Version of Record: https://www.sciencedirect.com/science/article/pii/S0049089X23000728
Manuscript_44025620a8e52c483bc268775eb6d0fd

Distinguishing Social Mechanisms of Membership Adoption in Emerging Technology Communities

Qianyi Shi[a]
Yongren Shi[b]

University of Iowa Departments of [a]Internal Medicine and [b]Sociology and Criminology

* both authors contribute equally.
Key words: Organizational Membership; Voluntary Associations; Social Networks; Knowledge Economy.

Please address correspondence to:
Yongren Shi
University of Iowa
Department of Sociology and Criminology,
401 North Hall
Iowa City, IA 52242
Phone: 319-467-0645
yongren-shi@uiowa.edu

Acknowledgement: Preliminary versions of this paper were presented at the Annual Meeting of the American Sociological Association in Los Angeles, CA. The authors would like to thank Freda Lynn, Jennifer Glanville, Bogdan Ion Vasi, and Elizabeth Menninga for their helpful suggestions. All remaining errors are strictly the responsibility of the authors. This research was supported by a grant from the National Science Foundation (2048670).

Distinguishing Social Mechanisms of Membership Adoption in Emerging Technology Communities

Abstract

Digital platforms that enable and foster associations and sharing among entrepreneurs and knowledge workers have become a vital part of the new knowledge economy, yet we know little about the new form of social organization of knowledge. This paper seeks to explore and evaluate two microscopic social mechanisms, namely network effect of recruitment and cultural affinity, that may produce knowledge clustering and differentiation within these communities. To understand the relative effect of mechanisms, we develop a novel estimation procedure that matches individual users based on their historical behavioral patterns. We collected and analyzed a large-scale event dataset from a digital platform for offline in-person meetups in two major U.S. cities, New York City and San Francisco Bay Area. We found that previous methods overestimate network effect in membership adoption decisions by 176%. Our findings show that the network effect is further amplified by varied levels of cultural affinity between individuals and groups, implying a clustering effect whereby individuals tend to gravitate towards groups that are culturally proximate. Implications for understanding social differentiation and knowledge economy are discussed.

1. Introduction

The emergence of knowledge economy is often intertwined with the differentiation and fragmentation of knowledge communities, wherein individuals possess specialized knowledge face growing challenges in communication and collaboration (Powell and Snellman 2004; Saxenian 1996; DellaPosta and Nee 2020). It is crucial to create a system that promotes the clustering and differentiation of knowledge domains, but at the same time, it is important foster effective convergence to keep individuals updated with new information. Understanding the mechanisms of social differentiation has been a longstanding goal for sociologists, with explanations ranging from the growing complexity and interdependence of social structures (Durkheim 1893; Spencer 1898) to cognitive mechanisms of knowledge creation and social influence (Carley 1991; Mark 1998). Recent studies have highlighted the role of population dynamics in engendering social differentiation. For instance, research on a tech professional community in New York City revealed a shift towards a more specialized ecosystem of knowledge and information, driven by the influx of new participants with different knowledge and interests (DellaPosta and Nee 2020).

The rise of digital technologies and platforms, including recommender systems, content sorting algorithms, search engines, and artificial intelligence, greatly enable and foster associations and sharing among entrepreneurs and knowledge workers. With these tools at their disposal, individuals are now able to quickly and effortlessly select niche topics and connect and engage with like-minded individuals, thereby potentially intensifying the process of knowledge differentiation and specialization (Bakshy, Messing, and Adamic 2015; Boyd and Ellison 2007; Flaxman, Goel, and Rao 2016). Thus, this technological transition brings forth new inquiries regarding the social mechanisms that underlie system differentiation within the emerging knowledge economy.

In this study, we focus on the Meetup.com platform as a case study and explore and evaluate the microscopic social mechanisms that may drive clustering and differentiation, namely the network effect of recruitment and the impact of cultural affinity on membership adoption. We expect that individuals who are closer to a group in the resource space, representing the configuration of technological topics, are more likely to assimilate into the cultural norms and values of the group and join it. In addition, the interpersonal networks formed in shared events are expected to have a positive effect on the spread of membership adoption between individuals. With advanced search functions, topic directories, and customizable filters, individuals can easily navigate and connect with like-minded communities. Thus, we anticipate that the network effect will be stronger between individuals and groups that share cultural affinity. In contrast, interpersonal connections become less influential when individuals and groups do not share cultural similarity, as individuals can explore diverse perspectives and engage with a global network of ideas and information.

We collect and analyze a large-scale event participation dataset from Meetup.com in two major U.S. cities, New York City and San Francisco Bay Area. To distinguish mechanisms, we develop a novel estimation procedure that matches individual users based on their historical behavioral patterns. Yet although many empirical studies have highlighted the impact of social connections on membership induction (McAdam 1990; McPherson, Smith-Lovin, and Cook 2001; Snow, Zurcher, and Ekland-Olson 1980), it is important to note that these studies often establish associational rather than causal relationships. It is challenging to ascertain a direct cause-and-effect relationship due to various confounding factors, such as shared interests and self-selection bias that can influence both social connection formation and membership decisions, leading to an overestimation of the effect. In addition, collecting members' behavioral data is notoriously hard. Previous data collection efforts on communes and social movements have predominantly relied on longitudinal surveys based on self-reported data (McPherson and Rotolo 1996; Zablocki 1980). Nevertheless, the collection of behavioral data in these contexts remains relatively limited and insufficiently explored within the existing literature. In contrast, research focusing on the causal identification in social networks has made progress in distinguishing network effects from homophily in causing behavioral clustering (Aral, Muchnik, and Sundararajan 2009; Shalizi and Thomas 2011). However, the studies often assume that demographic traits are the main drivers of homophilous connections, while overlooking the potential impact of behavioral preferences on tie formation in online contexts. By leveraging behavioral data, our estimation procedure significantly contributes to a vital and growing research area.

Herein, we discuss the importance of understanding knowledge economy through an ecological lens. We hypothesize that, in voluntary communities such as Meetup where digital technologies help organize in-person events, co-participation network ties still serve a positive role in membership recruitment. We further hypothesize that the effect is contingent on the user–group affinity. Then, a new causal identification method based on topic matching is proposed to estimate the network effect. In the remainder of the paper, we will discuss the data collection methods, analytical strategies, results, and implications for literatures on social influence, knowledge economies, sociology of organizational ecology.

1.1. Differentiation and Fragmentation of Knowledge-sharing Communities

In a knowledge-based economy, the acquisition of formal knowledge and informal knowhows is essential for both individual career advancement and the research and development efforts of high-tech companies (Powell and Snellman 2004). Creating a balanced ecosystem within knowledge-sharing communities is crucial, as it promotes both the centralization and development of knowledge domains through differentiation and clustering, while also foster effective communication and convergence to keep individuals updated with new information. Scholars have discovered that industrial clusters demonstrate diverse patterns of knowledge accumulation in non-adjacent geographical regions (Audretsch and Feldman 1996; Saxenian 1996). For example, Silicon Valley in the United States and the Cambridge Cluster in the United Kingdom are prime examples of highly differentiated and clustered knowledge economies (Saxenian 1996). These regions are renowned for their concentration of technology companies, research institutions, and skilled workforce, leading to the creation of innovative products and services. The notion of specialization has expanded beyond physical capabilities and assembly line processes. It now encompasses the cumulative growth of information and knowledge available within these clusters (Bell 1976; Castells 2011; Powell and Snellman 2004). The discoveries underscore the crucial role of differentiation and specialization as a fundamental component of the division of labor in modern economic domains propelled by information and technology.

The investigation of social differentiation has been a major goal of sociologists, dating back to the pioneering work of Spencer (1898) and Durkheim ([1893] 1893) over a century ago. Various explanations have been put forth to explain the emergence of social differentiation. Classic work by Durkheim ([1893] 1893) posits that social differentiation arises from the growing complexity and interdependence of social structures. Social constructivism (Carley

1986, 1991) offers explanations of how knowledge can be systemically differentiated based on minimal assumptions about human information processing and interaction. This differentiation can be generated in formal models and simulations (Mark 1998; DellaPosta, Shi, and Macy 2015) that rely on simple cognitive mechanisms of knowledge creation and forgetting, social influence, i.e., individuals who are connected tend to pass on information, and homophily, i.e., similarities between a pair of individuals create correlated outcome patterns without direct causal influence (Carley 1986; McPherson et al. 2001). In a recent study of a tech professional community in New York City, DellaPosta and Nee (2020) analyzes email exchanges within over seven years, revealing a shift towards a more specialized ecosystem of knowledge and information. The influx of new participants with different knowledge and interests drives this division of knowledge, but the community retains its ability to address diverse topics despite individual contributors sorting into specialized niches.

Digital platforms that host in-person events, such as Meetup.com and Facebook Group, have become popular among knowledge workers seeking to freely share and acquire new technologies and know-how, find startup partners, or simply network. These digital platforms aim to remove barriers to access for specialized formal organizations and provide opportunities for those who are traditionally marginalized and excluded from institutionalized knowledge to gain access to essential resources (Adams et al. 2019; DellaPosta and Nee 2020; Lupton 2014; Nee and Drouhot 2020). In this study, we focus on Meetup.com, a social networking website that facilitates *offline in-person* events, as a case study. The Meetup platform allows users to create groups based on interest and identity, and these groups organize social events (Paxton and Rap 2016). Despite sharing characteristics with other knowledge-based communities, the Meetup platform possesses unique features that are of sociological significance. Unlike other collaborative platforms, such as GitHub and Stack Overflow, that are entirely virtual, event participation on Meetup takes place face-to-face in physical space. Physical co-presence and face-to-face interactions provide participants with valuable information that cannot be obtained through virtual interactions. This information helps participants understand the topical focus of events and the behavioral norms associated with them. For example, the etiquette at a cocktail party for an entrepreneurship group may be different from that of a reading club. Frequent participants in these types of social events can develop a deep understanding and honed skills in managing impressions, enabling them to foster and maintain trusting relationships with their co-members, which is a critical asset in a dynamic business environment (Goffman 1983).

