Published in final edited form as:

Stat Med. 2013 February 20; 32(4): 591–599. doi:10.1002/sim.5653.

Inference for influence over multiple degrees of separation on a social network

Tyler J. VanderWeele

Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston Boston, Massachusetts (Tyler J. VanderWeele)

Abstract

States and behaviors of different individuals are expected to be correlated across a social network. Christakis and Fowler have proposed a 'three degrees of influence rule' to characterize the extent of such dependence. In this paper we discuss three distinct interpretations of such a rule, one involving only associations (which is the interpretation for which Christakis and Fowler give evidence), one involving actual causation, generally referred to as contagion or social influence, and one involving direct effects. We discuss analytic procedures appropriate for assessing evidence for each possible interpretation and the increasingly difficult methodological challenges present in each interpretation.

Keywords

Contagion; direct effects; homophily; sensitivity analysis; social influence

Introduction

In their book, *Connected* [1], and in the current paper under discussion [2], Christakis and Fowler make a claim that social influence spreads in general to three degrees of separation across a social network for a wide range of behaviors and states. In their book, the claim often reads as though the influence to three degrees of separation is to be interpreted causally. For example, on p. 27–28 of their book, Christakis and Fowler write: "Our own research has shown that the spread of influence in social networks obeys what we call the Three Degrees of Influence Rule. Everything we do or say tends to ripple through our network, having an impact on our friends (one degree), our friends' friends (two degrees), and even our friends' friends' friends (three degrees)." In the paper under discussion [2], Christakis and Fowler seem qualify this claim and interpretation. In this commentary, we consider three possible, and increasingly stronger, interpretations of this general claim and discuss some of the methodological issues that arise in attempting to provide evidence for each interpretation.

Three Degrees of Influence as Association

Christakis and Fowler note, in the paper under discussion [2], that their claim that influence extends to three degrees of separation on a social network was not necessarily intended to be indicative of contagion or actual social influence over three degrees of separation. The states of two persons may be associated, not simply because of social influence, but also because of shared environmental factors, or because of homophily - the phenomenon whereby persons with shared characteristics are more likely to become friends with one another [2]. All three of these may give rise to associations across a social network. As they note, the evidence they present in their papers for the 'three degrees of influence rule' is simply one of clustering or statistical association. They claim that behaviors and states of individuals separated by up to three degree of separation still manifest correlation, but they do not claim that this is necessarily actual social influence. It may also be environmental confounding or homophily. "Influence" here then must almost be understood in a predictive manner e.g. even if persons i and j are separated by three contacts on the social network, knowing person j 's state would *influence my prediction* of person i 's state. This interpretation of "influence" is one merely of association (though this is not how the word "influence" would ordinarily be interpreted). They do give evidence for actual causal social influence at one degree of separation [2–4], but their claim regarding three degrees of separation is only associational.

Under this very broad interpretation of "influence" simply, we may then be skeptical that the association persists to only three degrees of separation. We may expect dependence everywhere in the social network. And Christakis and Fowler do concede that with a sufficiently large sample size, we may indeed be able to detect association across additional degrees. However, they point out that beyond three degrees, the extent of the association is, in most of the states and behaviors they consider, sufficiently small as to be practically irrelevant. They thus grant that the 'three degrees of influence rule' in this associational sense, is a relative rule rather than absolute. In most cases, after three degrees of separation, the associations are generally sufficiently small to not be of substantive interest. This first interpretation then seems relatively uncontroversial.

Three Degrees of Influence as Causation

A stronger interpretation of the 'three degrees of influence rule' is that actual social influence extends to three degrees of separation. That is, if we were to change the state or behavior of various persons, then this intervention would also on average influence the friends' friends' friends of those upon whom we intervene. This would be a causal interpretation of the 'three degrees of influence rule.' This interpretation is not one that Christakis and Fowler explicitly defend in their paper [2]. They make a strong argument that the evidence for social influence at one degree of separation is substantial, using longitudinal data and directional ties. However, they limit their claims about the 'three degrees of influence rule' to that of clustering, association.

