Co-evolution models of longitudinally measured interactions

Brouwer, Jasperina; Froehlich, Dominik E.

Published in:
Analyzing Group Interactions

DOI:
10.4324/9780367321116-12

Publication date:
2020

10

CO-EVOLUTION MODELS OF LONGITUDINALLY MEASURED INTERACTIONS

Jasperina Brouwer and Dominik E. Froehlich

Introduction

Social network analysis (SNA) has gained increasing attention in any context where interactions take place, for example, in education among students or at the workplace among employees. These analyses may focus on status (e.g., friendship) or interactions (e.g., communication), the latter being emphasized in this chapter. SNA investigates the social structures, which implies the ties (connections) among the individuals (i.e., actors) within a certain network. The social network structures need to be examined together with important attributes of the actors who make up the social network. These attributes may be, for instance, background characteristics (e.g., gender, age, grades) but also to attitudes or behaviors (Borgatti, Everett, & Johnson, 2018)

Recent technological and methodological advancements have promoted a plethora of cross-sectional or longitudinal methods to study group interactions within the framework of SNA ranging from qualitative (Herz, Peters, & Truschkat, 2015) and mixed approaches (Froehlich, 2020) to quantitative ones (Snijders, Van de Bunt, & Steglich, 2010; Brouwer, Flache, Jansen, Hofman, & Steglich, 2018), some of which are presented in this volume. In this chapter, we will focus on the quantitative SNA approaches adopted for investigations of interactions over time. Although various SNA approaches can be applied to investigate interaction (see also in this volume Mejeh, 2020; Ryu, 2020; Froehlich & Van der Wilt, 2020), this chapter introduces an example of peer interactions within small group teaching to illustrate the method of analyzing longitudinal social networks. We discuss the implications of the different approaches for SNA findings and interpretations for the empirical investigation of (peer) interactions.

Methodological background

Most quantitative applications of SNA are rooted within the post-positivism paradigm, which assumes the existence of a manifest but merely stochastically comprehensible reality and aims at testing hypotheses (Guba & Lincoln, 1994). SNA may also be discussed under the banner of relational sociology, which promotes relational thinking over emphasizing individuals' (or other entities') attributes (Donati, 2010; Emirbayer, 1997).

One of the main mechanisms underlying interaction and network formation is homophily. Homophily refers to the likelihood that individuals are connecting to each other because they are similar to some degree in terms of characteristics or behavior (McPherson, Smith-Lovin, & Cook, 2001). This raises the question whether individuals either select others to interact with because they are similar and/or influence each other over time during their interaction. Selection and influence are two network processes that both can explain homophily. Selection means that individuals connect to each other based on certain characteristics (e.g., grades) and influence means that individuals become more similar over time because of their connections with others, who have certain characteristics (e.g., grades; Veenstra & Steglich, 2012).

To understand more about which process is predominant within a certain context and at a certain moment, we need to disentangle *selection* from *influence*. Investigation of longitudinally collected complete (full) social network data provides insight into selection and influence mechanisms. In complete networks, all network ties within a certain boundary (e.g., classroom) are measured (Steglich, Snijders, & Pearson, 2010). So far, various models have been developed for the analysis of longitudinal social networks (see Snijders, 2005). For disentangling selection from influence it is necessary that the evolution of networks within a continuous time frame is modeled given the changes in actor attributes (i.e., behavior) simultaneously (Ripley, Snijders, Boda, Vörös, & Preciado, 2016; Steglich et al., 2010). Stochastic actor-based modeling (SABM) is an inferential technique to test hypotheses in longitudinal data where the observations are interdependent. Actor attributes and ties are both independent and dependent variables when both changes are modeled (Snijders, 2001, 2005, 2011; Snijders et al., 2010; Steglich et al., 2010; Ripley et al., 2016).

