Social Networks 33 (2011) 211–218

Contents lists available at ScienceDirect

Social Networks

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

The "unfriending" problem: The consequences of homophily in friendship retention for causal estimates of social influence☆

Hans Noel[a,*], Brendan Nyhan[b]

[a] Government, Georgetown University, United States
[b] Robert Wood Johnson Scholar in Health Policy Research, University of Michigan, United States

ARTICLE INFO

Keywords:
Contagion
Social influence
Peer influence
Peer effects
Homophily
Friendship retention
Unfriending
Friendship attrition

ABSTRACT

An increasing number of scholars are using longitudinal social network data to try to obtain estimates of peer or social influence effects. These data may provide additional statistical leverage, but they can introduce new inferential problems. In particular, while the confounding effects of homophily in friendship formation are widely appreciated, homophily in friendship retention may also confound causal estimates of social influence in longitudinal network data. We provide evidence for this claim in a Monte Carlo analysis of the statistical model used by Christakis, Fowler, and their colleagues in numerous articles estimating "contagion" effects in social networks. Our results indicate that homophily in friendship retention induces significant upward bias and decreased coverage levels in the Christakis and Fowler model if there is non-negligible friendship attrition over time.

1. Introduction

Until recently, scholars have given relatively little attention to the influence of personal relationships on human behavior, instead studying people largely as atomistic individuals ripped from the social context in which they live. Thankfully, this impoverished approach has started to give way to an interdisciplinary movement seeking to understand the influence of social networks in domains ranging from health to politics. Results from cases in which peers were randomly or quasi-randomly assigned such as college roommates have provided credible evidence of such effects (e.g., Sacerdote, 2001; Zimmerman, 2003; Carrell et al., 2011; for a review, see Kremer and Levy, 2008).[1]

However, when random assignment of peers is not feasible, researchers must use observational data, which creates serious inferential problems (Manski, 1993). In particular, peers may behave similarly as a result of "correlated effects" such as common environmental shocks or shared characteristics rather than social influence. Given the likelihood that peers will be similar on a range of characteristics due to homophily (McPherson et al., 2001), distinguishing between homophily and peer effects has proven to be a very difficult challenge.[2]

Many scholars have therefore turned to use longitudinal network data to try to separate homophily from social influence effects. In principle, observing dyads over multiple periods seems as though it could help separate homophily in tie formation from subsequent peer influence. However, homophily may also affect whether social ties are *maintained* over time, confounding estimates of peer influence effects. We call this the "unfriending" problem in honor of the Facebook practice of removing a person from one's list of friends on the online social network site.

We illustrate the potential inferential consequences of this problem below in an analysis of the statistical model used to estimate "contagion" effects in a series of widely publicized studies by Christakis, Fowler, and their colleagues (hereafter CF). Our Monte Carlo simulations, which are adapted from those of CF, indicate that their model's estimates of social influence effects are unbiased and have accurate coverage levels when homophily in friendship retention is not present. However, when non-negligible attrition is present, estimates from the CF model show substantial upward bias and decreased coverage levels as homophily in friendship retention increases. In short, the "unfriending" problem can create spurious evidence of social influence when none exists.

☆ We thank Nicholas Christakis and James Fowler for generously sharing the simulation code adapted in this paper and for helpful comments and clarifications. We are also grateful to Megan Andrew, Jason Fletcher, Rob Franzese, Michael Heaney, Jonathan Ladd, Russell Lyons, Aya Kachi, Walter Mebane, Jacob Montgomery, David Nickerson, Edward Norton, Phil Paolino, Fabio Rojas, Betsy Sinclair and two anonymous reviewers for helpful suggestions and feedback.

* Corresponding author. Tel.: +1 734 615 4258/+1 919 452 6451.
E-mail addresses: hcn4@georgetown.edu (H. Noel), bnyhan@umich.edu (B. Nyhan).

[1] Related experimental studies by Nickerson (2008) and Fowler and Christakis (2010) provide suggestive evidence that such effects can extend two or more degrees, though the applicability of their results to non-experimental contexts is unclear.

[2] For a review of the literature, see Soetevent (2006).

0378-8733/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.socnet.2011.05.003

2. Leveraging dynamic networks: a solution?

The CF studies, which use longitudinal social network data from the Framingham Heart Study (FHS) and National Longitudinal Study of Adolescent Health (Add Health), make strong claims about the effects of one's friends[3] on a wide range of dependent variables: obesity (Christakis and Fowler, 2007; Fowler and Christakis, 2008b), smoking (Christakis and Fowler, 2008), happiness (Fowler and Christakis, 2008a), loneliness (Cacioppo et al., 2009), depression (Rosenquist et al., 2010b), alcohol consumption (Rosenquist et al., 2010a), sleep loss (Mednick et al., 2010), and divorce (McDermott et al., N.d.). Each CF paper uses the same approach, estimating versions of the following model for ego i and alter j observed at times t_0 and t_1:

$$Y_{i,t_1} = f(Y_{i,t_0}, Y_{j,t_0}, Y_{j,t_1}, \text{controls}) \quad (1)$$

These models are typically estimated using generalized estimating equations (GEE) with an independent correlation structure to account for repeated observations of the same ego (specifically, those who name or are named by more than one friend) and dyad (those who name each other and are thus included twice, one each as the ego and once as the alter). The functional form of the model varies depending on the distribution of the dependent variable (logistic regression if the dependent variable is binary; linear regression if it is continuous or quasi-continuous).[4]