[Figure 1 to be inserted here]

Our data is restricted to the technological groups on Meetup.com. The platform itself attracts tech-savvy individuals who are interested in leveraging technology for various purposes, including networking, skill development, and knowledge sharing. The accessibility and convenience of Meetup.com also make it an ideal platform for individuals with technological interests to create and join groups centered around specific technologies, programming languages, or industry sectors. Since its establishment, the tech community has undergone a rapid evolution characterized by a growing level of differentiation and specialization. Over the course of 15 years, various technological niches and specialized areas of expertise have emerged, reflecting the dynamic nature of the knowledge communities and the diverse interests of its participants. Figure 1 illustrates the structural configurations of technological topics in the San Francisco Bay Area (SFBA). Each point on the figure represents a technological topic, and their proximity indicates the level of shared participations in groups focused on those topics. The t-SNE algorithm (Maaten and Hinton 2008) is utilized to position the points in lower dimensions while preserving their similarities. The figure reveals a notable increase in both the number and diversity of topics in SFBA over a 10-year period. For instance, in 2009, SFBA had 818 topics listed by at least 5 groups, and a total of 112 active technological groups attracted 5,059 participants. In 2018, the number of topics remained at 818, but the active technological groups grew to 1,490, drawing a total of 105,149 unique participants. The largest topics in SFBA in 2018 were Software Development (48,989 participants), Open Source (42,175), and Computer Programming (36,371). The figure also demonstrates the presence of clustered and distinct groups of specialized topics like game design and computer graphics, blockchain and cryptocurrency, AWS and data center, big data analytics and artificial intelligence, indicating that participants with similar interests often attend social events within these specific clusters. However, there are also centrally positioned technological topics with broader appeal, such as E-commerce, Informatics, and Web Analysis, as well as less specialized topics that are often combined with other groups, such as Data Integration, Study Group, or Concurrent Programming.

1.2 Microscopic Mechanisms of Social Differentiation.

The objective of the study is to distinguish the microscopic mechanisms that may potentially drive clustering and differentiation in Meetup technological communities. An individual's decision of choosing group memberships is primarily constrained by the affinity of their technological interest to the range of activities provided by the groups, and by the induction through social connections which pass on group information from members to non-members. Although macro-level social differentiation can be influenced by external factors such as the influx of new participant cohorts (DellaPosta and Nee 2020) or external market forces (Fligstein and Dauter 2007), our focus is directed towards examining the endogenous mechanisms operating at the micro and behavioral level.

The cultural affinity measures the relative proximity of cultural profiles between an individual and a group in a resource space, which is a multidimensional space of resources available to groups within a given environment (Hannan and Freeman 1993). A resource space describes the configuration of understandings that are held and enacted by both individuals and groups within a knowledge field (Hannan and Freeman 1993; McPherson 2004; Mohr and Guerra-Pearson 2010). In the context of Meetup technological communities which are characterized by voluntary and geographically unconstrained participation with low barriers, the resource space is shaped by the perceptions and actions of community members towards technological topics. Figure 2(a) shows a stylized depiction of a resource space. Notably, a resource space shares similarities with Blau space in the McPhersonian framework of ecology of affiliation, where dimensions represent sociodemographic attributes that shape individuals' social connections (McPherson 1983, 2004; McPherson and Ranger-Moore 1991). However, the dimensions of a resource space represent technological topics, and the relative positions of individuals and groups in the space indicate their cultural affinity.

[Figure 2 to be inserted here]

The process of adopting a group membership encompasses not only obtaining nominal membership, but also embracing the cultural norms, values, and expectations of the group. In order to effectively integrate into the group, prospective members are expected to have certain familiarity with the group-specific culture, including its objectives, goals, rules, daily activities, and behavioral expectations. However, these cultural demands may limit the pool of potential members and restrict the group’s reach to a broad audience. Hence, it is anticipated that individuals who are in closer proximity to a group in the resource space are more likely to assimilate and adopt the cultural norms and values of that group, therefore, they are more likely to join the group.

H1: The likelihood of individuals adopting group membership is positively related to their level of affinity for the group.

Similar to the diffusion process observed in the adoption of new products or the spread of unpopular opinions, the diffusion of membership of conventional in-person groups relies on network connections (DiMaggio and Garip 2012; McAdam 1990; McPherson et al. 2001). We consider that co-participating ties, i.e., connections or relationships formed between individuals who engage in shared activities or events, is a special form of weak ties (Granovetter 1973). They typically involve frequent and sustained interaction, allowing for the exchange of information, resources, and support among the connected individuals. These ties serve as channels through which potential members receive information about the benefits, activities, and opportunities associated with joining a particular group. Social network ties also provide a sense of trust and credibility, as recommendations or endorsements from trusted individuals can increase the perceived value and attractiveness of group membership. Extensive research has documented the importance of peer influence in membership growth in a variety of organizational contexts, including voluntary associations, revolutionary insurgence (Gould 1991), social movements (McAdam 1990; Snow et al. 1980), online discussion boards (Backstrom et al. 2006), and so on.

A high level of network density can also create tensions and conflicts in members' personal networks due to the limited time and cognitive capacity to maintain relationships. Thus, an individual who has friends who participate in the same Meetup group would be pressured to join to avoid isolation from the group. By extension of the logic, the likelihood of inducing membership adoption increases as two individuals participate in more events together in the past. Figure 2(b) illustrates the potential impact of social influence that the focal individual may have, as a result of their multiple connections with ingroup members.

H2. The participation in shared events within voluntary communities has a positive relationship with the spread of membership adoption between individuals.

The network effect of recruitment is expected to be strong between individuals and groups that are proximate in the resource space due to several reasons. First and foremost, proximity enhances accessibility, making it easier for individuals and groups to connect and establish relationships. When individuals and groups are culturally proximate, they can more readily interact and communicate, leading to a higher likelihood of recruitment (Putnam 2000; Stolle 1998). This ease of accessibility facilitates the initial contact and fosters a sense of familiarity and proximity that can be conducive to recruitment efforts. Moreover, culturally proximate individuals and groups exhibit shared interests and objectives, fostering a natural affinity. This alignment enhances the appeal and relevance of groups, increasing the likelihood of recruitment. The resulting sense of belonging and purpose further reinforces the network effect.

Information flow plays a crucial role in driving the strong network effect of recruitment among proximate individuals and groups (Contractor, Wasserman, and Faust 2006). Proximity enables efficient exchange and dissemination of information, allowing for greater exposure to each other's behavior, activities, and achievements. This enhanced information flow amplifies awareness and visibility of proximate groups, contributing to the network effect. Proximity in resource space also provides access to relevant information, opportunities, and advancements, serving as a catalyst for recruitment efforts. The intensified information exchange strengthens the network effect as individuals recognize the benefits of participating in a well-connected and informed network. Thus, we expect the moderating effect of cultural affinity on network effect.

Hypothesis 3. In digital technology-assisted voluntary associations, cultural affinity compounds the network effect, producing a higher level of membership adoption.

2. Material and methods

2.1. The Meetup.com Dataset

By 2019, the Meetup platform had accumulated over 44 million users and over 330,000 groups hosting over 84,000 events per week, with a global reach similar to its competitors including Facebook, YouTube, Instagram, TikTok, and Twitter throughout 190 countries and 2,000 cities (PR Newswire 2020). In New York City, for example, membership and the number of events have both increased exponentially since the inception of the website. More than 50,000 members participated monthly, and the number of events rose to over 20,000 in 2019.

While digital technologies have greatly aided the access to valuable group information, forming social connections in in-person events is crucial in spreading groups' cultures, rules, atmosphere, etc. that are spread via word-of-mouth. However, few studies have systematically examined the extent to the influence of digital technology on offline membership recruitment. As a novel mode of participation and social networking, the Meetup platform is an environment in which innovation dissipates and entrepreneurship forms (Hsiao 2021). It has a large and growing user base actively participating in lively groups and communities. In addition, the website provides a multitude of categories and topics, from which anyone can find voluntary groups that interest them. Researchers have compared the categories on Meetup to voluntary association categories used in conventional surveys, such as the General Social Survey (GSS), and found that Meetup contains categories that are newer (e.g., environment groups and support groups) and more informal (e.g., information sharing or socializing) (Paxton and Rap 2016). Furthermore, the detailed temporal data at the individual, event, and group levels enable us to study the evolution of topics, identify the mechanisms by which ecological conditions affect the clustering of groups, and the interplay between the aggregated-level social structure and micro-level behaviors over time.

The dataset used in this study focuses on membership recruitment of technology groups between 2017 and 2018. We downloaded the data from Meetup.com using its Application Programming Interfaces (APIs) during the summer of 2019. We selected two high-tech industrial clusters located in New York City and the San Francisco Bay Area, which include the cities of San Francisco, San Jose, and Oakland. For each cluster, all the zip codes were identified manually and the APIs were used to find all the identification numbers of groups located within its 10-mile radius. By searching the group identification numbers, we collected group-level information, including the group bio, organizers, the groups' self-identified topics, and event information, such as the event's address, latitude and longitude, timestamp, descriptions, and attendee lists. With the attendee list for each event, we compiled groups' active membership. By regrouping the attendance data from the event, we can deduce the information on group memberships and event attendance on Meetup for each participant.

2.2.Methods

In this study, we aim to achieve two objectives using recruitment data from technology groups in the prominent technology hubs of SFBA and NYC during 2017-2018. Firstly, we aim to differentiate the impact of peer influence from homophily in shaping an individual's membership decision. Secondly, we seek to estimate the effect of user's affinity for groups on membership induction and its interaction with social influence. This section provides an in-depth explanation of the estimation strategies employed, the construction of key variables, and the statistical analyses conducted for this purpose.

2.2.1. Matching-based Estimation

Our estimation goal is to ascertain the probability of an individual user attending an event of a new group, given prior interactions with existing members. The key challenge lies in the confounding process of homophily, whereby individuals may join a group due to shared interests with unconnected members, rather than a genuine convergence of interests facilitated by network connections with members (i.e., previous co-participations in other groups). To distinguish network effect from homophily, we devise a matching-based method to match individuals based on their underlying preferences in technological interest, independent of their network status of having or not having connections with the members of the group.

To determine the network effect of recruitment to *a target group*[1], denoted as *G*, we categorize all individuals who were not members of the target group (i.e., those who had not participated in the target group's events in 2017) into two groups: a treated group and an untreated control group. We define the *treatment status* as those who had been active and never participated in the target group G, but had interacted with people who were members of the target group. We denote the year of 2017 as the network formation period, in which users' treatment status is assigned, and the year of 2018 as the recruitment period, in which adoption

[1] Groups with a size smaller than or equal to 100 have been omitted from the analytic sample to ensure variation in the group-specific user participation. Smaller groups have fewer participants to influence non-members to start with, making it more likely to have no variation in the adoption outcome in the prospective members. We provide the participation frequency distribution by group size in Appendix A.

outcomes are measured[2]. The estimation of the network effect aims to evaluate whether, on average, users in the treated group have a higher probability of participating in the target group than those who are untreated.

