



Modeling the coevolution of relational events and relational states in social networks

Christoph Stadtfeld^{1,2*}, Maria Eugenia Gil-Pallares¹ and Viviana Amati³

Handling Editor: Sabrina Gaito

*Correspondence:

christoph.stadtfeld@ethz.ch

¹Social Networks Lab, ETH Zürich, Zurich, Switzerland

²CIA, Corvinus University, Budapest, Hungary

Full list of author information is available at the end of the article

C. Stadtfeld and M.E. Gil-Pallares contributed equally to this work.

Abstract

We present a statistical model for the joint analysis of temporal social network data collected as relational events (e.g., digital traces) and as a series of relational states (e.g., repeated surveys). This model effectively combines behavioral and cognitive measures of social ties. It enables hypothesis tests on the coevolution between digital behavioral traces of relationships (e.g., social interactions online) and their cognitive perceptions (e.g., friendships). We introduce a new estimation routine, based on the expectation-maximization algorithm, to address the challenge posed by observing digital and survey data at different frequencies. We apply this framework in an empirical study of an emerging community of first-year undergraduate students at a Swiss university. We combine repeated survey measures of friendship perceptions with time-stamped social media connections collected through the Facebook API. Our results suggest that the online and offline networks coevolve: changes in one network provide valuable information for modeling the respective other. Additionally, indirect connections in the offline network were associated with the creation of direct ties in the online network. Both networks exhibit similarities in transitivity, homophily, and gender effects. However, they differ in preferential attachment: there is evidence for degree popularity online and against it in the offline network. The paper highlights the broad applicability of the model for future dynamic social network studies that combine relational events and relational states.

Keywords: Dynamic social networks; Statistical modeling; Relational events; Relational states; Friendship; Social media

1 Introduction

Humans increasingly live parts of their social lives on digital platforms, creating new and exciting opportunities for researchers to study and understand the dynamics of social networks on both a large scale and in microscopic temporal detail. The promise of the emerging scientific field of Computational Social Science (CSS) is that a single research study could investigate the networks of millions of users and use behavioral data for each user that was recorded to the millisecond. In their foundational paper, Lazer et al. [1, p.722] hail the opportunities provided by “the capacity to collect and analyze data with an unprecedented breadth and depth and scale”. In a related initial work, Conte et al. [2, p.342] argue that “the availability of unprecedented amounts of data about human interactions

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

[...] opens the possibility of using those data to leverage knowledge about social behavior beyond research on the scale of tens of people.”

However, since the early euphoria in the CSS community, researchers have increasingly recognized that merely studying large digital online platforms comes with limitations. In social network studies, it has become evident that automatically collected data of interpersonal interaction may be difficult to interpret and validate without knowing the qualitative contexts in which they appear [3, 4]. Similar-looking social media traces may relate to very different real-life (offline) relationships, such as friendship, work interactions, or even enmity. However, collecting additional cognitive measures of how individuals perceive their actions and social relationships often proves difficult, as data are commonly controlled by platform providers and are affected by platform algorithms [5, 6]. Cognitive data on social networks are best collected through surveys that represent the cognitive states of an individual’s social relations.

Dynamic social networks are commonly measured and represented in one of two ways. The first approach considers the states of social ties in a series of networks collected at discrete points in time. These data are represented as series of *relational states*. Relational states are commonly assessed in survey panels, where individuals’ perceptions of their own social relationships at the time of the survey are measured (e.g., asking “who are you friends with?”) rather than their perceived times of change, which are harder to recall (e.g., asking “when did you become friends?”) [7]. Digital data may also be collected as sequences of relational states, for example, when continuous measurement of an online platform is technically not feasible but it can be accessed at discrete time points. Observational and archival network data often take the form of relational states, for example, when researchers encode observed social interactions only in specific observation periods as a static network or when economic interactions are archived as annually aggregated data.

The second approach considers exact sequences of dyadic events that change the state of a social tie (e.g., creation or deletion), reflect its presence (e.g., instances of interaction signaling an underlying social tie), or represent other instances of dyadic interaction without necessarily assuming the presence of an underlying social tie (e.g., interpersonal communication on an online platform). These data are represented as sequences of *relational events*. Relational events are often time-stamped and are commonly collected in the context of digital data, for example, in communication networks (e.g., who called whom at what point in time) or on social media platforms (e.g., who started following whom or engaged with content at what point in time) [8]. Other types of dynamic network data (e.g., observed face-to-face interactions) or archival data (e.g., historical time-stamped economic transactions) may be collected as relational events [9]. The distinction between relational states data and relational events data is first and foremost a matter of data representation – are dynamic networks represented as sequences of state matrices or as sequences of relational events? The constructs they represent may be related. Relational events can, for example, indicate the onset or dissolution of a stable social tie or can have a duration [9, 10]. Similarly, relational states data may represent aggregates of relational events in a temporal period.

The decision between collecting relational event data or series of relational states often follows practical considerations. In dynamic social network studies, behavioral data are commonly collected as time-stamped relational events using digital platforms or dig-

ital measurement tools, whereas cognitive perception data are collected as a series of relational states through survey panels. The combination of different measurement approaches using relational events and relational states is rare but may provide a more nuanced picture of social network dynamics. First, they may capture the same relational construct from different perspectives, for example, the perceived and observed communication intensity between individuals. Second, they may capture different qualitative dimensions of the same construct, for example, the perceived interaction quality and the observed interaction frequency. Third, they may capture different but interdependent layers of relationships in the same setting, for example, perceived friendships and changes of social media connections. It is commonly the case that relational states collected through surveys represent a *cognitive* dimension of social ties, whereas relational events collected through digital means represent a corresponding *behavioral* dimension. The simultaneous collection of relational event data and relational states data is practically often difficult, and thus not yet common in empirical studies.

Only recently have new study designs emerged that combine cognitive and behavioral dimensions and measure both relational events and relational states in social networks [11, 12]. Indeed, researchers in social networks have long argued that social relationships are complex and consist of at least a cognitive and behavioral component that should be studied jointly [13, 14]. Large-scale data-driven approaches that rely exclusively on digital data may exhibit inherent limitations regarding the scientific insights and substantive impacts they generate. Consequently, there has been a growing demand for methods that integrate large-scale studies on digital platforms with deep cognitive and behavioral survey measurements developed by social and cognitive scientists [4, 15, 16]. However, one core challenge of conducting hybrid studies is the current absence of methodological frameworks that allow for the joint analysis of social network data derived from both digital platforms and survey instruments. In fact, recent review papers on relational event models have called for the development of such statistical tools [9, 17].

In this paper, we introduce a novel statistical and methodological framework for the study of the coevolution of social networks measured both as relational events (e.g., interactions on online social media platforms and other digital trace data), and relational states (relationships elicited through social network surveys). The utility of this analytical approach is empirically demonstrated through the investigation of emerging social networks in a community of first-year undergraduate students. We integrate relational event data on social media connections between students, collected through the Facebook API, with relational state data on perceived friendship relations, collected through a longitudinal survey panel among the same individuals. These data are modeled jointly to test a variety of dynamic mechanisms within and between the two networks.

2 Related work

The study of social network evolution has garnered significant attention across various disciplines in recent decades, with researchers inquiring about the circumstances under which social ties are most likely to change. Various theoretical explanations have been presented that are concerned with how existing network structures and individual attributes can affect the probability that new ties emerge or dissolve [18]. Among the most common theoretical explanations are the role of reciprocity, embeddedness and transitivity, popularity and preferential attachment, and homophily [7, 19–21]. The interdisciplinary aspect

of social network analysis has motivated the continuous development and refinement of methodologies that cater to the particularities of the observed data, behaviors, and events specific to different social groups.

