A Primer on Multilevel Mediation Models for Egocentric Social Network Data

Hyunjin Song

To cite this article: Hyunjin Song (2018) A Primer on Multilevel Mediation Models for Egocentric Social Network Data, Communication Methods and Measures, 12:1, 1-24, DOI: 10.1080/19312458.2017.1416343

To link to this article: https://doi.org/10.1080/19312458.2017.1416343

© 2018 Hyunjin Song. Published with license by Taylor & Francis.

Published online: 29 Dec 2017.

Submit your article to this journal

Article views: 391

View related articles

View Crossmark data
TEACHER’S CORNER

A Primer on Multilevel Mediation Models for Egocentric Social Network Data

Hyunjin Song

Department of Communication, University of Vienna, Vienna, Austria

ABSTRACT

The substantial increase in nested data such as egocentric network data pose unique challenges to researchers, including how to adequately model the mechanisms and processes in which hypothesized effects operates within the complex interdependencies among social actors. While there is a growing interest in identifying the causal mechanisms and their contingencies using such nested structured data, the direct application of principles of (moderated) mediation analysis within traditional multilevel modeling framework is ambiguous and controversial. In this article, I provide a tutorial illustrating an approach to assessing mediation hypotheses based on path-analytic framework popularized by Hayes (2013). Focusing on estimation and inference of indirect effects with regard to nested data structure, I provide examples of the analysis within the context of egocentric network data, and illustrate implementation using Mplus software.

Social actors are often embedded within a group, a network, or a context, where such groupings represent some substantive relationship and dependencies (or a lack of thereof) among actors. Such data can be viewed as multilevel, therefore representing the complex interdependencies among social actors. Such dependencies give rise to the problem of violating the basic assumption of statistical independence. Ordinary linear models, by definition, cannot handle such dependencies; adequately addressing such dependencies requires multilevel linear models (MLM: Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). MLM is now a popular area of interest across social scientific disciplines, addressing various types of social regularities such as voters nested within counties, longitudinal measurements within subjects, TV viewers within media markets, etc.

There is a parallel, yet growing interest in identifying the causal mechanisms and their contingencies in various fields of social science. Answering such questions, so far, has typically been based on formal tests of mediation (the causality underlying phenomena in question) as well as moderation (its boundary conditions). The analysis of mediation, according to Preacher (2011, p. 691), “is becoming a very useful and popular practice in social science.” In line with this growing interest, scholars are increasingly paying more attention to the possibility of testing the mediation hypothesis with complex, nested data structures—such as in traditional multilevel linear models (MLM) or with egocentric social network data. Yet nested structures of such data pose a unique challenge to a formal test of causal mechanisms. Traditionally available methods of accessing statistical mediation often rely on ordinary least square (OLS) regression, where data are collected at only a single level, and therefore it assumes independence of observations. For that matter, a direct application of the traditional “single-level” mediation analysis to the nested structure of data is often seen as difficult. Instead, researchers routinely rely on a series of logical inferences rather than
directly estimating the indirect effect in question (see Borgatti & Cross, 2003; or Christakis & Fowler, 2007), the method of which is popularized with the term *the causal steps approach* (Baron & Kenny, 1986), as described in Panel A of Figure 1 below. However, as argued by Hayes (2009, 2013), the causal step approach to testing mediation does not directly quantify the indirect effect in question. Instead, it is based on a series of logical inferences, which depend on the examination of each constituent path of a hypothesized indirect effect. Additionally, the causal steps approach could be problematic in that it requires researchers to find a significant total effect of $X$ on $Y$ (“$c$” path) to even begin the mediation analysis. As suggested elsewhere (e.g., Hayes, 2009, 2013; Preacher, Rucker, & Hayes, 2007), it is possible for an indirect effect to be significantly different from zero even if there is no evidence of the significant total effect of $X$ on $Y$. Furthermore, neither the changes of significance level nor the reduced magnitude of the *direct* effect of $X$ on $Y$, when compared to the *total* effect of $X$ on $Y$, meet the “statistical” requirement to establish evidence of a mediation effect. Therefore, when a better alternative is available, such as bootstrapping or Monte Carlo methods, the causal steps approach should be avoided, as to the case of OLS-based mediation analysis (Hayes, 2013; Preacher et al., 2007).

**Figure 1.** A conceptual depiction of the causal steps approach and the single-level mediation model.

**A mediation analysis using multilevel linear models**

In this article, I attempt to provide more accessible discussion of the application of mediation analysis using MLM framework, as described in Zhang, Zyphur, and Preacher (2009) or in Bauer,
Preacher, and Gil (2006). For a motivating example, I focus on its application to egocentrically collected network data, where typical survey respondents (hereafter denoted as “ego”) provide their own perceptions and attributes of their social surroundings, focusing on their interactions with their immediate personal network members (hereafter denoted as “alters”). An analysis of personal networks (hereafter denoted as “egocentric network analysis”) has a rich tradition in communication science, sociology, and political science literature, and is becoming an increasingly popular tool for understanding the impact of one’s contextual surroundings on perceptions, attitudes, and behaviors. Also, the typical structure of egocentric network data provides one of the ideal cases for reliably estimating multilevel mediation effects since a large number of focal survey respondents serving as upper-level observations. Although I ground this work within this specific setting, the general approach advanced in the present article can be easily extended and applied to other substantive domains of interest, such as organizational communication (e.g., Feeley & Barnett, 1997; Monge & Contractor, 2003), international/intercultural communication (e.g., Kim & Barnett, 1996), contextual influences in interpersonal, or mass communication processes (e.g., Kim et al., 2013; Parks & Floyd, 1996) and to any research on networks and social media (e.g., Ellison, Steinfield, & Lampe, 2007), provided that the theoretical perspective and data structure follows a standard multilevel framework.

Whole network data vs. egocentric network data

Before discussing the multilevel approach of assessing mediation hypothesis for egocentric network data, it is worth reviewing two different types of network data and measurement techniques—“whole network” and “egocentric network” (Marsden, 2011)—and their relationships with traditional MLM framework. With slight variations among various boundary specification methods (Laumann, Marsden, & Prensky, 1983; Marsden, 2011), whole network data generally aims to capture an entire social structure and complete interdependencies among actors, where all possible social relations are surveyed among a well-defined population. In contrast, egocentric network data aims to capture immediate social environments of “discrete” individuals, where the source of interdependency mainly arises from the relations between an ego and an ego’s immediate contacts (“alters”).

Due to a more complex interdependency structure in whole network data, traditional multilevel framework (which assumes no interconnections among higher units) has been largely regarded as inapplicable or even irrelevant for the whole network cases. Many of the whole network studies often take a case-study approach (statistically, N = 1 study) due to difficulties of collecting multiple whole networks. Even with the presence of multiple whole networks, researchers often employ side-by-side comparisons (see Eveland & Hutchens, 2013, for a Multiple Regression Quadratic Assignment Procedure, or MRQAP, context) or meta-analytic techniques (e.g., see Ripley, Snijders, Boda, Vörös, & Preciado, 2014, for a Stochastic Actor Oriented Model application) to evaluate the overall parameter estimates. A recent development in statistical modeling of whole network data seeks to advance its straightforward application to the multilevel modeling of social networks (e.g., Tranmer & Lazega, 2016; Wang, Robins, & Matous, 2016). Yet the multilevel analysis of whole network data and its statistical inferences have become available only relatively recently. A full discussion of such applications and its extension—especially for the statistical mediation analysis—is therefore beyond the scope of this manuscript, for reason which I will discuss later in the discussion section. Thus, I will explicitly devote our attention to the methods of assessing multilevel mediation (and potentially multilevel moderated mediation) for egocentric network data.

Egocentric network data and traditional MLM framework

Quite contrary to a case for whole networks, a more direct application of multilevel modeling to egocentric network data is now widely available after the pioneering work of Snijders, Spreen, and Zwaagstra (1995) and others (Van Duijn, Van Busschbach, & Snijders, 1999). When the traditional name generator method (Marsden, 2011) is used for gathering egocentric network data, each focal
respondent provides information regarding their alters. Therefore, the dyadic properties between an ego and her alter are regarded as micro-level variables, and the attributes of each ego are regarded as macro-level variables in a typical name-generator approach. Also, since the typical egocentric network data largely resembles the standard person-by-variable, nested data structure for MLM framework (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012), application of traditional MLM is relatively straightforward and therefore noncontroversial (De Miguel Luken & Tranmer, 2010; Lubbers et al., 2010).