[Figure 3 to be inserted here]

The estimation procedure, illustrated in Figure 3, is to calculate the probability of the non-member *i*, represented by blue circle, joining the target group *G* in 2018. The treatment status is indicated by the connections that user *i* has with ingroup members. The social connection between a pair of users is the frequency of co-participation in events in 2017. Specifically, individual *j* is considered an alter of *i* if *i* and *j* co-participated in at least two events in 2017. The binary adoption outcome is determined by user *i*'s event participation in the target group in 2018[3]. We repeat the estimation procedure for each group. Thus, individuals' treatment statuses are group-specific, implying that users' treatment statuses are assigned separately for each target group based on their participation history.

2.2.2. *Why Matching: How Matching Methods Improve Causal Inference*

In the counterfactual framework (Morgan and Winship 2014), the causal effect of a treatment is used to compare the outcome of an individual under the treatment condition to the outcome of the same person if they were not to receive the treatment. However, the same person cannot be assigned simultaneously to both treatment and control conditions. The missing observation of the counterfactuals poses a fundamental challenge in causal inference (Holland 1986). While it is infeasible for a person to experience both treatment and untreated conditions simultaneously, a

[2] The one-year time window for the treatment status data allows for the development of meaningful interactions between egos and alters, while ensuring that participation is not influenced by factors other than social influence from a particular alter.
[3] Details on the construction of the analytic sample are available in the Appendix.

reasonable approximation is to randomly assign individuals to treatment and untreated groups, ensuring that the persons have the identical background characteristics except for the treatment assignment.

The assumption of strongly ignorable treatment assignment is crucial in studies utilizing observational data. It requires that the treatment assignment and the outcomes are independent, given the confounding variables (Rosenbaum and Rubin 1983; Rubin 1978). In observational data, however, users assigned to treated and untreated groups may systematically differ based on treatment status, thus violating the key assumption. In the context of our Meetup dataset, confounding factors, such as interest in technological topics, geographic distance to event locations, and time availability may affect both the assignment of the treatment (i.e., exposure to influence from group members) and the outcome (i.e., participation in the new group), leading to potential bias in the estimation of the treatment effect (Aral et al. 2009; Holland 1986; Shalizi and Thomas 2011).

The fundamental premise of our estimation strategy is to match the treated users to their counterfactual counterparts who closely resemble them in terms of their propensity to receive the treatment (Rosenbaum and Rubin 1983; Stuart 2010) using behavioral data. After matching, any difference in the outcome between the treated and control groups can be attributed to the treatment status. In our specific study, we focus on estimating the average treatment effect on the treated (ATT), which captures the effect of treatment on users' participation outcomes only for those who, in fact, receive treatments (i.e., having non-zero adopter alters)[4].

[4] Given that we exclude unmatched treated users and those outside the common support, the effect we estimate technically represents the average treatment effect in the remaining matched sample (ATM). However, it is important to note that the calculation of marginal effects is based on the remaining treated sample, resulting in estimations similar to the average treatment effect.

2.2.3. *The Matching Procedure*

Using propensity score matching (PSM), we calculate the propensity scores for receiving a treatment at the individual level in a logistic regression, conditioned on a vector of matching variables, as discussed in more detail in the next section. Due to the group-specific nature of the analysis with the aim at estimating the effect at varied level of user-group affinity, the logistic regression is conducted separately for each group and at each level of affinity (i.e., low, medium, and high). The propensity score matching process is more likely to match treated and untreated individuals when their propensity scores have closer values[5].

Within each subsample, we conduct matching by pairing each treated user with up to five untreated users who have the closest propensity scores. The selection criteria also ensure that the distance between the treated and untreated users' propensity scores is within 0.2 times the standard deviation of the propensity scores within the treatment group (Austin 2011; Stuart 2010). Untreated users are selected with replacement in our study, allowing for multiple matches between untreated users and treated users within the same group. Unmatched users or observations that fall outside of the common support (i.e., non-overlapped areas within the overall propensity score distribution of the treated and the untreated) are dropped from the resultant matched samples. Each treated user is assigned a weight of one, while the weight assigned to each matched control user is inversely proportional to the number of control users,

---

of the treated. It is essential to acknowledge that these effects cannot be generalized to the entire treated population.

[5] To make sure that the chosen method results in the best matched sample, we tested the chosen method, nearest neighbor with replacement, against other matching methods including nearest neighbors without replacement, full matching, and optimal matching. The chosen method resulted in the most balanced sample with the least number of covariates with standardized mean differences between the treated and control sample greater than .1 after matching and the lowest average of standardized mean differences of the top 10 most unbalanced covariates.

aiming to account for heterogeneity in the number of controls across treatments (Stuart 2010). To improve the balance of the covariates, we also employ exact matching to match treated users with untreated users who have the same level of activeness[6]. The balance diagnostics, presented in Appendix B, demonstrate that the PSM approach effectively balances the matching variables based on treatment status.

2.2.4. *Constructing Matching Variables*

We use the technological topics to match users, and these topics are extracted from the groups in which each user has participated during the network formation period. In the context of the Meetup knowledge community, the past behavioral data might be more reliable indicators of individuals' understanding of the resource space than demographic traits, as the active population is relative homogeneous in race and education. The topic profile is operationalized as a user–topic matrix, with rows representing users and columns representing topics. An element of the matrix indicates a user's interest intensity in a particular topic. On the platform, the organizers of each group assign multiple topic labels that correspond to the technological topics around which the group's meetup events are organized.[7] Through participation in different groups, a user is viewed to exhibit varying probabilities or interests in the topics on Meetup.

We use a semi-supervised topic model called Labeled Latent Dirichlet Allocation (L-LDA) (Ramage et al. 2009) to derive individuals' topic profiles. A topic model is an algorithm

[6] We also considered gender as a variable on which to conduct exact matching. However, due to the low percentage of females in the sample, exact matching on gender did not improve balance of the matching.

[7] For example, a group might use labels such as "Data Science," "Women in Tech," "Arduino," and "Cryptocurrency." If a user attends events from a group whose topics include "Data Science," "Women in Tech," and "Data Mining" as well as from another group whose topics include "Data Science" and "Python Programming," then the user's topic profile would be ("Data Science":2, "Women in Tech":1, "Data Mining":1, "Python Programming":1).

that reveals the topics or themes within a collection of documents based on the distribution of words in these documents (Blei 2012). The model can cluster words into distinct topics by analyzing their co-occurrences in documents. In the context of Meetup groups, a group can be considered as a "document," while the attending members represent the "words" in that document. L-LDA, as a variant of the topic model, differs from the traditional topic model in that it solves the credit attribution problem in topic models, that is, how to determine within a document which set of words are responsible for a topic label (See Appendix C: LDA and Labeled LDA for extensive discussions on topic modeling). Following the procedure outlined in L-LDA, we first create a matrix with rows representing users and columns representing groups. The elements in the matrix represent the frequencies of participations in a specific group. Based on users' participation history in groups and the given correspondence between topic labels and groups, L-LDA generates a vector of weights for each user, with each weight being a topic interest probability that a person is associated with for a corresponding topic for all the available topics.

In addition to topic profiles, we include gender, geographic distance, and level of activity as matching variables. We infer gender of the users based on the first names of their registered username using an R package *gender* (Blevins and Mullen 2015). This package utilizes historical records from the Social Security Administration to determine the gender of a given name. In particular, the *gender* package provides the probabilities of a name given to a male and a female at birth, with a combined probability of one. The package then assigns the gender category with the higher probability to a given name. We define the distance between a user and a group as the distance between the user's average activity location and the average event location of the target group. We determine the average location by computing the mean latitude and longitude from the GPS coordinates of all events attended by the user or hosted by the group in 2016 and 2017. We then calculate the geodesic distance in kilometers between the user's mean activity location the group's mean event location. The level of activity is the number of events that a user attended in 2017.

2.2.5. *The Moderating Variable*

User–group affinity. To determine the affinity between an individual and a target group, we begin by computing the target group's topic profile as the mean over all its users' topic profiles in the network formation period of 2017. Then, the user's affinity for a group is calculated as the correlation between two topic profiles. Affinity levels below the 33rd percentile of the distribution are classified as low, those between the 33rd and 67th percentiles as medium, and all remaining levels as high. This categorization allows for a clear distinction between different levels of affinity within the dataset.

Table 1 shows the summary statistics of the main variables used in the matching procedure and statistical analyses. Table 1a shows the participation outcome frequencies by treatment status and user–group affinity for San Francisco and New York. Table 1b shows the descriptive statistics of the main matching variables.

[Table 1 to be inserted here]

2.3. Statistical Analysis

We proceed with a sequence of three analytical steps, each building on a previous step. The first step involves assessing the overall effect of social influence by comparing the network effect estimated using data where homophilous selection is not accounted for to the network effect that accounts for homophilous selection. To this end, we first estimate the social influence effect on membership adoption in treated users and random untreated users without accounting for the vector of topic interests and other matching variables (See Figure 3 top panel). In this *baseline matching*, each treated user is randomly matched with five control users in the same group without conditioning the match on any confounding covariates (Aral et al. 2009). Similar to PSM, each treated user is assigned a weight of 1, while each control user is assigned a weight equal to the inverse of the total number of controls matched to the corresponding treated user (i.e., 1/5 = 0.2). We estimate the effect of social influence by using the ratio between the weighted percentage of the treated adopters and the weighted percentage of the untreated adopters. Specifically, the ratio represents the percentage of treated users who ended up participating in a given group divided by that of the untreated users. A higher ratio indicates a larger effect of the treatment on participation, signifying a higher percentage of treated adopters of group *G* relative to the percentage of untreated adopter. Since baseline matching does not eliminate the dependency between the selected covariates and the treatment status, the resulting ratio is the total effect of social influence *and* homophilous selection in producing local clustering of participatory behavior (See Figure 4 upper panel).

Second, we estimate the effect of social influence on membership adoption by comparing the treated and untreated samples matched through PSM, which eliminates the confounding effects of homophily between social influence and membership adoption. Similarly, we calculate the ratio of the percentage of treated adopters and the percentage of untreated adopters as the estimate of social influence. To assess the upward bias caused by homophily, we compare the differences in the effects of social influence using the data generated by these two matching methods (See Figure 4 bottom panel).

