Quantitative Marketing and Economics (2020) 18: 381–417
https://doi.org/10.1007/s11129-020-09227-6

Social network design for inducing effort

Pinar Yildirim[1] · **Yanhao Wei**[2] · **Christophe Van den Bulte**[1] · **Joy Lu**[3]

Received: 26 June 2019 / Accepted: 15 April 2020 / Published online: 10 July 2020

Abstract

Many companies create and manage communities where consumers observe and exchange information about the effort exerted by other consumers. Such communities are especially popular in the areas of fitness, education, dieting, and financial savings. We study how to optimally structure such consumer communities when the objective is to maximize the total or average amount of effort expended. Using network modeling and assuming peer influence through conformity, we find that the optimal community design consists of a set of disconnected or very loosely connected sub-communities, each of which is very densely connected within. Also, each sub-community in the optimal design consists of consumers selected such that their "standalone" propensity to exert effort correlates negatively with their propensity to conform and correlates positively with their propensity to influence others.

Keywords Customer community design · Network design · Network optimization · Peer influence

JEL Classification C7 · D85 · D91 · Z13 · M31

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11129-020-09227-6) contains supplementary material, which is available to authorized users.

Yildirim: Assistant Professor of Marketing, The Wharton School, University of Pennsylvania, email: pyild@wharton.upenn.edu. Wei: Assistant Professor of Marketing, Marshall School of Business, University of Southern California. Van den Bulte: Gayfryd Steinberg Professor and Professor of Marketing, The Wharton School, University of Pennsylvania. Lu: Assistant Professor of Marketing, Tepper School of Business, Carnegie Mellon University. All correspondence about the manuscript can be directed to the first author. We thank Anthony Dukes, Upender Subramanian, Olivier Toubia, and the reviewer team for comments on an earlier draft.

✉ Pinar Yildirim
pyild@wharton.upenn.edu

[1] The Wharton School, University of Pennsylvania, Philadelphia, PA, USA

[2] Marshall School of Business, University of Southern California, Los Angeles, CA, USA

[3] Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, USA

1 Introduction

Over the last two decades, marketers have become increasingly keen on leveraging peer influence among customers. Numerous firms and platforms connect consumers who do not know each other into online communities and allow them to observe each others' activity to increase the level or consistency of the effort they exert to achieve a goal. Members of the ALS community on PatientsLikeMe share data allowing them to see what actions are taken by other patients and whether their progress is fast, slow, or about average. Opower, now part of Oracle, works with many utilities to share energy consumption information from similar households in one' s vicinity to reduce energy consumption. Online platforms like Crowdrise and Razoo allow users to observe each others' charitable donations.

In all these examples, a manager is interested in leveraging norm-based peer influence to increase the overall or average effort exerted towards a goal. As the examples demonstrate, in areas ranging from healthcare to energy consumption and re-use, many companies bring consumers together and create communities with the hope that the information exchange among them will improve their overall level of activity, motivate them to use products and services in more and better ways, and share information about their experiences in general. On platforms where individual users are strangers, marketers are increasingly involved in connecting users.

This notable change in marketers' mindset is due to the renewed appreciation that peer influence can be leveraged among people who hardly know each other (Uetake and Yang 2020). In the absence of a clear benchmark about what is appropriate or desirable, people tend to conform to what others do, even when those others are strangers. This insight goes back to famous experiments showing that whether people put into a group with strangers stated that a light dot was moving or not (Sherif 1935) or that one line was longer than another (Asch 1956) was influenced by what those strangers stated. More recently, field experiments have shown that people are more likely to re-use towels in a hotel, to make charitable donations, to consume less energy, to consume less water, to exercise, to pay overdue taxes, and to rate movies when they see or are told that others do so, even when they have no prior connection to these others (e.g., Frey and Meier 2004; oldstein et al. 2008; Martin and Randal 2008; Shang and Croson 2009; Chen et al. 2010; Allcott 2011; Ferraro and Price 2013; Ayres et al. 2013; Allcott and Rogers 2014; Hallsworth et al. 2017). Along similar lines, a recent field study documented that physicians were more likely to keep prescribing a new drug if their colleagues in the same practice or hospital did so, even if those were not the peers they turned to for discussing treatment options or referring patients (Iyengar et al. 2015).

Online settings offer an important advantage for social network design: Information about people's attributes can be restricted to those relevant for the specific purpose of the community (Centola 2018). In several online communities geared towards peer support, the social designer allows members to form their own connections but restricts the information about other members to only purpose-pertinent traits while omitting characteristics like age, gender or race. PatientsLikeMe is an example. Doing so can be useful for increasing the salience of peers most likely to induce the desired behavior. Many companies and organizations focusing specifically on leveraging social conformity rather than peer support, like Opower, go one step further. They even take the decision of whom to connect to entirely out of the members' hands and connect people with what is hoped to be the socially optimal set of peers. In communities designed and joined to achieve a specific purpose, that purpose becomes the focus for identity and other bases of tie formation lose salience. Consequently, members of highly focused and purpose-oriented communities are remarkably likely to accept an exogenously imposed network structure, unlike what we observe in purely affectively oriented friendship networks or in general-purpose platforms like Facebook (e.g., Centola and van de Rijt 2015).

In this paper, we study the optimal design of communities with the goal of improving overall community activity when consumers' effort or activity level can be manipulated through peer influence. Specifically, we address the following research questions: How should a manager or social planner design a community to maximize its total effort by connecting individuals to others to act as reference groups? Which consumer's information should be shared with which other consumer? Should a manager choose multiple small communities, or a single large community? More specifically, is it more effective to have all community members exchange information with every other member, to organize members in distinct sub-communities, or to have star-like structures where some members are more central than others? Should some consumers effort or activity be made more prominent than that of others? The answers to these questions are not trivial. Indeed, the need to study the optimal composition of individuals in such communities analytically have been highlighted by Uetake and Yang (2020).

Our analysis focuses on a community design problem with three properties. First, the community manager ("designer") seeks to maximize the average (or total) level of effort or activity that community members exert towards a goal, like losing weight or consuming less electricity. Second, the designer has the opportunity to set a reference group for each consumer by forming ties between consumers (e.g., by forming virtual or physical groups or by managing the person-to-person flow of information). Third, the decision to exert effort is subject to normative peer influence, such that people incur a disutility when they deviate from the descriptive norm set by the community members they are connected to.

The first two properties are present in the examples we provided, and need no further discussion. The third property, though quite familiar to behavioral researchers, may require elaboration. Social conformity to the descriptive norm set by peers' behavior occurs quite frequently when people are not sure about what the adequate or appropriate behavior is or when people do not want to deviate from their group's norm by over- or under-achieving because they may be embarrassed or they may be subject to criticism. People consorm even when deviating from the norm is not subject to punitive action or when status concerns and other signaling considerations are ruled out (e.g., Asch 1956; Cialdini and Goldstein 2004; Cialdini et al. 1990; Festinger 1954; Kelman 1958; Zafar 2011). As Bernheim and Exley (2015) note, conformity in such situations arises through a preference mechanism rather than a belief mechanism. Consequently, the consumer model in our analysis of optimal reference group design is set up differently from the model by Bernheim (1994) emphasizing belief rather than preference mechanisms of conformity.

Identifying the optimal design for a community is not trivial even under a fairly straightforward social influence relationship such as conformity. How should a community's social network be designed such that the overall activity of the community be maximized (or, in some cases, minimized)? It may intuitively seem like the community should be organized around the most influential and active individual, in a star-like structure. A closer examination, however, suggests that the immediate intuitive design solutions are not necessarily optimal. Moreover, a secondary intuition may suggest that individuals should be connected based on activity level, pairing high and low activity individuals directly to set the high activity individual as the reference. However, the optimal network configuration considers not only the direct impact of social influence when two individuals are connected, but also the indirect impact through the rest of the community.

Our analysis shows that even with the inclusion of the indirect impact, the optimal reference group design can be characterized and the following key results can be derived. (1) The optimal community consists of a set of disconnected or very loosely connected sub-communities, each of which is very densely connected within. (2) In this design, each sub-community consists of members selected such that their "standalone" propensity to exert effort correlates negatively with their propensity to conform and correlates positively with their propensity to influence others. In other words, the best selective matching is achieved within sub-communities between which few ties exist. Also, because every connection is a conduit of influence, conformity is leveraged to the maximum within sub-communities that are internally densely connected. Finally, the improvement from optimally linking community members increases with the size of the community and with the heterogeneity in members' (i) propensity to exert effort in isolation, (ii) social influence, and (iii) susceptibility to social influence. Intuitively, greater community size or heterogeneity offers greater degrees of freedom to the community designer in assigning specific people to sub-communities to achieve optimal matching.

These insights are derived from our benchmark model where peer influence operates through conformity. In the extensions, we also consider the case where individuals choose their behavior to be better than or different from others. When individuals seek to out-compete their peers rather than to conform, the optimal network structure for inducing maximum total or average effort is the completely connected network where every member is connected to every other member.

Putting people into groups where one is exposed to information from others can be done conveniently in online as well as offline networks. As to the propensities to exert effort, to be influential, and to be susceptible to influence, these can be assessed directly using questionnaires administered when people sign up to become members of the network (e.g., Flynn et al. 1996; Duckworth et al. 2007) or indirectly using easy-to-observe correlates of these propensities (e.g., Aral and Walker 2012).

Our findings are relevant for communities geared towards self-improvement, including branded online communities. From Lego to Harley Davidson to Sephora, many brands today are building the infrastructure to leverage the social connections among customers, converting them into sales dollars and affinity for the brand. In some cases like Nike's online community, consumers are encouraged to connect to others while using the brand to work to achieve a goal, such as keeping active, eating healthier, keeping attention at work, etc. We investigate how the community manager would like to design (or at least nudge) who gets connected with whom in such goal-oriented online brand communities. Like Bagozzi and Dholakia (2002) and Dholakia et al. (2004), we focus on social normative influence as the key mechanism through which peer influence operates in such online communities, but the behavior we focus on is effort exertion rather than participation and involvement in the community.

The rest of the paper is organized as follows. Section 2 provides a summary of the relevant literature and this paper's contributions to it. Section 3 presents the model setup. Section 4 characterizes the optimal community design. Section 5 describes under what circumstances managers in charge of designing communities can benefit the most from this exercise. Section 6 presents extensions and factors that influence the key results. Section 7 concludes with implications for practice and research.

2 Related Literature in Marketing and Network Science

2.1 Network Optimization

Network optimization has a long history in applied mathematics, graph theory and operations research, where the focus has been on transportation and distribution networks, and the key trade-off has been between travel time and cost (e.g., Newman 2010, pp. 541-548). Another area of application has been the optimal structure of electrical power grids, where the stability of the grid has been the key consideration (e.g., Lewis 2011, pp. 331-335). Vulnerability of networked physical systems to internal node or link failure or to outside attack resulting in node or link removal has also attracted quite a bit of attention in manufacturing, logistics, warfare, and counter-terrorism (e.g., Lewis 2011, pp. 375-430). Of course, how to optimally design networks taking into consideration speed vs. cost, stability, vulnerability, and the ability to fight off or recover from outside attack has also been of great interest in computer science (e.g., Goldenberg et al. 2005). Obviously, our work differs from this prior work in the planner' s objective and trade-offs. More fundamentally, it differs in that the nodes are sentient beings who act to achieve their own objectives and are subject to conformity. A partial exception is a paper by Cerdeiro et al. (2017) who consider a setting where the designer chooses a computer network architecture, nodes individually choose their security level, and a strategic attacker targets one node in order to minimize the connectivity of the structure. They show that decentralized security choices can lead to both over- and under-investment in security, and that network design can be a powerful tool to mitigate such protection inefficiencies. Even though there are similarities in setting (a central planner chooses structure whereas nodes choose action), key trade-off ("contagion" among nodes creating a trade-off by affecting low and high effort nodes in opposite directions) and main insight (optimal structure improves social welfare), their work does not consider how preference heterogeneity or heterogeneity in social conformity affects the optimal design. Consequently, the analyses identify different optimal designs.

