for link fitting.

Speed-up for Andorra Dataset. When the network is large (more than 10,000, e.g. Andorra) nodes, the computation of $\mathcal{L}_{\text{fp}}$ is very time-consuming. To speed up the learning, we could drop this term because empirically, this term is small for large-scale networks. For example, when learning endowment vectors for Andorra dataset, this term takes up less than 0.1% of the total loss even if the term is not optimized. Therefore, we drop $\mathcal{L}_{\text{fp}}$ in Andorra dataset to speed up the learning, but retain $\mathcal{L}_{\text{fp}}$ for other datasets to increase learning performance.

Choice of λ_{reg} and λ_{fp}. The value of λ_{reg} determines the constraint of the dimensionality and avoids overfitting. A large λ_{reg} will lead all b_k and c_k to be zero, while a zero-valued λ_{reg} fails to limit the dimensionality and the model may eventually overfit the data with an extremely high K. We select λ_{reg} *a priori* based on the complexity of the network. When the number of nodes is smaller than 1,000, we let $\lambda_{\text{reg}} = 0.1$; when the number of nodes is greater than 1,000 but smaller than 10,000, we set $\lambda_{\text{reg}} = 0.05$; otherwise we set $\lambda_{\text{reg}} = 0.01$. For λ_{fp}, we set it as 0.01 for all datasets. The good fitting and the prediction performance demonstrate that such selection is reasonable. Note that it is not necessary to obtain an optimal λ_{reg} or λ_{fp} if they result in satisfactory endowment vector estimations.

Learning rate and the number of iterations. There is no specific guidance to select these two terms *a priori*. Empirically, 0.1 is a reasonable initial learning rate, and it decays by 5% every 100 epochs. In addition, 5,000 can be a reasonable choice for the number of iterations. We will show in the next section that such choices result in sensible results.

Learning curves

Here we plot the learning curves for all datasets under the optimal settings. As described in *Methods*, we start from five random initial points to approximate the global optimum. As shown in 2-6 and 2-7, all runs converge during the first 1,000 iterations, meaning that our choices of the learning rate and the number of iterations are sensible. Moreover, most runs converge to similar losses and AUCs. This finding is important because it indicates that the local optimum is a good approximation for the global optimum.

Results of dimensionality selection

For each K, we have a $K^*_{\text{bnf}}(K)$ that results in the smallest loss. If a zero-valued b_k or c_k is learned at the setting of $K, K^*_{\text{bnf}}(K)$, we stop increasing K. Here we plot the losses for the best run for each pair of K, K_{bnf} for all datasets in 2-8 (except for Andorra) (we do not plot very small K for the convenience of visualization). The increment of K stopped at 4, 6, 5, 8, and 12 respectively for Karate, Trade, Synthetic, Movie and Company, respectively. When (K, K_{bnf}) is set to be $(4, 2)$, $(6, 3)$, $(5, 3)$, $(8, 5)$, and $(12, 4)$ for Karate, Trade, Synthetic, Movie and Company respectively, we derive the minimal losses. Then we drop degenerate dimensions: one costly dimension for Karate, Trade, Movie and Company; one beneficial dimension for synthetic. After deleting the degenerate dimensions, we derive (K^*, K^*_{bnf}): $(3, 2)$ for Karate, $(5, 3)$ for Trade, $(4, 2)$ for Synthetic, $(7, 4)$ for Movie, $(11, 4)$ for the Company dataset. For Andorra, because of the limit of computational resource, we are unable to enumerate for all (K, K_{bnf}). Instead, we set an arbitrary large $K = 16$ and reduce the dimensionality by the regularization term $\mathcal{L}_{\text{reg}}$, where we find one costly degenerate dimension. So we reduce one costly dimension, i.e. $K = 15$. Here we do not argue that 15 is an optimal choice for the Andorra dataset, but it is good enough for further agent-based modeling.

Graphical illustration

Illustration for the trade network

Since we have the trade network with only 100 countries, we can visualize the output of learned endowment vectors that is close to the optimal choice of dimensionality: $K = 6$ and $K_{\text{bnf}} = 4$. It is convenient for visualization to keep both K_{bnf} and K_{cst} even.

2-9 depicts the distribution of endowment vectors for all countries in the network. The first four dimensions are beneficial dimensions (with $b_k > 0$). We find that China, United Arab Emirates, the US, and Germany are high in these four dimensions respectively. The last two dimensions are costly dimensions (with $c_k > 0$). We find that geographically close countries are clustered on the space, meaning that the distance between countries plays an important role in the costs of network formation.

Illustration for synthetic dataset

As mentioned in the main article, nodes are located on a 50×50 grid. Each node can be either a seller or a buyer (with probabilities of 0.5 independently). We plot the learned endowments in 2-10. The plots show that we have recovered the data generating process. In the left panel, red and blue nodes are randomly distributed, which is consistent with the fact that buyers and sellers are randomly distributed on the grid. In the right panel, we observe that the degree of blue increases along the x-axis and that the degree of red increases along the y-axis. When both x and y are small, black is shown; when both x and y are large, purple is shown. In sum, the right panel demonstrates that we have successfully recovered the data generation of costly dimensions.

Comparison with network embedding algorithms

Although we have emphasized that the proposed model is not intended for dimension reduction of graph structures, we compare it with a classical network embedding algorithm, DeepWalk. Our goal is not to invent algorithms to outperform Deep-

Walk or other network embedding algorithms. In addition, to remain interpretability (link formation mechanism) and model simplicity, we do not enforce similarity among nodes on randomly sampled paths, which is why DeepWalk and other network embedding algorithms sometimes perform better in node classification tasks than we do, especially when similar nodes are clustered on the networks. As shown in 2.3, when the agents with this characteristic are densely clustered, the present model does not outperform DeepWalk, as expected. However, when characteristics do not show this "clustering" effect or may imply exchange effects (GDP in Trade, Type in Synthetic, Job in Movie), the present model *does* significantly outperform DeepWalk. This finding indicates that it is necessary to incorporate exchange effects along with coordination costs into network formation models, and that our model has a decent predictive power without losing interpretability in social science and link formation mechanisms.

"Social power or exclusion" for all networks

In the main article, we have presented the scatter plots about the correlation between social power or social exclusion and micro-level statistics of networks on the Andorra dataset. Here we show the correlations for other datasets in 2.4. Because the synthetic network is a bipartite graph and the movie network is almost a bipartite graph (some agents served as both directors and cast members), clustering coefficients are inapplicable. We find consistent results for most datasets with the results for the Andorra dataset. For the company data, we do not observe the significant correlation between social power and degree; we conjecture that this is because members in high status (managers) of the company do not need to communicate with every person.

Reconstruction of networks

Our methods to reconstruct the networks are as follows. Codes are provided online.

1. Learn the endowment vectors $\mathbf{W}$, and scaling parameters $\mathbf{b}$ and $\mathbf{c}$ with the aforementioned methods.

2. Rank the marginal utilities for all pairs (i, j) by equation (5) in main text (specifically, $\min(\Delta u_i(j), \Delta u_j(i))$) and establish a *reconstructed network* based on the top $|\mathcal{E}|$ pairs. Record the minimal utility $\min(\Delta u_i(j), \Delta u_j(i))$ in the reconstructed network, denoted by $\bar{u}$.

3. Multiple $\mathbf{c}$ by $1 - \alpha$, and repeat the previous step, to obtain *new reconstructed networks after decreasing* $\mathbf{c}$. In the reconstructed network, $\min(\Delta u_i(j), \Delta u_j(i)) \geq \bar{u}$.

4. Compute the statistics shown in Fig. (4) for each reconstructed network.

Robustness check: no splitting for beneficial and costly dimensions

We present the link fitting performance for the "split" (with beneficial and costly dimensions, as presented in the main article) and the "non-split" (b_k and c_k can be positive simultaneously) conditions. In 2.5, with the same number of dimensions, the AUC of the "split" condition is better than the "non-split" condition for all datasets. Note that because the AUCs of link fitting are typically very high, a significant decrease in AUC indicates a worse modeling fitting ability.

We also use the learned endowment vectors on the "non-split" condition to predict individual characteristics. We present the comparison between the "split" and "non-split" conditions in 2.6. We find that the prediction powers of endowments learned in the two conditions are comparable.

In addition, we also show that even if we do not force some b_k or c_k to be zero-valued in the "non-split" condition, we will still learn $\mathbf{b}^*$ and $\mathbf{c}^*$ such that for most k, either $b_k = 0$ or $c_k = 0$.

Robustness check: Variations of the functional form

We show the robustness of our model by using two variations of the functional form. The first one (Abs) is to change the G_i from ℓ_2 into an ℓ_1 norm:

$$\Delta u_i(j) = \sum_{k=1}^{K} b_k \max(w_{jk} - w_{ik}, 0) - \|\mathbf{c} \circ (\mathbf{w_j} - \mathbf{w_i})\|_1. \tag{2.19}$$

The second one ($Smooth$) is to use a smooth form of ReLU ($\max(x, 0)$) to assist the optimization:

$$\Delta u_i(j) = \sum_{k=1}^{K} b_k \log(1 + \exp(w_{jk} - w_{ik})) - \|\mathbf{c} \circ (\mathbf{w_j} - \mathbf{w_i})\|_2. \tag{2.20}$$

Note that when x is far away from 0, $\max(x, 0)$ is very close to $\log(1 + \exp(x))$. When x is around 0, $\log(1 + \exp(x))$ has a smoother gradient than $\max(x, 0)$, which is more convenient for the optimization. We can also consider $\log(1 + \exp(x))$ as a "noisy" form of $\max(x, 0)$: when x is close to zero, a small random shock will influence whether there is a benefit.

To show that these two versions of functional forms will not influence the fitting ability and the predictive power of the learned endowments, we present the fitting performance in 2.8 and the prediction performance in 2.9. We find that the performances for three versions of utility forms yield very similar results; the similar results mean that in the main article, it is reasonable to only present one functional form.

Chapter 3

Modeling dynamics of long ties with embedding techniques

3.1 Background

Social network analysis provides a powerful instrument to investigate the structure of society by aggregating interpersonal relationships among individuals [135, 34]. In the social network literature, a large body of research focuses on how tightly clustered social ties and groups are formed, as well as how they evolve, spread information and behaviors, and promote group solidarity [136, 258, 35, 163, 67, 93, 176, 88]. One central effect that can explain many social phenomena is homophily, a phenomenon whereby similar people tend to make friends with each other. For example, homophily serves as a reason for the formation of echo chambers, social segregation, and political polarization [217, 3, 134].

Meanwhile, a smaller but increasing number of studies focus on social ties that function as "bridges" between different communities, because of the unique roles they play in global network structures and information diffusion [160, 190, 56, 224]. This notion was originally put forward by Granovetter [111], who proposed the "weak tie hypothesis" and highlighted the potential value of these seemingly "weak ties" for spreading novel information across loosely connected communities. Additionally, Burt proposed a similar idea, which he named structural hole spanners, emphasizing

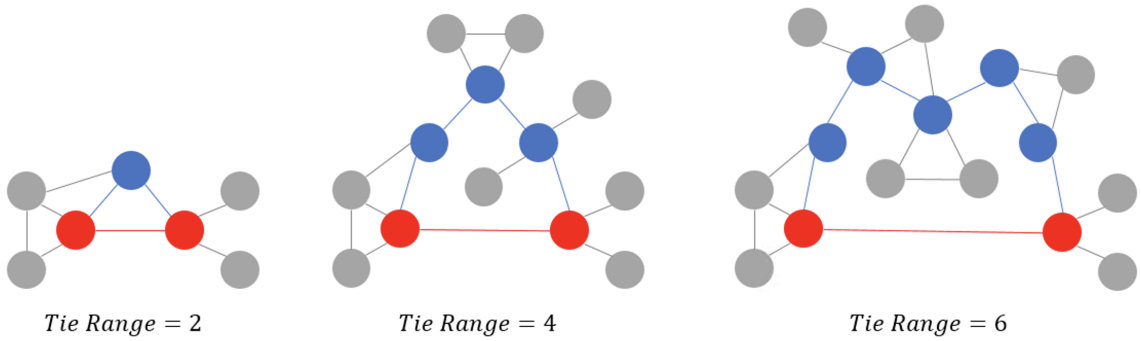


Figure 3-1: Tie range characterizes the length of the second shortest path between two connected nodes. The blue nodes are the nodes on the shortest path between the two red nodes except for their direct link.

the value gained by people situated in-between communities [52]. Moreover, the Watts–Strogatz network [258] highlights the role of these "bridges" in explaining the "small world" phenomenon [217]. In recent years, there has been a rapid growth in studies on weak ties from many different disciplines [160, 190, 274, 105, 107, 157].

One recent development in the literature is the concept of "long ties." They are social ties that have a large tie range. Tie range is measured by the length of the second shortest path between two connected nodes (see Fig. 3-1). Social ties with a large tie range may work as important social network bridges between different communities [111, 182, 206, 194, 244]. The idea of using tie range was already been discussed in the work of ref. [111]. In the decades since, there has been a debate on the effectiveness of long ties in spreading information and social contagion. While some researchers observe that long ties are crucial for the widespread of behaviors and information [258, 107, 243, 85], others find that long ties may have limitations for spreading novel information or continuing the social contagion of certain behaviors [57, 56].

When discussing long ties, researchers typically state that they are weak, meaning they have few interactions or low bandwidth. However, one recent study by Park (2018) shows that this is not necessarily the case; they find that not only are long ties not always weak, but that they can sometimes be much stronger than other social ties [194]. Despite this intriguing result, Park's paper also left open some interesting

questions regarding the corresponding mechanisms of long ties.

In the literature on weak and long ties, there is a lack of empirical evidence on their dynamics within temporal networks. Temporal networks, i.e., the network structures with the same set of nodes in a time series, are extensively studied in the complex network and graph data mining literature (e.g., [123, 161, 122, 148, 126]). However, few studies shed light on the dynamics of weak ties or long ties. One challenge is that large-scale dynamic networks without a profusion of missing nodes and links are surprisingly rare. This issue may result in the false identification of long ties since most publicly available temporal networks are derived from a subsample of nodes in a fully connected network. The paper that most closely aligns with the present study is ref. [181], in which the authors discuss the dynamics of tie strength. However, in their paper, the researchers did not approach the concept of tie range, nor did they use empirical data.

In the present study, we combine empirical analysis and computational modeling to understand the dynamics of long ties. First, we empirically examine the dynamic trends of long ties compared to other types of ties. Through this examination, we find that although long ties often initiate with weak tie strength, they become stronger and more persistent over time. Second, we further investigate the mechanisms of long ties, which ref. [194] did not fully address. In total, we discuss three possible mechanisms: degree heterogeneity, survival bias, and beneficial long ties [212, 259]. Firstly, we use empirical analysis to show that the first two mechanisms might not fully explain our main results. Next, we extend the interdisciplinary model (that uses game theory and node embedding techniques) proposed by ref. [268] to temporal networks. Our novel model incorporates the benefits that long ties bring into the utility function, such as novel information and diversity. The modeling results imply that the hypothesis of beneficial long ties might explain the patterns found in our study.

3.2 Results

3.2.1 Long ties last longer and become stronger

In this work, we employ *tie range* to quantify of the local network structure of a social tie. As the length of our data is two years, we partition the data into eight phases; however, our results are robust to other ways of partitioning, as well (see *SI*). To begin our analysis, we classify all social ties by tie range in the first phase, and then, we observe the evolution of those ties in the subsequent phases.

First, we examine the dynamics of interaction frequency (measured by the number of calls or texts) and interaction duration (measured by the total duration of the calls). While observing the magnitudes in just the first phase, we find a "U-shape" in the data that is consistent with the results of ref. [194]. Our result shows that interaction frequency and duration initially decrease with the tie range, but later increase with the tie range. We also find that long ties can be as intimate as short-range ties that are closely embedded in a social network.

By comparing the dynamics of short ties and long ties, we find that not only do long ties are initially stronger, but also that they persist longer. For example, in the long run, the average interaction duration and frequency of social ties with a tie range ≥ 6 even appear to be much larger than those with a tie range of 2. Furthermore, social ties with a tie range of 5 also appear to be stronger and more persistent than ties with a tie range of 3 or 4. In *SI*, we discuss the robustness of our findings by adjusting the time window that defines the length of each phase.

To understand what mechanisms drive the pattern shown in Fig. 3-2, we decompose the dynamics of interaction frequency or duration into persistence probability and interaction increments. We define y_t as the interaction frequency or duration in phase t, and we let the difference between phase t and 1 be $\Delta y_t = y_t - y_1$. Then

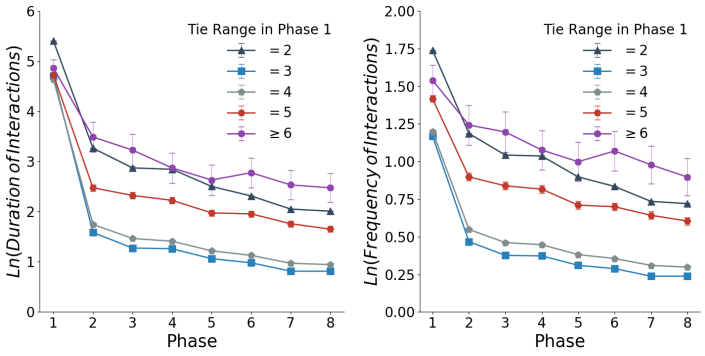


Figure 3-2: Evolution of both interaction frequency and interaction duration of ties throughout the eight seasonal snapshots. All ties are classified according to their tie range in the first season. Error bars are 95% confidence intervals.

persistence probability and interaction increments are defined as follows:

$$\begin{aligned}\mathbb{E}[y_t|y_1>0] &= \mathbb{E}[y_1+\Delta(y_t)|y_t>0,y_1>0]\mathbb{P}[y_t>0|y_1>0] \\ &= \big(\mathbb{E}[y_1|y_t>0,y_1>0] + \underbrace{\mathbb{E}[\Delta y_t|y_t>0,y_1>0]}_{\text{interaction increments}}\big)\ \underbrace{\mathbb{P}[y_t>0|y_1>0]}_{\text{persistence probability}}\end{aligned} \tag{3.1}$$

The dynamics of the persistence probability and interaction increments are presented in Fig. 3-3. As illustrated in the left panel in this figure, we find that social ties with a tie range ≥ 6 have the largest persistence probability in all subsequent phases, followed by closely embedded ties with a tie range of 2. Meanwhile, we find that social ties with a mid-sized tie range (i.e., 3 or 4) dissolve the fastest. This consistent trend with the overall effect presented in Fig. 3-2 suggests that the overall effect is largely driven by a different persistence probability in social ties with different tie ranges.

Regarding the interaction increments, we find that they generally increase with tie range, indicating that conditional on a persistent social tie, the interaction frequency and duration appear to be larger when this is a long tie. By contrast, social ties with a tie range of 2 have the smallest interaction increments. One of our conjectures is that the persistent short-range ties typically require less effort to maintain, as they can be indirectly maintained by their common friends; by contrast, we speculate that long ties require a lot of time investment to be able to be maintained.

In *SI*, we also conducted additional analysis to illustrate that in general, long ties

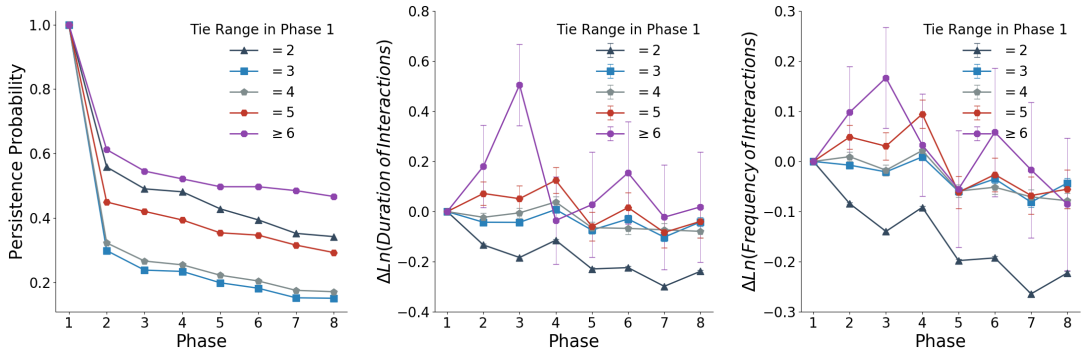


Figure 3-3: Dynamics of persistence probability and interaction increments conditional on the tie range in phase 1. Error bars are 95% confidence intervals.

have longer lifespans.

3.2.2 Dynamics of tie range

Next, we examine the dynamics of tie range. We first examine the dynamic trends of tie range in the first successive phase (the left panel of Fig. 3-4). Here, we only examine the social ties that persist in both phase 1 and phase 2. We present the transition probability matrix between tie ranges in successive phases in Fig. 3-4. As shown in the figure, for those social ties that persist, they have a large likelihood of evolving into short ties. In particular, for longer range ties, i.e., tie range of $= 5$ or ≥ 6), their probability of evolving into a tie range equal to 2 is the largest: 32% or 36%, respectively. Few short ties become long ties, since such an evolution requires that all their common neighbors dissolve with either of them. In addition, long ties seem also to be a stable status. For example, long ties with a social tie range ≥ 6 have a probability of 34% to have a tie range of 5, while 15% of them have a tie range of ≥ 6.

We next examine the tie range dynamics in phase 4 and phase 8, which are presented in the middle and right panels in Fig 3-4. We find the pattern in phases 4 and 8 are largely consistent with the pattern in phase 2. In particular, for those with tie range $= 5$ or ≥ 6 in phase 1, they have a probability of 26% or 38%, respectively, to persist with a tie range ≥ 5 in phase 4; they have a probability of 41% or 52%, respectively, to persist with a tie range ≥ 5 in phase 8. These results indicate that

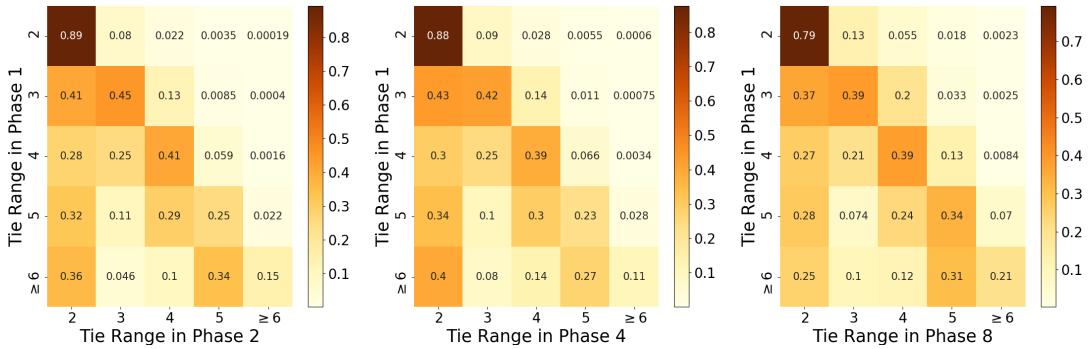


Figure 3-4: Transition probability matrix of tie range from phase 1 to a subsequent phase. Social ties that dissolved in the corresponding phase are disregarded in the analysis.

although long ties have a high probability of becoming short ties, they can also persist as long ties. This finding suggests that it is not necessary for a social tie to become a short-range tie to be long-lasting.

Next, we proceed to jointly investigate tie range and tie strength (i.e., the frequency and the total duration of interactions). As shown in Fig. 3-5, in general, those ties that become short-range (e.g., tie range = 2) are those with more interactions; for social ties that have a different initial tie range but that change to a tie range of 2, the interaction frequency or duration are always the greatest. For the persistence probability, the same trend generally holds. The one exception here is for those with a tie range ≥ 6: if they continue to be social ties with a tie range ≥ 6, their tie strength remains strong. Note that although we are only discussing phase 1 and phase 2, our results are equally robust when we examining any phase t and its first subsequent phase, $t+1$ (see *SI*).

3.2.3 Discussing the mechanisms

In the previous sections, we show that long ties are not only stronger but also last longer. Moreover, quite a few strong long ties continue to be long ties. To discuss the plausible explanations for the observed patterns, We next propose and discuss three hypotheses pertaining to degree heterogeneity, survival bias, and valuable long ties.

Degree heterogeneity. First, one plausible explanation for the observed patterns is

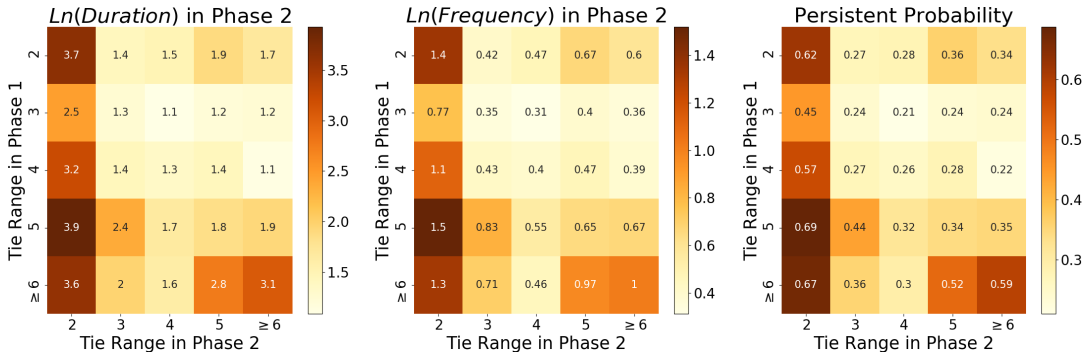


Figure 3-5: Interaction duration (a), frequency (b) and persistent probability (c) in the next phase when tie range evolves. The text above the figures indicates the meanings of the numbers.

degree heterogeneity. As shown in Fig. S8 in *SI*, we find that individuals who have fewer friends are more likely to have long ties. Thus, they tend to retain the relationships with their small number of friends, but with greater persistent probability and interaction increments.

To reduce the impact of degree heterogeneity, we plot the results conditional on the degree range. The results of this work are presented in Fig. S7 in *SI*. We find that the patterns observed in our main text are found in all degree subgroups. This finding shows that although degree heterogeneity may provide an explanation for the observed patterns, it does not fully explain our main results.

Survival bias. The second plausible explanation is survival bias – that only very valuable long ties survived – even though newly-formed long ties are likely weaker than newly-formed short ties. Therefore, surviving long ties tend to continue to persist, or perhaps even become stronger, while others dissolve rapidly. To test for this hypothesis, we need to examine (1) whether newly formed long ties are weaker than newly formed short ties in the beginning and (2) whether newly formed long ties have a smaller persistence probability, such that only very strong long ties survive. We find (1) is supported while (2) is not supported, and thus survival bias cannot fully explain our results.

To investigate these two ideas, we put social ties into one of two categories: existing ties, and new ties. An existing tie is one that as had any interactions in the previous

phase, while a new tie has had no such interactions. After separating all ties into existing or new ones, we perform the same analysis as that found in the previous sections. We use the tie range in phase 2 as the reference, and we investigate whether there was non-zero interaction frequency or duration in order to determine if it is a new or existing tie.

We first examine whether newly formed long ties are weaker initially than newly formed short ties. In Fig. S8 in *SI*, we show that while existing ties present a "U-shape" in the relationship between the interaction frequency (duration) and tie range in phase 2, this "U-shape" pattern does not hold for new ties. Instead, as indicated by Fig. S8 in *SI*, for new ties, the longer the new tie is, the fewer interactions the two people have in phase 2. This result supports our conjecture that newly formed long ties are likely to be weaker than newly formed short ties.

Next, we investigate whether newly formed long ties have a smaller persistence probability. However, as indicated by Fig. S9 in *SI*, we observe that for newly formed ties there exists a U-shape between tie range and persistence probability; importantly, newly formed long ties have the highest persistence probability. This finding contradicts our conjecture that the persistence probability of newly formed long ties would be the smallest. Thus, for the two notions we examined, we find that (1) is supported while (2) is not supported. Therefore, the survival bias hypothesis does not fully explain our main results.

Valuable long ties. Our last hypothesis is that long ties tend to be more valuable. This hypothesis is consistent with the weak tie theory and the roles of long ties, as previous studies conjecture [111, 258]. However, while most of computational models that simulate the real-world networks highlight homophily [177] – the phenomenon that individuals with similar attributes tend to be friends – previous models do not typically consider the benefits of social exchange between people with different skill or information sets [268]. Recent work, such as that by [268], provides an example of how one can consider homophily and social exchange jointly, but this work is restricted to static social networks. Below, we propose a computational model that combines game theory and machine learning in order to examine long tie dynamics and which

supports our hypothesis on valuable long ties.

3.2.4 The interdisciplinary model

In this section, we propose a game-theoretical computational model that simulates the dynamics of social networks. Specifically, the model combines the embedding techniques [197, 112, 144, 253] in machine learning and the strategic network formation in economics [136, 66]. Compared to the common network formation game models in the economics literature, our model stresses the high-dimensional heterogeneity, as well as the values of social exchange. Compared to network embedding techniques, our model helps understand the social network formation mechanisms. Ultimately, our model integrates the strategic network formation approach to explain the mechanisms, while the embedding techniques improve the predictability of the computational model. Our study echoes the recent paper that discusses the trade-off between explanation and prediction in computational social science [120].

Our model considers two procedures during the formation of social ties: the meeting procedure, and the choice procedure. This two-step model takes into account the dynamics of social ties – that people first meet others randomly, and then make their rational decisions about the choice of friends. The meeting procedure models the reality, wherein people meet each other at random. There may exist many potential neighbor candidates who are mutually beneficial (e.g., some potentially valuable long ties), but the extremely low meeting probability can prevent the social tie from being formed. Moreover, when first meeting a new neighbor, a person may lack sufficient information to assess the person, and they are unable to make a rational decision about the social tie. After getting to know a new friend over a period of time (one phase in our study), the individual can then start to make make a rational decision. The choice procedure assumes that individuals are rational when choosing their network neighbors and that each individual maximizes their utility function.

Formally, let $\mathcal{I}$ be the set of individuals and let i (or j, ℓ) be their index. Additionally, let t index the discrete time steps (or phases), and thus $t \in \mathbb{N}^+$. Let $\mathbf{A}^{(t)}$ denote the adjacency matrix in phase t. $\mathbf{A}^{(t)}_{ij} = 1$ indicates that i and j are connected

in phase t. $\mathbf{A}_{ij}^{(t)} = 0$ indicates that i and j are disconnected in phase t. For simplicity, we only consider an undirected network, i.e., $\mathbf{A}_{ij}^{(t)} = \mathbf{A}_{ji}^{(t)}$ for all $i, j \in \mathcal{I}$ and for all $t \in \mathbb{N}^+$. To account for the heterogeneity of individual attributes, we use the "endowment vector" $\mathbf{w}_i$, which is a K-dimensional vector as the embedding [197, 112]. As embedding techniques do, each dimension measures a certain latent attribute of an individual, such as a type of skill or useful information. A larger w_{ik} indicates that the individual retains a high endowment of the k^{th} dimension.

In each phase, the neighbor's set of i consists of two components: the new friend set $\mathcal{M}_i^{(t)}$, and the existing friend set $\mathcal{N}_i^{(t)}$. The new friend set is formed in the random meeting procedure. We assume each pair of individuals has a different meeting probability. The concept of the "meeting probability" is widely adopted in several econometric studies that aim to model social network formation [178, 66]. Specifically, for each pair of individuals (e.g., i and j), they have a probability of $p_{ij}^{(t)}$ to "meet" each other in phase t. If $\mathbf{A}_{ij}^{(t-1)} = 1$, that is, the two individuals were connected in phase $t-1$, then the $p_{ij}^{(t)}$ is a large probability. Otherwise, $p_{ij}^{(t)}$ is a small probability dependent on the network topology between i and j. We can imagine that if this is a long tie, the probability would be much smaller. Formally, we parametrize $p_{ij}^{(t)}$ as follows:

$$p_{ij}^{(t)} = \begin{cases} d_{t-1}(i,j) & \mathbf{A}_{ij}^{(t-1)} = 0 \\ q & \mathbf{A}_{ij}^{(t-1)} = 1 \end{cases} \tag{3.2}$$

The distance metric $d_{t-1}(i,j)$ depends on the network topology between individual i and individual j in phase $t-1$. We define the distance metric to be proportional to the probability of a random walk from i to j. Here, q is set to describe the probability of be maintained in the meeting procedure in phase t.

The second component is the existing friend set $\mathcal{N}_i^{(t)}$. It is a subset of all friends in phase $t-1$, i.e., $\mathcal{N}_i^{(t)} \in \mathcal{M}_i^{(t-1)} \cup \mathcal{N}_i^{(t-1)}$. This means that individuals make rational decisions after maintaining their friendships for a period of one phase. The rationale behind this notion is that individuals need a significant amount of time to observe and

understand the value of an existing friend, so the rational choice procedure happens in the phase immediately following the meeting procedure. For a connected social tie in phase $t-1$, the friendship must survive in both the meeting procedure (a random draw from Bern(q)) and the rational choice procedure. The choice procedure is modeled using the following utility function:

$$U_i^{(t)}(\mathbf{c}_i^{(t)}) = \sum_{j \in \mathcal{M}_i^{(t-1)} \cup \mathcal{N}_i^{(t-1)}} \left(c_{ij}^{(t)} \sum_k \left(\sigma(w_{jk} - w_{ik}) + \sum_{\ell \in \mathcal{M}_j^{(t-1)} \cup \mathcal{N}_j^{(t-1)}} \delta\sigma(w_{\ell k} - w_{ik}) \right) - \left(c_{ij}^{(t)} \right)^2 \right), \quad \text{where } \sum_j \left(c_{ij}^{(t)} \right)^2 = 1. \tag{3.3}$$

Here, $U_i^{(t)}$ is the utility function of individual i in phase t. $\mathbf{c}_i^{(t)} \in [0,1]^{\mathcal{M}_i^{(t-1)} \cup \mathcal{N}_i^{(t-1)}}$, which can be understood as a function that maps any j in the neighbor set in phase $t-1$, i.e., to map each element in $\mathcal{M}_i^{(t-1)} \cup \mathcal{N}_i^{(t-1)}$, to a real number in $[0,1]$. The utility function sums over all i's neighbors in phase $t-1$. σ is the ReLU function: if $w_{jk} - w_{ik} > 0$, it outputs $w_{jk} - w_{ik}$; otherwise 0. ℓ enumerates over all j's neighbors in phase $t-1$, which are also i's "friends' friends." The depreciation factor δ, which ranges in $(0,1)$, measures how the value of a potential friend depreciates as the distance on the network increases. We refer to $\sigma(w_{jk} - w_{ik}) + \sum_{\ell \in \mathcal{M}_j^{(t-1)} \cup \mathcal{N}_j^{(t-1)}} \delta\sigma(w_{\ell k} - w_{ik})$ as the benefit that j brings to i. In addition, we separate the benefit into two: the *direct benefit*, $\sigma(w_{jk} - w_{ik})$, and the *indirect benefit* $\sum_{\ell \in \mathcal{M}_j^{(t-1)} \cup \mathcal{N}_j^{(t-1)}} \delta\sigma(w_{\ell k} - w_{ik})$. The design of these benefit terms was intended for our valuable long tie hypothesis – we hope to observe that long ties have, on average, larger values in the direct benefit term.

$c_{ij}^{(t)}$ measures the time investment of i in j. A non-zero value of $c_{ij}^{(t)}$ indicates that j belongs to $\mathcal{N}_i^t$. The restriction of the sum of squared $c_{ij}^{(t)}$ reflects that people have limited time or energy to invest in their neighbors. The benefit of each neighbor is proportional to the time or energy investment in each neighbor j; this is why we

multiply the benefit term by $c_{ij}^{(t)}$. At the same time, the squared term $\left(c_{ij}^{(t)}\right)^2$ is used to measure the cost of time or energy. The design of $c_{ij}^{(t)}$ echoes our degree heterogeneity hypothesis – those with many ties may have less investment in any one individual neighbor.

By the Cauchy-Schwarz inequality, Equation (3.3) can be solved by

$$(c_{ij}^{(t)})^* \propto \sum_k \left(\sigma(w_{jk} - w_{ik}) + \sum_{\ell \in \mathcal{M}_j^{(t-1)} \cup \mathcal{N}_j^{(t-1)}} \delta\sigma(w_{\ell k} - w_{ik}) \right), \text{ and } \sum_j \left((c_{ij}^{(t)})^*\right)^2 = 1. \tag{3.4}$$

In particular,

$$j \in \mathcal{N}_i^{(t)} \text{ iff } \left(c_{ij}^{(t)}\right)^* > 0; j \notin \mathcal{N}_i^{(t)} \text{ iff } \left(c_{ij}^{(t)}\right)^* = 0. \tag{3.5}$$

In other words, if the optimal solution informs $\left(c_{ij}^{(t)}\right)^* = 0$, then this indicates that i and j are not longer connected. Otherwise, $\left(c_{ij}^{(t)}\right)^*$ is proportional to the duration during which i interacts with j.

This model provides major improvements based on the framework proposed by ref. [268]. First, different from their paper, we establish a model for network dynamics. In particular, we incorporate a meeting procedure: this addresses the phenomenon that in reality, there are many neighbor candidates who do not form links purely because they have no opportunity to meet. Second, our model also takes into account the "weight" (i.e., the interaction frequency or duration) of the links. This is different from ref. [268] where the weights between the links are binary. Third, ref. [268] assumes that the marginal utility of additional neighbors is not dependent on other existing neighbors; by contrast, our model does not incorporate this assumption, and it also accounts for the network externality (i.e., the benefits of friends of friends). We provide additional analyses to verify our modeling fitting capacity in *SI*.

Figure 3-6 provides the main implications derived from learning results of our model. We first present the average benefit, i.e., $\sigma(w_{jk}-w_{ik})+\sum_{\ell \in \mathcal{M}_j^{(t-1)} \cup \mathcal{N}_j^{(t-1)}} \delta\sigma(w_{\ell k} - w_{ik})$ given the different tie range in Panel (a). The average is taken over all candidate

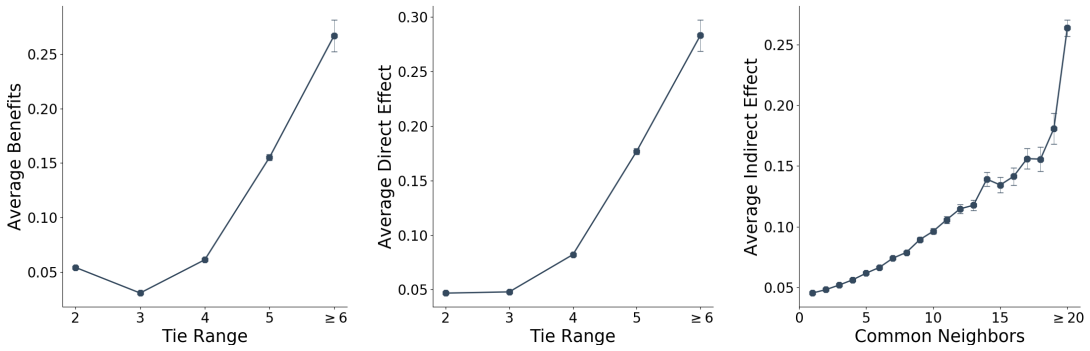


Figure 3-6: The results implied by our model. (a) The corresponding result of the model which balances the time investment and benefit. (b) Direct benefit from 1-neighbors. (c) Indirect benefit from common neighbors.

neighbors in $\mathcal{M}_j^{(t-1)} \cup \mathcal{N}_j^{(t-1)}$ given the tie range in phase $t-1$. From this, we find a U-shape, i.e., the average benefit decreases with tie range at the beginning, but later increases with the tie range. This is consistent with our previous findings regardng the U-shape between tie range and tie strength.

Next, we separate the benefits in Equation (3.3) into the direct effect and the indirect effect. We present the average direct effect, which is $\sigma(w_{jk} - w_{ik})$ in Panel (b). We observe an increasing pattern with the tie range, indicating that as the tie range increases, the average benefit that a tie brings also increases. This result supports our hypothesis that long ties tend to be more valuable, which explains the results in the previous sections. We also compute the average indirect effect, i.e., $\sum_{\ell \in \mathcal{M}_j^{(t-1)} \cup \mathcal{N}_j^{(t-1)}} \delta\sigma(w_{\ell k} - w_{ik})$. In our model, only social ties with common friends, i.e., those with a tie range of 2, have indirect effects. We plot the relationship between the number of common neighbors and the average indirect effect. As observed in Panel (c), we find an increasing pattern. In particular, by examining the first several data points in the plot, we observe a seemingly convex pattern, indicating the increasing marginal utility of common neighbors.

3.3 Discussion

In this study, we combine empirical data analysis and an interdisciplinary computational model to study the dynamics of long ties and their differences from other

types of ties. We find that very long ties tend to be stronger, and they persist longer than other, shorter-range ties, even though long ties may initially be weak. Moreover, we find that long ties tend to have a stable status, as many social ties persist to become long ties. We propose three hypotheses to investigate the mechanisms at play. After discussing hypotheses regarding degree heterogeneity and survival bias, we conclude that they might not fully explain our main results. Finally, we propose an interdisciplinary model that combines game theory and machine learning to account for heterogeneity in the values of social ties. The modeling results indicate that the hypothesis that long ties are more valuable than other ties might explain our results.

Our results also signal the importance of mixing diverse people in promoting social cohesion. For example, both our empirical analysis and modeling results indicate that people who are dissimilar in certain attributes or distant in a social network may have significant mutual benefits. However, the small likelihood of those people meeting can hinder the formation of their future interactions.

Based on this study, there are several interesting research directions that could be investigated. First, although we examine a large-scale social network with very few missing nodes, our dataset only reflects communication taking place over phones. Therefore, it would be interesting to examine the external validity of our results compared to offline social networks or online social media networks. Second, there may be interesting variants of our model. For example, our model only reflects the absolute advantages that other people bring, but it would be interesting to incorporate comparative advantages in our model, as well. Finally, It would be interesting to find a universal metric that combines tie range and tie strength when we assess the relationship between two nodes in social networks.

3.4 Methods and Materials

3.4.1 Data description

In our study, we use a nationwide call detail record dataset. Users' private information has been anonymized and thus we are able to identify them. This data provider is a company that functions as the service provider for most of the mobile phone users in that country. The time period covered by the data starts from Jan. 2015 to Dec. 2016. In the dataset, we retrieve the total number of calls, text, as well as the duration of calls between any two people in each month.

We establish a temporal social network with the dataset. We consider discrete time steps (or phases): for each phase, we construct a "snapshot" of the network, where the node indicates a user and edge represents the interaction between two users. A key question is how we determine the window of each phase. In our main results, we adopt a seasonal window – we treat every four months as a phase. In *SI*, we also use one month or six months to verify the robustness of our results.

There are tourists from foreign countries who also visited this country. To maintain a temporal network where the node set is stable and the global network structure does not change dramatically with the dynamics of a few nodes, we only consider the interactions among users who have at least one call or text in each phase. We construct a temporal directed network with 45192 nodes and 385533 edges on average for each phase.

In terms of the weight of the directed network, we consider two variables as mentioned in the main text: interaction frequency and duration. Interaction frequency is the total number of calls or text that node i sends to j; there are a few calls with zero-second duration and we filter those calls out. Interaction duration is the total time length that i calls j in each phase, and does not account for texting.

3.4.2 Tie range and long ties

Tie range[111, 194] is defined as the length of the second shortest path between two connected nodes (Fig 3-1). It indirectly reflects the network distance of the connection. Consistent with previous long tie studies [57, 194], there is no clear cutoff of tie range that decides whether a tie is short or long tie. A good reference is the Milgram experiment, which suggested that the average network distance between every two people is approximately 6. In our study, we treat social ties with a tie range of 2 as short ties, and ties with 5 or ≥ 6 as long ties.

3.4.3 Details in learning

Based on Equation (3.4), we construct the loss function to minimize the MSE Loss between c_{ij} and its right hand side. We use stochastic gradient descent to optimize the loss function. For each epoch, we construct our loss function as below:

$$\mathcal{L} = \mathcal{L}_{pos} + \mathcal{L}_{neg}, \tag{3.6}$$

The loss function is composed of the loss functions of positive (connected pairs), and negative samples (disconnected pairs).

$$\mathcal{L}_{pos} = \sum_{i \in N_{sampled}} \sum_{j \in \left(\mathcal{N}_i^{(t-1)} \cup \mathcal{M}_i^{(t-1)}\right) \cap \mathcal{N}_i^{(t)}} |\hat{c}_{ij}^{(t)} - c_{ij}^{(t)}| \tag{3.7}$$

$$\mathcal{L}_{neg} = \sum_{i \in N_{sampled}} \sum_{j \in \left(\mathcal{N}_i^{(t-1)} \cup \mathcal{M}_i^{(t-1)}\right) \setminus \mathcal{N}_i^{(t)}} \hat{c}_{ij}^{(t)} \tag{3.8}$$

For positive samples, we minimize the difference between $c_{ij}^{(t)}$, the time investment of i on j, and the predicted time investment denoted by $\hat{c}_{ij}^{(t)}$.

$$c_{ij}^{(t)} = \frac{\log\left(D_{ij}^{(t)} + 1\right)}{\sum_{j \in \mathcal{M}_i^{(t-1)} \cup \mathcal{N}_i^{(t-1)}} \log\left(D_{ij}^{(t)} + 1\right)} \tag{3.9}$$

where $D_{ij}^{(t)}$ is the interaction duration between i and j in phase t. To reduce the impact of extreme values, we take the logarithm of $D_{ij}^{(t)}$. Since $D_{ij}^{(t)} \geq 0$, $c_{ij}^{(t)} \geq 0$.

$$\hat{c}_{ij}^{(t)} = \frac{\exp\left\{\sum_k \left(\sigma(w_{jk} - w_{ik}) + \sum_{\ell \in \mathcal{M}_j^{(t-1)} \cup \mathcal{N}_j^{(t-1)}} \delta\sigma(w_{\ell k} - w_{ik})\right)\right\}}{\sum_{j' \in \mathcal{M}_i^{(t-1)} \cup \mathcal{N}_i^{(t-1)}} \exp\left\{\sum_k \left(\sigma(w_{j'k} - w_{ik}) + \sum_{\ell \in \mathcal{M}_{j'}^{(t-1)} \cup \mathcal{N}_{j'}^{(t-1)}} \delta\sigma(w_{\ell k} - w_{ik})\right)\right\}} \tag{3.10}$$

When minimizing the loss function, we treat the time investment of i in j, which is calculated by the interaction duration or frequency, as the input and endowment vectors in this loss function as the variables to be inferred. Note that the existence of the δ may result in an uncontrollable gradient issue. We thus use grid search for this variable and check the robustness of our results in *SI*. Moreover, we also pay attention to dimension selection of endowments. For more details, please see *SI*.