Third, having properly accounted for the confounding effects using PSM, we examine the effect of social influence on the participation outcomes using the matched sample from PSM in a series of logistic regressions. In the first model, we estimate the overall average treatment effect on the treated. In the second model, we estimate the moderating effect of user–group affinity on the treatment effect. For both models, we use cluster-robust standard error to account for heteroscedasticity in standard errors caused by repeated matched counterfactuals and also dependence between observations within matched pairs and potential group memberships.

3. Results

Figure 4 shows the treatment effects estimated through Baseline Matching vs. PSM at different treatment thresholds. The top panel illustrates the change in the relative ratios of the percentage of adopters in treatment groups and in control groups for baseline matching (blue dashed line) and PSM (red solid line). For example, in San Francisco, the baseline matching is 4.7 at the lowest level of the treatment threshold. This indicates that, on average, the percentage of members who ended up participating in a target group from the treated set is 4.7 times that of the untreated set when using baseline matching (i.e., treated and untreated are randomly matched). The horizontal axis represents the treatment threshold, which reflects the varying levels of exposure to peer influence (i.e., being treated). We define the treatment threshold as the minimum number of shared events required between two users for them to be considered connected and thus treated. By varying the treatment threshold from 2 to 100, we can observe the corresponding changes in the performances of two matching methods[8]. In San Francisco (upper left panel in Figure 4), when the treatment threshold is set at 2 (i.e., users are considered to be in

[8] In this section of the analysis, we incorporate multiple treatment levels. However, in subsequent analyses, we define treatment as having 2 or more shared events with an eligible alter.

the treatment group if they have 2 or more shared events with an alter who has participated in group *G* in 2017), the estimated ratio between the proportions of adopters in treatment groups and in control groups is 4.7 using baseline matching, compared to 1.7 using PSM, resulting an overestimation of 276%. This suggests that failing to account for between-user homophily in the estimation of social influence leads to upward bias. As the treatment threshold increases, the difference in percentages of adopters between treatment and untreated groups remains significantly large and stable. Similar patterns, albeit less pronounced, are observed in the estimation of social influence in New York City.

[Figure 4 to be inserted here]

The bottom panel of Figure 4 illustrates the magnitude of the upward bias that is observed in the baseline matching approach when estimating treatment effects, compared to the PSM sample. This is obtained by dividing the baseline matching line (in blue) by the PSM line (in red) from the top panel. For users with exposure to 2 or more shared events with adopter alters (i.e., the lowest treatment threshold), the treatment effect estimated using baseline matching (4.7) is 176% (=(4.7/1.7-1)*100) higher than that using PSM (1.7) in San Francisco and 94% (=(3.6/1.85 - 1)*100) higher in New York. The upward bias as ratio, as shown in the bottom panel of Figure 4, is highest when the threshold of number of shared events is low and decreases as that threshold increases. Despite the varying amount of upward bias, it is evident that homophily significantly contributes to the clustering of user's adoption of the same memberships. Failing to adequately account for it can lead to an exaggeration of treatment effects.

Using the matched samples via PSM, we formally analyze the effect of social influence on the adoption of the membership of a target group (Model 1) and the moderating effect of user-group affinity (Model 2). The non-linearity inherent in the logistic regressions suggests coefficients in logit or odds ratios do not correspond to a constant effect on the predicted probability of the dependent variable therefore should not be used for interpretation (Mize 2019). To better understand the substantive impact of the independent variables on outcome, we report the marginal effects in Table 2.

[Figure 5 to be inserted here]

[Table 2 to be inserted here]

Table 2 reports the marginal effect of treatment on the predicted probability of participation in target group *G*. Figure 5 shows the graphic presentation of results. In San Francisco, the predicted probability of participation without receiving any social influence is 0.00178 (that is a user without co-participations with existing members of *G* in 2017 has a probability of 0. 00178 of joining *G*), compared to 0.00299 for the same user who had exposed to the social influence. The average marginal effect of the treatment is 0.00121 (=0.00299-0.00178) with no overlap in the confidence intervals, suggesting a significant treatment effect. Similarly, in New York, the predicted probability of participation without receiving the social influence is 0.00174 compared to 0.00322 for the same individual if receiving the social influence. Consistent with the findings in San Francisco, individuals exposed to social influence have a higher probability of joining a group by 0.00148(=0.00322-0. 00174) compared to those who are not exposed. In summary, Hypothesis 2 is supported that participation in shared events within voluntary communities has a positive relationship with the spread of membership adoption between individuals.

Model 2 examines the interaction between group-specific treatment and user-group affinity on participation in group *G*. Table 3 shows the marginal effect of treatment on participation in *G* by levels of user-group affinity. Additionally, Table 4 reports the average marginal effect (i.e., the marginal effect of being treated minus that of being untreated) of treatment on participation in G by levels of user-group affinity. To visually depict these effects, Figure 6 plots the marginal effect by treatment status on the predicted probability of participating in *G* by levels of user-group affinity.

[Table 4 to be inserted here]

[Figure 6 to be inserted here]

For San Francisco, the effect of treatment on participation is amplified by the high levels of user-group affinity. As shown in Table 3 and Figure 6, among users with low affinity, the untreated group has a predicted probability of participation of 0.00050, whereas the treated group shows a probability of 0.00102, representing a substantial difference of 0.00052. Similarity, for users with medium affinity, the untreated group has a predicted probability of participation of 0.00123, while the treated group shows a probability of 0.00222, resulting in a significant difference of 0.00099. For users with high affinity, the untreated group has a predicted probability of participation of 0.00301, whereas the treated group exhibits a probability of 0.00484, a significant difference of 0.00182. Table 4 shows the average marginal effects of treatment by levels of affinity, underscoring the significant variation in treatment effects based on affinity levels. Specifically, results from Table 4 suggests that the treatment effect is more pronounced among users with higher levels of affinity towards the target group.

In the case of New York City, the effect of treatment on participation also varies by levels of user-group. As shown in Table 3, for low-affinity users, the untreated group has a predicted probability of participation is 0.00103 while the treated group has a probability of 0.00167, resulting in a significant difference of 0.00065. For medium-affinity users, the untreated group has a predicted probability of participation of 0.00160, whereas the treated group has a probability of 0.00287, a significant difference of 0.00127. For high-affinity users, the untreated group has a predicted probability of participation of 0.00250, while the treated group has a probability of 0.00493, a significant difference of 0.00243. Consistent with San Francisco, Table 4 further supports the finding that the treatment effect becomes significantly stronger among users with higher levels of affinity with the prospective group.

For both cities, exposure to treatment results in higher likelihood of participation when target group is similar to the user in the topic interests, supporting Hypothesis 3 which states that in digital technology-assisted voluntary associations, cultural affinity compounds the network effect, producing a higher level of membership adoption.

4. Discussions and Conclusions

Digital platforms that host in-person events, such as Meetup.com, have gained increasing popularity among entrepreneurs and knowledge workers seeking to exchange information and technology, as well as connect with potential startup partners or expand their professional network. However, the emergence of the new knowledge economy is intertwined with knowledge differentiation and fragmentation, wherein individuals possess specialized knowledge face growing challenges in communication and collaboration. To examine the microscopic mechanisms that can contribute to systemic differentiation, this study specifically investigates two social mechanisms influencing membership adoption: shared participation in events and cultural affinity between individuals and groups. Our results indicate that participation in shared events between a pair of individuals greatly increases the likelihood of contagion of membership of new groups, and the effect increases as the number of shared events increases. Our results further suggest that exposure to social influence from highly compatible groups significantly increases the probability of future participation.

Implications for Network Methodology

The methodological challenge in this study pertains to the accurate estimation of the network effect in co-participation relationships, which is a confluence of two mechanisms: network contagion, i.e., the diffusion of membership through the relationship, and homophily, i.e., membership adoption as a result of shared interests and overlapping group participation (McPherson, Smith-Lovin, and Cook 2001). The distinction between contagion and homophily in network diffusion has been established as a difficult task (Shalizi and Thomas 2011). In order to address this challenge, we have developed a matching-based inference procedure that enables us to estimate the probability of joining a tech group driven by co-participation formed in prior events. We model prior event participation in past using a semi-supervised topic model, specifically the Labeled Latent Dirichlet Allocation. The topic modeling approach reveals the themes and topics present in a collection of individuals' participation profiles through the distribution of topic labels in these profiles.

Our statistical procedure reveals a significant overestimation of the network effect in recruitment compared to previous methods. This finding highlights the importance of using rigorous procedure and behavioral data, rather than self-reported surveys, to obtain more accurate and reliable estimates. With this statistical tool now available, scholars have the capacity to explore a broader range of research questions regarding the underlying mechanisms that drive social differentiation in various domains, including voluntary associations, sociology of organizations, and related fields. By employing our procedure, researchers can gain valuable insights into the dynamics of social networks, information diffusion, and the formation of ties, thereby advancing our knowledge in these areas and fostering interdisciplinary collaborations.

Implications for Sociology of Diffusion and Social Integration

The findings of this research hold significant implications for distinguishing the social mechanisms for systematic differentiation and segmentation in user-driven voluntary association ecosystems. This study has attempted to estimate the effects of the shared event participation and individual-group cultural affinity as independent and distinct processes. However, it is acknowledged that they are likely to form a positive feedback loop, reinforcing each other's impact. For instance, from the standpoint of the organizers, preserving a stable and large membership is essential and can be accomplished through programs that reinforce the social networks within the membership. The shared history of participating in events can enhance the relationships between co-members, thereby increasing the likelihood of the transfer of information regarding new group initiatives or technological advancements. Concurrently, the occurrence of overlapping events and memberships can lead to an elevation in cultural and behavioral homogeneity. And as a result, when a co-member joins a new group, there is a high probability of a ripple effect of membership adoption via social connections, as the cultural disparity between the group and the individual is expected to be minimal.

Implications for Knowledge Economy

Our estimation approach has the potential to revise the understanding of the dynamics of knowledge economy, where informal associations of knowledge workers have become the crucial foci for the diffusion of informal knowledge, including know-hows, technology fads, entrepreneurial skills, social capitals, among many others. In the study, we aim to identify and analyze the social mechanisms that potentially contribute to knowledge differentiation, which is an essential part of the well-functioning economy. Establishing a balanced ecosystem within knowledge-sharing communities is essential. Thus, in order to prevent fragmentation within a knowledge system, it is crucial to employ an accurate estimation procedure that can effectively distinguish the relative influence of different microscopic mechanisms. However, further research is needed to validate the relative contributions of these mechanisms in producing knowledge formation and differentiation, as well as to explore the interplay between microscopic and macro-level mechanisms.