Over the last twenty years, the endogenous choice of both actions and links has attracted considerable attention among game theorists (for a review, see Vega-Redondo 2016). Our work also considers both dimensions of choice, but differs in a key respect. Previous analyses of optimal choice in action and link formation considered settings where both decisions are made by network members. Our work, in contrast, assumes that network members choose the action but the network manager chooses the network structure. For instance, a recent paper by Goyal et al. (2017) shares several features with our analysis. In both analyses, (i) both actions and links are chosen to maximize an outcome, (ii) network members experience peer influence leading them to prefer to conform to the behavior of their contacts, and (iii) differ on their preferred action. However, whereas they study a setting where individual utility-maximizing network members choose both their own actions and their own contacts, we consider a setting where a community manager sets the network structure to maximize social welfare (operationalized as total utility).[1] This distinction matters greatly. Goyal et al. (2017) find that, if left to their own devices, network members will tend to end up in a network structure of densely clustered subgraphs where they connect only with people like themselves. That is, preference heterogeneity combined with social conformity results in a highly Balkanized world.[2] Our analysis shows that a social planner can achieve higher social welfare by designing a network of densely clustered subgraphs consisting of members who are heterogeneous along both preference and influence/susceptibility dimensions.

2.2 Social Influence and Community Formation

Marketing practitioners and academics have long been keen on understanding how to best leverage peer influence in social groups. Research attention at first focused on documenting peer influence in field settings (e.g., Van den Bulte and Lilien 2001; Aral et al. 2009). More recent research has focused on identifying the mechanisms at work (e.g., Chen et al. 2011; Iyengar et al. 2015), characterizing which customers, if any, exert a disproportional amount of influence on others (e.g., Aral and Walker 2012; Goldenberg et al. 2009; Iyengar et al. 2011; Katona et al. 2011), and leveraging either homophily or peer influence through efficiently targeted marketing effort (e.g., Hill et al. 2006; Godes and Mayzlin 2009; Goel and Goldstein 2013; Bimpikis et al. 2016; Ascarza et al. 2017). This prior work relies on an existing community structure, be it static or evolving, as a given for the marketer. Most recently, research has started to view communities and social networks among individuals as an outcome or even a choice variable. Some of this work investigates how marketing actions might alter the network to the marketers' benefit (Ansari et al. 2018) and how network structures result endogenously from members' choices affected by a marketer's or platform manager's policies (Hartmann 2010; Iyer and Katona 2015; Phan and Godes 2018; Wei et al. 2016). Other work compares the effectiveness of various exogenously given

[1] Another difference is that in Goyal et al. action is binary, whereas in the present study it is continuous.
[2] Similar results are obtained by Bojanowski and Buskens (2011), Neary (2012), Advani and Reich (2015), and Ellwardt et al. (2016).

network structures for a particular marketing objective (Murtha et al. 2014; Peres and Van den Bulte 2014).

The present study raises the hitherto unaddressed question of optimal community design. More specifically, we investigate the following question: How can a manager or central planner design a community such that the overall effort expended by members is maximized? Unlike prior work involving comparative assessments of multiple network structures, we seek to directly characterize the optimal structure. Unlike prior work where the network is exogenously given, we endogenize network structure as a decision made by a central planner or community manager. Unlike prior work where marketers aim to improve their marketing ROI by targeting specific nodes within a fixed network and assume such targeting can generate peer influence, we allow peer influence to organically develop as an outcome of the endogenous grouping decision. We take the view of a marketer or community manager intervening by choosing the network structure, i.e., choosing the pattern of ties connecting the given set of nodes.

Researchers, managers, and policymakers are becoming increasingly interested in using interventions that manipulate individual behavior through social influence (Carrell et al. 2013; Cerdeiro et al. 2017; Haag and Lagunoff 2006; Kraut et al. 2012; Valente 2012). One recent example is the formation of communities where members can determine their individual behavior by using others' behaviors as a reference point. Our study is part of this emerging research on effective community design but differs in the intervention and objectives being analyzed.[3] Prior studies focused on seeding ideas or a marketing promotion within an existing social network to achieve virality. Such studies concluded that targeting particularly influential or influenceable individuals is useful to attain the virality goals. These studies, however, do not address the question about how communities should be designed, if possible, as such optimal designs can improve the gains above and beyond the improvement that comes from optimal seeding.

A key insight from prior research, which we accept as a premise to our model, is that endogenous tie formation by people subject to social conformity does not render community design ineffective. Allcott and Kessler (2015) show that a majority of customers receiving Opower social comparison reports for free from their electric utility are willing to pay for such information without being able to select the people they receive benchmark information about. Five field experimental studies in the realm of health behavior and physical fitness by Centola (2010), Centola (2011), Zhang et al. (2015a), Rovniak et al. (2016), Zhang et al. (2016) similarly document people's willingness to relinquish their agency in network formation and allow them-

[3]Community design is also an important area of study in organizational behavior and management literatures. There are, however, key differences with our work. First and most importantly, while these studies investigate links between network structure and outcomes, they do not attempt to optimize structure. Second, social conformity is not the root motivation for employee behavior. Financial incentives influence employee behavior. Finally, in optimization of collective employee effort, there is a concern for free-riding because firm objectives can be reached even when an individual is not exerting effort. In the context that we study, individuals' goals are not tied to an overall objective and they do not gain utility from others' progress unless they also exert effort.

selves to be placed in a network structure designed by the manager of a community they join. Another example is that every year, millions of freshmen students across the globe willingly agree to be assigned—often in explicitly random fashion—to dormitory rooms, student cohorts, study groups and course project teams by the educational institutions they join in their pursuit of an education. In short, community design is accepted by many people joining communities built and managed to help them reach a specific objective. Furthermore, there is ample evidence that exogenous assignment to particular apartments, offices, or dorm rooms has long-run effects on both the subsequent pattern of organically developed social ties (e.g., Festinger et al. 1950; Van den Bulte and Moenaert 1998; Hasan and Bagde 2013) and the level of achievement many months after the initial assignment (e.g., Hasan and Bagde 2013; Sacerdote 2001). In short, evidence shows that endogenous tie formation does not render community design ineffective.

That community or network design affects interaction and outcomes even after members form social ties organically is consistent with a core proposition of Festinger's (1954) theory of conformity through social comparison: People dislike being in a state of discordance with peers, and often resolve that state of cognitive dissonance by changing their attitude and behavior rather than severing the social ties with their discordant peers. It is this very tendency that underlies social conformity through descriptive norms, documented amply in both lab and field experiments. The same behavioral principle of discordance aversion explains the pervasive tendency for organic networks to exhibit tie transitivity, balance, and clustering (e.g., Cartwright and Harary 1956; Davis 1970).

The pattern of network formation among the 432 members of an online health-buddy community run by the Medical Office at MIT is quite informative (Centola and van de Rijt 2015). After joining, members were provided with a home page containing "a detailed dashboard with a real-time chart that kept track of their daily activities, exercise intensity levels, and exercise minutes." Members were assigned to one of six separate sub-communities of 72 people each, and were randomly assigned six anonymous health buddies from their own sub-community, of whom they could see the complete profile information (age, gender, race, BMI, exercise interests, dietary preferences, and overall level of fitness) as well the exercise records. Of particular interest is that, over the course of five weeks, program participants were allowed to add and drop buddies, i.e., to rewire their ego-network of buddies. Participants were not able to prevent anyone from linking to them. This set-up provided six independent observations of network evolution.

Strikingly, over the 5 weeks in which 432 member were tracked, only 51 tie changes were initiated. The propensity to change ties was not associated with activity level within the program (Centola and van de Rijt 2015, pg. 23). Community members quite accepted being placed in pre-formed networks. This shows that, once one has joined a community focused on behavior change, members can be much more accepting of their randomly allocated ties than theoretical models of highly strategic endogenous network formation imply. This is also supported by the experience of a similar health buddy program at the University of Pennsylvania, PennShape, where people were not allowed to change their network at all. In each of the four imposed network structures each comprising 186 members (for a total of 744 participants), the retention rate was above 95% by the end of eleven weeks (Centola 2018, pg. 165).

The pattern in the changes that did occur in the MIT experiment is also striking: "all six networks exhibited the same general pattern of social selection. In every case, health buddies were selected predominantly on the basis of similarities in … just three traits: age, gender, and BMI" (Centola 2018, pg. 150). Characteristics that were also observable and more directly associated with health and fitness, including "favorite exercise, dietary preferences, exercise intensity, typical exercise minutes, and exercise experience, were not significant factors for selection" (Centola 2018, pg. 150). That homophily was only muted in this online, purpose-oriented community need not be surprising. The more narrowly specified an explicitly designed online community is, " the less relevant every other feature of a person becomes until only a few features, or perhaps even one feature, defines who she is, how she behaves, and why she seeks social interaction" and with whom she interacts (Centola 2018, pg. 156). More surprising is that, given the members' goal of engaging in more rigorous exercise routines than they would otherwise follow, traits directly associated with health and fitness were not among the bases of homophily. Instead, community members, to the extent that they rewired their network, sought out demographic reference points rather than performance-based reference points (Centola and van de Rijt 2015, pg. 25). This shows that, once one has joined a community focused on health behavior change, peers' behavior or diligence need not be as strongly predictive of tie formation as theoretical models of strategic network formation imply (Centola and van de Rijt 2015, pg. 26). Along similar lines, studying the evolution of an organic email network of university students and faculty, Kossinets and Watts (2006) found that having a common interest and purpose (taking a common class) was strongly associated with tie formation, that homophily in status, gender and age was absent, and that people did "not strategically manipulate their networks" (pg. 90).

3 Model Setup and Equilibrium

We consider a set of n individuals or consumers who decide how much effort to exert in order to reach an objective they hold. A manager whom we will refer to as the "designer" wants to organize these individuals into a community and facilitate information exchange among them in order to maximize the total or, equivalently, the average amount of effort exerted. Since communities are networks of individuals, we will treat the problem as a network design problem and use terminology from network science. Any two individuals may be connected to each other, and if i and j are connected, we denote the tie by $i \sim j$. A network, denoted generically by G, is simply the collection of the individuals and their undirected ties. We start by characterizing the "out-of-community" effort in the absence of any reference groups, and then move to establishing the equilibrium effort when ties are present between community members. [4] Let the consumer's out-of-community utility from exerting effort at level y_i be:

$$u_i = \alpha_i y_i - \frac{1}{2} y_i^2 \tag{1}$$

where α_i is the idiosyncratic reward from working towards an objective, and $\frac{1}{2} y_i^2$ is the quadratic cost of effort. Consumers are heterogeneous with respect to how much they care about the objective. If the objective is preserving energy, for instance, consumers vary in how much they value the environmental and financial benefits of consuming less energy. We assume $\alpha_i \in [0, 1]$ but the results easily scale, mutatis mutandis, to larger bounded intervals for α_i.

A consumer exerting effort when there is a reference group gains utility not only from the benefits of working towards her objective, but also from the conforming to the activity of others in the community. We represent the in-community utility as:

$$u_i = \alpha_i y_i - \frac{1}{2} y_i^2 + \frac{1}{2} \sum_{j \sim i} [1 - \rho_{ij}(y_i - y_j)^2] \tag{2}$$

In the expression, $j \sim i$ denotes any individual j that is connected to i and vice versa, and $\rho_{ij} > 0$ calibrates the social influence of j on i and is common knowledge. The last term $\sum_{j \sim i} [1 - \rho_{ij}(y_i - y_j)^2]$ captures the utility from adhering to the descriptive norm, and is the component which represents peer influence. To make the problem as realistic as possible, we introduce two characteristics for social influence. First, it can be asymmetric between the individuals, as is often the case in real life. The influential bloggers are not necessarily influenced by the individuals who follow them. Formally, we allow the influence between any two connected consumers to be asymmetric ($\rho_{ij} \neq \rho_{ji}$) and also allow the activity of i to be greatly influenced by j's activity ($\rho_{ij} \gg 0$) or at a marginal rate ($\rho_{ij} \to 0$). ρ_{ij} need not equal ρ_{ik}, $k \neq j$, such that i may be influenced by different individuals in his network at different rates. Second, we do not assume that the degree of social influence is a function of one individual's personality traits but ρ_{ij} jointly reflects both the ability of j to influence others, and the susceptibility of i to be influenced by others:

$$\rho_{ij} = \lambda_i \cdot \chi_j, \tag{3}$$

where $\lambda_i > 0$ calibrates how sensitive i is to observed peer behavior and $\chi_j > 0$ calibrates how much influence j has over others. In reality, it would be unrealistic to assume that one individual has the same level of influence over every other individual. It would be equally unrealistic to assume that an individual is influenced by every other person to the same degree. This flexible modeling approach is important. While these individual-level influence and susceptibility parameters significantly increase the complexity in our derivations, they are central to the insights about characterization of the optimal network design. Note, ρ_{ij} and α_i are calibrated against the marginal cost of effort, which is assumed to take a constant value of 1. Also notice that the heterogeneity among individuals allows designers to take into account all types of individuals ranging in effort and social influence to design a community.

