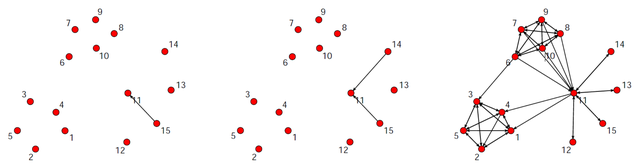
Modeling Relational Events via Latent Classes

Christopher DuBois, Padhraic Smyth

Department of Statistics and Computer Science, University of California, Irvine

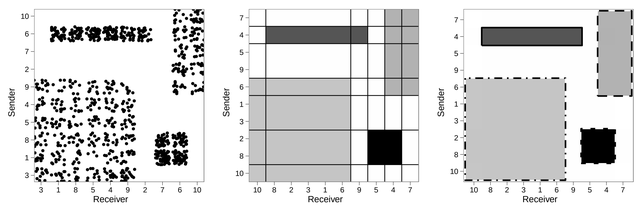
Motivation

- Network data often constructed from sequences of relational events (e.g. human communication data)



Goal: Predict the probability the next event involves sender s, recipient r, and action type a.

Latent Variable Modeling of Social Networks



- We propose the **Marginal Product Mixture Model** which instead considers *latent classes of events*.
- **Block models** instead assume a partition of the individuals and a model for group-wise interactions.
- The MPMM addresses the sparsity issue of direct probability estimation by using the marginal activity for the sender, for example.
- **Generative Model:** For each event
  - Pick latent class $c \sim \text{Categorical}(\pi)$
  - Pick sender $s|c \sim \text{Categorical}(\theta_c)$
  - Pick receiver $r|c \sim \text{Categorical}(\phi_c)$
  - Pick action type $a|c \sim \text{Categorical}(\psi_c)$
- **Inference:** Collapsed Gibbs Sampling (CGS) algorithm iteratively samples a latent class assignment for each observation and continues until convergence. Inner loop is simple and minimal bookkeeping is required.

Exploratory Data Analysis

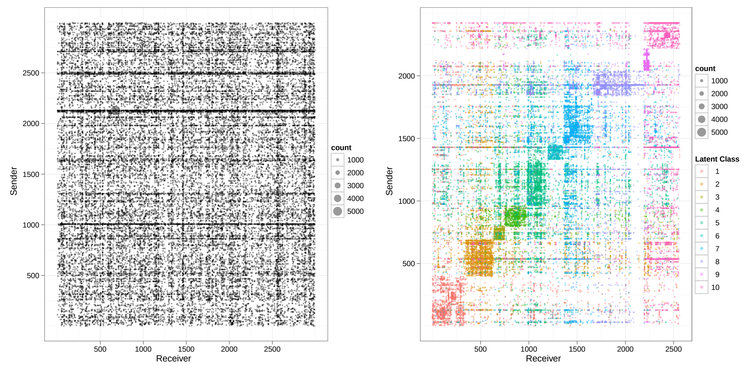


Figure: Left: Counts of email interactions among individuals in the Eckmann data set. Right: MPMM fit to data using $C = 10$. Each event is colored according to its assigned latent class. Rows and columns are rearranged according to which latent class an individual is assigned to most.

Class A

Top Senders	Pr.	Top Receivers	Pr.	Top Actions	Pr.
U.S. : Government agents	0.47	Greece : NA	0.05	Sports contest	0.59
U.S. : Athletes	0.29	Australia : Government agents	0.02	Agree or accept	0.14
U.S. : Nominal agents	0.04	United Kingdom : NA	0.02	Optimistic comment	0.04
U.S. : Police	0.04	Canada : Government agents	0.02	Comment	0.03
U.S. : Occupations	0.04	France : NA	0.01	Control crowds	0.03
U.S. : Ethnic agents	0.03	Belgium : Government agents	0.01	Improve relations	0.01

Class B

Top Senders	Pr.	Top Receivers	Pr.	Top Actions	Pr.
U.S. : Military	0.88	Iraq : Government agents	0.17	Comment	0.19
U.S. : Government agents	0.08	Iraq : National executive	0.07	Military raid	0.14
U.S. : Military hardware	0.01	Iraq : Military	0.05	Military clash	0.10
U.S. : Officials	0.00	Iraq : Ethnic agents	0.05	Military occupation	0.10
U.S. : Police	0.00	Iraq : Intangible things	0.04	Shooting	0.10
U.S. : Motor vehicles	0.00	NA : Insurgents	0.04	Political arrests and detentions	0.04

