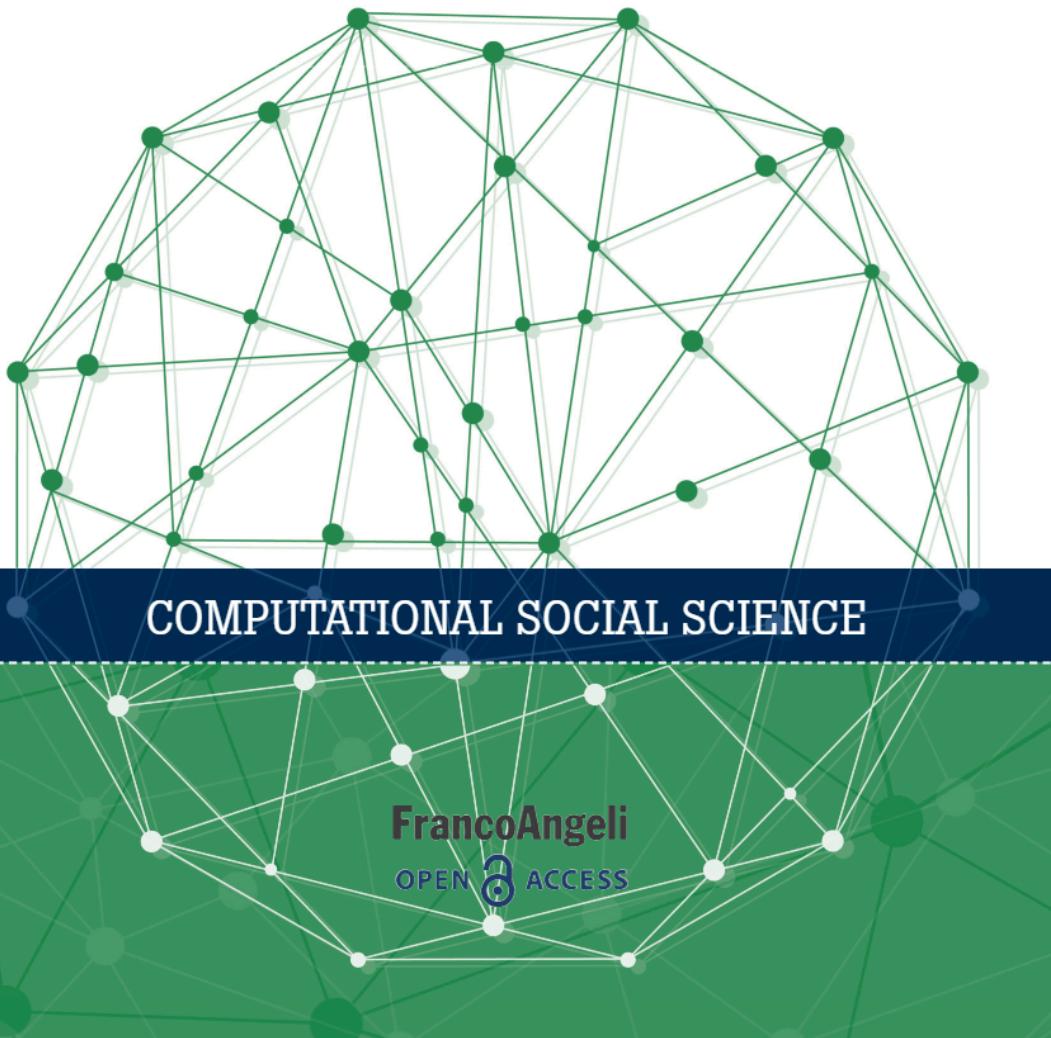


Edited by Giuseppe Giordano,
Marialuisa Restaino, Andrea Salvini



METHODS AND APPLICATIONS IN SOCIAL NETWORKS ANALYSIS

Evidence from Collaborative,
Governance, Historical
and Mobility Networks



Edited by Giuseppe Giordano,
Marialuisa Restaino, Andrea Salvini

METHODS AND APPLICATIONS IN SOCIAL NETWORKS ANALYSIS

Evidence from Collaborative,
Governance, Historical
and Mobility Networks

COMPUTATIONAL SOCIAL SCIENCE

FrancoAngeli
OPEN  ACCESS

*2. Examining the association between behavioral trajectory similarity clusters in social networks**

by Brandon Sepulvado**, Omar Lizardo***, Mike Wood****, Cheng Wang*****^{*}, David Hachen*****

1. Introduction

The increasing availability of fine-grained temporal data on human interaction in social networks (Bahulkar *et al.*, 2017; Kim & Anderson, 2012; Liu *et al.*, 2018; Miritello, 2013; Sekara *et al.*, 2016), augmented with equally fine-grained information on attitudes, habits, and practices collected via unobtrusive means (Purta *et al.*, 2016; Purta & Striegel, 2019), has opened up new opportunities to study the link between social networks and human behavior (Lazer *et al.*, 2009). For social networks, rather than thinking of pairwise interactions in the model of a static graph, there is now an emphasis on temporal dynamics and developing temporal versions of quantitative network structure (Dickison *et al.*, 2016; Holme & Saramäki, 2012). In the study of human behavior, there is a renewed emphasis on methods attentive to dynamic change, such as event-history and time-series analysis (Epskamp, 2020).

* Research reported in this paper was supported in full by the National Heart, Lung and Blood Institute of the National Institutes of Health under award number R01HL117757 (\$2,913,061). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

** Department of Methodology and Quantitative Social Sciences, NORC at the University of Chicago, 4350 East-West Highway, 8th Floor, Bethesda, MD 20814, USA, sepulvado-brandon@norc.org.

*** Department of Sociology, University of California, Los Angeles, 375 Portola Plaza, 264 Haines Hall, Los Angeles, CA 90095, USA, olizardo@soc.ucla.edu.

**** Department of Sociology, Brigham Young University, 2008 JFSB, Provo, UT 84602, USA, michaelwood@byu.edu.

***** Department of Sociology, 2257 Faculty Administration Building, Wayne State University, Detroit, MI 48202, USA, chengwang@wayne.edu.

***** Department of Sociology, University of Notre Dame, 4076 Jenkins Nanovic Hall, Notre Dame, IN 46556, USA, David.S.Hachen.1@nd.edu.

