Inferring Multilateral Relations from Dynamic Pairwise Interactions

Aaron Schein,[*] **Juston Moore, Hanna Wallach**
School of Computer Science
University of Massachusetts Amherst
Amherst, MA 01003
{aschein, jmoore, wallach}@cs.umass.edu

Abstract

Correlations between anomalous activity patterns can yield pertinent information about complex social processes: a significant deviation from normal behavior, exhibited simultaneously by multiple pairs of actors, provides evidence for some underlying relationship involving those pairs—i.e., a *multilateral relation*. We introduce a new nonparametric Bayesian latent variable model that explicitly captures correlations between anomalous interaction counts and uses these shared deviations from normal activity patterns to identify and characterize multilateral relations. We showcase our model's capabilities using the newly curated Global Database of Events, Location, and Tone, a dataset that has seen considerable interest in the social sciences and the popular press, but which has is largely unexplored by the machine learning community. We provide a detailed analysis of the latent structure inferred by our model and show that the multilateral relations correspond to major international events and long-term international relationships. These findings lead us to recommend our model for any data-driven analysis of interaction networks where dynamic interactions over the edges provide evidence for latent social structure.

1 Introduction

Social processes are complex because relationships between actors are not independent. In many domains, we only observe interactions between pairs of actors, but we are interested in learning about the latent structure underlying these pairwise relationships. For example, in international relations, treaties and coalitions influence the pairwise behavior of countries. A particular type of multilateral relation, a coalition, can result in a group of countries taking action toward another country. We are interested in characterizing *multilateral relations*, persistent relationships between pairs of actors in an interaction network. When multilateral relations are inherently latent, temporal correlations in interactions across edges in the network provide evidence for the underlying relations.

Every pair of actors has its own inherent normal rate of activity, which is not influenced by multilateral relations. Machine learning researchers have traditionally relied on normal activity patterns in order to infer latent structural information. There are many applications in which correlations in anomalous behavior yield more pertinent information than correlations in normal activity patterns. A significant deviation from normal behavior, exhibited by multiple pairs of actors, provides evidence for an underlying multilateral relation. Pairs involved in the same multilateral relations are likely to show similar anomalous deviations from their normal interaction rates at the same points in time. Although there are existing models for anomaly detection in complex social processes, most do not model similarities between actors and cannot detect multilateral relations. For example,

---

[*]Aaron Schein and Juston Moore are joint principal authors.

Heard et al.s model for anomaly detection in social networks [1] assumes that the events associated with each actor are independent, and therefore ignores any correlations between actors' anomalous activity patterns. Similarly, Kleinbergs model for identifying periods of bursty observation counts [2] finds only those events involving a single actor. We present a new nonparametric Bayesian latent variable model that explicitly captures correlations between anomalous interactions of pairs, thereby facilitating the detection and characterization of multilateral relations.

We model dynamic interactions between actors as a sequence of weighted graphs, where actors are represented as nodes and the weights on each directed edge represent the number of directed interactions in a particular time slice. This representation is equivalent to considering a single graph with edges annotated with a time series of interaction counts over discrete time slices. Grouping pairs based on anomalous interactions is equivalent to clustering annotated edges in this framework. A large amount of work in network science has been devoted to discovering groups of nodes in a network based on structural properties of the network, where edges are either binary or weighted. These clustering algorithms assign a node to a single cluster [3, 4] or infer a mixed-membership cluster assignment [5]. The existing work largely ignores attributes that might be attached to edges, and in particular doesn't include dynamics of interactions over edges. Instead, we choose to directly model the dynamics of interactions over edges. Our model explicitly groups the edges in an interaction network based on the dynamics of the associated time series. The resulting groups can be interpreted as mixed-membership assignments for the nodes by associating a node with every relation to which its outgoing edges are assigned. However, considering groupings of edges allows us to discover more nuanced patterns of interaction.

We describe the mathematical details of our new model in the next section and outline an efficient procedure for approximate inference in section 3. In section 4.1 we validate our inference procedure by demonstrating that it can recover the parameters used to generate synthetic data. In section 4.3, we apply our model to international relations data consisting of directed pairwise interactions automatically extracted from news articles from around the world between 1979 and the present, newly curated in the Global Database of Events, Language, and Tone (GDELT) [6]. This data is of interest to researchers in the social sciences, as well as government and business intelligence analysts because it provides a high-level view of the relationships between countries over the last several decades. We first show that our model discovers important international events that occurred during the last decade and further provide a detailed analysis of the latent structure inferred by the model.

2 Model

In this section, we present our new nonparametric Bayesian latent variable model for discovering multilateral relations from periods of correlated anomalous pairwise interaction.