The rising popularity of digital platforms like Meetup.com and Facebook Groups has revolutionized knowledge sharing among individuals in various fields, including knowledge workers, entrepreneurs, and innovators. These platforms serve as accessible spaces for free exchange of technological advancements, expertise, and networking opportunities. Our analytical approach sheds light on the transformative influence of digital platforms in rejuvenating the knowledge economy, offering a potential research program that link micro-level behavioral mechanisms and global communication patterns.

Limitations of the Study

This study has limitations that require further research. Firstly, shared event participation in is assumed to be uniformly distributed. However, future studies must validate this assumption by establishing the correlation between co-participation and actual interpersonal relationships. In this paper, a threshold of two co-participated events was set to define a valid social tie, under the premise that the likelihood of meaningful social interaction increases with the number of events co-participated in. Future studies could employ a more nuanced approach to measuring social ties, taking into account the probability of social interaction weighted by the size of the event.

Furthermore, qualitative research is necessary to understand the extent of social interaction among Meetup users and its impact on the decision to join a new group.

Another limitation of this study is that we do not incorporate information regarding socio-demographic information and actual friendship networks, except considering gender in the matching process. While we acknowledge that socio-demographic information and actual friendship networks could enhance the accuracy of estimation, we posit that their influence may be diminished in associations facilitated by digital platforms. In these contexts, individuals have the opportunity to discover and engage with groups that may not be limited by traditional demographic boundaries or existing friendship circles. However, the extent to which individuals can truly escape these factors remains an empirical question that warrants further research attention.

Our statistical analyses do not explicitly incorporate the influence of platform algorithms, such as recommender systems, which utilize user-provided interests and location data to suggest relevant groups and events. We didn't include it not because they are not important, rather, they are unobservable to researchers. However, we believe that certain aspects of the impact of algorithms, particularly recommender systems, can be encompassed within the estimation of cultural affinity. This is because these algorithms are designed to enhance the process of information discovery and reinforce existing user preferences, thereby potentially amplifying the effect of cultural affinity (Adomavicius and Tuzhilin 2005).

REFERENCES

Adams, Philippa R., Julie Frizzo-Barker, Betty B. Ackah, and Peter A. Chow-White. 2019. "Meetups: Making Space for Women on the Blockchain." in *Blockchain and Web 3.0*. Routledge.

Adomavicius, G., and A. Tuzhilin. 2005. “Toward the next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions.” *IEEE Transactions on Knowledge and Data Engineering* 17(6):734–49. doi: 10.1109/TKDE.2005.99.

Aral, Sinan, Lev Muchnik, and Arun Sundararajan. 2009. “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):21544–49. doi: 10.1073/pnas.0908800106.

Audretsch, David B., and Maryann P. Feldman. 1996. “R&D Spillovers and the Geography of Innovation and Production.” *The American Economic Review* 86(3):630–40.

Austin, Peter C. 2009. “Balance Diagnostics for Comparing the Distribution of Baseline Covariates between Treatment Groups in Propensity-Score Matched Samples.” *Statistics in Medicine* 28(25):3083–3107. doi: 10.1002/sim.3697.

Austin, Peter C. 2011. “Optimal Caliper Widths for Propensity-Score Matching When Estimating Differences in Means and Differences in Proportions in Observational Studies.” *Pharmaceutical Statistics* 10(2):150–61. doi: 10.1002/pst.433.

Backstrom, Lars, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. 2006. “Group Formation in Large Social Networks: Membership, Growth, and Evolution.” Pp. 44–54 in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, *KDD ’06*. New York, NY, USA: Association for Computing Machinery.

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348(6239):1130–32. doi: 10.1126/science.aaa1160.

Bell, Daniel. 1976. *Coming Post-Industral Society*. Basic Books.

Blei, David M. 2012. "Probabilistic Topic Models." *Communications of the ACM* 55(4):77–84. doi: 10.1145/2133806.2133826.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3(Jan):993–1022.

Blevins, Cameron, and Lincoln Mullen. 2015. "Jane, John … Leslie? A Historical Method for Algorithmic Gender Prediction." *Digital Humanities Quarterly* 009(3).

Boyd, Danah M., and Nicole B. Ellison. 2007. "Social Network Sites: Definition, History, and Scholarship." *Journal of Computer-Mediated Communication* 13(1):210–30. doi: 10.1111/j.1083-6101.2007.00393.x.

Carley, Kathleen. 1986. "Knowledge Acquisition as a Social Phenomenon." *Instructional Science* 14(3):381–438. doi: 10.1007/BF00051829.

Carley, Kathleen. 1991. "A Theory of Group Stability." *American Sociological Review* 56(3):331–54. doi: 10.2307/2096108.

Castells, Manuel. 2011. *The Rise of the Network Society*. John Wiley & Sons.

Contractor, Noshir S., Stanley Wasserman, and Katherine Faust. 2006. "Testing Multitheoretical, Multilevel Hypotheses about Organizational Networks: An Analytic Framework and

Empirical Example." *The Academy of Management Review* 31:681–703. doi: 10.2307/20159236.

DellaPosta, Daniel, and Victor Nee. 2020. "Emergence of Diverse and Specialized Knowledge in a Metropolitan Tech Cluster." *Social Science Research* 86:102377. doi: 10.1016/j.ssresearch.2019.102377.

DellaPosta, Daniel, Yongren Shi, and Michael Macy. 2015. "Why Do Liberals Drink Lattes?" *American Journal of Sociology* 120(5):1473–1511. doi: 10.1086/681254.

DiMaggio, Paul, and Filiz Garip. 2012. "Network Effects and Social Inequality." *Annual Review of Sociology* 38(1):93–118. doi: 10.1146/annurev.soc.012809.102545.

Durkheim, Émile. 1893. *The Division of Labor in Society*. New York: Free Press.

Flaxman, Seth, Sharad Goel, and Justin M. Rao. 2016. "Filter Bubbles, Echo Chambers, and Online News Consumption." *Public Opinion Quarterly* 80(S1):298–320. doi: 10.1093/poq/nfw006.

Fligstein, Neil, and Luke Dauter. 2007. "The Sociology of Markets." *Annual Review of Sociology* 33(1):105–28. doi: 10.1146/annurev.soc.33.040406.131736.

Goffman, Erving. 1983. "The Interaction Order: American Sociological Association, 1982 Presidential Address." *American Sociological Review* 48(1):1–17. doi: 10.2307/2095141.

Gould, Roger V. 1991. "Multiple Networks and Mobilization in the Paris Commune, 1871." *American Sociological Review* 56(6):716–29. doi: 10.2307/2096251.

Granovetter, Mark S. 1973. “The Strength of Weak Ties.” *American Journal of Sociology* 78(6):1360–80. doi: 10.1086/225469.

Hannan, Michael T., and John Freeman. 1993. *Organizational Ecology:* Cambridge, MA: Harvard University Press.

Holland, Paul W. 1986. “Statistics and Causal Inference.” *Journal of the American Statistical Association* 81(396):945–60. doi: 10.1080/01621459.1986.10478354.

Hsiao, Yuan. 2021. “Evaluating the Mobilization Effect of Online Political Network Structures: A Comparison between the Black Lives Matter Network and Ideal Type Network Configurations.” *Social Forces* 99(4):1547–74. doi: 10.1093/sf/soaa064.

Lupton, Deborah. 2014. *Digital Sociology*. London: Routledge.

Maaten, Laurens van der, and Geoffrey Hinton. 2008. “Visualizing Data Using T-SNE.” *Journal of Machine Learning Research* 9(86):2579–2605.

Mark, Noah. 1998. “Beyond Individual Differences: Social Differentiation from First Principles.” *American Sociological Review* 63(3):309–30. doi: 10.2307/2657552.

McAdam, Doug. 1990. *Freedom Summer*. Oxford, New York: Oxford University Press.

McPherson, J. Miller. 1983. “An Ecology of Affiliation.” *American Sociological Review* 48(4):519–32. doi: 10.2307/2117719.

McPherson, J. Miller. 2004. “A Blau Space Primer: Prolegomenon to an Ecology of Affiliation.” *Industrial and Corporate Change* 13(1):263–80. doi: 10.1093/icc/13.1.263.

McPherson, J. Miller, and James R. Ranger-Moore. 1991. "Evolution on a Dancing Landscape: Organizations and Networks in Dynamic Blau Space." *Social Forces* 70(1):19–42. doi: 10.2307/2580060.

McPherson, J. Miller, and Thomas Rotolo. 1996. "Testing a Dynamic Model of Social Composition: Diversity and Change in Voluntary Groups." *American Sociological Review* 61(2):179–202. doi: 10.2307/2096330.

McPherson, J. Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415–44. doi: 10.1146/annurev.soc.27.1.415.

Mize, Trenton. 2019. "Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects." *Sociological Science* 6:81–117. doi: 10.15195/v6.a4.

Mohr, John, and Francesca Guerra-Pearson. 2010. "The Duality of Niche and Form: The Differentiation of Institutional Space in New York City, 1888–1917." *Research in the Sociology of Organizations* 31:321–68. doi: 10.1108/S0733-558X(2010)0000031013.

Morgan, Stephen L., and Christopher Winship. 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. 2nd ed. Cambridge: Cambridge University Press.

Nee, Victor, and Lucas G. Drouhot. 2020. "Immigration, Opportunity, and Assimilation in a Technology Economy." *Theory and Society* 49(5):965–90. doi: 10.1007/s11186-020-09414-0.

Normand, Sharon-Lise T., Mary Beth Landrum, Edward Guadagnoli, John Z. Ayanian, Thomas J. Ryan, Paul D. Cleary, and Barbara J. McNeil. 2001. "Validating Recommendations for Coronary Angiography Following Acute Myocardial Infarction in the Elderly: A Matched Analysis Using Propensity Scores." *Journal of Clinical Epidemiology* 54(4):387–98. doi: 10.1016/S0895-4356(00)00321-8.

Paxton, Pamela, and Robyn Rap. 2016. "Does the Standard Voluntary Association Question Capture Informal Associations?" *Social Science Research* 60:212–21. doi: 10.1016/j.ssresearch.2016.05.003.