To facilitate the learning process, we apply mini-batch stochastic gradient descent with Adam optimizer.[143] Consistent with conventional network embedding algorithms, node sampling probability is proportional to node degree ($d^{\frac{3}{4}}$).[179] In this case, the endowment vectors of both these sampled nodes and their neighbors will be updated in each epoch in the gradient descent. We set 500 epochs and sample 1,000 nodes as the input of the model in each epoch. In *SI*, we show that our learning converges under this setting.

3.5 Appendix

Data processing and summary statistics

In our study, we use a nationwide mobile phone call dataset involving about 45 thousand (45192) people's phone call logs in 2 years from Jan. 2015 to Dec. 2016. The country is an European country with more than 50 thousand but fewer than 100 thousand citizens. We aggregate the monthly phone call and texting log for each pair of users. Then we take a series of snapshots by aggregating all temporal events happening in a time window. We have flexibility for the time window. We establish

a directed graph including all phone call logs in the time window. As mentioned in the main text, we mainly consider two types of edge weights, interaction frequency and interaction duration. Interaction frequency is the phone call counts between two people, and interaction duration is the sum of call volumes of all phone calls in an interval.

We next discuss how to select the time window. Note that the selection of the time window affects the distribution of tie range. A too narrow time window may result in each snapshot being so sparse that many short-range ties might be treated as long-range ties. A too wide time window may result in too few snapshots for us to analyze the network dynamics. Eventually, we choose a season (three months) as the time window for the main text. Each season or three months are regarded as a "phase".

We also believe monthly or semi-yearly time windows are reasonable choices. We illustrate distributions of the number of nodes and edges of each monthly, seasonal, and half-yearly snapshots induced from the dataset.

As the length of our data is two years, we partition the data into 8 phases (Seasonal snapshots). As the definition of tie range, we classify all connections with respect to tie range in each phase. Due to the small magnitude of ties over range 6, we merge them as ≥ 6. In addition, some ties with infinite tie range cannot be ignored. As illustrated in Tab. 3.1, social ties with a tie range of 5 or ≥ 6 only take a small proportion of all connections.

To test for the robustness of the choice of the time window, we further adjust the interval into a month or a half year. When the time interval is set as a month, we obtain 24 monthly snapshots. We respectively calculate the tie range of each edge in every snapshot. Consistent with the main text, we use the logarithm value of interaction frequency and duration so a few extreme values would not unreasonably affect the averages. Fig. 3-7(a&b), (c&d) present our main results after adjusting the time window. We observe a very similar trend with the results when the time window is four months. These results show that our main results are robust in terms of the time window.

Sensitivity check

Since tie range of an edge is easily impacted by another node or edge that is distant on the network, we need to conduct examine how our results are sensitive to the existence of a few nodes or edges. We examine the sensitivity of our results to the impacts of certain nodes or edges. We randomly drop a proportion (5%) of nodes or edges and then replicate our main result. As shown in Fig. 3-9, dropping either nodes or edges would not affect our main conclusions. This indicates that our results are not sensitive to a few nodes or edges happening to exist on the network.

Lifespan

In the main text, we use interaction frequency and interaction to investigate the dynamics of social ties. Here we use the "lifespan" as the other dimension to measure the dynamics. It is defined as the number of phases for which a pair has any interactions. As shown in Fig. 3-10, there are also U-shapes regarding the relationship between tie range and lasting phases. This result further verify our statement "long ties last longer."

Tie strength heterogeneity

Here we examine our main results by controlling tie strength i.e., interaction duration and frequency. This control reduces the impact of initial tie strength. By categorizing social ties by their strength in phase 1, we find that our main results are robust in each category (see Fig. 3-12) – for each tie strength subgroup, our main conclusion persists.

Degree heterogeneity hypothesis

Here we discuss our "degree heterogeneity" hypothesis. Again, we categorize social ties by degree, and plot the trends for each subgroup in Fig. 3-13. We find that our main results persist in all degree subgroups. Therefore, the degree heterogeneity hypothesis cannot fully explain our main results.

Survival bias hypothesis

To test for this hypothesis, we need to examine whether (1) newly formed long ties are weaker than newly formed short ties in the beginning; and (2) newly formed long ties have a smaller persistence probability such that only very strong long ties survive.

For (1), we plot the trends in Fig. 3-14. We find that for new ties, the tie strength is weakest for those with tie range ≥ 6. By contrast, for existing ties, the trend appears to be a "U-shape." Thus, we support "newly formed long ties are weaker than newly formed short ties in the beginning". For hypothesis (2), we re-conduct the analysis by decomposing the outcome into persistence probability and interaction increments. However, we find that newly formed ties still have the largest persistence probability. Thus (2) is not supported. We therefore believe the survival bias hypothesis cannot fully explain our main results.

Details in learning

Here we provide more technical details regarding the learning process of our proposed model. In our proposed model, we need to learn both hyper-parameter δ and endowments. However, simultaneously training δ and endowment vectors may cause an uncontrollable gradient issue. Therefore, we first try to find the optimal δ and then train endowment vectors by minimizing the loss. From the data, we observe there is a positive indirect effect from common friends, and thus δ should be a small positive value. As shown in Fig. 3-16, we find that the model performs better when we set δ as 0.2 than other options – the fit result $\hat{c}_{ij}$ is closest to the real-world data c_{ij}.

After determining the value of δ, we next infer the endowment vectors. To speed up the learning rate of the model, we adopt a sampling strategy. We set the maximum number of epochs as 500 and randomly sample 1000 nodes in each epoch. According to the loss function, sampled nodes and their neighbors will receive a gradient descent and endowment vectors of them will be updated in each epoch. We set a testing set of 1000 nodes to track the learning curve of the model. As shown in Fig. 3-17, the loss appears to converge to stable after 100 epochs.

Tie Range	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8
2	373,270	338,689	306,481	311,417	253,648	243,471	206,858	204,401
	71.2%	69.2%	68.2%	67.7%	64.3%	63.0%	61.0%	59.7%
3	105,438	102,713	93,316	100,617	88,051	91,261	77485	83,864
	20.1%	21.0%	20.8%	21.9%	22.3%	23.6%	22.9%	24.5%
4	40,729	42,366	43,115	41,968	44,102	43,540	43757	43,405
	7.77%	8.66%	9.59%	9.12%	11.2%	11.3%	12.9%	12.7%
5	4738	5264	6097	5,561	7,971	7636	9973	9727
	0.90%	1.08%	1.36%	1.21%	2.02%	1.98%	2.94%	2.84%
≥ 6	284	255	433	409	686	663	1004	1004
	0.05%	0.05%	0.10%	0.09%	0.17%	0.17%	0.30%	0.28%

Table 3.1: Statistics of ties with different range over 8 phases.

As to the dimension selection of endowment vectors, we investigate how different selection of the dimensions impact our main results. We test it from 2-dimensional to 5-dimensional endowment vectors. Note that a too large dimensionality may raise the issue of computational complexity. We present the results corresponding to Fig. 6 (a) in the main text in Fig. 3-18. As shown in the figure, the conclusions from different dimensions are largely similar. We therefore choose the dimensionality of four as an illustration in the main text.

We implemented our algorithm in PyTorch. The endowment vectors are implemented as embeddings in PyTorch, and we use Adam optimizer with regularization for the optimization.

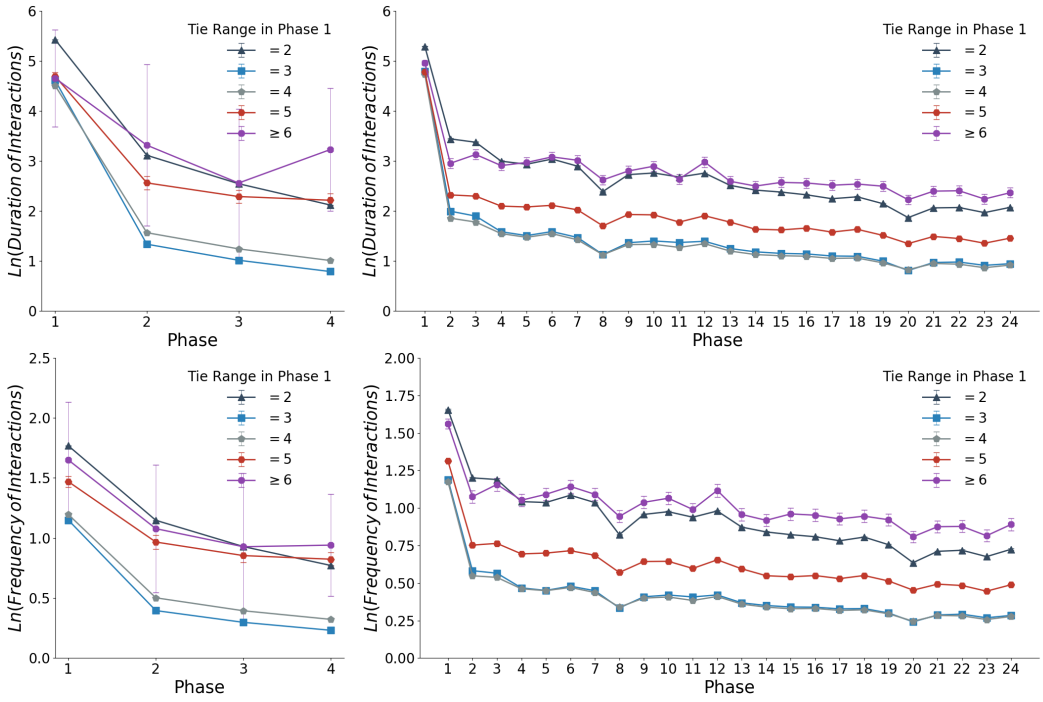


Figure 3-7: Evolution of both interaction frequency and interaction duration of ties throughout the four semiyearly (a&c) and twenty-four monthly (b&d) snapshots. All ties are classified according to their tie range in the first phase. Error bars are 95% confidence intervals.

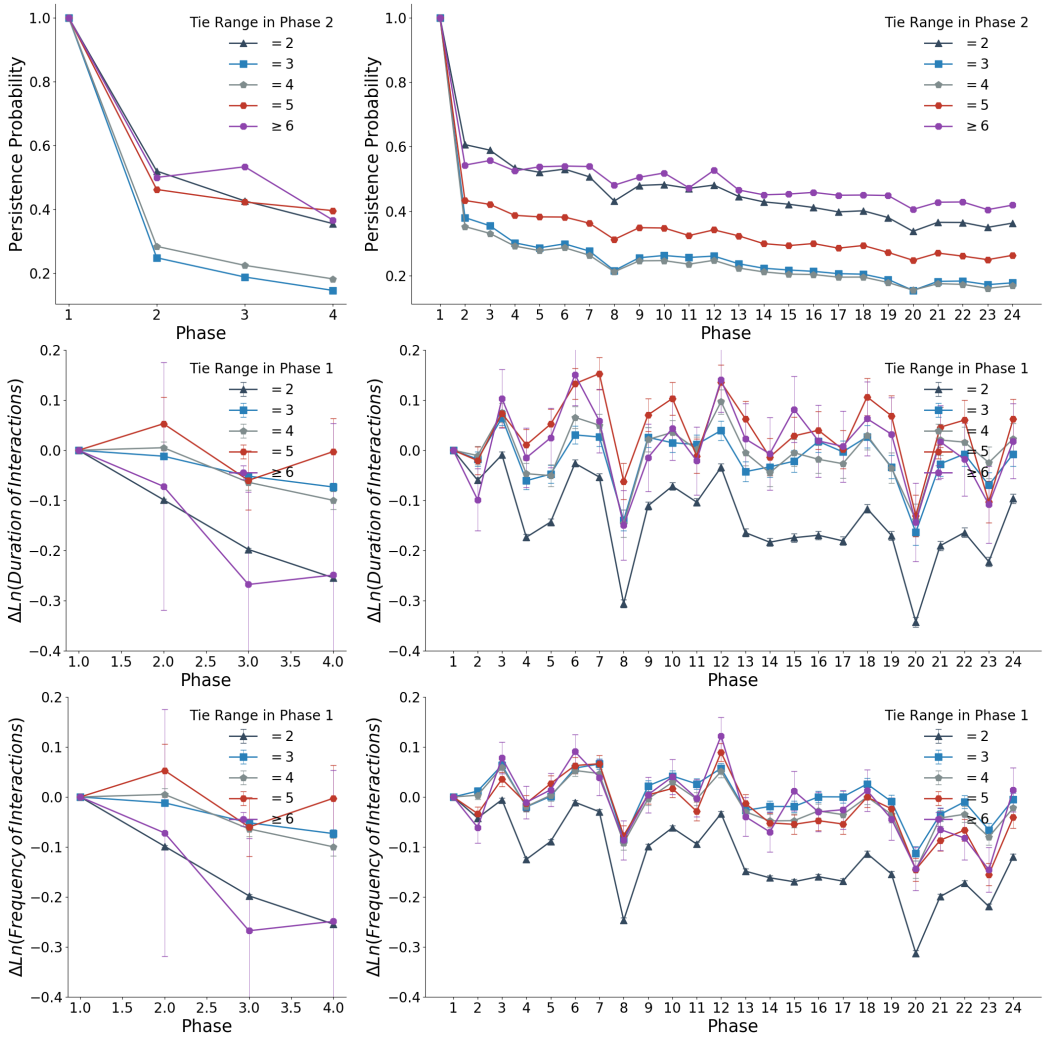


Figure 3-8: Dynamics of persistence probability and interaction increments conditional on the tie range in phase 1. Either a month (b,d&f) or a semi-year (a,c&e) is set as the time window. Error bars are 95% confidence intervals.

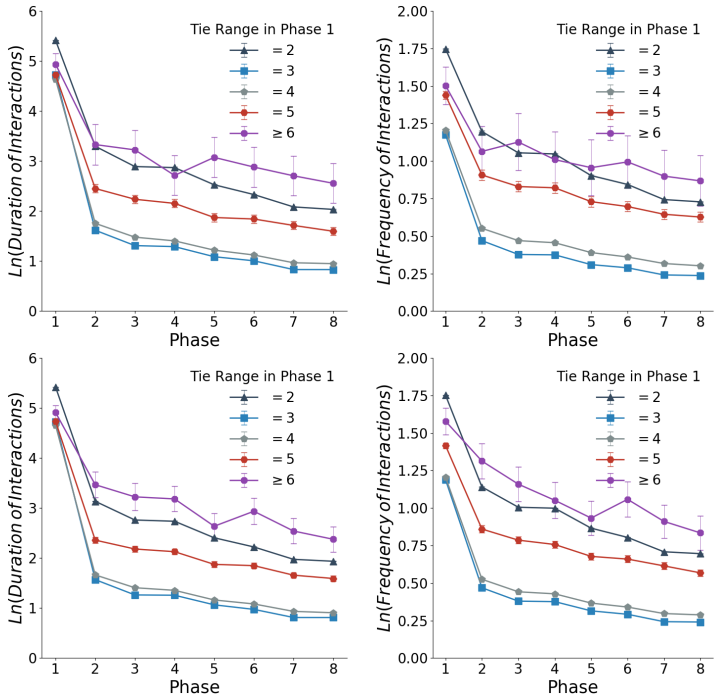


Figure 3-9: Sensitivity check by randomly dropping a proportion (5%) of nodes or edges.

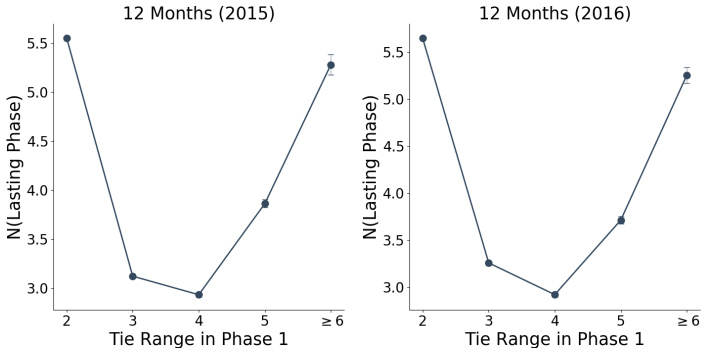


Figure 3-10: Lifespans (or the number of lasting phases) of ties with different tie range. We examine the two years separately.

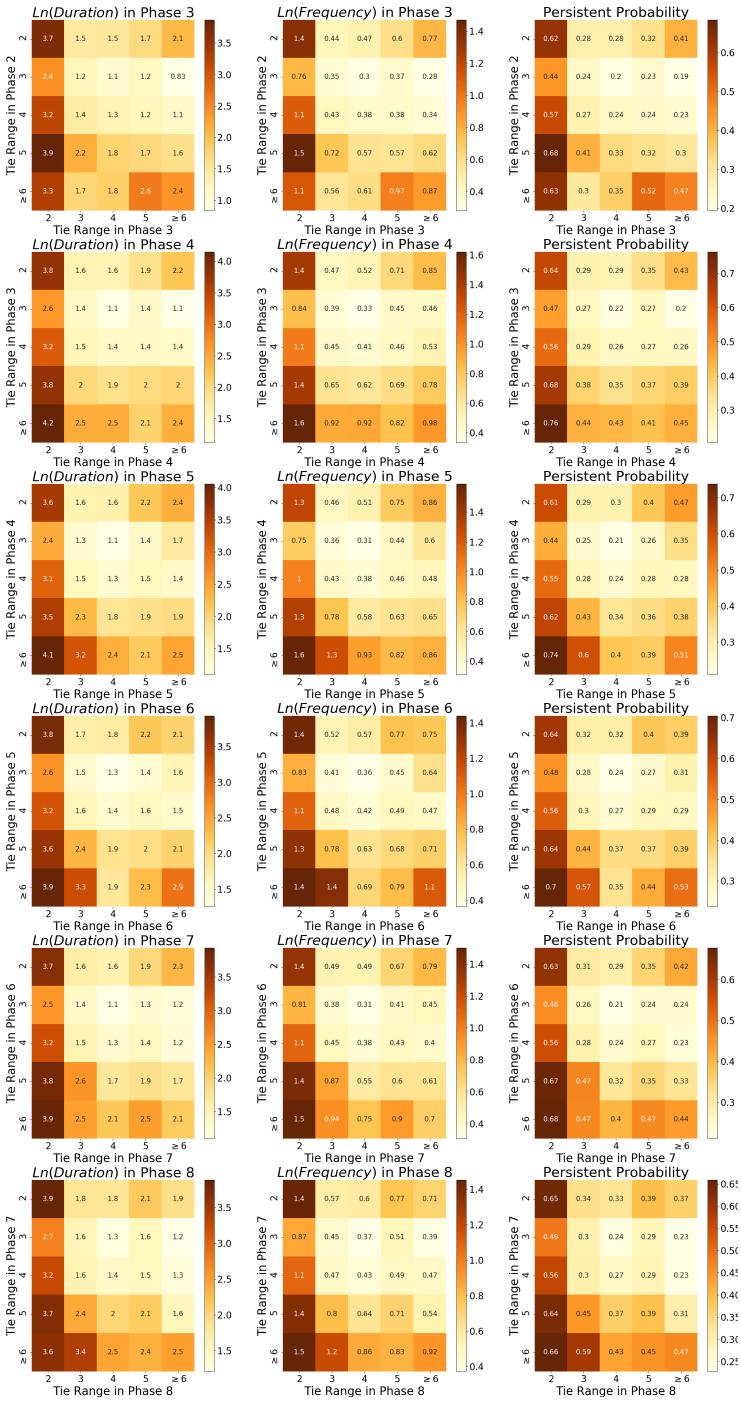


Figure 3-11: Interaction duration (a), frequency (b) and persistent probability (c) in the next phase when tie range evolves. (phase t vs phase $t+1$)

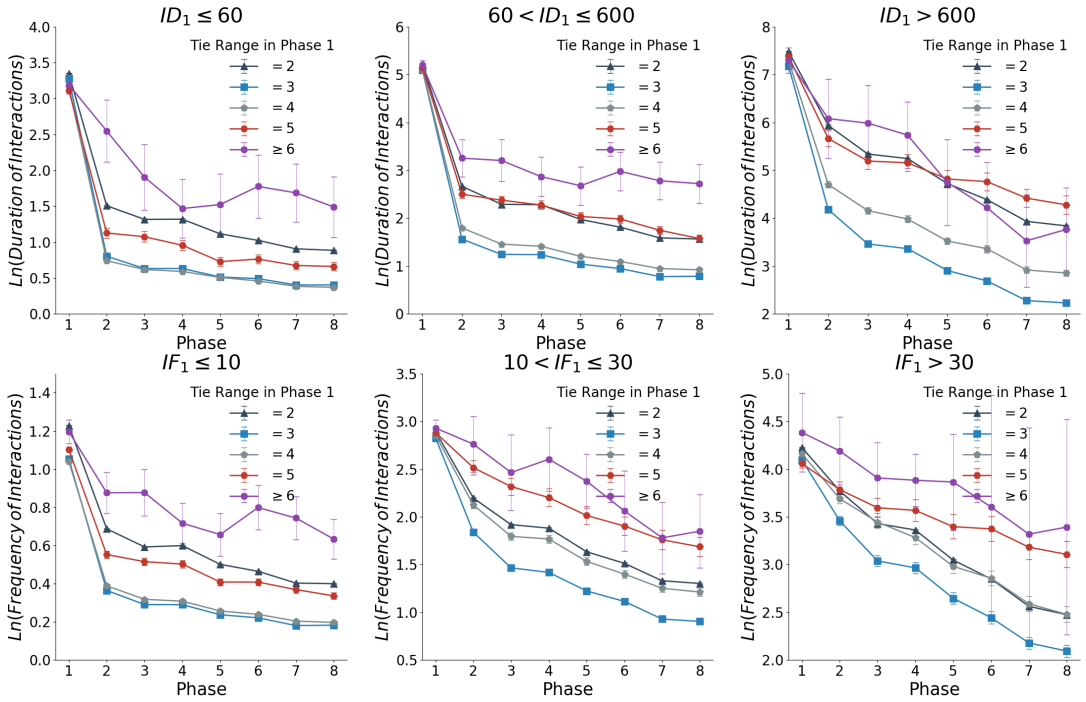


Figure 3-12: Evolution of ties with different ranges when we examine separately subgroups categorized by interaction duration (a-c) and frequency (d-f) in Phase 1 .

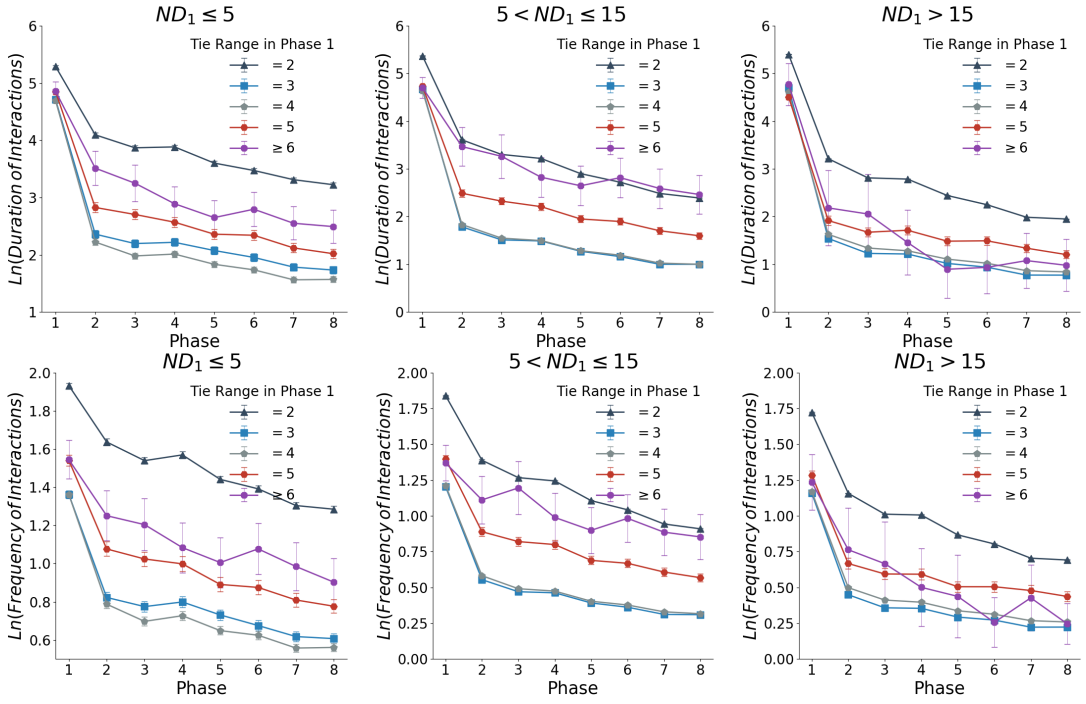


Figure 3-13: Evolution of interaction duration (a-c) and frequency (d-f) of ties with different ranges when we examine degree subgroups. ND indicates node degree. The medium node degree of phone call network in phase 1 is 12.

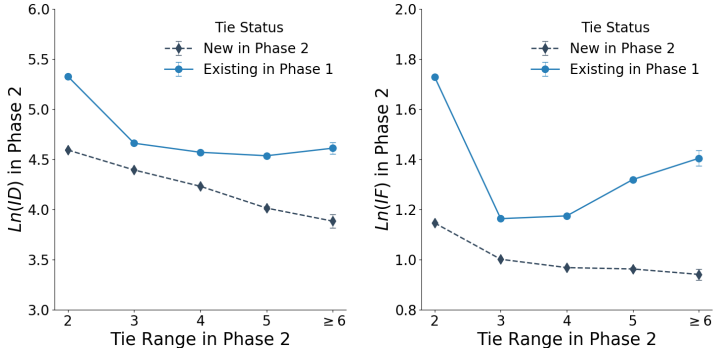


Figure 3-14: Interaction duration (a) and frequency (b) for newly formed ties and existing ties.

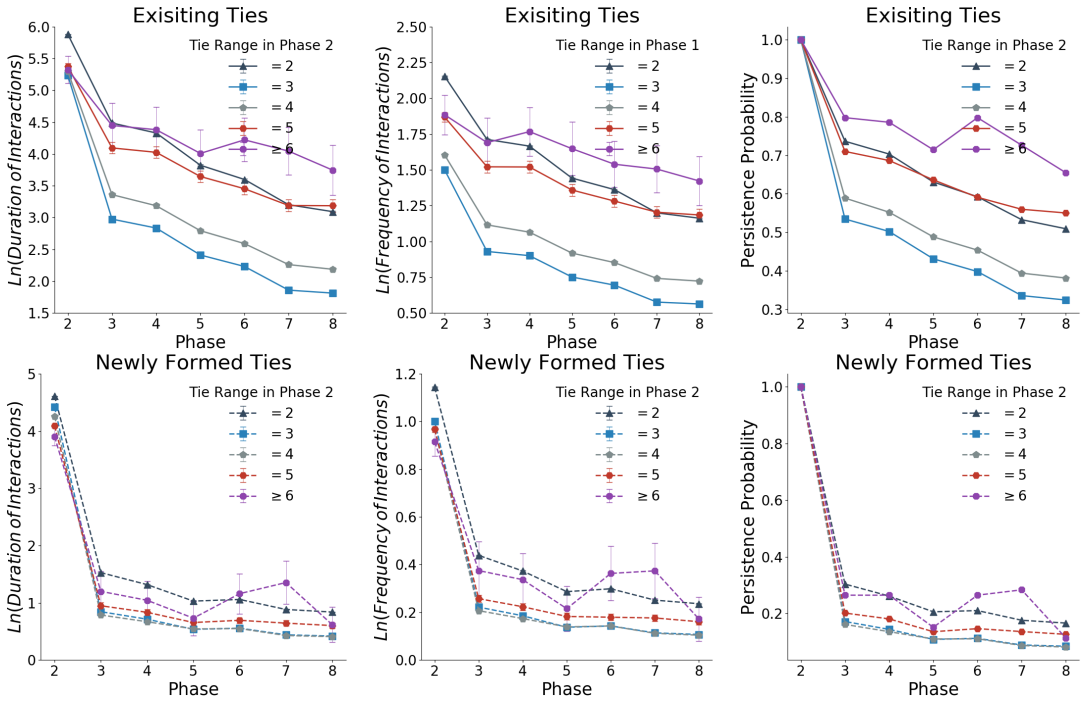


Figure 3-15: Dynamics of interaction frequency, interaction duration and persistent probability of survival (a-c) or newly-formed (d-f) ties throughout the next seven seasonal snapshots conditional on the tie range in phase 2. Error bars are 95% confidence intervals.

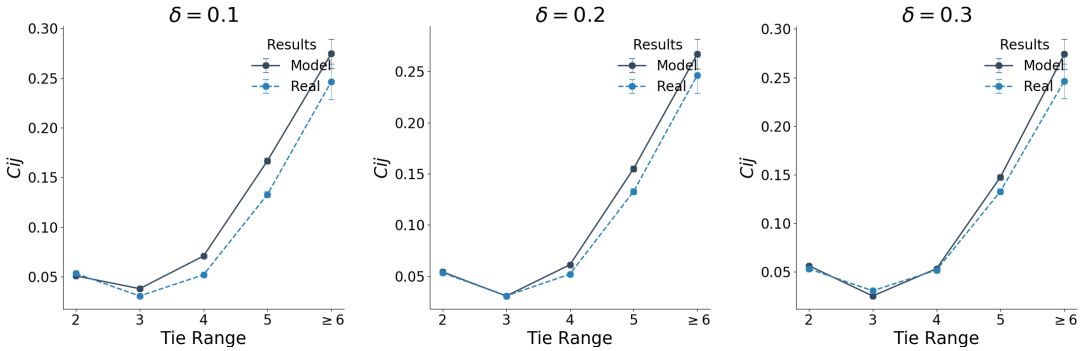


Figure 3-16: Choice of δ.

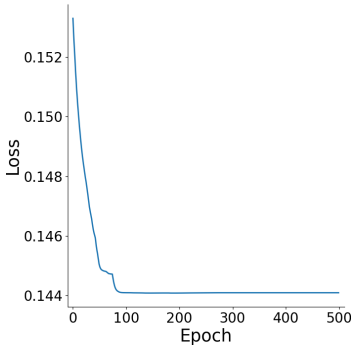


Figure 3-17: The learning curve.

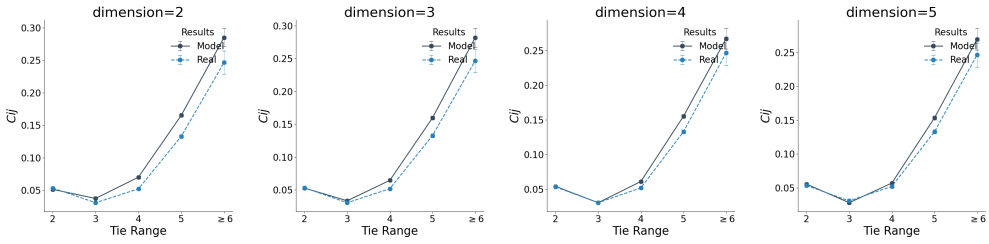


Figure 3-18: Results of choosing different dimensionality.

Chapter 4

Identifying gift contagion with structural causal models

4.1 Background

Individuals belong to many different social groups: kinship groups, friend groups, work groups or organizations, and interest groups. The collective identities developed in these groups deeply shape the behavior of their members [237, 69, 67, 62, 21, 47]. Nowadays, social groups are facilitated through digital platforms, especially social network platforms such as Facebook, Line, WeChat, and WhatsApp. These platforms support social groups for a variety of purposes, including relationship maintenance, opinion and information exchange, and event planning [252, 28, 193, 43, 166]. In particular, during the COVID-19 pandemic, online work group chats have substituted for conventional in-person meetings [49]; indeed, it was reported that 42 percent of the U.S. labor force worked from home full-time as of June 2020.[1] In China, WeChat groups are widely used for instant work-related communication [166, 201]. Although online work groups offer the convenience of long-distance communication and coordination, they may face challenges related to team building and group solidarity [124].

One way to promote group bonding is through the use of group gifts, which are the gifts sent by a group member without specifying recipients. Examples of group

[1] `https://news.stanford.edu/2020/06/29/snapshot-new-working-home-economy/`

gifts include the food items or souvenirs bought by a member to her work group after traveling abroad as well as the small gifts being exchanged at a Christmas or holiday party (the "white elephant" gift exchange). While prior literature focuses on one-to-one gifts and their role in creating interpersonal social bonds [173], few studies have investigated group gifts and their role in promoting in-group interactions and solidarity.

In this study, we observe the outbreak of sending group gifts in both online and offline settings, such as the "red packet rain" in online groups, indicating the presence of gift contagion (the social contagion of gift giving).[2] Social contagion is defined as "an event in which a recipient's behavior has changed to become 'more like' that of the actor" [261]. [18] has pointed to the importance of identifying causal effects in the process of social contagion. In the context of group gifts, gift contagion implies that people who receive larger amounts of gifts feel promoted to increase their own subsequent contributions. If gift contagion exists in groups, the actual impact of a given gift would be amplified, leading to stronger social bonds and feelings of group solidarity [169].

To quantify the effect of gift contagion, our study employs a large-scale dataset of 3.4 million users in a major Chinese social network platform. On the platform, users send online red packets to each other as a type of digital monetary gift. Online red packets, especially group red packets, swiftly became extremely popular after being released: more than 700 million people engaged in sending or receiving red packets during one week in 2019 [54, 264].

Methodologically, the causal identification of social contagion in observational data is notoriously challenging [18, 221]. In particular, the following two confounding factors may hinder valid causal identification of gift contagion. The first confounding is the "temporal clustering." Specifically, group members may send gifts within a short time period independently to celebrate a festival or an event [18]. The second is homophily, the phenomenon whereby individuals tend to befriend similar others [177]. For example, wealthy people tend to cluster in an online group and send larger

[2] https://digital.hbs.edu/platform-rctom/submission/making-it-rainchinese-red-packets/

amounts of gifts to each other.

We leverage a natural experiment to overcome the above challenges. Our natural experiment is enabled by a random gift amount allocation algorithm for group red packets. The algorithm splits a red packet into several shares and randomly determines the amount of each share. We utilize this random assignment of received cash amounts to identify the impact of the amount received on a participant's subsequent gifting behavior. We first examine the presence of gift contagion in online groups. On average, receiving one additional dollar causes a recipient to send 18 cents back to the group within the subsequent 24 hours. Moreover, we find that this overall effect is mainly driven by the extensive margin, i.e., receiving red packets significantly promotes the likelihood of giving. Second, we investigate heterogeneity in the effect size of gift contagion across different time periods (festival versus non-festival periods), and across different types of groups (e.g., relative versus classmate groups). Third, our analysis suggests that a social norm exists in that the luckiest draw recipient should send the very first subsequent red packet. Finally, we find that both individual-level network position and group-level network structure affect the strength of gift contagion, respectively.

4.1.1 Online red packets

In our study, we focus on the custom of sending monetary gifts to family members, friends, and other acquaintances in what is known as "red packets" (also known as "red envelopes," or "lucky money"). Red packets are typically sent to others as a way of commemorating festivals or important events. They also function as a means of tightening the social network in East and Southeast Asian cultures, known as the *renqing* and *guanxi* system [168, 256]. We summarize the history of red packets in Appendix.

The Chinese social network platform enables users to designate private contacts (we use the term "friends" throughout the paper) and to create group chats. Online groups are created for a wide variety of purposes, ranging from family members to coworkers and friends. The number of group members ranges between 3 and 500.

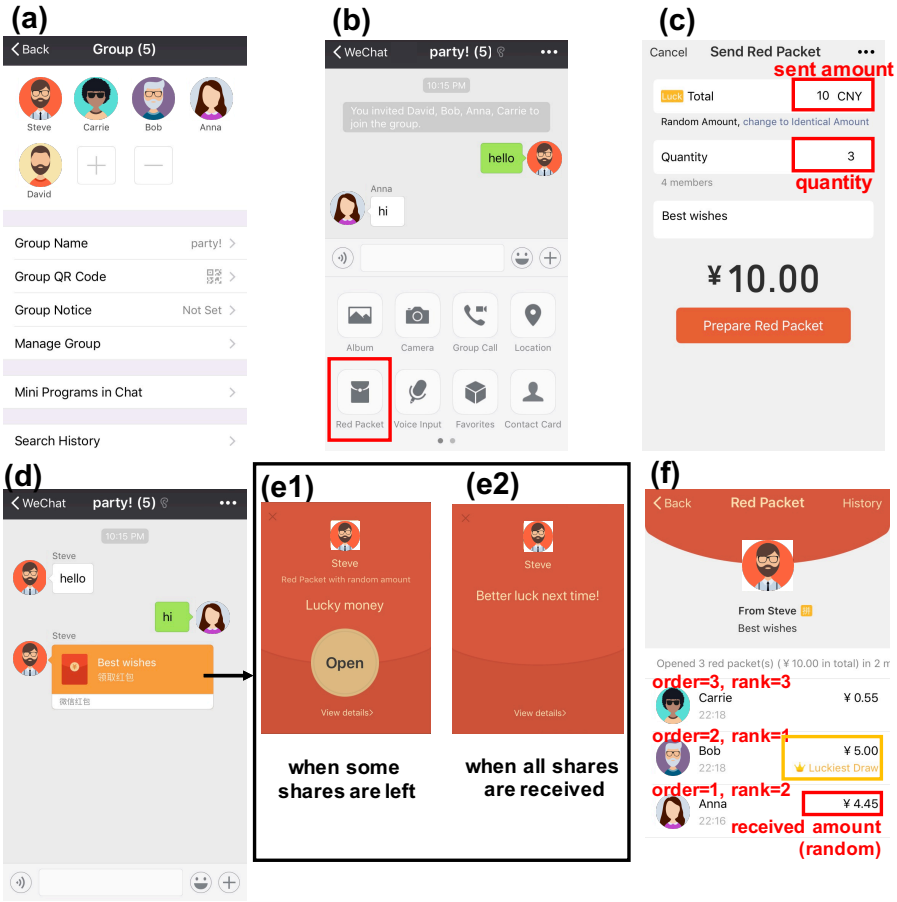


Figure 4-1: Illustration of group red packets with random amounts

The platform introduced its online red packet feature in 2014, allowing users to send monetary gifts to either a friend or a group. WeChat red packets were popularized during the Lunar New Year of 2015 — it was reported that 55% of the Chinese population sent and received red packets on that single day. Because of the popularity of red packets, WeChat also benefits from a rapid growth in its mobile payment market share. For example, in 2015, Tencent's mobile payments share grew from 11% to 20%, while Alipay's share dropped from 82% to 68%.[3] In 2019, its market share is 39.5% while Alipay is 54.5%.[4]

Noting that the platform allows two types of group red packets, our study focuses on the most popular one – the random-amount, group-designated red packet, as depicted in Figure 4-1.[5] In this example, Panel (a) provides the basic information about the group, the name of which is "Party!", with five group members in total. Panel (b) presents the user interface of this group, from which a group member can click the "Red Packet" button to send a red packet. Panel (c) shows that a group member (Steve in this example) can choose both the *total amount* of the red packet that he would like to send ("Total"=10 CNY) and the *number of recipients* ("Quantity"=3). Panel (d) is the interface for the red packet notification, from which a user can click the orange button to choose to receive the packet. Panel (e1) pops up when some shares of the red packet remain. In this example, only the first three users who click the "Open" button can receive a share of this red packet. Panel (e2) pops up when all shares of the red packet have been received by group members. Finally, Panel (f) shows the recipient list, which can be viewed by clicking "View details" in Panel (e). All of the group members, including senders, recipients, and non-recipients, can view the recipient list and see the amount obtained by each recipient. We define the *order of receiving time* as group members' respective places in the order at the time when they receive the red packet. In the above setting, the amount that each user receives is randomly assigned by the platform, and is a function of the total amount of the red packet, the number of recipients, and the order of the receiving

[3] `https://www.pymnts.com/news/b2b-payments/2016/barclays-corporate-banking-head-chief-resignation/`
[4] `https://www.businessofapps.com/data/wechat-statistics/`
[5] The other type is that senders can also choose to split red packets equally.

time. Moreover, the platform designates which user receives the largest amount of a red packet with a "Luckiest Draw" icon and the corresponding yellow text. All group members can observe who is the luckiest draw recipient.

4.2 Methods

4.2.1 Data collection

We collect a dataset consisting of randomly-sampled online groups with red packet activity from October 1, 2015 to February 29, 2016. To protect user privacy, users' identities were anonymized before we accessed the data. To avoid data sparsity, we restrict our analysis to online groups in which the number of red packets sent is at least three times the number of group members. We also filter out groups that might be used for online gambling based on the following criteria: (1) a number of red packets greater than 50 times the number of group members; (2) a name that suggests a gambling focus (containing words such as 元/块(Chinese yuan), 发(send), 红(red), 包(packet), 最佳(luckiest), 抢(grad), 赌(gamble), 钱(money), 福利(welfare), and 接龙(chain) or Arabic numerals (which indicate the default packet amount set for gambling); or (3) no designated group name, which could also be temporarily created for gambling.[6] In total, this selection process results in 174,131 groups with 3,466,928 group members (3,450,540 unique users).

In our main analyses, we include: (1) the characteristics of 174,131 online groups, including the number of group members, the total number of red packets, and the total cash value of the red packets; (2) 3,450,540 unique users in these online groups, along with their characteristics, such as the number of in-group friends are also retrieved; and (3) the attributes of each red packet, including the cash amount, the corresponding recipients and the opening time. In total, our sample comprises 36,608,864 red packets. Furthermore, we focus on recipients of "spontaneous red packets," which indicate that no group red packet is sent in the 24 hours prior to this type of red

[6]We show that groups identified as gambling groups appear to exhibit greater levels of gift contagion Appendix

packet. We conduct robustness checks by varying the time window and find that our main results remain (Appendix). In total, we identify 1,549,720 spontaneous red packets sent to 7,266,446 recipients.[7] Each observation refers to a user's received red packet and we have 7,266,446 observations in total. In Appendix we present a detailed description of the data.

4.2.2 Estimation strategy

Random assignment algorithm

Here we illustrate the random assignment algorithm for red packet amounts on the platform. First, the sender determines the total amount of the red packet ($a > 0$) and the number of recipients to receive a portion of the red packet ($n \geq 1$).[8] Then group members choose to open the red packet on a first-come, first-served basis. They do not know the values of a and n until they open the red packet. Let o denote the order of receiving time ($o = 1, 2, ..., n$). The amount received by the recipient with order o, denoted by V_o, is determined by the following algorithm:

1. When $o = 1$ and $o < n$: the amount obtained by the first recipient (order = 1) follows a uniform distribution on $(0, \frac{2a}{n}]$. The expected amount is:

$$\mathbb{E}[V_1] = \frac{1}{2} \times (0 + \frac{2a}{n}) = \frac{a}{n}.$$

   When $o = n = 1$, the amount received is a because the only recipient should take all the cash amount.

2. When $1 < o < n$: the amount received follows a uniform distribution on (0,

[7] We exclude observations in which the sender clicks the red packet and receives a share of her own red packet.

[8] In practice, the amount received is rounded to the nearest cent, and is set at least 0.01 CNY.

$\frac{2(a-V_1-...-V_{o-1})}{n-o+1}]$. We have:

$$\begin{aligned}\mathbb{E}[V_o] &= \mathbb{E}\Big[\mathbb{E}[V_o|V_1,...,V_{o-1}]\Big] \\ &= \mathbb{E}\left[\frac{1}{2}\times\Big(0+\frac{2(a-V_1-...-V_{o-1})}{n-o+1}\Big)\right] \\ &= \frac{a-\mathbb{E}[V_1]-...-\mathbb{E}[V_{o-1}]}{n-o+1}.\end{aligned}$$

We show that $\mathbb{E}[V_o]=\frac{a}{n}$ by induction:

First, we have already shown that $\mathbb{E}[V_1]=\frac{a}{n}$.

Second, assuming that we have $\mathbb{E}[V_{o'}]=\frac{a}{n}$ for all $o'<o$, we have $\mathbb{E}[V_o]=\frac{a-(o-1)\frac{a}{n}}{n-o+1}=\frac{a}{n}$.

3. When $o=n$: $V_o=a-V_1-...-V_{o-1}$, indicating that the last recipient takes the surplus. Then we have $\mathbb{E}[V_o]=a-\mathbb{E}[V_1]-...-\mathbb{E}[V_{o-1}]=\frac{a}{n}$.