[4] We use the terms effort and activity interchangeably.

This framework captures an individual' s desire to exert effort and to conform to the behaviors of others they are connected to. Specifically, conformity enters the utility function directly through the quadratic term, consistent with how Bernheim and Exley (2015) model descriptive normative influence as a preference mechanism. In Section 6.2, we investigate the optimal network design for settings in which individuals want to out-compete their connected peers by exerting more effort than them, rather than adhere to the descriptive norm set by their peers.

When a customer i exerts effort alone, his utility-maximizing level of activity is α_i. Thus we also refer to α_i as the out-of-community effort level. However, when he is maximizing the utility given in Equation (2), the chosen effort level is a weighted average of α_i and the activity of his contacts:

$$y_i^* = \frac{\alpha_i + \sum_{j\sim i} \rho_{ij} y_j}{1 + \sum_{j\sim i} \rho_{ij}} \tag{4}$$

The optimal individual effort in (4) is not specific to the utility function formulation in (2). Other formulations, for instance $u_i = \alpha_i y_i - \frac{1}{2} y_i^2 - \frac{1}{2} \sum_{j\sim i} \rho_{ij}(y_i - y_j)^2$, also lead to the expression in (4). All of our subsequent analyses are based on (4) and hold with any utility specification that gives (4). The utility specification in (2) has the benefit that the individual is never better off outside the community such that participation is guaranteed, if $\rho_{ij} \in [0, 1]$ for all i and j.

We consider a Nash equilibrium where all individuals choose their effort simultaneously. In what follows, y_i^* denotes the equilibrium effort of individual i and vector $\mathbf{y}^*$ collects the equilibrium effort levels of all the community members. Proposition 1 establishes the existence and uniqueness of the equilibrium.[5] All proofs are in the Online Appendix.

Proposition 1 *Given a set of out-of-network activity levels* $\boldsymbol{\alpha}$*, influence levels* $\boldsymbol{\rho} > 0$*, and a fixed community network* G*, there exists a unique equilibrium of individual activities,* $\mathbf{y}^*$*. Customer* i*'s equilibrium activity satisfies:*

$$y_i^* = \alpha_i + \sum_{j\sim i} \rho_{ij}(y_j^* - y_i^*) \tag{5}$$

The equilibrium is given by

$$\mathbf{y}^* = \mathbf{L}^{-1}\boldsymbol{\alpha} \tag{6}$$

[5] In a "larger" model, one may want to consider how individuals react to the choices of the community network designer, taking into account that the latter has connected him or her to a non-representative group of individuals. We abstract away from such considerations. In a model where individuals are fully aware of the non-representativeness of their ties and prefer to conform to the population average rather than their distinct peers, $\{\alpha_i, \lambda_i, \chi_i\}_{i=1}^n$ are common knowledge, and individuals act fully rationally accordingly, each individual should be able to infer and act towards an unbiased estimate of the population average effort. This seems to render the problem of network design much less relevant. Hence, the effectiveness of community network design relies on the premise that individuals will conform to their ties, even when they might be aware that these ties are non-representative.

where $\mathbf{L}$ *is a square "influence matrix" of size n, given by*

$$L_{ij} = \begin{cases} -\rho_{ij}, & \textit{if } i \sim j \\ 0, & \textit{otherwise} \end{cases}$$

$$L_{ii} = 1 - \sum_{j \sim i} L_{ij}.$$

Proposition 1 demonstrates (i) how an individual's in-community activity is linked to not only his own out-of-community activity, but also the activity of the other individuals, and (ii) how social influence is determined by both ρ_{ij} and the structure of the community. Equations 5 and Eq. 6 imply that an individual' s equilibrium in-community effort can sometimes be lower than his out-of-community effort, that the same is true for the overall level of effort, and that any connection can influence anyone' s equilibrium level of effort. Consequently, to increase the total collective effort exerted by customers or community members, managers may need to purposely design the community structure.

4 Optimal Community Design

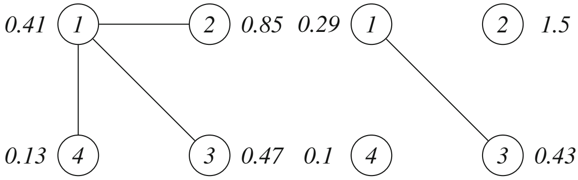
Given the utility formulation in (2), the total effort $T \equiv \sum_{i \in G} y_i^*$ can be maximized by determining which consumers' information should be shared with which other consumer, and the designer has to answer this question simultaneously for all community members. Suppose we consider only two individuals with $\alpha_1 > \alpha_2$. It is easy to see that placing a tie between them increases their joint activity iff consumer 1 influences consumer 2 more than consumer 2 influences consumer 1 (that is, iff $\rho_{21} > \rho_{12}$). When applying such a pairwise comparison rule for $n > 2$, any two individuals would be connected iff $(\rho_{ij} - \rho_{ji})(\alpha_j - \alpha_i) > 0$. However, the following example shows that this pairwise design rule could lead to worse outcomes than the empty network (i.e., where everyone is isolated), let alone a comprehensive design rule that considers not only the direct but also the indirect influence of consumers on each other.

Consider a 4-person community. The individuals are characterized by their individual parameters as follows:

$$\alpha = \begin{pmatrix} 0.0 \\ 1.5 \\ 0.5 \\ 0.1 \end{pmatrix}, \lambda = \begin{pmatrix} 2.0 \\ 1.5 \\ 0.5 \\ 0.1 \end{pmatrix}, \chi = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

The community design produced by the pairwise rule is given in the community below on the left. All consumers are placed in a star structure where consumer 1 is connected to every other consumer. The equilibrium activity of each consumer is written next to each node. Applying the pairwise design rule results in a network structure where the total effort exerted T is 1.86, an 11% decrease compared to the empty network where each individual is disconnected from everyone else (in which case $T = \sum_i \alpha_i = 2.1$).

Now consider the community below on the right where the total activity is 2.32, a 10% increase compared with the empty network. In this community, the only tie is between consumers 1 and 3. Connecting these two results in a net increase in their effort from 0.50 to 0.71. The community on the right is actually the optimal design as no other community configuration results in higher total activity.



This example illustrates that designing a community based on direct pairwise influence can be not only suboptimal but even detrimental to population-level effort. Optimal community design requires going beyond the immediate returns to adding a link between two individuals and taking into account the spillovers to other connected nodes.

Designing a community without any simplifications implies searching for the best *network*, and this is by no means a trivial task. For n consumers to select from, there are $2^{\frac{n(n-1)}{2}}$ possible network configurations, and we wish to find the network that leads to the highest collective effort.[6] Moreover, the problem is nonlinear, as evidenced by (6). To tackle this difficult problem, we temporarily relax the discreteness of tie creation. Specifically, we allow the firm to not only create a connection between two consumers but also do so at varying levels of tie strength. For instance, a social community may change the frequency at which consumers hear from each other. Although this relaxation enlarges the set of possible configurations, it facilitates the mathematical analysis. Let $s_{ij} = s_{ji} \in [0, 1]$ be the strength of the link between i and j. When $s_{ij} = 0$, the two consumers do not hear from each other, or, equivalently, there is no connection between them. The equilibrium activity of a consumer then becomes a variant of that in Equation (4), weighted for tie strength:

$$y_i^* = \frac{\alpha_i + \sum_{j=1}^{n} s_{ij}\rho_{ij} y_j}{1 + \sum_{j=1}^{n} s_{ij}\rho_{ij}} \tag{7}$$

The Nash result in Equation (6) can be similarly expressed as $\mathbf{L}^{-1}\boldsymbol{\alpha}$ and using the following influence matrix $\mathbf{L}$:

$$\begin{aligned} L_{ij} &= -\rho_{ij} s_{ij}, \\ L_{ii} &= 1 - \sum_{j \sim i} L_{ij} \end{aligned}$$

[6]This number is 64 for $n = 4$, larger than 32,000 for $n = 6$, and is considerably larger for any greater sized population (e.g., $n > 100$).

4.1 What is the Optimal Tie Strength for Inducing Effort?

We now assess how changing the strength of the tie between two consumers affects the overall effort level in the community.

Lemma 1 *The marginal effect of the strength of the tie between individuals i and j on the activity of each community member, $\partial \mathbf{y}^*/\partial s_{ij}$, is given by:*

$$\frac{\partial \mathbf{y}^*}{\partial s_{ij}} = \mathbf{L}^{-1}\mathbf{v}_{ij}^*, \tag{8}$$

where the vector $\mathbf{v}_{ij}^*$ is defined as below using $\mathbf{e}_i$ to denote the i-th basis vector and $\mathbf{e}_j$ to denote the j-th basis vector (i.e., $\mathbf{e}_i$ is all zeros except for the i-th entry):

$$\mathbf{v}_{ij}^* \equiv \rho_{ij}(y_j^* - y_i^*)\mathbf{e}_i + \rho_{ji}(y_i^* - y_j^*)\mathbf{e}_j.$$

The change in the overall community activity in response to a change in a specific tie's strength equals $\partial T/\partial s_{ij} = \sum_{k=1}^{n} \partial y_k^*/\partial s_{ij}$. Lemma 1 says that the marginal effect of increasing s_{ij} on the collective effort T is positive iff

$$\partial T/\partial s_{ij} = \mathbf{1}'\mathbf{L}^{-1}\mathbf{v}_{ij}^* > 0 \tag{9}$$

Linking two consumers exhibiting conformity raises the effort of the lower-activity consumer, and lowers that of the higher-activity consumer. Optimal community design consists of finding the best trade-off between these two opposing effects, operating not just directly on the two focal individuals who are being connected but also indirectly on others in the community who in turn affect the focal two. The totality of these effects is taken into account and summarized in Equation (9).

Lemma 2 *Hold constant all the tie weights in a community except for s_{ij}. If $\partial T/\partial s_{ij} > 0$ at any particular value of s_{ij}, then $\partial T/\partial s_{ij} > 0$ at all values of s_{ij}.*

If the marginal effect of s_{ij} on T is positive at some point, it will remain positive as one increases (or decreases) s_{ij}. This remarkable result implies that creating medium-strength ties is generally not optimal. If strengthening the connection between two consumers in a community increases the total activity (T) in the community, then the tie should be pushed to full strength $s_{ij} = 1$. Conversely, if weakening a tie between two individuals increases the activity T, the link should be removed completely. Put differently, the optimal tie strength is " bang-bang" and to maximize total activity, the community designer must add ties at the extensive rather than intensive margin: it is the presence or absence of a tie rather than its strength which helps to maximize total activity.

4.2 What is the Optimal Community Structure for Inducing the Maximum Effort?

We now turn to the key question: How to structure the community of n individuals to maximize the total or average level of effort exerted? We first introduce a result for the most dense community networks, i.e., complete networks where every two individuals are connected.

Lemma 3 *(i) For a community with a complete network of individuals characterized by* $\{\alpha_1, \lambda_1\}_{i=1}^n$ *and* $\chi_i = 1$*, the total effort is given by:*

$$T = \frac{\sum_{i=1}^n \frac{\alpha_i}{1+n\lambda_i}}{\frac{1}{n}\sum_{i=1}^n \frac{1}{1+n\lambda_i}}. \tag{10}$$

(ii) For a community with a complete network of individuals characterized by $\{\alpha_i, \chi_i\}_{i=1}^n$ *and* $\lambda_i = 1$*, the total effort is given by:*

$$T = \frac{\sum_{i=1}^n \alpha_i\,(1+n\chi_i)}{\frac{1}{n}\sum_{i=1}^n (1+n\chi_i)}. \tag{11}$$

Equations (10) and (11) imply that the average effort level within a complete network, $\frac{T}{n}$, is a weighted average of the stand-alone effort levels (α_i), where individuals who are less influenceable (λ_i) or more influential (χ_i) are given a greater weight.