Powell, Walter W., and Kaisa Snellman. 2004. "The Knowledge Economy." *Annual Review of Sociology* 30(1):199–220. doi: 10.1146/annurev.soc.29.010202.100037.

PR Newswire. 2020. "Meetup Spins Off From WeWork." Retrieved April 19, 2022 (https://www.prnewswire.com/news-releases/meetup-spins-off-from-wework-301031766.html).

Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. Simon and Schuster.

Ramage, Daniel, David Hall, Ramesh Nallapati, and Christopher D. Manning. 2009. "Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-Labeled Corpora." P. 248 in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing Volume 1 - EMNLP '09*. Vol. 1. Singapore: Association for Computational Linguistics.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1):41–55. doi: 10.2307/2335942.

Rubin, Donald B. 1978. "Bayesian Inference for Causal Effects: The Role of Randomization." *The Annals of Statistics* 6(1):34–58.

Rubin, Donald B. 2001. "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." *Health Services and Outcomes Research Methodology* 2(3):169–88. doi: 10.1023/A:1020363010465.

Saxenian, AnnaLee. 1996. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128, With a New Preface by the Author*. Harvard University Press.

Shalizi, Cosma Rohilla, and Andrew C. Thomas. 2011. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies." *Sociological Methods & Research* 40(2):211–39. doi: 10.1177/0049124111404820.

Snow, David A., Louis A. Zurcher, and Sheldon Ekland-Olson. 1980. "Social Networks and Social Movements: A Microstructural Approach to Differential Recruitment." *American Sociological Review* 45(5):787–801. doi: 10.2307/2094895.

Spencer, Herbert. 1898. *The Principles of Sociology*. D. Appleton.

Stolle, Dietlind. 1998. "Bowling Together, Bowling Alone: The Development of Generalized Trust in Voluntary Associations." *Political Psychology* 19:497–525. doi: 10.1111/0162-895X.00115.

Stuart, Elizabeth A. 2010. “Matching Methods for Causal Inference: A Review and a Look Forward.” *Statistical Science : A Review Journal of the Institute of Mathematical Statistics* 25(1):1–21. doi: 10.1214/09-STS313.

Zablocki, Benjamin David. 1980. *Alienation and Charisma: A Study of Contemporary American Communes*. Free Press.

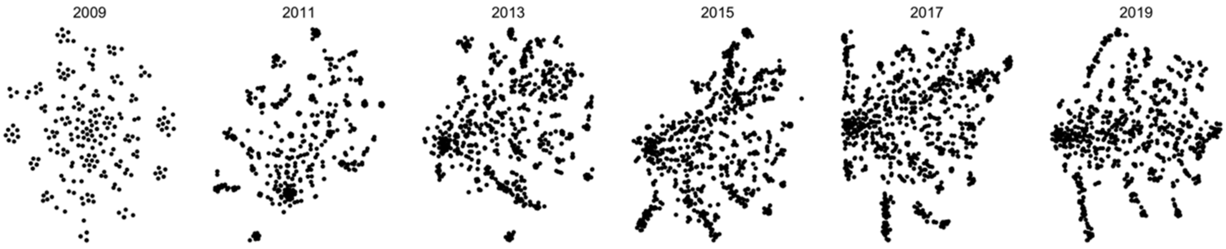

Figure 1. The Increasing Differentiation of Topic Structures in San Francisco between 2009 and 2019. Each dot represents a technological topic, and the spatial proximity between dots reflects the likelihood of overlapping participation by the same individuals across different topic groups.

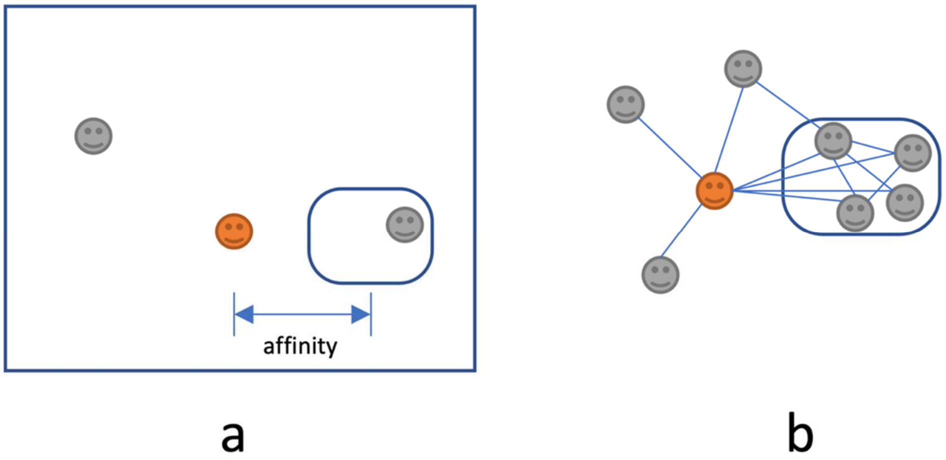

Figure 2. Resource Space and Social Structure. The resource space is a multidimensional space defined by technological topics that individuals are associated with. The social structure is inferred through co-participations in groups.

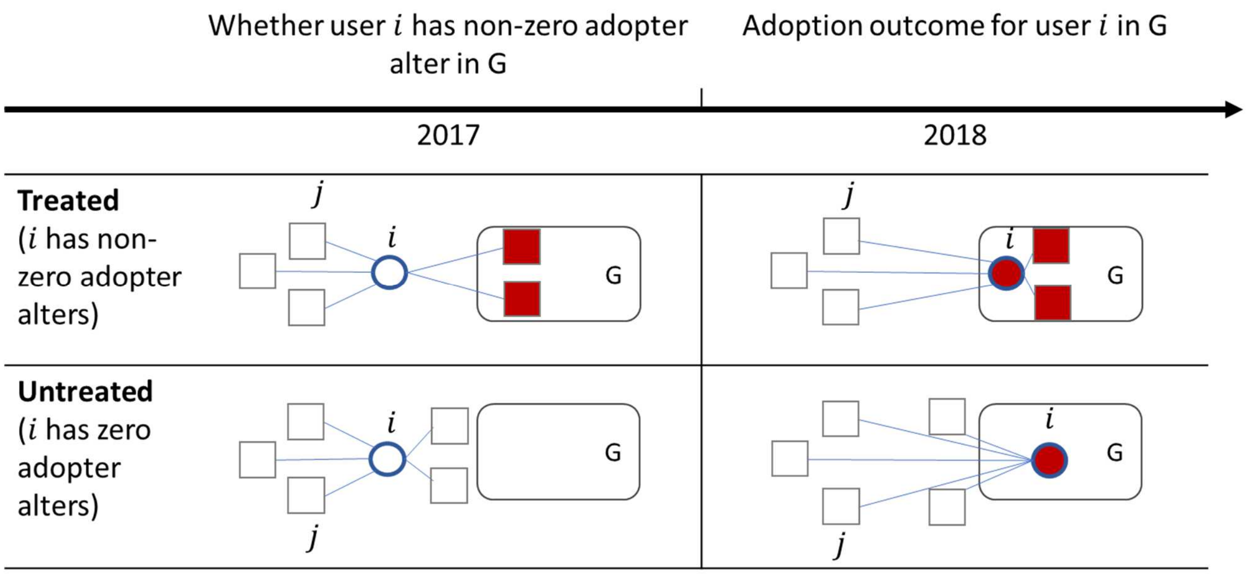

Figure 3. Estimation Procedure of Social Influence on Adoption of a Membership.

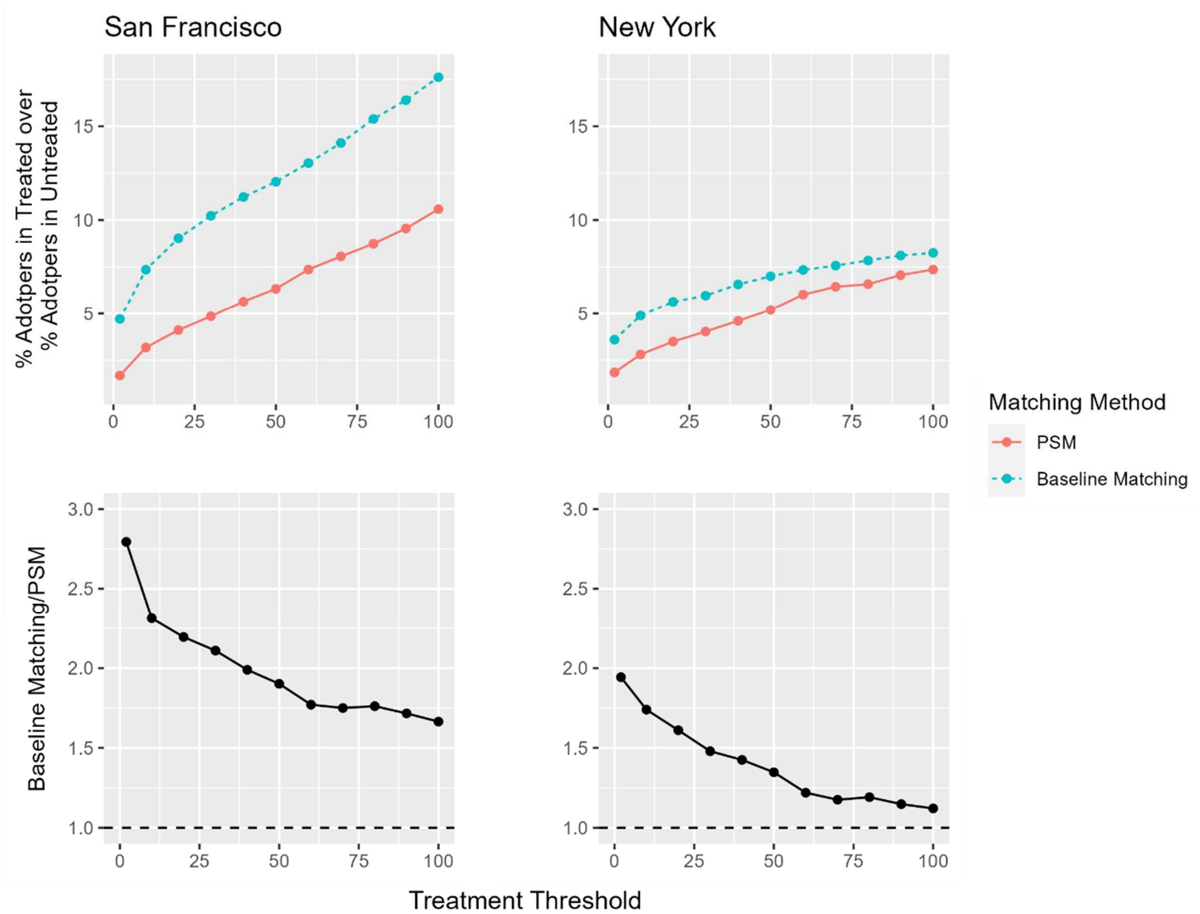

Figure 4. Treatment effects estimated using Baseline Matching vs. PSM by Influence Thresholds. Note: The treatment threshold indicates the minimum number of events a treated user has to co-participate with alters in order for the exposure to be defined as treated.