Therefore, the expectation of the received amount is the same: $\frac{a}{n}$. However, the variance in the amounts is not always the same. For example, when $n>2$,

$$\begin{aligned}\mathrm{Var}(V_1) &= \frac{1}{12}\times(\frac{2a}{n}-0)^2=\frac{a^2}{3n^2}; \\ \mathrm{Var}(V_2) &= \mathbb{E}\Big[\mathrm{Var}(V_2|V_1)\Big]+\mathrm{Var}\Big(\mathbb{E}[V_2|V_1]\Big) \\ &= \mathbb{E}\Big[\frac{1}{12}\times\big(\frac{2(a-V_1)}{n-1}\big)^2\Big]+\mathrm{Var}\Big(\frac{a-V_1}{n-1}\Big) \\ &= \mathbb{E}\Big[\frac{(a-V_1)^2}{3(n-1)^2}\Big]+\frac{1}{(n-1)^2}\mathrm{Var}(V_1) \\ &= \Big(-\frac{(a-\frac{2a}{n})^3}{9(n-1)^2}+\frac{a^3}{9(n-1)^2}\Big)\times\frac{n}{2a}+\frac{a^2}{3(n-1)^2n^2} \\ &= \frac{a^2}{3n^2}+\frac{4a^2}{9(n-1)^2n^2}>\mathrm{Var}(V_1).\end{aligned}$$

In addition, we provide the complete proof for variance differences in Appendix.

To show that the random assignment algorithm functions as described above, we compare the empirical distributions of received amounts from our data to the simulation results generated by the algorithm in Figure 4-2. In our first comparison

example in the upper two rows, we see that the total amount is 10 CNY and the number of recipients is 5 (108,560 observations). In our second comparison example in the bottom two rows, we see that the total amount is 5 CNY and the number of recipients is 3 (38,523 observations). We do not find significant differences between these two distributions generated by the simulation and our empirical data (p = 0.30 and 0.36 for the two cases, respectively, two-sided Kolmogorov-Smirnov tests). Additionally, consistent with the random assignment algorithm, the expectation of the amount received is solely determined by the total amount of the gift and the number of recipients ($\frac{10}{5} = 2$ and $\frac{5}{3}$ for the two cases, respectively). Examining the remaining data, we find that the results continue to hold.

Furthermore, we verify the randomization procedure and provide the results in Appendix. The results suggest that, conditional on the total amount of the red packet, the number of recipients, and the order of receiving time, the amount received is not significantly correlated with individual characteristics or historical behaviors. Altogether, these analyses confirm that the amount that a recipient obtains is solely determined by the following three variables: (1) the total amount of the red packet; (2) the number of recipients; and (3) the order of receiving time. This verification enables us to use the following empirical strategy to quantify the causal impact of gift contagion.

Empirical strategy

We next discuss our empirical strategy, which is used to quantify the impact of the amount received on the recipient's subsequent gifting behavior. We regard the random assignment of received amounts as a stratified randomized experiment [140, 129, 132, 26], where a stratum is uniquely determined by the total amount of the red packet, the number of recipients, and the order of receiving time. We apply the empirical strategy of stratified randomized experiments proposed by [132] and conduct the following regression analyses:

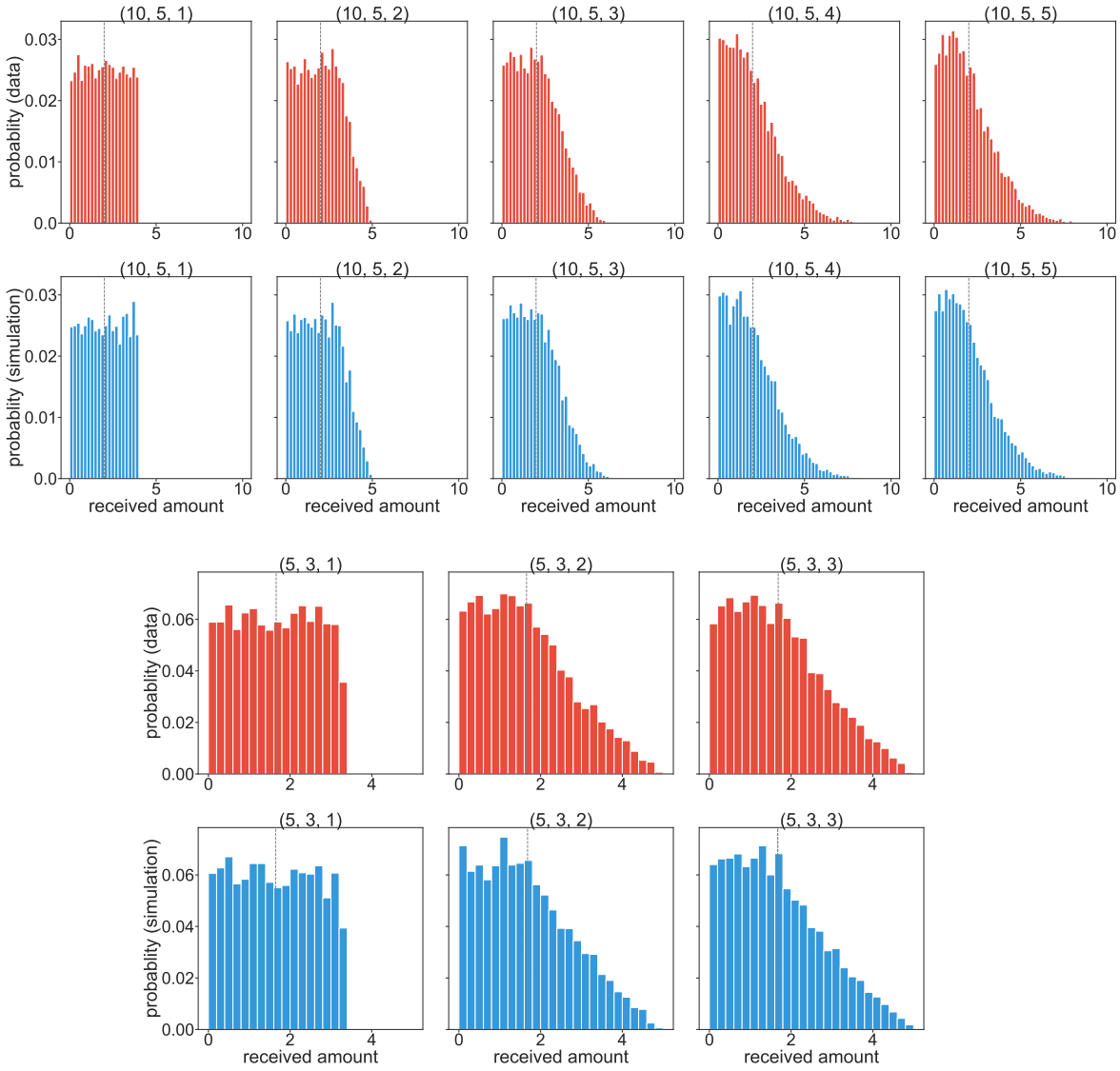


Figure 4-2: Distributions of received amounts in our dataset (red) and simulation (blue). The top two rows are those with 10 CNY and 5 recipients; The bottom two rows are those with 5 CNY and 3 recipients. A title with (a, n, o) indicates that the total amount of the gift is a CNY, the number of recipients is n, and the order of receiving time is o.

$$Y_{gir} = \beta T_{gir} + \sum_{s} \gamma_s B_s(A_r, N_r, O_{ir}) + \epsilon_{gir}. \tag{4.1}$$

In Equation (4.1), g denotes an online group, and i denotes a unique user who receives a share of a red packet r. ϵ_{gir} represents the random noise. The dependent

variable Y_{gir} is the amount sent by the recipient i in the time interval after receiving a red packet. The selected time intervals are 10 minutes, 1 hour, 3 hours, 6 hours, 12 hours, and 24 hours. The main independent variable T_{gir} is the amount received by user i from red packet r. β is the estimand of interest that specifies the linear relationship between T_{gir} and Y_{gir} and measures the degree of gift contagion. A_r, N_r, and O_{ir} refer to the total amount of the red packet r, its number of recipients, and user i's order of receiving time, respectively. Finally, $B_s(A_r, N_r, O_{ir})$ is a dummy variable indicating whether the value of $X_{gir} = (A_r, N_r, O_{ir})$ belongs to the sth stratum. The dummy variable helps control for stratum fixed effects. In total, we have 180,578 strata in our sample.

To address potential data interdependence, we focus on group- versus user-level interdependence, as only 3.1% of the users in our dataset belong to more than one group, data interdependence at the group level is the primary concern. To further address user-level interdependence, our bootstrap identifies any two groups containing the same user as a "cluster." We use the Poisson bootstrap [86] at the "cluster" level for 1,000 replicates to estimate the robust standard errors or 95% confidence intervals.

To depict the causal relationship examined by our empirical strategy, we use Pearl's directed acyclic graphs (DAGs) to visualize the causal relationship in our empirical strategy [196]. As shown in Figure 4-3, controlling for X blocks all of the "backdoor" paths from T to Y, which satisfies the backdoor criterion and allows us to identify the causal impact of T on Y. This process provides greater confidence that confounding factors (U), such as temporal clustering and homophily, would not bias our estimation.

This empirical strategy has two advantages in identifying a causal relationship. First, it enables us to fully control for the stratum fixed effect, without requiring a specific functional form for the impact of X. For example, a linear specification, i.e., adding A_r, N_r, and O_{ir} directly into the regression, would lead to an overestimated treatment effect (Appendix). Second, we realize that if most strata have few observations, we may fail to measure such within-stratum effects. Fortunately, our sample size is sufficiently large that we have a sufficient number of observations within each

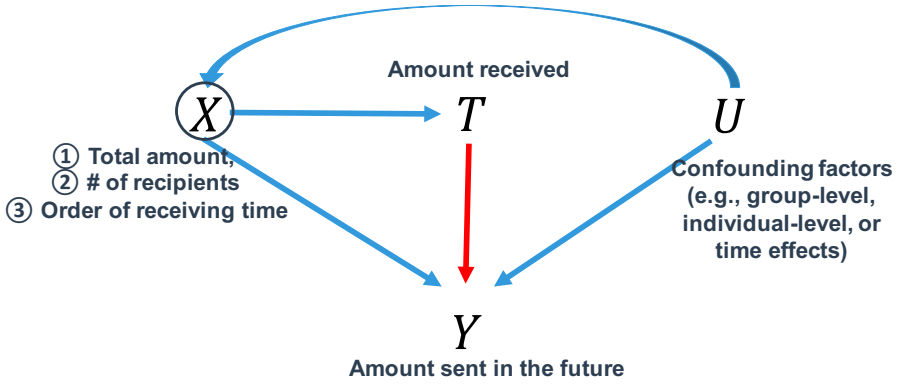


Figure 4-3: A directed acyclic graph illustrating the causal relationship

stratum. Note that the average number of observations in a stratum is 8.37.

4.3 Hypotheses

Both observational and experimental studies provide a wide body of support for the idea that various behaviors are found to be socially contagious [67, 96, 21, 47, 150, 19, 146]. In our setting, the phenomenon of gift contagion may be driven by multiple motivations. For example, reciprocity, including direct and indirect, can promote social contagion [9, 187, 188, 219]. Since any group members can receive a share of red packets in an online group, we conjecture that indirect reciprocity should play a more important role than direct reciprocity in our sample.[9] Furthermore, fairness concerns, or inequity aversion may also promote gift contagion [90, 45], as the random assignment of gift amounts generates an unequal distribution of cash amounts between recipients. Consequently, this perceived lack of fairness in distribution may motivate users who obtain larger amounts to increase their subsequent contributions in order to mitigate the inequity. Based on this conjecture, we form our first hypothesis:

Hypothesis 1 (Gift contagion) *The larger the amount a recipient obtains, the larger the amount she will send to the group.*

Next, we are interested in understanding whether the degree of gift contagion varies between different time periods. In East and Southeast Asia, red packets are

[9]This conjecture is further supported by our analysis in Appendix.

usually sent during the Lunar New Year and other festivals [222, 256]. We thus posit that gift contagion is stronger during festivals.

Hypothesis 2 (Festival effect) *Gift contagion is stronger during festivals than other time periods.*

In addition, we expect the degree of gift contagion to be contingent on group types. Red packets are historically sent among relatives (see Appendix for more details). We thus posit that we will observe more gift contagion in groups of relatives.

Hypothesis 3 (Group type effect) *Gift contagion is stronger in groups of relatives than other groups.*

Recall that the user interface highlights who is the luckiest draw recipient, and this information is observed by all group members. We thus posit that gift contagion is stronger for luckiest draw recipients. There are two possible reasons for this conjecture. The first one is the *amount effect*: luckiest draw recipients receive larger amounts and thus they may send a larger amount. The second is that the salience of the luckiest draw recipient information may motivate luckiest draw recipients to send red packets (referred to as the *luckiest draw effect*). Therefore, we propose the following hypothesis.

Hypothesis 4 (Luckiest draw effect) *Gift contagion is stronger for luckiest draw recipients than others.*

Finally, we are interested in the moderating effect of social network characteristics. There has been extensive literature on the effect of individual network positions on social contagion [258, 53, 133, 21, 247, 47, 32, 19]. For example, [21] conduct a large-scale randomized field experiment to examine Facebook users' product adoption decisions. They find that individuals who are susceptible to influence are less clustered on networks. In another study, [247] examine how individual positions on e-mail networks impact the adoption of a new product, and find that individuals who are less clustered on a network are more inclined to adopt a new product.

Furthermore, the number of friends (i.e., degree) may also affect the degree of gift contagion. Both the classic linear threshold model [139] and complex contagion theory [57] suggest that, regarding the adoption of new products or behaviors, people need to see more than one neighbor who has already adopted the product or behavior before being influenced to do likewise. An experiment examining information diffusion on Facebook also indicates that the probability of sharing information depends on the number of friends who already shared this information [31]. In our sample, those with more in-group friends generally have more friends who have adopted the gifting behavior, and thus may be more likely to be influenced. This leads to the following hypothesis:

Hypothesis 5 (Individual network position on gift contagion)

(a) Gift contagion is stronger for individuals who are less clustered in the group.

(b) Gift contagion is stronger for individuals who have more friends (higher degree) in the group.

In addition to individual-level network characteristics, the group-level network structure may also affect the degree of gift contagion as well [110, 258, 53, 56]. For example, the weak tie theory suggests that the existence of long-range ties helps the spread of social contagion; thus, social contagion is expected to be stronger in networks that are not tightly knit [110, 258, 53]. By contrast, [56] shows that a more clustered network spreads behavior more quickly because the adoption of a new behavior requires reinforcement from multiple influential network neighbors. Moreover, a tightly knit group network structure may indicates closer relationships within the group. Therefore, we are agnostic regarding the degree of gift contagion between groups with different group-level network structures.

Hypothesis 6 (Group network structure on gift contagion) *The strength of gift contagion differs between tightly and loosely knit groups.*

4.4 Results

4.4.1 Gift contagion in online groups

We first apply a simplistic, non-parametric approach to shed light on the causal effect of the amount received by a user within a group on the probability of that user sending the first subsequent red packet. We depict the probability of sending the first subsequent red packet for the recipients of a given red packet in Figure 4-4. From this figure, we see a decreasing trend related to the rank of received amounts: those who receive the largest amount have the highest probability of sending the first subsequent red packet. Moreover, the largest difference lies between those who receive the largest amount and those who receive the second largest amount, while the differences between other recipients are much smaller.

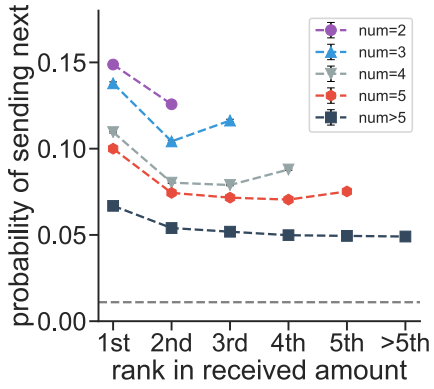


Figure 4-4: The recipients' probability of sending the first subsequent red packet. "Num" is the number of recipients of a given red packet. The x-axis indicates the rank of received amounts among recipients. For example, "1st" refers to the user who receives the largest amount, i.e., the luckiest draw recipient. ">5th" is the average probability among recipients whose rank is below the 5th position. The dashed gray line represents the average probability that a non-recipient sends the first subsequent red packet. The error bars, i.e., the 95% CIs, are much smaller than the markers, and become invisible.

Next, we apply the empirical strategy described in Section 4.2.2 to quantify the impact of the amount received on the subsequent amount sent, namely, we estimate

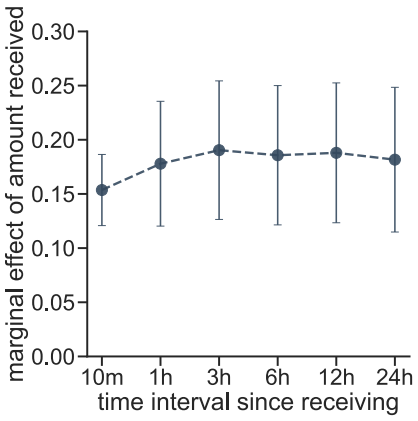


Figure 4-5: The marginal effects of the amount received on the amount sent within the corresponding timeframes. Error bars are the 95% CIs.

β in Equation (4.1). Figure 4-5 presents the marginal effects for different timeframes. From the figure, we see an increase in the effect size as the timeframe widens, with the effect stabilizing after the first three hours. In later analyses, we focus on regression results for 10 minutes and 24 hours, respectively.

Table 4.1: Regression analyses for gift contagion

	Overall		Extensive		Intensive	
	10 min	24 h	10 min	24 h	10 min	24 h
	(1)	(2)	(3)	(4)	(5)	(6)
Amount received	0.1559***	0.1853***	0.0031***	0.0032***	0.0202	-0.2284*
	(0.0166)	(0.0343)	(0.0001)	(0.0001)	(0.0788)	(0.1471)
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	7,266,446	7,266,446	7,266,446	7,266,446	1,060,746	1,370,741
Adjusted R^2	0.0394	0.0396	0.0211	0.0233	0.1826	0.1391

Note: The dependent variable (DV) for Columns (1) and (2) is the amount sent within the respective timeframe. It is coded as zero for those who do not send red packets. The DV in Columns (3) and (4) is the dummy variable for sending red packets. The DV in Columns (5) and (6) is the amount conditioning on sending red packets. Marginal effects are reported. Standard errors clustered at the group- and user-level are in parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

Result 1 (Gift contagion) *The larger the amount a recipient obtains, the larger the amount she sends to the group.*

Support. *As shown in Columns (1) and (2) in Table 4.1, the regression coefficients for the amount received are positive and significant at the 1% level (10 minute: 0.1559,* $p < 0.01$*; 24-hour: 0.1853,* $p < 0.01$*).*

By Result 1, we reject the null hypothesis in favor of Hypothesis 1. Since the gifting behavior is quantitatively measurable in our dataset, we are able to decompose the overall effect to the extensive and intensive margins. This decomposition has received little examination in the previous gift contagion literature. For example, in their study of gift contagion on Facebook, [146] reports only the overall effect of gift contagion. The extensive margin reflects whether the amount received increases the recipient's likelihood of sending a red packet, while the intensive margin indicates, conditional on sending a red packet, whether the amount received affects the amount sent. As shown in Columns (3) and (4), receiving one more CNY increases the recipient's probability of sending a red packet by 0.31% in 10 minutes ($p < 0.01$) and 0.32% in 24 hours ($p < 0.01$). By contrast, the intensive margin is not significant within 10 minutes ($p > 0.1$), and even becomes negative within 24 hours ($p = 0.061$). Therefore, we conclude that the primary driver of our observed overall effect of receiving a red packet on subsequent behavior is that users are more likely to send packets versus more likely to send a greater amount.

In addition, we test for generalized reciprocity [266, 186], i.e., whether receiving gifts in one group triggers the recipient to send a gift within another group. Again, we apply the estimation strategy in Equation (4.1) to estimate the effect, but with the dependent variable being the average amount that the user sends to other groups she belongs to in our sample. Since we sample our data at the group level, our test of generalized reciprocity is restricted to those who belong to multiple sampled groups, which yields 18,910 (3.1%) users in our sample.

Altogether, although the sign for the estimated coefficient is positive, it is not significant (see Table 4.13 in Appendix). This null result may be due to two factors. First, although the number of users is not small, a lack of within-stratum variation may underpower our analysis. Indeed, among 12,671 strata, 7,956 contain only one observation. Second, since users may belong to additional groups that are not in our sample, the lack of all sending and receiving history of a user leads to an underestimation of our effect.

4.4.2 Heterogeneous effect of gift contagion

The fine-grained information in our large dataset provides opportunities to examine how the effect size of gift contagion varies in multiple dimensions, which further deepens our understanding of gift contagion. For example, [146] only examine Facebook birthday gifts, while our sample includes the sending of red packets for different purposes, such as celebrating the Lunar New Year and job promotion. We first examine how the strength of gift contagion differs between festivals and non-festival seasons by running regressions separately, and report the main results below.[10]

Result 2 (Festival effect) *The effect of gift contagion is stronger during festival than non-festival seasons, and the difference is significant in the first 10 minutes.*

Support. *As shown in Columns (1) and (2) of Table 4.2, the size of the overall effect is larger during festival than non-festival seasons (10 minutes: 0.1948 versus 0.1207,* $p = 0.039$*; 24 hours: 0.2474 versus 0.1271,* $p = 0.106$*).*

By Result 2, we reject the null hypothesis in favor of Hypothesis 2. We also examine the extensive and intensive margins and present the results in Columns (3)-(6), Table 4.2. From these results, we see that the extensive margin is significant for both festival and non-festival seasons, although the differences are not significant at the 5% level (10 minutes: $p > 0.1$; 24 hours: $p = 0.087$). Additionally, although no significance is detected for the intensive margin, the intensive margin for festival season is larger than non-festival season.

Second, we examine whether the gift contagion effect varies across different types of groups. We identify three group types by inferring a group's composition from group names. (1) *Relative groups*: groups with names containing 家(family). (2) *Classmate groups*: groups with names containing [班(class), 小学/中学/初中/高中(elementary/secondary/low secondary/high secondary school, respectively), 大学(college/university), 校(school), 届/级(grade)]. (3) *Coworker groups*: groups with names containing 公司(company), 集团(corporate group), 工作(work), and 有限(limited liability). Table 4.3 reports the regression results for group type analysis.

[10]We consider all important dates that people celebrate in China including the Lunar New Year and other festivals.

Table 4.2: Regression analyses for gift contagion: festival versus non-festival seasons

	Overall		Extensive		Intensive	
	10 min	24 h	10 min	24 h	10 min	24 h
	(1)	(2)	(3)	(4)	(5)	(6)
			Festival			
Amount received	0.1948***	0.2474***	0.0031***	0.0031***	0.0771	-0.0237
	(0.0256)	(0.0533)	(0.0001)	(0.0001)	(0.0835)	(0.1442)
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	2,297,290	2,297,290	2,297,290	2,297,290	399,763	545,953
Adjusted R^2	0.0458	0.0493	0.0172	0.0222	0.1626	0.1260
			Non-festival			
Amount received	0.1207***	0.1271**	0.0032***	0.0034***	-0.1313	-0.5508
	(0.0251)	(0.0519)	(0.0001)	(0.0001)	(0.1577)	(0.3229)
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	4,969,156	4,969,156	4,969,156	4,969,156	660,983	824,788
Adjusted R^2	0.0342	0.0318	0.0196	0.0208	0.1861	0.1485

Note: The dependent variable (DV) for Columns (1) and (2) is the amount sent within the respective timeframe. It is coded as zero for those who do not send red packets. The DV in Columns (3) and (4) is the dummy variable for sending red packets. The DV in Columns (5) and (6) is the amount conditioning on sending red packets. Marginal effects are reported. Standard errors clustered at the group- and user-level are in parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

Table 4.3: Regression analyses for gift contagion by group types

	Overall		Extensive		Intensive	
	10 min	24 h	10 min	24 h	10 min	24 h
	(1)	(2)	(3)	(4)	(5)	(6)
			Relatives			
Amount received	0.1458***	0.1934***	0.0031***	0.0031***	0.0923	0.0481
	(0.0213)	(0.0441)	(0.0002)	(0.0002)	(0.0828)	(0.1262)
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	2,200,404	2,200,404	2,200,404	2,200,404	366,553	472,239
Adjusted R^2	0.0342	0.0318	0.0196	0.0208	0.1861	0.1485
			Classmates			
Amount received	-0.0207	-0.0825	0.0074***	0.0074***	-0.9680**	-1.0938**
	(0.0627)	(0.1229)	(0.0009)	(0.0009)	(0.4107)	(0.5300)
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	408,397	408,397	408,397	408,397	47,242	62,616
Adjusted R^2	0.0982	0.1206	0.0082	0.0148	0.2631	0.1956
			Coworkers			
Amount received	0.1101*	0.0498	0.0032***	0.0031***	-0.2257	-0.6819
	(0.0647)	(0.1068)	(0.0006)	(0.0006)	(0.3692)	(0.5326)
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	143,297	143,297	143,297	143,297	17,974	23,156
Adjusted R^2	0.1633	0.1723	-0.0067	0.0041	0.3694	0.3236

Note: The dependent variable (DV) for Columns (1) and (2) is the amount sent within the respective timeframe. It is coded as zero for those who do not send red packets. The DV in Columns (3) and (4) is the dummy variable for sending red packets. The DV in Columns (5) and (6) is the amount conditioning on sending red packets. Marginal effects are reported. Standard errors clustered at the group- and user-level are in parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

Result 3 (Group type effect) *The overall effect of gift contagion is stronger in relative groups than in classmate groups.*

Support. *The effect size is significantly greater in relative groups than in classmate groups (10 minutes:* 0.1458 *versus* -0.0207*,* $p = 0.012$*; 24 hours:* 0.1934 *versus* -0.0825*,* $p = 0.034$*). There is no significant difference for other pairwise comparisons.*

By Result 3, we reject the null hypothesis in favor of Hypothesis 3. In addition, as indicated by Table 3, we find that the overall effect is significant for relative groups, while the effects are not significant for the other two groups.[11] We also compare the extensive and intensive margins across group types and find that the extensive margin is significantly higher in classmate groups than in relative groups (10 minutes: 0.0074 versus 0.0031, $p < 0.01$; 24 hours: 0.0074 versus 0.0031, $p < 0.01$) or coworker groups (10 minutes: 0.0074 versus 0.0032, $p < 0.01$; 24 hours: 0.0074 versus 0.0031, $p < 0.01$). By contrast, the intensive margin for classmate groups is significantly smaller than that for relative groups (10 minutes: -0.9680 versus 0.0923, $p = 0.011$; 24 hours: -1.0938 versus 0.0481, $p = 0.036$).

4.4.3 "Luckiest draw" effect

As discussed in Section 4.3, we posit that luckiest draw recipients may exhibit stronger gift contagion. To examine the behavioral difference between luckiest and non-luckiest draw recipients, we run the regressions in Equation (4.1) for these two subgroups separately and report the results in Table 4.4.

Result 4 (Luckiest draw effect) *Gift contagion is stronger for luckiest draw recipients than non-luckiest draw recipients, and the difference is significant in the 10-minute timeframe.*

Support. *In Columns (1) and (2) of Table 4.4, the marginal effects for luckiest draw recipients are larger than those for non-luckiest draw recipients (10 minutes: 0.3251 versus 0.0984,* $p = 0.017$*; 24 hours: 0.3979 versus 0.1631,* $p > 0.1$*).*

[11] The overall effect is marginally significant for coworker groups only within 10 minutes ($p < 0.1$).

Table 4.4: Regression analyses for gift contagion: luckiest versus non-luckiest draw recipients

	Overall		Extensive		Intensive	
	10 min	24 h	10 min	24 h	10 min	24 h
	(1)	(2)	(3)	(4)	(5)	(6)
			Luckiest			
Amount received	0.3251***	0.3979**	0.0076***	0.0080***	-0.3366	-0.7459*
	(0.0889)	(0.1726)	(0.0004)	(0.0004)	(0.2886)	(0.4531)
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	1,923,297	1,923,297	1,923,297	1,923,297	296,799	371,698
Adjusted R^2	0.0640	0.0503	0.0348	0.0373	0.1844	0.1222
			Non-luckiest			
Amount received	0.0984***	0.1631**	0.0007***	0.0008***	0.2199	0.2012
	(0.0334)	(0.0814)	(0.0001)	(0.0001)	(0.1477)	(0.3272)
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	5,343,149	5,343,149	5,343,149	5,343,149	763,947	999,043
Adjusted R^2	0.0373	0.0413	0.0159	0.0184	0.1543	0.1181

Note: The dependent variable (DV) for Columns (1) and (2) is the amount sent within the respective timeframe. It is coded as zero for those who do not send red packets. The DV in Columns (3) and (4) is the dummy variable for sending red packets. The DV in Columns (5) and (6) is the amount conditioning on sending red packets. Marginal effects are reported. Standard errors clustered at the group- and user-level are in parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

By Result 4, we reject the null hypothesis in favor of Hypothesis 4. Moreover, we find that the extensive margin for luckiest draw recipients is almost ten times of that for non-luckiest draw recipients (10 minutes: 0.0076 versus 0.0007, $p < 0.01$; 24 hours: 0.0080 versus 0.0008, $p < 0.01$). Conditional on sending red packets, the marginal effect on the amount that a user sends is smaller for luckiest than non-luckiest draw recipients, although the difference is only marginally significant (Columns (5) and (6), 10 minutes: -0.3366 versus 0.2199, $p = 0.086$; 24 hours: -0.7459 versus 0.2012, $p = 0.090$).

It is possible that luckiest draw recipients send more simply because they receive more. To control for this "amount" effect, we implement the following matching procedure. Specifically, we match each luckiest draw recipient with non-luckiest draw recipients by holding the following variables constant: the total amount of the red packet, the number of recipients of that red packet, the order of receiving time, and the amount received by the corresponding recipient. Matching on the first three variables allows us to control for the effect of unobserved variables, as the backdoor criterion is satisfied [196]. Moreover, matching on the received amount allows us to

further control for the difference in the amount received.[12] Our matching procedure yields 668,936 luckiest draw recipients and 1,658,283 non-luckiest draw recipients, representing successful matching of 33.7% of our luckiest draw recipients. Additionally, we bootstrap for 1,000 replicates to construct the confidence intervals.

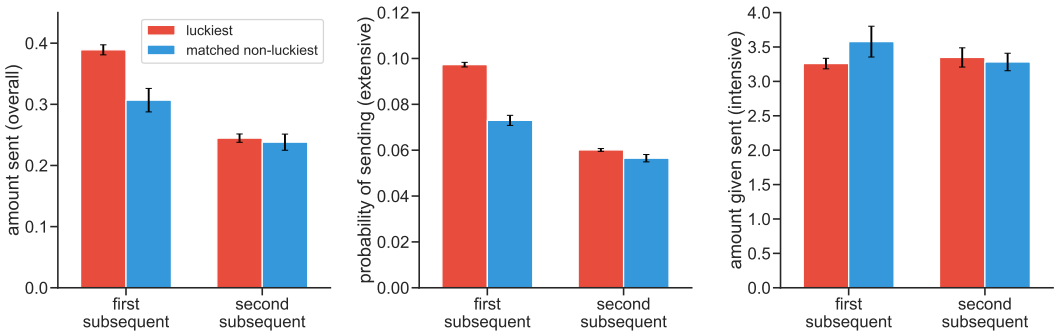


Figure 4-6: Comparisons between luckiest and non-luckiest draw recipients for the unconditional amount that a user sends (left), the probability of sending (middle), and the conditional amount that a user sends (right). Error bars are the 95% CIs.

As shown in the left panel of Figure 4-6, the cash amount sent in the first subsequent red packet increases by 0.082 CNY from non-luckiest to luckiest draw recipients, an effect that is significant at the 1% level. This result suggests that being the luckiest draw recipient alone promotes the gift contagion. By contrast, for the *second* subsequent red packet, we find a much smaller increase and the effect is no longer significant ($p = 0.37$).[13] We also decompose the overall effect into the extensive and intensive margins.[14] Again, we find a significant difference for the extensive margin, instead of the intensive margin.

Altogether, our results suggest the existence of a group norm whereby the luckiest draw recipients should send the first subsequent red packet. This norm can facilitate coordination among group members to maintain a chain of red packets [91, 219, 102]. Moreover, we find that the strength of such a group norm is contingent on

[12]We conduct one-to-many matching [232].

[13]There is no significant difference for the third and subsequent red packets.

[14]Note that the definitions here are slightly different: (1) overall: the amount sent in the kth subsequent red packet; (2) extensive margin: whether the recipient sent the kth subsequent red packet; and (3) intensive margin: the amount sent conditional on being the user who sends the kth subsequent red packet.

the discrepancy between the amounts received by the luckiest draw recipient and the amounts received by other recipients (see Appendix), suggesting that the fairness concern plays a role in influencing the strength of the luckiest draw effect [45, 46]. Additionally, gifts may pressure recipients into signaling their own virtue, especially for luckiest-draw recipients. However, since users endogenously decide whether they want to receive a red packet, they can avoid social pressure by not clicking on an offered packet. In addition, the upper limit of a red packet's cash amount is not very large – 200 CNY (roughly 30 USD) and the average cash amount for luckiest-draw recipients in our setting is 1.16 CNY. Therefore, we suspect that social pressure or reputational concerns do not play an important role in our setting.

4.4.4 Social contagion and social network

In this section, we apply social network analysis to understand how the group network structure affects the strength of our observed gift contagion. Since group members may or may not be private contacts on the platform ("friends"), we construct a relationship network among group members, with each edge indicating that two group members are contacts.

First, we examine how individual network positions affects gift contagion. Specifically,we focus on the clustering coefficient and degree. The clustering coefficient of user i in group g [121, 258], or the extent to which a user's friends are connected, is defined below:

$$\text{clustering coefficient}(i,g) = \frac{\sum_{j \in \mathcal{N}_i^g} \sum_{k \in \mathcal{N}_i^g, k \neq j} \mathbb{1}[k \in \mathcal{N}_j^g]}{|\mathcal{N}_i^g|(|\mathcal{N}_i^g| - 1)}, \tag{4.2}$$

where $\mathcal{N}_i^g$ is the set of a user i's in-group friends in group g. The value of the clustering coefficient ranges from $[0, 1]$; 0 indicates that none of i's friends are connected and 1 indicates that all of i's friends are connected in an online group. Moreover, we use the normalized degree in our analysis: $\frac{\text{degree}(i,g)}{\text{No. of group members}}$, with a range of $[0, 1]$.[15]

Table 4.5 reports the regression results adding the clustering coefficient, the nor-

[15]Compared to the (unnormalized) degree, normalized degree considers the effect of group size.

malized degree, and their interaction terms with the amount received as independent variables. We summarize the results below:

Result 5 (Individual network position on gift contagion)

(a) The overall effect of gift contagion is smaller for group members with a higher clustering coefficient.

(b) The normalized degree does not significantly impact the overall effect of gift contagion.

Support. *As shown in Columns (1) and (2) of Table 4.5, the interaction term between "Amount received" and "Clustering coefficient" is negative and significant at the 1% level (10 minutes:* -0.2961, $p < 0.01$*; 24 hours:* -0.7641, $p < 0.01$*), and the clustering coefficient itself is also negative and significant (10 minutes:* -0.1416, $p < 0.05$*; 24 hours:* -0.4770, $p < 0.01$*). The coefficient for normalized degree is positive and significant at the 1% level (10 minutes:* 0.9943, $p < 0.01$*; 24 hours:* 2.2320, $p < 0.01$*), although its interaction term with "Amount received" is not significant.*

By Result 5(a), we reject the null hypothesis in favor of Hypothesis 5(a). This finding is consistent with prior studies [21, 247]. Moreover, as shown in Columns (3)-(6), the interaction terms for the extensive and intensive margins are also negative and significant.[16] For the normalized degree, we do not find a salient interaction effect, and thus we fail to reject the null hypothesis in favor of Hypothesis 5(b).[17]

Next, we examine the effect of the group-level network structure on our observed gift contagion. We use *the average normalized degree*, or *network density* to measure the degree to which a network is tightly connected [184]:

$$\text{average normalized degree}(g) = \frac{\sum_{i \in \mathcal{G}} |\mathcal{N}_i^g|}{|\mathcal{G}| \times (|\mathcal{G}| - 1)}. \tag{4.3}$$

[16]The only exception is the intensive margin result for 10 minutes.

[17]We also examine the impact of centrality, in particular, the eigenvector centrality, which is widely used in the literature of networks [170, 135]. However, no significant overall effect for the interaction term is found (Table 4.17).

Table 4.5: Effect of individual in-group degree and clustering coefficient on gift contagion

	Overall		Extensive		Intensive	
	10 min	24 h	10 min	24 h	10 min	24 h
	(1)	(2)	(3)	(4)	(5)	(6)
Amount received	0.3544***	0.6852***	0.0085***	0.0088***	0.1642	0.4334
	(0.0875)	(0.2150)	(0.0004)	(0.0005)	(0.3349)	(0.6273)
Amount received × normalized degree	0.0612	0.1756	-00051***	-0.0052***	0.2748	0.2561
	(0.0.0951)	(0.2133)	(0.0004)	(0.0004)	(0.3739)	(0.6754)
Amount received × clustering coefficient	-0.2961***	-0.7641***	-0.0028***	-0.0030***	-0.4187	-1.021**
	(0.0768)	(0.1682)	(0.0004)	(0.0004)	(0.2842)	(0.4736)
Normalized degree	0.9943***	2.2320***	0.0631***	0.0872***	3.7129***	6.4928***
	(0.0787)	(0.1734)	(0.0027)	(0.0034)	(0.3997)	(0.6754)
Clustering coefficient	-0.1416**	-0.4770***	-0.0457***	-0.0661***	1.6278***	2.0105***
	(0.0553)	(0.1211)	(0.0021)	(0.0027)	(0.2953)	(0.5087)
Group size fixed effect	Y	Y	Y	Y	Y	Y
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	7,266,446	7,266,446	7,266,446	7,266,446	1,060,746	1,370,741
Adjusted R^2	0.0400	0.0403	0.0260	0.0308	0.1524	0.1102

Note: The dependent variable (DV) for Columns (1) and (2) is the amount sent within the respective timeframe. It is coded as zero for those who do not send red packets. The DV in Columns (3) and (4) is the dummy variable for sending red packets. The DV in Columns (5) and (6) is the amount conditioning on sending red packets. Marginal effects are reported. Standard errors clustered at the group- and user-level are in parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

$\mathcal{G}$ denotes the set of group g's members. The average normalized degree ranges from [0, 1]. We present results of the regression with average normalized degree and the interaction term in Table 4.6.

Result 6 (Group network structure on gift contagion) *Although there is no significant overall effect, the interaction effect between the amount received and the average normalized degree is negative and significant for the extensive margin.*

Support. *In Columns (3) and (4) of Table 4.6, the interaction term for "Amount received × avg normalized degree" is negative and significant (10 minutes:* -0.0046*,* $p < 0.01$*; 24 hours:* -0.0044*,* $p < 0.01$*), although there is no significant interaction effect for overall effects or intensive margins.*

By Result 6, we reject the null hypothesis in favor of Hypothesis 6 for the extensive margin. As shown in Table 4.6, although the "Amount received × avg normalized degree" are non-significant in Columns (1) and (2), they are negative and significant in Columns (3) and (4). We also examine the impact of overall clustering [135] and find similar results. The detailed analyses are reported in Appendix.

Table 4.6: Effect of average normalized degree in groups

	Overall		Extensive		Intensive	
	10 min	24 h	10 min	24 h	10 min	24 h
	(1)	(2)	(3)	(4)	(5)	(6)
Amount received	0.2221***	0.4167***	0.0062***	0.0062***	0.0187	0.1988
	(0.0592)	(0.1346)	(0.0003)	(0.0003)	(0.2079)	(0.3902)
Amount received × avg normalized degree	-0.0970	-0.3400	-0.0046***	-0.0044***	-0.0226	-0.6034
	(0.0890)	(0.2103)	(0.0004)	(0.0004)	(0.3082)	(0.6009)
Avg normalized degree	0.8392***	1.6332***	0.0150***	0.0145**	4.7210***	7.0219***
	(0.0898)	(0.2079)	(0.0048)	(0.0063)	(0.4379)	(0.7828)
Group size fixed effect	Y	Y	Y	Y	Y	Y
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	7,266,446	7,266,446	7,266,446	7,266,446	1,060,746	1,370,741
Adjusted R^2	0.0397	0.0399	0.0239	0.0272	0.1523	0.1100

Note: The dependent variable (DV) for Columns (1) and (2) is the amount sent within the respective timeframe. It is coded as zero for those who do not send red packets. The DV in Columns (3) and (4) is the dummy variable for sending red packets. The DV in Columns (5) and (6) is the amount conditioning on sending red packets. Marginal effects are reported. Standard errors clustered at the group- and user-level are in parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

Finally, we examine the impact of receiving gifts on network dynamics. We change the dependent variable in Equation (4.1) to the number of within-group edges added by a user after the user receives a red packet. Figure 4-7 presents the results, where the x-axis indicates different time intervals and the y-axis represents the marginal effect of the amount received (in CNY) on the number of new friends added by the recipient within the group. On average, receiving 100 CNY encourages the recipient to add 0.05 friends within the group in the subsequent seven days ($p < 0.01$). Although this appears to be a small effect, it reflects how in-group gifts can foster in-group interactions through establishing new connections.

In sum, our findings suggest that in-groups gifts not only promote gift contagion, but can also encourage within-group interaction and strengthen group solidarity.

4.5 Discussion

Taking advantage of the random assignment of red packet amounts to gift recipients, we leverage a natural experiment to quantify the strength of gift contagion within online groups. We document the presence of gift contagion and further find that the overall effect is driven primarily by the extensive margin, i.e., receiving red packets encourages more users to send packets. The degree of gift contagion varies across

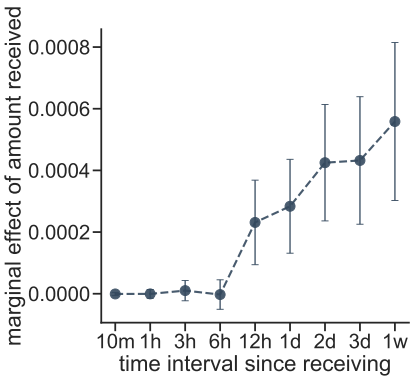


Figure 4-7: The marginal effect on the within-group edges added by the recipient within the group. Error bars are the 95% CIs.

different time periods and various groups. Moreover, we find evidence of a group norm whereby the luckiest draw recipients are expected to take the lead in sending the first subsequent red packet. Regarding the moderating effect of in-group social networks, we find that the higher a user's clustering coefficient is, the less susceptible she is to gift contagion. Additionally, there is a significantly negative interaction effect for the extensive margin between the amount received and how tightly knit a group network is. Altogether, our results, especially the analyses for the extensive and intensive margins, deepen our understanding of the social phenomenon of gift contagion.

There are several possible future directions based on our study. First, it would be interesting to examine how receiving red packets affects other types of user behaviors, such as group communication and liking others' feeds. Second, due to data constraints, we are not able to disentangle which mechanism, such as reciprocity or fairness concern, is the main driver for our observed gift contagion. Therefore, carefully design experimental studies are needed for future work to investigate the primary mechanism. Finally, as we are analyzing online gift contagion in East Asian culture, it would be interesting to explore whether our results can be generalized to offline settings or other culture groups.

4.6 Appendix

Table 4.7: Summary statistics of groups

Variables	Mean	Min	25%	50%	75%	Max
Group size	19.91	3	8.0	14.0	24.0	490
Total no. of red packets	210.24	9	54.0	115.0	253.5	8458
Total cash amount of red packets (CNY)	919.30	0.11	164.46	418.86	990.0	373679.07

Table 4.8: Summary statistics of users (1)

Variables	**Count**	**Proportion**
Gender		
Male	1,783,737	51.69%
Female	1,639,955	47.53%
Unreported	26,848	0.78%

Table 4.9: Summary statistics of users (2)

Variables	**Mean**	**Min**	**25%**	**50%**	**75%**	**Max**
Within-group degree	8.75	1	4	7	11	358
No. of private contacts	182.61	0	54	110	204	25,956
No. of groups that a user joins	38.99	1	9	20	40	16,945,750

Table 4.10: Summary statistics of red packets

Statistics	**Mean**	**Min**	**25%**	**50%**	**75%**	**Max**
Amount	4.37	0.01	0.5	1	5	200
No. of recipients	5.06	1	3	5	5	100
Time interval between red packets (in seconds)[a]	29304.07	0	46	130	938	12,475,671
Completion time (in seconds)[b]	1267.53	2	10	23	176	509,131

[a] Time interval indicates the time between the current and the preceding red packet.

[b] Completion time measures the time interval between the red packet's sending time and the time when the last share of this red packet was received. Red packets that are not received by anyone are excluded.

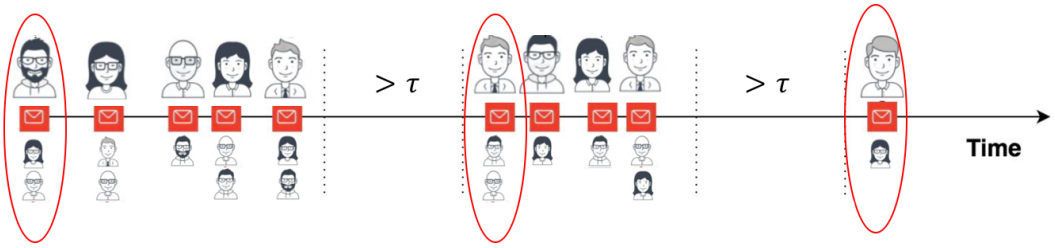


Figure 4-8: The illustration of spontaneous red packets (circled red packets) and sessions (three sessions split by dashed lines in this example).