Consider an example of van Duijn et al.’s (1999) seminal application of the MLM framework to egocentric network data, wherein the ego’s frequency of contact with alters (operationalized as the visits and telephone calls) is modeled. A multilevel model for egocentric social network data could be set up for the lower-level dependent variable, $Y_{ij}$ (e.g., the change in contact frequency with alters). This lower-level dependent variable is then explained as a function of the same lower-level independent variable, $X_{ij}$ (e.g., travel time between the ego and the alter), as well as the higher-level independent variable, $Z_j$ (e.g., focal respondents’ demographics). The regression equations at the first and the second level are:

$$Y_{ij} = dY_{ij} + c_j X_{ij} + e_{Y_{ij}}$$
$$dY_j = dY + Z_{Yj} + u_{Yj}$$
$$c_j = c + Z_{cj} + u_{cj},$$

where the regression coefficients $dY_j$ and $c_j$ are allowed to vary over higher-level unit (e.g., focal respondents), and the regression coefficients $dY$, $c$, and $Z_j$ are fixed over all higher units. Readers with familiarity to the traditional MLM may notice here that the above equation depicts a simple random-intercept-random-slope model. The effect of $X$ on $Y$ ($c_j$) is specified as the function of average effect across all clusters ($c$) and some systematic variations ($Z_{cj}$), along with the random variation that is unique to each upper unit of $j$ ($u_{cj}$). Likewise, the intercept of the regression can be broken down to the grand-mean intercept across all upper units of $j$ (i.e., $dY$); systematic variation of the slope factor as a function of the upper-level explanatory factor ($Z_{Yj}$), and the random variation unique to each unit $j$ ($u_{Yj}$).

Much of the existing literature utilizing MLM in egocentric network data is based on the traditional MLM analysis described above (such as in Lubbers et al., 2010 or Bello & Rolfe, 2014). There are many advantages in applying traditional MLM to egocentric network data, although it does not come without limitations, and it is worth revisiting what MLM can and cannot do. First, MLM assumes all of the observed variables are perfectly reliable (i.e., without measurement error), which in many of the cases may be unrealistic. Second, and perhaps the greatest limitation of traditional MLM framework, is it is impossible to accommodate an upper-level dependent variable, where the conceptual independent variable lies at the lower level and the dependent variable is in the upper level (Preacher, Zyphur, & Zhang, 2010). Such a “bottom-up inference” may be of considerable interest to communication scholars, yet because the MLM framework cannot accommodate the upper level dependent variable, some researchers instead rely on the aggregation of the lower-level information to create a summary measure defined at the upper-level. Such an analytical technique may be justified by a strong theoretical argument, and by analytical needs in particular, when standard mediation framework is required (e.g., Zhu, Woo, Porter, & Brzenzinski, 2013). However, one should bear in mind that such aggregation into a higher unit (e.g., deriving a single summary measure) is a considerable loss of information (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Recent application of structural equation modeling (SEM) to multilevel analysis can overcome aforementioned limitations, and is readily available in many instances (e.g., Du Toit & Du Toit, 2008; Preacher, 2011; Preacher et al., 2010). Yet, SEM frameworks generally rely on large sample procedures (Hayes, Montoya, & Rockwood, 2017). For the multilevel SEM framework, the sample size requirement to reliably detect true effects further increase, perhaps prohibitively, where upper-level sample sizes lower than 100 tend to lead to non-convergence and inaccurate parameter estimates (see Hox & Maas, 2001). Therefore, the multilevel SEM could be a less desirable option for a researcher in certain circumstances.
A basic framework of multilevel mediation models

Before discussing multilevel mediation in detail, it is worth revisiting a traditional statistical model of assessing mediation (i.e., single-level mediation analysis). In a typical model, a dependent variable, \( Y \), is assumed to be caused by two other same-level predictor variables \( X \) and \( M \). In addition, we further assume the presumed causal agent, \( M \), mediates the influence of \( X \) on \( Y \). The conceptual model is depicted in Panel B in Figure 1.

Let \( a \), \( b \), and \( c \) be quantifications of causal effects, where \( a \) be the effect of \( X \) on \( M \), \( b \) be the effect of \( M \) on \( Y \), \( c \) be the “total effect” of \( X \) on \( Y \), and \( c' \) be the “direct effect” of \( X \) on \( Y \). Then we could express the single-level mediation model, using a standard OLS or ML method, as follows:

**Total effect model:**

\[
Y_i = d_Y + cX_i + e_Y_i
\]  

(4)

**Indirect effect model:**

\[
M_i = d_M + aX_i + e_M_i
\]  

(5)

\[
Y_i = d_Y + c'X_i + bM_i + e_Y_i
\]  

(6)

where \( d \) denotes an intercept for \( M \) (\( d_M \)) or \( Y \) (\( d_Y \)), and \( e \) denotes the error terms in the equations, and subscript \( i \) denotes each observation. The direct effect of \( X \) on \( Y \) is denoted as \( c' \), and the indirect effect of \( X \) on \( Y \) through a mediator \( M \) is formally expressed as the product of \( a \) and \( b \), or \( ab \). The total effect of \( X \) on \( Y \), \( c \), would be mathematically identical to the sum of the direct effect and the indirect effect, which equals \( c' + ab \). When all variables are measured on a single level, we could obtain the estimates of such model parameters using a standard OLS-regression or ML methods. However, applying such techniques to hierarchically nested data (such as the egocentric network) while ignoring the nested structure leads to the downward bias in the estimation of standard errors, resulting an erroneous hypothesis test and the conclusion of its significance (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012).

Let us next consider a set of standard MLM equations predicting \( Y \) from \( X \) with random intercept and random slope factors. For the sake of simplicity, consider a simple model without any of the level-2 predictor variables. In such a model, a level-1 variable, \( Y \), is assumed to be caused by another level-1 variable \( X \). We further assume the effect of \( X \) on \( Y \) may vary across upper level (level-2) unit \( j \). Let \( c \) be the quantification of causal effects of \( X \) on \( Y \), and we could express a generic form of MLM predicting \( Y \) from \( X \) as follows.

**Level-1 equation:**

\[
Y_{ij} = d_Y + c_jX_{ij} + e_{Y_{ij}}
\]  

(7)

**Level-2 equation:**

\[
d_Y = d_Y + u_{Yj}
\]  

(8)

\[
c_j = c + u_{cj}
\]  

(9)

Notice here that the level-1 equation of the general MLM (Equation 7) is identical to a “total effect” model in mediation framework (Equation 4) except the second subscript \( j \), which refers to an upper level observation. This is the mathematical expression that the invariant constraint in Equation 4 is relaxed, such that any causal effects operating within the lower level are allowed to vary across the upper level.
Following the standard practice of MLM, we then add two level-2 equations expressing the intercept and the slope of the level-1 equation, yielding Equation (8) and (9). Next, we create the upper-level variables (“between-” components) that correspond to a lower-level X and M by aggregating the respective lower-level variables, and use the group-mean centering procedure to capture a strict “within-” component of variations. As such, we treat the within- and between-component of variations as two distinctive constructs, in order to statistically distinguish those components in indirect effect estimates (Preacher, 2011; Preacher et al., 2010; Zhang et al., 2009). We could then express a deconflated multilevel mediation model, conceptually described as in Panel A of Figure 2 below, which consists of the MLM equations predicting M from X, and the MLM equations predicting Y from the joint function of X and M, as follows.

**Level-1 “M” equation:**

\[ M_{ij} = d_M + a_j X_{ij} + e_{M_{ij}} \]  \hspace{1cm} (10)
Level-2 “M” equation:
\[
d_{M_j} = d_M + a_B \tilde{X}_j + u_{M_j}
\]
(11)
\[
a_j = a_W + u_a
\]
(12)

Level-1 “Y” equation:
\[
Y_{ij} = d_{Y_i} + c'_j \tilde{X}_{ij} + b_j \tilde{M}_{ij} + e_{Y_{ij}}
\]
(13)

Level-2 “Y” equation:
\[
d_{Y_j} = d_Y + c'_b \tilde{X}_j + b_B \tilde{M}_j + u_{Y_j}
\]
(14)
\[
c'_j = c_{W'} + u_{c'_j}
\]
(15)
\[
b_j = b_W + u_{b_j}
\]
(16)

where \(\tilde{X}_{ij}\) and \(\tilde{M}_{ij}\) denotes group-mean centered \(X\) and \(M\) at the lower level (i.e., \(\tilde{X}_{ij} = X_{ij} - \bar{j}\), and \(\tilde{M}_{ij} = M_{ij} - \bar{M}_j\), respectively), and \(\tilde{X}_j\) and \(\tilde{M}_j\) denote group-level means of \(X\) and \(M\) in the upper level. Therefore, \(\tilde{X}_{ij}\) and \(\tilde{M}_{ij}\) strictly capture “within-level” variations, whereas \(\tilde{X}_j\) and \(\tilde{M}_j\) capture “between-level” variations.²