A fundamental requirement for analyzing network evolution is the availability of temporally structured data whose collection is determined by the type of relationships and the mechanisms studied. Longitudinal data collection, such as repeated surveys, remains a widely adopted approach, with network observations collected in a survey panel design, balancing between granularity and feasibility [7]. Recently, with the world's digitalization, collecting temporally accurate interaction data, often referred to as relational events, has become an accessible and fast alternative [9]. Although survey data can be costly and challenging to collect, they can capture aspects that time-stamped data cannot. Individuals' perceptions, attitudes, beliefs, and behaviors outside the digital realm can be overlooked when collecting data merely as relational events. In contrast, high-frequency temporal data, such as online social media data, can uncover latent patterns in offline network dynamics or objective relational behavior that would otherwise remain hidden, affected by biases, or only partially revealed by survey data. The rapid rise of social media platforms such as Facebook, Twitter, and Instagram has fueled a growing interest in the study of online social networks and their dynamic evolution [8]. For instance, researchers have examined how users form and dissolve online friendships over time [22, 23] or how retweet and like networks evolve on online platforms [24–26]. Similarly, new digital measurement devices became available that allow continuous online tracking of individuals' behavior and interactions [27, 28].

In the presence of uneven temporal granularity, in particular, relational states and relational event data, a standard solution is to unify frequencies, for example by aggregating event data into time windows, or conduct separate analyses. The latter allows for the use of various models tailored to specific data types. Panel data analysis and modeling have been extensively studied, with established methods such as Stochastic Actor-oriented Models [SAOMs, 29] and Temporal Exponential Random Graph Models [TERGMs, 30] offering valuable insights into relational dynamics. Additionally, time-stamped relational data can be analyzed effectively through Relational Event Models [REMs, 9, 31] or their actor-oriented counterpart, Dynamic Network Actor Models [DyNAMs, 32]. However, a notable gap remains in the literature regarding the *coevolution* of relational states and relational events.

There appears to be a growing trend in empirical social network studies to combine cognitive and behavioral measures of social relationships. Some recent empirical studies discuss the co-occurrence of online and offline interactions [e.g., 33–35]. Other recent studies propose to combine survey data with data from digital traces [15, 16, 36, 37]. There is indeed evidence that online and offline social ties of the same individuals are coevolving and perceived and utilized differently. Subrahmanyam et al. [38] show that individuals might take advantage of differences in communication online and offline to convey concerns that otherwise would not be shared in person, thereby increasing their social capital. de Zúñiga et al. [39] test the empirical differences of social capital online and offline, highlighting the structurally unique nature of social media interactions. Cole and Griffiths [40] show that the relative anonymity and reduced social pressure of online environments foster a sense of safety, which allows users to express themselves more freely and facilitates the connection with communities of shared identity. Steinfield et al. [41] argue for the

additional value of social media for introverted people or those with lower self-esteem, showing how the use of Facebook contributed more significantly to their bridging social capital in comparison to other students. Spottswood and Wohn [42] present a collection of different approaches to the study of online social capital, encouraging researchers to consider the repercussions of offline events on the need and use of online social capital. Berger et al. [43] highlight the need for in-depth studies of the structural differences between digital and in-person interactions and the interconnection of users online and offline.

Survey-focused studies have increasingly acknowledged the role of social media in individuals' social lives by assessing aspects of its use and impact. Some studies have asked additional questions about the users' time spent on apps or the types of relationships created on them [12, 41, 44]. Surveys appear to be the preferred methodology for gathering data about the impact of social media usage on social networks, their perceptions, influence, and behaviors [45]. Al-Yazidi et al. [46] recently raised awareness about the complexity of the influence between networks. They argue that social networks are multi-dimensional and context-dependent and that these aspects are easily overlooked. Moreover, Kahai and Lei [47] encourage researchers not to assume beforehand the impact of online networks on offline relationships, but to understand under what circumstances this influence may be present.

3 Model

We propose a model for the joint analysis of temporal social network data collected simultaneously in the same social setting as relational events and relational states. A social setting may, for example, be an organization, a sub-community of an online platform, or – as in the empirical part of this paper – a group of university students who form interpersonal social ties while also interacting with one another online.

Relational events could, for example, represent automatically recorded time points at which individuals interact with each other on an online platform. We denote by $\mathcal{A}$ the set of actors in a specific social setting. A relational event $\omega = \{i, j, t : i, j \in \mathcal{A}\}$ is an instantaneous interaction characterized by the event sender i , the event receiver j and the time t at which the event occurred. The ordered sequence of relational events, denoted as $\Omega = \{\omega_1, \omega_2, \dots\}$ constitutes the relational event data [9].

Relational states could represent individuals' perceptions about their social ties in the same community of individuals. Social network surveys can be conducted repeatedly at $M > 1$ points in time [7] to assess, for example, who individuals perceive to be their friends at the time of data collection. The resulting data, denoted as $x = \{x(t), t = 1, \dots, M\}$, consist of ordered sequences of adjacency matrices $x(t)$ with elements $x_{kl}(t) = 1$ if individual k is tied to individual l at time t (for example, k reports l as a friend) and $x_{kl}(t) = 0$ otherwise. Hereafter, for simplicity, we assume that $M = 2$. However, the model we propose is applicable to longer panel designs.

Two consecutively observed networks of relational states are likely to differ due to tie changes consisting of tie creations and terminations. These changes are not directly observed due to the panel design. We conceptualize the sequence of tie changes as two sets of unobserved relational event lists: the sequence of tie creation $\tilde{\Omega}^c = \{\tilde{\omega}_1^c, \tilde{\omega}_2^c, \dots\}$ and the sequence of tie termination (or deletion) $\tilde{\Omega}^d = \{\tilde{\omega}_1^d, \tilde{\omega}_2^d, \dots\}$.

In the sequence $\tilde{\Omega}^c$ each event $\tilde{\omega}_h^c = \{k, l, \tilde{t}\}$ represents an ordered pair of nodes that got connected between the first and second observation, i.e., the following conditions hold:

$$\begin{aligned} t_1 &< \tilde{t} < t_2 \\ x_{kl}(t_1) &= 0 \\ x_{kl}(t_2) &= 1. \end{aligned}$$

Here, we consider only cases where ties are newly created based on the panel information, but the event list can be further augmented with self-canceling events, for example a tie being first created and then being dropped while not being present at both times t_1 and t_2 .¹ Similarly, the sequence $\tilde{\Omega}^d$ includes the unobserved sequence of tie deletions between two subsequent observations for cases of dyads x_{kl} where $x_{kl}(t_1) = 1$ and $x_{kl}(t_2) = 0$, $t_1 < \tilde{t} < t_2$.

Building upon earlier work [49], we consider changes in each of the event lists to follow a Dynamic Network Actor Model. Each event (an observed relational event, or an unobserved event of tie creation or deletion in the survey data) is assumed to be governed by a continuous-time Markov process in which the hazards of an event are constructed as a two-step process of actors' propensities to occur as the sender of events (for example, to initiate an online interaction or to create or delete a tie in the relational states network) and senders' choice of the event receiver (who to interact with or change a relationship with). The state space of the process is defined as the set of process states $y(t) \in Y$ at time t , which can entail various information, such as the state of one or multiple networks and the current attributes of the actor.

Since the model assumes that the sequence of events is the outcome of a Markov process, the waiting time between two consecutive events has an exponential distribution with rate modeled by the rate function ρ_i describing the frequency at which an actor i gets opportunities to send an event. The rate function is actor-specific and expressed by the linear function

$$\rho_i(y, \gamma) = \exp \left(\gamma_0 + \sum_{m=1}^{|r|} \gamma_m r_m(i, y) \right) \quad (1)$$

with $r(i, y)$ a vector of statistics capturing the dependence of the rate on actors' attributes and network positions weighted by parameter vector γ . The parameter γ_0 serves as a model intercept, expressing the baseline waiting time between two subsequent network changes of the same actor, assuming that all other statistics are zero. The statistics express the differential tendencies of actors to make network changes, for example, based on their demographics, psychological traits, or their degree and embeddedness in the network. Examples of statistics are the value of an attribute v , the out-degree and the number of transitive triads in which an actor i is embedded:

$$\begin{aligned} r_1(i, y) &= v_i \\ r_2(i, y) &= \sum_{j \in \mathcal{A}} x_{ij} \end{aligned}$$

¹Self-canceling ties may also be considered for cases where ties are present in two subsequent survey waves. Such cases are not considered in this paper, but the method can be straightforwardly generalized to accommodate self-canceling observations as in the method of Snijders et al. [48].

$$r_3(i, y) = \sum_{j \in \mathcal{A}} \sum_{k \in \mathcal{A} \setminus j} x_{ij} x_{jk} x_{ik},$$

with $v, x \in y$ and diagonal elements $x_{ii} = 0$.