What would be required to generate evidence for a causal interpretation of the 'three degrees of influence rule'? In their longitudinal analyses that support social influence for one degree of separation, Christakis and Fowler [3, 4] regress the ego's current state at time $t+1$, $Y_{i,t+1}$,

on the alter's state at time $t + 1$, $Y_{j,t+1}$, and at time t , $Y_{j,t}$, the ego's state at time t , $Y_{i,t}$, and of the covariates for the ego, X_i . They restrict the analysis to relationships of a particular type (e.g. ego-nominated friendship, alter-nominated friendship, mutual friend, spouse, etc.); if R_{ijt} is an indicator that i and j have a tie of the type under consideration at time t then the regression is restricted to pairs such that $R_{ijt} = 1$. For a binary outcome, for example, logistic regression is used:

$$\log \left(\frac{P(Y_{i,t+1}=1|Y_{j,t+1}, Y_{j,t}, Y_{i,t}, X_i, R_{ijt}=1)}{P(Y_{i,t+1}=0|Y_{j,t+1}, Y_{j,t}, Y_{i,t}, X_i, R_{ijt}=1)} \right) = \alpha + \beta_1 Y_{i,t} + \beta_2 Y_{j,t+1} + \beta_3 Y_{j,t} + \beta_4' X_i \quad (1)$$

In this model, Christakis and Fowler [3, 4] interpret the estimate of β_2 as their "contagion effect" or causal estimate of social influence. Generalized estimating equations is used to fit the model and obtain standard errors, clustering on the ego [2–4]. Although the modeling approach has come under some criticism due to issues of model inconsistency and the validity of standard errors [5], it does, at least in certain contexts (e.g. ego-nominated and alternominated friendships), still give a valid test of the null hypothesis of no contagion [6], provided control has been made for both environmental confounding and homophily, which are strong assumptions.

We could adapt this approach to attempt to assess social influence over multiple degrees of separation. Instead of letting $R_{ijt} = 1$ denote the presence, for example, of an ego-nominated friendship tie (at one degree of separation) between persons i and j at time t , we could let

$R_{ijt}^{(3)} = 1$ if person i names some person k_1 as a friend, k_1 names k_2 as a friend, and k_2 names j as a friend, so that i and j are separated by three ties; and we would let $R_{ijt} = 0$ otherwise. We

could likewise define $R_{ijt}^{(2)}$ as separation by two ties or $R_{ijt}^{(4)}$ as separation by four ties, etc. In the Framingham data each person names only one friend so we could not run a similar analysis with mutual ties separated by three degrees. In datasets in which persons are

allowed to name more than one friend, in defining the indicator $R_{ijt}^{(3)}$, one would want to restrict the analysis to pairs in which i names k_1 , k_1 names k_2 , and k_2 names j , without their being a tie between i and k_2 or between k_1 and j (to avoid shorter paths), and without there being other paths between i and j shorter than length 3. These subtleties do not arise in the Framingham data.

If, for example, we wanted to assess influence over two degrees of separation, we could proceed in very much the same way as in the original longitudinal analyses of Christakis and Fowler [3, 4], now running the regression:

$$\log \left(\frac{P(Y_{i,t+1}=1|Y_{j,t+1}, Y_{j,t}, Y_{i,t}, X_i, R_{ijt}^{(2)}=1)}{P(Y_{i,t+1}=0|Y_{j,t+1}, Y_{j,t}, Y_{i,t}, X_i, R_{ijt}^{(2)}=1)} \right) = \alpha + \beta_1 Y_{i,t} + \beta_2 Y_{j,t+1} + \beta_3 Y_{j,t} + \beta_4' X_i \quad (2)$$

only for those separated by two degrees. We could use generalized estimating equation, again clustering on the ego, to fit the model and then use β_2 to test whether there is social influence (interpreted now causally) over two degrees of separation. We could run a similar

analysis with $R_{ijt}^{(2)}=1$ replaced by $R_{ijt}^{(3)}=1$ to examine influence over three degrees of separation.