Our recent work has attempted to join these lines of work by taking the classic social network concept of dyadic behavioral similarity and looking at it from a temporal perspective that emphasizes dynamic trajectories (Sepulvado *et al.*, forthcoming). We depart from most prior work on similarity, which has looked at it statically with the concept of homophily. There has been a great deal of research attempting to both predict various types of homophily and assess its impact on network outcomes (Aral *et al.*, 2009; Kossinets & Watts, 2007; Ingram & Morris, 2007; Yang *et al.*, 2019; Lewis & Kaufman, 2018; Leszczensky & Pink, 2019). However, there has been very little work looking at how changes in similarity predict tie formation, and how tie formation results in changes in the similarity between two persons in their behaviors. In this respect, the “dynamic” turn toward the study of dyads in general and homophily in particular advocated by Rivera *et al.* (2010) about ten years ago has yet come to pass (but see Bahulkar *et al.*, 2017; Dokuka *et al.*, 2016; Schaefer & Kreager, 2020; Wang *et al.*, 2020).

2. Background

In an earlier paper (Sepulvado *et al.*, forthcoming), we developed the idea of dyadic similarity trajectories and proposed methods for assigning dyads into discrete, substantively meaningful clusters that help predict dyadic connectivity and dyadic matching (homophily) on key sociodemographic traits. This paper uses those methods to explore the predictive linkage between two similarity trajectories, one based on physical activity and the other on social activity.

The larger goal of this line of work is to model the association between various similarity trajectories in order to improve our understanding of how social ties in social networks form, evolve, and decay. In this paper, we focus on dyadic similarity trajectories based on similarity in the two vertices’ network position as measured by their *temporal outdegree*, the number of people that they communicate with per unit of time and an indicator of how socially active a person is in temporal social networks (Holme & Saramäki, 2012).

Of particular interest is the correspondence between trajectories based on similarity in network position and trajectories based on similarity in behavior; in our case, physical activity is measured by daily step counts. We use these two measures of similarity, how active people are physically and socially, to ascertain the extent to which similarity in outdegree is associated with both the probability of a tie and similarity in physical activity. Insights gained by analyzing the association between similarity in social and physical

activity will enhance our understanding of how behavioral similarity trajectories can help predict tie formation, evolution, and persistence.

We use data from the longitudinal *NetHealth* Study, which collected network, behavioral, and attitudinal data from smartphones, activity trackers, and surveys. From this rich-attribute, fine-grained longitudinal social network and behavior dataset, we generate dyadic communication similarity trajectories using daily counts of the number of outgoing (person-initiated) communication events obtained from smartphone logs. We compute daily similarity scores for each dyad, which we use to create a final “second-order” dyadic communication similarity trajectory. We then use clustering techniques sensitive to the temporal (i.e., time-series) structure of the data to generate dyadic clusters, showing that the clusters help predict dyadic connectivity. Finally, we look at the statistical association between dyadic clusters based on communication trajectories with clusters based on physical activity trajectories, and we ask whether this association is moderated by dyadic matching in sociodemographic traits.

2.1. Key Concepts

We begin by defining some key terms. This includes the network notion of a *dyad*, the idea of a *behavioral trajectory*, the concept of *behavioral trajectory similarity clusters* defined over dyads, and the *association* between distinct behavioral trajectory similarity clusters when considering two or more behaviors, which may include endogenous communicative activity in social networks. In the discussion, we elaborate on how our approach can be extended to other behavioral and non-behavioral temporal patterns of change and higher-order network motifs beyond dyads.

2.2. Dyads

We begin with the classic social network concept of the *dyad* (Wasserman & Faust, 1994, p. 505-ff). In social networks represented as a graph G with a set of undirected links E , a set of vertices V , and associated symmetric adjacency matrix \mathbf{A} , a dyad is defined as all unordered pair of actors in the system (in directed networks, the pairs are ordered); each cell (a_{ij}) in either the upper or lower triangle of the symmetric adjacency matrix ($a_{ij} = a_{ji}, \forall i, j \in V$) refers to a dyad in the network. In the simplest case of an undirected network, there are two mutually exclusive types of dyads: *connected* (in which case $a_{ij} = 1$)

and *null* (in which case $a_{ij} = 0$). Connected dyads are joined by an edge (link or tie) in the social network, whereas null dyads do not share an edge.

In the temporal network case, dyads must be indexed at each timepoint because actors may join or leave the system and because dyads may transition from one state (e.g., similar) to another (e.g., dissimilar) or from being connected to null (and back). Accordingly, the question of whether two dyads are similar on a given trait must be separated from whether they have a direct link in the temporal network. More formally, at any point in time (t), an edge between two actors, i and j , may or may not exist, which can be written, using matrix notation, as $a_{ijt} = 1$ if the edge exists at time t , $a_{ijt} = 0$ if the edge does not exist at time t . A dyad can be similar or dissimilar on a given time-varying trait and be either connected or disconnected in the temporal network. Whether connected dyads are more behaviorally similar is an empirical question.

2.3. Actor-Level Behavioral Trajectories

Each actor in a temporal social network is observed at multiple points in time. This means that information obtained on behaviors, traits, and habits of the actor can also dynamically change over the observation period (Christakis & Fowler, 2007; Fowler & Christakis, 2008; Lazer *et al.*, 2009; Lewis *et al.*, 2008). Consider the value of a given trait or behavior s , observed for actor i at time t . The time series of values $s_i = \{s_{it(1)}, s_{it(2)}, s_{it(3)} \dots s_{it(m)}\}$ across all time points m defines a *behavioral trajectory* for that actor on that trait. For instance, if the trait is something like physical activity, such a trajectory might indicate that an actor's physical activity level might increase, decrease, or vacillate between the two. In essence, behavioral trajectories can take multiple functional forms with respect to time (Sepulvado *et al.*, forthcoming).

2.4. Dyad-Level Behavioral Trajectory Similarity

Since each actor in the network has a behavioral trajectory, each dyad can be more or less similar on that trait at each point in time. More formally, for each pair of actors in a temporal network i and j , with behavioral trajectories on trait s , denoted by s_i and s_j , we can define the dyadic behavioral trajectory similarity at each point in time s_{ijt} as the absolute value of the difference between the value of the trait for each actor at that time, or $|s_{it} - s_{jt}|$. Note that just like for each behavioral trajectory, the dyadic similarity trajectory

defines a time-series over each observed time point for each dyad in the network, $s_{ij} = \{s_{ijt}(1), s_{ijt}(2), s_{ijt}(3) \dots s_{ijt}(m)\}$.