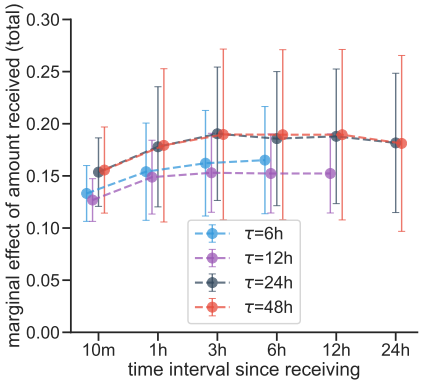


Figure 4-9: Treatment effects for different τ. Error bars are the 95% CIs clustered at the group- and user-levels.

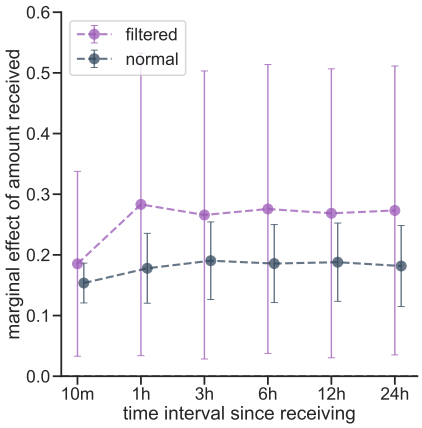


Figure 4-10: Treatment effects for normal groups studied in the main text and groups that were filtered out. Error bars are the 95% CIs clustered at the group- and user-levels.

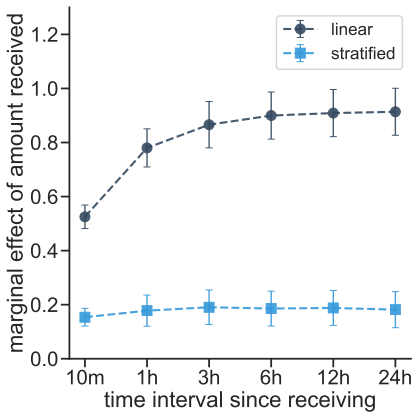


Figure 4-11: The regression results for the linear specification of the effect of (A_r, N_r, O_{ir}). "Linear" represents the results when (A_r, N_r, O_{ir}) is linearly specified. "Stratified" represents the results when (A_r, N_r, O_{ir}) is used to stratify data, as is in the main text. Error bars are the 95% CIs clustered at the group- and user-levels.

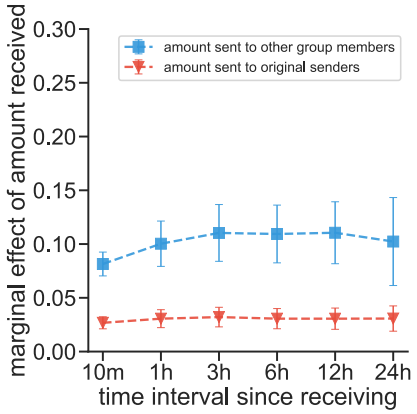


Figure 4-12: The x-axis represents the time interval since receiving a red packet. The y-axis represents the marginal effect of receiving red packets on the amount sent in the future. "Indirect" and "direct" refer to the ratio of the amount sent to group members except for the original sender and to the original sender, respectively. Error bars are the 95% CIs clustered at the group- and user-levels.

Table 4.11: Regression coefficients and the corresponding adjusted p values for red packets with the amount of 5 CNY and 3 recipients. (a, n, o) refers to the total amount, the number of recipients, and the order of receiving time.

Variables	(5,3,1)	Adj. p	(5,3,2)	Adj. p	(5,3,3)	Adj. p
female	0.0016	0.9365	0.0049	0.8036	0.0019	0.9296
degree	0.0520	0.9023	-0.0806	0.9023	0.0090	0.9365
fricnt	0.5963	0.9365	0.1410	0.9365	2.4340	0.8036
joincnt	0.3196	0.9023	0.1584	0.9365	0.3770	0.9023
history_sendamt	-0.6952	0.9365	0.1280	0.9365	0.1321	0.9365
history_sendcnt	-0.2336	0.9023	-0.0903	0.9365	0.1838	0.9023
history_recvamt	-0.9458	0.9023	0.1200	0.9365	0.2068	0.9365
history_recvcnt	-0.5519	0.9023	-0.0977	0.9365	0.4581	0.9023
groupamt	-3.6014	0.9365	-1.1416	0.9365	0.9528	0.9365
groupnum	-7.6666	0.9023	-5.3852	0.9296	3.9070	0.9365

Table 4.12: Regression coefficients and the corresponding adjusted p values for red packets with the amount of 10 CNY and 5 recipients. (a, n, o) refers to the total amount, the number of recipients, and the order of receiving time.

Variable	(10,5,1)	Adj. p	(10,5,2)	Adj. p	(10,5,3)	Adj. p	(10,5,4)	Adj. p	(10,5,5)	Adj. p
female	0.0006	0.9490	-0.0002	0.9490	-0.0007	0.9490	-0.0007	0.9490	0.0022	0.6569
degree	0.0922	0.8288	-0.0785	0.6569	0.003	0.9556	0.0048	0.9490	0.0112	0.9490
fricnt	-1.5021	0.6569	1.4011	0.6569	-1.1583	0.7047	-0.6723	0.8807	-0.4135	0.9490
joincnt	-0.0285	0.9490	0.0302	0.9490	-0.0339	0.9490	-0.2852	0.6569	-0.095	0.9490
history_sendamt	1.0972	0.8132	0.8839	0.8288	2.1079	0.6569	-0.2787	0.9490	-0.32	0.9490
history_sendcnt	0.0613	0.9490	0.2013	0.6569	0.1389	0.6569	-0.0471	0.9490	-0.0268	0.9490
history_recvamt	1.3355	0.6569	0.1375	0.9490	1.4951	0.6569	-0.5232	0.9119	0.0097	0.9875
history_recvcnt	0.6882	0.6569	0.478	0.7167	0.7242	0.6569	-0.2521	0.9119	-0.3717	0.7052
groupamt	29.2084	0.6569	4.5624	0.9490	23.1333	0.6569	-1.8278	0.9490	3.9539	0.9490
groupnum	7.2803	0.9490	13.0648	0.7052	3.8353	0.9490	-6.4186	0.9358	-7.3134	0.8807

A Brief History of Red Packets

Red packets are typically sent from older relatives to children or unmarried young people. Children and unmarried young people wish their older relatives a "Happy New Year," which is called *bainian* (拜年). Then, the older relatives give them cash gifts in exchange. Red packets are a traditional custom dating back to the Han dynasty (circa 50 BC - 100 AD) [222].

Lucky money was once called "压祟钱", the literal meaning of which is the money

Table 4.13: Regression results for generalized reciprocity

	Overall		Extensive		Intensive	
	10 min	24 h	10 min	24 h	10 min	24 h
	(1)	(2)	(3)	(4)	(5)	(6)
Amount received	0.0067	-0.0250	0.0003	0.0001	-5.5001	3.7199
	(0.0191)	(0.3870)	(0.0006)	(0.0015)	(58.5689)	(12.2207)
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	154,312	154,312	154,312	154,312	321	5182
Adjusted R^2	0.0665	0.1119	0.0665	0.1119	0.9775	0.5685

Note: The dependent variable (DV) for Columns (1) and (2) is the amount sent within the respective timeframe. It is coded as zero for those who do not send red packets. The DV in Columns (3) and (4) is the dummy variable for sending red packets. The DV in Columns (5) and (6) is the amount conditioning on sending red packets. Marginal effects are reported. Standard errors clustered at the group- and user-level are in parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

Table 4.14: Regression results for the ratio of the second largest to largest amount received

	Overall		Extensive		Intensive	
	10 min	24 h	10 min	24 h	10 min	24 h
	(1)	(2)	(3)	(4)	(5)	(6)
Amount received	0.3209***	0.3621*	0.0055***	0.0056***	-0.3529	-0.7868
	(0.0920)	(0.1862)	(0.0004)	(0.0004)	(0.3079)	(0.5367)
Ratio	0.0515	-0.1205	-0.0363***	-0.0379***	0.2693	-0.0584
	(0.1126)	(0.2379)	(0.0021)	(0.0022)	(0.4646)	(0.8487)
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	1,268,240	1,268,240	1,268,240	1,268,240	223,329	273,898
Adjusted R^2	0.0620	0.0469	0.0326	0.0350	0.1811	0.1208

Note: The dependent variable (DV) for Columns (1) and (2) is the amount sent within the respective timeframe. It is coded as zero for those who do not send red packets. The DV in Columns (3) and (4) is the dummy variable for sending red packets. The DV in Columns (5) and (6) is the amount conditioning on sending red packets. Marginal effects are reported. Standard errors clustered at the group- and user-level are in parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

Table 4.15: Heterogeneous treatment effects for recipient's gender

	Overall		Extensive		Intensive	
	10 min	24 h	10 min	24 h	10 min	24 h
	(1)	(2)	(3)	(4)	(5)	(6)
			Female recipient			
Amount received	0.1919***	0.1912**	0.0033***	0.0034***	-0.1933	-0.5985*
	(0.0407)	(0.0838)	(0.0001)	(0.0002)	(0.1830)	(0.3448)
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	3,870,582	3,870,582	3,870,582	3,870,582	551,408	711,086
Adjusted R^2	0.0450	0.0441	0.0188	0.0217	0.1867	0.1444
			Male recipient			
Amount received	0.1836***	0.1747**	0.0033***	0.0035***	-0.2100	-0.6796**
	(0.0398)	(0.0801)	(0.0002)	(0.0002)	(0.1728)	(0.3244)
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	3,380,557	3,380,557	3,380,557	3,380,557	506,889	656,573
Adjusted R^2	0.0439	0.0454	0.0194	0.0220	0.1587	0.1158

Note: The dependent variable (DV) for Columns (1) and (2) is the amount sent within the respective timeframe. It is coded as zero for those who do not send red packets. The DV in Columns (3) and (4) is the dummy variable for sending red packets. The DV in Columns (5) and (6) is the amount conditioning on sending red packets. Marginal effects are reported. Standard errors clustered at the group- and user-level are in parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

Table 4.16: Heterogeneous treatment effects for sender's gender

	Overall		Extensive		Intensive	
	10 min	24 h	10 min	24 h	10 min	24 h
	(1)	(2)	(3)	(4)	(5)	(6)
			Female sender			
Amount received	0.1421***	0.1825***	0.0037***	0.0039***	-0.1601	-0.3750
	(0.0292)	(0.0568)	(0.0002)	(0.0002)	(0.1446)	(0.2626)
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	2,960,098	2,960,098	2,960,098	2,960,098	478,056	609,017
Adjusted R^2	0.0541	0.0471	0.0211	0.0233	0.2318	0.1547
			Male sender			
Amount received	0.1622***	0.1873***	0.0028***	0.0029***	0.0583	-0.1988
	(0.0206)	(0.0403)	(0.0001)	(0.0001)	(0.0891)	(0.1639)
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	4,329,692	4,329,692	4,329,692	4,329,692	573,594	748,681
Adjusted R^2	0.0309	0.0295	0.0247	0.0272	0.1281	0.0995

Note: The dependent variable (DV) for Columns (1) and (2) is the amount sent within the respective timeframe. It is coded as zero for those who do not send red packets. The DV in Columns (3) and (4) is the dummy variable for sending red packets. The DV in Columns (5) and (6) is the amount conditioning on sending red packets. Marginal effects are reported. Standard errors clustered at the group- and user-level are in parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

Table 4.17: Effect of individual eigenvector centrality on gift contagion

	Overall		Extensive		Intensive	
	10 min	24 h	10 min	24 h	10 min	24 h
	(1)	(2)	(3)	(4)	(5)	(6)
Amount received	0.1653***	0.3093***	0.0061***	0.0061***	0.1938	0.5509
	(0.0588)	(0.1154)	(0.0003)	(0.0003)	(0.2212)	(0.3733)
Amount received × eigen	-0.0319	-0.3678	-0.0086***	-0.0085***	-0.4935	-2.1662*
	(0.1843)	(0.3798)	(0.0007)	(0.0008)	(0.6307)	(1.0945)
Eigen	2.6627***	6.1014***	0.2140***	0.2963***	5.2927***	8.9651***
	(0.1501)	(0.3054)	(0.0033)	(0.0037)	(0.7270)	(1.0232)
Group size fixed effect	Y	Y	Y	Y	Y	Y
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	7,266,446	7,266,446	7,266,446	7,266,446	1,060,746	1,370,741
Adjusted R^2	0.0398	0.0401	0.0308	0.0315	0.1520	0.1099

Note: The dependent variable (DV) for Columns (1) and (2) is the amount sent within the respective timeframe. It is coded as zero for those who do not send red packets. The DV in Columns (3) and (4) is the dummy variable for sending red packets. The DV in Columns (5) and (6) is the amount conditioning on sending red packets. Marginal effects are reported. Standard errors clustered at the group- and user-level are in parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

that drives away the demon *Sui* (祟) on the Lunar New Year's Eve [209]. "岁" (age) has the same pronunciation as "祟." Sui was a demon who would enter houses on New Year Eve's and deliberately terrify and harm children. Children would catch a terrible fever and even mental disorders if they became terrified. To protect the children, parents gave as an offering eight copper coins wrapped in red packets. It was believed that these eight coins would emit strong lights that would drive the demon away. These eight copper coins were considered the initial version of red packets.

In the 1900s, when the printing technique was popularized in China, red packets have been developed into their current form. Chinese characters symbolizing good wishes are printed on red packets. Red packets are no longer used only to ensure the safety of children for superstitious reasons. At present, they usually symbolize senders' wishes for successful fortune, health, studies, and career paths. The wrapping of red packets typically contains characters with such meanings.

In addition to the role of wishes for children or unmarried young people, red packets are also used as cash gifts on other occasions. Invitees to birthday parties, weddings, and funerals are expected to bring cash gifts, usually wrapped in red envelopes to the hosts. The amount of the gift represents the senders' evaluation of the strength of social bonds and relationships between senders and recipients. Receiving

Table 4.18: Effect of overall clustering in groups

	Overall		Extensive		Intensive	
	10 min	24 h	10 min	24 h	10 min	24 h
	(1)	(2)	(3)	(4)	(5)	(6)
Amount received	0.2838***	0.6047***	0.0078***	0.0080***	-0.0416	0.2730
	(0.0809)	(0.1741)	(0.0004)	(0.0004)	(0.2658)	(0.4472)
Amount received × overall clustering	-0.1654	-0.5448**	-0.0061***	-0.0063***	0.0600	-0.6614
	(0.1088)	(0.2407)	(0.0005)	(0.0006)	(0.3550)	(0.6304)
Average clustering coefficient	0.7165***	1.4052***	-0.0017	-0.0061	5.4598***	8.3258***
	(0.1109)	(0.2365)	(0.0062)	(0.0082)	(0.4829)	(0.7520)
Group size fixed effect	Y	Y	Y	Y	Y	Y
Stratum fixed effect	Y	Y	Y	Y	Y	Y
No. of observations	7,266,446	7,266,446	7,266,446	7,266,446	1,060,746	1,370,741
Adjusted R^2	0.0397	0.0399	0.0238	0.0272	0.1522	0.1099

Note: The dependent variable (DV) for Columns (1) and (2) is the amount sent within the respective timeframe. It is coded as zero for those who do not send red packets. The DV in Columns (3) and (4) is the dummy variable for sending red packets. The DV in Columns (5) and (6) is the amount conditioning on sending red packets. Marginal effects are reported. Standard errors clustered at the group- and user-level are in parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

cash gifts is regarded as owing "*renqing*" (favor), and recipients are strongly expected to send back the gift cash in the future (called to *return "renqing"*) [256, 51].

At present, with the proliferation of online platforms, red packets are commonly sent on these platforms. Red packets are used as convenient cash gifts, through either one-to-one or one-to-many channels. These online red packets are no longer only sent from older people to younger people, nor are they only used for the Lunar New Year or important events. The limit of online red packets (typically 200 CNY or approximately 30 USD) reduces the potential social pressure to reciprocate with large-amount cash gifts. Users even use them for entertainment. On other online platforms, red packets are a means of providing coupons to users. Incorporating good wishes for the customers, these red packets may encourage consumption and user engagement.

Sample Description

Our sample includes 3,450,540 unique users in 174,131 groups. For each user, we obtain the demographic information listed below. For variables that are updated monthly, we use the information retrieved in February 2016 for our analysis. More-

over, we identify friendships, i.e., whether users are contacts, between users in our dataset. We summarize our data below:

- Group
  - Group size: the number of group members in a group.
- Group members
  - Gender: self-reported by users.
  - Number of groups that a user joins.
  - Number of private contacts ("friends") that a user has.
  - Within-group degree: number of private contacts (or "friends") that a user has in one group. Note that it is possible that members of the same group may not be "friends."
  - Clustering coefficient: the extent to which a user's friends are connected in the group, as defined in the main text.
- Red packet sending variables
  - Sending time.
  - Sender ID.
  - Total cash amount of the red packet, determined by the sender.
  - The number of recipients, determined by the sender.
- Red packet receiving variables
  - Recipient ID.
  - Receiving time. The time interval between a red packet being sent and being received by the current recipient. A red packet expires 24 hours after being sent. We use the receiving time to infer the order of receiving time of a given red packet r.

– The cash amount received.

We report the summary statistics for group size, the total number of red packets, and the total cash amount of red packets for each group in Table 4.7. We also present summary statistics for users' gender age in Table 4.8. In Table 4.9, we further report information for within-group degree (how many private contacts, or "friends," a user has in a group), the number of private contacts, and the number of groups that she joins.

Finally, we summarize information on red packets (Table 4.10), including the cash amount, the number of recipients, the time interval between two successive red packets in a group, and the total completion time. We find that most red packets contain relatively small amounts (75% of them do not exceed 5 CNY). In addition, the time intervals between two successive red packets are generally small, with all of the money from a given red packet often being received within minutes.

Randomization Check

Conditional on the three variables that determine our stratification, we show that the received amount (T) is independent of the following variables: (1) whether the user is `female`; (2) within-group `degree`, or the number of "friends"; (3) the number of friends (denoted by `fricnt`); (4) the number of groups that the user joins (denoted by `joincnt`); (5) the total amount of red packets that the user has sent in the group (denoted by `history_sendamt`); (6) the total number of red packets that the user has sent in the group (denoted by `history_sendcnt`); (7) the total amount of red packets that the user has received in the group (denoted by `history_recvamt`); (8) the total number of red packets that the user has received in the group (denoted by `history_recvcnt`); (9) the total amount of red packets sent in the group by all group members historically (denoted by `groupamt`); and (10) the total number of red packets sent in the group by all group members historically (denoted by `groupnum`).

Specifically, we run simple OLS regressions for each stratum in which the dependent variable is one of the aforementioned variables, and the independent variable is

the cash amount received by a user. We present the corresponding coefficients and the adjusted p-values after implementing false discovery control[18] for the two representative cases in Tables 4.11 and 4.12, respectively. In summary, no significant correlation is found. We also check other combinations of the amount sent and the number of recipients, and no significance is found. Overall, our data pass the randomization check.

Calculation of the variance of the amount received

Here we provide a complete calculation for the variance of the amount received by the oth recipient. Although the expected amount is the same for different recipients, we show that their variance is generally different. Specifically, we show a non-decreasing trend for the variance with o. Let S_o denote the summation of the first o recipients' amounts received ($S_o = V_1 + V_2 + ...V_o$). Recall that a is the total amount of the red packet, n is the number of recipients, and V_o is the amount received by the oth recipient.

We first consider the case in which $o < n$:

$$\begin{aligned}\mathbb{E}[S_{o+1}^2] &= \mathbb{E}\big[(S_o + V_{o+1})^2\big] = \mathbb{E}[S_o^2] + 2\mathbb{E}\Big[S_o \times \frac{a - S_o}{n-o}\Big] + \frac{4}{3}\mathbb{E}\Big[\frac{(a-S_o)^2}{(n-o)^2}\Big] \\ &= \mathbb{E}[S_o^2]\Big(1 - \frac{2}{n-o} + \frac{4}{3}\frac{1}{(n-o)^2}\Big) + \mathbb{E}[S_o]\Big(\frac{2a}{n-o} - \frac{8}{3}\frac{a}{(n-o)^2}\Big) + \frac{4}{3}\frac{a^2}{(n-o)^2} \\ &= \mathbb{E}[S_o^2]\Big(1 - \frac{2}{n-o} + \frac{4}{3}\frac{1}{(n-o)^2}\Big) + a^2\Big(\frac{2o}{(n-o)n} - \frac{8o}{3(n-o)^2 n} + \frac{4}{3(n-o)^2}\Big).\end{aligned} \tag{4.4}$$

Note that $\mathbb{E}[S_o] = \mathbb{E}[V_1] + ... + \mathbb{E}[V_o] = \frac{ao}{n}$.

We next relate V_o, the amount received by the oth recipient, to S_o:

[18] We use the Benejamini-Hochberg procedure with $\alpha = 0.1$ because it is more conservative and generates smaller adjusted p values than methods such as the Bonferroni correction.

$$\mathrm{Var}(V_o) = \mathbb{E}\Big[\frac{1}{12} \times \frac{(2(a - S_(o-1)))^2}{(n-o+1)^2}\Big] = \frac{1}{3}\mathbb{E}\Big[\frac{(a-S_{o-1})^2}{(n-o+1)^2}\Big];$$
$$\mathrm{Var}(V_{o-1}) = \mathbb{E}\Big[\frac{1}{12} \times \frac{(2(a - S_(o-2)))^2}{(n-o+2)^2}\Big] = \frac{1}{3}\mathbb{E}\Big[\frac{(a-S_{o-2})^2}{(n-o+2)^2}\Big].$$

Dividing the first equation by the second, we obtain

$$\frac{\mathrm{Var}(V_o)}{\mathrm{Var}(V_{o-1})} = \frac{(n-o+2)^2}{(n-o+1)^2} \times \frac{\mathbb{E}[(a-S_{o-1})^2]}{\mathbb{E}[(a-S_{o-2})^2]}.$$

Then,

$$\begin{aligned}
\frac{\mathrm{Var}(V_o)}{\mathrm{Var}(V_1)} =& \frac{\mathrm{Var}(V_o)}{\mathrm{Var}(V_{o-1})} \times \frac{\mathrm{Var}(V_{o-1})}{\mathrm{Var}(V_{o-2})} \times ... \times \frac{\mathrm{Var}(V_2)}{\mathrm{Var}(V_1)} = \\
=& \frac{(n-o+2)^2}{(n-o+1)^2} \times \frac{(n-o+3)^2}{(n-o+2)^2} ... \times \frac{n^2}{(n-1)^2} \\
& \times \frac{\mathbb{E}[(a-S_{o-1})^2]}{\mathbb{E}[(a-S_{o-2})^2]} \times \frac{\mathbb{E}[(a-S_{o-2})^2]}{\mathbb{E}[(a-S_{o-3})^2]} \times ... \times \frac{\mathbb{E}[(a-S_1)^2]}{a^2} \\
=& \frac{n^2}{(n-o+1)^2} \frac{\mathbb{E}[(a-S_{o-1})^2]}{a^2}.
\end{aligned}$$

Since $\mathrm{Var}[V_1] = \frac{a^2}{3n^2}$, and $\mathbb{E}[S_{o-1}] = \frac{a(o-1)}{n}$, we obtain

$$\mathrm{Var}(V_o) = \frac{a^2}{3(n-o+1)^2}\Big(1 - \frac{2(o-1)}{n}\Big) + \frac{1}{3(n-o+1)^2}\mathbb{E}[S_{o-1}^2]. \tag{4.5}$$

Combining Equations 4.4 and 4.5, we have

$$\mathrm{Var}(V_{o+1}) = \Big(1 + \frac{1}{3(n-o)^2}\Big)\mathrm{Var}(V_o).$$

Therefore, we know $1 < o < n$:

$$\mathrm{Var}(V_o) = \mathrm{Var}(V_1)\prod_{k=1}^{o-1}\Big(1 + \frac{1}{3(n-k)^2}\Big) = \frac{a^2}{3n^2}\prod_{k=1}^{o-1}\Big(1 + \frac{1}{3(n-k)^2}\Big). \tag{4.6}$$

For $o = n$, because the last two recipients split the surplus uniformly at random, V_{n-1} and V_n are identically distributed. Therefore,

$$\mathrm{Var}(V_n) = \mathrm{Var}(V_{n-1}) = \frac{a^2}{3n^2}\prod_{k=1}^{n-2}\Big(1+\frac{1}{3(n-k)^2}\Big).$$

In summary,

$$\mathrm{Var}(V_o) = \begin{cases} 0 & n = 1 \text{ and } o = n \\ \frac{a^2}{3n^2}\prod_{k=1}^{o-1}\left(1+\frac{1}{3(n-k)^2}\right) & n > 1 \text{ and } o < n \\ \frac{a^2}{3n^2}\prod_{k=1}^{n-2}\left(1+\frac{1}{3(n-k)^2}\right) & n > 1 \text{ and } o = n \end{cases} \tag{4.7}$$

Furthermore, the variance increases with o when $o < n$.

Additional Analyses

Threshold selection

First, we illustrate the selection process of a spontaneous red packet (circled in Figure 4-8). If the interval of two consecutive red packets is greater than τ, we divide these two red packets into two "sessions." Therefore, in each session, the time interval between any two consecutive red packets is less than τ.

We use 24 hours for our timeframe in the main analyses because a red packet expires 24 hours after being sent. To examine the sensitivity of our results to the selection of a 24-hour window, we also select 6, 12, and 48 hours and re-run our analyses; their respective treatment effects are shown in Figure 4-9. We find similar results for a 48-hour time window, and a slightly smaller treatment effect for a 6- or 12-hour time window.

Gambling and unnamed groups

To explore the impact of our filtering process on our results, we re-run our regressions focusing on gambling groups, for which group names indicate red packet games or gambling, and unnamed groups with functions that are unclear. The results in Figure 4-10 show that the filtered groups appear to have a higher marginal effect. These results suggest that filtering out these gambling groups may have helped us obtain a

more accurate understanding of gift contagion.

Alternatives for the econometric model

If we apply a linear specification for the effect of the three-dimensional vector — the total amount of the red packet (A_r), the number of recipients (N_r), and the order of receiving time (O_{ir}), we obtain the following regression:

$$Y_{gir} = \beta T_{gir} + \gamma_1 A_r + \gamma_2 N_r + \gamma_3 O_{ir} + \epsilon_{gir}. \tag{4.8}$$

As shown in Figure 4-11, the regression results suggest a much larger marginal effect than the results in the main text. One possibility is that the effect of (A_r, N_r, O_{ir}) on Y_{gir} is not a linear combination of the three variables, which raises the issue of functional form misspecification.

Direct and indirect reciprocity.

Here we present an additional analysis to show that, compared to direct reciprocity, indirect reciprocity plays a dominant role in promoting gift contagion. We separate Y_{gir} in Eq. (4.1) into two components: $Y^{(1)}_{gir}$ and $Y^{(2)}_{gir}$. $Y^{(1)}_{gir}$ is the amount sent to the original sender (the sender of red packet r), which measures *direct reciprocity*. $Y^{(2)}_{gir}$ is the amount sent to other group members, which could be a proxy for *indirect reciprocity*. [19] As shown in Figure 4-12, on average, the marginal effect on the amount received by the original sender is 3.07% [SE=0.65%, $p < 0.001$] in the next 24 hours. By contrast, this effect size is much larger for the amount received by other group members: 10.25% [SE=2.10%, $p < 0.001$].

"Luckiest draw" and fairness concerns.

To investigate whether the fairness concern plays a role in affecting the amount that a user sends, we run the following regression for luckiest draw recipients:

[19] Note that a small proportion of the total amount does not belong to either $Y^{(1)}_{gir}$ or $Y^{(2)}_{gir}$ because the sender can also receive a share of her own red packet.

$$Y_{gir} = \beta T_{gir} + \alpha_{\text{ratio}} Z_{gir} + \sum_s \gamma_s B_s(A_r, N_r, O_{ir}) + \epsilon_{gir}. \tag{4.9}$$

Compared to Eq. (4.1), we include an additional independent variable: Z_{gir}. Let T'_{gir} is the second-largest amount received from the same red packet; and then $Z_{gir} = \frac{T'_{gir}}{T_{gir}}$ represents the ratio of the second-largest amount to the largest amount. We remove the observations (luckiest recipients) that do not have corresponding the second luckiest recipient. Table 4.14 reports the regression results.

We find that the ratio of the second-largest to the largest amount has a negative impact on the likelihood of sending red packets. For example, for the next 10 minutes and 24 hours, we have $\hat{\alpha}_{\text{ratio}} = -0.0363$ and -0.0379 for extensive margin ($p < 0.01$). This suggests that, when the cash amount received by the luckiest-draw recipient is much larger than that received by others, the recipient may feel more obligated to send red packets to the group because of her fairness concern.

Gender effects

We also examine gender differences. We do not find any significant gender differences when running regressions on female and male recipients separately (Table 4.15). As shown in Table 4.16, although we do not find a significant difference in the overall effect between female senders and male senders, we find that the red packets sent by female senders exhibit a higher extensive margin than those sent by male senders ($p < 0.01$ for both 10 minutes and 24 hours).

Effect of overall clustering.

Here we use overall clustering as an alternative to the average normalized degree, as a measure of group network structure. The overall clustering of group g is define as

$$\text{overall clustering}(g) = \frac{\sum_{i \in |\mathcal{G}|} \#\{(j,k) | j,k \in \mathcal{N}_i^g \text{ and } i \neq j \neq k \text{ and } k \in \mathcal{N}_j^g\}}{\sum_{i \in |\mathcal{G}|} \#\{(j,k) | j,k \in \mathcal{N}_i^g \text{ and } i \neq j \neq k\}}. \tag{4.10}$$

$\mathcal{G}$ denotes the set of members of group g, and $\mathcal{N}_i^g$ denotes the set of network neighbors of user i in group g.

We report the regression results in Table 4.18. As shown in Columns (1) and (2), a larger overall clustering predicts a larger amount sent within a group. However, the interaction terms are negative ($p > 0.1$ and $p = 0.024$ for 10 minutes and 24 hours, respectively), suggesting that groups with a larger overall clustering do not necessarily induce stronger gift contagion. In Columns (3) and (4), the interaction terms are significant and negative, suggesting that for groups with a larger overall clustering, there is generally a smaller extensive margin. For the intensive margin (Columns (5) and (6)), we do not find any significance in the interaction terms.

Chapter 5

Identifying prosocial incentive effectiveness on exercises with advanced causal inference techniques

5.1 Background

With the prevalence of smartphones, mobile applications have provided a variety of functions that support and facilitate user daily life, such as shopping, learning, and social networking [183, 89, 65, 50]. In particular, mobile health (mHealth) applications greatly benefit the social well-being of our societies by allowing individuals to monitor their health conditions and maintain healthy lifestyles [89, 228, 4, 229, 250, 230]. As of 2017, more than 318,000 mHealth applications have been released, guiding users' healthy diet, physical activities, and disease control.[1] Especially during the COVID-19 pandemic, many mobile applications have been utilized around the world for contact tracing, in an effort to contain the spread of the pandemic [79].

Despite the numerous benefits to society, many mHealth applications also face challenges in their efficacy to promote a healthier lifestyle. In general, the design of mHealth applications lacks support from evidence-based strategies [117]. In fact, a

[1] https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health

few medical studies have utilized small-size randomized control trials, but most of them find non-significant effects of mobile health applications on physical activities [246, 145, 10, 155, 109, 225, 195, 172, 108].

Therefore, we need more empirical studies to understand how to design effective incentives on the mobile platforms to promote physical activity. Previous studies have proposed several incentives that may encourage people to maintain healthy lifestyles and encourage physical activity. The first type is to use monetary incentives (e.g., giving out money) to encourage more physical activities [60, 41, 5]. However, there has been controversy regarding the long-term effect of monetary incentives [210]. The second type of incentive is the social incentive, including peer effects or social comparison. As many other types of human behaviors do, physical activities are also found contagious: people tend to exercise more after noticing or witnessing peers doing so [67, 19]. Another plausible reason is social comparison, whereby users may compete to champion or rank high on the ranking page [234, 263]. Moreover, some studies have jointly examined the efficacy of monetary and social incentives on encouraging physical activities, and they find that the effect of monetary incentive on fitness is larger when the bonus is received by a friend of the subject than received by the subject [6, 165].

Our study thus seeks to identify the efficacy of a third incentive – prosocial incentive – on fitness behavior [39]. Prosocial incentives highlight the benefits to others when engaging individuals in an activity. Prosocial incentives have been shown to be effective in many ways, such as improving mental health [80] or work performance[131]. Currently, several mobile phone applications have associated fitness with charitable giving. For example, CharityMiles links the distances of running or biking to donation amount to charities.[2] In Chile, there is a similar design named *burn to give* (or *betterfly*).[3] In China, several "giant" applications, such as WeChat and Ant Forest, employ charitable incentives to encourage walking. They have a feature known as step donation – users' step counts are associated with the amount

[2]https://charitymiles.org/
[3]https://betterfly.cl/

donated by a third party (usually a company). Although many applications have utilized this charitable incentive to encourage workouts, few studies have revealed its efficacy. It is possible that instead of encouraging users to exercise more, those online platforms only attract fitness enthusiasts who also care for charities, resulting in spurious correlation between physical activities and engagement in charitable behavior.

In this paper, we combine an online field experiment of approximately 40 million users and follow-up observational studies to examine the efficacy of prosocial incentives on fitness behavior. This is a collaboration with WeChat, the largest Chinese social networking platform. WeChat is a "giant" mobile application in China, which enables users to complete a variety of activities virtually, including social networking, remote working, reading news and online posts, conduct mobiles payments for groceries or airline tickets, and sending gifts to friends; however, very few studies have so far analyzed the behavior on this platform [127, 270]. In this study, we focus on WeRun, which is a mini-program on WeChat. WeRun records users' daily step counts and provides a personalized ranking page for each user, which displays the step counts of the user and his or her friends. In addition, WeRun has a "step donation" feature, which associates users' daily step count with the welfare of charities. Our study focuses on the effectiveness of the step donation feature and compares it with that of the ranking feature.

Our online field experiment employs an encouragement design [115], a widely adopted experimental approach when experimenters cannot directly manipulate the behaviors of subjects. We send different messages to users to remind them of the prosocial or social incentives on the platform; and thus users in different treatment groups may be encouraged to different extents to view the corresponding features. Our encouragement design increases the engagement in the step donation feature by 2.5 times. We find a significant effect in a subgroup of the population, who have historically experienced step donation. However, we do not observe significance for the effect on the whole population, possibly because of the difficulty in promoting users' prosocial behavior.

Next, to complement the field experiment, we perform a matching design and focus on a subgroup of the population where we believe prosocial incentives have a larger effect size. The matching provides us with more statistical power, and enables us to investigate the effect heterogeneity and long-term treatment effects. The results of the matching design are largely consistent with the findings from the field experiment. We also find that the prosocial incentive appears to have a larger effect than the social incentive for this subgroup of the population, and that the effect size increases as the users are more historically physically active or with more social connections on WeChat.

Finally, we discuss the underlying prosocial mechanisms. According to [37], drivers for prosocial behavior can be intrinsic, extrinsic, and reputational. On WeRun, engaging in step donation is not visible to other users; therefore, reputational concerns are not a driver in this setting. Moreover, there is no explicit extrinsic driver, such as rewards or punishments, in the step donation feature, extrinsic motivations are neither a driver for users' prosocial behavior. The design of step donation feature enables us to perform a regression discontinuity design to distinguish the effect of "warm glow giving" [14], or impure altruism as a contrast to pure altruism. Warm glow giving refers to the phenomenon that people engage in prosocial behavior to gain the joy of helping others; however, they do not actually take into account the welfare of charitable recipients [14]. We find that for users who are likely driven by warm glow giving, the prosocial incentive may only encourage the user to engage in prosocial activity when no additional efforts of physical activity are required.

At least two main advantages exist when we study WeRun. First, our sample is more able to represent the general public. Previous related studies have primarily used online apps that are primarily for fitness, and so their sample may only represent fitness enthusiasts or young people who rely more on electronic devices to track their fitness behavior. It is reported that many WeRun users are more than 50 years old [162], which is different from many other online fitness platforms (such as 23 in [21]). Second, we can compare the strength of charitable incentives versus social incentives. We collected both the behavior of step donation and of checking the ranking page. The

step donation page represents a charitable incentive but the ranking page indicates the existence of a social incentive.

WeRun is a mini-program on WeChat, and is listed in parallel with a user's recent contacts in the "Chats" panel. It can be understood as a bot-chat that sends routinely information to users. As of 2020, WeRun has 300 million daily active users. WeRun users routinely receive a message at a time between 10 pm to 11 pm in their local time. As shown in the left panel of 5-1, users receive their daily step count and rank among their WeChat friends who also use WeRun. At the same time, they are notified of who among their friends "championed" on their step count ranking page, i.e., the person who walked the largest number of steps among all his or her WeChat friends. If a user chooses to click on the daily message, he or she will be directed to the ranking page, which is presented in the middle panel of Figure 5-1. On the top of this page, the daily cumulative step count and rank of the user are presented. Below the user's own information, there is a detailed ranking of the step counts of the user's WeChat friends. Step counts that exceed 10k are marked in orange while step counts that below 10k are marked in green. Steps counts will be reset to zero in the midnight and every day a new step ranking for each user is generated. Users can like the step counts of their WeChat friends (the "heart" icon).

The right panel presents the "step donation" page. This page can either be entered through "..." on the top right on the ranking page, or through clicking their step profile (e.g., 4074 steps in this example). The step donation feature allows any users who walk more than 10k steps to ask a third party (usually a company) to donate a small amount to charities.[4]

On WeRun, the step donation feature has a daily quota. A third party, usually a company who would signal their corporate responsibility, eventually donates to charities after a user completes step donation. Therefore, these companies set an upper limit for the daily step donation total amount, which is approximately a few hundreds of thousands Chinese yuan. In fact, the total amount has almost been

[4]Although it is called "donation," users' step count would not be decreased by engaging in step donation.

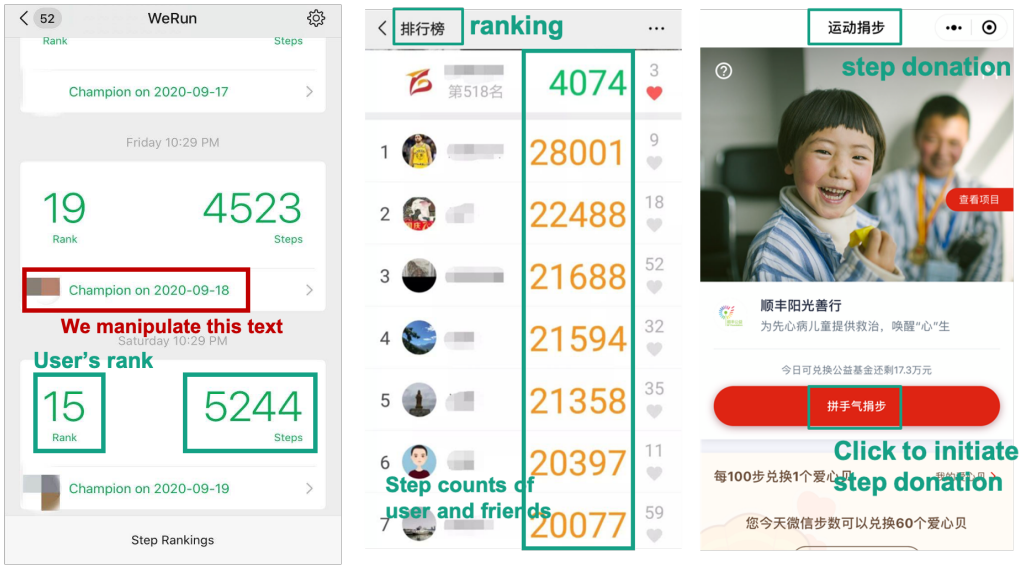


Figure 5-1: Illustration of step ranking page and step donation page.

reached everyday, with a deadline reached at some point in the evening. We utilize this deadline to deploy our regression discontinuity design.

5.2 Experimental study

The observational study suggests that both prosocial and social incentives can encourage users to work out more. Next, we design an online experiment to promote more workouts by encouraging users to get exposed to the prosocial and social incentives.

5.2.1 Experimental design

Through the collaboration with WeChat, we design and implement a field experiment. In our experiment, WeChat randomly manipulated the text in the left panel in Figure 5-1.[5] The treatment assignment is on the user level – once the user is assigned to an experimental group, the user always receives the text belonging to that group until the experiment terminates. This design helps eliminate the "carryover effects" [156] across different treatment assignments: imagine when the message one user received

[5]To make sure users are balanced across treatment groups, they did know what messages they will see before enter the page indicated by the left panel in Figure 5-1.

varies across days, the change in user behavior can be attributed to the messages received on any day so far. The experiment was run for 17 days from September 1st to September 17th, 2020.[6] We only analyze users in Mainland China and thus only one single time zone exists in our analysis. WeChat or the authors did not view or analyze the up-to-date data until the experiment terminated, and thus the optional stopping issue does not apply in our study.

Our study is essentially an encouragement design [115]: users are randomly sent different messages, which promote them with different probabilities of being engaged in the step donation or checking the step ranking page. We designed five treatment groups and one control group. Within each group, we also slightly change the text in case that results are completely driven by non-interesting factors, such as wording or text length. Such a variation may also generate heterogeneity in the first stage in the encouragement design – that is, the probability of viewing the step donation page or the step ranking page). Each treatment group contains approximately 6.3 million users. We present the summary of the message that each experimental group receives in Table 5.1. Details, such as the actual Chinese text that users receive and the slight change in text are presented in Table 5.6.

We next illustrate our rationale for designing different text for each group. Users in the `donate` group are reminded of the step donation feature; the design of this group is to examine the effect of the prosocial incentive. Note that engaging in step donation is not visible to other users, so such a design does not involve any hint on the potential benefits to the user. To further investigate how users' different prosocial motivations drive their prosocial behavior, we designed two other groups – `donate_achieve` and `donate_charity`. People engaged in prosocial behavior primarily for two reasons: impure altruism (warm glow giving) and pure altruism [37]. The theory of warm-glow giving indicates that people engage in prosocial behavior because they feel good about it, regardless of the actual impact of recipients [14]; by contrast, pure altruism

[6]We acknowledge that the experiment was implemented during the COVID-19 pandemic, which may impact user behavior. However, according to the JHU CSSE COVID-19 data, the 7-day average new case number is 12 in Mainland China as of September 17, 2020. Such a trend has been maintained since April 2020. This indicates that COVID-19 does have a severe impact on user behavior and internal validity of this study.

Table 5.1: Description of different treatment groups

Treatment group name	Message (translated)
`donate`	You can donate your steps.
`donate_achieve`	You can donate your steps to mark your achievement.
`donate_charity`	You can donate your steps to help charities.
`ranking`	Check your step count ranking among your friends.
`tomorrow`	Keep on it tomorrow.
`default` (control)	XXX champion today.

indicates that people engage in prosocial behavior because they hope to contribute to the well-being of the recipients. In our study, `donate_achieve` is designed to reflect impure altruism more, while `donate_charity` is designed to reflect pure altruism more. We conjecture that text designed with specific prosocial motivation only drives users with the corresponding motivation to engage in step donation.

In order to compare the efficacy of the prosocial incentive versus the social incentive, we designed the `ranking` group, where the text reminds the users of the ranking page. By doing so, we expect that users in this group will be more likely to check the ranking page and be influenced by the step count ranking. Moreover, we propose `tomorrow` as a benign intervention. This group may help eliminate, for example, the "novelty effect" [245, 199]: users may react to any change on the platform in the short term but this effect may disappear quickly. Finally, we hold some users as the control, named the `default` group. In this group, the message is what users normally received prior to our experiment:"who championed today" as what the control group received.

Our main outcome variable is the step count on a certain day. We use `step_count` to refer to the step count and `step_count_10k` to whether the user walks more than 10,000 steps. The choice of 10,000 steps as the cutoff is because users can only donate their steps when they walk more than 10,000 steps. In addition, we also examine whether a user engaged in step donation on a certain day, indicated by `step_donation`, or whether the user viewed the ranking page, indicated by `step_ranking`. These two outcomes are used to measure the effectiveness of the messages designed on encouraging users to either view the step donation page or the ranking page; even-

tually, we employ an IV design, whereby we treat `step_donation` and `step_ranking` on the first day of our experiment as the endogenous variable and `step_count` and `step_count_10k` on the second day as outcome variable.[7]

As discussed in Section 5.1, WeRun step donation has a daily quota for the total amount of step donation, which generates a daily deadline. Only users who arrive prior to the deadline can successfully complete their step donation. Because the deadline is reached prior to 10pm on all days during our 17-day experiment period, all users who were promoted to attempt step donation were notified that "the deadline has been passed for today and please try tomorrow." Therefore, our experiment is testing the effect of viewing the step donation with the information that the current day's quota has been reached. Users may be not motivated to walk more and engage in step donation on the next day after a failed attempt of step donation on the current day. This also shows our rationale of using the step count on the next day as the outcome variable. Details of our methods and the robustness check are presented in Appendix 5.6.

5.2.2 Results

Summary statistics

Before presenting our results, we present the basic summary statistics in treatment groups in Table 5.2. We use demographics as well as historical behavior in the past 29 days prior to the start day of our experiment.