Of course, as in the original analyses with one degree of separation [3, 4], we would still have to rule out environmental confounding and homophily. Interestingly, such problems arguably become less severe as we move to a higher number of degrees of separation. We would expect less environmental confounding and less homophily to be present for persons separated by three ties. However, the actual social influence is also likely to be less, rendering environmental confounding and homophily still very real threats. Controlling for the lagged state of person j , $Y_{j,t}$, partially helps to control for homophily [3, 4]. And sensitivity analysis [7] can be used to assess the extent to which homophily or environmental confounding could influence estimates.

However, when we move to this setting of multiple degrees of separation, some additional subtleties arise. Suppose another person, person h , names person i as a friend ($h \rightarrow i$) and person j in turn names h as a friend ($j \rightarrow h$). Although the shortest ego-nominated pathway between person i and person j would still be three degrees ($i \rightarrow k_1 \rightarrow k_1 \rightarrow j$), the state of person h might nonetheless influence the state of both person i and person j and thus be a confounder for the influence j exerts on i . This then creates new difficulties for inference over multiple degrees of separation. A similar problem can of course also arise in the original analyses of Christakis and Fowler [3, 4] when examining only a single degree of separation. However, importantly, in the context of only one degree of separation, if we are just interested in testing for any social influence, then, under the null hypothesis of no contagion anywhere in the network, the problem of confounding by a third person vanishes. Under the null of no contagion, person h will not influence i or j and will thus not be a confounder. When we are interested only in influence over one degree of separation, the analyses, for the purposes of testing, are robust to such confounding by a third person in a network. We do not have this same robustness when we are interested in inference over multiple degrees of separation. Such inference is more challenging.

As in the case of the associational interpretation of the 'three degrees of influence rule', we might also think that ultimately social influence should, to a certain extent, extend throughout the social network, to all degrees. Causation is generally, though not always, transitive. That is to say, if A is a cause of B , and B is a cause of C , then in general we expect A to be a cause of C . Thus if we have evidence for social influence for a particular state or behavior over one degree of separation in a network, then, by transitivity, we might expect that such influence should extend to all degrees of separation, albeit in muted form. Such transitivity can, however, sometimes fail. It can fail, if, for example, the values of B over which a change in A exerts a change in B , do not correspond to the values of B which exert a change on C .¹ However, we probably would not expect such failures in transitivity to always occur by precisely three degrees of separation. The type of dynamic and the network

¹For example, suppose that A has two levels, (0, 1), that B has three levels, (0, 1, 2), and C has two levels, (0, 1). If a change in A will change a value of B from 0 to 1 but exert no change in B if B is 1 or 2 already, and if a change in B from 1 to 2 will change C from 0 to 1, but a change in B from 0 to 1 has no effect on C , then A will be a cause of B (a change in A can change B) and B will be a cause of C (a change in B can change C) but A is not a cause of C (no change in A can influence C).

structure would have to be very peculiar indeed for this to occur. More likely than not then, here, as in the case of the associational interpretation, the 'three degrees of influence rule', if it is to hold, would have to be interpreted in a relative sense i.e. that with more than three degrees of separation, the actual social influence, although possibly present, is sufficiently muted that it is not of substantive relevance.

Three Degrees of Influence as a Direct Effect

Finally, let us consider a third, and even stronger possible interpretation of influence of over multiple degrees of separation. Suppose i names k as a friend, and k names j . Suppose we proceeded with the analyses suggested in the previous section in model (2) and thought that j exerted influence on i , even though they were separated by two degrees. We might then ask the further question of whether j influences i only through k 's state or whether j also exerts an additional 'direct influence' over i , independent of k 's state. For example, Christakis and Fowler [2, 3] discuss norms as a possible mechanism for influence. If person j becomes obese, this may influence person i even if person k 's weight does not change if, for example, person j 's obesity changes the norms of k , and thus changes the way k interacts with i e.g. because k has become accustomed to j 's being obese, k no longer criticizes person i for overeating. This would be an effect of j 's state on i 's state, not through k 's state; it would be a 'direct effect'. Of course, whether an effect is 'direct' or not depends on what is taken as the intermediate or the 'state' of k . If the 'state' of k were defined as all attitudes and behaviors of person k , then the effect of j on i might no longer be understood as 'direct'. But if the 'state' is simply taken as obesity, as in Christakis and Fowler [3], then j may influence i not through k 's obesity status. And we could of course extend such reasoning over multiple degrees of separation. If i names k_1 as a friend, and k_1 names k_2 , and k_2 names j as a friend, so that i and j are separated by three ties, we could ask whether j influences i other than through the state of persons k_1 and k_2 . This would be an interpretation of the 'three degrees of influence rule' as one of a direct effect.