The parameter γ_m measures the strength and the direction of the dependence of the rate on the statistic r_m . For example, if v_i is the level of extraversion of actor i , a positive estimated parameter γ_1 associated with $r_1(i, y)$ suggests that extraverted actors are more likely to initiate relational events.

Conditioning on actor i becoming active and sending an event, the choice of the receiver is modeled as a multinomial choice model [50]. Thus, the probability that actor i selects actor j as receiver is:

$$p_{ij}(y, \beta) = \frac{\exp\left(\sum_{m=1}^{|s|} \beta_m s_m(i, j, y)\right)}{\sum_{k \in \mathcal{A} \setminus i} \exp\left(\sum_{m=1}^{|s|} \beta_m s_m(i, k, y)\right)}, \quad (2)$$

where β_m are the parameters and the functions s_m are statistics describing how attributes of the event receiver, similarities between the sender and receiver, or the position of the i - j dyad in the current network affect the choice of the receiver. Examples of statistics describing the tendency towards attribute homophily, preferential attachment, and triadic embeddedness in receiver choices are:

$$\begin{aligned} s_1(i, j, y) &= \mathbf{I}(v_i = v_j) \\ s_2(i, j, y) &= \sum_{k \in \mathcal{A} \setminus \{i, j\}} x_{kj} \\ s_3(i, j, y) &= \sum_{k \in \mathcal{A} \setminus \{i, j\}} x_{ik} x_{kj} \end{aligned} \quad (3)$$

with $v, x \in y$.

The vectors of statistics r and s express what Stadtfeld and Amati [18] defined as *network mechanisms*: structural tendencies that are regularly found in social networks and that can be linked to one or multiple causal explanations of tie formation.

Under the assumption of conditional independence, given the process state y , the transition rates of an event with sender i and receiver j are:

$$\lambda(i, j|y, \gamma, \beta) = \rho_i(y, \gamma) p_{ij}(y, \beta). \quad (4)$$

The dynamics of social networks collected as relational events and as relational states can be modeled by this transition rate, which, together with the process state space Y , constitutes a continuous-time Markov process. However, in the case of relational states, the actual transitions are not observed and only the initial and final network states $y(t_1)$ and $y(t_2)$ are known. The process state resulting from the transition at time t shown above is denoted by $y^{i \rightarrow j} \in Y$. It is identical to y except for changes triggered by the event $\{i, j, t\}$ (e.g., a new tie from i to j being added to a network x). In the simplest model, we consider three competing Markov processes λ_{ev} , λ_{st}^c , λ_{st}^d . These processes govern the tendency of individuals to send relational events, to create and delete ties in the states network respectively. We denote the set of all model parameters as

$$\theta = (\gamma_{ev}, \gamma_{st}^c, \gamma_{st}^d, \beta_{ev}, \beta_{st}^c, \beta_{st}^d) \quad (5)$$

which are jointly estimated given the data. For a set of event sequences $(\Omega, \tilde{\Omega}^c, \tilde{\Omega}^d)$, we can formulate the likelihood function $L(\theta|\Omega, \tilde{\Omega}^c, \tilde{\Omega}^d, y(t_1))$ for the vector of parameters θ , based on the (augmented) event data and the initial information available in the process state y at time t_1 . Since the sub-models are conditionally independent given the process state, the likelihood functions can be derived and maximized separately for all coefficients. For the receiver choice models, the likelihoods follow from the choice probabilities in Equation (2). For example, we have:

$$L^\beta(\Omega) = \sum_{\omega \in \Omega} p_{ij}(y(t_\omega), \beta), \quad (6)$$

where $y(t_\omega)$ denotes the process state just before the observed relational event ω . It should be noted that y may also be altered by the augmented event sequences $\tilde{\Omega}^c$ and $\tilde{\Omega}^d$, which means that while the likelihoods are statistically independent, the processing of states requires a joint computation of chains. The likelihood formulation of the individual rate models $L^\gamma(\Omega)$ is derived similarly from exponential waiting time probability models and considers both the time intervals between two consecutively observed events within a chain (e.g., the waiting times between relational events), as well as the right-censored time intervals that are caused by events in other chains (e.g., creations and deletions of the relational state network). A detailed derivation and discussion of the likelihood functions $L^\gamma(\Omega)$ and $L^\beta(\Omega)$ for DyNAM models are provided in [49].

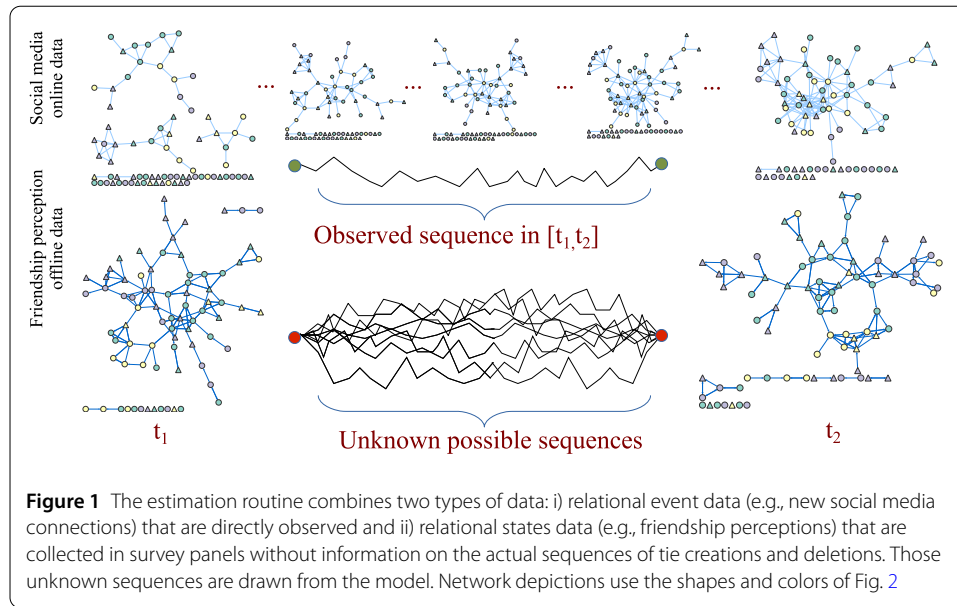
The proposed actor-oriented representation of network changes – considering actor activities and their tie choices in two steps – has been introduced by Snijders et al. [29] for the case of relational states data collected in network panels and has later been applied to the case of relational event data [51].

4 Methods

Estimating the vector of parameters θ is relatively straightforward when working with observed relational event sequences. Parameters are estimated using a maximum likelihood approach [49, 51]. However, the challenge arises when we incorporate data on relational states, as the sequences or chains of certain event types ($\tilde{\Omega}^c, \tilde{\Omega}^d$) are unknown. This is because we only observe networks at discrete time points, rather than the step-wise changes that occur between those points.

Related methods for the integration of survey data and time-stamped data (that are not both network data, as in our approach) have recently been discussed [36, 52]. In our case, the unknown event sequence $\tilde{\Omega} = \{\tilde{\Omega}^c, \tilde{\Omega}^d\}$ representing the creation and deletion of offline ties can be treated as missing or latent data during the estimation process. Figure 1 demonstrates how this can be accomplished. The upper part of the illustration represents the online data, from which we have completely observed the sequence of relational events. The lower part depicts two observed networks in a panel design. Since the sequence of relational states is unknown, we augment the data by stochastically generating possible sequences, as described in the following lines.

We take advantage of the fact that the probability of a given chain can be expressed in a closed form, similar to the likelihood function but depending on the parameters θ and the known data $y(t_1)$: $P(\tilde{\Omega}|\theta, \Omega, y(t_1))$. The data can thus be augmented by imputing missing chains that are weighted with a function of their respective probability, according to the principles of importance sampling. Samples of sequences of unobserved events are



obtained through an iterative process, in which the events of creation and deletion are sampled following a two-step approach. A sender, from the observed ones, is drawn, with probability given by rate model specification. Similarly, a receiver is drawn and a time for the event is simulated from the exponential distribution.