Loosely following the notation suggested by Kenny, Korchmaros, and Niall (2003), a set of equations presented above represent a 1–1–1 (deconfounded) “multilevel mediation” models with all effects are assumed to be random across all upper unit \(j\). Here, the three numbers in “1–1–1” simply indicate that the presumed cause \((X_{ij})\), mediator \((M_{ij})\), and the outcome \((Y_{ij})\) variables are all measured at the lowest level (i.e., level-1) of the data (Krull & MacKinnon, 2001). For the present context of modeling egocentric network data where the dyadic properties are regarded as lower-level variables and the attributes of each ego are regarded as upper-level variables, a 1–1–1 model means that all three constituent paths of the proposed mediation model are measured at each dyad \(i\), with dyadic observations are clustered within each focal ego \(j\). Since this multilevel mediation model is based on a general MLM framework, all of the standard assumptions of MLMs are equally applicable when estimating multilevel mediation models. First, we assume the predictors \((X \text{ and } M)\) are uncorrelated with their residuals in both levels. Second, residuals at level-1 \((e_{M_{ij}} \text{ and } e_{Y_{ij}})\) should be normally distributed with its expected value of zero, and should be uncorrelated with one another. Finally, the random effects (i.e., level-2 residuals) are assumed to be normally distributed, with their expected value equal to their average effect across all clusters (Bauer et al., 2006).

Having made such assumptions, now let us consider the statistical inferences regarding the average indirect effect of \(X\) on \(Y\) through a mediator \(M\) across all level-2 units. Note that in a simple, single-level traditional mediation model (Equations 4–6 above), a statistical inference regarding its indirect effect is done by assessing whether the parameter estimate quantifies the proposed indirect effect (e.g., the product of \(a_i\) path and \(b_i\) path in panel B in Figure 1) is significantly different from zero. The same principle could be equally applied to a 1–1–1 multilevel mediation model (as in Figure 2), and by doing so we can directly quantify the “average” indirect effects. Specifically, the product term between the average effect of \(X\) on \(M\) (i.e., “\(a_i\)” ) and the average effect of \(M\) on \(Y\) (i.e., “\(b_i\)” ), which can be formally expressed as \(E(a_i b_i) = a_W b_W + a_{a,b_i}\), directly quantifies the within-component of the indirect effect in question.³ Since the expected value of \(a_i\) and \(b_i\) only quantifies the extent to which a within-level variation of \(X\) and \(M\) \((\tilde{X}_{ij} \text{ and } \tilde{M}_{ij})\) affect \(M\) and \(Y\), respectively, this strictly captures a within-level indirect effect of \(X\) on \(Y\) through \(M\). In contrast, \(E(ab_{ij}) = a_B b_B\)
quantifies between-level indirect effects of X on Y through M, given the between group effects are fixed and, thus, cannot covary.

To make inferences regarding the significance of an average indirect effect, two approaches have been proposed. First, with the assumption that \( a_j \) and \( b_j \) follow a normal distribution (similar to a \( \text{Sobel} \)-type test), one can obtain CIs for the indirect effect estimate by deriving its standard errors using the sampling variance of the average indirect effect. Most of the multilevel programs or the covariance structure programs provide an option to display a variance-covariance matrix, and particularly, some covariance structure modeling programs such as \( \text{Mplus} \) implement this normal-theory based SE derivation by default. Yet just as the \( \text{Sobel} \) test, the normal-theory based approach is relying on the unrealistic assumption that the sampling distribution of the average indirect effect follows a normal distribution, therefore many researchers have warned against its use for hypothesis testing (MacKinnon, Lockwood, & Williams, 2004; Hayes & Scharkow, 2013; Preacher & Selig, 2012).

Another approach has been suggested, which is superior to the other alternative and therefore a recommended practice, is using either cases bootstrap (Efron & Tibshirani, 1994) or Monte Carlo (MC) simulations (Preacher & Selig, 2012). The cases bootstrap requires no \( \text{a priori} \) assumption of its distributional form regarding indirect effect, but it empirically simulates its sampling distribution from repeatedly drawing model estimates over a large number of replications (say, \( k \)). In contrast, MC methods derive sampling distributions of the indirect effect by simulating \( k \) values of its constituent components from a multivariate normal distribution, as described in Preacher and Selig (2012).\(^4\) Once the distribution of indirect effects are estimated from either the cases bootstrap or from MC method, one could obtain CIs for the average indirect effect by taking the value corresponding to, for instance, 2.5\(^{th}\) and 97.5\(^{th}\) percentiles to form a 95\% confidence interval. Later, I will revisit this statistical inference of the indirect effect with the MC method in greater detail, and with a step-by-step demonstration using an example dataset, discussing the software implementation of the multilevel mediation models. Readers should also bear in mind that fundamentals of estimations and statistical inferences regarding the indirect effect discussed so far are equally applicable to various models that we will discuss in the following section.

**A “2-1-1” multilevel mediation model**

Similar to the case of a “1-1-1” model, we could express a “2-1-1” multilevel mediation model predicting \( Y \) as a function of \( X \) and \( M \) as the special case of a 1-1-1 model. Again, the three numbers in “2-1-1” indicate that the presumed cause (\( X_{ij} \)) is measured at the cluster or group level (i.e., level-2), whereas mediator (\( M_{ij} \)) and the outcome (\( Y_{ij} \)) variables are measured at the lowest level (i.e., level-1) of the data. Such a model, as shown in panel B in Figure 2, may be of considerable interest to scholars who wish to model the effects of contextual-level characteristics on certain dyadic-level outcomes. For instance, one can assess how the impact of one’s egocentric network properties (such as degree of integration measured as density, communities, or the diversity of one’s network) affect dyadic-level outcomes via other dyadic-level mediator variables, provided that such contextual characteristics are unique to each individual. Such conceptual questions could be investigated using a 2-1-1 multilevel mediation model (as described in Zhang et al., 2009):

**Level-1 “M” equation:**

\[
M_{ij} = d_{M_i} + e_{M_i} \tag{17}
\]

**Level-2 “M” equation:**

\[
d_{M_i} = d_M + aX_j + u_{M_i} \tag{18}
\]
Level-1 “Y” equation:

\[ Y_{ij} = dY_i + b_j \tilde{M}_{ij} + e_{Y_{ij}} \]  

(19)

Level-2 “Y” equation:

\[ dY_i = dY_i + c'X_j + b_B \tilde{M}_j + u_{Y_i} \]  

(20)

\[ b_j = b_W + u_{b_j} \]  

(21)

where \( \tilde{M}_{ij} \) again denotes group-mean centered \( M \) as with the previous case (i.e., \( \tilde{M}_{ij} = M_{ij} - \bar{M}_j \)), and \( \bar{M}_j \) denotes group-level means of \( M \) in the upper-level. The coefficients \( b_j = b_W + u_{b_j} \) quantify the effect of \( M \) on \( Y \) at the within-level (i.e., “\( W \)” subscript) and at the between-level (i.e., “\( B \)” subscript). The direct effect of \( X \) on \( Y \), expressed as \( c' \), quantifies the effect of the upper-level value \( X \) on the mean value of \( Y \) that is estimated as the fixed effect. Since the effect of presumed cause, \( X \), only varies between “upper” levels, any indirect effects should be confined to only between-level estimates. The average between-group indirect effect of \( X \) on \( Y \) through \( M \) is, therefore, formally expressed as \( ab_B \) where within-group effect (in the 1-1 relationship) is deconflated from the estimated indirect effect.

**Multilevel mediation analysis of personal networks: an empirical example**

Having outlined basic forms of multilevel mediation models, I will now implement the models and procedures outlined above with data concerning the egocentric political network of individuals that are assessed in the 2000 ANES times series study, using Mplus, in order to demonstrate the statistical procedures outlined above.

**Data and measures**

The data used in this manuscript (hereafter denoted as “the example dataset”) originates from a 2000 ANES times series study, part of the most widely used nationally representative surveys that contain social network modules. Researchers often investigated the extent to which one’s personal networks are related to various political or social behaviors of respondents, and the present dataset provides some unique opportunities. In typical social network data such as the ANES time series studies (or any equivalent egocentric network-based datasets such as GSS or CNEP), one’s discussants (“an alter” are conceptually regarded as being nested under each of the focal respondents (“an ego”), since each ego serves as an informant regarding his or her alter. Therefore, the responses of alters are clustered within each ego. Typically, this is done by asking “name generator” and “name interpreter” questions, often the focal respondents provide from 3–5 “important matter” or “political matter” discussants within their immediate social network peers (Klofstad, McClurg, & Rolfe, 2009; Marsden, 2011).