Our sample is balanced in gender (a female ratio of 49%). The average number of days with step donation is as small as 0.06; this is partially because of the nature of the WeRun design, i.e, the ranking page is highlighted but step donation is not designed as the main feature. Among the 29 days, the average number of days with viewing the ranking page is 9.20, suggesting that many users are active users who frequently view the ranking page. We perform a balance check across those covariates

[7]We choose the subsequent day to measure outcome because most users receive our treatment after 10pm in their local time, and suspect that most users would not be motivated to walk more steps after that time point.

Table 5.2: Summary statistics of the covariates in each treatment group. Sample size and the means of pre-treatment covariates are presented. Standard errors are in parentheses.

Treatment group name	Sample size	Female ratio	Average steps	Days with step donation	Days with step ranking
donate	6,888,743	49.34%	6,770.58	0.0617	9.291
		(0.02%)	(0.005)	(0.0003)	(0.004)
donate_achieve	6,891,377	49.39%	6,768.53	0.0616	9.292
		(0.02%)	(0.005)	(0.0003)	(0.004)
donate_charity	6,890,329	49.38%	6,771.79	0.0620	9.297
		(0.02%)	(0.005)	(0.0003)	(0.004)
ranking	6,889,882	49.37%	6,770.41	0.0630	9.295
		(0.02%)	(0.005)	(0.0003)	(0.004)
tomorrow	6,888,094	49.39%	6,767.85	0.0620	9.290
		(0.02%)	(0.005)	(0.0003)	(0.004)
default (control)	6,888,092	49.39%	6,771.58	0.0626	9.292
		(0.02%)	(0.005)	(0.0003)	(0.004)

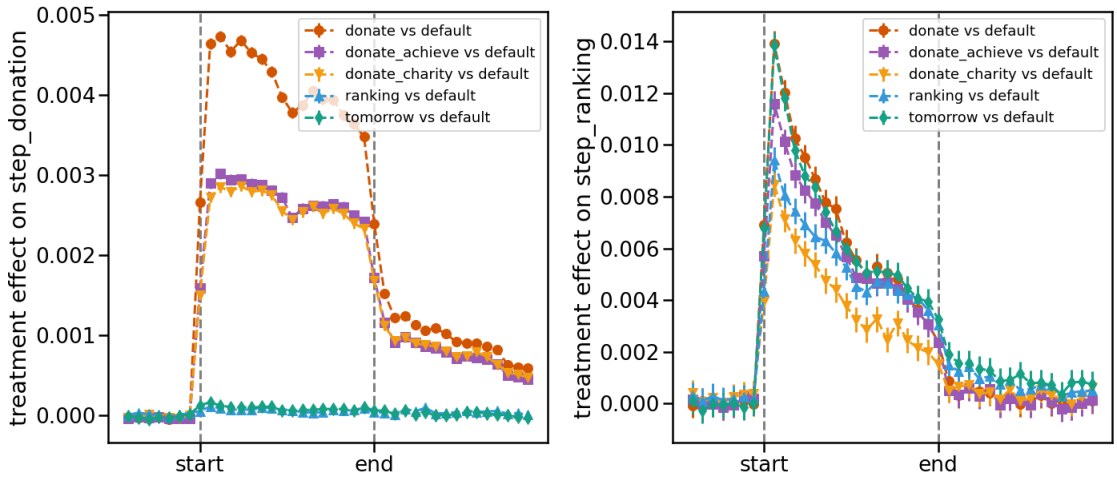


Figure 5-2: Treatment effects on `step_donation` and `step_ranking`. The experiment lasted for 17 days. Each error bar indicates the effect for one day. We also report the effects on the 7 days before and 15 days after the 17-day experimental time period. Error bars are 95% CIs.

in Table 5.7 in Appendix, and our analysis indicates covariates are balanced.

Effect on viewing step donation page and ranking page

As a first step, we examine the effectiveness of the messages designed on encouraging users to view the step donation page or the ranking page. Despite the challenges in promoting prosocial behavior [153, 154, 152], a few previous studies indicate that mobile messaging may be an effective way [267, 236].

First, we examine the impact of engaging in the step donation feature. We present the daily treatment effects in the left panel of Figure 5-2 by comparing each treatment

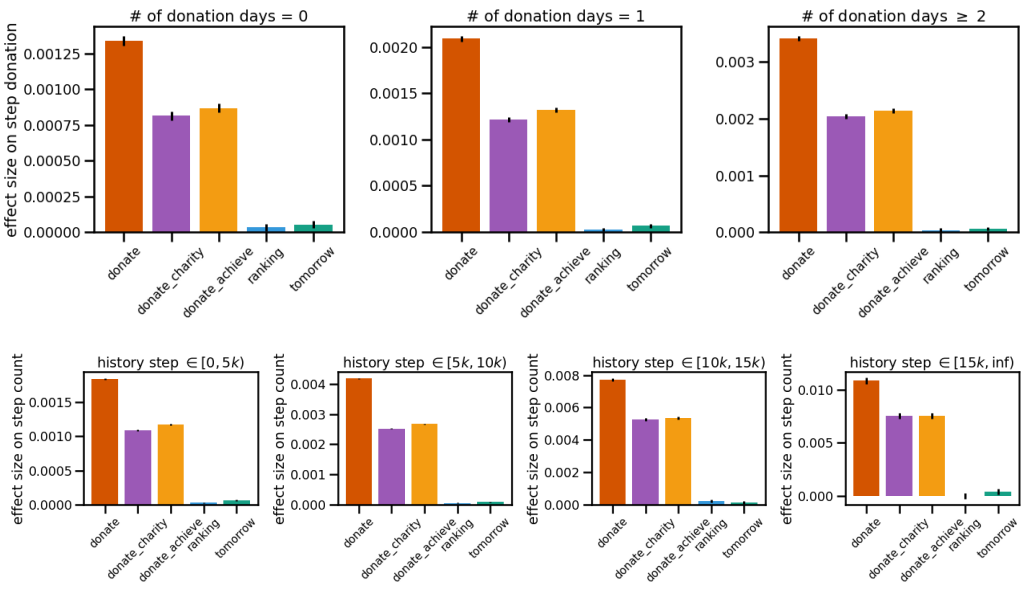


Figure 5-3: Heterogeneity of the effectiveness of different treatment messages on engaging in step donation. The upper panels are conditional on different step donation history, and the lower panels are conditional on different historical average steps. Error bars are 95% CIs.

group versus the control group.[8] We find that `donate` promotes the step donation activities the most: there was a 0.5% increase in the probability of engaging step donation in the `donate` group compared to the `default` group on the first day. The effect size decays slowly until 0.35% at the end of the 17-day experiment period. The slow decay indicates that the effect is not mainly driven by the novelty effect. The effect size of `donate` is in fact large in the relative scale – as shown in Figure 5-13 in Appendix, `donate` increases the likelihood of engaging in step donation by 2.5 times compared to the default group. Moreover, there appears to be a long-term effect – after we terminated the experiment and all users returned to receiving the default message, the effect is still significant for users who were assigned to `donate`.[9]

Interestingly, the `donate_achieve` and `donate_charity` groups are much less effective than the `donate` group. We provide two explanations for this result. One conjecture is that users may be driven by pure altruism or warm-glow giving to be-

[8] To illustrate the effect size on the absolute scale, we also present the means in Figure 5-13 in Appendix.

[9] The same with `donate_achieve` and `donate_charity`.

havior prosocially. Highlighting one type of motivation may drive users with the corresponding motivation only. Based on this conjecture, there may be an approximately equal number of users who are driven by pure altruism or warm-glow giving, leading the similar trends of donate_achieve and `donate_charity` in the left panel of Figure 5-2.

The effects of `ranking` and `tomorrow` are small but significant. This could be explained by a second order effect – those people are more likely to enter the ranking page than `default`, and thus eventually are more likely to enter the step donation page subsequently.

We also investigate heterogeneity in the effect size and present the results in Figure 5-3. We find that users who have historically engaged in step donation are more likely to be encouraged to do it again. The effect size also increases with the historical average step count – those who typically walk more steps are more likely to be encouraged. This is because those users are more likely to exceed the 10k cutoff on the step count. These results suggest that the messages designed are more likely to affect those people who are quantified to engage in step donation or more familiar with the feature. Later in our instrumental variable design, we also leverage this first stage heterogeneity to improve the precision of our estimation.;

Next, we examine the impact of our messages on viewing the step ranking page. The effect of `ranking` is small compared to other treatment messages, although it is intended to encourage users to view the ranking page. A plausible explanation is that this may have a very similar meaning to `default` – promoting users to view the ranking page – and thus users may not be very responsive to this treatment.

We find that `donate` promotes viewing the step ranking page the most – It promotes 1.4% more users on WeRun than `default` to view the step ranking page on the first day of the experiment; this appears to be a novelty effect – the effect decays rapidly and reaches 0.3% at the end of the experiment. However, surprisingly, a trivial message such as `tomorrow` can also have a similar novelty effect. Moreover, as for long term effects, the effect of `tomorrow` appears to persist as 0.1% after the experiment terminated, but the effect of `donate` is much smaller than the experiment

terminated. These results suggest that although reminding of step donation encourages views of step ranking, this effect is unsustainable and may be purely caused by the novelty effect.

Moreover, `donate_achieve` promotes fewer users to view the ranking page than `donate`, and`donate_charity` promotes even fewer. We conjecture that this is because `donate_achieve` may encourage additional users to check their ranking page, while `donate_charity` may only encourage those users who would engage in step donation.

These results suggest that `ranking` may not be an effective message to motivate more views on the step ranking page. The `donate` group and other donation related groups are also driven by the novelty effect to view the ranking page, but the effect size is no greater than the `tomorrow` group, which we believe to only generate a novelty effect.

Effect on step counts.

We proceed to examine the impact on step counts. As shown in Figure 5-4, we do not find any significant impact on either the step count or the probability of whether step count is greater than 10,000. We also perform subgroup analysis conditional on different historical average step counts or historical step donation, or perform variance reduction by accounting for important covariates and examine heterogeneity given historical activity in Appendix 5.6. However, we do not find any significant effects.

To further measure the impact of viewing the step donation page on step counts, we employ an instrumental variable design. We retrieve users who were assigned to "`donate`", "`donate_achieve`", "`donate_charity`", and "`default`" groups as our observations in the instrumental variable. The underlying assumption is that receiving messages about step donation does not have a direct effect on users' step count. – viewing the step donation page exists on the causal path between these two variables. Specifically, viewing the step donation on day 1 of the treatment is the endogenous variable that is located at the causal path from the random assignment to the step

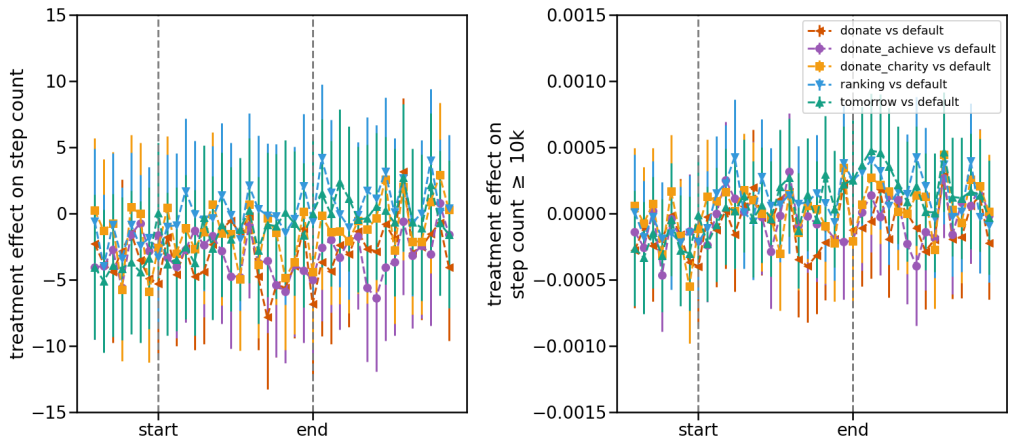


Figure 5-4: Treatment effects on step counts. The experiment lasted for 17 days. Each error bar indicates the effect of one day. We also report the effects before and after this period. Error bars are 95% CIs.

count on day 2.[10]

For the endogenous variable, we leverage two types. The first type is whether the user engaged in step donation on day 1; and the other is whether the user engaged in step donation on day 1 after the deadline. [11] We employ three sets of instrumental variable designs. First ("simply"), we employ the selected seven types of random assignments as instruments (Column (1)).[12] Second ("polynomial"), to leverage first-stage heterogeneity, we add the interaction terms between the random assignments and exogenous variables into the first stage, and then use lasso [242, 38] to retain important instruments with non-zero regression coefficients (Column (2)). Third ("stratum"), we create 25 stratum by the five subgroups divided by historical step donation count times the five subgroups divided by historical step count. We then add the interaction between the stratum indicators and the 7 treatment assignment indicators, and then use lasso to retain instruments that have non-zero regression

[10]We do not account for observations after day 2 because the step counts may be affected by the behavior of viewing the step donation page happening on any day before. We do not include the "`ranking`" and "`tomorrow`" because their corresponding messages may directly affect the step count through motivations that are not related to step donation.

[11]Note that since design messages were sent after 10 pm and all deadlines happened prior to 10 pm, we cannot use whether the user engaged in step donation on day 1 after the deadline as another endogenous variable.

[12]There are two different messages for each among `donate`, `donate_charity`, and `donate_achieve`, which are discussed in Appendix 5.6 and `default` is the seventh.

Table 5.3: Regression results of the instrumental variable design.

	simple		**polynomial**		**stratum**	
	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: step count on day 2					
donate day 1	865.78		319.42		810.40	
	(622.48)		(539.36)		(565.58)	
donate day 1 after cap		855.06		322.03		927.71*
		(608.21)		(543.85)		(547.82)
First-stage F statistic	1,788.76	2,593.40	654.21	891.50	37.42	9.80
Adjusted R^2	0.58	0.58	0.58	0.58	0.58	0.58
No. instruments	6	6	101	101	98	98
No. observations	27,558,541	27,558,541	27,558,541	27,558,541	27,558,541	27,558,541
	Dependent variable: step count $\geq$ 10k on day 2					
donate day 1	0.0885		0.0201		0.1255**	
	(0.0619)		(0.0353)		(0.0510)	
donate day 1 after cap		0.0872		0.0203		0.1198**
		(0.0604)		(0.0356)		(0.0485)
First-stage F statistic	1,788.76	2,593.40	654.21	891.50	37.42	9.80
Adjusted R^2	0.36	0.36	0.36	0.35	0.38	0.38
No. instruments	6	6	101	101	98	98
No. observations	27,558,541	27,558,541	27,558,541	27,558,541	27,558,541	27,558,541

*Note: Robust standard errors are in parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.*

coefficients (Column (3)).

The results of the instrumental variable design on all users are presented in Table 5.3. As indicated, when we use the first and second sets of instruments, we do not find significant effects despite the large first stage F statistics. In Columns (5) and (6), although the result indicates that the step donation engagement on day 1 increases the step count on day 2, the small F-statistics in the first stages indicate the presence of weak instruments. Therefore, their significance may not fully support a clear causal relationship between step donation and the step count.

We also deploy the instrumental variable design on users who engaged at least once in step donation in the past 29 days prior to the experiment only. We do so because the heterogeneity analysis in Figure 5-3 indicates that those users have a larger effect size on encouraging viewing the step donation page. As presented in Columns (1)-(2) and (5)-(6) of Table 5.4, we find a significant effect when the dependent variable is whether the step count is greater than 10,000, the effects are significant at the 5% and 1% levels, respectively. Their F statistics are greater than 10 (e.g., they are 68 and 169 in Columns (1) and (2)). Columns (3) and (4) do not show significance, but the coefficients are not significantly different from the other four columns. These results indicate that for those users who are familiar with viewing the step donation

Table 5.4: Regression results of the instrumental variable design for users with step donation history.

	simple		**polynomial**		**stratum**	
	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: step count on day 2					
donate day 1	1327.21		903.85		1642.55	
	(1030.45)		(1034.31)		(1034.31)	
donate day 1 after cap		1122.63		1580.88		1698.49*
		(895.81)		(1148.34)		(1007.10)
First-stage F statistic	68.04	168.51	26.15	29.42	18.83	23.47
Adjusted R^2	0.60	0.58	0.60	0.58	0.60	0.60
No. instruments	6	6	47	87	45	106
No. observations	558,240	558,240	558,240	558,240	558,240	558,240
	Dependent variable: step count $\geq$ 10k on day 2					
donate day 1	0.2104**		0.0935		0.2329***	
	(0.0901)		(0.0568)		(0.0650)	
donate day 1 after cap		0.1698**		0.0723		0.1472***
		(0.0783)		(0.0568)		(0.0563)
First-stage F statistic	68.04	168.51	26.15	29.42	18.83	23.47
Adjusted R^2	0.39	0.39	0.39	0.39	0.40	0.40
No. instruments	6	6	45	101	44	106
No. observations	558,240	558,240	558,240	558,240	558,240	558,240

*Note: Robust standard errors are in parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.*

page, they can be encouraged to exceed 10k steps on the next day. We do not find significant effects when the dependent variable is the step count, except for Column (6) with significance at the 10% level, although the signs of regression coefficients are all positive.

We also test the effect of viewing the ranking page on the step count on the subsequent day. Here we only employ the "`ranking`", "`tomorrow`", and "`default`" groups, and the endogenous variable becomes whether the user views the step ranking page on day 1. We also create the stratum differently – instead of dividing the sample by the step donation history, we change it to the step ranking history. We present the results in Table 5.8 in Appendix. Overall, we do not find a significant effect of viewing the ranking page on the step count on the subsequent day.

5.3 Matching design

Although our experimental study, especially the instrumental variable design restricted to users with step donation history, provides evidence of the effect of the prosocial incentive on physical activity, it may have two limitations. The first limitation is the statistical significance – our experimental results are not significant in

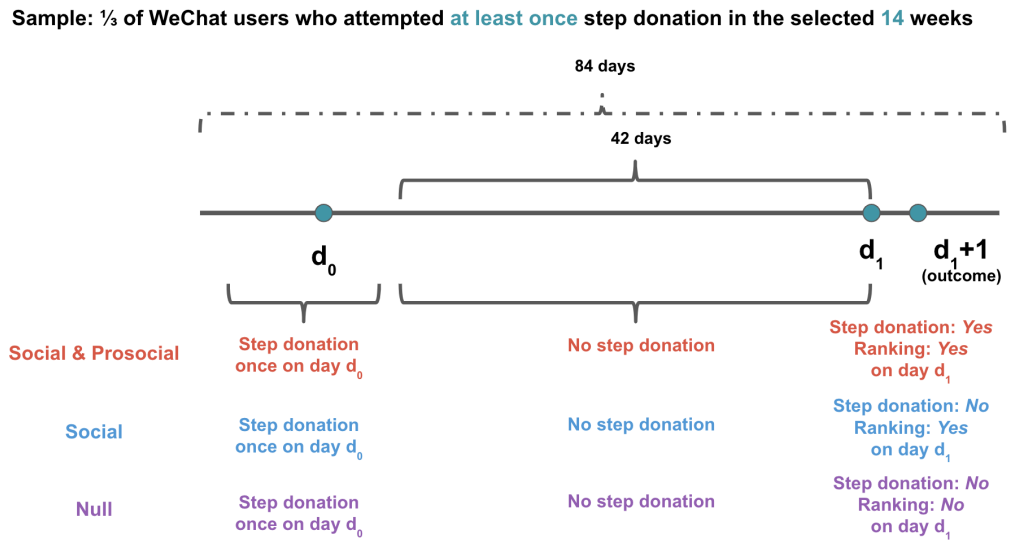


Figure 5-5: Illustration of our matching procedure.

terms of the reduced form (Figure 5-4) or the instrumental variable design for the whole population (Table 5.3). Second, in the experiment, because of low statistical power, we are unable to investigate fine-grained heterogeneous effects of the prosocial incentive or social incentive on physical activity. Here by using observational study, we proceed to further understand how effect size varies across different user subgroups.

In this section, we employ a matching design with observational data and focus our analysis on users who have at least one-time engagement in step donation in the sample.

5.3.1 Data collection and matching design

Collaborating with WeChat, we first access 1/3 of the whole population of the WeRun users with de-identified information. We observe a 84-day window of the behavior of WeRun users, from September 8^{th} to November 30^{th}, 2019. We then focus our analysis on a subsample that is used for the matching procedure.

Specifically, we employ matching to identify the causal effect of viewing the prosocial and social incentives. In our analysis, each observation corresponds to the user

of user i on day d. We first construct a treatment group, referred to as the `social & prosocial` group. If user i viewed the step ranking page (the middle panel of Figure 5-1) and step donation page (the right panel of Figure 5-1) on day d, this observation belongs to the treatment group. We construct another two control groups, `social` group and `null` group. In the `social` group, user i did not view the step donation page on day d, but they viewed the ranking page (the middle panel of Figure 5-1). In the `null` group, user i did not view the step donation page on day d, or the ranking page. We match each observation from the `prosocial & social` to the `social` or `null` group.[13] Our analysis is focused on the average treatment effect on the treated – we match each observation from the treatment group (`social & prosocial`) to the two control groups (`social` and `null`). Comparing the `prosocial & social` group and the matched `social` group estimates the effect of the step donation page (prosocial incentive); we denote it by τ_p. Comparing the `prosocial & social` group and the matched `null` group estimates the effect of the step donation page and the ranking page (prosocial incentive and social incentive); we denote it by $\tau_p + \tau_s$. In this way, we can also indirectly estimate the effect of viewing the ranking page (social incentive), which is τ_s.

In particular, we require that observations in these three groups did not engage in the step donation page in the past 42 days, but did so at least once prior to this 42-day time window.[14] Please refer to Figure 5-5 for illustration. This requirement has two benefits. First, engaging in step donation after not doing so for a long time window implies that the current step donation is more likely to be triggered by an exogenous shock. This may reduce the influence of selection bias on our matching procedure. Second, step donation might not be very well-known among WeRun users. Requiring engaging in step donation at least once in the two control groups helps us to exclude users who are completely unfamiliar with step donation users, which also reduces the impact of the potential selection bias resulting from unfamiliarity

[13]We do not have users who view the step donation page without viewing the ranking page.

[14]Since the farthest time we can retrieve in the data is a successive 84-day time window, the last day of step donation must be one day from the first day of our sample until the start of the 42-day window.

Table 5.5: Summary statistics of the covariates in each group in the matching. Sample size and the means of pre-treatment covariates are presented. Standard errors are in parentheses.

Treatment group name	Sample size	Female ratio	Average steps	Days with step donation	Days with step ranking
prosocial & social	55,504	45.10%	9440.22	2.00	25.85
		(0.21%)	(19.893)	(0.00)	(0.061)
social	13,097,865	46.21%	9424.78	1.00	32.74
		(0.01%)	(1.265)	(0.00)	(0.003)
null	12,905,418	43.84%	7851.40	1.00	11.54
		(0.01%)	(1.090)	(0.00)	(0.003)

with the WeRun step donation feature. In total, we retrieve 55,504 observations in the `prosocial & social` group, 13,097,865 observations in the `social` group, and 12,905,418 observations in the `null` group. Each observation corresponds to a user i on day d. Table 5.5 presents a table of summary statistics.

We measure the outcome by the behavior on the next day. The main outcome variable is the user's step count on the next day, and whether it exceeds 10k.[15] In addition, we measure the user's engagement in step donation or the step ranking page on the next day as outcome variables. The covariates are high-dimensional, which include basic information such as gender, the numbers of WeChat contacts, and the number of interactions; moreover, we retain user histories such as the step counts and the number of viewing WeRun rankings. The comprehensive description of covariates are presented in Appendix 5.6.

We present the summary statistics of the data in Table 5.5. As shown in the table, the number of observations that belong to the `prosocial & social` group is much smaller than the numbers in the other two groups. This imbalance also illustrates the rationale why we regard the `prosocial & social` as the treatment group and match every observation in the treatment to the other two control groups: because of the large imbalance in sample size, it is much easier for us to use a matching algorithm to find a high-quality match in the control group. We examine the quality of our matching results in Appendix .

We mainly employ nearest neighbor matching [273, 241] to estimate the treatment

[15]Again, we choose the subsequent day to measure outcomes because most users receive our treatment after 10pm in their local time, and we assume that most users would not be motivated to walk more steps within the time window from 10pm to midnight. Moreover, we only collected the daily maximum steps and cannot measure the step count increase from the time when they receive the notification to the midnight of that day.

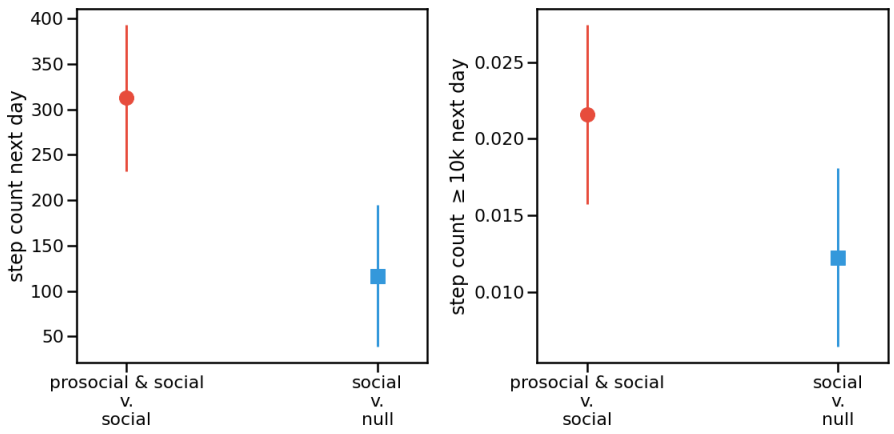


Figure 5-6: Treatment effects on the step count outcomes by comparing the `prosocial & social` versus the matched `social` group, or comparing the matched `social` group and the matched `null` group. Error bars are 95% CIs.

effects of viewing the step donation page or the step ranking page. Since our covariates are high-dimensional, we rescale the covariate space so that the matching quality is better in important covariates. A covariate is more important when it is scaled up, i.e., when it is multiplied by a number that is greater than 1. Learning an effective metric is an open question, and we employ MALTS [192] to learn the optimal metric for the covariate space. The algorithm stretches each dimension by its predictability of the treatment variable and the outcome variable. In the main text, we employ one-to-one matching but our results are robust when it is extended to one-to-many matching.

5.3.2 Matching results

Step donation effect and ranking effect.

After matching is employed, we estimate the difference in means between `prosocial & social` and `null` ($\tau_p + \tau_s$ conditional on `prosocial & social`) as well as the difference in means between `social` and `null` (τ_s conditional on `prosocial & social`). After that, we derive τ_p and τ_s, which correspond to the effect of viewing the step donation page and the ranking page, respectively.

We first examine the impact of viewing the step donation page on the step count

on the subsequent day. As shown in Figure 5-6, users who viewed the step donation page (the `prosocial & social` group) walk more 312.48 steps than users from the matched `social` group on the subsequent day ($p < 0.001$). The likelihood of walking at least 10,000 in the `prosocial & social` group is 2.16% greater than `social` group ($p < 0.001$). The effect sizes are consistent with the results derived from the instrumental variable design in Table 5.3, although they are smaller in the matching results.[16] This result once again suggests that the prosocial incentive, i.e., the step donation page, may encourage users to work out more.

We next examine the effect of the ranking page by comparing the matched `social` group and the matched `null` group, i.e., the estimation of τ_s. As shown in the left panel of Figure 5-6, the `social` group walks 116.77 more steps than the `null` group ($p = 0.003$). In the right panel, the probability of exceeding 10,000 steps in the `social` group is 1.22% larger than the `null` group ($p < 0.001$). This result shows that the effect size of the ranking page is smaller than the step donation page, indicating that the prosocial incentive is more effective in encouraging workouts in our sample. The matching result is also consistent with the experimental design: the confidence intervals of the corresponding IV design (Table 5.8) covers the effects we present in the matching results.

Moderator analysis

In the controlled experiment, we do not investigate the heterogeneous effects due to limited "first-stage" effects. Here, we use matching to investigate the moderating effects of average steps, gender, and WeChat friend count. These analyses further help reveal the moderating factors that affect the effect size.

We first leverage the average steps, i.e., the average in the past 42 days prior to day d when the user triggers a step donation. We categorize observations into four groups – $0 \sim 5000$, $5000 \sim 10000$, $10000 \sim 15000$, and > 15000, and present the results in Figure 5-7.[17] As shown in the left panel of Figure 5-7, the effect

[16] Reasons such as different subgroups of the sample can explain this difference.

[17] We set a 10k cutoff because users who exceed this cutoff view different text on the step donation page, and also their step count is highlighted in orange in the ranking page.

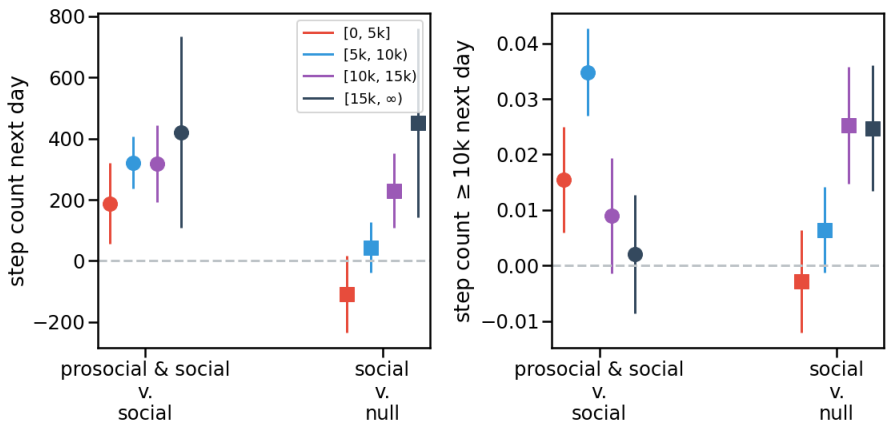


Figure 5-7: Moderating effect conditional on historical average steps. Error bars are 95% CIs.

of the step donation page on the subsequent day appears to increase as a user's average step count increases, despite the overlapping confidence intervals. This result indicates that step donation might encourage users who walk more to exercise even more frequently. On the right panel, however, The effect of the step donation page on the likelihood of exceeding 10k does not show an increasing trend. Interestingly, the 5000 $\sim$ 10000 subgroup has the largest effect, indicating that the design of the step donation page may strongly incentivize users to exceed 10k steps. We conjecture that this is because for users who typically walk more than 10k steps every day, their probability of exceeding 10k is already large, and thus step donation would not change their probability of 10k steps but only change their step counts.

As for the effect of the ranking page, it appears only effective only for users whose average step counts are greater than 10,000 steps. We conjecture that this is because the WeRun ranking page has another incentive to encourage users to exceed 10k steps – those who exceed 10k steps are marked yellow in the ranking page. Moreover, the overall trends are increasing for both outcome variables.

We next investigate gender effects and present the results in Figure 5-8. We find that the effect size of the step donation page (`prosocial & social` versus `social` group) is larger for male users than female users – the effect size of the prosocial incentive is 435.38 for male users, while it is 162.87 for female users ($p < 0.001$). We

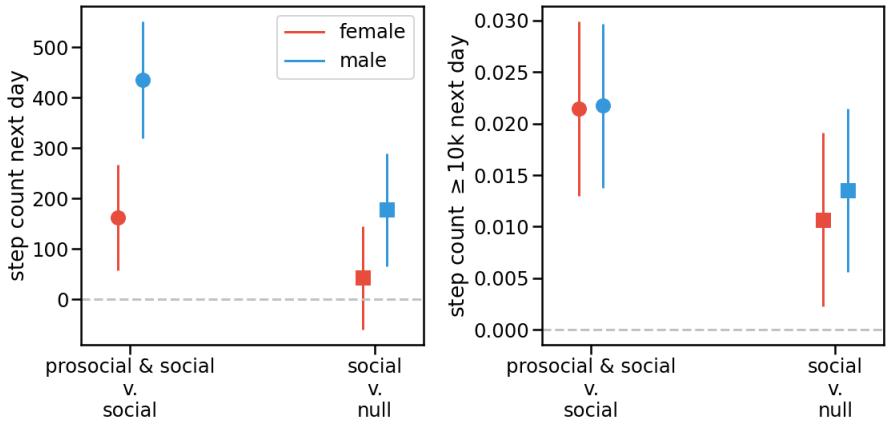


Figure 5-8: Gender difference in the effect size. Error bars are 95% CIs.

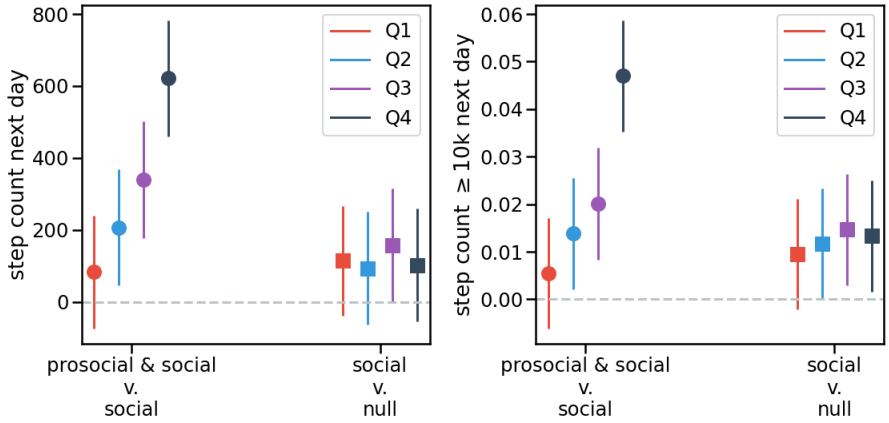


Figure 5-9: Heterogeneity of effect sizes with respect to quartiles of users' WeChat friend counts. Error bars are 95% CIs.

conjecture that the gender difference results from the fact that male users in general walk more than female users, and thus the effect size on the step count is larger for males. In terms of the likelihood of exceeding 10k steps, we do not find significant gender differences ($p > 0.1$). Moreover, we also observe marginally significant gender difference in the effect of the ranking page; for example, the effect for male users is 177.61 while the effect for female users is 42.57 ($p = 0.081$).

We proceed to examine the moderating effect of friend count. We partition users to four quartiles by WeChat the friend count. The cutoffs for the quartiles are 125, 231, and 433. As shown in Figure 5-9, the effect size of the prosocial incentive on step

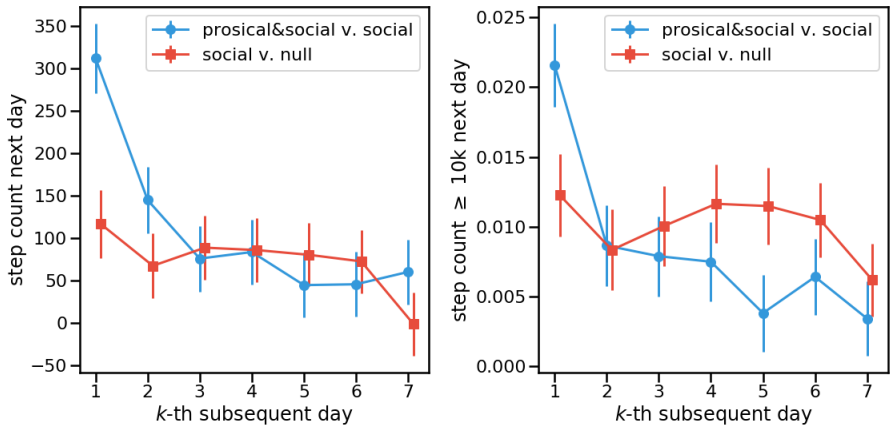


Figure 5-10: The temporal trend of effect sizes. Each error bar indicates the difference between the two groups in comparison. Error bars are 95% CIs.

count for the four quartiles is increasing, indicating step donation has a larger effect for users with more WeChat friends than those with fewer friends. The increasing trend might come from the fact that users with more WeChat friends are in general more responsive to the design on the platform. We do not find the increasing trend in the ranking page effect, which does not support our conjecture that users with more friends may be more responsive to social incentive.

5.3.3 Long-term effect

Since in our controlled experiment, the randomization is only on the user level, we are unable to test the long term effect of experiencing the step donation once. Here we can simply change the outcome variable from the step count on the next day to any subsequent day.

We present the results in Figure 5-10. We find that the step donation page has a persistent effect on the step count but the effect size decays. – the effect of the step donation page on the next day is 312.49 on the next day but it decays to 145.02 on the second subsequent day. All effects of the step donation are significant in the subsequent seven days. The effect of the ranking page on the next day is 116.77 on the next day but it decays to 67.48 on the second subsequent day. All effects of the step donation are significant in the subsequent six days but not significant for the seventh

day. This result indicates that the step donation feature may have a long term effect – it encourages users to exercise more persistently. By contrast, the long-term effect of the ranking feature might be smaller.

5.4 Regression discontinuity design

In both the experiment and the matching results, we do not discuss how different prosocial motivations change users' behavior. Here we perform a regression discontinuity design to distinguish a portion of users who are very likely to be driven by impure altruism (from pure altruism) and investigate the impact of experiencing step donation on their subsequent behavior.

5.4.1 Data description and RDD rationale

The data was collected in the same time period with the matching design. That is, we first randomly sample 30% of the WeRun anonymized users and retain those users who have viewed the step donation page during that time period. This process results in 4.7 million users in this sample. We retrieved the timestamps when they entered the right panel of in Figure 5-1, as well as the step donation deadline time for each day. If a user entered the step donation page multiple times on a single day, we only count the last time. We present a detailed description of the step donation feature and the daily deadline in Figure 5-20 in Appendix. Unfortunately, we are unable to retrieve whether the user clicked the buttons on Panels (b) or (c) in Figure 5-20 and thus we use the last time of a user entering the step donation page as the time they engage in step donation.[18]

We next introduce the rationale of our regression discontinuity design. The users who arrived immediately before and after the discontinuity point can be assumed homogeneous, since they cannot anticipate the daily deadline of step donation. Then

[18]We only know a statistic – among people who exceeding 10,000 steps when entering Panel (a), their probability of clicking on the button is 62.40%. Therefore, the effect size in the rest of this section may be underestimated by roughly 40%.

we can compare the difference in the subsequent behavior of users who arrived before or after the deadline.

We include the following variables for our regression discontinuity design. The running variable is defined as the time of attempted step donations since the discontinuity point on day d. The unit is in seconds. The binary treatment variable indicates whether a user enters the step donation page *after* the daily deadline. We examine the following outcome variables: (1) a dummy variable indicating whether the user attempts step donation on day $d+1$ before the deadline on day $d+1$; (2) a dummy variable indicating whether the user attempts step donation, regardless of before or after the deadline; (3) whether the step count on day $d+1$ is greater than 10k; and (4) the step count on day $d+1$. Covariates mentioned in Section 5.3 are also employed to improve precision.

5.4.2 RDD results

Before presenting the results, we employ the McCrary test [174] to examine whether the marginal density of the running variable is continuous without manipulation. Figure 5-21 in Appendix shows the density curve to be smooth around the discontinuity point, with $p = 0.88$ from the McCrary Test. This result supports the validity of our regression discontinuity design.

We first present a graphical illustration of our regression discontinuity design in Figure 5-11. By comparing the minutes before or after the deadline We find that missing the deadline significantly increases the next day's probability of step donation before the deadline by 2.35% percent ($p<0.01$). It also appears to increase the next day's probability of attempting step donation by 1.12% percent, but the effect is not significant ($p > 0.1$).[19] As for the effect on step count, we do not observe a significant effect. In other words, missing the deadline does not cause these users to walk more or less on the subsequent day.

We further apply the non-parametric regression to estimate the causal effect. We employ Imbens-Kalyanaraman optimal bandwidth [113] to retain observations used in

[19] We compare the outcomes in the one minute before or after the deadline

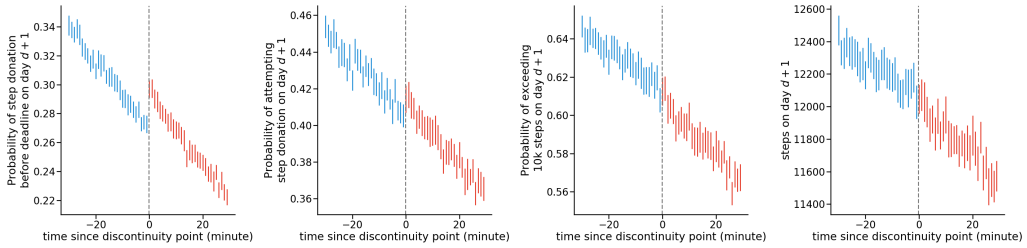


Figure 5-11: Graphical illustration of the effect of discontinuity points on the four outcome variables. Error bars are 95% CIs clustered at the user level.

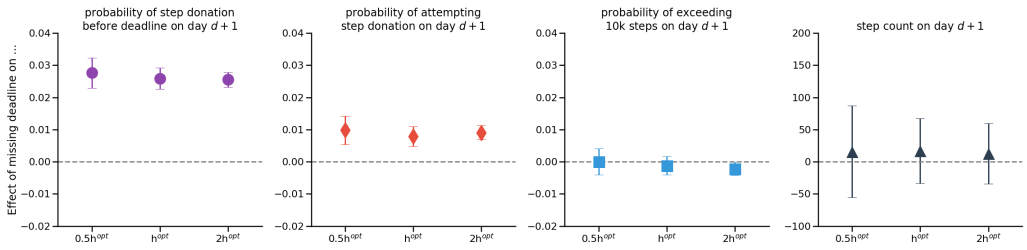


Figure 5-12: Non-parametric estimates of the effect of missing a deadline. h^{opt} refers to the Imbens-Kalyanaraman optimal bandwidth. To show the robustness, we also present the results when the bandwidth is $\frac{1}{2}h^{\text{opt}}$ and $2h^{\text{opt}}$. Error bars are 95% CIs clustered at the user level.

the non-parametric regression. We present the results in Figure 5-12, where h^{opt} is the optimal bandwidth. We also use $\frac{1}{2}h^{\text{opt}}$ and $2h^{\text{opt}}$ to test the robustness of our results. To address data interdependence, all standard errors are bootstrapped on the user level. As shown in the figure, we find that missing the deadline increases the user's probability of making the deadline on the next day by 2.59% [95%CI=2.26%, 2.92%]. Moreover, missing the deadline also increases the user's probability of attempting step donation on the next day by 0.79% [95%CI=0.48%, 1.11%]. These results mean that after missing today's deadline, users will be more likely to attempt step donations on the next day, and they will also try to attempt step donation earlier on the next day.

For the step count and its probability of exceeding 10k steps, we do not find a significant effect ($p > 0.1$). This result could be caused by the fact that human mobility has regularity and cannot be easily impacted by such a prosocial motivation. In other words, impure altruism may only encourage users who are qualified to engage in step donation to complete the prosocial action; however, impure altruism may not

be able to encourage users to walk more steps to complete the requirement of the prosocial incentive of the mobile platform.

We also investigate the heterogeneity of the effect of missing the deadline. For example, we examine the heterogeneity conditional on the number of steps when entering the step donation page. We find that the effect of attempting step donation promptly is generally significant only when a user's number of steps exceeds 10k (Figure 5-22). This result indicates that missing the deadline has the largest effect for those who were qualified to donate but failed because of missing the deadline.

5.5 Discussion

Understanding incentives that encourage exercises has numerous social impacts. Our study investigates the under-studied incentive for physical activity, i.e., prosocial incentives. We employ mixed methods by combining findings of a large-scale experiment based on encouragement design and observational studies. From the experiment, our encouragement design through mobile messaging promotes users to experience the prosocial incentive. Our analysis suggests that prosocial incentives appear to promote more physical activity for users who are likely to be encouraged to experience the prosocial incentive. We do not find significant effects for the whole population, possibly due to the lack of statistical power or weak instruments (i.e., the mobile messaging can encourage a small proportion of users to experience step donation).

To complement this experiment, we perform an observational study with a high-dimensional matching strategy. We find that consistent with the experimental result, users who engaged in step donation are likely to walk more steps. Such an effect is as strong as the effect of the social incentive resulting from the ranking page. In addition, the result of a regression discontinuity design indicates that impure altruism drives some users to engage in the prosocial incentive, but we do not find evidence that the impurely altruistic motivation also drives the increase of step counts.

Our study does have limitations. First, as many other empirical studies do, the external validity might be a concern. However, as WeChat and WeRun users represent

a large and relatively comprehensive population in China, we believe our results at least reflect a nationwide pattern. We acknowledge the design of the step donation and ranking pages, such as the 10k threshold and the quota and deadline in the daily step donation, may generate platform-specific conclusions. We leave it as an open question about how those designs would affect user behavior. Second, although we perform a high-dimensional matching design and carefully examine the matching quality, we may still suffer from selection bias. However, we believe that the consistent results between the experiment and the matching provide us with more confidence.

5.6 Appendix

Supplemental information for the experiment

In the main text, we mentioned that the table displayed was simply English translation. Moreover, within each treatment group, we slightly change the text to reduce the effect that users' behavior changes according to uninteresting factors such as text length. Table 5.6 provides the details of the notifications that users receive. The English version is translated by the authors.

Table 5.6: Original Chinese text users receive in the experimental design.