A typical way to assess direct effects is to simply include the intermediate, person k 's state, in the regression. We might control for person k 's lagged state as well, thus modifying the model in (2) to:

$$\log \left(\frac{P(Y_{i,t+1}=1|Y_{j,t+1}, Y_{j,t}, Y_{i,t}, Y_{k,t+1}, Y_{k,t}, X_i, R_{ijt}^{(2)}=1)}{P(Y_{i,t+1}=0|Y_{j,t+1}, Y_{j,t}, Y_{i,t}, Y_{k,t+1}, Y_{k,t}, X_i, R_{ijt}^{(2)}=1)} \right) = \alpha + \beta_1 Y_{i,t} + \beta_2 Y_{j,t+1} + \beta_3 Y_{j,t} + \beta_4' X_i + \beta_5 Y_{k,t} + \beta_6 Y_{k,t+1}. \quad (3)$$

We might then be tempted to take β_2 as our measure of the direct social influence of person j on i not through k . We could potentially extend this analysis further over three or more degrees of separation by e.g. including both the current and lagged state of persons k_1 and k_2 on the path from j to i if i and j are separated by three degrees.

Unfortunately, such approaches are subject to an even wider array of methodological challenges than those encountered in the original analyses of Christakis and Fowler [3, 4] over one degree of separation, or those considered in the previous section for overall influence over multiple degrees of separation. The challenges are analogous to those that arise in the causal mediation analysis literature on direct and indirect effects [8–10].

Effectively here, we are taking $Y_{j,t+1}$ as the exposure, $Y_{k,t+1}$ as the intermediate, and $Y_{i,t+1}$ as the outcome. For such analyses, control needs to be made not only for 'exposure-outcome' confounding but also 'mediator-outcome' confounding. We thus need to worry about (i) environmental confounding, (ii) homophily, and (iii) confounding by a third person, not only for the relationship between persons j and i , but also between persons k and i . The use of sensitivity analysis and the interpretation of sensitivity analysis parameters also becomes more complex in the case of assessing direct effects [10–12]. Further complexities can arise if the 'exposure' and 'mediator' interact e.g. if person j 's state and person k 's interact, through reinforcement, in their influence on person i [9].

And yet further difficulties also emerge in the context of social networks. The original analyses of Christakis and Fowler [3, 4] came under critique by Lyons [5] because, under the alternative hypothesis of social influence, the models employed are in fact incompatible because of the use of contemporaneous ego-alter data so that certain variables appear as an outcome in one regression and as an independent variable in another regression in the same time period. Fortunately, in spite of these problems of model incompatibility, the analyses of Christakis and Fowler [3, 4] will still often constitute valid tests of the null of no social influence [6]. The original analyses of Christakis and Fowler [3, 4] allow one to do testing but not estimation. Estimation would require lagging the alter's state by an additional period [5, 6] so that e.g. the variables $(Y_{j,t}, Y_{j,t-1})$ instead of $(Y_{j,t+1}, Y_{j,t})$ would be included in the regression model (1). In principle, this could feasibly be done, both in examining influence over one degree of separation and also for overall influence over multiple degrees of separation as considered in the previous section. This issue of lagging and model compatibility is, however, compounded when we are interested in direct influence over multiple degrees of separation (e.g. of j 's state on i 's state, not through k 's state). If we lag j 's state by an additional period, but not k 's state, we may end up with similar difficulties of model incompatibility because of using contemporaneous data between k and i . If we lag j 's state and i 's by an additional period, we may end up with issues of model incompatibility because of using contemporaneous data between j and k . We could lag j 's state by two additional periods and k 's by only one, but it is then no longer entirely clear that we are capturing the sort of direct influence that is ultimately intended. Such issues would require further attention before such an approach is used in practice. In related work [13], we have also considered the effect of an intervention on one person on another's outcome and the extent to which this effect is or is not mediated by the outcome of the person who received the intervention. In any case, the methodological issues that would arise if one were to attempt to assess the 'three degree of influence rule' interpreted as 'direct influence' are challenging and merit further consideration.