In this application, several available dyadic-level variables in the original ANES 2000 survey were utilized to answer some substantive questions. An ego’s perception of alters’ political knowledge (RKNOW) was anchored on a 5-point scale ranging from “Not much at all” (1) to “A great deal” (5), with the “Average amount” (3) being the middle point of the scale. Alters’ political discussion frequencies with the ego (RFREQ) were measured on a 7-point scale (from “Never” to “Often”) largely the same way as the ego’s perception of alters’ knowledge was gauged. An ego’s proximity with an alter (RCLOSE) was coded as 1 if the alter is unrelated by blood, coded as 2 if the alter is a relative, and coded as 3 if the alter is spouse or domestic partner.

Several respondent-level (“level-2”) variables were utilized to capture the systematic variations occurring at the main respondent-level. Respondents’ standard demographics such as gender,
ethnicity, age, education attainment, and household income were assessed. Next, respondents’ media consumption was tapped by asking how many days in the past week they had watched national TV news and/or read a newspaper. Respondents’ factual political knowledge was assessed by asking whether or not they could correctly identify the job or political office held by several political figures, with correct responses coded as 1 and incorrect responses as 0 (including Don’t Know). Additionally, respondents were asked to evaluate the ideological standings of Gore and Bush on a seven-point scale from “extremely liberal” to “extremely conservative,” where any response naming Gore as being more liberal than Bush was regarded as the correct answer (Eveland, 2004). Finally, respondents’ self-placement of their ideological orientation was assessed on a 7-point scale from “Extremely liberal” (0) to “Extremely conservative” (6).

In most of the cases, egocentric network data are structured as a standard “wide” format, where each row represents focal respondents and some variables are recoded in a series of discrete variables, repeated for each of the alters. In order to employ MLM, it is required to be converted to a “long” format, such that each dyad is nested within the focal respondent, and each row of the long dataset contains the level-1 variables. In addition, the level-2 variables should be repeated for all rows corresponding to that focal respondent.

**Do respondents perceive their strong ties to be politically knowledgeable?**

The first model estimates the impact of a dyadic relationship’s proximity (measured as “RCLOSE”) on an ego’s perception of alters’ level of political knowledge (“RKNOW”), which is being mediated through frequency of political discussion with such alters (“RFREQ”). In this model, all key predictor, mediator, and dependent variables are assessed in the dyadic level that is nested under the individual-level, along with a set of controls. In addition, all presumed causal effects are set to vary across individual levels, as shown in panel A in Figure 2. An example Mplus code for this “1-1-1” deconflated multilevel mediation model is shown in the Appendix A, along with the abbreviated outputs from this model. Detailed results can be found in Table 1.

From Table 1, the average effects of proximity of alters (X) on discussion frequency with that alter (M) was statistically significant, $b = .194$, $SE = .026$, $p < .001$, meaning that across all respondents, those with closer ties are more likely to discuss politics. Likewise, the average effect of discussion frequency with alters (M) on the perception of alters’ knowledge level (Y) was statistically significant, $b = .278$, $SE = .040$, $p < .001$, meaning that respondents perceive them to be politically knowledgeable to the extent that they discuss politics more often. The direct effect of proximity of alters on ego’s perception of alters knowledge was not statistically significant at the dyadic level, $b = -.018$, $SE = .020$, $p > .05$, nor at the focal respondent level, $b = -.011$, $SE = .028$, $p > .05$. Note that these effects do not significantly differ across focal respondents, meaning that there is no evidence of the heterogeneity of causal effect across individuals.

The formal test of the indirect effect is provided by specifying a new parameter estimate, as defined in **MODEL CONSTRAINT** command in the syntax (as can be seen in the appendix). The indirect effect is parameterized as $E(a_jb_j) = a_Wb_W + \sigma_{a_ib_i}$ for the within-level indirect effect (where $\sigma_{a_ib_i}$ equals the covariance of the two random components), and $E(\sigma_{a}\sigma_{b}) = a_Bb_B$ for the between-level indirect effect. **Mplus** automatically derives its standard error under the assumption that the product terms (“$a_ib_i$” and “$a_Bb_B$”) are normally distributed, which can be found in the standard error for the indirect effects from the output (=.011 for the between-level, and .018 for the within-level, respectively).

As discussed previously, this normal-theory based approach of the statistical inference could be inaccurate since it relies on the unrealistic assumption that the sampling distribution of indirect effect terms are normal (Hayes, 2013; Preacher & Selig, 2012). Instead, scholars increasingly rely on the method that empirically generates a distribution of the indirect effect in question from cases bootstrap or from a large number of random draws, known as the **Monte Carlo method** (Bauer et al., 2006; MacKinnon et al., 2004; Preacher & Selig, 2012). Within the multilevel context, the MC
method has a unique advantage over other bootstrapping-based approaches, where choice of exact resampling method is ambiguous due to the nested data structure. In contrast, the MC method does not require researchers to have raw data at hand to estimate the SEs for indirect effects. Yet, estimating MC confidence intervals could be complicated due to a complex variance-covariance matrix required to estimate a MC distribution (see Preacher & Selig, 2012 for a detailed discussion).

A general form of variance-covariance matrix needed to estimate a MC distribution for within-level indirect effect in a 1-1-1 multilevel mediation model is defined as follows:

$$
\begin{pmatrix}
V(a_W) & Cov(a_W, b_W) & Cov(a_W, Cov(a_j, b_j)) \\
Cov(a_W, b_W) & V(b_W) & Cov(b_W, Cov(a_j, b_j)) \\
Cov(a_W, Cov(a_j, b_j)) & Cov(b_W, Cov(a_j, b_j)) & V(Cov(a_j, b_j))
\end{pmatrix}
$$

In the model syntax, we labeled the average within-level effect of X on M (a_W) and the average within-level effect of M on Y (b_W) as s_a and s_b respectively in the model syntax (see the Appendix). In the tech1 output (Figure A2 in the Appendix), the s_a path is identified as 4, and the s_b is identified as 5, and the covariance of the two paths (i.e., Cov(a_j, b_j)) is identified as 31.5 In the tech3 output, the variance of the s_a (i.e., V(a_W)) is the cell entry of (4,4) in the estimated variance-covariance matrix, and the variance of the s_b (i.e., V(b_W)) is the cell entry of (5,5), and the variance of the Cov(a_j, b_j) (i.e., V(Cov(a_j, b_j))) is the cell entry of (31,31) of the variance-covariance matrix. Subsequently, one could construct a variance-covariance matrix for estimating MC intervals using the same principle of locating the specific value of cell entries:

**Table 1.** Parameter estimates for 1–1–1 multilevel mediation model.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mediator (RFREQ)</td>
<td>Dependent (RKNOW)</td>
</tr>
<tr>
<td><strong>Dyadic-level (L1) Effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean intercept</td>
<td>2.640 (.115)***</td>
<td>1.169 (.121)***</td>
</tr>
<tr>
<td>Proximity of alter (&quot;RCLOSE&quot;)</td>
<td>.194 (.026)***</td>
<td>-.018 (.028)</td>
</tr>
<tr>
<td>Discussion frequency (&quot;RFREQ&quot;)</td>
<td>-</td>
<td>.278 (.040)***</td>
</tr>
<tr>
<td><strong>Individual-level (L2) Effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>.013 (.042)</td>
<td>.097 (.032)***</td>
</tr>
<tr>
<td>Age</td>
<td>.001 (.01)</td>
<td>.000 (.001)</td>
</tr>
<tr>
<td>White</td>
<td>-.092 (.040)*</td>
<td>-.008 (.030)</td>
</tr>
<tr>
<td>Black</td>
<td>-.054 (.094)</td>
<td>-.047 (.073)</td>
</tr>
<tr>
<td>Education</td>
<td>.020 (.014)</td>
<td>.027 (.011)*</td>
</tr>
<tr>
<td>Income</td>
<td>.000 (.006)</td>
<td>.013 (.005)***</td>
</tr>
<tr>
<td>TV news exposure</td>
<td>.032 (.007)***</td>
<td>-.009 (.006)</td>
</tr>
<tr>
<td>Newspaper exposure</td>
<td>.010 (.007)</td>
<td>-.008 (.006)</td>
</tr>
<tr>
<td>Level of ego’s political knowledge</td>
<td>.064 (.016)***</td>
<td>.034 (.012)*</td>
</tr>
<tr>
<td>Proximity of alter (&quot;RCLOSE&quot;)</td>
<td>.028 (.037)</td>
<td>-.011 (.028)</td>
</tr>
<tr>
<td>Discussion frequency (&quot;RFREQ&quot;)</td>
<td>-</td>
<td>.304 (.031)***</td>
</tr>
<tr>
<td><strong>Variance of Random Components</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximity of alter (&quot;RCLOSE&quot;)</td>
<td>.093 (.022)***</td>
<td>.008 (.028)</td>
</tr>
<tr>
<td>Discussion frequency (&quot;RFREQ&quot;)</td>
<td>-</td>
<td>.036 (.045)</td>
</tr>
<tr>
<td>Random intercept</td>
<td>.279 (.013)***</td>
<td>.047 (.011)***</td>
</tr>
<tr>
<td>Residual variance</td>
<td>.146 (.009)***</td>
<td>.266 (.017)***</td>
</tr>
<tr>
<td>Direct effect, between-level</td>
<td>- .011 (.028)</td>
<td></td>
</tr>
<tr>
<td>Direct effect, within-level</td>
<td>- .018 (.028)</td>
<td></td>
</tr>
<tr>
<td>Indirect effect, between-level</td>
<td>.009 (.011)</td>
<td></td>
</tr>
<tr>
<td>Indirect effect, within-level</td>
<td>.042 (.018)*</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>8898.356</td>
<td>9118.773</td>
</tr>
<tr>
<td>BIC</td>
<td>9118.773</td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>L1 = 1827/L2 = 764</td>
<td></td>
</tr>
</tbody>
</table>