We model pairwise interactions occurring over directed edges between a set of N actors. Each directed edge from actor i to actor j is associated with a T-length vector of interaction counts, $\boldsymbol{y}_{ij} = \left\{ y_{ij}^{(t)} \right\}_{t=1}^{T}$, recording the number of times i acted toward j in each discrete time slice t. We assume that edges are partitioned into latent groups. Each edge (ij) has a group assignment z_{ij}; two edges are in the same group if their group assignments are the same. Given its group assignment, each edge draws a number of interactions counts at each time slice t, from a Poisson distribution parameterized by the product of an edge-specific interaction base rate λ_{ij} and a group- and time-specific deviation factor $\delta_{z_{ij}}^{(t)}$: that is, $y_{ij}^{(t)} \sim \mathrm{Pois}\left(\lambda_{ij} \, \delta_{z_{ij}}^{(t)} \right)$. The base rate for each edge describes the expected number of interactions inherent to that edge during periods of normal activity. Each base rate is drawn from a global Gamma prior, with shape and scale parameters $\boldsymbol{\gamma} = \{\gamma_1, \gamma_2\}$; we in turn assume a noninformative prior over each of these hyperparameters. When a group's deviation factor $\delta_{z_{ij}}^{(t)}$ is equal to one, the count of interactions occurring over each edge (ij) in that group is drawn from a Poisson distribution with mean equal to the base rate, λ_{ij}; when the deviation factor increases (or decreases) from one, the mean of the Poisson distribution is increased (or decreased) accordingly. The deviation factors for each group therefore capture shared temporal correlations in the counts of interactions occurring over the edges in that group.

We assume that most of the deviation factors are close to one, and that only the anomalous time slices correspond to deviation factors much greater than or less than one. Our model considers

each deviation factor $\delta_g^{(t)}$ to be a drawn from one of two types of distributions: a *spike* distribution, which is highly peaked around one, and one or more *slab* distributions, which have support above or below one in order to model anomalously high counts, or low counts, respectively. In order to capture periods of anomalous activity, we encode the assumption that anomalies follow anomalous time slices, and normal time slices follow other normal time slices by modeling the deviation factors $\boldsymbol{\delta}_g = \left\{ \delta_g^{(t)} \right\}_{t=1}^{T}$ as "observations" from a K-state hidden Markov model (HMM). The latent states are denoted by $\boldsymbol{s}_g = \left\{ s_g^{(t)} \right\}_{t=1}^{T}$, and the transition probabilities are encoded in a matrix Θ, where the entry $\theta_{ss'}$ gives the probability of a transition from state s at time t to state s' at time $t + 1$. We place an asymmetric Dirichlet prior distribution over each row in Θ (parameterized by a vector $\boldsymbol{\beta}_s$), and assume that the first state $s_g^{(1)}$ is preceded by $s_g^{(0)} = 0$. The HMM emission distribution for state $s_g^{(t)} = 0$ is a low-variance spike distribution so that all corresponding deviation factors will be close to one. The emission distributions corresponding to the other states $s_g^{(t)} \in \{1, \ldots, K - 1\}$ provide support greater than, or less than, one to model increased or decreased values of $\delta_g^{(t)}$, respectively. Each emission distribution for state s is assumed to be a Gamma distribution with shape parameter c_s and scale parameters d_s.

Since it is often impossible to determine the number of multilateral relations underlying observed interactions, we allow the model to infer an appropriate number of groups. The partition of edges implied by their group assignments $\mathcal{Z} = \{z_n\}_{n=1}^{N}$ is drawn from a Chinese restaurant process prior [7] with concentration parameter α. By determining the number of groups in the partition, this prior causes edges to be grouped together only when they exhibit anomalous deviations from their normal interaction base rates at similar points in time.

3   Inference

Given an observed time series of interaction counts for each edge in an interaction network, $\mathcal{Y} = \left\{ \boldsymbol{y}_{ij} \right\}_{i \neq j}$, and hyperparameters $\mathcal{H} = \left\{ \{\boldsymbol{\beta}_s, c_s, d_s\}_{s=1}^{K}, \alpha \right\}$, our model permits inference of the edge-specific interaction base rates $\Lambda = \{\lambda_{ij}\}_{i \neq j}$, the group-and time-specific deviation factors $\Delta = \{\boldsymbol{\delta}_g\}_{g=1}^{G}$, the states of the Hidden Markov Model $\mathcal{S} = \{\boldsymbol{s}_g\}_{g=1}^{G}$, and the group assignments $\mathcal{Z} = \{z_{ij}\}_{i \neq j}$. For convenience, we will index the groups in the partition implied by $\mathcal{Z}$ as $g \in \{1, \ldots, G\}$.