To integrate the imputed event sequences, we apply the Expectation-Maximization [EM, 53] algorithm, an iterative optimization approach particularly effective for problems involving latent variables. The EM algorithm consists of two steps repeated until convergence. In the Expectation step (E-step), the objective function Q , is calculated as the conditional expected value of the complete data log-likelihood given the observed data and the current parameter value. The Maximization step (M-step) updates the parameter estimate by maximizing the expected log-likelihood computed in the E-step. When direct computation in the E-step is challenging, the Monte Carlo EM (MCEM) algorithm provides a practical alternative by approximating the objective function through averaging the log-likelihoods over samples of the latent variable drawn from a distribution parameterized by the previous iteration's estimates. Similarly, when the M-step cannot be solved analytically, optimization methods such as Stochastic Gradient Descent [SGD, 54] can be employed for parameter estimators updates.

In our implementation, missing event data are drawn using an improving sequence of intermediate parameter estimates $\theta^1, \theta^2, \dots$, and the estimators are subsequently updated using the information from the generated sample, until the algorithm converges towards a satisfactory estimate $\hat{\theta}$.

The implemented algorithm is based on the Ascent-based Monte-Carlo Expectation-Maximization algorithm [55] and is summarized in Alg. 1. The algorithm operates on the principle that at each iteration, the objective function $\tilde{Q}^2$ is approximated with the generated unknown chains of formations and dissolutions in the relational states network. The asymptotic convergence of $\Delta\tilde{Q}(\theta^t, \theta^{t-1})$ to a normal distribution, with approximated

² $\tilde{Q}(\theta, \theta') = \frac{\sum_i w_i(\theta') \ell(\theta)}{\sum_i w_i(\theta')}$, with $w_i = \frac{f(\theta, \tilde{\Omega}_i)}{f(\theta, \tilde{\Omega}_i)}$ the importance sampling weights for a sampled sequence of events $\tilde{\Omega}_i$ obtained from the parameters θ ; f the likelihood of the observed data; and ℓ the complete data log-likelihood.

Algorithm 1 Estimation algorithm

Input: Relational event data, panel data, model specification.

Parameter: Optional list of initial parameter estimates, maximum number of iterations (max_it), tolerance (tol), quantiles z_{α_1} and z_{α_2} , and parameters for the SGD routine.

Output: Parameter estimates and standard errors.

```

1: Let  $it = 0$ .
2: Generate samples (event sequences).
3: while  $it < max\_it$  do
4:   SGD: update parameter values.
5:   if  $\Delta \tilde{Q} - z_{\alpha_1} ASE < 0$  then
6:     Increase pool of samples, update importance sampling weights.
7:   else
8:     Accept parameter update,  $it \leftarrow it + 1$ , check if pool of samples needs to be increased, update importance sampling weights.
9:   if  $\Delta \tilde{Q} + z_{\alpha_2} ASE < tol$  then
10:    end
11:   end if
12: end if
13: end while

```

standard error (ASE), allows for testing the significance of the improvements in the objective function between steps. If the improvement is statistically significant at a level α_1 , the parameter estimate update is accepted. If there is no improvement, the sample pool size is increased to account for the possible Monte Carlo approximation error, and the step is repeated. Similarly, the upper confidence limit at level α_2 (potentially distinct from α_1) is computed to determine the end of the algorithm when the improvements are negligible. Reference values for the algorithm's parameters, such as significance levels, are provided in [55] and [56]. The algorithm is implemented as an extension to the goldfish software package in R [57].

To reduce computational load, the generated sequences of events for the relational states network are reused throughout the iterations of the algorithm, adapting the importance sampling weights accordingly [58]. To be able to combine samples obtained from different distributions, according to the value of the estimators in the iteration they are drawn, we follow the simplest correction scheme of the importance sampling weights presented in [59]. These weights are also used in the SGD routine. In particular, motivated by the ideas on [60, 61], we perform mini-batch SGD, and choose the sequences of the batches based on their importance weight.

5 Empirical setting and data

To demonstrate the practical applicability of the model, we utilize a subset of the Swiss StudentLife Study, which is a longitudinal social network study conducted in Switzerland within a STEM-focused undergraduate environment [12, 62]. The data collection of this study included an “offline” longitudinal survey component as well as an online component of social media connections. Friendship perceptions (the offline network) were collected as part of a repeated web survey in which students were asked whom among their fellow students they consider a friend. At each survey, students could nominate up to 20 people.

Overall, 218 students participated in the survey part of the social networks study. A subset of $N = 78$ participants further volunteered to give access to their Facebook API, through which we recorded Facebook “friendships” with others (the online network). Data could only be accessed if both Facebook friends gave access to their API. At the time of data collection in 2016, Facebook was among the main social media platforms used by students. While collecting data in early 2017, Facebook changed its access policy for researchers so that the Facebook data collection had to be discontinued.³ However, we were able to collect three waves of survey data while having access to the Facebook API, two of which are jointly analyzed through a statistical network model for the first time in this paper.

The data thus comprise a set of 78 actors for whom changes in both their offline friendships and online social media ties are recorded simultaneously. While Facebook connections were referred to as friendships on the platform, we do not consider the online and offline networks to be different measurements of the same underlying construct of friendship, but rather two distinct relationship layers that are expected to coevolve but have a distinct function and meaning for the individuals in our study. The data collection occurred in a period shortly after the participants joined an undergraduate STEM program and the resulting networks are thus expected to be highly dynamic.

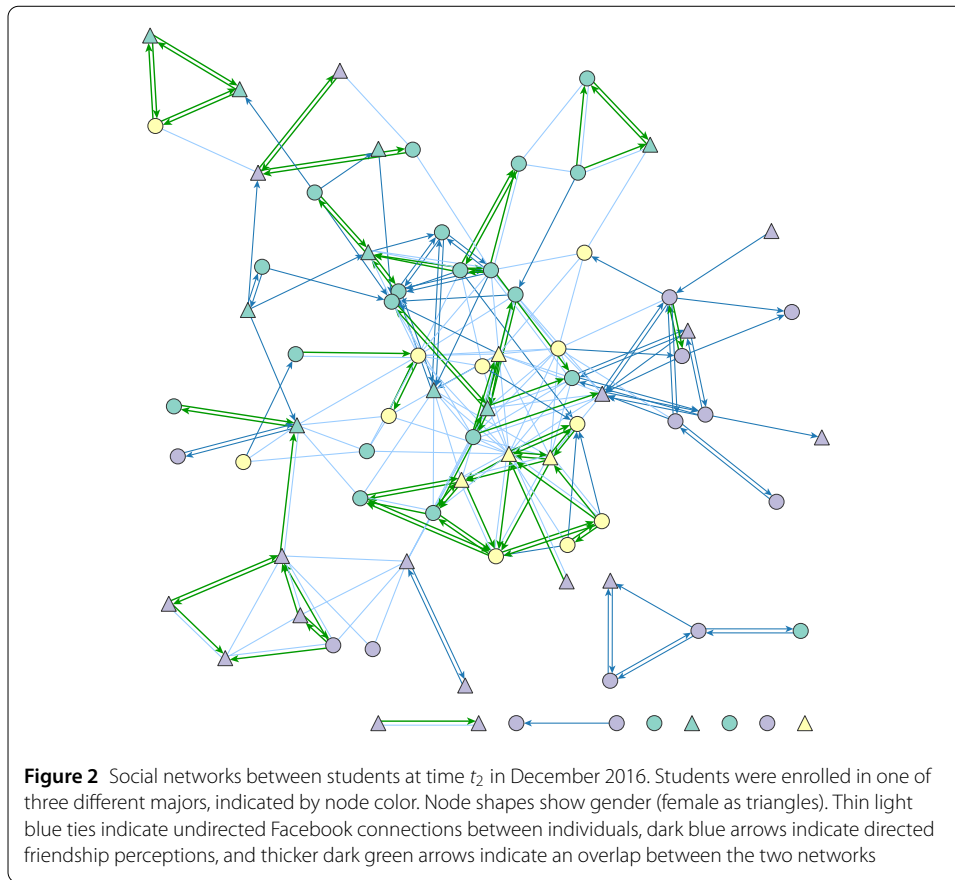
The study period covers data from October 17 to December 19, 2016, which started approximately one month after students joined the university in September, when they primarily met others from their study program for the first time [12]. The offline friendship network was collected in two panels at the beginning and end of the analyzed period. The first friendship network included 148 connections, while the second network consisted of 149 connections. However, the friendship network was highly dynamic: 61 of the ties observed in December did not exist in October, and 60 of the ties reported in October were no longer present in December.