*Note: Model predicts egos’ perception of alters’ political knowledge level. Maximum likelihood estimation with robust standard errors (MLR) was used in estimation.*** p < .001, ** p < .01, * p < .05.*
Once constructing such a variance-covariance matrix, one could get a MC confidence interval for the indirect effect using an available program, as in Preacher and Selig’s (2012) interactive tool, which generates a MC confidence interval for the indirect effect based on a large number of replications. For this application I’ve used 20,000 random draws, and their corresponding confidence intervals at .95 levels are shown in the left panel of Figure 3. The mean of the distribution of indirect effects is .0346, with the 2.5 percentiles equal to .0317 and the 97.5 percentile equal to .0801. Therefore, the CI for within-level indirect effect does not include zero at 95% confidence level, meaning that the effect of relational proximity on the ego’s perception of alters’ political knowledge is significantly mediated through the discussion frequency for each alter (b = .0346, 95% Monte Carlo CIs = [.0317, .0801]). Likewise, one can construct the between-level indirect effect largely based on the sample principle, with the exception that covariance of the two between-level regression coefficients should be set to zero, given the between group effects are fixed and, thus, cannot covary. The estimated between-level indirect effect was not statistically significant, b = .0029, 95% MC CIs = [-.0024, .0108], as can be seen in right panel of Figure 3.

Further considerations: multilevel moderated mediation

In the previous application, we have found no evidence that an indirect effect varies across higher-level units. Yet, one might wish to explicitly model such possible contingencies to express boundary conditions of observed causal processes. Testing multilevel “moderated mediation” faithfully represents such analytical imperatives, in that systematic variations in constituent paths of between or within indirect effects by a level-2 predictor, W, would mean that the strength of the indirect effect itself depends on the value of W.

Let d be quantifications of causal effects, where d be the extent to which the intercept and the slope vary as a function of a level-2 predictor W. This could be done by adding a new level-2 predictor variable W into a simple 1-1-1 multilevel mediation model above. Then we could express a generic form of a multilevel “moderated mediation” model as follows.

\[
\begin{pmatrix}
(4, 4) & (4, 5) & (4, 31) \\
(4, 5) & (5, 5) & (5, 31) \\
(4, 31) & (5, 31) & (31, 31)
\end{pmatrix}
= \begin{pmatrix}
0.001 & 0.000 & 0.000 \\
0.000 & 0.002 & 0.000 \\
0.000 & 0.000 & 0.000
\end{pmatrix}.
\]
**Level-1 “M” equation:**

\[ M_{ij} = d_{Mj} + a_j \tilde{X}_{ij} + e_{Mj} \]  

**Level-2 “M” equation:**

\[ d_{Mj} = d_M + d_{ij} W_j + a_B \tilde{X}_j + u_{iMj} \]  

\[ a_j = a_W + d_{2j} W_j + u_{aj} \]  

**Level-1 “Y” equation:**

\[ Y_{ij} = d_{Yj} + c_j' \tilde{X}_{ij} + b_j \tilde{M}_{ij} + e_{Yj} \]  

**Level-2 “Y” equation:**

\[ d_{Yj} = d_Y + d_{3j} W_j + c_B' \tilde{X}_j + b_B \tilde{M}_j + u_{iYj} \]  

\[ c_j' = c_{W'} + d_{4j} W_j + u_{c_j'} \]  

\[ b_j = b_W + d_{5j} W_j + u_{b_j} \]  

The conditional direct effect of \( X \) on \( Y \) (\( =c_j' \)), and the conditional indirect effect of \( X \) on \( Y \) through a mediator \( M \) at the value of \( W \) (\( =a_j b_j \)) are expressed as follows:

\[ c_j' = E(c_j'|W_j = w) = c_{W'} + d_{4j} W \]  

\[ a_j b_j = E(a_j b_j|W_j = w) = (a_W + d_{2j} W)(b_W + d_{5j} W) + \sigma_{u_a u_b}. \]  

From this general formation, it could be further ascertained if the indirect effect itself is being moderated by a third variable \( W \). For instance, one may wish to evaluate whether the impact of relational closeness (\( RCLOSE \)) on perceived knowledge of alters (\( RKNOW \)), which is being mediated through dyadic discussion frequency (\( RFREQ \)), is in turn conditional on ego’s party identification (“\( PID \)”). A full \textit{Mplus} code for this multilevel moderated mediation is presented in the Appendix B. Central to our interest, conditional indirect effects at values of \( PID \) (from “strong Democrat” = 0 to “strong Republican” = 6, \( M = 2.89, SD = 2.37 \)) are probed by invoking the above Equation 30. Here, rejecting the null hypothesis of either \( d_{2j} \) (as in “\( a_j \)”) or \( d_{5j} \) (as in “\( b_j \)”) equal to zero satisfies the requirement of a moderated mediation effect to occur strictly at within-level. As shown in Table 2, we found no evidence that ego’s party identification (\( PID \)) significantly moderates the relationship between relational closeness and dyadic frequency, \( b = -.004, SE = .011, p > .05 \), nor between dyadic discussion frequency and perceived knowledge of alter, \( b = -.008, SE = .015, p > .05 \). Yet we found evidence that the conditional indirect effects are significant when an ego is “strong Democrat” (\( PID = 0 \), indirect = .049, \( SE = .023, p < .05 \)) and when an ego is Independent (\( PID = 3 \), indirect = .041, \( SE = .018, p < .05 \)), but not significant when an ego is “strong Republican” (\( PID = 6 \), indirect = .034, \( SE = .022, p = .11 \)), using a normal-theory based approach as described earlier. Yet Monte Carlo CIs for indirect effects reveal that all of the 95% CIs exclude zero at each of three values of \( PID \) (0, 3, and 6), as depicted in Figure 4. A formal test of the index of moderated mediation also supported the conclusion that the within-level indirect effect does not vary as a function of the moderator variable.\(^7\)
Table 2. Parameter estimates for 1–1–1 multilevel moderated mediation model.