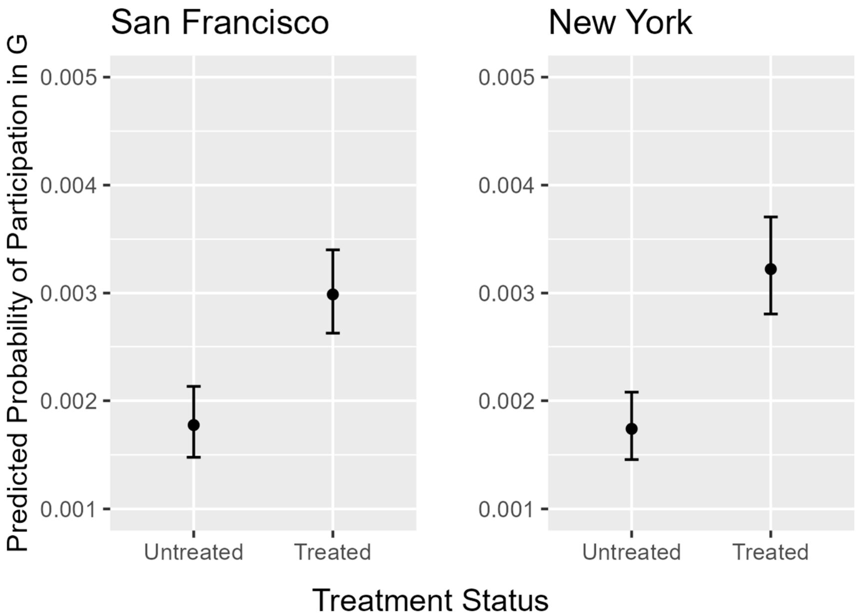

Figure 5. Overall Treatment Effect.

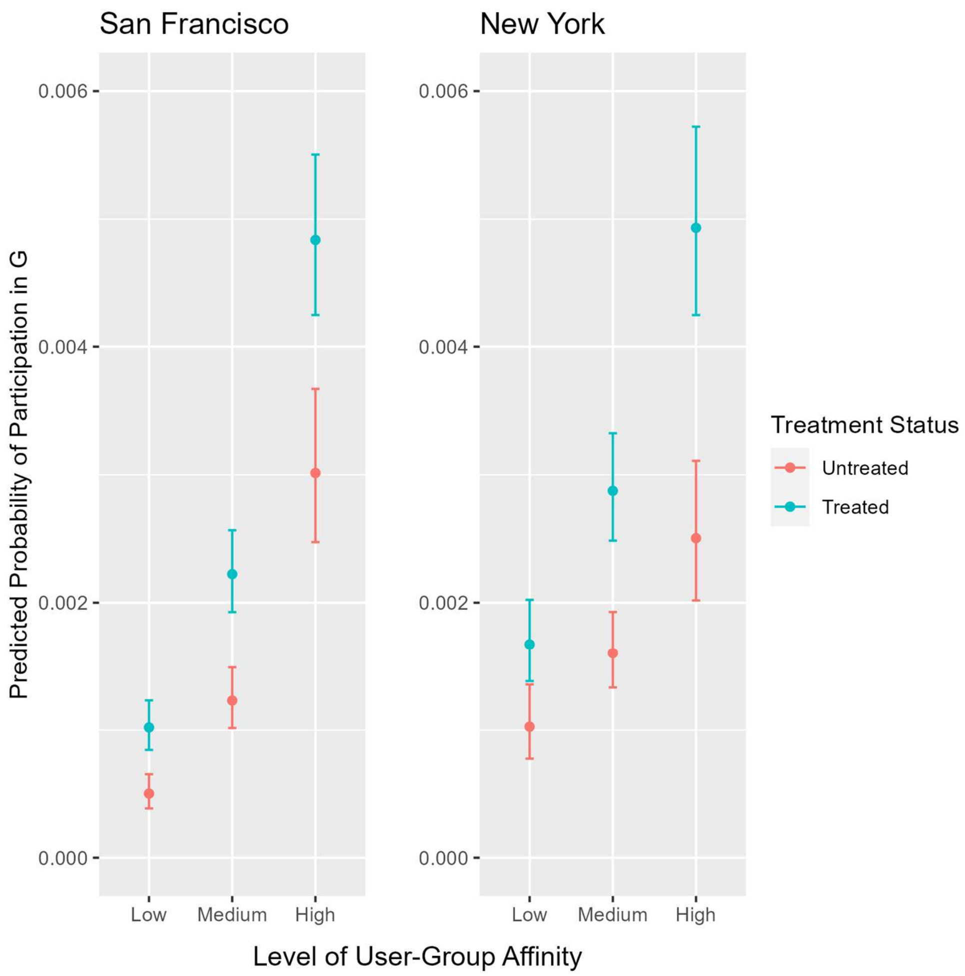

Figure 6. Marginal Effects of Treatment by Levels of User-Group Affinity.

Table 1. Summary Statistics

1a. Participation Outcomes by Treatment Status and User-Group Affinity

	Treatment Status	User-Group Affinity	N Participated	Total	% Participated
San Francisco	Untreated	Low	326	505,500	0.064%
		Medium	708	690,273	0.103%
		High	2,549	822,312	0.310%
	Treated	Low	600	505,500	0.119%
		Medium	1,367	690,273	0.198%
		High	4,062	822,312	0.494%
New York	Untreated	Low	348	312,111	0.111%
		Medium	510	352,187	0.145%
		High	915	354,280	0.258%
	Treated	Low	570	312,111	0.183%
		Medium	917	352,187	0.260%
		High	1,795	354,280	0.507%

Note: N refers to the effective sample size, i.e., the number of treated users matched with an equal weighted number of counterfactuals.

1b. Descriptive Statistics of the Matching Variables in the Matched Sample

	Variable	Mean	Median	Sd	Min	Max
San Francisco	Female	0.22	0	0.42	0	1
	Activeness	9.2	8	6.4	2	191
	User-Group Distance	155	31	704	0	9,353
New York	Female	0.25	0	0.43	0	1
	Activeness	8.3	7	5.1	2	94
	User-Group Distance	83	2.3	608	0	12,214

Note: Means are based on the number of treated users matched with an equal weighted number of counterfactuals

Table 2. Model 1. Marginal Effect of Treatment on Participation in *G*

	Treatment Status	ME of Treatment	P-value	95% CI	
San Francisco	Untreated	0.00178	<0.001	0.00148	0.00213
	Treated	0.00299	<0.001	0.00262	0.0034
New York	Untreated	0.00135	<0.001	0.00114	0.00161
	Treated	0.00338	<0.001	0.00292	0.00392

Table 3. Model 2: Marginal Effect (ME) of Treatment by Levels of User-group Affinity

	Treatment Status	Affinity	ME of Treatment	P-value	95% CI	
San Francisco	Untreated	Low	0.00050	<0.001	0.00039	0.00065
	Treated	Low	0.00102	<0.001	0.00084	0.00123
	Untreated	Medium	0.00123	<0.001	0.00102	0.00149
	Treated	Medium	0.00222	<0.001	0.00193	0.00257
	Untreated	High	0.00301	<0.001	0.00247	0.00367
	Treated	High	0.00484	<0.001	0.00425	0.00550
New York	Untreated	Low	0.00103	<0.001	0.00078	0.00136
	Treated	Low	0.00167	<0.001	0.00138	0.00202
	Untreated	Medium	0.00160	<0.001	0.00133	0.00193
	Treated	Medium	0.00287	<0.001	0.00249	0.00332
	Untreated	High	0.00250	<0.001	0.00202	0.00311
	Treated	High	0.00493	<0.001	0.00425	0.00572

Table 4. Model 2: Average Marginal Effect of Treatment by Levels of User-group Affinity.

	Affinity	AME of Treatment	P-value	95% CI	
San Francisco	Low	0.00052	<.001	0.00039	0.00065
	Medium	0.00099	<.001	0.00082	0.00117
	High	0.00182	<.001	0.00132	0.00233
New York	Low	0.00065	<.001	0.00040	0.00089
	Medium	0.00127	<.001	0.00099	0.00155
	High	0.00243	<.001	0.00183	0.00302

Appendices

Appendix A. Frequency of Participation by Group Sample Size

We removed groups with fewer than 100 unique participants in 2017 from our analytical samples. Small groups often have limited connections with potential recruits, making it challenging to create matched samples of treatment and control groups, as the assignment of treatment status is determined by whether a potential recruit has certain numbers of connections with members of the target group.

The density plot below illustrates the distribution of group sizes. Excluding small groups reduces the number of target groups by roughly half. However, because we conducted estimations separately for each group, we are confident about the robustness of our results and their independence from the sample construction.

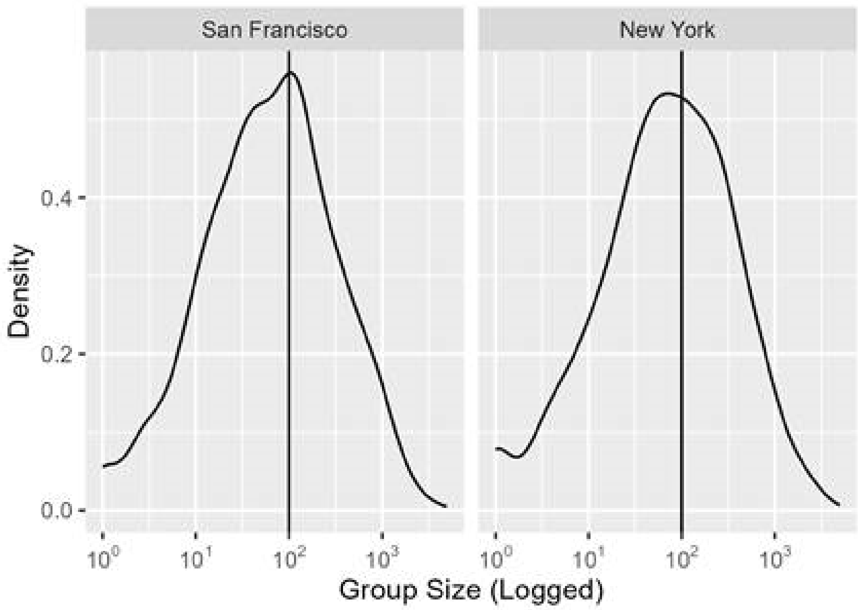

Figure A1. Density Plots of Group Size.

