Social Networks 43 (2015) 121–135

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Relational event models for social learning in MOOCs

Duy Vu[a,*], Philippa Pattison[b], Garry Robins[a]

[a] University of Melbourne, VIC 3010, Australia
[b] University of Sydney, NSW 2006, Australia

ARTICLE INFO

Keywords:
Relational event models
Social learning analytics
MOOCs

ABSTRACT

We propose three extensions for the relational event framework to model the co-evolution of multiple network event streams which are increasingly available thanks to the explosive growth of online applications. Firstly, a flexible stratification approach is considered to allow for more complex data structures with many types of nodes and events. Secondly, an inference method that combines nested case–control sampling with stratification is discussed to scale the approach to very large data sets. Finally, a suite of new temporal and network statistics is introduced to reflect the potentially complex dependencies among events and observed heterogeneities on nodes and edges.

The empirical value of new extensions is demonstrated through an analysis of social learning in Massive Open Online Courses (MOOCs). In particular, three modeling problems are considered from the network perspective: (1) the utility of social factors, performance indicators, and clickstream behaviors in the prediction of course dropout, (2) the social and temporal structure of learner interactions across discussion threads, and (3) the forms of mutual dependence of social learning interactions on prior learning success, and future learning success on forms of prior social learning interaction.

1. Introduction

The increasing growth of online social platforms such as Facebook, MOOCs, and GitHub has provided people with more opportunities to connect, learn, and collaborate. Large activity data streams generated by different parts of an application or across multiple applications allow us to explore in detail the dynamics of user interactions and behaviors, which in turn can help to design better support tools and improve their future experience. In MOOCs, for example, besides user activities on forums and assignments, clickstream data of their learning behaviors such as video interaction events can also be observed. To understand the social aspect of these online learning processes and hence improve learner engagement, it is important to model the interdependence among learners' behaviors and interactions across these event streams. This paper will demonstrate how relational event models (Butts, 2008) can provide us with a methodological framework to achieve this goal, even with complex data sets at the Web scale.

Currently, there are three statistical modeling approaches that can take account of local network structures. Each framework aims at a different network data type. First, Frank and Strauss (1986) proposed exponential random graph models (ERGMs) for cross-sectional networks where the general Markov dependence assumption allows us to model both local structures and other observed characteristics of nodes and edges (Robins et al., 2007). The inadequacies of Markov dependence and newer social circuit dependence specifications can be found in Snijders et al. (2006) while a thorough discussion on computationally intensive Markov Chain Monte Carlo inference is given by Hunter and Handcock (2006). Recent advances in ERGMs for multi-level network data of different types of edges among different sets of nodes have been discussed in Wang et al. (2013, 2013).

The second modeling approach is stochastic actor-oriented models (SAOMs) (Snijders, 2001) which are suitable for panel network data. Based on the Markov process assumption, this framework models the probabilities of edge and behavior changes as functions of the current network itself and other observed characteristics of nodes and edges. Consequently, competing structural and behavioral tendencies that simultaneously drive the dynamics of network processes can be jointly estimated under SAOMs. A non-technical introduction of SAOMs and its representative applications can be found in Snijders et al. (2010). Similar to ERGMs, statistical inference for SAOMs is also carried by time-consuming MCMC procedures that simulate micro changes between discrete-time network observations (Schweinberger and Snijders, 2007).

* Corresponding author. Tel.: +61 3 8344 5550.
E-mail addresses: duy.vu@unimelb.edu.au (D. Vu), pip.pattison@sydney.edu.au (P. Pattison), garrylr@unimelb.edu.au (G. Robins).

http://dx.doi.org/10.1016/j.socnet.2015.05.001
0378-8733/

The third modeling approach for time-stamped network data is relational event models (REMs) (Butts, 2008). Recent advances in statistical inference for REMs have been explored by Perry and Wolfe (2013) while a range of its different applications have been demonstrated in Brandes et al. (2009), Vu et al. (2011), DuBois et al. (2013), Quintane et al. (2013), Lomi et al. (2014). Compared to ERGMs and SAOMs, statistical inference of REMs is less intensive thanks to the tractability of partial likelihoods (Cox, 1972) and the sparsity of network statistic changes (Vu et al., 2011; Perry and Wolfe, 2013). Online social networks of ten thousands of nodes have been successfully analyzed using this partial likelihood inference method (Salathé et al., 2013). The most appealing feature of REMs, however, is its capability in taking full advantage of time-stamped activity data which are continuously monitored and recorded in online applications. Using ERGMs or SAOMs, and collapsing these fine-grained temporal data into one cross-sectional or some discrete-time network snapshots could result in the loss of information on the dynamics of network processes. Consequently, the estimation of network effects could be severely biased.