<table>
<thead>
<tr>
<th></th>
<th>Model 1 Mediator (RFREQ)</th>
<th>Model 2 Dependent (RKNOW)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dyadic-level (L1) Effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean intercept</td>
<td>2.623 (.118)***</td>
<td>1.174 (.123)***</td>
</tr>
<tr>
<td>Proximity of alter (&quot;RCLOSE&quot;)</td>
<td>.205 (.043)***</td>
<td>-.012 (.043)</td>
</tr>
<tr>
<td>x Party identification (&quot;PID&quot;)</td>
<td>-.004 (.011)</td>
<td>-.002 (.011)</td>
</tr>
<tr>
<td>Discussion frequency (&quot;RFREQ&quot;)</td>
<td>-</td>
<td>.301 (.062)***</td>
</tr>
<tr>
<td>x Party identification (&quot;PID&quot;)</td>
<td>-</td>
<td>-.008 (.015)</td>
</tr>
<tr>
<td><strong>Individual-level (L2) Effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>.014 (.042)</td>
<td>.098 (.033)**</td>
</tr>
<tr>
<td>Age</td>
<td>.001 (.001)</td>
<td>.000 (.001)</td>
</tr>
<tr>
<td>White</td>
<td>-.092 (.040)*</td>
<td>-.007 (.030)</td>
</tr>
<tr>
<td>Black</td>
<td>-.042 (.097)</td>
<td>-.051 (.075)</td>
</tr>
<tr>
<td>Education</td>
<td>.020 (.014)</td>
<td>.027 (.011)*</td>
</tr>
<tr>
<td>Income</td>
<td>-.001 (.006)</td>
<td>.013 (.005)**</td>
</tr>
<tr>
<td>TV news exposure</td>
<td>.032 (.007)**</td>
<td>.009 (.006)</td>
</tr>
<tr>
<td>Newspaper exposure</td>
<td>.032 (.007)</td>
<td>-.008 (.006)</td>
</tr>
<tr>
<td>Level of ego's political knowledge</td>
<td>.063 (.016)**</td>
<td>.034 (.012)**</td>
</tr>
<tr>
<td>Proximity of alter (&quot;RCLOSE&quot;)</td>
<td>.029 (.037)</td>
<td>-.012 (.029)</td>
</tr>
<tr>
<td>Discussion frequency (&quot;RFREQ&quot;)</td>
<td>-</td>
<td>.305 (.031)*****</td>
</tr>
<tr>
<td>Party identification (&quot;PID&quot;)</td>
<td>.005 (.009)</td>
<td>-.001 (.007)</td>
</tr>
<tr>
<td><strong>Variance of Random Components</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximity of alter (&quot;RCLOSE&quot;)</td>
<td>.093 (.022)**</td>
<td>.008 (.028)</td>
</tr>
<tr>
<td>Discussion frequency (&quot;RFREQ&quot;)</td>
<td>.037 (.045)</td>
<td>.047 (.011)**</td>
</tr>
<tr>
<td>Random intercept</td>
<td>.279 (.013)***</td>
<td>.266 (.017)**</td>
</tr>
<tr>
<td>Residual variance</td>
<td>.146 (.009)***</td>
<td></td>
</tr>
</tbody>
</table>

When PID = 0 ("Strong Democrat")
When PID = 3 ("Independent")
When PID = 6 ("Strong Republican")

Between direct effect
Within direct effect

Between indirect effect
Within indirect effect

AIC: 8898.092
BIC: 9145.987
Sample size: L1 = 1824/L2 = 763

Note: Model predicts egos’ perception of alters’ political knowledge level. Maximum likelihood estimation with robust standard errors (MLR) was used in estimation.

*** p < .001, ** p < .01, * p < .05.

Figure 4. Monte Carlo confidence intervals of the conditional indirect effect at various values of moderators (W = 0, 3, and 6), from the 1-1-1 multilevel moderated mediation model.
Various issues in testing multilevel (moderated) mediation

Sample size and number of random effects

The identification of variance components in multilevel regressions heavily depends upon the number of level-1 observations per level-2 units, whereas the accuracy of such variance component estimates heavily depend on the number of observations ascertained in the upper level (Bauer et al., 2006; Hox, 2010). Often, sample size considerations in level-1 and level-2 units work in opposite directions, such that researchers must choose whether or not to maximize their level-2 observations while sacrificing the average number of cluster size (available level-1 units per each level-2 unit). Also, unique to a personal network data, most of the name-generator methods typically constrain the number of alters to be 3–5 at most. This may present unique challenges in some model specifications due to the limited variance of within-level predictors, especially for the random effect components.

In general, the estimation becomes much more difficult and demanding when there are more random effect components in the model (Hox, 2010; Snijders & Bosker, 2012), and this is no exception for multilevel mediation models discussed here (Bauer et al., 2006). The estimation of random effects becomes even more complex when researchers proceed to a multilevel moderated mediation model. Therefore, researchers may wish to first estimate a simple random coefficient model without any predictors (such as the simple 1-1-1 model above, possibly with or without any random effects) to evaluate whether any of the constituent paths of a direct and an indirect effect significantly vary across the higher units. Then in estimating more complex models, one may choose to fix some constituent paths of direct/indirect effect to be invariant across higher units when there is no evidence of significant random effects. Although the unique structure of egocentric network data provides some advantages over ordinary multilevel-type data due to its large number of level-2 units (i.e., individual-level), there is no guarantee that the estimation would be successful if the theoretical model involves many of the random components, especially when there is limited number of level-1 units per cluster. Ideally, the empirical power of multilevel models, including multilevel mediation models, becomes greater as both level-2 and nested level-1 sample size increases. For a general recommendation, readers should consult available literatures such as Hox (2010), Snijders and Bosker (2012), or Raudenbush and Bryk (2002).

Complex data structure and whole network data

Social networks often represent a complex nested structure that involves more than two levels. For instance, one may construct a three-level nested structure such that one’ egocentric network is nested within each of the focal respondents, and these respondents are in turn nested within social groups, entities, or certain geographical boundaries (e.g., The National Longitudinal Study of Adolescent Health Study). Theoretically, the general principle described here could be extended to yield a three-level (moderated) mediation model without substantial modification. Yet the analysis of three-or more MLM has only just begun, and only a few existing literatures have explored this possibility (e.g., Preacher, 2011) due to the complexity of fitting and evaluating such models.

For whole network data, even less attention has been paid to the formal test and inferences. One promising possibility, at least theoretically, is that one could effectively model a whole network using a cross-classified random coefficient model, where each \( N^2 \) number of dyadic observations of whole network data \( (X_{ij}) \) is regarded as being nested within crossing of focal respondents \( (X_i) \) and their network alters \( (X_j) \) (see Tranmer, Steel, & Browne, 2013). In principle, one can get a point estimate of the indirect effect in question (“ab”), using any of sensible method for dealing with whole network data (such as Stochastic Actor Oriented Model, QAP regressions, network autocorrelation models, or cross-classified random coefficient models) by piece-wise estimations of the models predicting a mediator and a focal dependent variable. Then, one could further obtain empirical distributions of the indirect effect using resampling-like procedure while taking the observed interdependency of the
network data into account, such as using matrix random permutations (see O’Malley, 2013, for a brief discussion). Yet for any of the cases, because it is limited to a piece-wise method at best, the extent of bias and coverage rates of the possibly incorrectly estimated indirect effect relative to a true indirect effect is largely unknown. For that matter, a method of directly estimating and establishing evidence of statistical mediation for whole network data is not yet available.

Software implementation

All procedures covered in this manuscript were based on Mplus 7.0 software, one of the most widely used covariance structure modeling programs for general multilevel SEM framework (Preacher, 2011). Yet there are a couple of other alternatives; the R based OpenMX package (Neale et al., 2016) offers a most similar, if not identical, functionality of estimating multilevel mediation models described here, including SEM-based extensions. For Bayesian multilevel mediation, the bmlm package (Vuorre, 2017) offer similar functionality. Readers who are interested in causal mediation analysis (which is based on Rubin’s counterfactual framework) would find the R package mediation (Imai, Keele, & Tingley, 2010) to be useful, although it does not appear to report separate between- and within-level mediation effects (i.e., conflated estimates). One can also estimate the equivalent models using the piece-wise approach based on a standard MLM package such as lme4 (Bates et al., 2015). Yet covariance of the random slopes cannot be estimated in such cases, therefore estimated indirect effect may be misleading to some degree. Finally, a recently developed SPSS-based MLmed macro (Rockwood, 2017) offers many similar functions without requiring the knowledge of complex syntax language and pre-data management, greatly simplifies the analysis while supporting the most widely available software for communication researchers (i.e., SPSS). Yet some features in MLmed (e.g., number of maximum covariates in the model) are limited compared to other cases. Most of the mentioned R packages and MLmed macro offer detailed vignettes, which interested readers can further consult if they wish to estimate the model for their own.

Concluding remarks

In this article, a series of analytical and modeling strategies to assess mediation hypothesis have been proposed, focusing on its application to personal social network data (i.e., egocentric social network). An analytical and methodological improvement has been heralded by social networks scholars in favor of a more accurate understanding of causal mechanisms within and across individuals’ social networks (Fowler, Heaney, Nickerson, Padgett, & Sinclair, 2011), and there still remains a considerable possibility for further methodological development (Wang et al., 2016). Traditionally, multilevel modeling and the formal test of the mediation and the moderation have been proven to be useful across various disciplines of social sciences. The analytical integration of the two—as advocated in this manuscript—opens up a considerable possibility for statistical analysis of egocentric network data regarding the formal test of causal mechanisms within and across networks. Although what has been presented in this manuscript has been brief and non-exhaustive, I hope this article will serve as the first steps in a full elaboration of a multilevel mediation modeling to network analysis using egocentric networks.