The last key building block to our main result is stated in Lemmas 4 and 5, characterizing the optimal community structure for the extreme case where the propensity to exert stand-alone effort (α_i) and the propensity to influence (χ_i) or be influenced (λ_i) are perfectly correlated. Note that we will represent results related to parameter χ_i in terms of χ_i^{-1} in Lemma 5 for mathematical tractability.

Lemma 4 *Consider a set of individuals characterized by* $\{\alpha_i, \lambda_i\}_{i=1}^n$ *and* $\chi_i = 1$.

(i) If α_i *and* λ_i *are perfectly correlated, there is a unique locally optimal network which is also the globally optimal network.*

(ii) If α_i *and* λ_i *are perfectly negatively correlated (or positively correlated, respectively), then the complete (empty) network is globally optimal.*

Lemma 5 *Consider a set of individuals characterized by* $\{\alpha_i, \chi_i\}_{i=1}^n$ *and* $\lambda_i = 1$.

(i) If α_i *and* χ_i^{-1} *are perfectly correlated, there is a unique locally optimal network which is also the globally optimal network.*

(ii) If α_i *and* χ_i^{-1} *are perfectly negatively correlated (or positively correlated, respectively), then the complete (empty) network is globally optimal.*

Of course, perfect correlations are unlikely to occur within the customer or membership base of any company or organization. However, community designers may be able to organize the community members into multiple mutually disconnected sub-communities or sub-graphs within which such correlations hold. Part (i) of Lemmas 4 and 5 state that in such sub-communities, the globally optimal and locally optimal community designs are identical. Part (ii) of the lemmas states that these sub-communities should be fully connected within when the perfect correlation between both α and λ and α and χ^{-1} is negative.

Intuition suggests, and Proposition 6 presented later confirms, that members of such sub-communities should have high variance in stand-alone effort, influence, and influenceability. We formalize that insight and combine it with previous results in Propositions 2 and 3, characterizing the optimal community structure in the absence of perfect correlation. Unlike Lemmas 4 and 5 these two propositions pertain to local rather than global optima. It is possible that there are multiple local optima, but numerical analyses presented below show that the characterization also holds when searching numerically for the globally optimal community design.

Proposition 2 ***(Membership to Sub-communities in the Optimal Design)*** *Consider $n \to \infty$ consumers who are distributed according to density $(\alpha_i, \lambda_i) \sim \Phi$ over the support $[\underline{\lambda}, \overline{\lambda}] \times [\underline{\alpha}, \overline{\alpha}]$. Under a mild regularity condition on Φ, it is locally optimal to organize consumers into sub-communities where any two consumers are linked iff they belong to the same sub-community. Each sub-community consists of consumers located along a line segment parallel to a diagonal of the support.*

Proposition 3 ***(Membership to Sub-communities in the Optimal Design)*** *Consider $n \to \infty$ consumers who are distributed according to density $(\alpha_i, \chi_i^{-1}) \sim \Phi$ over the support $[1/\overline{\chi}, 1/\underline{\chi}] \times [\underline{\alpha}, \overline{\alpha}]$. Under a mild regularity condition on Φ, it is locally optimal to organize consumers into sub-communities where any two consumers are linked iff they belong to the same sub-community. Each sub-community consists of consumers located along a line segment parallel to a diagonal of the support.*

Propositions 2 and 3 state that in the optimal design for a very large community, the members will be sorted into a set of mutually disconnected sub-communities (i.e., graph components) where each sub-community is internally fully connected (i.e., the link density is 1 within a sub-community). Put differently, the complete or near-complete networks described in Lemma 4 still arise, but at the level of the smaller sub-communities. This implies that the optimal community network exhibits perfect transitivity (if $i \sim j$ and $j \sim k$, then $i \sim k$) and clustering.

As noted earlier, the key challenge in optimal community design under conformity is to balance the two opposite effects of conformity. The effort of low-α types gets pulled up towards the norm but that of high-α types tends to get pulled down. Proposition 2 states that this trade-off is handled optimally by combining influential high-α types and non-influential low-α types into heterogeneous sub-communities parallel to the main diagonal in Fig. 1. Similarly, Proposition 3 states that this trade-off is handled best by combining non-susceptible high-α types and susceptible low-α types into heterogeneous sub-communities parallel to the main diagonal in Fig. 2.

Figure 3 provides numerical examples to illustrate the results in Propositions 2 and 3. Specifically, for a fixed total number of links, we trace the gain in collective effort as we gradually increase the fraction of within-subcommunity links (i.e., link allocation). The sub-communities are the ones in the optimal network. We carry out this exercise for two levels of the total number of links (i.e., a dense network and a sparse network). Two observations emerge. First, consistent with the Propositions, an increasing fraction of within-relative to across-subcommunity links is associated with higher gains from community design. Second, choosing the right placement of links is more consequential than choosing the number of links, which highlights the importance of the Propositions.

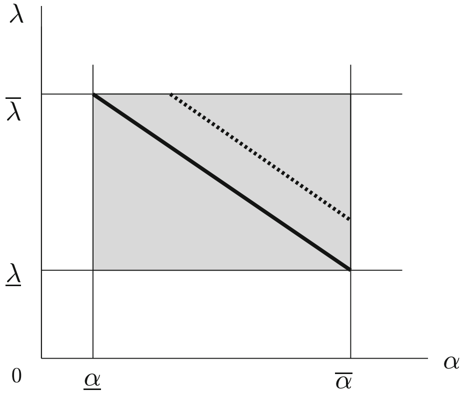


Fig. 1 A Sub-community of Individuals Located Along a Line on the α-λ Plane

Identifying the optimal structure of very large ($n \to \infty$) rather than small communities is the main contribution of our paper. While it may be straightforward for a designer to identify the optimal configuration of a small community through enumeration or experimentation, this is not true for very large communities with thousands of members. [7]

To identify and visually illustrate the typical globally optimal community structure in moderately sized networks, we numerically investigate seven scenarios in which we fix the network size to $n = 200$ and vary the relations between α_i, λ_i, and χ_i^{-1} among the members.

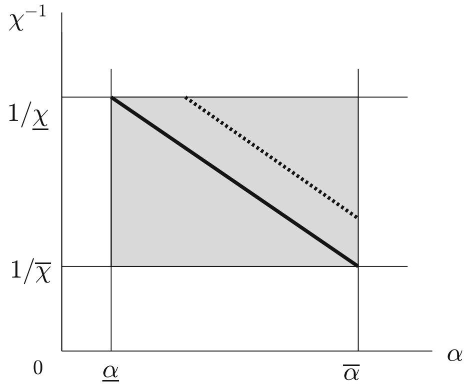
In the first scenario, we fix χ_i (ability to influence) to 1 and let α_i and λ_i vary independently of each other. In the second scenario, we fix λ_i (susceptibility to influence) to 1 and let α_i and χ_i^{-1} vary independently of each other. In the third scenario, we vary all three parameters independently from each other. In the fourth scenario, we again vary all three parameters, but induce a positive correlation between $\chi_i{}^{-1}$ and λ_i. Several theories imply such a negative correlation between influence and susceptibility to influence (e.g., Van den Bulte and Joshi 2007), but empirical evidence typically shows little or only moderate negative correlation (e.g., Aral and Walker 2012; Iyengar et al. 2011). Consequently, we deem the correlation structures of both scenarios 3 and 4 to be fairly realistic. In the fifth scenario, susceptibility to influence (λ_i) exhibits an inverse-U relation with respect to inability to influence (χ_i^{-1}). This is consistent with the theory and evidence of middle-status conformity (e.g., Hu and 604 Van den Bulte 2014; Iyengar et al. 2015; Phillips and Zuckerman 2001).

We operationalize these first five scenarios under the assumption that α_i, λ_i and χ^{-1} are smoothly (continuously) distributed. [8] In the sixth scenario, we fix χ_i (ability to influence) to 1 and let α_i and λ_i vary such that individuals are clumpy in the α_i -

[7] Similarly, theoretical network science studying equilibrium behavior and identifying theoretical properties of graphs and processes evolving on graphs typically does so for $n \to \infty$ (e.g., Acemoglu et al. 2011; Vincent and Ismael 2017; Iijima and Kamada 2017).

[8] Specifically, when α, λ or χ^{-1} are assumed to vary, we draw each from a Beta(2, 2) distribution featuring values bounded between 0 and 1. When either λ or χ^{-1} is assumed not to vary, we set its value to 1. In scenario 4, we induce a positive correlation of 0.5 between λ and χ^{-1} using the method of Plackett (1965) . In scenario 5, we draw α and χ^{-1} each from a Beta(2, 2) distribution, and then set $\lambda_i = 0.6 - 2(\chi_i^{-1} - 0.5)^2$.

Fig. 2 A Sub-community of Individuals Located Along a Line on the α-χ^{-1} Plane



λ_i space. In the seventh scenario, we fix λ_i (susceptibility to influence) to 1 and let α_i and χ_i^{-1} vary such that individuals are clumpy in the α_i - χ_i^{-1} space.

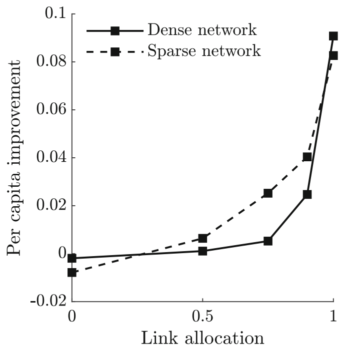
We use simulated annealing to find the globally optimal network structure in each of these scenarios. This numerical optimization method provides excellent approximations to the global optimum in network design problems (McAndrew et al. 2015; Kirkpatrick et al. 1983; Gastner and Newman 2006; Randall et al. 2002).[9] After having found the globally optimal network design using simulated annealing, we identify the various sub-communities in the optimal network using the community detection algorithm developed by Newman (2006).

We report the optimized graph using different colors to indicate different sub-communities. For each of these seven scenarios, we describe the optimal community design by means of three graphs (Fig. 4). The first is a graph showing the nodes and ties, in which the nodes of each sub-community have a different color. This allows us to check whether the optimum indeed consists of sub-communities that are disconnected from each other but densely connected within. The second is a plot of the nodes with α on the horizontal axis and with λ_i, χ^{-1} or $\lambda_i + \chi_i^{-1}$ on the vertical axis. Since the nodes are color-coded by sub-community, this plot shows to what extent each sub-community consists of individuals located on a downward sloping line or band in the parameter space, consistent with Figs. 1 and 2. The third graph is a plot showing how the level of effort exerted by each node in the optimal network, y_i*, varies with α_i, λ_i and χ_i^{-1}. Since the nodes are color-coded by sub-community, this plot also shows how the average level of effort exerted in each sub-community varies.

As the community graphs in the first column of Fig. 4 show, the globally optimal community structure is remarkably similar across these seven scenarios and consistent with the locally optimal community structure characterized by Propositions 2 and 3. In each scenario, the optimal community consists of a small number of

[9]The simulated annealing algorithm is implemented closely following Gastner and Newman (2006). In each step, one or multiple network links can be added, removed, or re-wired. We use a very slow cooling schedule; we run the algorithm for 10^7 steps and the final temperature is 10^{-7} of the initial temperature. The objective function is $\frac{1}{n}\sum_{i=1}^{n}(y_i^* - \alpha_i)$ and the initial temperature is set fairly high at 0.01.

Fig. 3 Per Capita Improvement as Links Allocated to Sub-communities



Notes: There are $n = 200$ individuals; α_i, λ_i, and χ_i^{-1} are each drawn from Beta(2,2). Link allocation is the fraction of within-subcommunity links; the rest of the links are cross-subcommunity. The subcommunities are the ones in the optimal network. The dense network fixes the number of links such that when link allocation equals 1, each subcommunity is fully connected. Sparse network has a quarter of the links in the dense network. Per capita improvement is measured relative to out-of-network effort.

sub-communities that are highly densely interconnected within but not, or only very weakly, connected across. [10]

As the plots in the middle column of Fig. 4 show, each sub-community consists of individuals located on a downward sloping band in the parameter space of α versus λ, χ^{-1} or their sum. The optimal structure in scenarios 6 and 7 is especially noteworthy: It involves not only combining people from different clusters into the same sub-community but also allocating members of the same cluster to different sub-communities.