This time series may itself display substantively meaningful behavior. For instance, a dyad may become more similar or dissimilar on the observed trait or display fluctuating similarity and dissimilarity patterns over time. Dyadic behavioral trajectory similarities and dissimilarities may also be affected by a variety of dyad-level factors, inclusive of dyadic similarity on other (both time-varying and -invariant) traits, such as whether the two actors in the dyad have similar value on sociodemographic attributes (e.g., race, religion) or environmental factors (e.g., propinquity) relevant to connectivity in social networks (Feld, 1982; McPherson *et al.*, 2001; Rivera *et al.*, 2010; Sepulvado *et al.*, forthcoming).

We can link the dynamics of dyadic behavioral trajectory similarity with other dynamics in a social network, most importantly the dynamic *transition* of dyads from null to connected (or vice versa) and, in directed networks, dynamic transitions of dyads from null, to asymmetric, to mutual (Wasserman & Faust, 1994, p. 505-ff). This can advance core issues in social network analysis, such as whether similarities precede connectivity (transition of dyad from a null to a connected state), whether connectivity is a causal input into increasing similarity, and whether increasing dissimilarity leads to disconnectivity (transition of a dyad from a connected to null state) or vice versa (Bahulkar *et al.*, 2017; Lewis & Kaufman, 2018; Noel & Nyhan, 2011; Schaefer & Kraeger, 2020).

2.5. Dyad-Level Behavioral Trajectory Similarity Clusters

The total possible number of dyads in a network increases super-linearly in the number of actors. As is well-known, for an undirected network represented by a graph of order n , there are $\frac{1}{2}n(n-1)$ possible dyads (Wasserman & Faust, 1994, p. 515). This means that the number of dyadic similarity trajectories to be considered will similarly increase. When examining dyadic behavioral similarity trajectories in temporal networks, it is desirable to look for a way to cluster dyads into a smaller set of classes. The idea is to assign dyads to the same class when they have similar (e.g., substantively the same except for measurement error and small statistical deviations) behavioral similarity trajectories (e.g., all the dyads in the same class become more similar over time). Such a possible set of dyadic classes or clusters in social networks have been called *behavioral trajectory similarity clusters* (Sepulvado *et al.*, forthcoming).

Although previous studies have examined temporal changes in tie characteristics (Krackhardt & Hancock, 2006; Martin & Yeung, 2006; Morgan *et al.*, 1997; Schaefer & Kraeger, 2020) or the emergence of collective similarity due to the diffusion of contagions (Centola, 2010; Christakis & Fowler, 2007, 2008, 2013; Fowler & Christakis, 2008b, 2008a), the concept of behavioral similarity trajectory clusters solves two fundamental problems (Sepulvado *et al.*, forthcoming). First, many studies researching similarity dynamics in social networks condition on a dyad already being connected (Fowler & Christakis, 2008b, 2008a), leading to biased conclusions about how ties impact dyadic similarity (e.g., Cohen-Cole & Fletcher, 2008). Second, similarity on behavioral and/or attitudinal traits is often considered at only a single time point (e.g., Aral *et al.*, 2009; Feiler & Kleinbaum, 2015; Lewis *et al.*, 2012), yet this stasis is not a valid assumption for many individual characteristics. For example, political attitudes can change, and behavioral patterns evolve with life events (Bidart & Lavenu, 2005), and as seasons change. Beyond addressing common issues with network analysis studies, such trajectory clusters enable theoretical innovation.

In this paper, we extend our previous work in two significant ways. First, we extend the notion of behavioral trajectory similarity from exogenous (e.g., physical activity) to *endogenous* social network traits. Namely, we use the *communicative activity* level to define a behavioral temporal degree trajectory for each actor in the network (Miritello, 2013; Miritello *et al.*, 2013; Raeder *et al.*, 2011). In a temporal network, each actor i may communicate (send ties) to a set of actors k at time t . The quantity $k(i)_t$ defines the level of communication of that person at that time, a measure of temporal network centrality referred to as *temporal outdegree* (Kim & Anderson, 2012). This measure of centrality is an indicator of a person's sociability and varies both between persons (in time-averaged slices) and within persons over time.

Absolute temporal degree *differences* between the members of the dyad in the network ($s_{ijt} = |k(i)_t - k(j)_t|$) thus define a set of behavioral trajectory similarity series based upon communication for each dyad. These, in turn, can be used to assign each dyad to a behavioral trajectory similarity cluster, as defined earlier. The second way in which we extend our previous work is by looking at multiple behavioral trajectory similarities in the same analysis. To that end, we construct behavioral trajectory similarity clusters based on an exogenous trait (daily step counts) and an endogenous trait (temporal degree). We then look at whether these two dyadic class assignments are statistically dependent upon one another using methods to detect a statistical association in categorical (polytomous) variables (Powers & Xie, 2008). In this way, we can examine the linkage between multiple dyadic behavioral trajectory similarity clusters.

3. Data

We use data from a study called NetHealth that followed a cohort of 625 undergraduates at the University of Notre Dame (Purta *et al.*, 2016). To examine the relationship between health behaviors, communication activities, social networks, and other actor-level traits, students were equipped with smartphones and activity trackers worn on the wrist (i.e., Fitbits) and surveyed before matriculation and every semester after that. Although the entire study period was from Fall 2015 to Spring 2019, only the Fall 2015 data are included in these analyses. We exclude data before 1 September 2015 because students were still picking up their Fitbit devices.

Communication data come from smartphones, specifically “metadata” from calls, text messages, and WhatsApp messages. The metadata includes timestamped information of communication events and the numbers of the caller/sender and receiver. We use the data to construct a temporally aggregated social network among study participants using a threshold: treating an edge as existing if there was any communication attempt in the period between any two pairs of actors i and j . For physical activity data, we use daily steps as measured by the Fitbits.

The communication and physical activity data are used to create distinct trajectories: first at the actor level and then for dyads. Each participant has a number of initiated communication events for each day in the 109-day Fall 2015 period. This daily communication activity is directed. We use this information to compute each person’s temporal (daily) degree in the network, defined as the number of other actors in the network contacted at time t . Similarly, each person has a daily total number of steps for each day in the period. Actor-level behavioral trajectory time series are constructed from both daily communications and steps. To derive dyadic similarity trajectories, we compute the absolute daily difference in degree and steps for each dyad.

There are 195,000 possible dyads in the NetHealth Study network, but not all of these are included in analyses. If at least one of the actors in a dyad had missing data for a given day, we treated the dyad’s steps as missing for that day. Study participants could have no missing data on communication activity because days without communication were coded as zero. We excluded dyads that were missing over 25% of their daily step differences, and, for those dyads with up to 25% missing data, conducted linear interpolation. After filtering out missing data, 32,872 dyads remain.