In the same period, 80 pairs of individuals created a new Facebook connection. These data were recorded as time-stamped events. At the time of the first survey wave in October, 52 Facebook ties already existed and – as no social media tie was deleted during the study period – in December the online network consisted of 132 ties. The initial and final networks are plotted separately for each network type in Fig. 1, which also includes three intermediate states of the Facebook network. The two types of relationships exhibit some overlap, with 29 cases of overlapping ties at the first survey wave in October and a slightly higher overlap of 41 ties at the second survey wave in December. The second observed network, consisting of both Facebook ties and friendship perceptions, is shown in Fig. 2.

During the period preceding the October survey, 2.08 Facebook events per day over the whole population (std. dev. 2.96) were observed, with a minimum of 0 and a maximum of 11 events in a single day. During the analyzed period between October and December, the observed rate decreased to 1.25 tie creations per day on average (std. dev. 2.15), ranging from 0 to 15 events per day.

Information on research ethics and IRB approval, enumeration and incentivization, data availability, and additional details on the data collection and study design are available in Vörös et al. [12].

³Facebook data were collected from September 2016 to April 2017, but then stopped as Facebook restricted its policy for research access. Three waves of friendship survey data are available during this period, which were collected simultaneously with the online data. Here, we use two of those data waves. The offline study continued for another three years, but unfortunately, no additional online data could be collected through the Facebook API.



6 Results

The results of the new events-states model are presented in Table 1. These results represent various network mechanisms that are included in the activity rate of individual actors in the upper part of the table (modeled as in Equation (1)) or in the choices of tie receivers in the lower part (modeled as in Equation (2)). The proposed new model is represented by the first three columns of coefficients, which are estimates of the parameter vector θ in Equation (5), referring to the tendencies to create Facebook ties, create friendship ties, and delete friendship ties. An alternative to jointly modeling the two networks is to conduct two separate analyses: one focusing on online dynamics and the other on offline dynamics as the dependent variable. In this approach, information from the other network is treated as static and is considered exogenous. Consequently, we compared the proposed model to a Dynamic Actor Network Model (DyNAM) for the online dynamics, and a Stochastic Actor-oriented Model (SAOM) for the network panel data. The results from these models are displayed in the fourth and last two columns of Table 1. The DyNAM models the relational event dynamics of the growing Facebook network, and considers the survey-based friendship data from the first observation as a static network covariate. The SAOM is the most commonly applied statistical network model for survey-based offline data and uses the online information available at the first observation as a static network covariate. Our attention will be on the sign and significance of the estimators, rather than their size.

The *Rate* (upper part of Table 1) of the three proposed submodels consists of an intercept that governs individuals' general tendency to create and delete ties in the respective

Table 1 Regression coefficients from the proposed new model and from a comparable Dynamic Network Actor Model (DyNAM) and a Stochastic Actor-oriented Model (SAOM). *Online* coefficients refer to the dynamics of the Facebook network, *offline* coefficients to those of the Friendship survey network. Some effects consider the respective *other* network. Significance: † $p < 0.1$, * $p < 0.05$, ** $p < .01$, *** $p < .001$

Submodel	Effect	Symbol	Events-States model			DyNAM		SAOM	
			Online	Offline	Deletion	Online	Creation	Offline	Deletion
Rate	Intercept		-16.087 (0.185)	-15.791 (0.235)	-16.959 (0.329)	-16.411 (0.182)	-	-	-
	Outdegree		0.305*** (0.035)	-0.001 (0.083)	0.472*** (0.073)	0.348*** (0.034)	-	-	-
	Ego female		0.386† (0.232)	0.118 (0.260)	0.210 (0.281)	0.366 (0.233)	-	-	-
	Tie other		2.047*** (0.388)	2.500*** (0.684)	0.199 (0.609)	1.754*** (0.418)	4.080*** (0.909)	0.321 (0.971)	0.321 (0.971)
Choice	Reciprocity		-	3.614*** (0.404)	-1.022* (0.397)	-	3.658† (2.105)	-1.445 (1.941)	-
	Transitivity		0.697** (0.238)	2.019*** (0.331)	-0.496 (0.314)	0.737** (0.232)	1.307*** (0.334)	0.210 (0.534)	0.210 (0.534)
	Trans. other		0.548* (0.272)	-0.679 (0.731)	0.012 (0.439)	0.323 (0.317)	-1.114 (1.623)	-2.239 (1.936)	-2.239 (1.936)
	Indegree		0.223*** (0.051)	0.066 (0.110)	0.016 (0.140)	0.202*** (0.051)	0.065 (0.245)	-0.120 (0.274)	-0.120 (0.274)
	Indeg. other		0.031 (0.081)	-0.138 (0.100)	0.017 (0.131)	0.100 (0.083)	-0.065 (0.253)	0.407* (0.202)	0.407* (0.202)
	Same gender		-0.299 (0.239)	-0.127 (0.304)	-0.162 (0.408)	-0.339 (0.237)	-0.551 (0.386)	-0.472 (0.460)	-0.472 (0.460)
	Same major		0.772** (0.246)	1.848*** (0.380)	0.594 (0.472)	0.945*** (0.241)	1.582* (0.770)	0.272 (0.847)	0.272 (0.847)
	Alter female		0.406† (0.241)	0.730* (0.308)	-0.553 (0.426)	0.440† (0.242)	0.392 (0.458)	-0.478 (0.981)	-0.478 (0.981)

network as well as two mechanisms relating to actors' outdegree and gender. The *Out-degree* effect reveals that individuals with a larger online network are more likely to form additional ties (0.305***) while individuals with larger friendship networks are more likely to terminate ties (0.472***). Female students (*Ego female*) were more actively involved in the creation of new ties online (0.386[†]), but gender had no significant effect on the rate of tie creation and tie deletion offline. The *Intercept* effects have no substantive interpretation besides governing the baseline tendency of individuals to change their networks.

The *Choice* (lower part of Table 1) of the three submodels was specified using mechanisms related to ties between sender and receiver, transitive structures involving a third node, popularity structures, and individuals' attributes. The *Tie other* effect indicates that the presence of a tie between two individuals in one network had a significant effect on the creation of a tie in the other network (2.047***, 2.500***). However, being connected online did not reduce the probability of deleting an offline friendship tie. While *Reciprocity* was not applicable in the online context, as ties are inherently two-sided, it was significant in the offline network. This indicates that an incoming perception of friendship was associated with a higher probability of forming a friendship tie (3.614***) and a lower likelihood of terminating it (−1.022*). Evidence for *Transitivity* in tie creation was found in both networks. Individuals were more likely to create an online tie if they were connected online through a third node (0.697**) and to become friends offline when they had a common offline friend (2.019***). The *Transitivity other* effect was significant in the online model, indicating that individuals were likely to connect on Facebook with friends of their offline friends (0.548*). Transitive structures had no significant effect on maintaining offline ties. Evidence for degree popularity or preferential attachment was only found in the Facebook network, as indicated by the *Indegree* effect (0.223***). There was further no evidence for popularity effects spilling over from the respective other network (*Indegree other*). Gender homophily (*Same gender*) had no significant effect in any of the submodels. Being in the *Same major* of the study program and thus sharing a higher number of courses had a significant effect on the creation of online ties (0.772**) and offline ties (1.848***). Being in the same major did not alter the probability of deleting an offline friendship tie. Female individuals (*Alter female*) were more likely to be chosen as Facebook connections (0.406[†]) and more likely to be nominated as friends (0.730*).

The comparison with the DyNAM model reveals small differences with regard to two coevolutionary online-offline mechanisms. The *Tie other* effect is also significant in the DyNAM model with a slightly smaller coefficient (1.754*** vs. 2.047***). This indicates that the friendship network at the first time point already entails relevant information for the model. In contrast, the *Transitivity other* effect is not significant in the DyNAM but larger and significant in the proposed model (0.323 vs. 0.548*). This suggests that transitive closure of online ties follows two-paths in the friendship network, but these two-paths are partially generated only during the modeled period. An alternative way of considering the offline network in a DyNAM model is not to rely only on the survey data at t_1 but to generate uniformly distributed chains of tie creations and deletions connecting the two survey networks. These chains are void of information from the relational states model but update the network state to reach the empirically observed friendship network at time t_2 . We performed 100 such random draws of chains and re-estimated the DyNAM model. The *Transitivity other* effect was positive and significant in 25% of these chains and

insignificant in 75% of the cases. Other coefficients that do not use information from the friendship network are relatively similar in size and equal in terms of significance.