CF argue that this specification controls for initial homophily (i.e., the likely similarity between Y_{i,t_0} and Y_{j,t_0}), allowing us to identify the causal effect of *changes* in the alter's trait from t_0 to t_1 by estimating the effect of Y_{j,t_1} controlling for Y_{j,t_0}. In Christakis and Fowler (2007), they write that including alters' lagged obesity as a covariate "controlled for homophily" (373). In later work, the language is somewhat more hedged—for instance, they write in Christakis and Fowler (2008, 2251) that a lagged measure of alter smoking "*helped* to account for homophily" (our emphasis)—but the suggestion that the coefficient for Y_{j,t_1} is a causal estimate of peer effects remains. Christakis and Fowler (2009) expands on these claims, stating that observed clustering at up to three degrees of separation reflects "Three Degrees of Influence" for happiness (51), obesity (108), and smoking (116) and asserting that we "now know that obesity is contagious" (111).

Cohen-Cole and Fletcher (2008a,b) and Halliday and Kwak (2009) question whether CF's model adequately controls for homophily, which has been shown to be significant for weight status (Trogdon et al., 2008; Halliday and Kwak, 2009; Valente et al., 2009),[5] and suggest that their model may generate spurious inferences (see also Shalizi and Thomas, 2011; Lyons, 2011; Ellen, 2009).[6] In response, CF describe Monte Carlo simulation results "documenting that homophily (ranging from no homophily to complete homophily) does not result in bias in the estimates of induction in this model specification" (Fowler and Christakis, 2008b, 1404).

CF's Monte Carlo results, which are presented in Fowler and Christakis (2008b) and in a very similar form in Fowler et al. (2011), are derived from a stylized model in which a population of individuals with a randomly chosen value on some characteristic of interest form friendships and then influence each other or not (we discuss the procedure in more detail below). CF find that estimates of this influence coefficient are approximately unbiased across varying levels of homophily when the true peer effect is 0 and have a slight downward bias when the true peer effect is 0.1. On this basis, they conclude that "This simulation evidence suggests that the [Cohen-Cole and Fletcher] assertion that homophily causes us to overestimate the size of the induction effect is false." However, as we discuss below, their simulation does not incorporate friendship *attrition* and thus fails to fully account for the effects of homophily.

3. The "unfriending" problem in longitudinal data

Due to the prevalence of cross-sectional data and interest in fixed characteristics such as race and sex, scholars of social networks have tended to think about homophily in relatively static terms and to analyze it as a propensity to form ties with others who share similar characteristics. However, social networks are actually the result of a dynamic process of friendship *formation* and *dissolution*.

As a result, while relatively few longitudinal network studies have been conducted, most report substantial levels of friendship dissolution between survey waves. For instance, Mollenhorst (2009) finds that about half of adult friendships were replaced over the seven years between the two waves of his survey. For the adolescents in Add Health, Moody (N.d.) found approximately half of the friends named by respondents during in-school interviews were named again during in-home interviews 6–8 months later. Studies of social networks among younger children have found rates of attrition that are even higher still (Hallinan and Williams, 1987; Cairns and Cairns, 1995). A partial exception is the FHS data used by CF, which was conducted in a relatively stable community. O'Malley and Christakis (N.d.) report that 82% of friendships were maintained between waves in FHS, which could be a result of asking for "close friends" who could help the researchers contact the participant in the future.[7]

When friendship attrition is present, homophily is likely to be a factor. Just as people who are similar are more likely to be friends, friends who are less similar are more likely to *stop* being friends. Most of us have had friends from whom we have grown apart in this way. As we have less in common with those people, we stop spending time with them and eventually fall out of touch. In some cases, one person may deliberately end the relationship as a result of differences in political views, alcohol consumption, or other behaviors or characteristics.

Numerous examples of homophily in tie dissolution have been documented in the sociology literature (see Burt, 2000; McPherson et al., 2001 for reviews). One well-known example is a two-wave study of adolescent friendships by Kandel (1978). She describes homophily in friendship retention based on both initial characteristics and subsequent behavior (430):

> At time 1, prior to any subsequent change, pairs that will remain stable over time are much more similar in their behaviors and values [marijuana use, educational goals, political views, and delinquency] than the subsequently unstable pairs... At time 2, homophily among former friends is lower than among new friendship pairs or stable pairs.

[3] The studies typically also estimate social influence effects among family members; we do not consider the validity of those estimates here.

[4] For other approaches to obtaining causal estimates of peer effects in longitudinal or repeatedly sampled network data, see Anagnostopoulos et al. (2008), Bramoullé et al. (2009), Aral et al. (2009), and Lazer et al. (2010).

[5] de la Haye et al. (2010) finds homophily in obesity-related behaviors as well. For further explorations of possible social transmission of obesity or weight status, see Anderson (N.d.); Barnes et al. (N.d.); Brown et al. (N.d.); McFerran et al. (2010a,b); Campbell and Mohr (forthcoming).

[6] There are other concerns with this statistical model such as possible simultaneity bias and environmental confounding that we do not discuss here.

[7] CF report that they treated friendship ties as maintained when a friend as named at t_1 and t_3 but tie information was missing at t_2 (personal communication). Under this definition, 96% of friendship ties were maintained between waves. However, since missing tie information was often the result of missing an exam altogether, friendship retention in their statistical analyses is likely to be lower.