Participants completed surveys collecting demographic information used in the analysis. We use this information to create three binary variables indicating whether both actors in a dyad have the same responses for race/ethnicity.

ty, gender, and religious identity. The final sample of 282 participants includes 156 male, 126 female students, 192 students identifying as white, 15 students identifying as Black, 25 identifying as Asian, 36 identifying as Latino, one student who did not identify with any of the racial identities, and 13 who were born outside the U.S. The sample includes 211 Catholics, 36 Protestants, 10 students identifying with “Other Religion” and 25 identifying as not religious. 16,563 dyads identify as the same gender, 16,602 identify with the same race/ethnicity, and 19,359 identify with the same religious identity.

4. Methods

To identify dyadic *classes* of (dis)similarity trajectories, we turn to unsupervised learning. Yet, given that we start with trajectories (i.e., time series) of 109 daily differences in steps and communication activity for all dyads, most traditional clustering methods are unsuitable. We thus rely upon time series clustering methods. There are three general approaches within this class of clustering methods: clustering on specific time points, clustering on subsequences of observations, or clustering the entire time-series (Aghabozorgi *et al.*, 2015, pp. 18-19). We choose the third strategy because we are interested in the evolution of dyadic similarity across the full Fall 2015 semester.

Rather than clustering using model-based methods (e.g., clustering on output from ARIMA models) or feature-based methods (e.g., clustering lower-dimensional representations of time series), we use a shape-based method that clusters multiple time series based upon the similarity of their shapes because previous evaluations have demonstrated that shape-based methods exhibit superior performance (Aghabozorgi *et al.*, 2015, pp. 18-19; Liao, 2005; Paparrizos & Gravano, 2017). Because it outperforms various alternatives for time series data (Paparrizos & Gravano, 2015, 2017), we choose the *k*-shape algorithm. *k*-shape clustering is based on the *k*-means algorithm and considers multiple known distortions that are frequently present in time-series data (Paparrizos & Gravano, 2015, p. 1859). For example, time-series’ features might be scaled differently (i.e., scaling invariance), and specific subsequences of the two time-series might have a similar shape while other subsequences considerably diverge (i.e., shift invariance).

We assess the meaningfulness and validity of the dyadic trajectory similarity clusters, based on temporal degree and obtained using the method described above in two ways. First, we examine whether dyads assigned to the same temporal degree trajectory similarity class are more or less likely to transition from the null to the connected state during the observation period.

For this analysis, we specify a model predicting a dyad's probability of being connected from the assigned similarity trajectory cluster membership and compute predicted probabilities of having a tie for each cluster. We use logistic regression with Firth (1992) penalization to account for any bias produced by the "rare-event" nature of the dyadic outcome (there is a larger class imbalance, with null dyads outnumbering connected dyads by a factor of 49 to 1). Additionally, we tested a Firth-penalized model with intercept correction (not presented; Puhr *et al.*, 2017), which did not substantively change model performance. The Firth logistic regression treats whether a dyad has a tie (i.e., 1 if the two individuals communicated during the Fall 2015 semester) as the dependent variable and cluster membership as the dependent variables.

Second, and this is the crucial innovation introduced in the paper, we examine the question of whether dyadic trajectory similarity classes based on one (endogenous) trait (i.e., temporal degree) are associated with dyadic trajectory similarity classes based on another (exogenous) trait (daily steps). We use log-linear models to examine the association between cluster membership along these two dyadic trajectory similarity dimensions (Powers & Xie, 2008). Additionally, we examine the extent to which any association between the two dyadic clusters is due to dyadic matching on time-invariant sociodemographic factors, which has been called "homophily" in the social networks literature (McPherson *et al.*, 2001; Rivera *et al.*, 2010). To do this, we use log-linear models of the three-way association between dyadic temporal degree and step similarity trajectory cluster, and three types of sociodemographic homophily – gender, race, and religion. These models allow us to ascertain the extent to which the association between step and communication activity clusters is partially the result of the association between demographic similarity and both step- and communication-based dyadic similarity trajectory cluster membership.

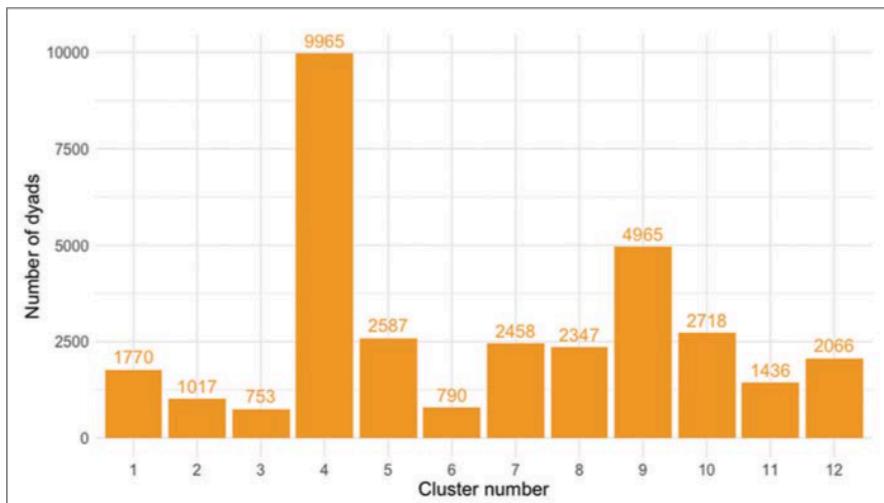
We first estimate a Poisson model using the steps and communication clusters to predict cell count (i.e., the number of dyads falling into each combination) and identify systematic associations by examining which cells have high residuals (i.e., where cell count differs from what would be expected by chance). Next, we run a series of nested log-linear models (Powers & Xie, 2008) to examine the three-way association between demographic homophily measures (same gender, same race, or same religion), physical activity cluster membership, and temporal degree cluster membership. The baseline (independence) model fits the marginal distributions for the steps clusters, the communication clusters, and the homophily measure. The second model fits parameters for each homophily factor's association with the steps and the communication clusters. The final model fits the two-way marginal between step and temporal degree cluster membership. These models regress

cell count on different specifications of the step trajectory clusters, communication trajectory clusters, and each sociodemographic variable; each model's exact specification may be found in the "Equation" column of Table 1.