SABM makes use of simulation. The model changes based on the perspective of the actor and that an actor makes a "decision" of initiating the changes of the ties, that is, maintaining, dissolving, or creating a tie (Ripley et al., 2016). It does not mean that this is an active decision, but the actor controls the outgoing ties based on the network position, attributes, and the perceived others in the network. This changing network is the result of a Markov process implying that the network state at a future time point can only be predicted probabilistically as a function of the current network state. Therefore, it is important to include all meaningful information as independent information (Snijders et al., 2010). The changes in a network take place when actors have the chance to change their outgoing ties or behavior at a certain point in time. This moment of change in the outgoing connections is captured with the *rate parameter*. The fundamental moment for change is the *ministep*, which means that not more than one behavior or tie variable of one actor can change. Changes of ties are decomposed in ministeps and modeled with a probabilistic function. The moment that an actor in a network modifies one of his connections may depend on the social network structure and actor attributes (see for more details Snijders, 2005). The likelihood of the network change in a certain way and from an actor's perspective is determined by the *objective function*, referring to the possible states of the social network, which are, in turn, dependent on the ties and actor and/ or dyadic covariates (attributes of one actor and/or between a pair of actors). An example of an objective function as a linear combination of elements is the tendency of similarity/ homophily expressed in the ties from the focal actor towards actors with the same grades (Snijders et al., 2010).

Applicability and requirements

The data to be collected needs to be in a relational format that is able to find interdependencies between the observed cases. One of the main challenges is dealing with missing data. Missing data of 10 percent can be acceptable; the model may converge and the estimates are not too much biased (Ripley et al., 2016). Also, consider ethical issues when doing (longitudinal) SNA. See for more information about ethical issues when applying SNA, Borgatti and Molina (2003), Korir, Mittelmeier, and Rienties (2020), Palonen and Froehlich (2020). Quantitative longitudinal network analysis is also dependent on your capabilities in applying statistical and mathematical functions. The algorithms of analysis presented in what follows are relatively complex and under current development. As such, you need to be able (and confident!) to navigate in an environment where ready-made solutions are potentially not (yet) available.

The software you need for this analysis is R (R Core Team, 2014). R is freely available and is supported across operating systems. A disadvantage of using R is that the researcher needs to learn the script language, which can be time-consuming depending on previous program experience. For stochastic actor-based modeling should be conducted with the R package SIENA (Simulation Investigation for Empirical Network Data; Ripley et al., 2016). SIENA is a simulation method that allows researchers to investigate the co-evolution of social network structures and actor behavior over time (Veenstra & Steglich, 2012).

Workflow

Step 1: Check whether your research question aligns with the method

Research questions addressed with quantitative longitudinal SNA emphasize mutual impact of changing networks (i.e., relationships or interactions) and changing personal attributes. Research questions differ from cross-sectional social network data, since one time point cannot inform us about change and, hence, influence and selection effects. The advanced techniques for analyzing social network data provide unique insights in changes in relationships or interactions over time given the changes in the social network positions and behavioral attributes (e.g., academic performance, smoking). Actors can become central over time by network-related structural processes but also because of their personal characteristics. For instance, consider the research questions posed by Van Duijn et al. (2003, p. 155): "What kinds of individual and network variables explain changes over time within a friendship network?" The question is at what stages and why these variables are important. Such research questions can be reformulated in hypotheses that are testable with quantitative complete network data.

In the illustrative example, Brouwer et al. (2018) investigated whether students select each other based on their grades and addressed the following research question: "With whom do freshmen connect when they need study related support during their first academic year?" They controlled for friendship, because friends might be more willing to help. Based on the homophily principle (McPherson et al., 2001) it is likely that friends achieve similarly. The next research question was: "Do they ask more often a similar-achieving friend or a higher-achieving fellow student who is not a friend?" Another example of a research question about school networks, focusing on selection and influence, is: "To what degree can influence and selection account for the co-evolution of substance use and friendship ties?" (Steglich et al., 2010, p. 363). Research questions like these are addressed by applying stochastic actor-based modeling (Snijders et al., 2010).

Step 2: Define your network's boundary

Longitudinal social network research generally use clearly predefined boundaries by the researcher (Wasserman & Faust, 1994). For disentangling selection from influence, networks need to have specified boundaries. The researcher needs to define the network boundary prior to the start of the research project. The boundary specification determines which actors are approached and included as participants. Why is it so important to collect quantitative data of a complete network given the predefined network boundary? When a researcher wants to disentangle selection from influence effects, insight into whom is nominating whom (selection) but also who is not nominated by the other participant (non-selection) is needed. Often, these complete networks are collected with surveys and to a lesser extent with interviews (see Wasserman & Faust, 1994).