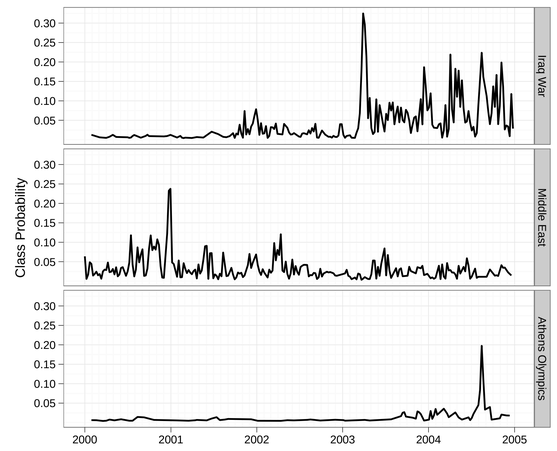


Figure: Exploratory analysis after fitting MPMM with $C = 50$ to a dataset of international political events. Top: Excerpts from distributions of senders, receivers, and action types for each latent class. Bottom: Number of events per week assigned to these three latent classes.

Experimental Results

University Email [1] **International Political Events** [2]

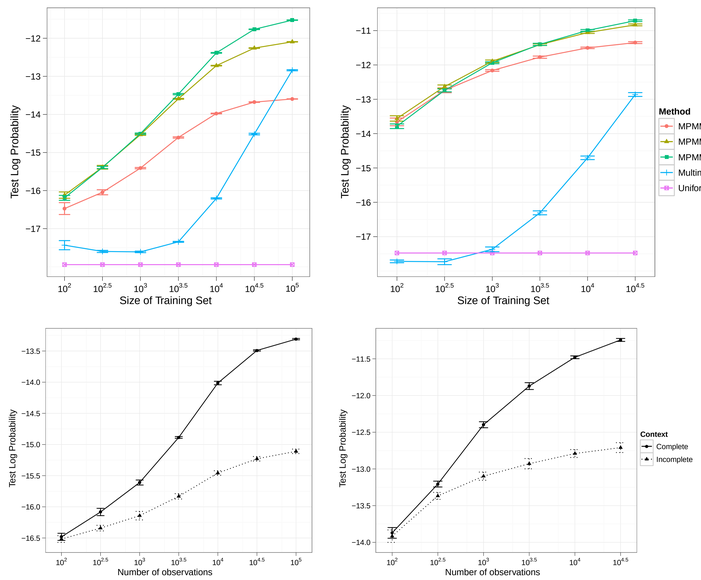


Figure: Top: Test log probability as size of the training set and number of latent classes vary. Baselines include a multinomial-Dirichlet and uniform. Bottom: Effect of removing recipient information on predictive accuracy.

Summary

- Useful method for relational data
  - scalable
  - interpretable
  - robust to missing data and sparse data
  - able to share statistical strength over similar individuals/events
- Extensions
  - D-dimensional categorical data; here we used only $D = 3$.
  - Flexible number of latent classes.
  - Incorporating Beta random variables is straightforward. Helpful to obtain each class's distribution over time.
  - HMM at the class level is a simple extension for another type of time-dependence.
  - Smoothing that is more specific to social networks (e.g. friend-of-a-friend effects).

[1] J.-P. Eckmann, E. Moses, and D. Sergi. Entropy of dialogues creates coherent structures in e-mail traffic. *Proceedings of the National Academy of Sciences of the United States of America, 101(40):14333–7, October 2004.*

[2] G. King,W Lowe. An automated information extraction tool for international conflict data with performance as good as human coders: A rare events evaluation design. International Organization, 57:617-642, 2003.

UNIVERSITY *of* CALIFORNIA • IRVINE

Supported by: Office of Naval Research award N00014-08-1-1015 and a National Defense Science and Engineering Graduate Fellowship.

CENTER FOR MACHINE LEARNING AND INTELLIGENT SYSTEMS
Donald Bren School of Information and Computer Sciences
University of California, Irvine