5. Results

The bar plot in Figure 1 shows the sizes of the clusters derived from outgoing communication, while Figure 2 presents the typical trajectories for each dyadic class identified by the k -shape clustering algorithm. The line in each plot in Figure 2 represents the centroid trajectory of a cluster. The x -axes indicate the day (with month labels), and the y -axes indicate the z-normalized value for the absolute difference in temporal degree for all the dyads in that class on that day. Recall that each dyad trajectory is a time series that was initially the absolute difference in daily activity for the actors in a dyad.

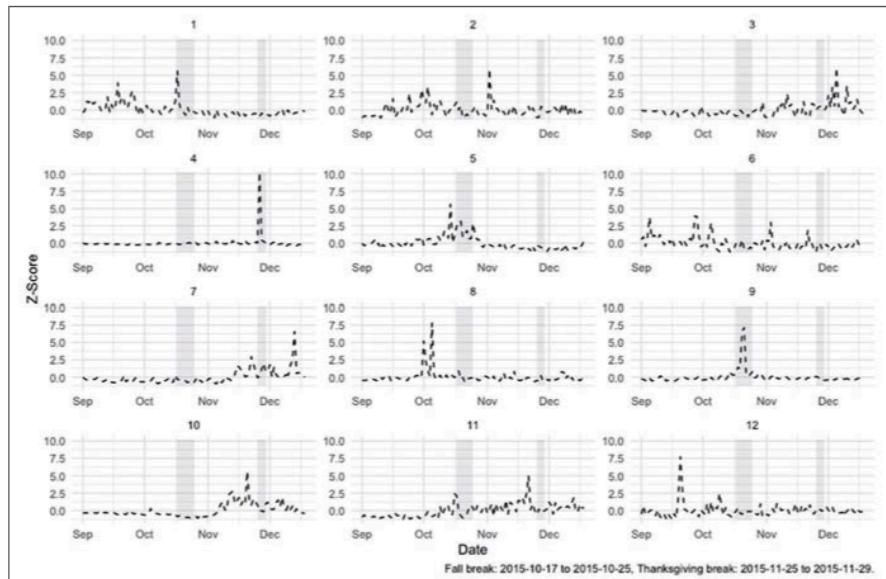
Fig. 1 – Sizes of degree-based clusters



There are two clusters (4 and 9) with a constant standardized difference score of zero across all days except for one large peak. Each peak occurs during a holiday break period where students typically leave campus and visit home, as noted by the figures' shaded area (Fall Break and Thanksgiving, respectively). These are the two largest clusters, containing 14,930 dyads (almost 50%). The peaks indicate periods when there is a great deal more communication heterogeneity in the population, which results in dyads between people with sub-

stantially different activity levels. There are also peaks evident in other cluster profiles (e.g., 8 and 12), but the z-score is not flat outside those peak periods.

Fig. 2 – Cluster centroids for degree-based clusters

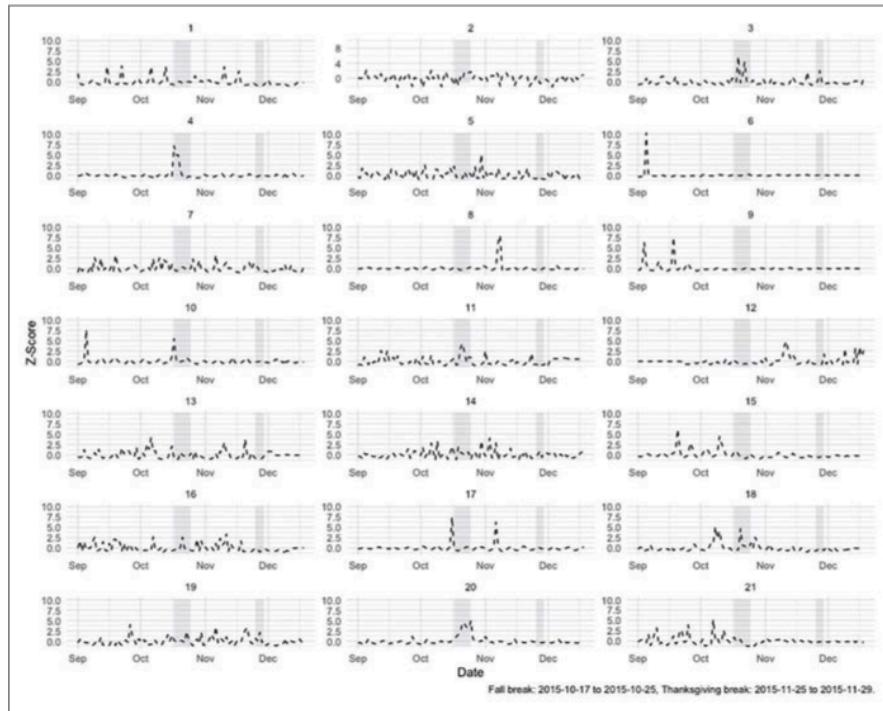


If we focus on non-peak periods, other patterns are evident. In clusters 1 and 2, there is a good deal of difference early in the period, but scores eventually stabilize (indicating convergence in temporal degree over time). For clusters 3 and 7, it is the reverse (indicating divergence). Dyads in clusters 5 and 10 exhibit an inverted-U shape pattern with a period of dissimilarity in the middle. Clusters 6, 11, and – to some extent – 12 have more of an erratic pattern with periods in which the z-scores are below zero, thus indicating above-average similarity between the dyads.

Figure 3 shows centroid plots for the 21 dyadic trajectory similarity clusters based upon steps. The CVIs suggested 21 steps-based similarity trajectories, and the number of dyads within each cluster is distributed much more evenly than with the communication similarity trajectories. Clusters 3, 4, 10, 16, and 18 tend to have extreme dissimilarity during the mid-semester break, though they each exhibit different combinations of similarity and dissimilarity throughout the rest of the semester. Clusters 3 and 19 peak during the Thanksgiving break, but dyads in 3 remain otherwise much more similar than those in 19. We refer readers to Sepulvado *et al.* (forthcoming) for a more detailed description of the clusters.

Figure 4 presents the results of the logistic regression predicting the tie probability based on cluster membership. We present the predicted probabilities of a tie derived from model parameter estimates. Most predicted probabilities are around .02, which is the NetHealth social network's overall density in Fall 2015 (equivalent to the base probability of being a connected dyad). Clusters 2 and 3 have higher than expected probabilities, while dyads in clusters 10 and 11 are less likely to be connected than expected by chance.