Treatment group name	Proportion	$\geq/<$ 10k	Chinese original text	English translation
donate_S	1%	$\geq$	可以捐步了	You can donate your steps
		$<$	满万步可以捐步	You can donate your steps if you reach 10k
donate_L	1%	$\geq$	已满万步，可以捐步了	You have reached 10k steps and so you can donate your steps
		$<$	满万步可以捐步	You can donate your steps if you reach 10k
donate_achieve_S	1%	$\geq$	捐步记录成就	You can donate your steps to mark your achievement
		$<$	满万步可捐步记录成就	You can donate your steps to mark your achievement if you reach 10k
donate_achieve_L	1%	$\geq$	已满万步，捐步记录成就	You have reached 10k steps and so you can donate your steps to mark your achievement
		$<$	满万步可捐步记录成就	You can donate your steps to mark your achievement if you reach 10k
donate_charity_S	1%	$\geq$	捐步助力公益	You can donate your steps to help charities
		$<$	满万步可捐步助力公益	You can donate your steps to help charities if you reach 10k
donate_charity_L	1%	$\geq$	已满万步，捐步助力公益	You have reached 10k steps and so you can donate your steps to help charities
		$<$	满万步可捐步助力公益	You can donate your steps to help charities if you reach 10k
ranking_S	1%	$\geq$	查看好友排行	Check your step count ranking among your friends
		$<$	查看好友排行	Check your step count ranking among your friends
ranking_L	1%	$\geq$	已满万步，查看好友排行	You have reached 10k steps. Check your step count ranking among your friends
		$<$	未满万步，查看好友排行	You have not reached 10k steps. Check your step count ranking among your friends
tomorrow_S	1%	$\geq$	明天继续加油	Keep on it tomorrow
		$<$	明天继续加油	Keep on it tomorrow
tomorrow_L	1%	$\geq$	已满万步，明天继续加油	You have reached 10k steps. Keep on it tomorrow
		$<$	未满万步，明天继续加油	You have not reached 10k steps. Keep on it tomorrow
default (control)	2%	$\geq$	XXX 获得今日冠军	XXX champion today
		$<$	XXX 获得今日冠军	XXX champion today

Note: "Proportion" refers to the proportion of users who were randomly assigned to a certain treatment group among all WeRun users. The messages received were affected by whether the user's daily step count is no less than 10k by the time the routinely message was sent.

Table 5.7: Results of t tests for the balance check of the covariates in the experiment. p values are unadjusted for multiple hypothesis testing.

Comparison	Statistic	Female ratio	Average steps	Step donation	Step ranking
donate v. default	mean diff	-0.047%	-1.00	-0.0009	-0.001
	S.E.	0.027%	2.77	0.0005	0.006
	p value	0.0782	0.7178	0.0502	0.8820
donate_achieve v. default	mean diff	-0.005%	-3.05	-0.0010	0.001
	S.E.	0.027%	2.77	0.0005	0.006
	p value	0.8454	0.2705	**0.0323**	0.9037
donate_charity v. default	mean diff	-0.013%	0.21	-0.0007	0.005
	S.E.	0.027%	2.77	0.0005	0.006
	p value	0.6414	0.9382	0.1687	0.3844
ranking v. default	mean diff	-0.023%	-1.16	0.0004	0.004
	S.E.	0.027%	2.77	0.0005	0.006
	p value	0.3843	0.6743	0.4544	0.5114
tomorrow v. default	mean diff	-0.005%	-3.73	-0.0006	-0.002
	S.E.	0.027%	2.77	0.0005	0.006
	p value	0.8528	0.1777	0.2041	0.7773

Balance check. We conduct a balance check of the experiment for the statistics in Table 5.2 in the main text. We compare the mean of the treatment group versus the mean of the control group. Among the 4×5 comparisons, only one of them is significant at the 5% level – donate_achieve v. default for step donation ($p = 0.0323$). However, after we apply the Benjamini-Hochberg procedure [40] to account for false discovery, the smallest p value in 25 tests should be $0.05/20 \times 1 = 0.0025$ to be significant. Therefore, we consider the p value of 0.0323 as false positive.

Step donation and ranking effects. The main text presents the treatment effects of different treatments on viewing the step donation page or the ranking page. To illustrate the effect sizes on the relative scale, we present the means of the outcome variables in Figure 5-13.

Step counts. First, we also present the means of the step counts (or the dummy indicating its being greater than 10k) for each treatment group in Figure 5-14. This helps show the effect size on the relative scale. We cannot observe economically sizable effects from these figures.

As discussed in the main text, we examine heterogeneity of the effect size given users' history in step donation or step counts. Results are presented in Figure 5-15. As shown in the figure, we do not observe significant effects across different subgroups. The occasional significant effects may result from multiple hypothesis testing or imperfect random assignments.

In addition, we perform variance reduction by adding covariates into the regression. The covariates include the demographic features and user donation history and

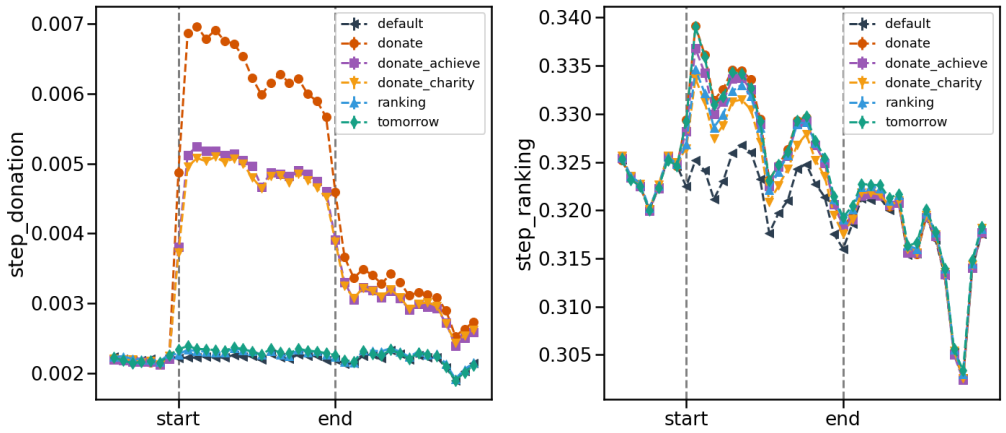


Figure 5-13: Means of on `step_donation` and `step_ranking` for different treatment groups. The experiment lasted for 17 days. Each error bar indicates the mean for the day. We also report the effects before and after this period. Error bars are standard errors.

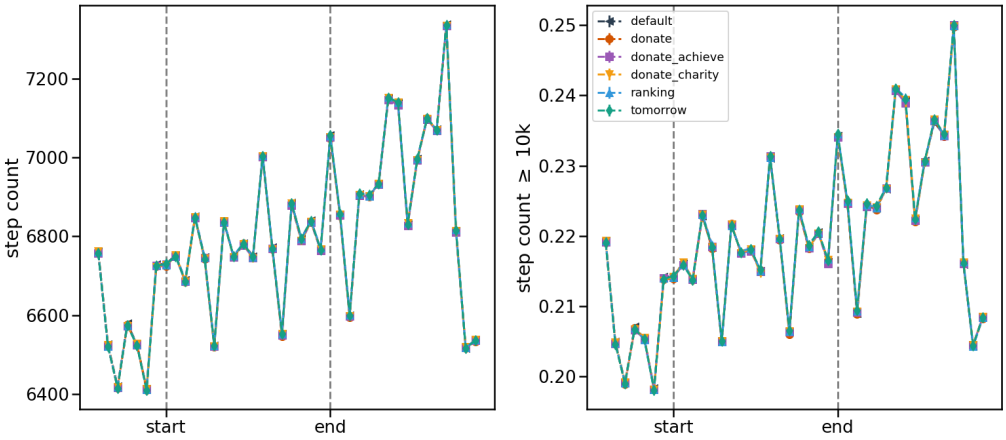


Figure 5-14: Means of step counts and the corresponding dummy variable for different treatment groups. The experiment lasted for 17 days. Each error bar indicates the mean for the day. We also report the effects before and after this period. Error bars are standard errors.

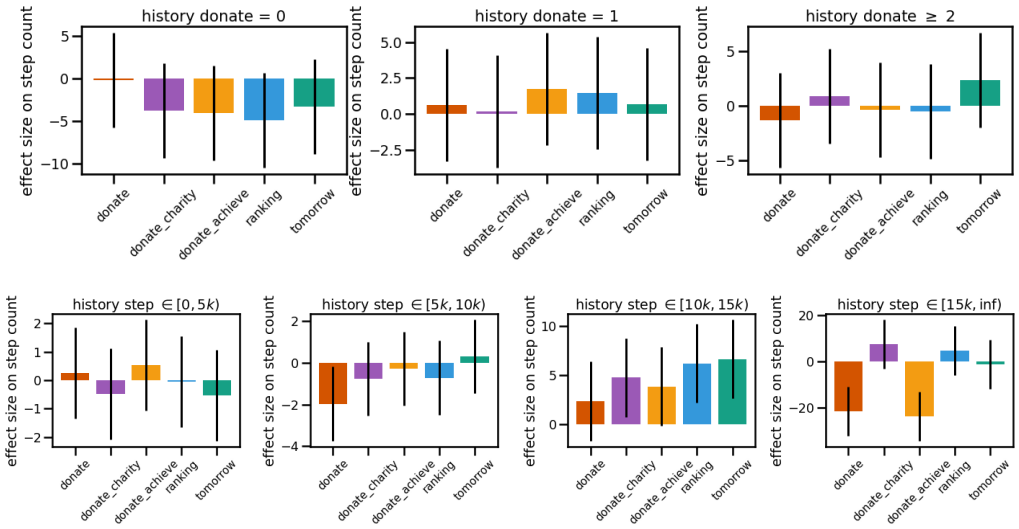


Figure 5-15: Heterogeneity of the effectiveness of different treatment messages on step counts. The upper panels are conditional on different step donation history, and the lower panels are conditional on different historical average steps. Error bars are 95% CIs.

historical average step count in the past 29 days. The results are plotted in Figure 5-16. Although the standard errors of the treatment effects have been reduced, we do not observe clear patterns of significant effects.

In the main text, we perform an instrumental variable design to test the effect of viewing the step donation page. Here we use the variable indicating viewing the ranking page as the endogenous variable. We only use users assigned `tomorrow`, `ranking`, or `default` in the regression. The results are presented in Table 5.8. We also create the stratum in a different way, instead of stratifying the step donation history, we stratify step ranking check history. As shown in the table, we do not observe significant effects of the endogenous variable – check the ranking page.

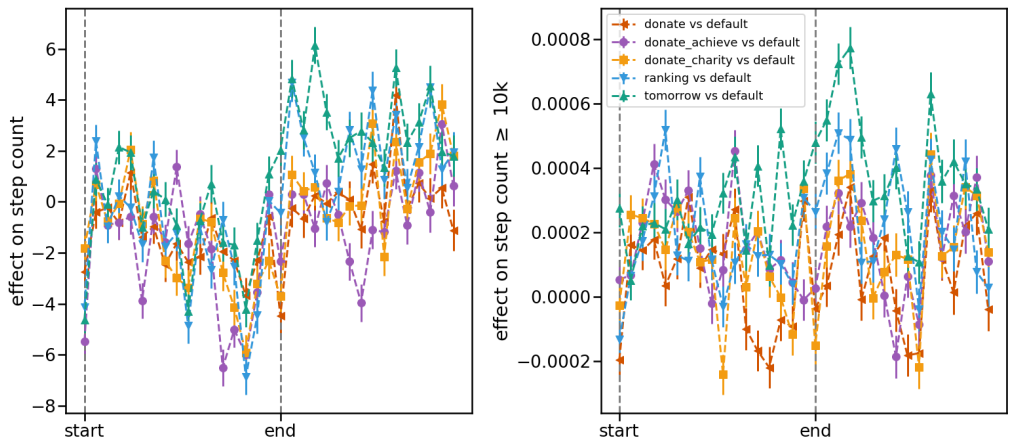


Figure 5-16: Daily treatment effects after controlling for covariates. Error bars are 95%.

Table 5.8: Regression results of the instrumental variable design for the effect of checking the ranking page.

	simple	**polynomial**	**stratum**
	(1)	(2)	(3)
Dependent variable: step count on day 2			
check ranking day 1	-206.88	208.42	-253.38
	(218.87)	(210.97)	(191.64)
First-stage F statistic	459.39	32.81	3.14
Adjusted R^2	0.58	0.58	0.58
No. instruments	4	101	101
No. observations	20,666,068	20,666,068	20,666,068
Dependent variable: step count $\geq$ 10k on day 2			
check ranking day 1	-0.009	0.006	-0.037**
	(0.022)	(0.019)	(0.018)
First-stage F statistic	459.39	32.81	3.14
Adjusted R^2	0.35	0.36	0.37
No. instruments	4	101	101
No. observations	20,666,068	20,666,068	20,666,068

*Note: Robust standard errors are in parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.*

Table 5.9: Robustness check for changing the number of matched observations.

Matching approach	`prosocial & social`	`prosocial`	`social`
Dependent variable: average outcome (step count)			
1:1 matching	10396.49	10035.13	9993.00
	(29.61)	(28.21)	(28.04)
1:3 matching	10396.49	10024.74	9964.68
	(29.61)	(16.16)	(16.04)
1:5 matching	10396.49	10011.27	9953.83
	(29.61)	(12.49)	(12.40)
Dependent variable: average outcome (step count $\geq$ 10k)			
1:1 matching	0.4596	0.4348	0.4271
	(0.0021)	(0.0021)	(0.0021)
1:3 matching	0.4596	0.4351	0.4269
	(0.0021)	(0.0012)	(0.0012)
1:5 matching	0.4596	0.4341	0.4261
	(0.0021)	(0.0009)	(0.0009)

Supplemental information for matching

List of covariates. We employ the following covariates in the matching. (1) Time features: the day when the user engaged in this and last step donation. (2) User information: gender, economic level of the city, friend count, and message count. (3) WeRun history: the step count, the views of step ranking page, and the number likes received on day d, $d-1$, $d-2$, $d-3$ the averages from $d-7$ to $d-4$, $d-14$ to $d-8$, $d-21$ to $d-15$, $d-27$ to $d-22$, $d-35$ to $d-28$, and $d-42$ to $d-36$. In total, we have 37 covariates.

One-to-many matching. In the main text, we mainly employ one-to-one nearest matching because of the imbalance in sample sizes between the treatment group and the two control groups. This can help us to match the best quality observation, but matching to too few observations may increase the variance. Here we instead employ one-to-three matching and one-to-five matching. The results are presented in Table 5.9. We do not observe noticeable changes in the effect sizes, but one-to-many matching can reduce the variance of the means of the outcome variables.

Matching quality. To demonstrate the robustness of the matching results, we filter out matched pairs with the largest distances in the covariate space. Specifically, we filtered out the matched pairs with the top 0%, 20%, 40%, 60%, 80% matching distances. Because in general the match quality is worse for users whose step count is greater, we tend to filter out physically active users. We thus analyze four subgroups – conditional on users' step counts on the current day – separately. As shown in

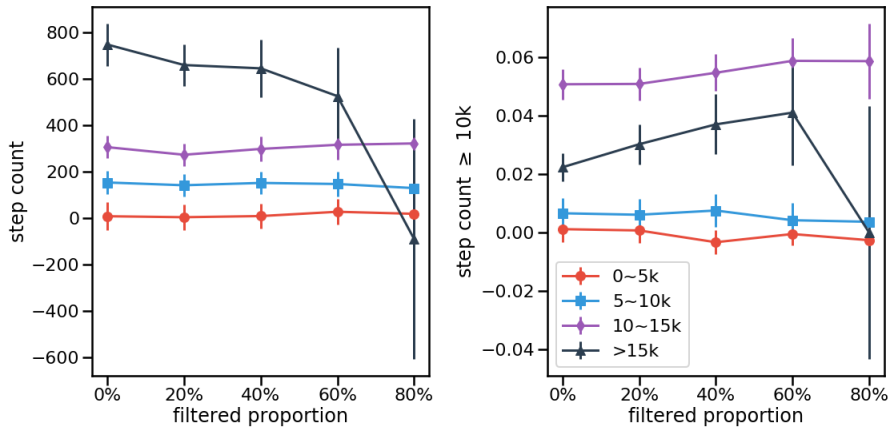


Figure 5-17: Matching quality check when we compare the `prosocial & social` group versus the `social` group.

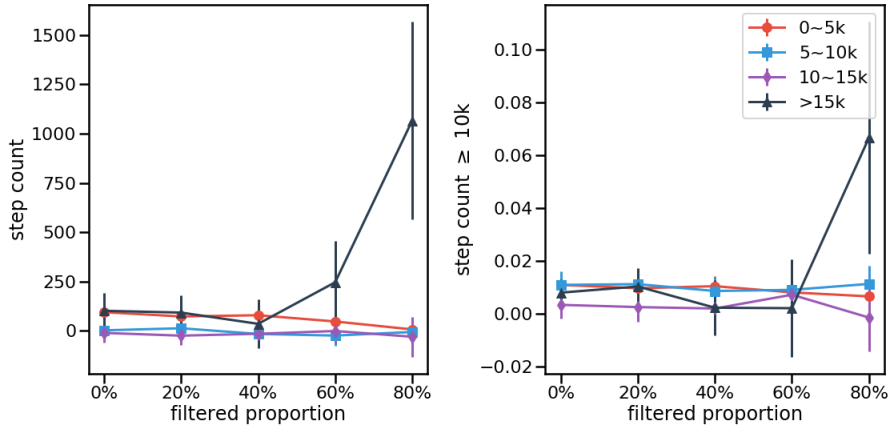


Figure 5-18: Matching quality check when we compare the `social` group versus the `null` group.

Figures 5-17 and 5-18, filtering out low-quality matched pairs do not affect the effects for the first three subgroups. It does appear to affect the $\geq$ 15k subgroup when we filter too much proportion. This is because when we filter out 80%, most observations in the $\geq$ 15k subgroup would be removed, which leads to a large variance.

Variance reduction. We first examine the robustness of the results by adding covariates to regressions, which improve the precision of our estimation. We present the coefficients from regressions with or without the control for covariates in Table 5.10. Covariates are identical to the ones used for matching.

We find that after controlling for covariates, the effects sizes of both `prosocial`

`& social` and `social` incentives on step counts are reduced but remain statistically significant (Columns (1) and (2)). Controlling for covariates does not heavily affect the regression coefficients when the dependent variable is the step donation next day (Column (3)) but reduces the effect size when we examine checking the ranking page as the outcome. This further verifies the robustness of our main results.

Table 5.10: Regression coefficients after controlling for covariates.

	Dependent variable:			
	step count next day (1)	step count $\geq$10k next day (2)	step donate next day (3)	check ranking next day (4)
Prosocial&Social	429.25***	0.0338***	0.0367***	0.4200***
	(41.752)	(0.003)	(0.001)	(0.003)
Social	116.77***	0.0123***	0.0060***	0.4950***
	(41.339)	(0.003)	(0.000)	(0.003)
Covariate control	N	N	N	N
Observations	166,206	166,206	166,206	166,206
Adjusted R^2	0.001	0.001	0.016	0.195
Prosocial&Social	285.66***	0.0231***	0.0347***	0.3422***
	(31.223)	(0.003)	(0.001)	(0.003)
Social	88.71***	0.0077***	0.0043***	0.4048***
	(30.22)	(0.003)	(0.001)	(0.003)
Covariate control	Y	Y	Y	Y
Observations	166,206	166,206	166,206	166,206
Adjusted R^2	0.513	0.289	0.020	0.262

Propensity score matching. We also employ propensity score matching [208] to perform the analysis. We first conduct a Logistic regression to estimate the propensity score for each observation. We then match each observation in the `prosocial & social` group with an observation in the `social` group with the closest propensity score. The regression results are presented in Table 5.11. We find that adding covariates drastically changes the regression coefficients, indicating that propensity score matching may not produce high quality results. This can result from the common issues of propensity score matching [142]. Therefore, we should not trust the results of propensity score matching.

More outcome variables. To be consistent with the analysis of the experiment, we use the engagement of the step donation page and the ranking page as the outcomes and perform the same matching procedure. The results are presented in Figure 5-19.

As expected, as shown in the left panel of Figure 5-19, users who viewed the step donation page (`prosocial & social` group) are more likely to be engaged in the step donation on the next day than otherwise (`social` group or `null` group). This result suggests that experiencing the step donation feature once promotes the person to do

Table 5.11: Regression coefficients after controlling for covariates with propensity score matching.

	Dependent variable:			
	step count next day (1)	step count ≥10k next day (2)	step donate next day (3)	platform engagement next day (4)
Prosocial&Social	4011.5888***	0.4583***	0.0387***	-0.2976***
	(34.463)	(0.002)	(0.001)	(0.004)
Social	4211.8033***	0.4618***	0.0000	-0.2278
	(34.463)	(0.003)	(0.000)	(0.004)
Covariate control	N	N	N	N
Observations	166,206	166,206	166,206	166,206
Adjusted R^2	0.103	0.221	0.026	0.110
Prosocial&Social	289.0539***	0.0772***	0.0313***	0.3629***
	(289.193)	(0.022)	(0.002)	(0.028)
Social	71.6937	0.0577***	-0.0073***	0.4160***
	(236.590)	(0.022)	(0.002)	(0.028)
Covariate control	Y	Y	Y	Y
Observations	166,206	166,206	166,206	166,206
Adjusted R^2	0.582	0.446	0.030	0.206

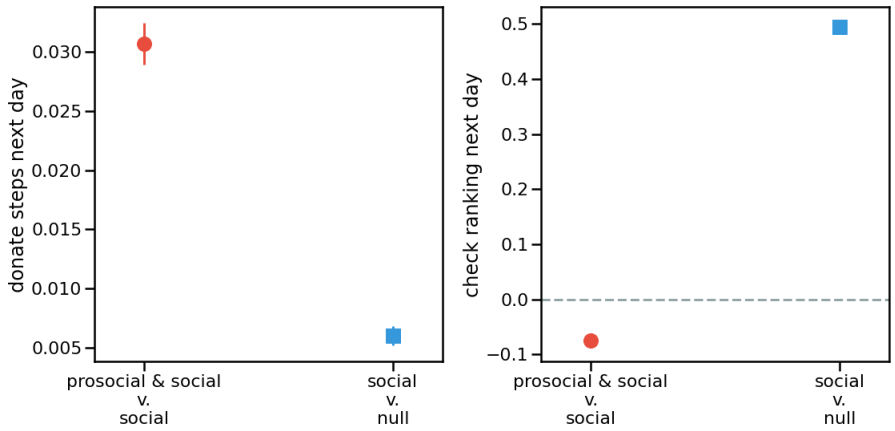


Figure 5-19: Averages of outcome variables in the `prosocial & social` group, the matched `social` group, and the matched `null` group. Error bars are 95% standard errors.

it over again.

As shown in the right panel of Figure 5-19, users who viewed the ranking page (`prosocial & social` and `social` group) are much more likely to view the step ranking page on the next day. We observe a greater likelihood in the `social` group of viewing the ranking page on the next day than in the `prosocial & social` group. We conjecture that this is because in the `prosocial & social` group, there are users who are only interested in the step donation feature but uninterested in the step ranking feature.

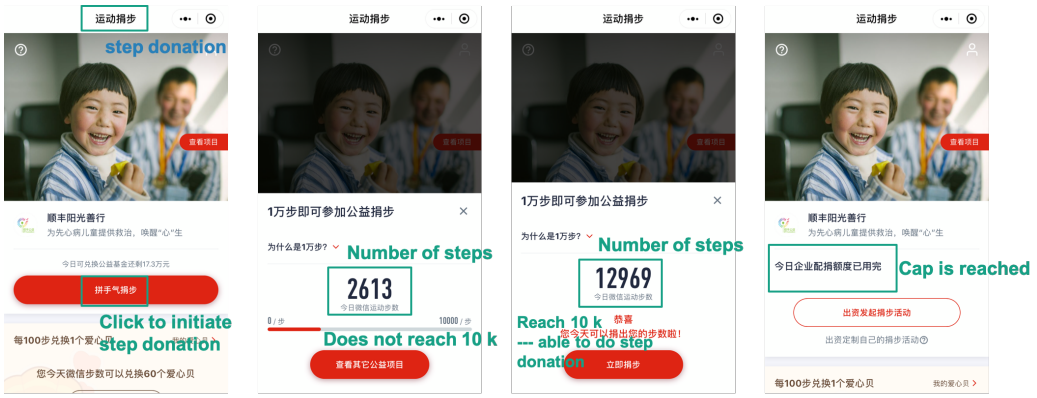


Figure 5-20: Illustration of the step donation.

Supplemental information for RDD

Details of step donation and deadline. Figure 5-20 illustrates the process of engaging in step donation. Users may enter the step donation mini program on WeChat at any time. After seeing the Panel (a), a user may click the red button in the middle to attempt a step donation. If her number of steps is smaller than ten thousand, the platform will inform the user that she is not qualified for today (Panel (b)). If her number of steps achieves ten thousand, the platform will inform the user that she is qualified and may go further to do the step donation (Panel (c)). However, if the user arrives late, i.e., after the capping time, the platform will inform the user that today's quota is capped (Panel (d)). Note that users' step donation behavior will not be observed by any other user. There is no badge design or any other reputation gain or rewards attached to it.

Density Test A common way to support the validity of a regression discontinuity design is the density test. That is, if the density curve of the sample around the discontinuity point does not change smoothly, this curve suggests that some agents may be able to deliberately place themselves before or after the discontinuity point, resulting in non-homogeneity between treatment and control groups.

Figure 5-21 shows the density curve to be smooth around the discontinuity point, with $p = 0.88$ from the McCrary Test [174]. This result supports the validity of our regression discontinuity design.

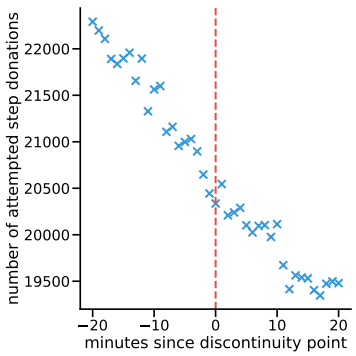


Figure 5-21: Density check of our regression discontinuity design. We do not find a clear discontinuity at 0, suggesting that users cannot well predict the capping time and our regression discontinuity design is valid.

By changing the outcome variable to covariates including demographical features such as gender, click history and step history, we also find that important covariates do not change abruptly around the discontinuity point. Thus, comparisons between samples before and after the discontinuity point would provide us trustworthy causal evidence.

We primarily use the Imbens-Kalyanaraman optimal bandwidth but also test the robustness of our results with a half of or double the optimal bandwidth. For the three outcome variables, we derive $h^{\text{opt}} = 1693, 2040$, and 2239, respectively. These results indicate that our choice of the bandwidth is sensible.

Heterogeneity. We further examine the treatment effects conditional on the number of steps when entering the step donation page. We find that the effect of attempting step donation promptly is generally significant only when a user's number of steps does not reach 10k. For the probability of attempting step donation, we find that the effect is in general significant when the number of steps is smaller than 17.5k steps. When the number is larger, the effect becomes non-significant, partially because their probability of exceeding 10k steps with or without missing the deadline is already large enough.

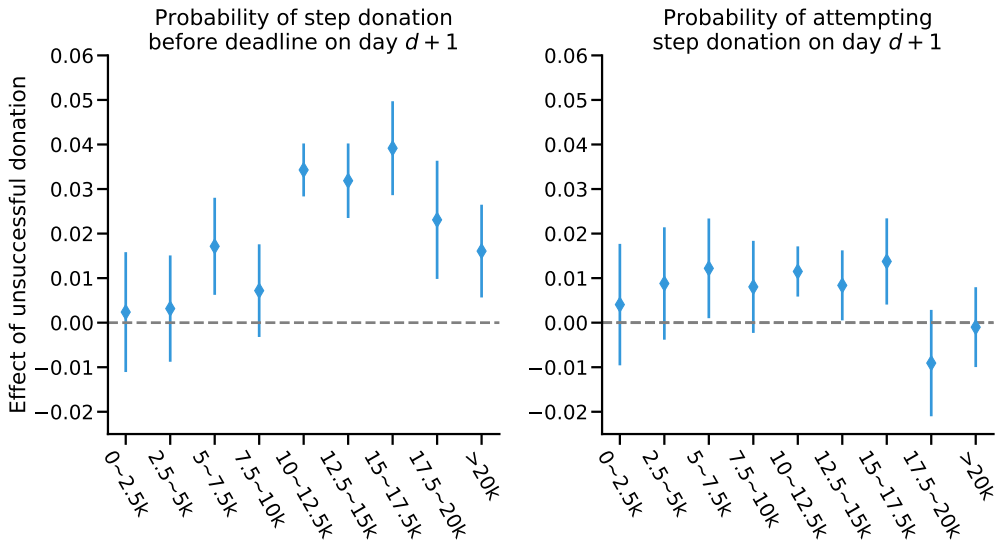


Figure 5-22: Heterogeneous treatment effects conditional on the number of steps when entering the step donation page. Error bars are 95% CIs.

Chapter 6

Mitigating network interference with network motifs and machine learning

6.1 Background

Randomized control trials, or "A/B tests", have been crucial to understanding the impact of an intervention, such as a new policy [119, 114], product intervention [20, 29], or medical treatment [16, 167]. Randomized control trials can estimate the causal effect of a treatment intervention by ensuring that treatment and control assignments are independent of other variables. Increasingly, causal inference methods have had to adapt to modern A/B test settings where there are high-dimensional features [216, 204, 87, 83], computational and algorithmic considerations [262], and network interference concerns [17, 84, 215].

For addressing interference, traditional causal inference methods rely on a critical assumption called the "stable unit treatment value assumption" (SUTVA) [92, 227]. SUTVA is in fact a strong assumption, requiring that a unit's outcome is only affected by its own assignment conditions, regardless of the assignment conditions of all other observations. However, this can be an unrealistic assumption in many settings, including social networks, where user's are influenced by one another [200, 48, 24]. We refer to this as *network interference* [17].

An increasing number of approaches are aimed at dealing with network inter-

ference [207, 240, 248, 149, 265, 22, 36, 215, 128]. Existing methods for addressing networked interference can be categorized into two main strategies. The first one is to improve the random assignment strategy. Cluster random assignment treats observations at the level in which observations have strong interdependence (e.g., assigning on the class or city level) [207, 106, 36]. Graph cluster randomization is a special case for social networks [248, 84]. It first runs a graph clustering algorithm and then randomizes treatment assignment on the graph cluster level. Recent work proposes random clustering instead of a fixed clustering [249]. Another type of approach, including ours, relaxes SUTVA by allowing the potential outcome to be defined as a function of the assignment conditions of the ego observation and its (n-hop) neighbors [251, 159]. However, few studies have utilized the network structure in the n-hop neighbors (i.e. local network structure) to further characterize different interference conditions.

From the empirical end, many social network studies have highlighted why local structures should be considered to address network interference. For example, the structural diversity hypothesis [247, 233] claims the likelihood of product adoption is largely dependent on the degree to which a unit's neighbors who have adopted are disconnected. By contrast, complex contagion theory [57, 56] or the "echo chamber" effect [30, 94], suggest an individual is most likely to adopt a behavior when she is clustered in the network of multiple neighbors who have adopted this behavior. Figure 6-1 illustrates four examples of network interference that could be captured by the local networks structure and treatment assignment. The first two conditions are simply the cases where all neighbors are treated or non-treated, followed by the important network interference conditions suggested by structural diversity and complex contagion, respectively. In the case of structural diversity and echo chamber settings, the ego node in (c) and (d) has 1/2 neighbors treated but exhibit very different local structures and the ego's outcome may be different in these settings; and we do not know which one is the dominant factor that drives most of the variance in the outcome.

Our study provides a tool for experimenters and practitioners to account for impor-

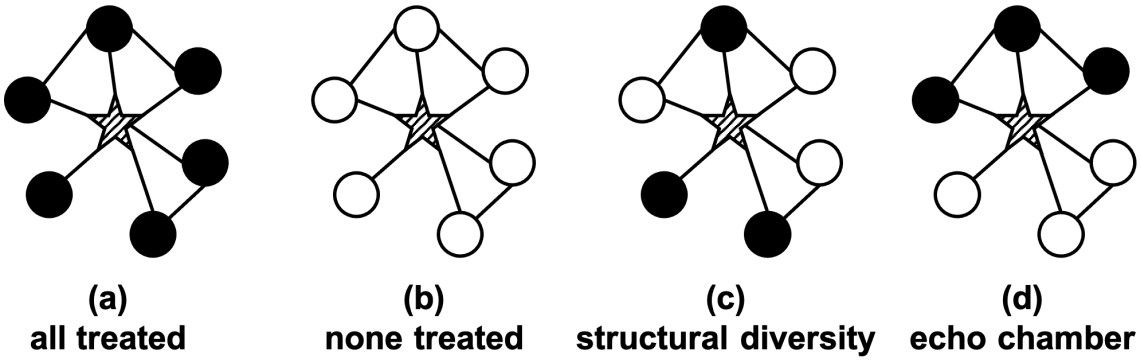


Figure 6-1: **Examples of network interference conditions across different local network structures.** The star indicates a user and a circle represents a user's friends. Solid circles indicate that a friend is in treatment and hollow circles indicate a friend is in control. For stars, the shaded indicates that it could be treated or control.

tant network interference conditions without necessarily specifying a particular dominant social science theory. We rely on network motifs [180, 11] (or graphlets [213]) to characterize both the local structure and treatment assignment conditions to account for the different types of (network interference) exposure conditions. We can think of network motifs as "label-independent structure" since the features do not depend on treatment labels and network motifs with treatment assignment conditions as "label-dependent structure" since they depend on labels [103]. We refer to network motifs with treatment assignment labels as *causal network motifs*. Prior work has considered label-dependent network motifs for use in protein function prediction applications [61] and has developed computationally tractable ways for detecting labeled motifs [203]. Another example is [23], which uses network features such as transitivity in a propensity score matching framework, with a focus on observational data. Our proposal is distinct in that we focus on experimental settings where causal network motifs can quantify the causal impact of different interference conditions.

In this study, we provide an approach to automatically categorize network interference conditions based on local network neighborhood structures. We develop our approach based on the framework proposed by [22], where the authors regard different network interference conditions[1] as unique treatments. However, it is unclear

[1]In this study, we use exposure conditions and network interference conditions interchangeably.

how experimenters or practitioners should a priori define different network interference conditions that are suitable for specific data and experiment. Moreover, manually determining different network interference conditions may make more severe the issue of a large number of researcher degrees of freedom.

The following sections will describe our two-step solution to identify different network interference conditions. First, we construct network motifs with assignment labels, referred to as *causal network motifs*, to characterize complex network interference conditions. In our study, we not only characterize the network structure by the number of different network motifs, but also characterize the assignment conditions in the local network neighborhood by the causal network motifs. Second, using these causal network motif features, we develop a tree-based algorithm to cluster different exposure conditions. Each leaf or terminal node in the tree corresponds to a network interference condition and can be used to estimate the average potential outcome given that condition. We cannot directly apply conventional decision trees or other machine learning algorithms, since we need to adjust those algorithms to address issues, such as selection bias and positivity as will be explained [260, 63, 64]. Moreover, as a common goal in causal inference, we aim to estimate the average potential outcome given an interference condition to quantify causal impacts, rather than predicting the potential outcome and the treatment effect for every observation [211].

6.2 Methods

6.2.1 Causal Inference Setup

Let i (or j) index individuals in a social network. $\mathcal{U}$ is the set of individuals in the population. $\mathcal{N}_i$ be the neighbor set of i, and this can be extended beyond immediate neighbors. Y_i is the observed potential outcome. $Z_i \in \{0, 1\}$ denotes the random treatment assignment for i. Whenever applicable, we use upper-case letters to represent variables or vectors that could be intervened by experimenters or affected by the

intervention; and we use lower-case letters for other variables that are not affected by the intervention (such as demographics). Sets are in calligraphy.

Potential outcomes framework for network interference

We start from the potential outcomes framework [211]: The potential outcomes framework defines that a unit's outcome is a function of assignment conditions.

$$Y_i = y_i(Z_i). \tag{6.1}$$

$y_i(1)$ and $y_i(0)$ are called potential outcomes and we only ever observe one of them. This function implies no interference, meaning that the potential outcome Y_i does not depend on the treatment assignment of other users. However, this is often an unrealistic assumption in network settings, since the unit's outcome can be dependent on the treatment assignment conditions of neighbors or even any other unit, through, for example, social contagion.

To overcome this unrealistic assumption, our approach is based on [22] which introduces *exposure mapping*. An exposure mapping is defined as a function that maps the treatment assignment vector to a value; the fraction of treated friends [248] is an example of a simple exposure mapping. Each exposure mapping corresponds to a condition that considers both treatment assignment conditions of both the ego node and her (n-hop) neighbor. Under this framework, researchers or practitioners usually need to define a priori a set of *network interference conditions* (also known as *exposure conditions*). We denote this set by $\mathcal{D}$. Then in this framework, the outcome variable is expressed by

$$Y_i = \sum_{d \in \mathcal{D}} y_i(d) \mathbb{1}[D_i = d]. \tag{6.2}$$

Each d corresponds to a pre-specified exposure condition (interference condition). D_i is a random variable indicating i's exposure condition. Analyzing different exposure conditions can offer insights into how observations react to different network interference settings, such as comparing (with selection bias corrected) users in control with

treated friends to users in control with only control friends.

Let us introduce three examples. First, SUTVA is a special case under this exposure mapping framework (where $\mathcal{D} = \{ego\_treated, ego\_control\}$). The second example is the existence of direct and indirect effects, as introduced in [22]: that $\mathcal{D} = \{no\_exposure, direct\_exposure, indirect\_exposure, direct + indirect\_exposure\}$. Direct effect means the ego or user is treated and indirect effect means any neighbor is treated; thus we have four conditions in total. In our study, we use direct effect to refer to the effect of changing Z_i from 0 to 1 on the outcome variable; and indirect effect refers to any consequence that result from the assignment conditions besides ego node i, i.e. Z_{-i}.[2] The third example, the k (or q-fractional) neighborhood conditions in [248] fits this framework, with (2×2) four different exposure conditions — whether the ego node is treated $\times$ whether more than k (or q-fractional) of her neighbors are assigned to the same treatment condition with the ego node.

Distinguishing between correlation and causation

The indirect effect may contain a large degree of heterogeneity based on how many (n-hop) neighbors are treated and how they are connected. Some of these conditions are theorized in [57, 247] while other cannot. It is usually challenging to distinguish between what is the causal impact of certain exposure conditions, versus what is usually confounded by ego's local network structure (for example, users with more friends tend to have their friends less clustered) [233]. In other words, selection bias may result from simply taking average over observations within a certain interference condition. For example, imagine that we want to quantify the impact of an indirect effect, so simply taking the average over observations who have at least one treated friend versus taking average over observations with no treated friends. This average may be biased towards observations who have more friends. We illustrate this selection bias issue in Figure 6-4.

[2]Indirect effects will be further specified by $\mathbf{X}_i$ in later sections.

Mathematically, our goal is to estimate the average potential outcome

$$\bar{y}(d) = \frac{1}{|\mathcal{U}|} \sum_i y_i(d), \text{ for all } d \in \mathcal{D}. \tag{6.3}$$

This is one of the core goals proposed in [22]. To correct the aforementioned selection bias, we employ inverse probability weighting, such as Horvitz–Thompson estimator and Hajek estimator [218] Although the Horvitz–Thompson estimator is an unbiased estimator for the average potential outcome, it empirically has unaffordably high variance. Therefore, following [84, 22], we use the Hajek estimator to estimate the average potential outcome for exposure condition d since its small bias can usually be ignored in a large sample:

$$\hat{\bar{y}}_{Hajek}(d) = \frac{\sum_{i\in\mathcal{U}} \frac{1_i(d) y_i(d)}{\pi_i(d)}}{\sum_{i\in\mathcal{U}} \frac{1_i(d)}{\pi_i(d)}} = \frac{\sum_{i\in\mathcal{U}, D_i=d} \frac{y_i(d)}{\pi_i(d)}}{\sum_{i\in\mathcal{U}, D_i=d} \frac{1}{\pi_i(d)}}, \tag{6.4}$$

where we denote $\pi_i(d) = \mathbb{P}[D_i = d]$ and $1_i(d) = \mathbb{1}[D_i = d]$. $\pi_i(d)$ is the *inclusion probability* of the exposure condition of i being d, and is a generalization of propensity scores [208]. It can also be understood as a weighted average over observations with $1_i(d) = 1$, where the weight is $1/\pi_i(d)$. Therefore, it can be estimated through weighted linear regressions: the coefficient for the constant is the Hajek estimator. Its variance can be estimated via Taylor linearization [214].

The probability $\pi_i(d)$ is often challenging to compute analytically. We therefore use Monte Carlo for sufficiently large replicates to approximate $\pi_i(d)$. Specifically, we re-run the treatment assignment procedure for R replicates to obtain the empirical distribution of $(Z_i, \mathbf{X_i})$.[3] Therefore, we can derive the estimated inclusion probability $\hat{\pi}_i(d)$ for any exposure condition d. For example, we can let $\hat{\pi}_i(d) = \frac{\sum_r \mathbb{1}[D_i^{(r)}=d]+1}{R+1}$; and we substitute the $\pi_i(d)$ in Equation 6.4 with $\hat{\pi}_i(d)$. The relative bias $\hat{\bar{y}}_{Hajek}(d)$ to $\bar{y}(d)$ diminishes exponentially as R increases. The details have been discussed in [22].

While the exposure mapping framework introduces a tractable way to take into

[3]It is challenging to derive this distribution analytically, especially when the random assignment is deployed on the individual level.

account network interference, it is an open question how we should then account for the treatment assignment conditions of the n-hop neighbors and their network connections in Figure 6-1. We hope to provide an approach that generates exposure conditions that are suitable for a given experiment and dataset, and that avoids manually defining exposure conditions a priori. Our study provides a two-step solution to automatically identify different exposure conditions while overcoming selection bias concerns, as will be explained in more detail in the next sections. First, for an A/B test on a network, we construct network motif features with treatment assignment conditions to provide a fine-grained characterization of the local network structure and potential interference conditions. Second, using the network motif characterization as input, we develop a tree-based algorithm to perform clustering and define the set $\mathcal{D}$ rather than allowing practitioners to explore that.

6.2.2 Causal Network Motifs

Network motifs are a way to characterize all patterns of smaller network features among a set of nodes [180]. We introduce "causal network motifs," which differ from conventional network motifs in two primary aspects. First, we focus on (1-hop) ego networks that include the ego node, with the methods generalizing to higher n-hop ego networks for $n>1$. Second, we consider the treatment assignment conditions of the user and their n-hop connections. We use the terminology "network motifs" to refer to conventional motifs without treatment assignment labels (or assignment conditions) and causal network motifs to refer to ones with assignment conditions. Examples of network motifs are illustrated in Figure 6-2. We use these counts on an n-hop ego network to characterize the exposure condition of each observation.

Experimenters and practitioners need to determine a priori the region of ego networks (the n for n-hop ego networks which means the path length between an ego node and another node is no greater than n) and the network motifs that matter. For example, Figure 6-2 specifies 1-hop ego network and uses the assignment conditions of dyads, triads, and open tetrads as features in the interference vector.

Using causal network motifs implies the following two assumptions:

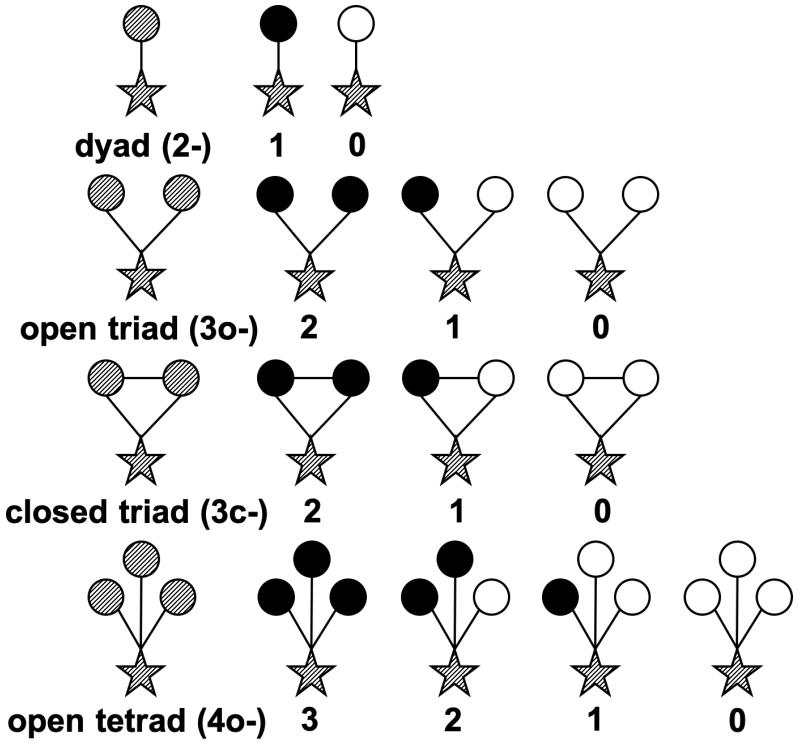


Figure 6-2: **Examples of causal network motifs.** Stars represent egos and circles represent alters. Solid indicates the node being treated, hollow indicates control, and shaded indicates that it could be treated or control. The first patterns in each row are conventional network motifs without assignment conditions, or just called *network motifs*, followed by corresponding network motifs. Our interference vector is constructed by dividing the count of a causal network motif by the count of the corresponding causal network motif. The labels below each network motif indicate the naming: for example, an open triad where one neighbor is treated is named 3o-1.

1. **(n-hop ego networks)** We assume that an ego node's outcome can only be affected by its own assignment condition and its (n-hop) neighbors' assignment conditions. This is a common assumption in the prior network interference literature, and is sometimes called *the stable unit treatment on neighborhood value assumption* [95, 159]. A larger n implies a more relaxed assumption.

2. **(Specified network motifs)** Given the specified n-hop ego networks and causal network motifs, we assume that only the assignment conditions of specified network motifs affect the outcome of the ego observation. This assumption implies that we do not distinguish two ego networks with identical counts of specified network motifs but with different network structures. Considering more and higher-order would mitigate this issue.