The plots in the third column in Fig. 4 show how the amount of effort exerted in the optimized network, y_i^*, varies as a function of not only consumers' behavioral parameters α_i, λ_i, and χ_i^{-1} but also their location in the network. These plots in the third column exhibit three patterns consistent with the analytical results and their intuition. The first pattern is that members of the same sub-community exhibit very little variation in y^*. This, of course, is consistent with conformity in a fully connected subgraph. The second pattern is that the amount of effort exerted in the optimized network tends to increase with the combination of stand-alone tendency to exert effort (α), the susceptibility to peer influence (λ), and the inability to exert influence (χ^{-1}). The third pattern is that people with exactly the same value of $\alpha + \lambda + \chi^{-1}$ exert different levels of effort y^* depending on which sub-community they are allocated to.

[10] Such networks are colloquially referred to as caveman graphs (e.g., Watts 2004, pp. 43-44 and 103), an extreme case in the small-world network literature. There is also a clear parallel between synchrony among oscillators (e.g., chirping crickets) studied in that literature and the social conformity among people we consider. However, small-world networks have not been studied as resulting from a centralized optimization.

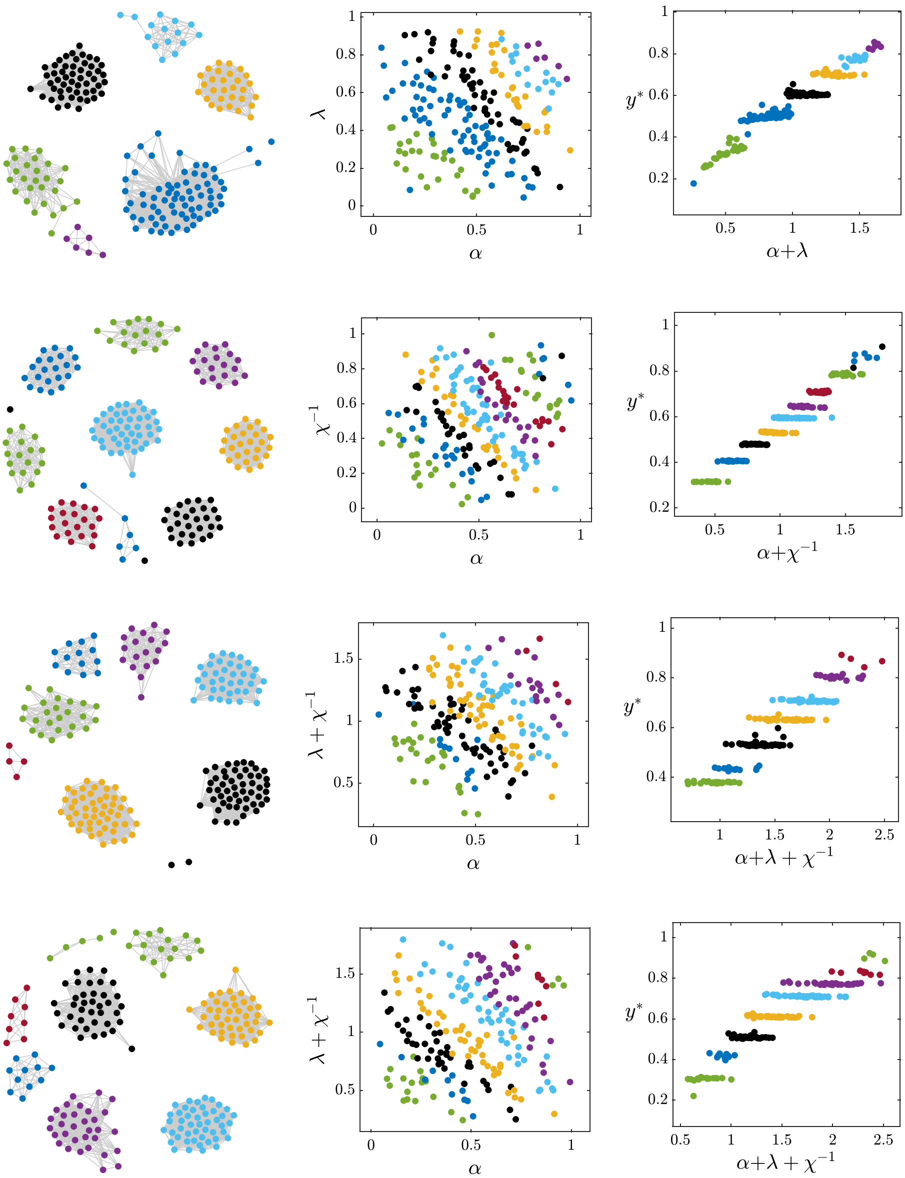


Fig. 4 Optimal Community Structure for Each of Seven Scenarios with 200 Individuals. Notes. Each row characterizes the optimal community structure for one of the seven scenarios, ordered from 1 to 7. Each row displays three plots. From left to right, the plots show (i) the optimal network where sub-communities are visually separated by different colors (but same color does not necessarily imply same sub-community), (ii) the location of individual community members in the parameter space, and (iii) the individual activity levels y_i plotted against the key parameters

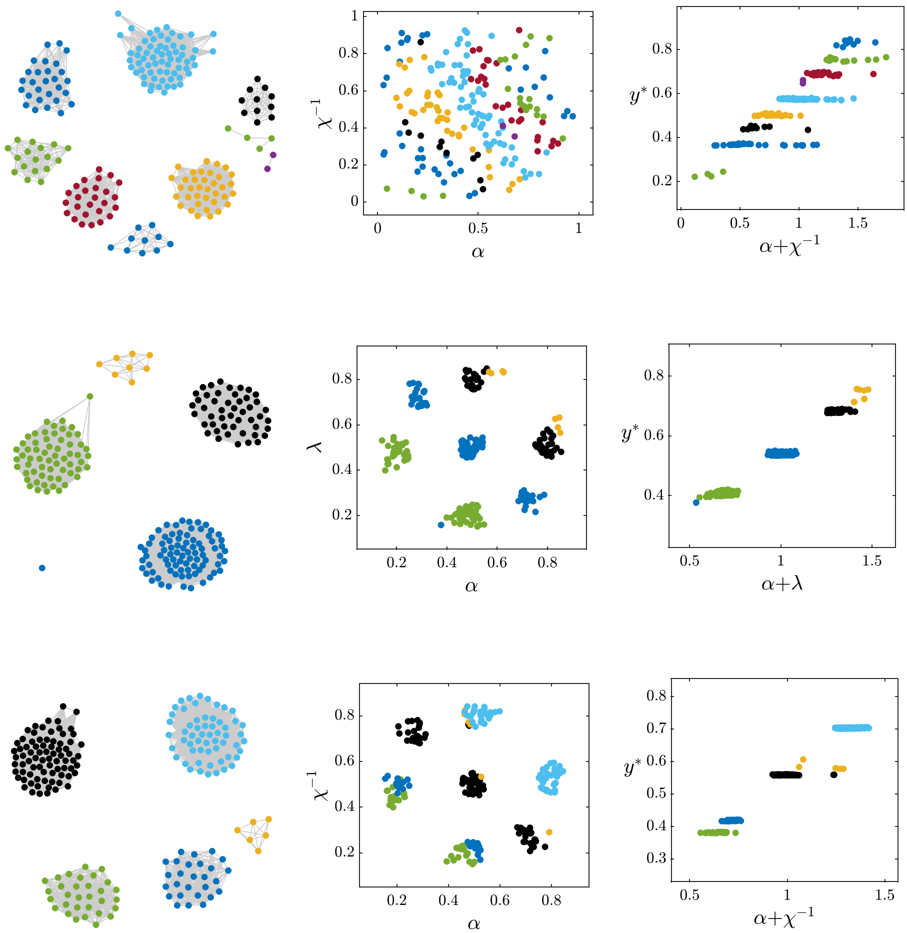


Fig. 4 (continued)

Overall, these numerical results for networks with $n = 200$ and various constellations of α_i, λ_i, and χ_i^{-1} parameters exhibit globally optimal community structures very similar to those identified in Propositions 2 and 3 for $n \to \infty$. Hence, one can view these numerical results as a corroboration of Propositions 2 and 3 for finite networks. Since simulated annealing is known to provide very good approximations to global rather than merely local optima, one can also view the numerical results as indicating that the community structures characterized in Propositions 2 and 3 are not only locally but likely also globally optimal.

For each scenario, we computed the fraction of customers exerting greater effort after being placed in the optimized network than when being disconnected, i.e., the fraction for which $y^* > \alpha_i$. For scenarios 1 through 7, these fractions were 66%, 69% 68%, 73%, 72%, 59%, and 57%. As we show next, it is not a coincidence that the improvement is greater in the first five scenarios where the parameters are distributed smoothly and exhibit greater variation than in the last two scenarios where the customers are clumped together and the parameters exhibit less variation.

For each of the seven scenarios, we also calculate how well the optimal network structure performs compared to a set of predefined structures. First is the global-star structure in which the individual with the highest $\alpha_i \chi_i / \lambda_i$ is connected to everyone else, and no other ties exist. Next are two global-core networks in which the top 10% or 20% individuals in terms of $\alpha_i \chi_i / \lambda_i$ form a core; individuals in the core are connected with each other; and any person not in the core is connected to a random person in the core. The last benchmark is the fully connected or full network.

Table 1 shows that the optimal network results in higher average or total effort exerted than any of these four benchmark structures, and does so in each of the seven scenarios. Table 1 also shows that, in the first five scenarios, the network structure that maximizes average effort tends to outperform the four benchmark structures also in terms of the fraction of people who increase their effort level. As one would expect, the largest increases in effort are achieved by people in the lowest quartile of α_i (the lowest stand-alone effort individuals) and the gains decrease as α_i increases. Again as expected, people in the highest quartile of stand-alone effort even decrease their effort after being connected to others. That is so for the optimal network as well as the four benchmark structures, but the downward pull is the smallest in the optimal structure, for six of the seven scenarios.

In some situations, like schooling and professional board certification, the goal is not to maximize effort per se but to bring as many people as possible above a threshold in effort or passing rate in outcome. To assess how well the various network structures compare on this metric, we proceed as follows. First, we define the threshold as a quantile of the α-distribution, and let it range from zero to unity in 5% increments. Then, we compute how many people reach that threshold in each network structure and scenario. Finally, we average the performance across scenarios.

Figure 5 graphs the performance for the optimal network and the global-core (20%) structure, which is the second-best performer in Table 1. A greater number of people meet the threshold in the optimal network than in the disconnected or empty network, at least as long as the threshold is below the 80th percentile value of α. Once the threshold is 80% or higher, the optimal network does slightly worse than the empty network, but only by a slim margin. This again shows how the optimal network manages the trade-off between pulling up the effort or outcome of low-α types and pushing down that of high-α types. Fig. 5 also shows that the optimal network in essence meets or beats the global-core (20%) structure for any threshold value between 0 and 1. Of course, a network designed to maximize the passing rate above a specific threshold will outperform the network design that maximizes average effort, but Fig. 5 suggests that the latter performs fairly well even on that alternative objective.