Appendix B. Assessment of Covariate Balance After Matching.

To assess the quality of the matching, we check the balance of the matching variables before and after matching to make sure that we have removed the confounding effects (Stuart 2010). That is, whether the distributions of the matching variables are similar between the treated and the untreated samples. We perform the balance assessment for each group and affinity level by examining the absolute standardized difference in means, d, between the treated and the untreated samples on each matching variable and the number of imbalanced variables after the matching. Due to large number of groups in the sample, we only present the summary statistics of the two tests.

Figure A2 shows the average standardized difference in means for the 10 most imbalanced variables over all the groups for San Francisco and New York. To obtain this distribution, we calculate the standardized mean difference, denoted as d, between the treated and untreated samples within each subgroup (i.e., group and user-group-affinity combination). We identify the top 10 variables with the largest absolute values of d, indicating the most imbalanced variables on which treated and untreated samples are matched. We then calculate the mean of these d values. A smaller value in d suggests a smaller difference between the treated and the untreated in terms of the distribution on the focal covariate, better balance, and less confounding effect from the focal covariate. If any matching variable within a group-user-group-affinity subgroup has a standardized difference in means greater than 0.2, we exclude the entire subgroup from the analysis [9].

For both San Francisco and New York, the means of d's for the 10 most balanced variables for the before-matching sample (density distribution in red) are higher than those for

[9] A total of 130 out of 1042 and 150 out of 1449 group-affinity combinations are dropped from New York and San Francisco, respectively due to imbalance.

the matched one (density distribution in blue). In the before-matching group, the means of d's for the 10 most imbalanced variables are higher than the more relaxed absolute standardized differences of means threshold of 0.25 (Rubin 2001; Stuart 2010) for most of the group-group-user-affinity subsamples. However, in the matched sample, the majority are below 0.1, indicating strong balance across all the group-affinity combinations overall.

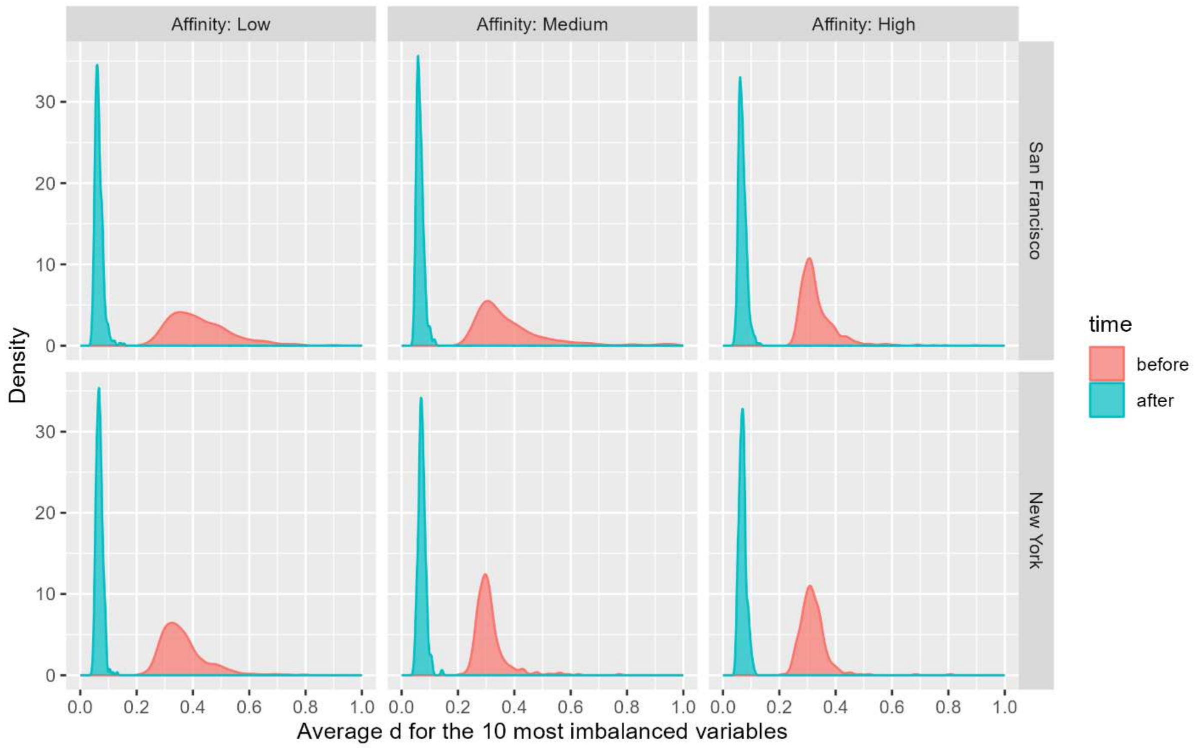

Figure A2. Average d for the 10 most imbalanced variables by level of affinity and city

In addition, Figure A3 presents the boxplots of variables exceeding the standardized difference threshold of 0.1, across groups-user-group-affinity subsamples (Austin 2009; Normand et al. 2001). In both cities, the matching significantly reduces the number of variables whose absolute standardized means difference between the treated and the untreated exceeds the specified threshold. The median number of variables exceeding 0.1 is 0 for low and medium user-group affinity levels in San Francisco and 1 for high user-group affinity level in San Francisco and in New York regardless of affinity levels (boxplots on the right in each panel).

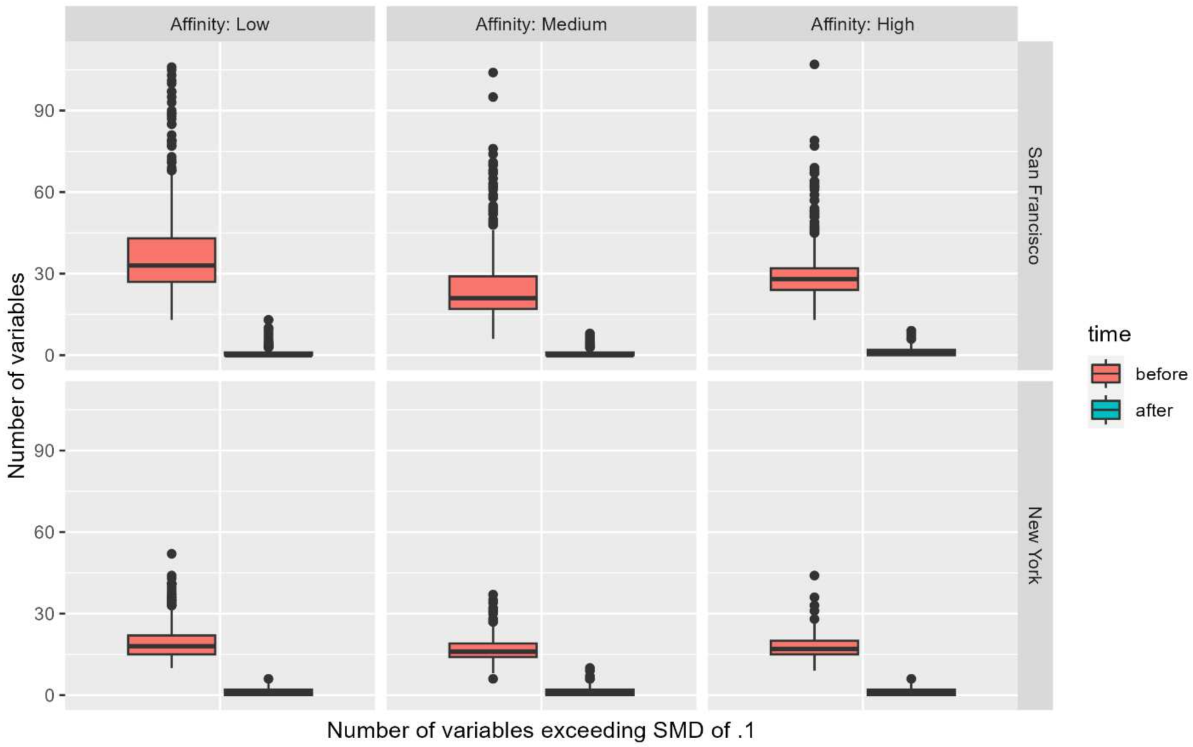

Figure A3. Number of Imbalanced Variables by Threshold, Level of Affinity, and City

Appendix C. LDA And Labeled LDA

Topic models are algorithms utilized to identify the primary themes within a large and unstructured collection of documents (Blei 2012; Blei, Ng, and Jordan 2003). While typically applied to collections of textual documents, Latent Dirichlet Allocation (LDA) can be used to uncover themes within social networks (Blei 2012; Blei et al. 2003). Labeled LDA has similar assumptions as LDA but “incorporates supervision by simply constraining the topic model to use only those topics that correspond to a document’s (observed) label set” (Ramage et al. 2009:249).

Labeled LDA includes the observed label set, Λ, which is a list indicating the presence and absence of the topic in a document. Consequently, the topic proportion of a document is constraint rather than inferred, ensuring that the topics assignment generated by the algorithm matches with the topic label of the document. Given that the group topics in our data are known rather than inferred, Labeled LDA provides a more accurate representation of the data structure compared to LDA.

Reference:

Blei, David M. 2012. "Probabilistic Topic Models." *Communications of the ACM* 55(4):77–84. doi: 10.1145/2133806.2133826.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3(Jan):993–1022.

Ramage, Daniel, David Hall, Ramesh Nallapati, and Christopher D. Manning. 2009. "Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-Labeled Corpora." P. 248 in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing Volume 1 - EMNLP '09*. Vol. 1. Singapore: Association for Computational Linguistics.