Conclusion

We have considered three possible interpretations of influence over multiple degrees of separation on a social network: an associational interpretation, a causal interpretation, and an interpretation as direct influence. The associational interpretation, which is all that Christakis and Fowler [2] discuss in their defense of the 'three degrees of influence rule' seems relatively uncontroversial, if interpreted in the relative sense of the associations generally being sufficiently small to not be of substantive interest beyond three degrees of

separation. The causal interpretation makes a stronger claim that there is actual social influence (not just clustering and association) over multiple degrees of separation on the network. This stronger interpretation would also need to be interpreted in a relative sense of the effects simply being too weak beyond a certain point to be of interest. However, progress could potentially be made for empirically testing claims of actual social influence over multiple degrees of separation using an analytic approach not dissimilar to that of the original Christakis and Fowler [3, 4] analyses of smoking and obesity. Unfortunately, however, such analyses are not as robust to potential confounding by a third person as those aiming only to test for social influence at one degree of separation. A yet stronger interpretation of influence over multiple degrees of separation would be one of direct influence, not through the state of the persons constituting intermediate social ties. Analyses of such 'direct effects' are subject to all of the same difficulties as the analyses for one degree of separation, but these are compounded by the use of data on the states of three or more persons simultaneously. Current evidence for the associational interpretation of influence over multiple degrees of separation seems quite strong [1–4]. Further work could examine empirical evidence for influence over multiple degrees of separation in the causal or 'direct influence' sense.

Acknowledgments

The author thanks the editor and reviewers for helpful comments. This research was supported by Grant # R01-ES01787.

References

1. Christakis, NA.; Fowler, JH. *Connected: The surprising power of our social networks and how they shape our lives*. Little Brown, New York: 2009.
2. Christakis NA, Fowler JH. Social contagion theory: examining dynamic social networks and human behavior. *Statistics in Medicine*, to appear. 2011
3. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*. 2007; 357:370–379. [PubMed: 17652652]
4. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. *New England Journal of Medicine*. 2008; 358:2249–2258. [PubMed: 18499567]
5. Lyons R. The spread of evidence-poor medicine via flawed social-network analyses. *Statistics, Politics and Policy*. 2011; 2(1):1–26. Article 2.
6. VanderWeele TJ, Ogburn EL, Tchetgen Tchetgen EJ. Why and when “flawed” social network analyses still yield valid tests of no contagion. Submitted to *Statistics, Politics and Policy*.
7. VanderWeele TJ. Sensitivity analysis for contagion effects in social networks. *Sociological Methods and Research*. 2011; 40:240–255.
8. Pearl, J. *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence*. San Francisco: Morgan Kaufmann; 2001. Direct and indirect effects; p. 411–420.
9. VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *American Journal of Epidemiology*. 2010; 172:1339–1348. [PubMed: 21036955]
10. VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology*. 2010; 21:540–551. [PubMed: 20479643]
11. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychological Methods*. 2010; 15:309–334. [PubMed: 20954780]
12. Bullock, JG.; Shang, EH. Mediation analysis is harder than It looks. In: Druckman, James N.; Green, Donald P.; Kuklinski, James H.; Lupia, Arthur, editors. *Cambridge Handbook of Experimental Political Science*. New York: Cambridge University Press; 2011.

13. VanderWeele TJ, Tchetgen Tchetgen EJ, Halloran ME. Components of the indirect effect in vaccine trials: identification of contagion and infectiousness effects. *Epidemiology*. 2012; 23:751–761. [PubMed: 22828661]