5 When is Community Design Most Beneficial?

Having characterized the optimal community structure for any given community, we now present results on how the total effort exerted in an optimized community varies with some important community characteristics: (i) the members' propensities to

Table 1 Comparison of Gains from Various Community Structures

	Community Network Structures				
	Optimal	Global-star	Global-core (10%)	Global-core (20%)	Full network
Scenario 1					
Effort lift, average individual	0.067	0.010	0.016	0.008	0.011
% individuals with effort increase	65.5%	53.5%	59.5%	58.0%	51.5%
Effort lift, 1st-quartile α_i	0.233	0.105	0.116	0.124	0.297
Effort lift, 2nd-quartile α_i	0.090	0.033	0.051	0.053	0.080
Effort lift, 3rd-quartile α_i	0.015	-0.017	-0.011	-0.017	-0.073
Effort lift, 4th-quartile α_i	-0.072	-0.080	-0.094	-0.129	-0.261
Scenario 2					
Effort lift, average individual	0.063	-0.005	-0.004	-0.003	-0.005
% individuals with effort increase	68.5%	49.0%	49.5%	50.0%	49.0%
Effort lift, 1st-quartile α_i	0.196	0.261	0.231	0.224	0.266
Effort lift, 2nd-quartile α_i	0.090	0.069	0.065	0.066	0.070
Effort lift, 3rd-quartile α_i	0.031	-0.076	-0.063	-0.061	-0.077
Effort lift, 4th-quartile α_i	-0.066	-0.274	-0.247	-0.240	-0.280
Scenario 3					
Effort lift, average individual	0.090	0.009	0.011	0.010	0.011
% individuals with effort increase	68.0%	51.0%	54.0%	55.5%	52.5%
Effort lift, 1st-quartile α_i	0.242	0.254	0.199	0.185	0.283
Effort lift, 2nd-quartile α_i	0.130	0.075	0.068	0.077	0.086

Table 1 (continued)

	Community Network Structures				
	Optimal	Global-star	Global-core (10%)	Global-core (20%)	Full network
Effort lift, 3rd-quartile α_i	0.043	-0.058	-0.038	-0.032	-0.060
Effort lift, 4th-quartile α_i	-0.059	-0.237	-0.185	-0.189	-0.262
Scenario 4					
Effort lift, average individual	0.116	-0.009	0.002	0.001	-0.006
% individuals with effort increase	73.0%	48.0%	51.5%	51.5%	49.5%
Effort lift, 1st-quartile α_i	0.269	0.275	0.220	0.207	0.293
Effort lift, 2nd-quartile α_i	0.152	0.070	0.067	0.073	0.076
Effort lift, 3rd-quartile α_i	0.083	-0.098	-0.055	-0.061	-0.098
Effort lift, 4th-quartile α_i	-0.049	-0.284	-0.224	-0.216	-0.297
Scenario 5					
Effort lift, average individual	0.093	-0.038	-0.037	-0.037	-0.043
% individuals with effort increase	72.0%	42.5%	42.0%	43.5%	42.0%
Effort lift, 1st-quartile α_i	0.249	0.232	0.204	0.183	0.244
Effort lift, 2nd-quartile α_i	0.126	0.037	0.032	0.035	0.035
Effort lift, 3rd-quartile α_i	0.045	-0.109	-0.097	-0.094	-0.119
Effort lift, 4th-quartile α_i	-0.055	-0.312	-0.285	-0.274	-0.336

Table 1 (continued)

	Community Network Structures				
	Optimal	Global-star	Global-core (10%)	Global-core (20%)	Full network
Scenario 6					
Effort lift, average individual	0.047	0.005	0.009	0.007	0.012
% individuals with effort increase	58.5%	56.5%	67.0%	69.0%	57.0%
Effort lift, 1st-quartile α_i	0.221	0.095	0.104	0.105	0.265
Effort lift, 2nd-quartile α_i	0.048	0.012	0.022	0.032	0.045
Effort lift, 3rd-quartile α_i	0.032	-0.004	0.004	0.008	-0.012
Effort lift, 4th-quartile α_i	-0.113	-0.082	-0.095	-0.117	-0.249
Scenario 7					
Effort lift, average individual	0.050	0.012	0.013	0.012	0.014
% individuals with effort increase	57.0%	60.0%	63.0%	63.5%	59.5%
Effort lift, 1st-quartile α_i	0.232	0.238	0.239	0.237	0.285
Effort lift, 2nd-quartile α_i	0.041	0.038	0.041	0.043	0.045
Effort lift, 3rd-quartile α_i	0.040	-0.022	-0.018	-0.020	-0.026
Effort lift, 4th-quartile α_i	-0.113	-0.207	-0.212	-0.211	-0.249

The scenarios correspond to the rows in Figure 4.

influence, be influenced, and exert stand-alone effort, (ii) the heterogeneity across the members, and (iii) the size of the community. These results help designers understand under what circumstances they can benefit the most from purposively configuring the network structure.

Consider two platforms with a consumer base identical in size (n) and in the isolated activity levels of their members ($\boldsymbol{\alpha} \equiv \{\alpha_i\}_{i=1}^n$). If the influence potential ($\boldsymbol{\rho} \equiv \{\rho_{ij}\}_{i,j=1}^n$) among the consumers of the first platform is greater than that of the second, the maximum level of activity that can be attained by the second platform can never exceed that of the first. The following proposition formalizes this.

Proposition 4 ***(Strength of Conformity and Gain from Community Design)*** *Consider the consumers in two communities, A and B, characterized by $A = (\boldsymbol{\alpha}, \boldsymbol{\rho})$ and $B = (\boldsymbol{\alpha}, \boldsymbol{\rho}')$. Suppose $\rho'_{ij}/\rho_{ij} = \rho'_{ji}/\rho_{ji} \geq 1$ for all (i, j) pairs. Then, with each community being given its optimal configuration, the total effort level in B is greater than or equal to that in A.*

This proposition means that stronger conformity provides greater opportunities for optimal design to boost the total level of effort exerted. The community designer can always find a structure that leverages greater conformity, and consequently can always find a design in which the positive motivating effect of conforming to others' effort weakly dominates the negative demotivating effect of conformity.

The following proposition quantifies the gain from placing consumers in a well-designed community structure as opposed to keeping them in isolation.

Proposition 5 ***(Heterogeneity and Gain from Community Design)*** *Consider a consumer base $\{\alpha_i, \{\rho_{ij}\}_{j \neq i}\}_{i=1}^n$. Regardless of the network configuration,*

$$\frac{1}{n}\left(T - \sum_{i=1}^{n} \alpha_i\right) \leq \frac{n-1}{2}\rho^+\alpha^+$$

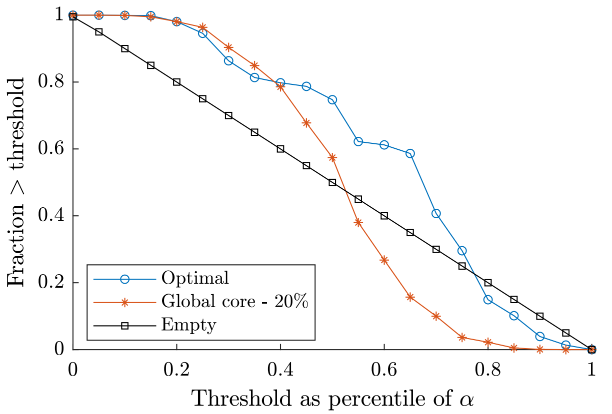


Fig. 5 "PASS THE THRESHOLD" PERFORMANCE OF OPTIMAL NETWORK AND GLOBAL-CORE (20%) STRUCTURE

where $\rho^+ \equiv \max\{|\rho_{ij} - \rho_{ji}|, i \neq j\}$ *and* $\alpha^+ \equiv \max\{|\alpha_i - \alpha_j|, i \neq j\}$ *are maximum differences across all* (i, j) *pairs. In words, the upper bound on the improvement of an average consumer depends on the heterogeneity among consumers in their propensities to influence, be influenced, and exert effort stand-alone. The greater the heterogeneity, the higher the upper bound.*

The proposition implies that the gain from designing any community is zero when each member has the same influence χ and the same susceptibility to influence λ. The proposition also implies that communities showing little variation in α provide lower potential for improvement through peer influence than heterogeneous ones. Hence, social networks formed organically where people tend to connect with others with a similar α (tie formation exhibiting homophily in α) provide little benefit from networking and may even perform worse than randomly connected networks (Carrell et al. 2013). This insight emphasizes the importance of purposive design or "central planning" for effective communities.

The proposition also has important implications for customer targeting. Many large firms prefer to segment consumers into externally heterogeneous and internally homogeneous groups, and then target each segment with possibly tailored offerings. Altogether, identifying and targeting internally homogeneous segments of customers is a good practice. In the setting we study, in contrast, maintaining or even seeking heterogeneity among one's customers or members is beneficial and a firm or service provider with a homogeneous customer or membership base will see smaller gains from being in a community, as will its average customer or member.

The importance of seeking heterogeneity within sub-communities was also illustrated in the numerical analyses of the two scenarios exhibiting clumping in the $\alpha - \lambda - \chi$ parameter space. The globally optimal community structure identified by simulated annealing in scenarios 6 and 7 involved not only combining people from different clusters into the same sub-community, but also allocating members of the same cluster to different sub-communities.

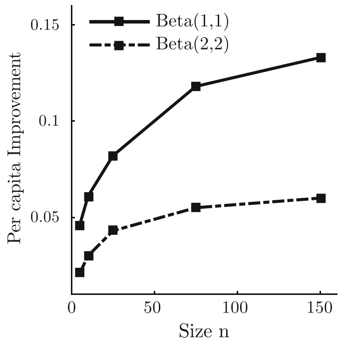
Next, we show that the size of the customer or membership base also affects the gains from community design, independently of the heterogeneity of the base. To this end, we consider a community with n consumers whose characteristics $(\alpha_i, \lambda_i, \chi_i)$ are heterogeneous and drawn from a non-degenerate distribution Φ. For any set of n consumers, there is an optimal design that leads to an improvement in the community activity, $\sum_i (y_i^* - \alpha_i)$. We denote by $Q(\Phi, n)$ the expectation of this improvement over the distribution Φ, and by $q(\Phi, n) \equiv Q(\Phi, n)/n$ the expected per capita improvement.

Proposition 6 ***(Community Size and Gain from Community Design)*** *The expected per capita improvement in effort exerted from optimal community design satisfies the following inequality:*[11]

$$q(\Phi, n+1) > \frac{\sum_{k=1}^{n} q(\Phi, k)}{n} \quad , \; n = 1, 2, ...$$

[11]In fact, a stronger result also holds: $q(\Phi, n+1) > \frac{\sum_{k=1}^{n} k \cdot q(\Phi, k)}{\sum_{k=1}^{n} k}$. The right hand side of this inequality is a weighted average and assigns more weight to the larger communities.

Fig. 6 Expected per Capita Improvement $q(\Phi, n)$



Notes: $q(\Phi, n)$ is evaluated by averaging across 100 communities. For each community, members are drawn from a distribution Φ and the optimal community network is computed. Φ is specified as follows: for the first column, $(\alpha_i, \lambda_i) \sim \text{Beta}(1,1) \times \text{Beta}(1,1)$ and $\chi_i = 1$; for the second column, $(\alpha_i, \lambda_i) \sim \text{Beta}(2,2) \times \text{Beta}(2,2)$ and $\chi_i = 1$. Per capita improvement in effort is measured relative to out-of-network effort.

The proposition provides a lower bound for the expected per capita improvement from optimally arranged communities. It states that the per capita improvement in a population of size $n + 1$, i.e., $q(\Phi, n + 1)$, exceeds the average of the per capita improvements of smaller communities: $q(\Phi, n)$, $q(\Phi, n-1)$, ..., and $q(\Phi, 1)$. Therefore, working with more customers or members will generate greater improvements in the average level of effort exerted after the optimization.

Figure 6 illustrates the analytical results of Propositions 5 and 6 with a numerical example. It plots the expected average benefit $q(\Phi, n)$ as a function of n for two different heterogeneity distributions, where one has greater variations in α_i and λ_i than the other.[12] Clearly, the per capita improvement from placing consumers in optimized community structures increases with the number of members (Proposition 6) and with the heterogeneity in the stand-alone effort α_i and the influenceability λ_i of these members (Proposition 5).

The intuition is straightforward. Larger and more heterogeneous sets of nodes provide more opportunities for the designer to identify and select a structure where, on balance, the positive motivating effect of conformity on low α_i nodes outweighs its negative demotivating effect on high α_i nodes.

[12] In the Figure we assume that $(\alpha_i, \lambda_i) \sim \text{Beta}(2, 2) \times \text{Beta}(2, 2)$ and $\text{Beta}(1, 1) \times \text{Beta}(1, 1)$. Note that Beta(2, 2) is a symmetric bell shaped distribution over the unit interval centered around 0.5, appropriate for populations where there is a greater mass of individuals carrying close to median characteristics in the importance they give to their goals and the degree of susceptibility they are subject to. In contrast, Beta(1, 1) is the uniform distribution, suggesting a more homogeneous spread of individual types.

6 Alternative Models

6.1 Minimizing Activity in Communities

In some contexts, activity can be harmful (e.g., smoking, gambling, addictive gaming) and a community designer may want to minimize rather than to maximize it. It is easy to show that an activity-minimizing community for $(\boldsymbol{\alpha}, \boldsymbol{\rho})$ is, at the same time, an activity-maximizing community for $(-\boldsymbol{\alpha}, \boldsymbol{\rho})$. So it is straightforward to transform a minimization problem into a maximization problem, and the results presented in Sections 4 and 5 apply as long as the negative value of the individual activity $(-\boldsymbol{\alpha})$ is used in the calculations and community design.

