Creating Social Network Models from Sensor Data

Danny Wyatt Tanzeem Choudhury Jeff Bilmes

1 Introduction

Complex macro-social phenomena can arise from simple micro-level behavior without any global coordination (e.g. racial segregation in neighborhoods can occur simply from individuals wanting to avoid being in the minority even in a population that prefers diversity [6]). Such simple local rules and assumptions are only rarely empirically grounded because they require studying a given social system at an individual scale and at a spatial and temporal resolution not possible using manual non-automated techniques. As a result, few studies have attempted to analyze social interactions that involve micro-level behavior. We are interested in grounding simple behavioral assumptions with real observations, and in testing the effects of these behaviors at the macro-level dynamics of groups and larger social systems. To this end, we instrumented a subject population with wearable sensors that recorded them as they went about their lives over the course of 9 months. We believe this longitudinal dataset will offer fertile opportunities to explore the type of questions described above. By continuously studying individuals in a social system, our goal is to discover the simple rules people follow or the behaviors they exhibit and how these affect the structure and dynamics of social networks. For example, is the interaction likelihood of two people proportional to their physical proximity, or instead to the stylistic similarity of their conversational dynamics [1] or body language? How predictive are the social network topologies that will emerge?

2 The Dataset

We recruited 24 subjects from the incoming class of the graduate program of a single department at a large research university. To collect data, each subject wore an HP iPAQ Personal Digital Assistant (PDA) with an attached multi-sensor board (MSB) containing 8 different sensors. The MSB was worn on the front of the wearer's shoulder, similar in placement to a lapel microphone (see 1)–for long-term usability, a close-talking microphone was not appropriate. Of the 8 sensors, the microphone is clearly the most important sensor for conversation detection. In our earlier [9, 8] work we developed algorithms to automatically segment multi-person conversations from the streaming sensor data and segment speakers within those conversations. From the segmented conversations, a variety of attributes about the conversational dynamics (which maybe be indicative of the network dynamics) can be computed, such as turn-taking, duration of voiced segments, pauses, speaking rate, mean and variance of pitch, mean and variance of the signal amplitude, duration of conversation, etc.

The MSB also contains 7 other sensors that sample at varying rates: triaxial accelerometer (550 Hz), visible light (550 Hz), digital compass (30 Hz), temperature and barometric pressure (15 Hz), infrared light (5 Hz), and humidity (2 Hz). These sensors can be used to infer the wearer's physical activity (e.g. walking, sitting, standing) and whether she is indoors or outdoors [5]. In addition to the data gathered via the MSB, the PDA records (at 0.5 Hz) the MAC addresses and signal strengths of the 32 strongest WiFi access points nearby. This WiFi data can be used to determine the wearer's location [4]. Data was collected during working hours for one week each month over the 9 month course of an academic year. At the end of every collection week each subject filled out a survey. The survey asked questions about: which other participants the subject interacted with and how (e.g. homework collaboration, research collaboration, socially, etc.), which sub-areas of the discipline the subject was interested in pursuing, which faculty members the subject had collaborated with, etc. Once each term, the survey also asked which classes the subject was taking, how she was funded, and whom she considered her advisor.

3 The Model

Given the sensor data, our goal is to infer the underlying static network structure (and ultimately the underlying temporal dynamics). We generalize the standard exponential random graph model [2, 7] to handle missing information that is specific to our dataset but is likely to be common in other multi-person sensor streams as well. First, we do not directly measure social ties–we only observe information about the duration that pairs of individuals spend in conversation over time window samples. Second, the time windows are not always fully observed – thus, the true proportion of time spent together conversing is not entirely known.