Notes

1. The Quadratic Assignment Procedure (QAP) tests are permutation-based tests, where a multiple regression model is repeatedly estimated with randomly permuted (original) matrix data in order to yield a reference distribution of regression coefficients, the method of which takes the observed interdependency of the network data into account. See Dekker, Krackhardt, and Snijders (2007) or Cranmer, Leifeld, McClurg, and Rolfe (2017) for a detail.

2. Since the single indirect effect estimate contains two sources of variations in multilevel mediation models, we cannot make a definite claim regarding the nature of such indirect effects if within- and between-
components are conflated within a single indirect effect estimate. This may be seen as a rather trivial issue, yet it poses some concerns regarding the precision of the statistical inferences and the validity of the corresponding interpretation of the model parameters when such models are used, therefore literature strongly advocates the use of deconflated models over conflated models (Lüdtke, Mars, Robitzsch, & Ulrich, 2008; Preacher et al., 2010).

3. Because \(a_i\) and \(b_i\) are not independent, the expected value of \(a_i b_i\) is the function of the average effect of \(X\) on \(M\) (\(a_i = a_W\)), the average effect of \(M\) on \(Y\) (\(b_i = b_W\)), and the covariance of the two random effects (\(a_{ab}\)). In some cases this covariance term is not statistically significant, yet the covariance term should always be included regardless of its statistical significance when assessing indirect effects for multilevel mediation. See Bauer et al. (2006) for a detail.

4. In the single-level mediation case, the MC method replaces covariance between two constituent components (\(a_{ab}\)) with zero for simplicity, with no a priory assumption about the distribution about \(a\overline{b}\), nor about \(a_{ab}\). For the covariance of non-zero random slopes (as in MLM), often Wishart distribution is assumed for its sampling distribution. However, although certain parametric assumptions can be made for a covariance component, their joint asymptotic distribution over a large number of random draws would be close to normal according to the Central Limit Theorem.

5. A tech1 output of Mplus presents sequential identification numbers (e.g., 1, 2, 3, …) for the estimated model parameters, and this identification numbers are used in tech3 output to provide the cell entries of the variance-covariance matrix of all parameter arrays. A sample of tech1 and tech3 output is presented in the Appendix C. In tech1 output, “Alpha” matrix contains the means and intercepts of the continuous latent variables, and “PSI” matrix contains the variance and residual variances of continuous latent variables (see “PARAMETER ARRAYS” in the Mplus user guide: Muthén & Muthén, 1998-2012).

6. SPSS MCMED macro (Hayes, 2013) also provides similar functionality.

7. Since the proposed (continuous) moderator both affects the \(X \rightarrow M\) and \(M \rightarrow Y\) paths, the conditional indirect effect becomes the non-linear function of a moderator. The procedure described in Edwards and Lambert (2007) was used for probing whether conditional indirect effects at different values of moderators are significantly different from each other.

Acknowledgments

I would like to thank Dr. Andrew F. Hayes and Dr. Neha Gondal for their helpful comments and suggestions on earlier version of the manuscript. I also thank the action editor Dr. “Chris” Fei Shen and three anonymous reviewers for their constructive comments and suggestions on improving the manuscript. All remaining errors are the sole responsibility of the author.

ORCID

Hyunjin Song http://orcid.org/0000-0001-7752-3035

References


Appendix A: Mplus syntax for the sample deconflated 1-1-1 model

TITLE: 1-1-1 DECONFLATED model with covariates.

DATA:
! text file containing raw data in long format
file is 'data.txt';

VARIABLE:
! variables in the dataset
names are CaseID PID VOTE TVNEWS NPNEWS Know Age Edu Income gender White
Black AlterID Rclose Rgender RKnow Rfreq RVtSm Rclosemn RKnowmn Rfreqmn;
! declare variables to be used in the analysis
usevariables are TVNEWS NPNEWS Know Age Edu Income gender White Black
CaseID Rclose RKnow Rfreq Rclosenum Rfrequm;
! missing values are coded as -9999 in the dataset
missing are all \( -9999 \);

! level-2 grouping variable is “CaseID”
Cluster is CaseID;
! identify variables that only have within-variances
within are Rclos\(e\) Rfreq;
! define variables for between-variance
! \( M(\text{group mean}) = Rfreqmn \) and \( X(\text{group mean}) = Rclos\(e\)mn \)
! those two variables should be pre-calculated before analysis.
between are Rclos\(e\)mn Rfreqmn TVNEWS NPNEWS Know Age Edu Income gender White Black;

**DEFINE:**
! group mean-center \( X \) and \( M \)
center Rclos\(e\) Rfreq (groupmean);

**ANALYSIS:**
! multilevel analysis, using robust SE, multico\(r\)e processing
Type = twolevel random;
estimator = MLR;
processor = 4;

**MODEL:**
! \( X = Rclo\(s\), \ M = Rfreq, \ Y = RK\(n\)ow; \)
%WITHIN% ! model for within effect
s_a | Rfreq ON Rclos\(e\); \! regress \( m \) on \( x \), call the random slope “\( S_a \)”
s_b | RK\(n\)ow ON Rfreq; \! regress \( y \) on \( m \), call the random slope “\( S_b \)”
s_cp | RK\(n\)ow ON Rclos\(e\); \! regress \( y \) on \( x \), call the random slope “\( S_{cp} \)”
[Rclos\(e\)@0]; \! \( X \) was group-mean centered, so fix its mean to zero
[Rfreq@0]; \! \( M \) was group-mean centered, so fix its mean to zero
%BETWEEN% ! model for between effect
Rfreqmn RK\(n\)ow \! estimate residual variance of \( M(\text{mean}) \) and \( Y \)
Rfreqmn ON Rclos\(e\)mn (ab); \! regress \( M(\text{mean}) \) on \( X(\text{mean}) \), between “\( a = ab \)”
RK\(n\)ow ON Rfreqmn (cpb); \! regress \( Y \) on \( X(\text{mean}) \), between “\( \text{direct} = cpb \)”
RK\(n\)ow ON Rfreqmn (bb); \! regress \( Y \) on \( M(\text{mean}) \), between “\( b = bb \)”
\! Regress \( Y \) on L2 covariates
RK\(n\)ow ON TVNEWS NPNEWS Know Age Edu Income gender White Black;
\! Regress \( M(\text{mean}) \) on L2 covariates
Rfreqmn ON TVNEWS NPNEWS Know Age Edu Income gender White Black;
\[ s_a \] (a); \! mean of within-level variation, path “\( aw \)”
\[ s_b \] (b); \! mean of within-level variation, path “\( bw \)”
\[ s_{cp} \] (cp); \! mean of within-level variation, path “\( cpw \)”
\! freely estimate covariance between random slopes and intercepts
RK\(n\)ow WITH s_a s_b s_{cp};
\! freely estimate covariance between random slopes
s_a WITH s_b (covab); \! covariance of two random slopes, call “\( \text{covab} \)”
s_a WITH s_{cp};
s_b WITH s_{cp};

**MODEL CONSTRAINT:**
! deconf\(l\)icted within- and between- direct/indirect effect
new (ind_w direct_w direct_w); \! define new variables
ind_w = a*b+covab; \! within-indirect effect
ind_b = ab*bb; \! between-indirect effect
direct_w = cp; \! within-direct effect
direct_b = cpb; \! between-direct effect

**OUTPUT:**
! request parameter specification (tech1)
! request variance-covariance matrix (tech3)
te\(ch1\) tech3;
Appendix B: Mplus syntax for the sample deconfounded 1–1–1 moderated mediation model

TITLE: 1-1-1 DECONFLATED moderation mediation model with covariates.