In the illustrative example, Brouwer et al. (2018) defined their boundary as follows. They investigated learning communities, but they were also interested in help-seeking relationships and friendships beyond the learning communities. Therefore, they defined the boundary as one cohort of students that belong to one faculty where they can all meet (and nominate) each other. The assigned learning community was included as an attribute variable.

Step 3: Safeguard high response rates

As discussed previously, the problem of missing data is particularly serious in social network research, because the study outcomes can be biased if important ties are missing (Ripley et al., 2016). The main reasons for missing data is non-response because they feel that the questions are too personal or attrition of the study (Borgatti et al., 2018). To prevent missing data, participants should be informed about the importance of complete network data and confidentiality should be safeguarded. A small amount of missing data (<10 percent) will not result in biased results. See for more information Huisman and Snijders (2003) and Ripley et al. (2016).

In the illustrative example, Brouwer et al. (2018) had a high response rate of 90 percent in their study. This may be the result of intensive contact with the participants during the study. Make sure to inform them very carefully about the added value of longitudinal SNA, about the aim, and the duration of the study. It is also important to inform participants about preliminary results and eventually reward them with a voucher (or another small present). The researcher should be active and available when participants email their questions or concerns about privacy or other issues. Ensure that the survey is not too long – not more than four social network questions.

Step 4: Have a timely follow-up

Longitudinal data measure a social networks in at least two waves, but it is preferred that more time points are measured. The measurement points should be not too far apart from each other. However, the time gap between the first and final measurement must be large enough to provide insight in the changes of a network (Snijders, 2009). This implies that network data are collected over time but also individual characteristics, attitudes, motivation, among others, such as personal attributes and depending on the focus of the research.

Step 5: Organizing and analyzing the data

The models are estimated with the data-analysis package RSIENA (Simulation Investigation for Empirical Network Analysis) and allows us to test binary complete network data including existing ties (coded as 1) or no ties (coded as 0). The social networks data set should therefore be dichotomized in zeros (non-tie) and ones (tie) and stored in adjacency matrices. The behavior data can be stored with one row for each participant consisting of the dichotomized or ordinal values for the attributes (e.g., attitudes, motivation, grades). The network data set and the attribute data set should be separately imported in R. Always check whether your network and attribute files consist of the same number of actors and whether the nodes are in the same order in attributes file as in the rows and columns of the network matrix. See Ripley et al. (2016) for how to create the SIENA objects in R.

The data sets can be first analyzed in a descriptive way by making graphs of each time point and by calculating network descriptives, such as in- and outdegree, density, and reciprocity. The changing peer networks in a complete program and changing levels of the attributes can be jointly modeled in a co-evolution model. In a co-evolution model, we test selection and influence effects by including endogenous structural social network effects and exogenous cross-network effects and covariates (such as measures of learning performance). The analyses produce parameter estimates of the dynamics of social networks (structural network and attribute-dependent selection effects) and behavior (behavior trends and influence effects; Ripley et al., 2016). Check prior to the analysis whether enough stability exists to see whether the SIENA method is appropriate for the data set. This can be done by calculating the Jaccard index. As a rule of thumb, Jaccard values of .30 or higher are good, whereas lower than .20 the estimation of the models might be problematic. The model specification needs to be done based on theory and prior knowledge (see Ripley et al., 2016). The structural effects, such as outdegree (tendency to form outgoing ties), reciprocity (mutual ties), and (at least one of the) transitivity parameters (two actors have a shared common tie) are always included. Think also about the specification of control variables (e.g., age, gender), degree-based parameters related to the actor covariates (e.g., indegree achievement; outdegree achievement effect), and dyadic covariate effects (same achievement effect; Snijders, 2005; Snijders et al., 2010). See Veenstra and Steglich (2012) for a more detailed explanation of a specification of a selection and influence model. For the networks, the actor-based model gives information about tie creation, maintenance, and termination of ties. In case of a co-evolution model instead of merely a selection model, the behavior in the actor-based model gives information about increasing, decreasing, or maintaining a certain level of behavior (e.g., achievement, smoking). Exogeneous individual and dyadic covariates are not modeled but can be used as an explanation of network or behavioral changes. After the modeling, the researcher checks whether the model converges. As a rule of thumb, the t-ratios for convergence should be less than 0.1 and the overall maximum convergence ratio of the model should be less than 0.25 (although these are not strict rules). When this is not the case, re-running the analysis is one of the possibilities. Check the RSIENA manual (Ripley et al., 2016) for more options. The t-ratios of the estimates are based on an approximate normal distribution. The estimate is divided by the standard error to get the t-value and belonging p-value to get informed about the significance level (see for more details Snijders, 2005; Snijders et al., 2010). After interpreting the results of the final model, the researcher describes the results.