Ideally, we should consider a large n and network motifs with more nodes because these are more relaxed assumptions. However, two issues may arise. First, it is typically computationally expensive to count network motifs of many nodes in an ego network of large n. There are many possible network motif patterns and potentially large counts. Second, related to the positivity requirement in the next section, we need all (or almost all) observations to contain all specified network motifs. Specification of too many network motifs may exclude a significant proportion of samples from the analysis.

Note that although all our examples are in the undirected setting, it can be easily extended to directed networks (e.g., a directed edge from i to j indicates that i reached j). In either undirected or directed networks, we should be cautious that the networks are pre-treatment so that the network structure is not affected by the treatment assignments. For example, when we use a network where edges represent whether the two nodes have interactions, the interactions used should have happened before the treatment was assigned; otherwise this would bring about an issue known as post-treatment bias [15].

After counting causal network motifs for each ego node in our network, our next step is to convert the counts to features, which will be used in the next section. Let

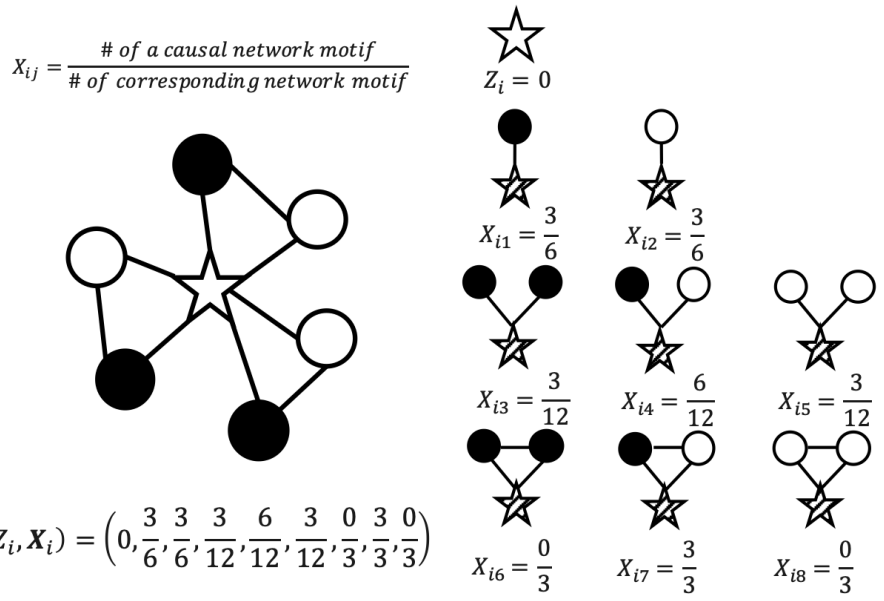


Figure 6-3: An example of ego network with treatment assignments and the corresponding interference vector. Stars represent egos and circles represent alters. Solid indicates the node being treated, hollow indicates control, and shaded indicates that it could be treated or control.

$\mathbf{X}_i$ denote an m-dimensional random vector, referred to as *interference vector*. The interference vector has an important requirement: Each element of the random vector is *"intervenable"* — that is, the random treatment assignment affects the value of each element of the vector. The requirement addresses the selection bias issue when we estimate the average potential outcomes.

We construct the interference vector in the following way. For each observation, for the count for each causal network motif (e.g., 2-1, 2-0, ..., 3o-2, 3o-1, ...), we normalize it by the count of the corresponding network motifs (e.g., dyads, open triads, closed triads, ...)[4]. In this way, each element of $\mathbf{X}_i$ is intervenable and the support for each element is in $[0, 1]$. Note that when considering a network motif with many nodes, some observations may not have certain network motifs, and normalization cannot be performed. In these scenarios, we can either exclude this network motif from the interference vector, or drop these observations if they take a really small proportion. Please refer to Figure 6-3 for an illustration of constructing the interference vector.

[4]The q-fractional neighborhood conditions are considered special cases in our approach where only dyad motifs are used.

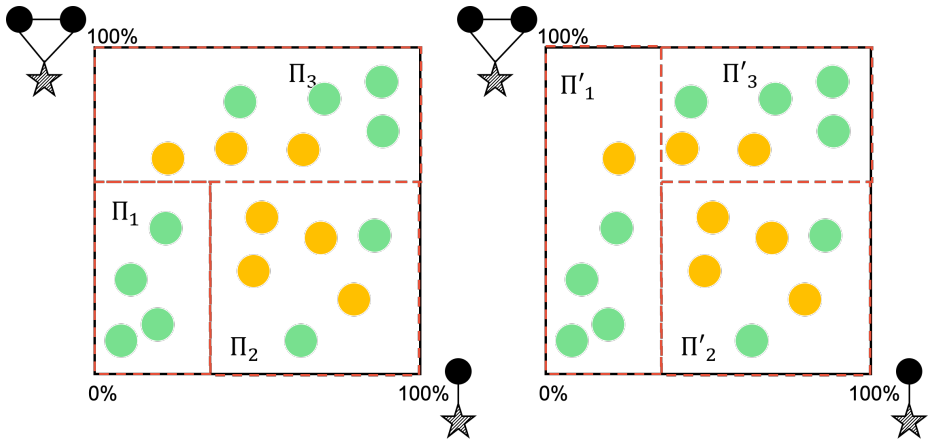


Figure 6-4: **Illustration of selection bias and positivity**. x-axis and y-axis represent the fraction of the given causal network motif among the corresponding network motifs, assuming that we only specify these two causal network motifs. The positions of nodes indicate observed values for each observation (not the probability distribution). Π_1, Π_2, Π_3 or Π'_1, Π'_2, Π'_3 represent a plausible partitioning. Imagine green and yellow nodes represent two types of observations (e.g. green for observations with fewer neighbors and yellow for observations with more neighbors). In independent random assignments, green nodes are more likely to have extreme values in the x- or y- axis, while yellow nodes are more likely to be centered around the mean. The left partitioning may violate positivity because yellow nodes may have zero or very small probability to belong in Π_1; by contrast, the right partitioning is feasible. In the right partitioning, simply taking the average is still problematic because yellow nodes have a smaller probability to belong in Π'_i. Therefore, we need inverse probability weighting (e.g., Hajek estimator) to correct this selection bias.

We combine the ego node's own assignment condition Z_i and interference vector $\mathbf{X}_i$ as the features for our tree-based algorithm described in the next section to determine exposure conditions. $(Z_i, \mathbf{X}_i) \in [0,1]^{m+1}$. Related to [22], our approach is mathematically equivalent to, by an abuse of notation, $D = f(Z_i, \mathbf{X}_i)$ such that $\bar{y}(D) = \bar{y}(f(Z_i, \mathbf{X}_i))$. $\mathcal{D}$ is equivalent to partitioning $[0,1]^{m+1}$ to $\mathcal{X}_1 \cup \mathcal{X}_2 \cup ... \cup \mathcal{X}_{|\mathcal{D}|}$. Z_i impacts direct effects, and $\mathbf{X}_i$ corresponds to indirect effects.

6.2.3 A Tree-Based Partitioning Approach

The next question is how to design an algorithm to determine exposure conditions by $(Z_i, \mathbf{X}_i)$ (i.e., partitioning its support $[0,1]^{m+1}$). By an abuse of notation, we replace any exposure condition d by the corresponding partition $\mathcal{X}$ (where $\mathcal{X} \subset [0,1]^{m+1}$): $\pi_i(\mathcal{X}) = \mathbb{P}[(Z_i, \mathbf{X}_i) \in \mathcal{X}] = \mathbb{P}[f(Z_i, \mathbf{X}_i) = d] = \pi_i(d)$, and similarly

$1_i(\mathcal{X}) = \mathbb{1}[(Z_i, \mathbf{X}_i) \in \mathcal{X}] = 1_i(d)$, or the potential outcome $y_i(\mathcal{X}) = y_i(d)$.

Our approach partitions $[0,1]^{m+1}$ and determines exposure conditions based on a decision tree regression [42].[5] Decision trees can be used for clustering [164] and typically have good interpretability in the decision making process [202]. Thus, it is a proper machine learning algorithm to solve the partitioning problem. Each leaf of the decision tree corresponds to a unique exposure condition (partition). Compared with conventional decision tree regression, we need to have the following revisions:

1. **(Positivity)** Positivity ensures observations have a non-zero chance of being in an exposure condition [260]. In our setting, it means that for all $i \in \mathcal{U}$, $d \in \mathcal{D}$, $\pi_i(\mathcal{X}) > 0$. If any partition d would lead to the existence of any observation i that $\pi_i(d) = 0$, we would be unable to estimate the average potential outcomes. Mathematically, it would make the denominator in Equation 6.4 zero. Note that the requirement is to set all observations ($\mathcal{U}$), rather than just the observations that are randomly assigned to $\mathcal{X}$.

   This is also part of the reason why we normalize the interference vector. Imagine we use the number of each causal network motif as the elements of the interference vector. If we decided a partition in which the number of 3c-2 (closed triads with fully treated neighbors) is greater than 10, then all observations with fewer than 10 open triads would have zero probability of belonging to this partition. The necessity of positivity is illustrated in Figure 6-4.

   The tree algorithm should not split a node if such splitting will lead to any child node corresponding to an exposure condition in which any observation has zero probability to belong. Since $\pi_i(\mathcal{X})$ is sometimes not solvable analytically, we use Monte Carlo to approximate it ($\hat{\pi}_i(\mathcal{X})$) as mentioned previously for the estimation of ($\hat{\pi}_i(d)$). Moreover, we adjust the positivity requirement to non-trivial probability, which allows a very few observations to have zero or near-zero probability; non-zero δ and ϵ introduce a small bias but allow partitioning more features.

[5] We only illustrate regressions because they can be generalized to binary outcome variables, but our approach could be easily extended to classification.

$$\sum_{i \in \mathcal{U}} \mathbb{1}[\hat{\pi}_i(\mathcal{X}) \leq \epsilon] \leq \delta |\mathcal{U}|. \tag{6.5}$$

It means that the fraction of observations with $\hat{\pi}_i(\mathcal{X}) \leq \epsilon$ is smaller than δ.

2. **(Honest splitting)** One pitfall is that when we estimate the variance, the algorithm is choosing a threshold to partition so that it would minimize its objective function; however, it may overfit the training data, selecting an improper threshold when splitting, and eventually overestimating the difference between the average potential outcomes indicated by the two child nodes. We thus split the original training set into training and estimation sets — the training set is used for tree partitioning and a separate estimation set is used for estimating the mean and variance. This is a common approach to correcting confidence intervals when using machine learning for causal inference [25, 151].

3. **(Weighted sum of squared errors)** Hajek estimator can be derived by minimizing weighted sum of squared errors given a candidate exposure condition, which corresponds to a subset of $[0,1]^{m+1}$ (denoted by $\mathcal{X}$).

$$\hat{\bar{y}}_{Hajek}(d) = \arg\min_{y} \sum_{i \in \mathcal{U}, 1_i(\mathcal{X})=X} \frac{1}{\pi_i(\mathcal{X})} (Y_i - y)^2. \tag{6.6}$$

In conventional decision tree regression, one common criterion that determines whether the algorithm will continue to split a node is the reduction in sum of squared errors. Since Hajek estimators can be solved through weighted linear regressions, it is sensible to use a *weighted sum of squared errors* (WSSE). It can be computed through a weighted linear regression where the weight is $1/\pi_i(\mathcal{X})$.[6]

$$\text{WSSE}(\mathcal{X}) = \sum_{i \in \mathcal{U}, 1_i(\mathcal{X})=1} \frac{1}{\pi_i(\mathcal{X})} (Y_i - \hat{\bar{y}}_{Hajek}(\mathcal{X}))^2. \tag{6.7}$$

When considering splitting the partition $\mathcal{X}$ to sub-partitions $\mathcal{X}_l$ and $\mathcal{X}_r$, we

[6] In practice, we replace it by $1/\hat{\pi}_i(\mathcal{X})$.

cannot simply compare $\text{WSSE}(\mathcal{X}_l) + \text{WSSE}(\mathcal{X}_r)$ versus $\text{WSSE}(\mathcal{X})$. This is because $\mathbb{E}[\sum_{i\in\mathcal{U},1_i(\mathcal{X})=1} \frac{1}{\pi_i(\mathcal{X})}] = \mathbb{E}[\sum_{i\in\mathcal{U},1_i(\mathcal{X}_l)=1} \frac{1}{\pi_i(\mathcal{X}_l)}] = \mathbb{E}[\sum_{i\in\mathcal{U},1_i(\mathcal{X}_r)=1} \frac{1}{\pi_i(\mathcal{X}_r)}] = |\mathcal{U}|$, $\text{WSSE}(\mathcal{X}_l) + \text{WSSE}(\mathcal{X}_r)$ is generally greater than $\text{WSSE}(\mathcal{X})$. We thus adjust the splitting criterion by taking a weighted average between $\text{WSSE}(\mathcal{X}_l)$ and $\text{WSSE}(\mathcal{X}_r)$:

$$
\begin{aligned}
(\mathcal{X}_l^*, \mathcal{X}_r^*) = &\arg\min_{X_l, X_r} \frac{\sum_i 1_i(\mathcal{X}_l)}{\sum_i 1_i(\mathcal{X})}\text{WSSE}(\mathcal{X}_l) + \\
&\frac{\sum_i 1_i(\mathcal{X}_r)}{\sum_i 1_i(\mathcal{X})}\text{WSSE}(\mathcal{X}_r), \\
\text{where } &\frac{\sum_i 1_i(\mathcal{X}_l)}{\sum_i 1_i(\mathcal{X})}\text{WSSE}(\mathcal{X}_l) + \frac{\sum_i 1_i(\mathcal{X}_r)}{\sum_i 1_i(\mathcal{X})}\text{WSSE}(\mathcal{X}_r) \\
&< \text{WSSE}(\mathcal{X}) - \gamma \\
\text{and } &\sum_i 1_i(\mathcal{X}_l) \geq \kappa \text{ and } \sum_i 1_i(\mathcal{X}_r) \geq \kappa.
\end{aligned}
\tag{6.8}
$$

γ is a hyper-parameter used to require non-trivial reduction in WSSE. Similar to conventional decision trees, we can also set the minimum leaf size (κ) to prevent the tree from growing unnecessarily deep.[7] However, if the constraints cannot be satisfied, $(\mathcal{X}_l^*, \mathcal{X}_r^*)$ is an empty set and the algorithm would not further split $\mathcal{X}$.

Our algorithm is implemented by recursion. SPLIT is a procedure used to partition a given space. One can use $\text{SPLIT}([0,1]^{m+1})$ to start the recursion algorithm. The data used for the algorithm is a random half of the original training set. We then use the separate half to estimate the mean and variance. Empirically, direct effects (the impact of Z_i) are usually larger than indirect effects $\mathbf{X}_i$ (the impact of $\mathbf{X}_i$). Thus, the first partition is more inclined to split on Z_i: that is, it splits into two trees, in which one corresponds to treated observations and the other corresponds to non-treated observations.

[7] An empirically effective alternative is to require $\sum_{i\in\mathcal{U},1_i(\mathcal{X})=1} \frac{1}{\hat{\pi}_i(\mathcal{X})}$ to be very close to $|\mathcal{U}|$; e.g., $(1-\phi)|\mathcal{U}| \leq \sum_{i\in\mathcal{U},1_i(\mathcal{X})=1} \frac{1}{\hat{\pi}_i(\mathcal{X})} \leq (1+\phi)|\mathcal{U}|$, where ϕ is a small number. This helps us to avoid partitions with a small number of observations, which may have a high degree of randomness in $\sum_{i\in\mathcal{U},1_i(\mathcal{X})=1} \frac{1}{\hat{\pi}_i(\mathcal{X})}$ and consequently, $\hat{\bar{y}}_{Hajek}(\mathcal{X})$.

Algorithm 1 Implementation for the tree-based algorithm

procedure SPLIT($\mathcal{X}$)
  $\mathcal{X}_l^*, \mathcal{X}_r^* = \emptyset$
  Compute $\hat{\bar{y}}(\mathcal{X})$ and its variance using estimation set
  Compute WSSE($\mathcal{X}$) using training set
  $\text{WSSE}^*(\mathcal{X}) = \text{WSSE}(\mathcal{X})$
  for $k \in [1, m+1]$ **do**
    for $i \in \mathcal{U}$ **do**
      $\theta \leftarrow \mathbf{X}_{ik}$
      $\mathcal{X}_l \leftarrow \{j | \mathbf{X}_{jk} \leq \theta \ \ \& \ \ j \in \mathcal{X}\}$
      $\mathcal{X}_r \leftarrow \{j | \mathbf{X}_{jk} > \theta \ \ \& \ \ j \in \mathcal{X}\}$
      if Eq 5 is true for both $\mathcal{X}_l$ and $\mathcal{X}_r$ **then**
        Compute WSSE($\mathcal{X}_l$), and WSSE($\mathcal{X}_r$) using training set
        if Constraints in Eq. 8 are satisfied **then**
          $\mathcal{X}_l^* \leftarrow \mathcal{X}_l$
          $\mathcal{X}_r^* \leftarrow \mathcal{X}_r$
          Update $\text{WSSE}^*(\mathcal{X})$ using Equation 8
        end if
      end if
    end for
  end for
  if $X_l^* \neq \emptyset$ and $X_r^* \neq \emptyset$ **then** ▷ If find any satisfy Eq. 8, choose the one with minimum WSSE.
    SPLIT($\mathcal{X}_l^*$)
    SPLIT($\mathcal{X}_r^*$)
  else
    Add new exposure condition to $\mathcal{D}$, corresponding to $\mathcal{X}$
  end if
end procedure

Figure 6-5: Implementation for the tree-based algorithm

The bottleneck for computational efficiency is Line 5 — the loop that iterates every element in $\mathcal{U}$. An improvement is to replace this for-loop by randomly choosing η observations in $\mathcal{U}$. This may not help select the optimal cutoff but would provide good cutoffs if η is not too small. We also have several hyperparameters — γ, κ, δ, ϵ, and η. To tune those parameters, we can apply cross-validation to find the best choice that minimizes $\text{WSSE}^*([0,1]^{m+1})$.

In addition to estimating the average potential outcome, our approach can also be applied to estimating global average treatment effects and heterogeneous direct effects with indirect effects fixed as explained:

1. **Global average treatment effects.** The global average treatment effect of an experiment is the difference between the average outcome in a counterfactual world where everyone is treated versus that in a counterfactual world where everyone is non-treated [248]. This is fundamentally unsolvable unless assumptions such as SUTVA or our assumptions are made. In our approach, these two counterfactual worlds belong to two separate partitions. We can just identify these two partitions and compare the difference in the two average potential outcomes to compute the global average treatment effect.

   Specifically, when the tree algorithm terminates, we have two special subsets of $[0,1]^{m+1}$, denoted by $\mathcal{X}_0$ and $\mathcal{X}_1$ respectively. $\mathcal{X}_0$ is the partition that contains the cases where $Z_i = 0$, all elements of $\mathbf{X}_i$ that represent fully non-treated neighborhood (2-0, 3o-0, 3c-0, 4o-0, ...) are equal to 1, and the rest equal 0; $\mathcal{X}_1$ is the partition that contains the cases where $Z_i = 1$ and all elements of $\mathbf{X}_i$ that represent fully treated neighborhood (2-1, 3o-2, 3c-2, 4o-2, ...) are equal to 1, and the rest equal 0. Then we can use $\hat{\bar{y}}(\mathcal{X}_1) - \hat{\bar{y}}(\mathcal{X}_0)$ to estimate the global average treatment effect. Variance can be estimated by the method proposed in [22].

2. **Heterogeneous direct effects with indirect effects fixed.** As discussed, our algorithm tends to split on Z_i first if direct effects are larger than indirect effects. Once the tree grows to two sub-trees that correspond to treated and non-

treated observations, respectively, the further partitioning is not synchronized in these two trees. However, what if we want to understand how direct effects vary across different network interference conditions? Let $\mathcal{X}^{(1)}$ and $\mathcal{X}^{(0)}$ be the two spaces such that $Z_i^{(1)} = 1$ for all $(Z_i^{(1)}, \mathbf{X}_i^{(1)}) \in \mathcal{X}^{(1)}$; and $Z_i^{(0)} = 0$ for all $(Z_i^{(0)}, \mathbf{X}_i^{(0)}) \in \mathcal{X}^{(0)}$. $\mathcal{X}^{(1)}$ and $\mathcal{X}^{(0)}$ are complementary if for all $(0, \mathbf{X}_i^{(0)}) \in \mathcal{X}_i^{(0)}$, we have $(1, \mathbf{X}_i^{(0)}) \in \mathcal{X}_i^{(1)}$, and for all $(1, \mathbf{X}_i^{(1)}) \in \mathcal{X}_i^{(1)}$, we have $(0, \mathbf{X}_i^{(1)}) \in \mathcal{X}_i^{(0)}$.

We can then revise our tree algorithm by only partitioning on a space $[0, 1]^m$ (i.e., on $\mathbf{X}_i$) and not partitioning on Z_i. In each node, including leaves, the goal is no longer estimating the average potential outcome. Instead, we estimate the heterogeneous direct effects ($\tilde{\mathcal{X}} = \{\mathbf{X} | (1, \mathbf{X}) \in \mathcal{X}^{(1)}\}$)

$$\tau_{direct}(\tilde{\mathcal{X}}) = \bar{y}_{Hajek}(\mathcal{X}^{(1)}) - \bar{y}_{Hajek}(\mathcal{X}^{(0)}) \tag{6.9}$$

$\mathcal{X}^{(1)}$ indicates the cases where neighborhood assignment conditions are in a given region (e.g., high structural diversity, high echo chamber, or any other interference conditions) and the ego is treated; and $\mathcal{X}^{(0)}$ indicates the cases where the ego is non-treated but neighborhood interference conditions are the same. In this case, $\tau_{direct}(\tilde{\mathcal{X}})$ represents the average direct effect under that interference conditions.

If estimating heterogeneous direct effects with indirect effects fixed, two main revisions in Figure 6-5 are made. First, we adjust the weighted linear regressions. Remember that the coefficient for the constant variable represented the Hajek estimator; in this case, we need to add the Z_i into the regression, and report the coefficient for this variable to estimate $\tau_{direct}(\tilde{\mathcal{X}})$. WSSE is still derived from the error term for the weighted linear regression. The partitioning space becomes $[0, 1]^m$. Second, the variance of the estimator is more complicated to estimate. Essentially, the variance estimation is similar to the methods proposed in [25].

6.3 Results

We evaluate our algorithm in a synthetic Watts-Strogatz (WS) network where we can verify our approach recovers the ground-truth which we know, and a real-world A/B test of a new product feature on Facebook to illustrate the scalability of our approach.

Watts-Strogatz Simulation Network.

In studies of causal inference, a common challenge is that the ground-truth potential outcomes under either treatment or control for a given unit i is missing — that is, for each observation, we can only observe one single potential outcome given either a treatment or control assignment; with presence of network interference, we can only observe the outcome under one exposure condition. Therefore, we rely on a simulation study, which is often used to verify the effectiveness of causal inference methods.

We generate a Watts-Strogatz network [258] with $|\mathcal{U}|$=200,000. The Watts-Strogatz model is a random graph generator that preserves network properties such as clustering and the "small-world phenomenon." To ensure a large variation in the local structure of individuals' neighborhood, we set a high rewiring rate for edges — 50%. We set the number of replicates $R = 100$.

To be as general as possible and also complement the real-world independent assignment experiment, we use graph cluster randomization [248] in the simulation study. We use a simple clustering approach — we cluster every 10 nodes on the ring of the WS network, and we assign the treatment randomly and independently on the cluster level. We consider the following four different data-generating processes for outcomes, assuming that ε_i is randomly drawn from Gaussian distribution with the mean of 0 and the variance of 1:

1. **Cutoff Outcome**.

$$Y_i^{(1)} = 0.1|\mathcal{N}_i| + gender_i + 2Z_i \times \mathbb{1}[X_{i,3c-2} > 0.7] + \varepsilon_i. \tag{6.10}$$

$gender_i$ is a covariate independent of any other independent variables, which is randomly assigned to 1 or 0. Since this function has clear cutoffs, we can use it to validate that our tree-based algorithm splits on the corresponding cutoff.

2. **Causal Structural Diversity Outcome**.

$$Y_i^{(2)} = 0.1|\mathcal{N}_i| + gender_i + structural\_diversity\_of\_treated + Z_i \times structural\_diversity\_of\_treated + \varepsilon_i \tag{6.11}$$

Structural diversity is the number of disjointed components in an observation's neighborhood (or ego network). "*structural_ diversity_ of_ treated*" is defined as the structural diversity given the set of treated neighbors. Although it was found that structural diversity predicts a higher product adoption rate [247], it is unclear whether this is a causal impact or just correlation. A causal impact means it is actually the structural diversity of the treated neighbors that matters, while correlation means that the behavior is reflected by network structure, regardless of assignment conditions [233]. Our approach can analyze whether the causal impact of structural diversity exists in experimental data. Especially, the numbers of fully treated open triads or tetrads indicate the structure diversity of treated neighbors (3o-2 or 4o-3), and we expect our algorithm to split on these features.

3. **Correlational Structural Diversity Outcome**.

$$Y_i^{(3)} = 0.1|\mathcal{N}_i| + gender_i + structural\_diversity + Z_i \times structural\_diversity + \varepsilon_i \tag{6.12}$$

"*structural_ diversity*" is the structural diversity given all neighbors, regardless of their assignment conditions. Therefore, we do not expect the tree algorithm to split on any features.

4. **Validation Under Null Effect**.

$$Y_i^{(4)} = |\mathcal{N}_i| + \varepsilon_i. \tag{6.13}$$

As a sanity check, we use as a covariate – the number of neighbors as the outcome variable. We do not expect the tree algorithm to split on any features.

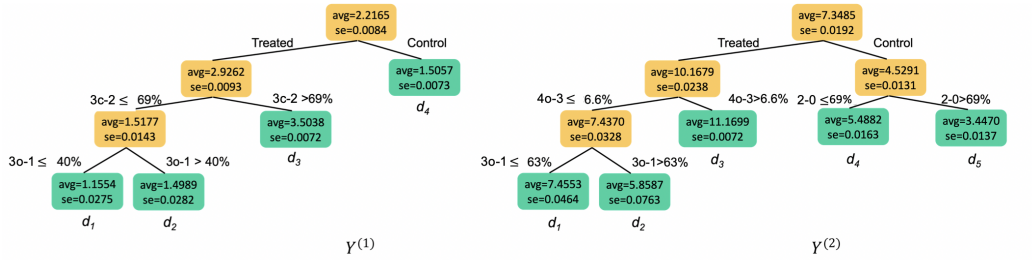


Figure 6-6: **The result trees for the simulation experiment using all specified network motifs.** The two trees represent $Y^{(1)}$ or the Validation Cutoff Outcome (left) and $Y^{(2)}$ or the Causal Structural Diversity Outcome (right), respectively. The numbers in each leaf represents the average potential outcome and standard error (square root of variance) of the corresponding partition (exposure condition).

Results are presented in Figure 6-6 and summarized below:

1. **Cutoff Outcome** ($Y_i^{(1)}$): As expected, after splitting treated versus control, it splits on fully treated closed triads (3c-2), with a threshold of 69%. This is consistent with the parameter 0.7 set in Equation 6.10. Also, as expected, it does not further split on control groups. However, the algorithm has an unexpected split on 3o-1, although the difference between the resulting d_1 and d_2 is not statistically significant. With further investigation, we find that this is because when partitioning the tree using the training set, it overfits the noise; however, since we use the estimation set for the average potential outcome and its variance (i.e. honest splitting), the resulting tree does not show this significant difference. This result also demonstrates the importance of honest splitting.

2. **Causal Structural Diversity Outcome** ($Y^{(2)}$): The result is presented in the right panel in Figure 6-6. It first splits on treated versus control. For treated observations, it first splits on fully treated open tetrad (4o-3), which is positively correlated with the degree of structural diversity of treated neighbors. When the fraction of fully treated open tetrad is greater than 6.6%, the algorithm terminates splitting, providing a partition (exposure condition) with the largest average potential outcome. If the fraction of fully treated open tetrad is smaller than 6.6%, it further splits on open triads (3o-1). If the fraction of 3o-1 is greater than 63%, the resulting exposure condition gives the smaller average potential outcome within treated observations. For control observations, partially because the structural diversity plays a smaller effect in the outcome function, it simply splits on the dyad level and terminates.

3. **Correlational Structural Diversity Outcome** ($Y^{(3)}$) and **Validation Under Null Effect** ($Y^{(4)}$): As expected, it terminates after splitting samples into treated and control groups. Note these examples have nothing to illustrate.

We compare our results with a baseline method used in [22]. The study examines four exposure conditions: no effect, direct effect, indirect effect, and direct + indirect effects. For example, when $Y_i^{(1)}$ is the outcome, most observations belong to either "indirect effect" (1.4981 $\pm$ 0.0035, 48.1%) or "direct + indirect effects" (2.9182 $\pm$ 0.0034, 50.0%). Similar results are derived in other outcomes. Therefore, this specification approach is not suitable for our experimental data. Moreover, such a partition cannot reveal the effects of specific network motifs or understand theories such as structural diversity.

We also compare our result to the specification of the fractional q-neighborhood exposure condition in [248], which is equivalent to the proposed algorithm using dyad features only. This comparison can help highlight the importance of accounting for network motifs rather than simply counting treated friends. When using only dyad features, the algorithm splits treated observations on 2-1 (i.e. fraction of treated neighbors), resulting in an average of potential outcome of 2.0472 ($\pm$0.0189) and

3.2137 ($\pm$0.0085) for less than or equal to, or greater than 60%, respectively. It does not split the control observations either.

We use estimated global treatment effects [248, 64] as a reference for the advantage of using causal network motif over simply accounting for proportions of treatment neighbors. Here we use $Y_i^{(1)}$ but similar results are derived in other settings. Using dyads features only (i.e., the proportion of treated neighbors, see Figure 6-7) gives ($3.2137 - 1.4976 =$) 1.7161 ($\pm$0.0112) while using all proposed motif features provides (d_3 and d_4 in the left panel of Figure 6-6, $3.5038 - 1.5057 =$) 1.9981 ($\pm$0.0102); as a baseline, the true global treatment effect is 2.0000, and the average treatment effect under SUTVA is ($2.9262 - 1.5057 =$) 1.4205 ($\pm$0.0118). In sum, using network motifs helps reduce more bias than using dyads features only when we estimate the global treatment effects because it characterizes network complex structure rather than purely counting the fraction of treated neighbors.

We also investigate heterogeneous direct effects given different network interference conditions. As shown in Figure 6-8, for both $Y^{(1)}$ and $Y^{(2)}$, it splits on the correct threshold: for $Y^{(1)}$, Z_i only matters then the fraction of 3c-2 is greater than 70%; for $Y^{(2)}$, it splits on important feature (4o-3), and the heterogeneous direct effect is larger when this fraction is greater.[8] As expected, no effect is observed for $Y^{(3)}$ and $Y^{(4)}$.

Care Reaction Rollout.

Finally, we apply our approach on a real-world A/B test of a product launch at Facebook. Facebook launched the 7th reaction, "Care".[9] When the reaction was launched, users in the control group (50%) could not use the Care reaction but could still observe other users reacting with care. The treated group (the other (50%), could both use and see the Care reaction. The sample size is approximately 5% of the Facebook population, and we only take into account users and neighbors who are in this 5% sampled users. Note that we cannot compare treated users in this

[8]Note that for the tasks of heterogeneous direct effects with indirect effects fixed, the average direct effect for a parent node does not necessarily lies between the values of its two child nodes.

[9]`https://www.facebook.com/careers/life/the-story-of-facebooks-care-reaction`

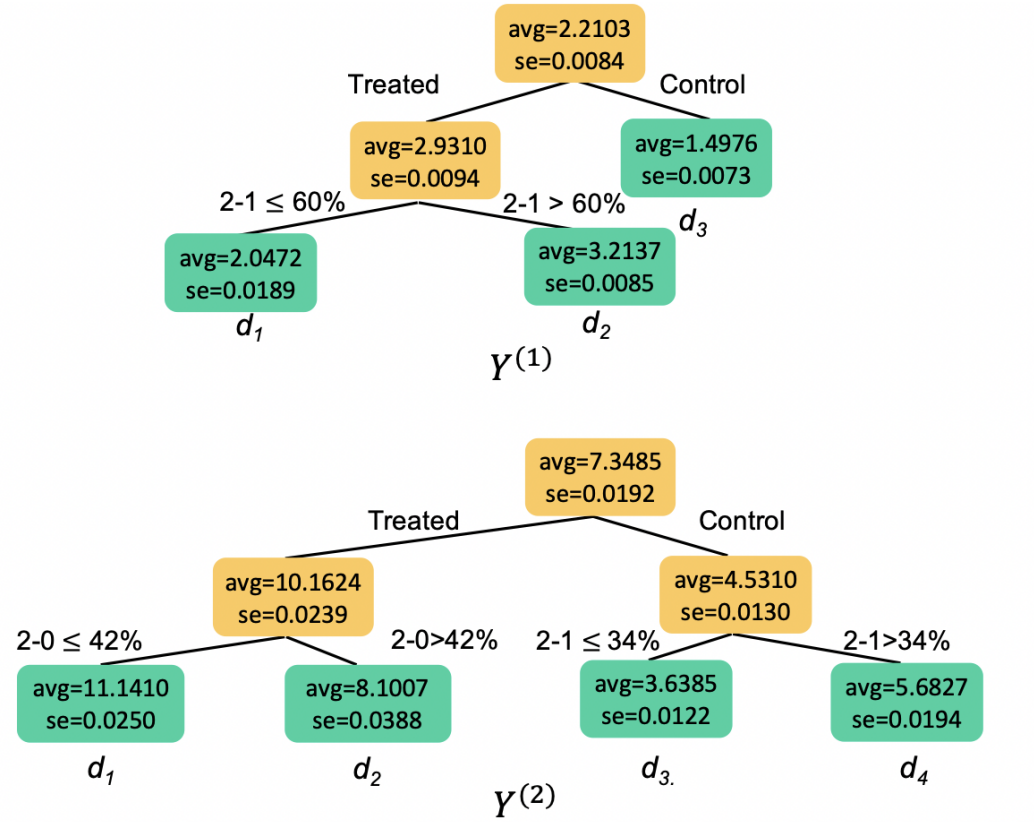


Figure 6-7: The result trees for the simulation experiment using dyad network motifs only. The two trees represent $Y^{(1)}$ or the Cutoff Outcome (top) and $Y^{(2)}$ or the Causal Structural Diversity Outcome (bottom), respectively. The numbers in each leaf represents the average potential outcome and standard error (square root of variance) of the corresponding partition (exposure condition).

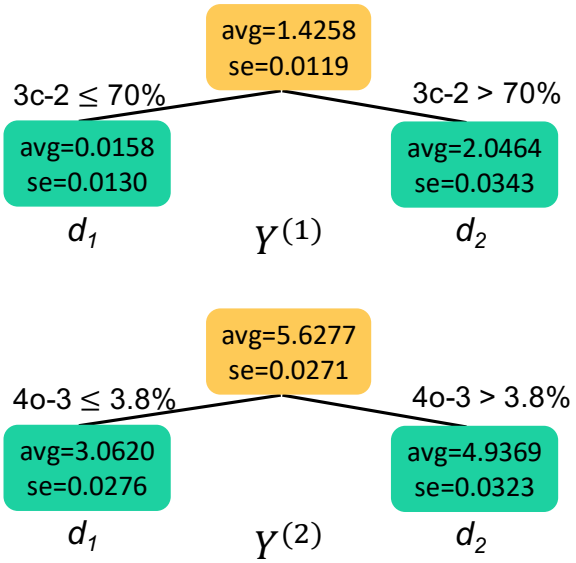


Figure 6-8: **The result trees for heterogeneous direct effects for the simulation experiment using all specified network motifs.** The numbers in each node or leaf represents the average direct effect given the network interference condition and standard error (square root of variance) of the corresponding partition (exposure condition).

5% versus all the rest of Facebook population because the database does not log non-compliance: those who were assigned to the 5% rollout but did not activate the treatment assignment (non-active users who did not attempt to use Care) are not recorded in the database and thus such a comparison is biased by non-compliance.

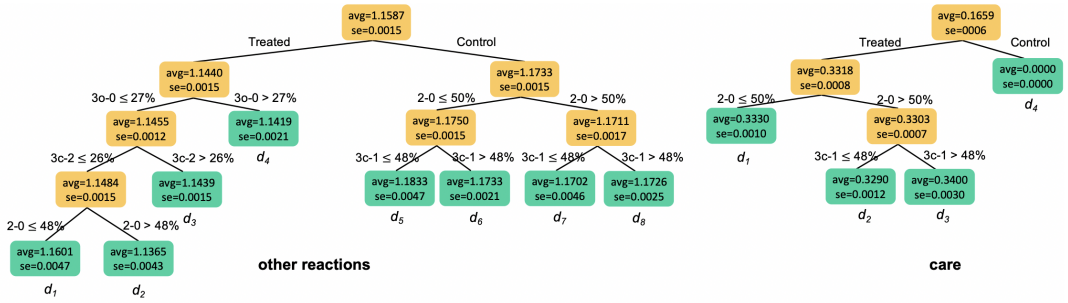


Figure 6-9: The result trees for the Care experiment using all specified network motifs. The left panel presents the result for the use of other reactions, and the right panel presents the result for the use of Care. The numbers in each leaf represents the average potential outcome and standard error (square root of variance) of the corresponding partition (exposure condition).

In this experiment, we expect a user's use of reactions to be impacted by the number of their Facebook friends who could use Care and how those friends are

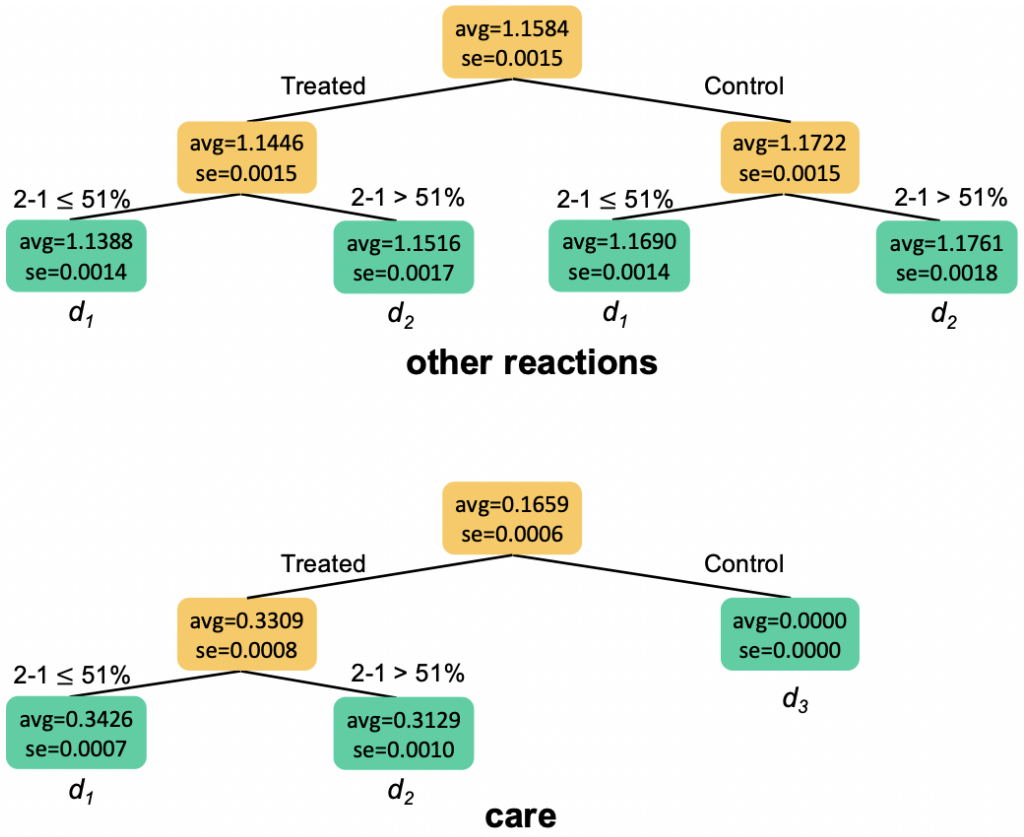


Figure 6-10: **The result trees for the Care experiment using dyads only.** The upper panel presents the result for the use of other reactions, and the lower panel presents the result for the use of Care. The numbers in each leaf represents the average potential outcome and standard error (square root of variance) of the corresponding partition.

connected which might impact their usage with each other. If a user has more friends using Care, or a group of friends able to use Care, she might use more Care reactions. There are two main outcome variables in this A/B test: (1) the number of Care reaction uses during the week after the experiment was launched. In the control group, the number of Care reaction uses is always zero; and (2) the number of other reaction uses. To prevent the impact from extreme points, outcome variables are set on the log scale ($\log_{10}(x+1)$).

We randomly split our observations into approximately equal training and estimation sets of sufficiently large sample sizes. We use the training set to partition the tree and the estimation set for the average potential outcome and standard errors in each leaf node. We compute only dyad and triad network motifs, which is mainly

restricted by computational resources.

The results are presented in Figure 6-9. There are eight exposure conditions determined by the algorithm when the outcome is other reactions, and four when outcome is use of Care. Again, we observe heterogeneity among different network interference conditions. For example, on the left panel, d_1 and d_2 have a difference of $(1.1601 - 1.1365 =)\ 0.0236 \pm 0.0064$. As a reference, directly comparing the averages between treated and control groups gives -0.0293, which shows the importance of distinguishing between different network interference conditions.

Again, we also compare the results using all specified network motifs (Figure 6-9) versus the results using only dyads (Figure 6-10) to illustrate the importance of using network motifs beyond dyads. In the left panel of Figure 6-9, d_3 contains the scenario of a fully treated neighborhood and d_5 the scenario of a fully non-treated neighborhood. Therefore, we can estimate the global average treatment as $(1.1419 - 1.1833) = -0.0414\ (\pm 0.0049)$. The baseline average treatment effect by directly comparing the averages between treated and control groups is $-0.0293\ (\pm 0.0021)$. Thus, simply comparing the difference in means between treated and control groups could underestimate the true treatment effect by 25%. By contrast, in the upper panel of Figure 6-10, comparing d_2 and d_3 only gives $(1.1516 - 1.1690 =)\ -0.0174\ (\pm 0.0022)$. Similar analysis and conclusions also apply when the outcome is Care.

We also examine heterogeneous direct effects under different network interference conditions. Since use of Care is always zero for the control group, this analysis is only meaningful when the outcome is other reactions. The algorithm splits on 3o-2 at 23% but does not provide significant differences between the two child nodes; we do not provide a figure here.

6.4 Discussion

Network interference is much more complicated than simply being described as the "indirect effect." To examine and analyze heterogeneity of indirect effects in experimental datasets, we provide a two-step solution. We first propose and employ the

causal network motifs to characterize the network interference conditions, and then develop a tree-based algorithm for partitioning. Our tree-based algorithm is interpretable in terms of highlighting which exposure conditions are important for defining potential outcomes, addresses selection bias and positivity issues, and avoids incorrect standard error concerns via honest splitting.

Given the large number of researcher degrees of freedom in existing approaches for network interference such as choosing the threshold for an exposure condition, our approach provides a simple way to automatically specify exposure conditions. In this way, researchers no longer need to define exposure conditions a priori, and the exposure conditions generated by the algorithm are suitable for the given data and experiment. We believe that methodological innovation for addressing network interference concerns in A/B tests on networks will continue to be an important area for development, and accounting for network motifs with treatment assignment conditions provides a useful way to detect heterogeneous network interference effects.

Our work also falls at the intersection of graph mining and causal inference. Graph mining methods are aimed at analyzing network properties such as the average number of connections, the number of communities, and more [58, 72]. We focus on one type of graph property called a motif [180, 7, 2, 103]. However, we are not aware of any applications that have used labeled network motifs in causal inference applications, which is the focus of this work.

Moreover, we highlight our approach should not be confused with the Causal Tree and relevant methods [25, 255]. The goal of our approach is to partition on the space of random assignments, while the Causal Tree and similar methods partition only on the covariate space. The critical difference in our setting is that in each partition of the treatment space, all observations should have a (almost) non-zero probability of belonging to that partition. In this way, we can construct certain unbiased estimators, such as Hajek, to estimate the average outcome under certain treatment regions and thus quantify a causal impact. By contrast, in Causal Tree and relevant methods, covariates are not intervened by the experiment and each observation has only probability of zero or one of belonging to each partition. Therefore, their methods are

primarily aimed at partitioning the covariate space (i.e., to identify heterogeneous treatment effects for sub-groups in the sample) while our approach is to partition the treatment space (i.e, to identify the causal effects of different treatment or more specifically, exposure conditions). [27] is probably the most relevant recent paper, which also characterizes ego network structure by motifs or other network features. However, they focus on using matching to estimate average direct effects, and the treatment assignments of neighbor nodes are not their main focus.

Practitioners using our approach may obtain important insights. For example, they could understand how to utilize social contagion for product promotion when they have constraints on the number of promos. Researchers may identify important network interference conditions that are not theorized in certain experimental settings.

There remain many open questions or future directions based on our approach. First, we can incorporate covariates into the algorithm such as demographic features. One way is to also allow partitioning on the covariate space as well as the treatment space as in this work. However, once the algorithm splits on a covariate, all the descendants of that node only estimate the average potential outcome for the subsample that satisfies the criterion. The other way is to incorporate covariates in the weighted linear regressions in our algorithm. This helps reduce the variance for the estimators and improve the precision when estimating average potential outcomes. We may also want to account for tie strength as well.