DATA:
! text file containing raw data in long format
file is 'data.txt';

VARIABLE:
! variables in the dataset
names are CaseID PID VOTE TVNEWS NPNEWS Know Age Edu Income gender White Black
AlterID Rclose Rgender RKnow RFreq RVtSm Rclosemn RKfrequm;  
! declare variables to be used in the analysis
usevariables are TVNEWS NPNEWS Know Age Edu Income gender White Black
CaseID Rclose RKfrequm Rclosemn PID;
! missing values are coded as −9999 in the data
missing are all (−9999);
! level-2 grouping variable
Cluster is CaseID;
! identify variables that only have within-level variances
within are Rclose RFreq;
! define variables for between-level variance
! M(group mean) = Rfrequm, and X(group mean) = Rclosemn
! those two variables should be pre-calculated before analysis.
between are Rclosemn RFrequm TVNEWS NPNEWS Know Age Edu Income gender White Black PID;

DEFINE:
! group mean-center X and M to capture strict within-variance
center Rclose RFreq (groupmean);

ANALYSIS:
! multilevel analysis, using robust SE, multicore processing
Type = twolevel random;
estimator = MLR;
processor = 4;

MODEL:
! X = Rclose, M = RFreq, Y = RKfrequ;
%WITHIN% ! model for within effect
s_a | RFreq ON Rclose; ! regress m on x, call the random slope “S_a”
s_b | RKfrequ ON RFreq; ! regress y on m, call the random slope “S_b”
s_cp | RKfrequ ON Rclose; ! regress y on x, call the random slope “S_cp”

[Rclose@0]; ! X was group-mean centered, so fix its mean to zero
[RFrequ@0]; ! M was group-mean centered, so fix its mean to zero

%BETWEEN% ! model for between effect
RFrequm RKfrequ ! estimate residual variance of M(mean) and Y
RFrequm ON Rclosemn (ab); ! regress M(mean) on X(mean), between “a” = ab
RKfrequ ON Rclosemn (cpb); ! regress Y on X(mean), between “direct” = cpb
RKfrequ ON RFrequm (bb); ! regress Y on M(mean), between “b” = bb
! Regress Y on L2 covariates
RKfrequ ON TVNEWS NPNEWS Know Age Edu Income gender White Black PID;
! Regress M(mean) on L2 covariates
RFrequm ON TVNEWS NPNEWS Know Age Edu Income gender White Black PID;

[s_a] (aw); ! conditional mean of within-level variation, path “a”
[s_b] (bw); ! conditional mean of within-level variation, path “b”
[s_cp] (cpw); ! conditional mean of within-level variation, path “cp”
s_a ON PID (d2); ! within-level variation of a as a function of PID, “d2”
s_b ON PID (d5); ! within-level variation of b as a function of PID, “d5”
s_cp ON PID (d4); ! within-level variation of cp as a function of PID, “d4”
! freely estimate covariance between random slopes and intercepts
RKnow WITH s_a s_b s_cp;
! freely estimate covariance between random slopes
s_a WITH s_b (covab); ! covariance of two random effects, call “covab”
s_a WITH s_cp;
s_b WITH s_cp;

Model constraint:
! deconflated within- and between- direct/indirect effect
new (wmodval1 direct1 indi1);
wmodval1 = 0; ! when PID = 0 “Strong Democrats”, set modal value of w as 0
indi1 = (aw+d2*wmodval1)*(bw+d5*wmodval1)+covab;
direct1 = cpw+d4*wmodval1;
new (wmodval2 direct2 indi2);
wmodval2 = 3; ! when PID = 3 “Independents”, set modal value of w as 3
indi2 = (aw+d2*wmodval2)*(bw+d5*wmodval2)+covab;
direct2 = cpw+d4*wmodval2;
new (wmodval3 direct3 indi3);
wmodval3 = 6; ! when PID = 6 “Strong Republicans”, set modal value of w as 6
indi3 = (aw+d2*wmodval3)*(bw+d5*wmodval3)+covab;
direct3 = cpw+d4*wmodval3;
! formal test of differences between two specific indirect effect
! when any of those indices are different from zero, it satisfies the requirement
! of moderated mediation (see Edwards & Lambert, 2007)
new (index1 index2 index3);
index1 = indi1 - indi2;
index2 = indi1 - indi3;
index3 = indi2 - indi3;
! between-level direct/indirect effect
new (direct_b indirect_b);
direct_b = cpb;
indirect_b = ab*bb;

OUTPUT:
! request parameter specification (tech1)
! request variance-covariance matrix (tech3)
tech1,tech3;
Appendix C: Additional figures

<table>
<thead>
<tr>
<th>Within Level</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Means</td>
<td></td>
</tr>
<tr>
<td>RCLOSE</td>
<td>0.000</td>
</tr>
<tr>
<td>Intercepts</td>
<td></td>
</tr>
<tr>
<td>RFREQ</td>
<td>0.000</td>
</tr>
<tr>
<td>Variances</td>
<td></td>
</tr>
<tr>
<td>RCLOSE</td>
<td>0.254</td>
</tr>
<tr>
<td>Residual Variances</td>
<td></td>
</tr>
<tr>
<td>RKNOW</td>
<td>0.266</td>
</tr>
<tr>
<td>RFREQ</td>
<td>0.146</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Between Level</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RFREQMN ON</td>
<td></td>
</tr>
<tr>
<td>RCLOSEMN</td>
<td>0.028</td>
</tr>
<tr>
<td>TVNEWS</td>
<td>0.032</td>
</tr>
<tr>
<td>NFINES</td>
<td>0.010</td>
</tr>
<tr>
<td>NFINM</td>
<td>0.001</td>
</tr>
<tr>
<td>AGE</td>
<td>0.001</td>
</tr>
<tr>
<td>EDU</td>
<td>0.020</td>
</tr>
<tr>
<td>INCOME</td>
<td>0.000</td>
</tr>
<tr>
<td>GENDER</td>
<td>0.013</td>
</tr>
<tr>
<td>WHITE</td>
<td>-0.092</td>
</tr>
<tr>
<td>BLACK</td>
<td>-0.054</td>
</tr>
<tr>
<td>RKNOW ON</td>
<td></td>
</tr>
<tr>
<td>RCLOSEMN</td>
<td>-0.011</td>
</tr>
<tr>
<td>RFREQMN</td>
<td>0.304</td>
</tr>
<tr>
<td>TVNEWS</td>
<td>-0.011</td>
</tr>
<tr>
<td>NFINES</td>
<td>-0.008</td>
</tr>
<tr>
<td>NFINM</td>
<td>0.034</td>
</tr>
<tr>
<td>AGE</td>
<td>0.000</td>
</tr>
<tr>
<td>EDU</td>
<td>0.027</td>
</tr>
<tr>
<td>INCOME</td>
<td>0.013</td>
</tr>
<tr>
<td>GENDER</td>
<td>0.097</td>
</tr>
<tr>
<td>WHITE</td>
<td>-0.008</td>
</tr>
<tr>
<td>BLACK</td>
<td>-0.047</td>
</tr>
<tr>
<td>RKNOW WITH</td>
<td></td>
</tr>
<tr>
<td>S_A</td>
<td>0.011</td>
</tr>
<tr>
<td>S_B</td>
<td>-0.001</td>
</tr>
<tr>
<td>S_CP</td>
<td>0.000</td>
</tr>
<tr>
<td>S_A WITH</td>
<td></td>
</tr>
<tr>
<td>S_B</td>
<td>-0.014</td>
</tr>
<tr>
<td>S_CP</td>
<td>0.000</td>
</tr>
<tr>
<td>S_B WITH</td>
<td></td>
</tr>
<tr>
<td>S_CP</td>
<td>0.008</td>
</tr>
<tr>
<td>Means</td>
<td></td>
</tr>
<tr>
<td>S_A</td>
<td>0.194</td>
</tr>
<tr>
<td>S_B</td>
<td>0.207</td>
</tr>
<tr>
<td>S_CP</td>
<td>-0.018</td>
</tr>
<tr>
<td>Intercepts</td>
<td></td>
</tr>
<tr>
<td>RFREQMN</td>
<td>2.640</td>
</tr>
<tr>
<td>RKNOW</td>
<td>1.169</td>
</tr>
<tr>
<td>Variances</td>
<td></td>
</tr>
<tr>
<td>S_A</td>
<td>0.093</td>
</tr>
<tr>
<td>S_B</td>
<td>0.036</td>
</tr>
<tr>
<td>S_CP</td>
<td>0.008</td>
</tr>
<tr>
<td>Residual Variances</td>
<td></td>
</tr>
<tr>
<td>RFREQMN</td>
<td>0.279</td>
</tr>
<tr>
<td>RKNOW</td>
<td>0.047</td>
</tr>
<tr>
<td>New/Additional Parameters</td>
<td></td>
</tr>
<tr>
<td>IND_B</td>
<td>0.009</td>
</tr>
<tr>
<td>IND_W</td>
<td>0.042</td>
</tr>
<tr>
<td>DIRECT_B</td>
<td>-0.011</td>
</tr>
<tr>
<td>DIRECT_W</td>
<td>-0.018</td>
</tr>
</tbody>
</table>

Figure A1. An abbreviated Mplus output from the 1–1–1 multilevel mediation model, with all effects are random.
Figure A2. Sample Mplus tech1 output from the 1–1–1 model.

Figure A3. Sample Mplus tech3 output from the 1-1-1 multilevel mediation model, with all effects are random. Cell entries of (4,4) in this covariance matrix is the variance of $a_w$ ("a_w") parameter, which is identified as "4" in the tech1 output above.