6.2 Social Competition

The model and analysis assume that peer influence operates through social conformity, specifically the desire to adhere to descriptive social norms. It is possible, however, that social influence operates through other mechanisms. In this section, we consider social competition or the desire to out-perform others, which may be relevant for engaging in intense physical exercise, at least to some people (Aral and Nicolaides 2017) as opposed to managing one's weight and fitness through mindful eating and moderate exercise, reducing one's electricity consumption, reducing one's tobacco or drug use, making charitable donations, and studying. To account for peer influence through social competition rather than conformity, we modify the model setup given in equations 2 and 4 into:

$$u_i = \alpha_i y_i - \frac{1}{2} y_i^2 + \sum_{j \sim i} \left[\delta_{ij} (y_i - y_j) \right], \tag{12}$$

where $\delta_{ij} > 0$ measures how much individual i cares about doing better than individual j. For instance, he may want to exercise harder than his peers or outperform other salespeople at work. He may care greatly about exceeding the level of effort or activity of some peers but care little about others. So, relationships are heterogeneous and $\delta_{ij} \neq \delta_{ik}$. The implied effort chosen by individual i is:

$$y_i = \alpha_i + \sum_{j \sim i} \delta_{ij}. \tag{13}$$

When peer influence operates through social competition, effort is always higher in-community than out-of-community effort, social information should always be shared, and the completely connected community is the optimal structure. Social competition thus presents less of a design challenge and limited trade-offs compared to social conformity, and therefore is more straightforward from a community design perspective.

6.3 Maximizing Outcome

In some situations, there may be a necessity to separate measures for outcome (e.g., test score) and effort. Suppose that an outcome measure, t_i, follows a mapping of effort that is heterogeneous among individuals:

$$t_i = \sigma_i y_i.$$

where $\sigma_i > 0$. Individuals with a large σ_i find it easier to achieve a better outcome with the same amount of effort. Our model accommodates situations where one wants to maximize the collective outcome instead of efforts: $\sum_i t_i$. Specifically, the utility function should be modified as follows:

$$u_i = \alpha_i t_i - \frac{1}{2} y_i^2 - \frac{1}{2} \sum_{j \sim i} \lambda_i \cdot \chi_j \cdot (t_i - t_j)^2, \tag{14}$$

where the individual reward and social conformity are both based on outcomes, unlike the cost of exerting effort. Because utility can be scaled up or down without affecting behaviors, the above utility is equivalent to

$$u_i = \alpha_i' t_i - \frac{1}{2} t_i^2 - \frac{1}{2} \sum_{j \sim i} \lambda_i' \cdot \chi_j \cdot (t_i - t_j)^2, \tag{15}$$

where $\alpha_i' \equiv \sigma_i^2 \alpha_i$ and $\lambda_i' \equiv \sigma_i^2 \lambda_i$. Consequently, we have

$$t_i = \frac{\alpha_i' + \sum_{j \sim i} \lambda_i' \cdot t_j}{1 + \sum_{j \sim i} \lambda_i' \cdot \chi_j}. \tag{16}$$

From this point on, one can apply all of our results on $\{\alpha_i', \lambda_i', \chi_i\}_{i=1}^n$ to maximize $\sum_i t_i$.

6.4 Uni-directional Connections

Our analysis assumed that influence flows both ways over social connections, even though the strength of influence need not be symmetric. What happens when the community manager has the ability to make the influence within communities flow in only one direction, e.g., by informing i about j's effort or actions, but not vice versa? The community manager will want to connect every higher-α type to every lower-α type. Since the downward pull of lower-α types can be avoided, there is no reason to have separate sub-communities, and the optimal community structure will be a single large fully connected graph component. The ability to make ties uni-directional renders the optimal community structure very easy to identify.

6.5 Rewiring Connections

Consistent with the empirical evidence reviewed in Section 2, we assumed that community members accept their assigned location in the community designed to help them achieve a specific goal. How critical is that assumption? Specifically, how different would the optimal community design be when, in a first stage, the designer imposes a structure and people join and, in a second stage, community members experiencing tension from non-conforming peers are able to rewire their connections at a cost reflecting inertia and the hassle cost of identifying other potential partners?

One possible assumption is that in the second stage people seek to achieve conformity by connecting to those with a first-stage effort level very similar to theirs. Since in the current optimal design, within-community variance in y^* is small, and especially so in the large communities on the main diagonals, there will be only limited rewiring.

An alternative assumption is that people will seek to form connections with others having a parameter α similar to theirs. We expect more rewiring in this second scenario, but still expect that the optimal design will not be radically different from the one we have identified. In the current design rules, high-α types with low influence are already paired with other high-α types, and low-α types with low susceptibility are already paired with other low-α types. So, the communities consisting of people located towards the off-diagonal corners are already fairly homogeneous in α. The key question is how much the heterogeneous sub-communities on the main diagonals change once rewiring starts. That is likely to be a function of the cost of rewiring relative to the tension induced by the difference in α between two connected community members. Hence, rewiring does not drive the total benefit from community optimization to zero. Rather, the total gets diluted and the amount of dilution is a function of the strength of homophily in organically formed social networks, the presence of not only high-α and low-α but also medium-α types in mixed sub-communities (Carrell et al. 2013), the cost of rewiring, the time lapsed until rewiring takes place, and the community designer’ s discount factor. Most importantly, we see no reason to expect the optimal design rules in this two-stage set-up to differ markedly from those we have identified.

7 Conclusion

In the absence of a clear benchmark about what is appropriate or desirable, people tend to conform to what others, even strangers, do. Conformity to descriptive social norms has a sizable effect on the effort people exert in areas like energy conservation, health and charitable help, and is increasingly being leveraged by companies and platforms like Opower, Crowdrise and BuddyUp. Motivated by these important developments, and the broader emerging interest in community design, we address the question of how to design the reference groups among community members to maximize the total effort exerted. The key trade-off to be managed is that social conformity pushes up the effort of those who would otherwise exert relatively little effort while pushing down the effort of those who would otherwise exert relatively greater effort.

7.1 Recommendations

Our analysis results in three specific recommendations. First, the community should be organized as a set of sub-communities, each of which is fully connected within itself but disconnected from other sub-communities. Second, each sub-community should consists of members selected such that their susceptibility to influence (propensity to influence) and their "standalone" propensity to exert effort are negatively (positively) correlated. Third, there is no benefit in varying the strengths of connections across dyads, e.g., by varying their visual salience. Our analysis also shows that the benefits of following these design rules are greater in communities that count more members or exhibit greater variation in their members' social influence, susceptibility to social influence, and propensity to exert effort in isolation. The reason is that larger size and greater variance provides the designer greater degrees of freedom when managing the key trade-off.

7.2 Implementation

One appeal of our recommendations is that the structure is easy to implement not only for online but also for offline communities. Organizing the overall community into sub-communities that are fully connected within but disconnected between amounts to creating distinct groups where every member is connected to every other member of their own group and to no one else. Not having to vary the visual salience, interaction frequency, or other facets of tie strength across dyads also makes the optimal community design easy to implement.

Assigning specific members to specific sub-communities requires information about their stand-alone propensity to exert effort, their social influence, and their susceptibility to influence. One practical way to obtain such information is through self-reports at the time people join the community. Stand-alone effort is likely correlated with both aptitude and grit, the latter being the passion and motivation for the particular goal the community is designed to foster (Duckworth et al. 2007). Influence and susceptibility to influence can be measured through simple questionnaires similar to those used to measure opinion leaders and seekers (Flynn et al. 1996), though using easy-to-observe correlates of influence and influenceability identified by analyzing prior member behavior may be a better procedure for managers of established communities (Aral and Walker 2012; Iyengar et al. 2011; Iyengar et al. 2015).

A final implementation issue deals with how sustainable the network design is. Left to their own devices, people tend to form networks that exhibit two key structural features: clustering and homophily. The tendency of organic networks to be clustered ("friends of friends are friends") does not pose a challenge, since the proposed design involves complete clustering. Three important elements imply that our design rules are likely to be robust to varying levels of homophily. First, in our analysis, people prefer being connected to others, even if they have a different α, as long as $\rho \leq 1$ (see Equation 2). Consequently, being connected to peers with a different α is no reason for anyone to prefer not to be part of the designed community. Second, the empirical evidence from the literature reviewed in Section 2 shows that (i) people joining an organization or interest-based community in pursuit of a goal are often willing to accept an imposed group or network structure and that (ii) such an imposed structure often has a lasting effect on subsequent network structure and individual behavior, even after people are allowed to rewire their network connections freely (e.g., Centola 2011; Festinger et al. 1950; Hasan and Bagde 2013; Sacerdote 2001; Zhang et al. 2015a). Third, the optimal structure never forces high-α and low-α types together without medium-α types also being present and providing cohesion even in the presence of homophily (Carrell et al. 2013). Consequently, though a purely homophily-based rewiring will dilute the benefits of network design, high levels of homophily will not negate them.

7.3 Suggestions for Future Research

Community and network design is an emerging area of research (Cerdeiro et al. 2017; Valente 2012), and many questions remain open. We studied the problem of a designer, a manager, or a policy maker who seeks to maximize total effort, in a setting where community members conform to descriptive social norms, and the key trade-off involves balancing the positive conformity effect on people who would exert less effort when disconnected from others against the negative conformity effect on those who would exert more effort when disconnected.

The design problem we studied is applicable to many areas of marketing, presenting different combinations of design objective, nature of social interaction, and key trade-off to be managed. One such area is ideation platforms, where the manager brings customers or employees together to generate ideas. Connected people may improve on each other' s ideas, and the trade-off is to connect people with experience and expertise levels that are sufficiently different but not too different (De Vaan et al. 2015; Toubia and Netzer 2016; Uzzi et al. 2013). Another area is learning about new products and technologies when the objective is knowledge spread and acquisition, customers learn through observation, and the trade-off is to have informative social learning that does not overpower private learning (Zhang et al. 2015b). Yet another area is managing the trade-off between motivation and demotivation in performance posting, e.g., when publicly ranking the performance of salespeople or customer representatives (Cai 2015).

There are some limitations in our model that future research may seek to address. We make the assumption that individuals wish to conform to social norms, $\rho > 0$. In Section 6.2, we also consider the case of social competition. However, it is possible that individuals may want to differentiate themselves ($\rho < 0$). Moreover, our model implicitly assumes that the effort exerted or output achieved is observable. In several domains such as exercise and energy conservation, it is actually possible for companies to observe individual effort as workout and energy conservation can be measured by a third party and reported to the user's network. In the Opower example, energy consumption is measured by the utility provider. In the workout examples, typically smart devices measure and report effort. In other situations, however, effort may not be observable, or may be only partially observable, potentially creating a moral hazard problem. Finally, our work considers continuous rather than discrete effort. When effort is discrete, multiple equilibria may occur (Hartmann 2010; Cerdeiro et al. 2017; Goyal et al. 2017).

Community design need not be restricted to choosing a structure and may, for instance, also include a policy of incentives given to high-α types. This was not important in our analysis where people prefer being connected to others, even if they have a different α, as long as $\rho \leq 1$, but it may be useful when $\rho > 1$. Giving incentives to high-α types may also be effective in settings where the social influence mechanism is not one of conformity but of status or knowledge spillover, such that people like to connect to high-α types but the latter may need additional incentives to accept invitations to connect from low-α types (Wei et al. 2016). Such incentives may include public tokens of status, special benefits like pre-release access to new products, membership in advisory councils, and various other incentives already implemented in many communities. As companies and other platforms increasingly rely on building and managing communities, network design is bound to offer increasing opportunities to improve business outcomes and consumer welfare.

References

Acemoglu, D., Dahleh, M.A., Lobel, I., Ozdaglar, A. (2011). Bayesian learning in social networks. *The Review of Economic Studies*, *78*(4), 1201–1236.

Advani, A., & Reich, B. (2015). Melting pot or salad bowl: The formation of heterogeneous communities. Technical report, IFS Working Papers.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9), 1082–1095.

Allcott, H., & Kessler, J.B. (2015). The welfare effects of nudges: A case study of energy use social comparisons. Technical Report w21671, National Bureau of Economic Research.

Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, *104*(10), 3003–3037.

Ansari, A., Stahl, F., Heitmann, M., Bremer, L. (2018). Building a social network for success. *Journal of Marketing Research*, *55*(3), 321–338.

Aral, S., Muchnik, L., Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, *106*(51), 21544–21549.