Second, we may consider alternative machine learning algorithms. Decision trees are not the only choice in our setting. We use a decision tree partially because it is an interpretable machine learning algorithm and does not involve functional form specification, except for assuming constant potential outcome. Our tree-based algorithm can in fact be improved by the Hoeffding tree [78], which provides a streaming algorithm to perform the partitioning efficiently. Instead of using sample split, we may improve our methods by the conformal prediction theory [220]. Moreover, we can imagine using nearest neighbor based algorithms and local regression instead to estimate the potential outcome given any point or region in $[0,1]^{m+1}$. In addition, parametric methods can also be used when the goal is specific about estimating the

potential outcome at certain points (e.g., fully treated neighbors) or estimating the global treatment effect [64].

Finally, our approach can be extended to any experimental data with multiple variables or continuous treatment variables, which are not necessarily only controlled experiments on social networks. While existing causal inference literature has primarily studied single binary or categorical treatments, fewer studies have approached continuous or multiple variables [118, 130]. Our approach provides a way to automatically convert multiple or continuous treatment variables to categorical treatments. It would be interesting to further investigate how machine learning can be applied to causal inference for continuous or multiple treatments as well as adapting this approach to observational causal inference settings.

Chapter 7

Conclusion and future plans

My PhD thesis discusses the explorations I have conducted in the interaction between social network analysis and computational techniques. Specifically, I have contributed to both the understanding of social networks and the approaches that facilitate social network analysis. The first part of my thesis focuses on understanding how social exchange and long ties play important roles in social network formation and dynamics; the second part focuses on products on online social media and their impacts on social interactions and networks.

In the future, I plan to develop my research towards the following three directions.

Social Interactions and Social Networks for the Good

I am especially interested in how to utilize social preference and social contagion to promote positive social interactions, and how social networks have shaped human behavior and can be reshaped by digital technologies.

In the future, I will continue to use social network perspectives to understand how to effectively promote various types of positive social interactions, including cooperation, prosocial behavior, and intergroup contact. For example, I am interested in how information and resources are diffused unequally through social networks and how social platforms can reshape the structure of social networks to mitigate such inequality. Another direction is to study the potential negative impact of social

contagion of "good actions." Although good actions may be contagious, they may also increase the inequality between different social groups: in a segregated social network, social contagion may further promote the level of prosociality in groups with a strong prosocial and cooperative culture, but this may exacerbate the gap between the rich and poor. Finally, as a long-term goal, I plan to design and implement mechanisms on online social platforms that can effectively promote positive social interactions and reshape the structure of social networks in our societies.

Complex Explanatory Models

Recent advances in machine learning have demonstrated the potential of complex models with high-dimensional hypothesis space in prediction-based tasks. By contrast, explanatory models, which are intended to describe the mechanism of a phenomenon, usually avoid such complexity. Take economic models for social networks as an example. Explanatory models in their literature are parsimonious and tractable, but typically do not account for heterogeneity in individual preferences and attributes and cannot model the complexity of real-world social systems. Hence, when used to predict individual decisions, these models may not achieve satisfactory performance. My research thus aims to develop explanatory models with individual-level heterogeneity and predictability.

In the future, I will continue to explore the applications of advanced machine learning techniques in explanatory models, with a focus on social interactions and social networks. Potentially useful techniques include multi-agent reinforcement learning, deep learning (graph neural networks), and online learning.

Complex causal inference

In the big data era, computational scientists utilize cutting-edge computational and machine learning techniques to detect and analyze interesting patterns. However, the lack of exogeneity in observational studies poses challenges to valid causal identifications. Although causal inference is stressed in quantitative social science, it is not

stressed enough in computer science and big data analytics. Importantly, existing approaches may not be naturally applicable to large-scale data with complex data structure. For example, the independent variable of interest (treatment variable) in causal inference is typically binary (i.e., simply treatment or control), but in big data the treatment variable may be continuous and high-dimensional, or partially random.

For observational data from online social platforms, randomness from computer algorithms can manipulate treatment variables of interest to facilitate causal inference. For example, in Chapter 4, the monetary gift amount is (partially) randomly determined by a computer algorithm algorithm, where the treatment variable is the gift amount received. I use a directed acyclic graph (DAG) to illustrate the complex causal relationships among observed and unobserved variables, guiding me to conduct valid causal inference.

For experimental data, I have developed a machine learning approach to analyze complex patterns of social contagion (in Chapter 6). In an experiment on social networks, the outcome of an individual is affected both by her own treatment assignment, her network neighbor's treatment assignment (e.g., Facebook friends), and their social network structure. The treatment space in this case is high-dimensional and contains network structure. This study provides a solution by combining network motif analysis and machine learning to convert the high-dimensional treatment assignments to low-dimensional ones.

In the future, I will continue to study interesting open questions in applied causal inference, including identifying complicated heterogeneous treatment effects, temporal patterns of treatment effects, and the interaction effect between multiple treatments.

Bibliography

[1] Martín Abadi and Agarwal A Barham P TensorFlow. Large-scale machine learning on heterogeneous distributed systems. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation*, pages 265–283, 2016.

[2] Christoph Adami, Jifeng Qian, Matthew Rupp, and Arend Hintze. Information content of colored motifs in complex networks. *Artif Life*, 17(4):375–390, 2011.

[3] Lada A Adamic and Natalie Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43, 2005.

[4] Sasan Adibi. *Mobile health: a technology road map*, volume 5. Springer, 2015.

[5] Idris Adjerid, George Loewenstein, Rachael Purta, and Aaron Striegel. Gain-loss incentives and physical activity: The role of choice and wearable health tools. *Management Science*, 2021.

[6] Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland. Social fmri: investigating and shaping social mechanisms in the real world. *Pervasive Mob Comput*, 7(6):643–659, 2011.

[7] Nesreen K Ahmed, Jennifer Neville, Ryan A Rossi, Nick G Duffield, and Theodore L Willke. Graphlet decomposition: Framework, algorithms, and applications. *KAIS*, 50(3):689–722, 2017.

[8] Edoardo M Airoldi, David M Blei, Stephen E Fienberg, and Eric P Xing. Mixed membership stochastic blockmodels. *J Mach Learn Res*, 9(Sep):1981–2014, 2008.

[9] Richard D Alexander. *The Biology of Moral Systems*. Transaction Publishers, 1987.

[10] Jerilyn K Allen, Janna Stephens, Cheryl R Dennison Himmelfarb, Kerry J Stewart, and Sara Hauck. Randomized controlled pilot study testing use of smartphone technology for obesity treatment. *Journal of obesity*, 2013, 2013.

[11] Uri Alon. Network motifs: theory and experimental approaches. *Nat Rev Genet*, 8(6):450–461, 2007.

[12] Mark I Alpert and W Thomas Anderson. Optimal heterophily and communication effectiveness: some empirical findings. *J Commun*, 23(3):328–343, 1973.

[13] Yehuda Amir. Contact hypothesis in ethnic relations. *Psychological bulletin*, 71(5):319, 1969.

[14] James Andreoni. Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal*, 100(401):464–477, 1990.

[15] Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press, 2008.

[16] Elliott M Antman, Joseph Lau, Bruce Kupelnick, Frederick Mosteller, and Thomas C Chalmers. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: treatments for myocardial infarction. *JAMA*, 268(2):240–248, 1992.

[17] Sinan Aral. Networked experiments. *The Oxford Handbook of the Economics of Networks*, pages 376–411, 2016.

[18] Sinan Aral, Lev Muchnik, and Arun Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51):21544–21549, 2009.

[19] Sinan Aral and Christos Nicolaides. Exercise contagion in a global social network. *Nature Communications*, 8(1):14753, 2017.

[20] Sinan Aral and Dylan Walker. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Manage Sci*, 57(9):1623–1639, 2011.

[21] Sinan Aral and Dylan Walker. Identifying influential and susceptible members of social networks. *Science*, 337(6092):337–341, 2012.

[22] Peter M Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *Ann Appl Stat*, 2017.

[23] Bruno Arpino, Luca De Benedictis, and Alessandra Mattei. Implementing propensity score matching with network data: The effect of gatt on bilateral trade. 2015.

[24] Susan Athey, Dean Eckles, and Guido W Imbens. Exact p-values for network interference. *JASA*, 113(521):230–240, 2018.

[25] Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *PNAS*, 2016.

[26] Susan Athey and Guido W Imbens. The econometrics of randomized experiments. In *Handbook of Economic Field Experiments*, volume 1, pages 73–140. Elsevier, 2017.

[27] Usaid Awan, Marco Morucci, Vittorio Orlandi, Sudeepa Roy, Cynthia Rudin, and Alexander Volfovsky. Almost-matching-exactly for treatment effect estimation under network interference. In *International Conference on Artificial Intelligence and Statistics*, pages 3252–3262. PMLR, 2020.

[28] Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 44–54. ACM, 2006.

[29] Eytan Bakshy, Dean Eckles, Rong Yan, and Itamar Rosenn. Social influence in social advertising: evidence from field experiments. In *EC*, pages 146–161, 2012.

[30] Eytan Bakshy, Solomon Messing, and Lada A Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.

[31] Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st International Conference on World Wide Web*, pages 519–528. ACM, 2012.

[32] Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. The diffusion of microfinance. *Science*, 341(6144):1236498, 2013.

[33] Albert-László Barabási. Scale-free networks: a decade and beyond. *Science*, 325(5939):412–413, 2009.

[34] Albert-László Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013.

[35] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[36] Guillaume Basse and Avi Feller. Analyzing two-stage experiments in the presence of interference. *JASA*, 113(521):41–55, 2018.

[37] C Daniel Batson and Adam A Powell. Altruism and prosocial behavior. *Handbook of psychology*, pages 463–484, 2003.

[38] Alexandre Belloni, Daniel Chen, Victor Chernozhukov, and Christian Hansen. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429, 2012.

[39] Roland Bénabou and Jean Tirole. Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678, 2006.

[40] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

[41] John Beshears, Hae Nim Lee, Katherine L Milkman, Robert Mislavsky, and Jessica Wisdom. Creating exercise habits using incentives: The trade-off between flexibility and routinization. *Management Science*, 2020.

[42] Christopher M Bishop. *Pattern Recognition and Machine Learning*. 2006.

[43] Nicholas Bloom, James Liang, John Roberts, and Zhichun Jenny Ying. Does working from home work? evidence from a chinese experiment. *The Quarterly Journal of Economics*, 130(1):165–218, 2015.

[44] Marián Boguná, Romualdo Pastor-Satorras, Albert Díaz-Guilera, and Alex Arenas. Models of social networks based on social distance attachment. *Phys Rev E*, 70(5):056122, 2004.

[45] Gary E Bolton and Axel Ockenfels. Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166–193, 2000.

[46] Gary E Bolton and Axel Ockenfels. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: comment. *American Economic Review*, 96(5):1906–1911, 2006.

[47] Robert M Bond, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle, and James H Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295, 2012.

[48] Jake Bowers, Mark M Fredrickson, and Costas Panagopoulos. Reasoning about interference between units: A general framework. *Polit Anal*, pages 97–124, 2013.

[49] Erik Brynjolfsson, John J Horton, Adam Ozimek, Daniel Rock, Garima Sharma, and Hong-Yi TuYe. Covid-19 and remote work: An early look at us data. Technical report, National Bureau of Economic Research, 2020.

[50] Erik Brynjolfsson, Yu Jeffrey Hu, and Mohammad S Rahman. *Competing in the age of omnichannel retailing*. MIT Cambridge, MA, 2013.

[51] Erwin Bulte, Ruixin Wang, and Xiaobo Zhang. Forced gifts: The burden of being a friend. *Journal of Economic Behavior & Organization*, 155:79–98, 2018.

[52] Ronald S Burt. *Structural holes*. Harvard university press, 1992.

[53] Ronald S Burt. *Structural Holes: The Social Structure of Competition*. Harvard University Press, 2009.

[54] Chen Cai, Hongyan Shang, and Tansiny Lertsiriworapong. A study of motivation factors of "grab red envelopes" phenomenon: A case study of wechat. In *4th International Conference on Management Science, Innovation, and Technology*, 2017.

[55] Dorwin Cartwright and Frank Harary. Structural balance: a generalization of heider's theory. *Psychol Rev*, 63(5):277, 1956.

[56] Damon Centola. The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197, 2010.

[57] Damon Centola and Michael Macy. Complex contagions and the weakness of long ties. *American journal of Sociology*, 113(3):702–734, 2007.

[58] Deepayan Chakrabarti and Christos Faloutsos. Graph mining: Laws, generators, and algorithms. *ACM computing surveys (CSUR)*, 38(1), 2006.

[59] Arun G Chandrasekhar and Matthew O Jackson. A network formation model based on subgraphs. preprint at `https://arxiv.org/abs/1611.07658`, 2016.

[60] Gary Charness and Uri Gneezy. Incentives to exercise. *Econometrica*, 77(3):909–931, 2009.

[61] Jin Chen, Wynne Hsu, Mong Li Lee, and See-Kiong Ng. Labeling network motifs in protein interactomes for protein function prediction. In *ICDM*, pages 546–555, 2007.

[62] Yan Chen and Sherry Xin Li. Group identity and social preferences. *American Economic Review*, 99(1):431–57, 2009.

[63] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey. Double/debiased/neyman machine learning of treatment effects. *Am Econ Rev*, 2017.

[64] Alex Chin. Regression adjustments for estimating the global treatment effect in experiments with interference. *JCI*, 7(2), 2019.

[65] Eunjoon Cho, Seth A Myers, and Jure Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090, 2011.

[66] Nicholas Christakis, James Fowler, Guido W Imbens, and Karthik Kalyanaraman. An empirical model for strategic network formation. In *The Econometric Analysis of Network Data*, pages 123–148. Elsevier, 2020.

[67] Nicholas A Christakis and James H Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370–379, 2007.

[68] Nicholas A Christakis, James H Fowler, Guido W Imbens, and Karthik Kalyanaraman. An empirical model for strategic network formation. preprint at `http://www.nber.org/papers/w16039`, 2010.

[69] Robert B Cialdini and Noah J Goldstein. Social influence: Compliance and conformity. *Annual Review of Psychology*, 55:591–621, 2004.

[70] Ethan Cohen-Cole and Jason M Fletcher. Detecting implausible social network effects in acne, height, and headaches: longitudinal analysis. *Bmj*, 337, 2008.

[71] Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. Political polarization on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, 2011.

[72] Diane J Cook and Lawrence B Holder. *Mining graph data*. John Wiley & Sons, 2006.

[73] Karen S Cook, Coye Cheshire, Eric RW Rice, and Sandra Nakagawa. Social exchange theory. *Handbook of social psychology*, pages 61–88, 2013.

[74] Karen S Cook and Toshio Yamagishi. Power in exchange networks: a power-dependence formulation. *Soc Networks*, 14(3-4):245–265, 1992.

[75] Sergio Currarini, Matthew O Jackson, and Paolo Pin. An economic model of friendship: homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045, 2009.

[76] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559, 2016.

[77] Anne-Ly Do, Lars Rudolf, and Thilo Gross. Patterns of cooperation: fairness and coordination in networks of interacting agents. *New J Phys*, 12(6):063023, 2010.

[78] Pedro Domingos and Geoff Hulten. Mining high-speed data streams. In *KDD*, pages 71–80, 2000.

[79] David A Drew, Long H Nguyen, Claire J Steves, Cristina Menni, Maxim Freydin, Thomas Varsavsky, Carole H Sudre, M Jorge Cardoso, Sebastien Ourselin, Jonathan Wolf, et al. Rapid implementation of mobile technology for real-time epidemiology of covid-19. *Science*, 2020.

[80] Elizabeth W Dunn, Lara B Aknin, and Michael I Norton. Spending money on others promotes happiness. *Science*, 319(5870):1687–1688, 2008.

[81] Nathan Eagle, Michael Macy, and Rob Claxton. Network diversity and economic development. *Science*, 328(5981):1029–1031, 2010.

[82] Nathan Eagle, Alex Sandy Pentland, and David Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences*, 106(36):15274–15278, 2009.

[83] Dean Eckles and Eytan Bakshy. Bias and high-dimensional adjustment in observational studies of peer effects. *JASA*, 2020.

[84] Dean Eckles, Brian Karrer, and Johan Ugander. Design and analysis of experiments in networks: Reducing bias from interference. *JCI*, 5(1), 2016.

[85] Dean Eckles, Elchanan Mossel, M Amin Rahimian, and Subhabrata Sen. Long ties accelerate noisy threshold-based contagions. *Available at SSRN 3262749*, 2019.

[86] Bradley Efron. Bootstrap methods: another look at the jackknife. In *Breakthroughs in Statistics*, pages 569–593. Springer, 1992.

[87] Naoki Egami, Christian J Fong, Justin Grimmer, Margaret E Roberts, and Brandon M Stewart. How to make causal inferences using texts. *arXiv preprint*, 2018.

[88] Barbara Entwisle, Katherine Faust, Ronald R Rindfuss, and Toshiko Kaneda. Networks and contexts: Variation in the structure of social ties. *American Journal of Sociology*, 112(5):1495–1533, 2007.

[89] Deborah Estrin and Ida Sim. Open mhealth architecture: an engine for health care innovation. *Science*, 330(6005):759–760, 2010.

[90] Ernst Fehr and Klaus M Schmidt. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868, 1999.

[91] Daniel C Feldman. The development and enforcement of group norms. *Academy of Management Review*, 9(1):47–53, 1984.

[92] RA Fisher. *The design of experiments.* Number 2nd Ed. Oliver & Boyd, Edinburgh & London., 1937.

[93] Andreas Flache and Michael W Macy. The weakness of strong ties: Collective action failure in a highly cohesive group. In *Evolution of social networks*, pages 27–52. Routledge, 2013.

[94] Seth Flaxman, Sharad Goel, and Justin M Rao. Filter bubbles, echo chambers, and online news consumption. *Public Opin Q*, 80(S1):298–320, 2016.

[95] Laura Forastiere, Edoardo M Airoldi, and Fabrizia Mealli. Identification and estimation of treatment and interference effects in observational studies on networks. *JASA*, pages 1–18, 2020.

[96] James H Fowler and Nicholas A Christakis. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *British Medical Journal*, 337:a2338, 2008.

[97] James H Fowler and Nicholas A Christakis. Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 107(12):5334–5338, 2010.

[98] Ove Frank and David Strauss. Markov graphs. *Journal of the american Statistical association*, 81(395):832–842, 1986.

[99] John RP French, Bertram Raven, and D Cartwright. The bases of social power. *Classics of Organization Theory*, 7:311–320, 1959.

[100] Noah E Friedkin. An expected value model of social power: predictions for selected exchange networks. *Soc Networks*, 14(3-4):213–229, 1992.

[101] Heider Fritz et al. The psychology of interpersonal relations, 1958.

[102] Simon Gächter, Daniele Nosenzo, and Martin Sefton. Peer effects in pro-social behavior: Social norms or social preferences? *Journal of the European Economic Association*, 11(3):548–573, 2013.

[103] Brian Gallagher and Tina Eliassi-Rad. Leveraging label-independent features for classification in sparsely labeled networks: An empirical study. In *SNAKDD*, pages 1–19. Springer, 2008.

[104] Guillaume Gaulier and Soledad Zignago. Baci: international trade database at the product-level (the 1994-2007 version), 2010.

[105] Laura K Gee, Jason J Jones, Christopher J Fariss, Moira Burke, and James H Fowler. The paradox of weak ties in 55 countries. *Journal of Economic Behavior & Organization*, 133:362–372, 2017.

[106] Alan S Gerber and Donald P Green. *Field experiments: Design, analysis, and interpretation*. WW Norton, 2012.

[107] Golnaz Ghasemiesfeh, Roozbeh Ebrahimi, and Jie Gao. Complex contagion and the weakness of long ties in social networks: revisited. In *Proceedings of the fourteenth ACM conference on Electronic Commerce*, pages 507–524, 2013.

[108] Anindya Ghose, Xitong Guo, and Beibei Li. Empowering patients using smart mobile health platforms: Evidence from a randomized field experiment. *MIS Quarterly*, 2021.

[109] Liam G Glynn, Patrick S Hayes, Monica Casey, Fergus Glynn, Alberto Alvarez-Iglesias, John Newell, Gearóid ÓLaighin, David Heaney, Martin O'Donnell, and Andrew W Murphy. Effectiveness of a smartphone application to promote physical activity in primary care: the smart move randomised controlled trial. *British Journal of General Practice*, 64(624):e384–e391, 2014.

[110] Mark Granovetter. The strength of weak ties: A network theory revisited. *Sociological Theory*, pages 201–233, 1983.

[111] Mark S Granovetter. The strength of weak ties. *American journal of sociology*, 78(6):1360–1380, 1973.

[112] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.

[113] Jinyong Hahn, Petra Todd, and Wilbert Van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209, 2001.

[114] Cassandra Handan-Nader, Daniel E Ho, and Becky Elias. Feasible policy evaluation by design: A randomized synthetic stepped-wedge trial of mandated disclosure in king county. *Eval Rev*, 2020.

[115] David R Heise. Problems in path analysis and causal inference. *Sociological methodology*, 1:38–73, 1969.

[116] César A Hidalgo and Ricardo Hausmann. The building blocks of economic complexity. *Proc Natl Acad Sci USA*, 106(26):10570–10575, 2009.

[117] John P Higgins. Smartphone applications for patients' health and fitness. *The American journal of medicine*, 129(1):11–19, 2016.

[118] Keisuke Hirano and Guido W Imbens. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, volume 226164, chapter The propensity score with continuous treatments, pages 73–84. 2004.

[119] Daniel E Ho. Does peer review work: An experiment of experimentalism. *Stan L Rev*, 69:1, 2017.

[120] Jake M. Hofman, Duncan J. Watts, Susan Athey, Filiz Garip, Thomas L. Griffiths, Jon Kleinberg, Helen Margetts, Sendhil Mullainathan, Matthew J. Salganik, Simine Vazire, Alessandro Vespignani, and Tal Yarkoni. Integrating explanation and prediction in computational social science. *Nature*, 595:181–188, 2021.

[121] Paul W Holland and Samuel Leinhardt. Transitivity in structural models of small groups. *Comparative Group Studies*, 2(2):107–124, 1971.

[122] Petter Holme. Analyzing temporal networks in social media. *Proceedings of the IEEE*, 102(12):1922–1933, 2014.

[123] Petter Holme and Jari Saramäki. Temporal networks. *Physics reports*, 519(3):97–125, 2012.

[124] Judith A Holton. Building trust and collaboration in a virtual team. *Team Performance Management: An International Journal*, 7(3/4):36–47, 2001.

[125] Lu Hong and Scott E Page. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proc Natl Acad Sci USA*, 101(46):16385–16389, 2004.

[126] Hong Huang, Yuxiao Dong, Jie Tang, Hongxia Yang, Nitesh V Chawla, and Xiaoming Fu. Will triadic closure strengthen ties in social networks? *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 12(3):1–25, 2018.

[127] Shan Huang, Sinan Aral, Yu Jeffrey Hu, and Erik Brynjolfsson. Social advertising effectiveness across products: A large-scale field experiment. *Marketing Science*, 39(6):1142–1165, 2020.

[128] Kosuke Imai, Zhichao Jiang, and Anup Malani. Causal inference with interference and noncompliance in two-stage randomized experiments. *JASA*, pages 1–13, 2020.

[129] Kosuke Imai, Gary King, and Elizabeth A Stuart. Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society: Series A*, 171(2):481–502, 2008.

[130] Kosuke Imai and David A Van Dyk. Causal inference with general treatment regimes: Generalizing the propensity score. *JASA*, 99(467):854–866, 2004.

[131] Alex Imas. Working for the "warm glow": On the benefits and limits of prosocial incentives. *Journal of Public Economics*, 114:14–18, 2014.

[132] Guido W Imbens and Donald B Rubin. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.

[133] Raghuram Iyengar, Christophe Van den Bulte, and Thomas W Valente. Opinion leadership and social contagion in new product diffusion. *Marketing Science*, 30(2):195–212, 2011.

[134] Shanto Iyengar and Kyu S Hahn. Red media, blue media: Evidence of ideological selectivity in media use. *Journal of communication*, 59(1):19–39, 2009.

[135] Matthew O Jackson. *Social and economic networks*. Princeton University Press, 2010.

[136] Matthew O Jackson and Asher Wolinsky. A strategic model of social and economic networks. *J Econ Theory*, 71(1):44–74, 1996.

[137] Matthew O Jackson and Yiqing Xing. Culture-dependent strategies in coordination games. *Proc Natl Acad Sci USA*, 111(Supplement 3):10889–10896, 2014.

[138] Neil F Johnson, Chen Xu, Zhenyuan Zhao, Nicolas Ducheneaut, Nicholas Yee, George Tita, and Pak Ming Hui. Human group formation in online guilds and offline gangs driven by a common team dynamic. *Phys Rev E*, 79(6):066117, 2009.

[139] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003.

[140] Walter N Kernan, Catherine M Viscoli, Robert W Makuch, Lawrence M Brass, and Ralph I Horwitz. Stratified randomization for clinical trials. *Journal of Clinical Epidemiology*, 52(1):19–26, 1999.

[141] Daichi Kimura and Yoshinori Hayakawa. Coevolutionary networks with homophily and heterophily. *Physical Review E*, 78(1):016103, 2008.

[142] Gary King and Richard Alexander Nielsen. Why propensity scores should not be used for matching. *Political Analysis*, 2019.

[143] Diederik Kingma and Jimmy Ba. Adam: a method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations*, 2015.

[144] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[145] Morwenna Kirwan, Mitch J Duncan, Corneel Vandelanotte, and W Kerry Mummery. Using smartphone technology to monitor physical activity in the 10,000 steps program: a matched case–control trial. *Journal of medical Internet research*, 14(2):e55, 2012.

[146] René F Kizilcec, Eytan Bakshy, Dean Eckles, and Moira Burke. Social influence and reciprocity in online gift giving. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 126. ACM, 2018.

[147] Jon Kleinberg and Éva Tardos. Balanced outcomes in social exchange networks. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, pages 295–304. ACM, 2008.

[148] Teruyoshi Kobayashi, Taro Takaguchi, and Alain Barrat. The structured backbone of temporal social ties. *Nature communications*, 10(1):1–11, 2019.

[149] Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, and Nils Pohlmann. Online controlled experiments at large scale. In *KDD*, pages 1168–1176, 2013.

[150] Adam DI Kramer, Jamie E Guillory, and Jeffrey 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):8788–8790, 2014.

[151] Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *PNAS*, 116(10):4156–4165, 2019.

[152] Nicola Lacetera, Mario Macis, and Angelo Mele. Viral altruism? generosity and social contagion in online networks. *Sociological Science*, 2016.

[153] Nicola Lacetera, Mario Macis, and Robert Slonim. Will there be blood? incentives and displacement effects in pro-social behavior. *American Economic Journal: Economic Policy*, 4(1):186–223, 2012.

[154] Nicola Lacetera, Mario Macis, and Robert Slonim. Economic rewards to motivate blood donations. *Science*, 340(6135):927–928, 2013.

[155] Brian Yoshio Laing, Carol M Mangione, Chi-Hong Tseng, Mei Leng, Ekaterina Vaisberg, Megha Mahida, Michelle Bholat, Eve Glazier, Donald E Morisky, and Douglas S Bell. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. *Annals of internal medicine*, 161(10_Supplement):S5–S12, 2014.

[156] Nan M Laird, Jonathan Skinner, and Michael Kenward. An analysis of two-period crossover designs with carry-over effects. *Statistics in Medicine*, 11(14-15):1967–1979, 1992.

[157] Jennifer M Larson. The weakness of weak ties for novel information diffusion. *Applied network science*, 2(1):1–15, 2017.

[158] David Lazer, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al. Life in the network: the coming age of computational social science. *Science (New York, NY)*, 323(5915):721, 2009.

[159] Michael P Leung. Treatment and spillover effects under network interference. *Rev Econ Stat*, 102(2):368–380, 2020.

[160] Daniel Z Levin and Rob 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.

[161] Aming Li, Sean P Cornelius, Y-Y Liu, Long Wang, and A-L Barabási. The fundamental advantages of temporal networks. *Science*, 358(6366):1042–1046, 2017.

[162] Wentao Liang. Study on the status quo of "werun" application operation. Master's thesis, Beijing Sport University, 2017.

[163] David Liben-Nowell and Jon Kleinberg. The link-prediction problem for social networks. *Journal of the American society for information science and technology*, 58(7):1019–1031, 2007.

[164] Bing Liu, Yiyuan Xia, and Philip S Yu. Clustering through decision tree construction. In *CIKM*, pages 20–29, 2000.

[165] Che-Wei Liu, Guodong Gordon Gao, and Ritu Agarwal. Reciprocity or self-interest? leveraging digital social connections for healthy behavior. *Leveraging Digital Social Connections for Healthy Behavior (December 13, 2019)*, 2019.

[166] Shengming Liu, Ye Zhang, Lifan Chen, Li Guo, and Dongling Yu. Enterprise wechat groups: Their effect on work-life conflict and life-work enhancement. *Frontiers of Business Research in China*, 9(4):516, 2015.

[167] Anne B Loucks and Jean R Thuma. Luteinizing hormone pulsatility is disrupted at a threshold of energy availability in regularly menstruating women. *J Clin Endocrinol Metab*, 88(1):297–311, 2003.

[168] Yadong Luo. The changing chinese culture and business behavior: The perspective of intertwinement between guanxi and corruption. *International Business Review*, 17(2):188–193, 2008.

[169] Barry Markovsky and Edward J Lawler. A new theory of group solidarity. 11:113–137, 1994.

[170] Peter V Marsden. Egocentric and sociocentric measures of network centrality. *Social Networks*, 24(4):407–422, 2002.

[171] Andreu Mas-Colell, Michael Dennis Whinston, Jerry R Green, et al. *Microeconomic theory*, volume 1. Oxford university press New York, 1995.

[172] Gemma Flores Mateo, Esther Granado-Font, Carme Ferré-Grau, and Xavier Montaña-Carreras. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. *Journal of medical Internet research*, 17(11):e253, 2015.

[173] Marcel Mauss. *The gift: The Form and Reason for Exchange in Archaic Societies*. Routledge, 2002.

[174] Justin McCrary. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714, 2008.

[175] Alan J McKane and Barbara Drossel. Models of food web evolution. *Ecological Networks: linking Structure to Dynamics in Food Webs*, pages 223–243, 2006.

[176] J Miller McPherson, Pamela A Popielarz, and Sonja Drobnic. Social networks and organizational dynamics. *American sociological review*, 57:153–170, 1992.

[177] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

[178] Angelo Mele. A structural model of dense network formation. *Econometrica*, 85(3):825–850, 2017.

[179] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[180] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[181] Giovanna Miritello, Esteban Moro, and Rubén Lara. Dynamical strength of social ties in information spreading. *Physical Review E*, 83(4):045102, 2011.

[182] James D Montgomery. Weak ties, employment, and inequality: An equilibrium analysis. *American Journal of Sociology*, 99(5):1212–1236, 1994.

[183] Luvai F Motiwalla. Mobile learning: A framework and evaluation. *Computers & education*, 49(3):581–596, 2007.

[184] Mark Newman, Albert-Laszlo Barabasi, and Duncan J Watts. *The Structure and Dynamics of Networks*. Princeton University Press, 2006.

[185] Martin A Nowak. Five rules for the evolution of cooperation. *Science*, 314(5805):1560–1563, 2006.

[186] Martin A Nowak and Sébastien Roch. Upstream reciprocity and the evolution of gratitude. *Proceedings of the Royal Society B: Biological Sciences*, 274(1610):605–610, 2007.

[187] Martin A Nowak and Karl Sigmund. Evolution of indirect reciprocity by image scoring. *Nature*, 393(6685):573, 1998.

[188] Martin A Nowak and Karl Sigmund. Evolution of indirect reciprocity. *Nature*, 437(7063):1291, 2005.

[189] Hisashi Ohtsuki, Christoph Hauert, Erez Lieberman, and Martin A Nowak. A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441(7092):502–505, 2006.

[190] J-P Onnela, Jari Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, Kimmo Kaski, János Kertész, and A-L Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the national academy of sciences*, 104(18):7332–7336, 2007.

[191] Scott E Page. *The difference: how the power of diversity creates better groups, firms, schools, and societies*. Princeton University Press, 2008.

[192] Harsh Parikh, Cynthia Rudin, and Alexander Volfovsky. Malts: Matching after learning to stretch. *arXiv preprint arXiv:1811.07415*, 2018.

[193] Namsu Park, Kerk F Kee, and Sebastián Valenzuela. Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes. *Cyberpsychology & behavior*, 12(6):729–733, 2009.

[194] Patrick S Park, Joshua E Blumenstock, and Michael W Macy. The strength of long-range ties in population-scale social networks. *Science*, 362(6421):1410–1413, 2018.

[195] Partridge, Kevin McGeechan, Lana Hebden, Kate Balestracci, Annette Wong, Elizabeth Denney-Wilson, Mark Harris, Philayrath Phongsavan, Adrian Bauman, and Margaret Allman-Farinelli. Effectiveness of a mhealth lifestyle program with telephone support (txt2bfit) to prevent unhealthy weight gain in young adults: Randomized controlled trial. *Jmir mhealth and uhealth*, 3(2), 2015.

[196] Judea Pearl. *Causality*. Cambridge University Press, 2009.

[197] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710, 2014.

[198] Tuan Q Phan and Edoardo M Airoldi. A natural experiment of social network formation and dynamics. *Proceedings of the National Academy of Sciences*, 112(21):6595–6600, 2015.

[199] Jordan Poppenk, Stefan Köhler, and Morris Moscovitch. Revisiting the novelty effect: When familiarity, not novelty, enhances memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5):1321, 2010.

[200] Jean Pouget-Abadie, Guillaume Saint-Jacques, Martin Saveski, Weitao Duan, S Ghosh, Y Xu, and Edoardo M Airoldi. Testing for arbitrary interference on experimentation platforms. *Biometrika*, 106(4):929–940, 2019.

[201] Jiezhong Qiu, Yixuan Li, Jie Tang, Zheng Lu, Hao Ye, Bo Chen, Qiang Yang, and John E Hopcroft. The lifecycle and cascade of wechat social messaging groups. In *Proceedings of the 25th International Conference on World Wide Web*, pages 311–320. ACM, 2016.

[202] J. Ross Quinlan. Induction of decision trees. *Mach Learn*, 1986.

[203] Pedro Ribeiro and Fernando Silva. Discovering colored network motifs. In *Complex Networks*, pages 107–118. Springer, 2014.

[204] Margaret E Roberts, Brandon M Stewart, and Richard A Nielsen. Adjusting for confounding with text matching. *AJPS*, 2018.

[205] Everett M Rogers and Dilip K Bhowmik. Homophily-heterophily: relational concepts for communication research. *Public Opin Q*, 34(4):523–538, 1970.

[206] Daniel M Romero, Brendan Meeder, and Jon Kleinberg. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proc. 20th International Conference on World Wide Web*, pages 695–704, 2011.

[207] Paul R Rosenbaum. Interference between units in randomized experiments. *JASA*, 102(477):191–200, 2007.

[208] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[209] Christian Roy. *Traditional festivals: a multicultural encyclopedia*, volume 1. Abc-Clio, 2005.

[210] Heather Royer, Mark Stehr, and Justin Sydnor. Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company. *American Economic Journal: Applied Economics*, 7(3):51–84, 2015.

[211] Donald B Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

[212] Francisco C Santos, Jorge M Pacheco, and Tom Lenaerts. Cooperation prevails when individuals adjust their social ties. *PLoS Comput Biol*, 2(10):e140, 2006.

[213] Anida Sarajlić, Noël Malod-Dognin, Ömer Nebil Yaveroğlu, and Nataša Pržulj. Graphlet-based characterization of directed networks. *Sci Rep*, 6:35098, 2016.

[214] CE Särndal, B Swensson, and J Wretman. *Model assisted survey sampling Springer*. Springer, 1992.

[215] Martin Saveski, Jean Pouget-Abadie, Guillaume Saint-Jacques, Weitao Duan, Souvik Ghosh, Ya Xu, and Edoardo M Airoldi. Detecting network effects: Randomizing over randomized experiments. In *KDD*, 2017.

[216] Fredrik Sävje, Michael J Higgins, and Jasjeet S Sekhon. Generalized full matching and extrapolation of the results from a large-scale voter mobilization experiment. *arXiv preprint*, 2017.

[217] Thomas C Schelling. Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186, 1971.

[218] Shaun R Seaman and Ian R White. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*, 2013.

[219] Ingrid Seinen and Arthur Schram. Social status and group norms: Indirect reciprocity in a repeated helping experiment. *European Economic Review*, 50(3):581–602, 2006.

[220] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *JMLR*, 9(3), 2008.

[221] Cosma Rohilla Shalizi and Andrew C Thomas. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, 40(2):211–239, 2011.

[222] Kin Wai Michael Siu. Red packet: A traditional object in the modern world. *The Journal of Popular Culture*, 35(3):103–125, 2001.

[223] Brian Skyrms and Robin Pemantle. A dynamic model of social network formation. In *Adaptive networks*, pages 231–251. Springer, 2009.

[224] Mario Luis Small. Weak ties and the core discussion network: Why people regularly discuss important matters with unimportant alters. *Social networks*, 35(3):470–483, 2013.

[225] Jordan J. Smith, Philip J. Morgan, Ronald C. Plotnikoff, Kerry A. Dally, Jo Salmon, Anthony D. Okely, Tara L. Finn, and David R. Lubans. Smartphone obesity prevention trial for adolescent boys in low-income communities: The atlas rct. *Pediatrics*, 134(3):723–731, 2014.

[226] Tom AB Snijders. Stochastic actor-oriented models for network change. *Journal of mathematical sociology*, 21(1-2):149–172, 1996.

[227] Jerzy Splawa-Neyman, Dorota M Dabrowska, and TP Speed. On the application of probability theory to agricultural experiments. *Stat Sci*, pages 465–472, 1990.

[228] Steven R Steinhubl, Evan D Muse, and Eric J Topol. The emerging field of mobile health. *Science translational medicine*, 7(283):283rv3–283rv3, 2015.

[229] Jeroen Stragier, Mariek Vanden Abeele, Peter Mechant, and Lieven De Marez. Understanding persistence in the use of online fitness communities: comparing novice and experienced users. *Computers in Human Behavior*, 64:34–42, 2016.

[230] Jeroen Stragier, Mariek Vanden Abeele, and Lieven De Marez. Recreational athletes' running motivations as predictors of their use of online fitness community features. *Behaviour & Information Technology*, 37(8):815–827, 2018.

[231] Richard S Strauss and Harold A Pollack. Social marginalization of overweight children. *Arch Pediatr Adolesc Med*, 157(8):746–752, 2003.

[232] Elizabeth A Stuart. Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1):1, 2010.

[233] Jessica Su, Krishna Kamath, Aneesh Sharma, Johan Ugander, and Sharad Goel. An experimental study of structural diversity in social networks. In *ICWSM*, volume 14, pages 661–670, 2020.

[234] Jerry Suls, Rene Martin, and Ladd Wheeler. Social comparison: Why, with whom, and with what effect? *Current directions in psychological science*, 11(5):159–163, 2002.

[235] Jimeng Sun and Jie Tang. A survey of models and algorithms for social influence analysis. In *Social network data analytics*, pages 177–214. Springer, 2011.

[236] Tianshu Sun, Guodong Gao, and Ginger Zhe Jin. Mobile messaging for offline group formation in prosocial activities: A large field experiment. *Management Science*, 65(6):2717–2736, 2019.

[237] Henri Tajfel, John C Turner, William G Austin, and Stephen Worchel. An integrative theory of intergroup conflict. *Organizational Identity: A Reader*, pages 56–65, 1979.

[238] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1067–1077. ACM, 2015.

[239] Jie Tang, Tiancheng Lou, Jon Kleinberg, and S Wu. Transfer link prediction across heterogeneous social networks. *ACM Trans Inf Syst*, 9(4), 2010.

[240] Eric J Tchetgen Tchetgen and Tyler J VanderWeele. On causal inference in the presence of interference. *Stat Methods Med Res*, 2012.

[241] Felix J Thoemmes and Eun Sook Kim. A systematic review of propensity score methods in the social sciences. *Multivariate behavioral research*, 46(1):90–118, 2011.

[242] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[243] Yasuyuki Todo, Petr Matous, and Hiroyasu Inoue. The strength of long ties and the weakness of strong ties: Knowledge diffusion through supply chain networks. *Research Policy*, 45(9):1890–1906, 2016.

[244] Penny Trieu, Joseph B Bayer, Nicole B Ellison, Sarita Schoenebeck, and Emily Falk. Who likes to be reachable? availability preferences, weak ties, and bridging social capital. *Information, Communication & Society*, 22(8):1096–1111, 2019.

[245] Endel Tulving and Neal Kroll. Novelty assessment in the brain and long-term memory encoding. *Psychonomic Bulletin & Review*, 2(3):387–390, 1995.

[246] Gabrielle Turner-McGrievy and Deborah Tate. Tweets, apps, and pods: Results of the 6-month mobile pounds off digitally (mobile pod) randomized weight-loss intervention among adults. *Journal of medical Internet research*, 13(4):e120, 2011.

[247] Johan Ugander, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences*, 109(16):5962–5966, 2012.

[248] Johan Ugander, Brian Karrer, Lars Backstrom, and Jon Kleinberg. Graph cluster randomization: Network exposure to multiple universes. In *KDD*, pages 329–337, 2013.

[249] Johan Ugander and Hao Yin. Randomized graph cluster randomization. *arXiv preprint arXiv:2009.02297*, 2020.

[250] Jose Van Dijck and Thomas Poell. Understanding the promises and premises of online health platforms. *Big Data & Society*, 3(1):2053951716654173, 2016.

[251] Tyler J VanderWeele. Ignorability and stability assumptions in neighborhood effects research. *Stat Med*, 2008.

[252] Elizabeth S Veinott, Judith Olson, Gary M Olson, and Xiaolan Fu. Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 302–309, 1999.

[253] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

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

[255] Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *JASA*, 2018.

[256] Cheng Lu Wang, Noel YM Siu, and Bradley R Barnes. The significance of trust and renqing in the long-term orientation of chinese business-to-business relationships. *Industrial Marketing Management*, 37(7):819–824, 2008.

[257] Warren E Watson, Kamalesh Kumar, and Larry K Michaelsen. Cultural diversity's impact on interaction process and performance: comparing homogeneous and diverse task groups. *Acad Manage J*, 36(3):590–602, 1993.

[258] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440, 1998.

[259] Lilian Weng, Márton Karsai, Nicola Perra, Filippo Menczer, and Alessandro Flammini. Attention on weak ties in social and communication networks. In *Complex Spreading Phenomena in Social Systems*, pages 213–228. Springer, 2018.

[260] Daniel Westreich and Stephen R Cole. Invited commentary: positivity in practice. *Am J Epidemiol*, 2010.

[261] Ladd Wheeler. Toward a theory of behavioral contagion. *Psychological Review*, 73(2):179, 1966.

[262] Jeffrey C Wong. Computational causal inference. *arXiv preprint*, 2020.

[263] Yue Wu, Atreyi Kankanhalli, and Ke-wei Huang. Gamification in fitness apps: How do leaderboards influence exercise? In *Thirty Sixth International Conference on Information Systems*, 2016.

[264] Ziming Wu and Xiaojuan Ma. Money as a social currency to manage group dynamics: Red packet gifting in chinese online communities. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2240–2247. ACM, 2017.

[265] Ya Xu, Nanyu Chen, Addrian Fernandez, Omar Sinno, and Anmol Bhasin. From infrastructure to culture: A/b testing challenges in large scale social networks. In *KDD*, pages 2227–2236, 2015.

[266] Toshio Yamagishi and Karen S Cook. Generalized exchange and social dilemmas. *Social Psychology Quarterly*, pages 235–248, 1993.

[267] Erez Yoeli, Jon Rathauser, Syon P Bhanot, Maureen K Kimenye, Eunice Mailu, Enos Masini, Philip Owiti, and David Rand. Digital health support in treatment for tuberculosis. *New England Journal of Medicine*, 381(10):986–987, 2019.

[268] Yuan Yuan, Ahmad Alabdulkareem, and Alex Sandy Pentland. An interpretable approach for social network formation among heterogeneous agents. *Nature communications*, 9(1):1–9, 2018.

[269] Yuan Yuan, Kristen Altenburger, and Farshad Kooti. Causal network motifs: Identifying heterogeneous spillover effects in a/b tests. In *Proceedings of the Web Conference 2021*, pages 3359–3370, 2021.

[270] Yuan Yuan, Tracy Liu, Chenhao Tan, Qian Chen, Alex'Sandy' Pentland, and Jie Tang. Gift contagion in online groups: Evidence from wechat red packets. *Available at SSRN*, 2021.

[271] Yuan Yuan, Tracy Xiao Liu, Chenhao Tan, and Jie Tang. Online red packets: a large-scale empirical study of gift giving on wechat. *arXiv preprint arXiv:1712.02926*, 2017.

[272] Wayne W Zachary. An information flow model for conflict and fission in small groups. *J Anthropol Res*, 33(4):452–473, 1977.

[273] TL Zakrison, PC Austin, and VA McCredie. A systematic review of propensity score methods in the acute care surgery literature: avoiding the pitfalls and proposing a set of reporting guidelines. *European Journal of Trauma and Emergency Surgery*, 44(3):385–395, 2018.

[274] Jichang Zhao, Junjie Wu, and Ke Xu. Weak ties: Subtle role of information diffusion in online social networks. *Physical Review E*, 82(1):016105, 2010.