Social Science & Medicine 125 (2015) 19–31

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Choosing your network: Social preferences in an online health community

Damon Centola [a,*], Arnout van de Rijt [b]

[a] Annenberg School & School of Engineering, University of Pennsylvania, Rm. 306, 3620 Walnut St., Philadelphia, PA 19104, USA
[b] Department of Sociology & Institute for Advanced Computational Science, SUNY Stony Brook, Stony Brook, NY 11794, USA

ARTICLE INFO

Article history:
Available online 14 May 2014

Keywords:
Networks
Homophily
Selection
Social support
Fitness
Weight loss
Internet

ABSTRACT

A growing number of online health communities offer individuals the opportunity to receive information, advice, and support from peers. Recent studies have demonstrated that these new online contacts can be important informational resources, and can even exert significant influence on individuals' behavior in various contexts. However little is known about how people select their health contacts in these virtual domains. This is because selection preferences in peer networks are notoriously difficult to detect. In existing networks, unobserved pressures on tie formation – such as common organizational memberships, introductions to friends of friends, or limitations on accessibility – may mistakenly be interpreted as individual preferences for interacting/not interacting with others. We address these issues by adopting a social media approach to studying network formation. We study social selection using an *in vivo* study within an online exercise program, in which anonymous participants have equal opportunities for initiating relationships with other program members. This design allows us to identify individuals' preferences for health contacts, and to evaluate what these preferences imply for members' access to new kinds of health information, and for the kinds of social influences to which they are exposed. The study was conducted within a goal-oriented fitness competition, in which participation was greatest among a small core of active individuals. Our results show that the active participants displayed indifference to the fitness and exercise profiles of others, disregarding information about others' fitness levels, exercise preferences, and workout experiences, instead selecting partners almost entirely on the basis of similarities on gender, age, and BMI. Interestingly, the findings suggest that rather than expanding and diversifying their sources of health information, participants' choices limited the value of their online resources by selecting contacts based on characteristics that are common sources of homophily in offline relationships. In light of our findings, we discuss design principles that may be useful for organizations and policy makers trying to improve the value of participants' social capital within online health programs.

1. Introduction

Over the last decade, the Internet has become an increasingly important domain for health (Fogel et al., 2002; Thackeray et al., 2008; Chou et al., 2009; Hawn, 2009; McNab, 2009; Pampel et al., 2010; Salathe and Khandelwal, 2011). Recent surveys of Internet use for health estimate that 23% of US patients living with chronic illnesses, such as high blood pressure, diabetes, heart conditions, or cancer, use peer-to-peer online resources to help support their medical treatment and discovery processes (Fox, 2011). Even more striking, among populations with chronic diseases who are seeking "practical advice for coping with day-to-day health situations," patients were overall more likely to seek out informal sources of peer-to-peer assistance than consult with medical professionals (Fox, 2011). As this trend increases, social scientists interested in the social dimensions of health are increasingly concerned with characterizing the online social networks that people use. In particular, recent research has begun to explore the question of how online social networks influence the spread of health information and behavior change (White and Dorman, 2001; Japuntich et al., 2006; Hawn, 2009; Centola, 2010, 2011). Centola (2010, 2011) uses controlled online experiments to demonstrate the effects of both network structure and homophily in promoting the contagious spread of health behaviors. However, relatively little is known about how online health communities form, and what kinds of networks people "create" in these often anonymous environments (Wellman and Hampton, 1999; Wellman, 2001). Given the variety of online health contexts for information exchange and influence (Fox, 2011), we focus

* Corresponding Author.
E-mail address: dcentola@asc.upenn.edu (D. Centola).

http://dx.doi.org/10.1016/j.socscimed.2014.05.019
0277-9536/

our study on the increasingly popular domain of online fitness programs, which are designed to promote exposure to health information and increased fitness through peer to peer interaction (Centola, 2013).

The popular bromide that people select ties "homophilously" – i.e., based on preferences for others with similar characteristics – was formally introduced over a half a century ago by Lazarsfeld and Merton (1954). The goal of their study was to determine why strong correlations were regularly observed between people with specific demographic characteristics and those exhibiting certain beliefs, attitudes and behaviors. Their explanatory strategy was first to show that people with similar demographic traits selectively formed ties to one another, and then to show that people who were socially connected influenced each other's beliefs. However, while they found that friends influence friends, support for homophily in tie formation (henceforth "choice homophily") was variable, occurring in some situations, but not in others. As Lazarsfeld and Merton put it, "[T]he problem of selection [is] not adequately formulated by the familiar and egregiously misleading question: When it comes to close friendships, do birds of a feather actually flock together? Rather it is a more complex problem of determining the degree to which such selectivity varies for different kinds of social attributes, how it varies within different kinds of social structure, and how such selective patterns come about." (Lazarsfeld and Merton, 1954:18).