The comparison with the SAOM model reveals some differences with respect to the size and significance of the coefficients. The proposed model generally has smaller standard errors, suggesting that the additional information from the online data may have led to a more precise estimation in some cases. In the SAOM model, we found no evidence that reciprocity was associated with a higher stability of friendship ties, or that female individuals were more popular in the friendship network. The SAOM also shows significant evidence for a tendency for students to befriend those with whom they had a Facebook connection at time t_1 (*Tie other*: 4.080***), for reciprocal and transitive tie creation (3.658[†], 1.307***), and for friendships with people from the same study major (1.582*).

Interestingly, and somewhat counterintuitively, the SAOM also suggests a tendency to avoid maintaining friendship ties with those who are more popular on Facebook (0.407*). This finding may be an artifact of the early popularity structures present in the initial Facebook network, which, unlike the proposed model, do not change during the estimation process. The current implementation of the SAOM Maximum Likelihood estimation routine further did not allow us to include tests for outdegree and gender effects in the rate.

7 Discussion

7.1 Empirical insights

We demonstrated the applicability of the proposed framework using a small data set that combines a longitudinal survey panel with social media data collected via the Facebook API. We showed that the social network dynamics, both online and offline, follow similar patterns. For example, both exhibit tendencies towards reciprocation, transitive closure, female popularity, and study program homophily. At the same time, the two types of networks coevolve: being connected on Facebook is associated with becoming offline friends later, and being friends in real life is predictive of connecting on Facebook. Additionally, one type of transitive closure seems to operate across the two networks, as indicated by evidence of connecting on social media with friends of friends in the offline network. This effect is found above and beyond the friendship transitivity mechanisms, indicating that social media ties may signal the presence of weaker interpersonal ties that close transitive friendship structures. From this perspective, the mechanism may be interpreted as a weak transitive closure, i.e., becoming an acquaintance of friends of friends. Alternatively, the social media tie may be the result of visibility and meeting opportunities that arise from being friends with the same person. Although the network mechanisms in each layer (online and offline) were similar, they were not identical. The most notable differences pertained to preferential attachment or degree popularity [20]. These were only observed in the online context, where popular nodes were likely to attract and send more ties, but a reverse mechanism appeared offline, where individuals with many connections were in fact more likely to drop some of those over time, rather than creating or attracting even more. This pattern seems intuitive, as offline relationships require more effort than online ones, which leads to the situation at hand: the density of the offline network mostly maintains itself during the studied period, with almost the same number of events of creation and deletion. In contrast, the online network kept growing and no ties were deleted in the observed period.

Our model was designed to integrate multiple event sequences, one taken from digital traces and one inferred from repeated survey measures. We also applied this feature to separate the mechanisms of offline tie creation and offline tie deletion, treating them as separate sequences (similar to the approach of [30]). This led to some additional interesting findings. Reciprocity was the only mechanism that was significantly associated with the creation of ties and the avoidance of tie deletions. Other mechanisms, such as the effect of being connected on Facebook or belonging to the same study program, were associated with the creation of ties but did not appear to prevent the deletion of a tie. In fact, the deletion parameters in those cases were positive (but insignificant), suggesting that existing ties with individuals from a *different* major or with whom one does *not* connect on social media, while being less likely to be created in the first place, had a tendency towards being more stable.

When comparing the model to two methodological standard approaches, actor-oriented relational event models and stochastic actor-oriented models, we could further demonstrate some benefits of jointly modeling the data. The cross-network transitivity effect could not be picked up by a DyNAM model that treated the offline friendship network as fixed. In a comparison with SAOMs, it appeared that the additional layer of information available through online dynamics supported the estimation of coefficients and prevented the potential misinterpretation of structural artifacts when treating social media data as fixed.

7.2 Future directions

A number of open questions warrant future work: The model that we propose assumes that tie creation is not coordinated and follows an individual's preferences for network structures. However, in the case of social media data, we indeed often have an underlying process in which ties are proposed and then accepted, effectively considering the preferences of both the tie sender and receiver. Future work could consider the coordination model of Stadtfield et al. [32] or similar approaches. We further ignored the missing information on the directions of proposal and acceptance, merely randomizing their order. Future attempts could discuss the identifiability of tie directions and consider using missing information on the sender-receiver directions as part of the EM estimation routine. The presented algorithm is a fully working prototype that requires additional research to improve its computational and numerical efficiency.⁴ Initial research has shown that approaches using different importance sampling weighting schemes [59], adaptive step sizes on the stochastic gradient descent routine [63, 64], or Markov Chain Monte Carlo methods for the generation of offline sequences [48] may be a promising venue for future developments. The implementation of more sophisticated strategies may require longer running times or higher demands on computational capacity. Thus, developments should be accompanied by improvements in resource efficiency, yielding a versatile algorithm. We also kept the comparison to existing methods for statistical social network analysis rather brief. We did not consider alternative model specifications and goodness-of-fit assessments, which we recommend studying in more depth in future work on this topic.

Another challenge for future studies using the proposed method is the collection of suitable social network data combining relational events with relational states. Collecting the

⁴The algorithm ran for 2.5 days, using 26 nodes, and consumed a total of 444.60 GB of memory. Improvements to the algorithm to achieve efficiency and less memory usage are currently under development.

data in this paper was not straightforward and required the recruitment of a population of participants with a high level of trust in the research team. By providing access to their online data and repeatedly answering a long and detailed survey, the participants offered us highly valuable insights into their personal social lives and the dynamics of the community as a whole. An obvious drawback is that the number of participants was very small compared to large-scale studies using digital trace data [1]. However, at the same time, we could collect deep and personal data, including social perceptions, personality traits, psychological states, and contextual factors (for details on the additional variables collected, see [12]). It is unclear how research designs could look in empirical contexts in which establishing a similar level of trust may be more challenging (e.g., professional organizations, existing online communities), particularly if the aim is to scale up to much larger populations [15]. Designing appropriate research studies is also difficult due to the fact that direct data access is often limited and controlled by companies, such as Facebook/Meta [6]. We experienced this in our own work when Facebook closed the API access for researchers like us mid-way through our network study. We were somewhat fortunate to collect social media information in parallel to three waves of survey data, but had initially planned to follow the social media evolution over the course of the whole Bachelor program. Recent proposals using data donations or automated tracking may enable promising alternative research designs comparable to the API-based study presented in this paper [65].

We motivated the development of our model by reviewing trends and limitations in large-scale social network analysis in the field of Computational Social Science [66, 67]. Although the relational events part of the model can be scaled up to large-scale data sets [see, e.g. 68, 69], due to the computational challenge of augmenting relational states data, the model is currently only applicable to small-scale social networks combining relational events and relational states. Empirical explorations with the related RSiena package suggest that practical limits might be in the hundreds or thousands of nodes [e.g. 70], which is well below large-scale data sets in Computational Social Science. In our empirical setting, we augmented a survey-based social network study with additional measures of digital behavioral traces [12], resulting in a small but complete social network data set with two layers of relational states and relational events. Augmenting potentially large-scale social network studies of digital traces with a sample of survey data is an interesting alternative to explore [15, 16]. In practice, this might mean that researchers recruit a sub-sample of a few hundred or thousand users from a large-scale online platform into a survey panel and then jointly analyze the large-scale relational event data with the sub-sample of relational states. Additional research is necessary to extend the model to such study designs. Statistical sampling strategies could be used to improve computational efficiency in the case of large-scale social network data sets [68]. This second sampling strategy has been considered in the modeling of relational events, but not yet in the modeling of relational states data and could further improve computational efficiency, but probably not to a point where the models can be fit to large-scale network data as discussed in the introduction. We hope to encourage future longitudinal social network studies that tackle the challenge of collecting survey data and digital data in the same social context, as well as future research on how the proposed model can be scaled up to larger empirical contexts.