Fig. 3 – Cluster Centroids for daily step dyadic trajectory similarity clusters



5.1. Association Between Steps and Temporal Degree Clusters

Next, we turn to the log-linear models. With residual deviance of 4664.2 on 220 degrees of freedom, the Poisson model predicting the number of dyads falling into each step and communication cluster combination has a p-value < 0.001. Figure 5 visualizes the residuals from the predicted counts under the independence model: $(\text{observed count} - \text{predicted count})/\sqrt{\text{predicted count}}$. The largest positive residual values (> 5) are found in 6% of the 252

cells. There are fewer negative residual cells (2% with residuals < -5), and all these cases involve communication clusters which have a stable daily z-scores except for one peak (clusters 4, 5, 9, 12), suggesting it is unlikely this step cluster type is associated with a flat temporal degree difference profile.

Fig. 4 – Predicted probability of connected dyads by temporal degree cluster

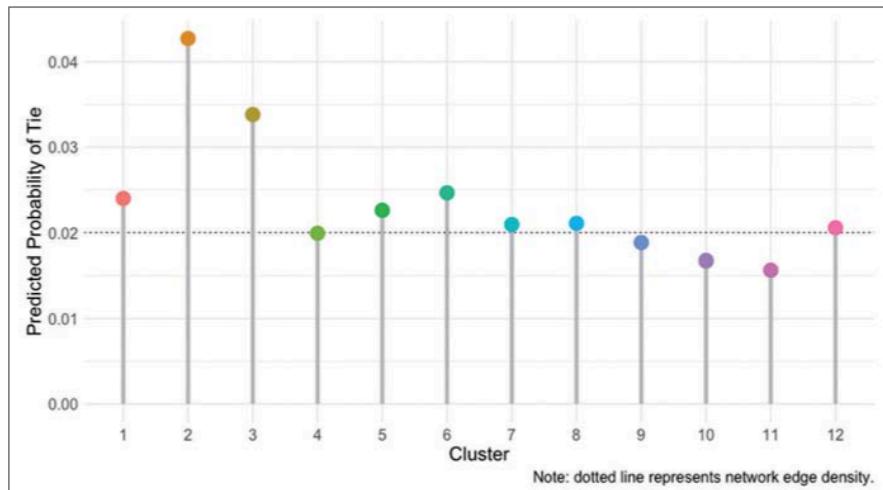
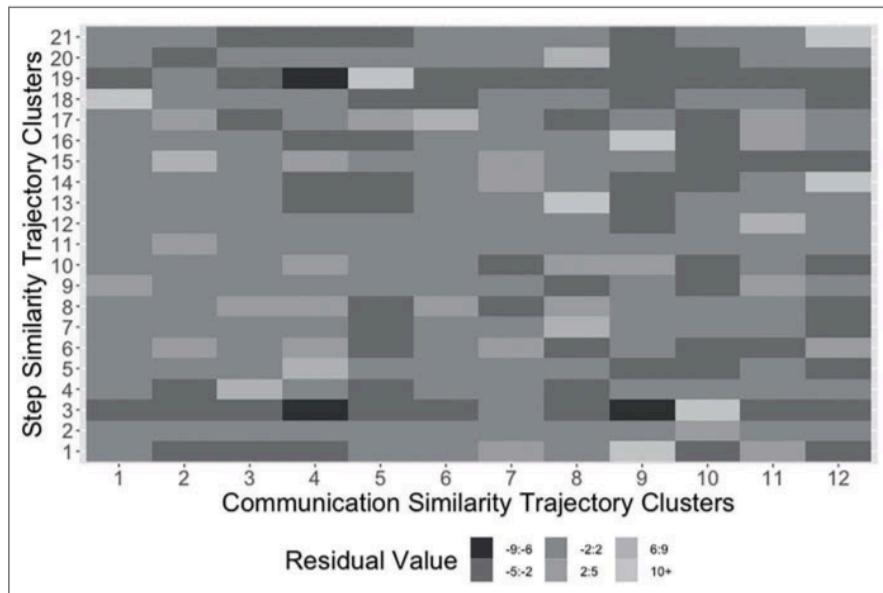


Fig. 5 – Heatmap of the association between trajectory clusters



We ran a series of nested log-linear models (Powers & Xie, 2008) to examine the three-way association between demographic homophily measures (same gender, same race, or same religion) and step and communication activity cluster membership. Table 1 shows the results for gender, race, and religious homophily. Recall that the baseline (independence) model fits the marginal distributions for the steps clusters, the communication clusters, and the homophily measure and that the second model fits parameters for each homophily factor's association with the steps and the communication clusters. As shown in the table, adding the two-way marginals that fit the association of gender, race, and religious homophilous dyads with step and communication trajectory similarity clusters statistically improves the fit of the model using conventional criteria of significance ($p < 0.01$). This indicates that dyads assigned to the same temporal degree and step trajectory similarity class are also more (or less) likely than expected by chance to match on key sociodemographic characteristics.

However, looking at the *substantive* improvement in model fit (indicated by the proportion reduction in deviance), we can see that the association between step and communication trajectory similarity clusters and race and religious homophily is much stronger than the corresponding improvement in model fit for gender (15% and 20% versus 1.6% respectively). This indicates that, while statistically discernible, the association between dyadic degree trajectory similarity cluster and gender is much weaker than for the other homophily dimensions. This is consistent with results reported in our previous work, which showed dyadic behavioral trajectory similarity cluster membership to have a weak relationship with gender homophily.

The final model in each panel fits the two-way marginal between step and temporal degree cluster membership. As already noted, there is a good deal of association between these two cluster memberships. This is reflected in the substantial improvement in model fit. Furthermore, the residual deviance in these models is very low, indicating that a model with three-way interactions allowing the association between step and communication clusters to vary by homophily of the dyad would not yield much of an improvement in fit.