A large literature has since emerged on homophily in social relations. As the terminology has evolved, the term "homophily" has now come to refer to the observed population-level regularity that people within a community tend to be socially connected to others who are more similar to themselves than would be expected by random chance (Coleman, 1958). Researchers in this tradition have identified several, very different, mechanisms that can generate this regularity. The most obvious mechanism, initially identified by Lazarsfeld and Merton (1954), is "choice homophily": People preferentially make ties to others who are similar to themselves. However, inferring individual choice homophily from population level homophily risks running afoul of the ecological fallacy since choice homophily can be completely absent at the individual level even when populations exhibit high levels of observed homophily. This disjuncture between individual behavior and collective outcome is due to the variety of other mechanisms that can produce similar population-level patterns. For instance, a second mechanism, which has recently been widely discussed in the literature on networks and health is the process of social influence (McPherson and Smith-Lovin, 1987; Popielarz and McPherson, 1995; Christakis and Fowler, 2007). While homophily on some traits, like race and gender, cannot emerge through social influence, interpersonal correlations on other health characteristics, such as obesity, heart disease, or smoking, can be linked to social influences between contacts (Christakis and Fowler, 2007). Recent research has emphasized these mechanisms as competing explanations for patterns of observed homophily on obesity, giving rise to a dichotomization of the literature on homophily and health into the competing positions of "social influence" vs. "choice homophily". However, the scope of the problem of the origins of interpersonal correlations on health characteristics is actually much broader. Other explanatory mechanisms, which operate at the level of social structure rather than at the level of the individual or the dyad, are equally important factors in the emergence of correlations in social networks.

For instance, organizational and institutional sorting processes at schools and workplaces typically determine the set of potential social contacts that an individual is exposed to within a given context (Feld, 1982; McPherson et al., 2001; Moody, 2001; Ruef et al., 2003; Bertrand and Mullainathan, 2004). These structures often implicitly "preselect" individuals into homophilous groups (by race, class, gender, educational background, and so forth), thereby eliminating opportunities for heterophilous tie formation (Blau, 1977; Blau and Schwartz, 1984; McPherson and Smith-Lovin, 1987). These social processes can force homophilous tie formation even when the members of a population lack any particular preference for homophilous ties (McPherson and Smith-Lovin, 1987). Similarly, homophily can also emerge from the process of friends introducing friends to one another, or "triadic closure" in social networks (Kossinets and Watts, 2009). For instance, if a pair of friends, A and B, are homophilous, and B also has a friend C with whom she is similarly homophilous, then A may become friends with C by virtue of B's introduction. A homophilous tie between A and C can thus form by virtue of social structure, without A having any particular interest in "finding" someone similar to herself. More importantly, homophily can emerge in social networks even when individuals consciously prefer heterophily. In friendship networks, competitive preferences to form ties with the most healthy, most physically attractive or most successful individuals can create patterns of observed homophily via the endogenous exclusion of low-health or low-attractiveness members of the population, who are then forced to form ties with one another (Ali et al., 2012). Crosnoe et al. (2008) show that this mechanism of social exclusion can generate explicit patterns of homophily on obesity. More generally, across a broad array of social characteristics in which actors have "aspirational" preferences to form ties to "desirable" alters, patterns of systematic exclusion of the less desirable individuals can lead to the false appearance of choice homophily in domains such as health (Ali et al., 2012), online dating (Hitsch et al., 2010), marriage markets (Mare, 1991; Kalmijn, 1994), scientific collaboration (Dahlander and McFarland, 2013), and residential segregation (Van de Rijt et al., 2009). Finally, selection on an unobserved trait may be mistaken for a selective preference for a correlated trait that is observed (Yamaguchi and Kandel, 1993; Kalmijn and Vermunt, 2007). For example, as fitness is related to age, a tendency for individuals to choose ties to others of a similar fitness observed in a study that measures subject fitness but not age may in actuality represent an unobserved tendency for subjects to select on the basis of age. Consequently, in evaluating the implications of social networks for health communications, observed patterns of homophily on health characteristics do not provide clear evidence for individuals' selective preferences for health contacts.

These issues become particularly salient in contexts where the selections that people make are typically sought after as informational or motivational resources. Within online fitness programs, the selection of health contacts explicitly serves the goal of providing a reference point for achievement within the program, and establishing a standard against which to evaluate success. Our goal is to determine how people select ties in these contexts, and thereby to understand how social selection both frames the scope of participants' exposure to novel and productive health information, and provides a motivational frame for future health. In particular, we are interested in whether participants select online health contacts who have levels of fitness and "status" on health characteristics that suggest aspirational goals in establishing ties, or whether ties are formed primarily to contacts with similar levels of fitness as themselves. This difference between "aspirational" tie formation, vs. "homophilous" tie formation is important for understanding the ultimate impact of online health networks on participants' health. One of the primary incentives for forming contacts within an online health program is because they provide a means for discovering new ways to lead a healthier lifestyle by providing exposure to new health information. Another reason that participants form ties is because they are seeking