In the illustrative example showcased in this chapter, Brouwer et al. (2018) tested in two steps whether students selected other fellow students for academic help based on friendship, preference for collaboration, and grades, while controlling for gender and self-efficacy. The hypothesis is that students ask for academic help from their similar-achieving friends instead of their non-similar achieving fellow students. The final model showed that when students are friends, they are more likely to ask academic help from each other. This can be derived from the positive friendship effect on academic help networks (b = 0.93; SE = 0.22). Positive effects are also found for achievement ego (b = 0.41; SE = 0.12) and achievement similarity (b = 2.10; SE = 0.10). This means that the higher a student achieves, it is more likely that he or she will ask others for academic support but also that students are more likely to approach each other for academic help when they have similar levels of achievement.

Strengths and weaknesses

There are few alternatives to doing relational or structural analysis across multiple time points, especially when the benefits of quantitative research (e.g., handling of big volumes of data) are useful for a particular research project. When it comes to model testing, the longitudinal aspect is especially helpful for making more confident conclusions about temporal causality. The advantages of these models are that they assume a continuous time frame between the different discrete measurements, take into account interdependencies between observations, and control for the structural effects (Veenstra & Steglich, 2012).

Despite these advantages, all the limitations of any longitudinal approach are applicable when studying co-evolution models of longitudinally measured interactions. This includes, for instance, attrition among study participants (which is especially important given SNA's high affordance in terms of network coverage) or questions about the timing of measurements: What is the "right" lag between the measurements points in relation to the effect that is being studied? What is a meaningful start and end date? What is the meaning of time, to begin with (Pettigrew, 1990)?

There are some specific limitations of this approach. First, as stated previously, the method itself is still under rigorous development. This means that not all procedures are fully in place. This means that not all procedures are fully in place yet (see Niezink, 2018). Second, some assumptions of these models do not fully represent reality. For example, the assumption that actors act independent from the other actors in the network. Therefore, it is difficult to investigate group phenomena and collective action with the analysis is RSIENA. Furthermore, you conduct this type of analysis in a complete network where all actors can nominate each other, for example friends outside the university (Baerveldt, Völker, & Van Rossem, 2008). In reality, an actor will also select people outside this network. Fourth, the analysis might be complicated for a beginning network researcher. However, many R packages are already available, as well as introductory material, and online support tools. Researchers who start with these types of analysis can also connect to other experts in the field and ask for support directly and/ or contact the authors of this book chapter.

Further reading

Ripley, R. M., Snijders, T. A. B., Boda, Z., Vörös, A., & Preciado, P. (2016). *Manual for RSiena*. Oxford: University of Oxford, Department of Statistics, Nuffield College.

Snijders, T., Steglich, C., & Schweinberger, M. (2017). Modeling the coevolution of networks and behavior. In K. Monfort, J. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences* (pp. 41–71). London: Routledge.

References

Baerveldt, C., Völker, B., & Van Rossem, R. (2008). Revisiting selection and influence: An inquiry into the friendship networks of high school students and their association with delinquency. *Canadian Journal of Criminology and Criminal Justice*, *50*(5), 559–587.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). *Analyzing social networks* (2nd ed.). London: Sage.

Borgatti, S. P., & Molina, J. L. (2003). Ethical and strategic issues in organizational social network analysis. *Journal of Applied Behavioral Science*, *39*(3), 337–349.