8 Conclusion

The paper presents an analytical framework for the joint analysis of longitudinal network data collected both through surveys (as relational states) and digital data sources (as rela-

tional events), together with an empirical application in the context of an emerging social network between university students in a STEM undergraduate program. We motivated the timeliness of the framework in the context of newly emerging social network studies that consider both behavioral and cognitive aspects and presented a statistical model formulation together with the basic estimation principle. The latter is implemented in a software package in R [57].

Traditionally, social network researchers have relied heavily on statistical tools and research designs that are tailored to survey data [7]. Indeed, survey data are still most valuable when research focuses on the mental representations of individuals' social networks. No digital interface or online database can reliably determine with whom individuals are friends, or who they like, dislike, or remember at all in a given context. At the same time, social ties are increasingly created, experienced, and maintained through digital means. Digital behavioral data are of crucial importance in social networks and may indeed affect and be affected by cognitive social representations of social ties. Social network scholars have long pointed out that social ties have at least a cognitive and behavioral component [13, 14]. We argue that one can be best assessed through surveys, while the other can often be collected automatically and with high temporal granularity, for example, by using online data, as in the case of our empirical study. The resulting data are combinations of relational states and relational events. We believe that future empirical social network studies should ideally consider multiple layers of social networks and longitudinally measure relational events and relational states, even when their temporal granularity differs. We hope that the analytical framework and empirical study presented in this paper can inspire future studies on social network dynamics in an increasingly digital world.

Acknowledgements

The authors thank Timon Elmer for leading the API-based Facebook data collection in 2016 as part of the Swiss StudentLife Study. The authors thank all other members of the Swiss StudentLife Study for their contributions to collecting and preparing the data. Earlier versions of this article were presented at the INSNA Sunbelt Conferences in 2019, 2022 and 2025, the ARS workshop on social network analysis 2025, and multiple Duisterbelt meetings; the feedback provided by the audiences greatly helped us develop the article. The authors acknowledge additional valuable feedback received from Tom Snijders, current and former members of the Social Networks Lab at ETH Zürich and two anonymous reviewers. CS thanks the Corvinus Institute for Advanced Studies (CIAS) at Corvinus University for hosting his research stay during which the article was finalized and prepared for submission.

Author contributions

CS and MEGP contributed equally as first authors to the article. CS contributed to acquiring funding, conceptualization, research design, data collection, methods development, software development, analysis, preparation of the manuscript, supervision. MEGP contributed to conceptualization, research design, methods development, software development, analysis, preparation of the manuscript. MEGP was leading the development and implementation of the proposed algorithm and conducted the presented empirical analyses. VA contributed to conceptualization, research design, methods development, analysis, preparation of the manuscript, supervision.

Funding information

Open access funding provided by Swiss Federal Institute of Technology Zurich. The Swiss StudentLife Study was supported by Swiss National Science Foundation Grants 10001A_169965 and 10DL17_183008; and the rectorate of ETH Zürich.

Data availability

In agreement with the Ethics Commission of ETH Zürich (approval 2016-N-27), we follow a strict protocol ensuring the anonymity of the study participants. Minimal replication data and code for the presented analyses can be obtained from the authors on request. Data for replication and original research studies are generally made available by the StudentLife team to external researchers. Additional information and data request forms can be found on the website of the Swiss StudentLife project [71].

Declarations

Ethics approval and consent to participate

The Swiss StudentLife study was reviewed and approved by the Ethics Commission of ETH Zürich in 2016 (approval 2016-N-27). We followed a detailed informed consent protocol, including an information and consent procedure prior to each survey for active study participants. There was a separate informed consent procedure for the API Facebook data collection. Information about Facebook ties was only collected when both connected individuals had given their consent. Further details on data collection and ethics are provided in Vörös et al. [12].

Competing interests

The authors declare no competing interests.

Author details

¹Social Networks Lab, ETH Zürich, Zurich, Switzerland. ²CIAS, Corvinus University, Budapest, Hungary. ³Department of Statistics and Quantitative Methodology, University of Milano-Bicocca, Milan, Italy.

Received: 2 December 2025 Accepted: 23 June 2026 Published online: 03 July 2026

References

- Lazer D, Pentland A, Adamic L, Aral S, Barabási A-L, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, et al (2009) Computational social science. *Science* 323(5915):721–723
- Conte R, Gilbert N, Bonelli G, Cioffi-Revilla C, Deffuant G, Kertész J, Loreto V, Moat S, Nadal J-P, Sanchez A, et al (2012) Manifesto of computational social science. *Eur Phys J Spec Top* 214(1):325–346
- Edelmann A, Wolff T, Montagne D, Bail CA (2020) Computational social science and sociology. *Annu Rev Sociol* 46(1):61–81
- Elmer T (2023) Computational social science is growing up: why puberty consists of embracing measurement validation, theory development, and open science practices. *EPJ Data Sci* 12(1):1–19
- Ntoutsis E, Fafalios P, Gadiraju U, Iosifidis V, Nejdil W, Vidal M-E, Ruggieri S, Turini F, Papadopoulos S, Krasanakis E, et al (2020) Bias in data-driven artificial intelligence systems—an introductory survey. *Wiley Interdiscip Rev Data Min Knowl Discov* 10(3):1356
- Schoch D, Chan C-h, Wagner C, Bleier A (2024) Computational reproducibility in computational social science. *EPJ Data Sci* 13(1):75
- Robins GL (2015) Doing social network research: network-based research design for social scientists. Sage, Thousand Oaks
- Kane GC, Alavi M, Labianca G, Borgatti SP (2025) Guest editorial: the evolution of social media networks: revisiting and extending 10 years later. *MIS Q* 49(4):ix–xvi
- Butts CT, Lomi A, Snijders TAB, Stadtfield C (2023) Relational event models in network science. *Netw Sci* 11(2):175–183
- Meijerink-Bosman M, Back M, Geukes K, Leenders R, Mulder J (2023) Discovering trends of social interaction behavior over time: an introduction to relational event modeling: trends of social interaction. *Behav Res Methods* 55(3):997–1023
- Lazer DM, Pentland A, Watts DJ, Aral S, Athey S, Contractor N, Freelon D, Gonzalez-Bailon S, King G, Margetts H, et al (2020) Computational social science: obstacles and opportunities. *Science* 369(6507):1060–1062
- Vörös A, Boda Z, Elmer T, Hoffman M, Mephram K, Raabe IJ, Stadtfield C (2021) The Swiss StudentLife study: investigating the emergence of an undergraduate community through dynamic, multidimensional social network data. *Soc Netw* 65:71–84
- Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. *Science* 323(5916):892–895
- Kitts JA (2014) Beyond networks in structural theories of exchange: promises from computational social science. In: *Advances in group processes*, vol 31. Emerald Group Publishing Limited, Leeds, pp 263–298
- Stier S, Breuer J, Siegers P, Thorson K (2020) Integrating survey data and digital trace data: key issues in developing an emerging field. *Soc Sci Comput Rev* 38(5):503–516
- Beuthner C, Breuer J, Jünger S (2021) Data linking—linking survey data with geospatial, social media, and sensor data. *GESIS Survey Guidelines*
- Schechter A, Contractor N (2024) Relational event models. In: McLevey J, Scott J, Carrington PJ (eds) *The sage handbook of social network analysis*, 2nd edn. Sage, London. Chap. 35
- Stadtfield C, Amati V (2021) Network mechanisms and network models. In: *Research handbook on analytical sociology*. Edward Elgar, Cheltenham Glos, pp 432–452
- Granovetter MS (1973) The strength of weak ties. *Am J Sociol* 78(6):1360–1380
- Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512
- McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. *Annu Rev Sociol* 27(1):415–444
- Mesch GS, Talmud I (2006) Online friendship formation, communication channels, and social closeness. *Int J Int Sci* 1(1):29–44
- Lin H, Wang Y, Lee J, Kim Y (2023) The effects of disagreement and unfriending on political polarization: a moderated-mediation model of cross-cutting discussion on affective polarization via unfriending contingent upon exposure to incivility. *J Comput-Mediat Commun* 28(4):1–13
- Macskassy S, Michelson M (2011) Why do people retweet? Anti-homophily wins the day! In: *Proceedings of the international AAAI conference on web and social media*, vol 5, pp 209–216
- Wachs J, Hannák A, Vörös A, Daróczy B (2017) Why do men get more attention? Exploring factors behind success in an online design community. In: *Proceedings of the international AAAI conference on web and social media*, vol 11, pp 299–308
- Uzaheta A, Amati V, Stadtfield C (2025) Modeling the duality of content niches and user interactions on online social media platforms. *Soc Netw* 83:152–172