Tab. 1 – Fit-statistics of log-linear models of the association between step and communication trajectory similarity clusters

Resid	Df	Resid. Dev	Df	Deviance	Pr(>Chi)	Eq.
<i>Homophily: Gender</i>						
471	4,984.68					S + C + G.H.
440	4,907.10	31	77.58	0.00	0.02	(S * G.H.) + (C * G.H.)
220	240.73	220	4,666.38	0.00	0.94	(S * G.H.) + (C * G.H.) + (S * C)
<i>Homophily: Race/Ethnicity</i>						
465	6,320.90					S + C + R/E.H.
434	5,372.17	31	948.73	0.00	0.15	(S * R/E.H.) + (C * R/E.H.)
214	616.37	220	4,755.80	0.00	0.75	(S * R/E.H.) + (C * R/E.H.) + (S * C)
<i>Homophily: Religion</i>						
468	6,286.57					S + C + Relig.H.
437	5,036.20	31	1,250.37	0.00	0.20	(S * Relig.H.) + (C * Relig.H.)
217	415.23	220	4,620.96	0.00	0.74	(S * Relig.H.) + (C * Relig.H.) + (S * C)

Note: Within each section, the p-value is for the comparison of one row with the preceding row. G.H. stands for gender homophily, R/E.H. stands for race/ethnicity homophily, and Relig. H. stands for religion homophily. S represents steps cluster, and C represents communication cluster.

6. Discussion

Dyads are an important component of social structure and, for some, the building block of social networks (Wasserman *et al.*, 1994, p. 505-ff; Rivera *et al.*, 2010). The analytic approach outlined here provides a way to take measures of time-varying traits defined on actors in a dynamic social network and quantify aspects of dyadic network evolution. This allows us to extend to the temporal case the basic notion of dyadic similarity; this notion is based on such basic social network constructs as homophily. Considering similarity to be a time-varying attribute of dyads yields the notion of a temporal *dyadic similarity trajectory*, which helps to specify the dynamic evolution of each pair of actors in the network concerning how similar (or dissimilar) they are on a target attribute. This attribute, as we have shown here, can be either exogenous or endogenous to the network. Building on this, we define the idea of *dyadic similarity trajectory cluster* as a mapping that assigns each dyad to a data-derived class based upon whether they share a temporal similarity pattern with other dyads in the same class.

Our empirical analysis, both in previous work and in this paper, shows that these clusters encode essential information, allowing us to predict both temporal network dynamics and whether a dyad is homophilous on a given set of (time-constant) traits. The results reported in this paper show that the dyadic trajectory similarity cluster approach previously shown to be fruitful when considering behavioral traits exogenous to the network (Sepulvado *et al.*, forthcoming) can be usefully extended to actor-level traits endogenous to the network, such as communicative activity (temporal degree). Our results indicate that dyadic trajectory similarity classes derived from this type of endogenous trait also encode useful information about network dynamics (e.g., helping us predict whether a dyad is connected or not) and are statistically associated with dyadic classes obtained from trajectory similarity based on exogenous traits. These findings indicate that insights into when and how social network ties form, how they evolve, and how long they persist can be garnered by constructing both behavioral and network position similarity trajectories for pairs of persons and ascertaining how these similarity patterns change when a tie is formed, during the life of the tie, before its decay, and after the tie no longer exists.

Future research should extend the approach proposed in this paper in several ways. First, temporal degree is only one of many endogenous time-varying traits defined on actors in a social network. Multiple indicators of an actor's position in a temporal network at a given time – defining a type of “temporal centrality” (Kim & Anderson, 2012) – yield an actor-level trajec-

tory (Liu *et al.*, 2018), from which one may derive a corresponding dyadic similarity trajectory and a related cluster assignment for each dyad. Future work can thus investigate whether dyadic trajectory similarity clusters based on other endogenous measures of actors' position provide substantively relevant network information (e.g., helping predict other dyadic properties). Such work can also examine the link between these other trajectory clusters and other fine-grained actor-level traits.

Additionally, the approach proposed here can be extended to other network building blocks or “motifs” (Milo *et al.*, 2002) beyond dyads, such as triads and higher-level structures. After all, a dyad is a subgraph of size 2, and it is possible to extend the notions of similarity, similarity trajectory, and similarity trajectory clusters for subgraphs of larger size. For instance, triadic dissimilarity can be treated as an additive function of the dissimilarity between its three constituent dyads. The temporal evolution of this quantity thus gives triadic dissimilarity trajectories. Triads can then be assigned to triadic trajectory similarity clusters, and these could be used to help predict whether given triads belong to (or are more likely to transition into) well-known triadic connectivity classes (Wang *et al.*, 2014), such as the null, open, or closed triad.

References

- Aghabozorgi S., Shirkhorshidi A.S., Wah T.Y. (2015), “Time-series clustering. A decade review”, *Information System*, 53, pp. 16-38.
- Bahulkar A., Boleslaw K.S., Chawla N., Lizardo O., Chan K. (2017), “Influence of personal preferences on link dynamics in social networks”, *Complexity*, Article ID 4543563.
- Bidart C., Lavenu D. (2005), “Evolutions of personal networks and life events”, *Social Networks*, 27 (4), pp. 359-376.
- Centola D. (2010), “The spread of behavior in an online social network experiment”, *Science*, 329 (5996), pp. 1194-1197.
- Christakis N.A., Fowler J.H. (2007), “The spread of obesity in a large social network over 32 years”, *The New England Journal of Medicine*, 357 (4), pp. 370-379.
- Christakis N.A., Fowler J.H. (2008), “The collective dynamics of smoking in a large social network”, *New England Journal of Medicine*, 358, pp. 2249-2258.
- Christakis N.A., Fowler J.H. (2013), “Social contagion theory: Examining dynamic social networks and human behavior”, *Statistics in Medicine*, 32, pp. 556-577.
- De Nooy W. (2011), “Networks of action and events over time. A multilevel discrete-time event history model for longitudinal network data”, *Social Networks*, 33 (1), pp. 31-40.