Motivated by a new problem of network analysis in social learning, this paper seeks to make four significant contributions in relational event modeling. Firstly, based on the counting process approach in survival and event history analysis (Andersen et al., 1993), we discuss a flexible stratification method to model multi-mode and multiplex network event streams. Such data are increasingly available in online applications, but relational event models for them have not been fully considered. In online learning systems, for example, many relational event streams on user activities can be simultaneously recorded, including direct instant messages among learners, forum posts between learners and discussion threads, user assignment submission and video interaction events. These diverse collections of learner interactions and behaviors are interdependent and need to be analyzed in a joint modeling framework. For example, by modeling both post events from users to discussion threads and submission events between users and quizzes, the bi-directional relationship between learning performance and social interactions in forums can be uncovered.

Our second contribution is a faster estimation procedure to address the computational challenge in the relational event framework (Butts, 2008, Section 2.3). Although recent progress has been made based on the sparsity property of count-based network statistics (Perry and Wolfe, 2013; Vu et al., 2011), our introduction of new temporal statistics requires other efficient alternatives. In particular, we discuss the application of nested case–control sampling (Borgan et al., 1995) in relational event modeling and its combination with stratification that could help to increase the sampling efficiency.

Our third contribution is a set of advanced network statistics that can take account of observed heterogeneities on nodes and edges as well as complex relational and temporal dependencies among event streams. For example, to model the interdependence among different interaction processes in MOOCs such as forum activities and quiz submissions, we introduce a set of network statistics that are calculated from multiple event types on different sets of nodes. These statistics, for instance, allow us to test whether high-performance learners tend to engage more with each other over time in discussion threads of common interests than low-performance ones. Another example is our application of recency statistics which are broadly used in survival analysis (Aalen et al., 2008) to test for the tendency that learners' behavior events are clustered rather than equally distributed over time.

Our last contribution is a novel application of REMs in social learning analytics (Shum and Ferguson, 2012). In order to explore the role of social structures in emergent networked learning environments, we consider three modeling problems of course dropout, quiz performance, and forum discussion. Our analysis, for example, shows that learners with high cumulative quiz scores are more likely to engage in forum discussions. However, high activity in forums is not associated with better quiz scores though it predicts a lower likelihood of dropping out. The analysis also uncovers interesting network structures in forum discussion such as four-cycle closure effects where learners, especially those with high quiz scores, tend to maintain their knowledge exchange collaboration over time.

The rest of this paper is organized as follows. In Section 2, we briefly introduce the emerging field of social learning analytics and the structure of relational event streams in MOOCs using a case-study data set from Coursera. The section concludes with an outline of main research questions that will be considered in our social learning analysis. The flexible stratification approach for multi-mode and multiplex event streams is then presented in Section 3 followed by a discussion in Section 4 on the nested case–control sampling and stratification procedure that can scale statistical inference for REMs to large data sets. Section 5 explores a set of advanced network statistics that substantially extend the modeling capability of REMs. Our analysis of social learning in MOOCs is presented in Section 6 to demonstrate the empirical value of all proposed extensions for REMs. Section 7 will sketch out some research problems in relational event modeling that will be investigated in our future work to further promote its applications.

2. Social learning in MOOCs

The field of learning analytics in higher education is burgeoning (Siemens and Gasevic, 2012; Siemens, 2013). Interest has been fueled by the increasing use of learning management systems (LMSs) in universities and other post-secondary educational institutions and the recognition – and early demonstration – that the data recorded in these systems can be used to improve student retention and student learning (Tanes et al., 2011; Arnold and Pistilli, 2012). It has also been boosted by the very large volume of continuous-time clickstream data generated by learners in Massive Open Online Courses. MOOC data are distinctive not only because they capture a substantial proportion of students' learning-focused interactions in online learning communities that are literally distributed around the globe, but also because of the large numbers of learners involved.

While much learning analytics literature has focused on the effect of the conditions created for learning as well as individual learner behavior on successful learning, research has increasingly turned to the role of social interaction in understanding students' learning behavior and learning success (Dawson, 2010; Haythornthwaite and Andrews, 2011). As educational theorists have long argued, learning is situated within a particular sociocultural context, and peer-to-peer interaction plays a vital role not only in building and maintaining engagement with learning activities but also in supporting forms of collaborative exchange that promote learning (Blackmore, 2010; Laurillard, 2001). Shum and Ferguson (2012) propose the term "social learning analytics" to refer to a distinctive subset of learning analytics that is socially situated. Social learning analytics draws on the substantial body of work demonstrating that new skills and ideas are not solely individual achievements, but are developed, carried forward, and passed on through interaction and collaboration. It goal is to identify behavior and patterns of interaction that promote learning and is important from both theoretical and practical perspectives: theoretically, because it explicates the important role of social interactions in creating opportunities for learning; practically, because current social interactions are a potentially powerful predictor of future social interactions and hence future learning opportunities and outcomes.