Brouwer, J., Flache, A., Jansen, E., Hofman, A., & Steglich, C. (2018). Emergent achievement segregation in freshmen learning community networks. *High Educ, 76*, 483–500. https://doi.org/10.1007/s10734-017-0221-2

Brouwer, J., Jansen, E. P. W. A., Flache, A., & Hofman, W. H. A. (2018). Longitudinal peer network data in higher education. In J. Huisman & M. Tight (Eds.), *Theory and method in higher education research* (Vol. 4, pp. 145–162). Bingley: Emerald Publishing Limited. ISBN 978-1-78769-278-7/ doi: 10.1108/S2056-375220180000004010

Donati, P. (2010). *Relational sociology: A new paradigm for the social sciences*. London: Routledge.

Emirbayer, M. (1997). Manifesto for a relational sociology. *The American Journal of Sociology*, *103*(2), 281–317.

Froehlich, D. E. (2020). Exploring social relationships in "a mixed way": Mixed structural analysis. In D. E. Froehlich, M. Rehm, & B. C. Rienties (Eds.), *Mixed methods social network analysis: Theories and methodologies in learning and education*. London: Routledge.

Froehlich, D. E., & Van der Wilt, F. (2020). Mixed structural group interaction analysis. In M. Huber & D. E. Froehlich (Eds.), *Analyzing group interactions: A researcher's guidebook* (pp. 116–123). London: Routledge.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). Thousand Oaks, CA: Sage.

Herz, A., Peters, L., & Truschkat, I. (2015). How to do qualitative structural analysis: The qualitative interpretation of network maps and narrative interviews. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, *16*(1).

Huisman, M. E., & Snijders, T. A. B. (2003). Statistical analysis of longitudinal network data with changing composition. *Sociological Methods & Research*, *32*(2), 253–287.

Korir, M., Mittelmeier, J., & Rienties, B. C. (2020). Is mixed methods social network analysis ethical? In D. E. Froehlich, M. Rehm, & B. C. Rienties (Eds.), *Mixed methods social network analysis: Theories and methodologies in learning and education*. London: Routledge.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, *27*(1), 415–444.

Mejeh, M. (2020). Cross-sectional social network analysis. In M. Huber & D. E. Froehlich (Eds.), *Analyzing group interactions: A researcher's guidebook* (pp. 85–95). London: Routledge.

Palonen, T., & Froehlich, D. E. (2020). Mixed methods social network analysis to drive organizational development. In D. E. Froehlich, M. Rehm, & B. C. Rienties (Eds.), *Mixed methods social network analysis: Theories and methodologies in learning and education*. London: Routledge.

Pettigrew, A. M. (1990). Longitudinal field research on change: Theory and practice. *Organization Science*, *1*(3), 267–292.

R Core Team. (2014). *R: A language and environment for statistical computing*. Retrieved from www.r-project.org

Ripley, R. M., Snijders, T. A. B., Boda, Z., Vörös, A., & Preciado, P. (2016). *Manual for RSiena*. Oxford: University of Oxford, Department of Statistics, Nuffield College. Retrieved from www.stats.ox.ac.uk/~snijders/siena/RSiena_Manual.pdf

Ryu, S. (2020). The combination of critical discourse analysis and social network analysis. In M. Huber & D. E. Froehlich (Eds.), *Analyzing group interactions: A researcher's guidebook* (pp. 96–106). London: Routledge.

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, *31*, 361–395.

Snijders, T. A. B. (2005). Models for longitudinal data. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 215–247). New York: Cambridge University Press.

Snijders, T. A. B. (2009). Longitudinal methods of network analysis. In R. Meyers (Ed.), *Encyclopedia of complexity and systems science* (pp. 5998–6013). New York: Springer.

Snijders, T. A. B. (2011). Statistical models for social networks. *Annual Review of Sociology*, *37*, 131–153.

Snijders, T. A. B., Van de Bunt, G. G., & Steglich, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, *32*, 44–60.

Steglich, C., Snijders, T. A. B., & Pearson, M. (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, *40*(1), 329–393.

Van Duijn, M. A., Zeggelink, E. P., Huisman, M., Stokman, F. N., & Wasseur, F. W. (2003). Evolution of sociology freshmen into a friendship network. *Journal of Mathematical Sociology*, *27*(2–3), 153–191.

Veenstra, R., & Steglich, C. (2012). Actor-based model for network and behavior dynamics. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 598–618). New York: The Guilford Press.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.