27. Pentland A (2010) *Honest signals: how they shape our world*. MIT Press, Cambridge
28. Stopczynski A, Sekara V, Sapiezynski P, Cuttone A, Madsen MM, Larsen JE, Lehmann S (2014) Measuring large-scale social networks with high resolution. *PLoS ONE* 9(4):95978
29. Snijders TAB, Van De Bunt GG, Steglich CEG (2010) Introduction to stochastic actor-based models for network dynamics. *Soc Netw* 32(1):44–60
30. Krivitsky PN, Handcock MS (2014) A separable model for dynamic networks. *J R Stat Soc, Ser B Stat Methodol* 76(1):29–46
31. Butts CT (2008) A relational event framework for social action. *Sociol Method* 3:155–200
32. Stadtfield C, Hollway J, Block P (2017) Dynamic network actor models: investigating coordination ties through time. *Sociol Method* 47(1):1–40
33. Hofstra B, Corten R, Van Tubergen F, Ellison NB (2017) Sources of segregation in social networks: a novel approach using Facebook. *Am Sociol Rev* 82(3):625–656
34. Kwak D, Kim W (2017) Understanding the process of social network evolution: online-offline integrated analysis of social tie formation. *PLoS ONE* 12(5):0177729
35. Standlee A (2019) Friendship and online filtering: the use of social media to construct offline social networks. *New Media Soc* 21(3):770–785
36. Schecter A, Contractor N (2016) Modeling the joint dynamics of relational events and individual states. In: 2016 IEEE/ACM international conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, pp 1087–1094
37. Yang C, Mousavi S, Dash A, Gummadi KP, Weber I (2025) Studying behavioral addiction by combining surveys and digital traces: a case study of TikTok. *Proc Int AAAI Conf Web Soc Media* 19:2106–2123
38. Subrahmanyam K, Reich SM, Waechter N, Espinoza G (2008) Online and offline social networks: use of social networking sites by emerging adults. *J Appl Dev Psychol* 29(6):420–433
39. Zúñiga HG, Barnidge M, Scherman A (2017) Social media social capital, offline social capital, and citizenship: exploring asymmetrical social capital effects. *Polit Commun* 34(1):44–68
40. Cole H, Griffiths MD (2007) Social interactions in massively multiplayer online role-playing gamers. *CyberPsychol Behav* 10(4):575–583
41. Steinfield C, Ellison NB, Lampe C (2008) Social capital, self-esteem, and use of online social network sites: a longitudinal analysis. *J Appl Dev Psychol* 29(6):434–445
42. Spottswood EL, Wohn DY (2020) Online social capital: recent trends in research. *Curr Opin Psychol* 36:147–152
43. Berger K, Klier J, Klier M, Probst F (2014) A review of information systems research on online social networks. *Commun Assoc Inf Syst* 35:145–172
44. Prochnow T, Patterson MS (2024) It's not just a game: social networks, isolation and mental health in online gamers. *J Public Ment Health* 23(2):140–153
45. Masrom MB, Busalim AH, Abuhassna H, Mahmood NHN (2021) Understanding students' behavior in online social networks: a systematic literature review. *Int J Educ Technol Higher Educ* 18(1):6
46. Al-Yazidi S, Berri J, Al-Qurishi M, Al-Alrubaihan M (2020) Measuring reputation and influence in online social networks: a systematic literature review. *IEEE Access* 8:105824–105851
47. Kahai SS, Lei Y (2019) Building social capital with Facebook: type of network, availability of other media, and social self-efficacy matter. *Int J Hum-Comput Stud* 130:113–129
48. Snijders TAB, Koskinen J, Schweinberger M (2010) Maximum likelihood estimation for social network dynamics. *Ann Appl Stat* 4(2):567–588
49. Stadtfield C, Block P (2017) Interactions, actors, and time: dynamic network actor models for relational events. *Sociol Sci* 4:318–352
50. McFadden D (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) *Frontiers in econometrics*. Academic Press, New York, pp 105–142. Chap. 4
51. Stadtfield C, Geyer-Schulz A (2011) Analyzing event stream dynamics in two-mode networks: an exploratory analysis of private communication in a question and answer community. *Soc Netw* 33(4):258–272
52. Wang J-L, Zhong Q (2025) Joint modeling of longitudinal and survival data. *Annu Rev Stat Appl* 12(1):449–476
53. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B, Methodol* 39(1):1–22
54. Robbins H, Monro S (1951) A stochastic approximation method. *Ann Math Stat* 22(3):400–407
55. Caffo BS, Jank W, Jones GL (2005) Ascent-based Monte Carlo expectation-maximization. *J R Stat Soc, Ser B Stat Methodol* 67(2):235–251
56. Ruth W (2024) A review of Monte Carlo-based versions of the EM algorithm. *arXiv* <http://arxiv.org/abs/2401.00945>
57. Hollway J, Stadtfield C, Hoffman M, Uzaheta A (2025) Goldfish: statistical network models for dynamic network data. R package version 1.6.12. <https://CRAN.R-project.org/package=goldfish>
58. Levine RA, Casella G (2001) Implementations of the Monte Carlo EM algorithm. *J Comput Graph Stat* 10(3):422–439
59. Elvira V, Martino L, Luengo D, Bugallo MF (2019) Generalized multiple importance sampling. *Stat Sci* 34(1):129–155
60. Needell D, Srebro N, Ward R (2016) Stochastic gradient descent, weighted sampling, and the randomized Kaczmarz algorithm. *Math Program* 155(1–2):549–573
61. Csiba D, Richtárik P (2018) Importance sampling for minibatches. *J Mach Learn Res* 19(27):1–21
62. Stadtfield C, Vörös A, Elmer T, Boda Z, Raabe IJ (2019) Integration in emerging social networks explains academic failure and success. *Proc Natl Acad Sci USA* 116(3):792–797
63. Li X, Orabona F (2019) On the convergence of stochastic gradient descent with adaptive stepsizes. In: The 22nd international conference on artificial intelligence and statistics, pp 983–992. PMLR
64. Bottou L (2012) Stochastic gradient descent tricks. In: Montavon G, Orr GB, Müller K-R (eds) *Neural networks: tricks of the trade*, 2nd edn. Springer, Berlin, pp 421–436
65. Ohme J, Araujo T, Boeschoten L, Freelon D, Ram N, Reeves BB, Robinson TN (2024) Digital trace data collection for social media effects research: apis, data donation, and (screen) tracking. *Commun Methods Meas* 18(2):124–141
66. Lazer D, Pentland A, Adamic L, Aral S, Barabási A-L, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, et al (2009) Computational social science. *Science* 323(5915):721–723.

67. Conte R, Gilbert N, Bonelli G, Cioffi-Revilla C, Deffuant G, Kertesz J, Loreto V, Moat S, Nadal J-P, Sanchez A, et al (2012) Manifesto of computational social science. *Eur Phys J Spec Top* 214(1):325–346
68. Overgoor J, Pakapol Supaniratisai G, Ugander J (2020) Scaling choice models of relational social data. In: *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. KDD '20. Assoc. Comput. Mach.*, New York, pp 1990–1998
69. Lerner J, Lomi A (2020) Reliability of relational event model estimates under sampling: how to fit a relational event model to 360 million dyadic events. *Netw Sci* 8(1):97–135
70. Veenstra R, Bertogna T, Laning-Wijnen L (2023) The growth of longitudinal social network analysis: a review of the key data sets and topics in research on child and adolescent development. In: *Feinberg ME, Osgood DW (eds) Teen friendship networks, development, and risky behavior. Oxford University Press, Oxford*, pp 326–352
71. Social Networks Lab, ETH Zürich The Swiss StudentLife Study Website. <https://sn.ethz.ch/research/studentlife>

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.