In order to reliably learn network structure under such missing information, we consider all edges as hidden (similar to [3]) and represent observational uncertainty in a Markov random field identically to the way virtual evidence is used in Bayesian networks. Specifically, the degree to which a portion of a given time window is missing is proportional to the confidence we place on the observation encoding the relative amount of time a given pair spends together in that window. Handling such missing data through a combination of virtual evidence and hidden edges is a novel aspect of our ERGM model.

$$p(y) = \frac{1}{Z} e^{\theta^{\mathsf{T}} \phi(y)} \tag{1}$$

$$p(x) = \frac{1}{Z} \sum_{y} e^{\theta^{\mathsf{T}} \phi(x,y)} \tag{2}$$

$$p(d) = \frac{1}{Z} \sum_{x} c(d,x) \sum_{y} e^{\theta^{\mathsf{T}} \phi(x,y)} \tag{3}$$

$$= \frac{1}{Z} \sum_{x,y} e^{\theta^{\mathsf{T}} \phi(x,y) + \log c(d,x)} \tag{4}$$

Equation 1 is the likelihood for a traditional ERGM where y represents the edges of the network, ϕ are features of the network, θ are weights to be learned, and Z is a normalizing term. Equation 2 is the likelihood for a model where the edges are hidden and interact indirectly via the observations x. Equation 3 is the likelihood for our model: evidence is only partially observed via d and a confidence function c weights values of x based on d. This is equivalent to 4, a multi-layer undirected graphical model where weights for fixed "virtual evidence features" are untrained and clamped to 1.

Figure 1: The MSB–microphone is on the upper right.

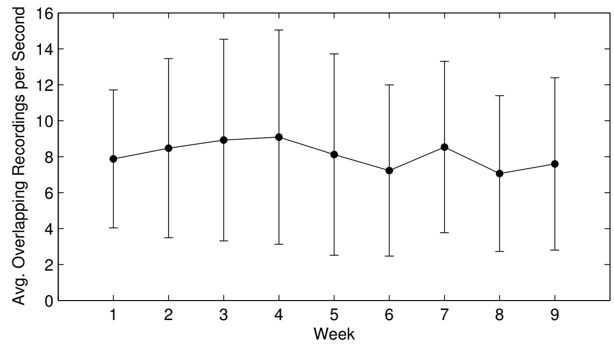

Figure 2: Average overlap in recording over all subjects. The error bars show standard deviation. Thus, it is highly likely that at any given time the data from some subjects will be missing.

With synthetic data, we have had promising initial results reconstructing the hidden network as well as (to a lesser degree) recovering the model parameters. We are currently working on a feature set to be used with our real-world conversation data. We will present results on the conversation data at the workshop.

References

[1] T. Choudhury and S. Basu. Modeling conversational dynamics as a mixed memory markov process. In *Proc. of NIPS*, 2004.

[2] O. Frank and D. Strauss. Markov graphs. *Journal of the American Statistical Association*, 81:832–842, 1986.

[3] F. Guo, S. Hanneke, W. Fu, and E. P. Xing. Recovering temporally rewiring networks: A model-based approach. In *Proc. of ICML*, 2007.

[4] J Letchner, D. Fox, and A. LaMarca. Large-scale localization from wireless signal strength. In *Proc. of AAAI*, 2005.

[5] L. Liao, T. Choudhury, D. Fox, and H. Kautz. Training conditional random fields using virtual evidence boosting. In *Proc of IJCAI*, 2007.

[6] T. Schelling. Dynamic models of segregation. *Journal of Mathematical Sociology*, 1:143–186, 1971.

[7] S. Wasserman and P. Pattison. Logit models and logistic regressions for social networks: I. an introduction to markov graphs and $p*$. *Psychometrika*, 61:401–425, 1996.

[8] D. Wyatt, T. Choudhury, and J. Bilmes. Conversation detection and speaker segmentation in privacy-sensitive situated speech data. In *Proc. of Interspeech*, 2007.

[9] D. Wyatt, T. Choudhury, J. Bilmes, and H. Kautz. A privacy sensitive approach to modeling multi-person conversations. In *Proc. of IJCAI*, 2007.