- Dickison M.E., Magnani M., Rossi L. (2016), *Multilayer social networks*, Cambridge University Press, Cambridge.
- Dokuka S., Valeeva D., Yudkevich M. (2016), “Homophily evolution in online networks: Who is a good friend and when?”, *International Conference on Analysis of Images, Social Networks and Texts*, Springer, Cham, pp. 91-99.
- Epskamp S. (2020), “Psychometric network models from time-series and panel data”, *Psychometrika*, 85 (1), pp. 206-231.
- Feld S.L. (1982), “Social Structural Determinants of Similarity among Associates”, *American Sociological Review*, 47 (6), pp. 797-801.
- Firth D. (1992), “Bias reduction of maximum likelihood estimates”, *Biometrika*, 80 (1), pp. 27-36.
- Fowler J.H., Christakis N.A. (2008), “Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study”, *British Medical Journal*, 337, a2338.
- Fowler J.H., Christakis N.A. (2008a), “The dynamic spread of happiness in a large social network”, *British Medical Journal*, 337, a2338.
- Fowler J.H., Christakis N.A. (2008b), “Estimating peer effects on health in social networks: A response to Cohen-Cole and Fletcher; and Trogdon, Nonnemaker, and Pais”, *Journal of Health Economics*, 27, pp. 1400-1405.
- Holme P., Saramäki J. (2012), “Temporal networks”, *Physics Reports*, 519 (3), pp. 97-125.
- Ingram P., Morris M.W. (2007), “Do people mix at mixers? Structure, homophily, and the ‘life of the party’”, *Administrative Science Quarterly*, 52 (4), pp. 558-585.
- Kim H., Anderson R. (2012), “Temporal node centrality in complex networks”, *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, 85 (2 Pt 2), 026107.
- Kossinets G., Watts D.J. (2009), “Origins of homophily in an evolving social network”, *American Journal of Sociology*, 115 (2), pp. 405-450.
- Krackhardt D., Handcock M.S. (2006, June), “Heider vs Simmel: Emergent features in dynamic structures”, in E. Airoldi, D.M. Blei, S.E. Fienberg, A. Goldenberg, E.P. Zing, A.X. Zheng (eds.), *ICML Workshop on Statistical Network Analysis*, Springer, Berlin, pp. 14-27.
- Lazer D., Pentland A.S., Adamic L., Aral S., Barabasi A.L., Brewer D., Christakis N., Contractor N., Fowler J., Gutmann M. et al. (2009), “Life in the network: the coming age of computational social science”, *Science*, 323 (5915), p. 721.
- Leszczensky L., Pink S. (2019), “What drives ethnic homophily? A relational approach on how ethnic identification moderates preferences for same-ethnic friends”, *American Sociological Review*, 84, 3, pp. 394-419.
- Lewis K., Kaufman J. (2018), “The Conversion of Cultural Tastes into Social Network Ties”, *The American Journal of Sociology*, 123 (6), pp. 1684-1742.
- Lewis K., Kaufman J., Gonzalez M., Wimmer A., Christakis N. (2008), “Tastes, ties, and time: A new social network dataset using Facebook.com”, *Social Networks*, 30 (4), pp. 330-342.

- Liu S., Hachen D., Lizardo O., Poellabauer C., Striegel A., Milenković T. (2018), “Network analysis of the NetHealth data: exploring co-evolution of individuals’ social network positions and physical activities”, *Applied Network Science*, 3 (1), p. 45.
- Martin J.L., Yeung K.-T. (2006), “Persistence of closer personal ties over a 12-year period”, *Social Networks*, 28, pp. 331-362.
- McPherson M., Smith-Lovin L., Cook J. (2001), “Birds of a feather: Homophily in social networks”, *Annual Review of Sociology*, 27 (1), pp. 415-444.
- Milo R., Shen-Orr S., Itzkovitz S., Kashtan N., Chklovskii D., Alon U. (2002), “Network motifs: simple building blocks of complex networks”, *Science*, 298 (5594), pp. 824-827.
- Miritello G. (2013), *Temporal Patterns of Communication in Social Networks*, Springer Science & Business Media, Berlin.
- Miritello G., Lara R., Cebrian M., Moro E. (2013), “Limited communication capacity unveils strategies for human interaction”, *Scientific Reports*, 3.
- Morgan D.L., Neal M.B., Carder P. (1997), “The stability of core and peripheral networks over time”, *Social Networks*, 19 (1), pp. 9-25.
- Paparrizos J., Gravano L. (2015), “k-Shape: Efficient and Accurate Clustering of Time Series”, in *SIGMOD ’15: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Victoria, Australia*, Association for Computing Machinery, New York, pp. 1855-1870.
- Paparrizos J., Gravano L. (2017), “Fast and accurate time-series clustering”, *ACM Transactions on Intelligent Systems and Technology*, 42 (2), pp. 1-49.
- Noel H., Nyhan B. (2011/7), “The ‘unfriending’ problem: The consequences of homophily in friendship retention for causal estimates of social influence”, *Social Networks*, 33 (3), pp. 211-218.
- Powers D., Xie Y. (2008), *Statistical Methods for Categorical Data Analysis*, Emerald Group Publishing, London.
- Puhr R., Heinze G., Nold M., Lusa L., Geroldinger A. (2017), “Firth’s logistic regression with rare events: accurate effect estimates and predictions?”, *Statistics in Medicine*, 36, pp. 2302-2317.
- Purta R., Mattingly S., Song L., Lizardo O., Hachen D., Poellabauer C., Striegel A. (2016), “Experiences measuring sleep and physical activity patterns across a large college cohort with fitbits”, *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, pp. 28-35.
- Purta R., Striegel A. (2019), “Predicting Friendship Pairs from BLE Beacons Using Dining Hall Visits”, *2019 28th International Conference on Computer Communication and Networks (ICCCN)*, pp. 1-9.
- Raeder T., Lizardo O., Hachen D., Chawla N.V. (2011), “Predictors of short-term decay of cell phone contacts in a large scale communication network”, *Social Networks*, 33 (4), pp. 245-257.
- Rivera M.T., Soderstrom S.B., Uzzi B. (2010), “Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms”, *Annual Review of Sociology*, 36 (1), pp. 91-115.

- Schaefer D.R., Kreager D.A. (2020), "New on the Block: Analyzing Network Selection Trajectories in a Prison Treatment Program", *American Sociological Review*, 85 (4), pp. 709-737.
- Secara V., Stopczynski A., Lehmann S. (2016), "Fundamental structures of dynamic social networks", *Proceedings of the National Academy of Sciences of the United States of America*, 113 (36), pp. 9977-9982.
- Sepulvado B., Wood M.L., Fridmanski E., Wang C., Chandler C.J., Lizardo O., Hachen D. (2020), "Predicting Homophily and Social Network Connectivity From Dyadic Behavioral Similarity Trajectory Clusters", *Social Science Computer Review*, forthcoming.
- Wang C., Lizardo O., Hachen D.S. (2014), "Triadic evolution in a large-scale mobile phone network", *Journal of Complex Networks*, 3 (2), pp. 264–290.
- Wang C., Lizardo O., Hachen D.S. (2020), "Neither influence nor selection: Examining co-evolution of political orientation and social networks in the NetSense and NetHealth studies", *PLoS One*, 15, 5, e0233458.
- Wasserman S., Faust K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge.
- Yang Y., Nitesh V.C., Uzzi B. (2019), "A network's gender composition and communication pattern predict women's leadership success", *Proceedings of the National Academy of Sciences*, 116, 6, pp. 2033-2038.