Aral, S., & Nicolaides, C. (2017). Exercise contagion in a global social network. *Nature Communications*, *8*, 14753.

Aral, S., & Walker, D. (2012). Identifying influential and susceptible members of social networks. *Science*, *337*(6092), 337–341.

Ascarza, E., Ebbes, P., Netzer, O., Danielson, M. (2017). Beyond the target customer: Social effects of customer relationship management campaigns. *Journal of Marketing Research*, *54*(3), 347–363.

Asch, S.E. (1956). Studies of independence and conformity: i. A minority of one against a unanimous majority. *Psychological Monographs:, General and Applied*, *70*(9), 1–70.

Ayres, I., Raseman, S., Shih, A. (2013). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *Journal of Law, Economics, and Organization*, *29*(5), 992–1022.

Bagozzi, R.P., & Dholakia, U.M. (2002). Intentional social action in virtual communities. *Journal of Interactive Marketing*, *16*(2), 2–21.

Bernheim, B.D. (1994). A theory of conformity. *Journal of Political Economy*, *102*(5), 841–877.

Bernheim, B.D., & Exley, C.L. (2015). Understanding conformity: An experimental investigation. Technical report, Working paper 16-070, Harvard Business School.

Bimpikis, K., Ozdaglar, A., Yildiz, E. (2016). Competitive targeted advertising over networks. *Operations Research*, *64*(3), 705–720.

Bojanowski, M., & Buskens, V. (2011). Coordination in dynamic social networks under heterogeneity. *The Journal of Mathematical Sociology*, *35*(4), 249–286.

Cai, J. (2015). *Silent or salient? Perks and perils of performance posting*. Working paper: Texas A&M University.

Carrell, S.E., Sacerdote, B.I., West, J.E. (2013). From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica*, *81*(3), 855–882.

Cartwright, D., & Harary, F. (1956). Structural balance: A generalization of Heider's theory. *Psychological Review*, *63*(5), 277–293.

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, *329*(5996), 1194–1197.
Centola, D. (2011). An experimental study of homophily in the adoption of health behavior. *Science*, *334*(6060), 1269–1272.
Centola, D. (2018). *How Behavior Spreads: The Science of Complex Contagions*. Princeton, NJ: Princeton University Press.
Centola, D., & van de Rijt, A. (2015). Choosing your network: Social preferences in an online health community. *Social Science & Medicine*, *125*, 19–31.
Cerdeiro, D.A., Dziubiński, M., Goyal, S. (2017). Individual security, contagion, and network design. *Journal of Economic Theory*, *170*, 182–226.
Chen, Y., Harper, F.M., Konstan, J., Xin Li, S. (2010). Social comparisons and contributions to online communities: A field experiment on MovieLens. *American Economic Review*, *100*(4), 1358–1398.
Chen, Y., Wang, Q., Xie, J. (2011). Online social interactions: a natural experiment on word of mouth versus observational learning. *Journal of Marketing Research*, *48*(2), 238–254.
Cialdini, R.B., & Goldstein, N.J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*, 591–621.
Cialdini, R.B., Reno, R.R., Kallgren, C.A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015–1026.
Davis, J.A. (1970). Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices. *American Sociological Review*, *35*(5), 843–851.
De Vaan, M., Stark, D., Vedres, B. (2015). Game changer: the topology of creativity. *American Journal of Sociology*, *120*(4), 1144–1194.
Dholakia, U.M., Bagozzi, R.P., Pearo, L.K. (2004). A social influence model of consumer participation in network-and small-group-based virtual communities. *International Journal of Research in Marketing*, *21*(3), 241–263.
Duckworth, A.L., Peterson, C., Matthews, M.D., Kelly, D.R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087.
Ellwardt, L., Hernandez, P., Martınez-cánovas, G., Munoz-Herrera, M. (2016). Conflict and segregation in networks: An experiment on the interplay between individual preferences and social influence. *Journal of Dynamics and Games*, *3*(2), 191–216.
Ferraro, P.J., & Price, M.K. (2013). Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. *Review of Economics and Statistics*, *95*(1), 64–73.
Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140.
Festinger, L., Schachter, S., Back, K. (1950). *Social Pressures in Informal Groups: A Study of Human Factors in Housing*. Palo Alto, CA: Stanford University Press.
Flynn, L.R., Goldsmith, R.E., Eastman, J.K. (1996). Opinion leaders and opinion seekers: Two new measurement scales. *Journal of the Academy of Marketing Science*, *24*(2), 137–147.
Frey, B.S., & Meier, S. (2004). Social comparisons and pro-social behavior: Testing “conditional cooperation” in a field experiment. *American Economic Review*, *94*(5), 1717–1722.
Gastner, M.T., & Newman, M. (2006). Optimal design of spatial distribution networks. *Physical Review E*, *74*(1), 016117.
Godes, D., & Mayzlin, D. (2009). Firm-created word-of-mouth communication: Evidence from a field test. *Marketing Science*, *28*(4), 721–739.
Goel, S., & Goldstein, D.G. (2013). Predicting individual behavior with social networks. *Marketing Science*, *33*(1), 82–93.
Goldenberg, J., Han, S., Lehmann, D.R., Hong, J.W. (2009). The role of hubs in the adoption process. *Journal of Marketing*, *73*(2), 1–13.
Goldenberg, J., Shavitt, Y., Shir, E., Solomon, S. (2005). Distributive immunization of networks against viruses using the ‘honey-pot’ architecture. *Nature Physics*, *1*(3), 184.
Goldstein, N.J., Cialdini, R.B., Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, *35*(3), 472–482.
Goyal, S., Hernández, P., Martínez-cánovasz, G., Moisan, F., Munoz-Herrera, M., Sanchez, A. (2017). Integration and diversity. working paper.
Haag, M., & Lagunoff, R. (2006). Social norms, local interaction, and neighborhood planning. *International Economic Review*, *47*(1), 265–296.

Hallsworth, M., List, J.A., Metcalfe, R.D., Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, *148*, 14–31.

Hartmann, W.R. (2010). Demand estimation with social interactions and the implications for targeted marketing. *Marketing Science*, *29*(4), 585–601.

Hasan, S., & Bagde, S. (2013). The mechanics of social capital and academic performance in an Indian college. *American Sociological Review*, *78*(6), 1009–1032.

Hill, S., Provost, F., Volinsky, C. (2006). Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science*, *21*(2), 256–276.

Hu, Y., & Van den Bulte, C. (2014). Nonmonotonic status effects in new product adoption. *Marketing Science*, *33*(4), 509–533.

Iijima, R., & Kamada, Y. (2017). Social distance and network structures. *Theoretical Economics*, *12*(2), 655–689.

Iyengar, R., Van den Bulte, C., Lee, J.Y. (2015). Social contagion in new product trial and repeat. *Marketing Science*, *34*(3), 408–429.

Iyengar, R., Van den Bulte, C., Valente, T.W. (2011). Opinion leadership and social contagion in new product diffusion. *Marketing Science*, *30*(2), 195–212.

Iyer, G., & Katona, Z. (2015). Competing for attention in social communication markets. *Management Science*, *62*(8), 2304–2320.

Katona, Z., Zubcsek, P.P., Sarvary, M. (2011). Network effects and personal influences: The diffusion of an online social network. *Journal of Marketing Research*, *48*(3), 425–443.

Kelman, H.C. (1958). Compliance, identification, and internalization three processes of attitude change. *Journal of Conflict Resolution*, *2*(1), 51–60.

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P. (1983). Optimization by simulated annealing. *Science*, *220*(4598), 671–680.

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

Kraut, R.E., Resnick, P., Kiesler, S., Burke, M., Chen, Y., Kittur, N., Konstan, J., Ren, Y., Riedl, J. (2012). *Building Successful Online Communities: Evidence-based Social Design*. Cambridge, MA: MIT Press.

Lewis, T.G. (2011). *Network Science: Theory and Applications*. New York, NY: John Wiley & Sons.

Martin, R., & Randal, J. (2008). How is donation behaviour affected by the donations of others? *Journal of Economic Behavior & Organization*, *67*(1), 228–238.

McAndrew, T.C., Danforth, C.M., Bagrow, J.P. (2015). Robustness of spatial micronetworks. *Physical Review E*, *91*(4), 042813.

Murtha, B.R., Bharadwaj, S.G., Van den Bulte, C. (2014). *Interlocking networks: How and when do connections between buying and selling teams affect customer solutions? 14-120*. Cambridge, MA: Marketing Science Institute Report.

Neary, P.R. (2012). Competing conventions. *Games and Economic Behavior*, *76*(1), 301–328.

Newman, M. (2010). *Networks: An Introduction*. Oxford, UK: Oxford University Press.

Newman, M.E. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, *74*(3), 036104.

Peres, R., & Van den Bulte, C. (2014). When to take or forgo new product exclusivity: Balancing protection from competition against word-of-mouth spillover. *Journal of Marketing*, *78*(2), 83–100.

Phan, T.Q., & Godes, D. (2018). The evolution of influence through endogenous link formation. *Marketing Science*, *37*(2), 259–278.

Phillips, D.J., & Zuckerman, E.W. (2001). Middle-status conformity: Theoretical restatement and empirical demonstration in two markets. *American Journal of Sociology*, *107*(2), 379–429.

Plackett, R.L. (1965). A class of bivariate distributions. *Journal of the American Statistical Association*, *60*(310), 516–522.

Randall, M., McMahon, G., Sugden, S. (2002). A simulated annealing approach to communication network design. *Journal of Combinatorial Optimization*, *6*(1), 55–65.

Rovniak, L.S., Kong, L., Hovell, M.F., Ding, D., Sallis, J.F., Ray, C.A., Kraschnewski, J.L., Matthews, S.A., Kiser, E., Chinchilli, V.M., et al. (2016). Engineering online and in-person social networks for physical activity: A randomized trial. *Annals of Behavioral Medicine*, *50*(6), 885–897.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics*, *116*(2), 681–704.

Shang, J., & Croson, R. (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *Economic Journal*, *119*(540), 1422–1439.

Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology*, *187*, 23–46.
Toubia, O., & Netzer, O. (2016). Idea generation, creativity, and prototypicality. *Marketing Science*, *36*(1), 1–20.
Uetake, K., & Yang, N. (2020). Inspiration from the ”biggest loser”: Social interactions in a weight loss program. *Marketing Science*, *39*(3), 487–499.
Uzzi, B., Mukherjee, S., Stringer, M., Jones, B. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468–472.
Valente, T.W. (2012). Network interventions. *Science*, *337*(6090), 49–53.
Van den Bulte, C., & Joshi, Y.V. (2007). New product diffusion with influentials and imitators. *Marketing Science*, *26*(3), 400–421.
Van den Bulte, C., & Lilien, G.L. (2001). Medical innovation revisited: Social contagion versus marketing effort. *American Journal of Sociology*, *106*(5), 1409–1435.
Van den Bulte, C., & Moenaert, R.K. (1998). The effects of R&D team co-location on communication patterns among R&D, marketing, and manufacturing. *Management Science*, *44*(11-part-2), S1–S18.
Vega-Redondo, F. (2016). Links and actions in interplay. In Bramoullé, Y., Galeotti, A., Rogers, B. (Eds.) *Oxford Handbook of the Economics of Networks, Oxford: Oxford University Press*.
Vincent, B., & Ismael, M. (2017). My friend far, far away: A random field approach to exponential random graph models. *The Econometrics Journal*, *20*(3), S14–S46.
Watts, D.J. (2004). *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton, NJ: Princeton University Press.
Wei, Y., Yildirim, P., Van den Bulte, C., Dellarocas, C. (2016). Credit scoring with social network data. *Marketing Science*, *35*(2), 234–258.
Zafar, B. (2011). An experimental investigation of why individuals conform. *European Economic Review*, *55*(6), 774–798.
Zhang, J., Brackbill, D., Yang, S., Becker, J., Herbert, N., Centola, D. (2016). Support or competition? How online social networks increase physical activity: A randomized controlled trial. *Preventive Medicine Reports*, *4*, 453–458.
Zhang, J., Brackbill, D., Yang, S., Centola, D. (2015a). Efficacy and causal mechanism of an online social media intervention to increase physical activity: Results of a randomized controlled trial. *Preventive Medicine Reports*, *2*, 651–657.
Zhang, J., Liu, Y., Chen, Y. (2015b). Social learning in networks of friends versus strangers. *Marketing Science*, *34*(4), 573–589.