Gamma-Poisson conjugacy allows us to marginalize out all the deviation factors, giving the following expression for the joint distribution over the observed counts and the latent variables of interest[1]:

$$P\left(\mathcal{Y}, \Lambda, \mathcal{S}, \mathcal{Z} \mid \mathcal{H}\right) = \int \mathrm{d}^{G \times T} \Delta \; P\left(\mathcal{Y}, \Lambda, \Delta, \mathcal{S}, \mathcal{Z} \mid \mathcal{H}\right)$$

$$= \prod_{g=1}^{G} \prod_{t=1}^{T} \left(\frac{\left(d_{s_g^{(t)}} \right)^{-c_{s_g^{(t)}}}}{\Gamma\left(c_{s_g^{(t)}} \right)} \frac{\Gamma\left(c'_g^{(t)} \right)}{\left(d'_g^{(t)} \right)^{-c'_g^{(t)}}} \prod_{(ij) \in g} \left(\frac{\lambda_{ij}^{y_{ij}^{(t)}}}{\Gamma\left(y_{ij}^{(t)} + 1 \right)} \right) \right)$$

$$\prod_{n=1}^{N} \mathrm{Gamma}\left(\lambda_n \mid \boldsymbol{\gamma} \right) \prod_{s=1}^{S} \mathrm{P\acute{o}lya}\left(\langle n_{s1}, \ldots, n_{sK} \rangle \mid \boldsymbol{\beta}_s \right) \mathrm{CRP}(\mathcal{Z} \mid \alpha) \tag{1}$$

where $c'_g^{(t)} = c_{s_g^{(t)}} + \sum_{(ij) \in g} y_{ij}^{(t)}$ and $d'_g^{(t)} = \left(\frac{1}{d_{s_g^{(t)}}} + \sum_{(ij) \in g} \lambda_{ij} \right)^{-1}$ are sufficient statistics, and $n_{ss'} = \sum_{g=1}^{G} \sum_{t=2}^{T} \mathbf{1}\left(s_g^{(t-1)} = s \right) \mathbf{1}\left(s_g^{(t)} = s' \right)$ is the total number of state transitions from state s to state s' in the HMM.[2] The term CRP refers to the likelihood of a partition drawn from a

---

[1]The Pólya distribution is also known as the Dirichlet-multinomial distribution.

[2]The function $\mathbf{1}(\cdot)$ evaluates to one if its argument evaluates to true and evaluates to zero otherwise.

Chinese Restaurant Process, given by:

$$\text{CRP}(\mathcal{Z} \,|\, \alpha) = \frac{\Gamma(\alpha)}{\Gamma(\sum_{g=1}^{G} n_g + \alpha)} \alpha^G \prod_{g=1}^{G} \Gamma(n_g), \qquad (2)$$

where n_g is the number of entities associated with group g and G is the number of groups in the partition implied by $\mathcal{Z}$.

We use Markov chain Monte Carlo to sample typical values for each of latent variables in our model from their joint posterior given $\mathcal{Y}$. In each iteration, we sample a value for each base rate λ_{ij} as well as typical hyperparameter values for γ_1, γ_2 from their posterior distributions using a slice sampling algorithm [8]. We sample each group assignment z_n from its conditional posterior using Neal et al.'s "Algorithm 8" for Gibbs sampling from a nonconjugate Dirichlet process [9]. We draw a sample for the HMM transition matrix Θ from its posterior distribution conditioned on $\mathcal{S}$, using Dirichlet-multinomial conjugacy. Finally, we jointly sample values for $\boldsymbol{s}_g = \left\{ s_g^{(1)}, \ldots, s_g^{(T)} \right\}$ from their exact conditional distribution by applying the forward-filtering backward-sampling algorithm for inference in HMMs [10].

4 Results

In this section we describe synthetic experiments that establish the validity of our model and its approximate inference algorithm, and we showcase our model's exploratory capabilities by performing an in-depth analysis of a large interaction network from international relations.

4.1 Synthetic experiments

We generated four synthetic datasets, with the same setting of parameters[3], for all combinations of 1000 or 2000 edges and 52 or 365 time slices (corresponding to daily and weekly binning of a single year). We selected numbers of actors and time slices, as well as hyperparameters, that would generate synthetic data similar to the real data we are interested in modeling. In each synthetic experiment, we ran our model over one of these four synthetic datasets and plotted the error in the inferred latent variables under various settings of the hyperparameters. The results show that our model is capable of recovering the latent parameters in these synthetic datasets. In each synthetic experiment, we ran our model over one of these four synthetic datasets and plotted the error in the inferred latent variables under various settings of the hyperparameters to confirm that we recovered the true parameterization. We display an example of these plots in Figure 1.

4.2 Global Database of Events, Language, and Tone (GDELT)

GDELT [6] is a new data set consisting of over 200 million records of real-world events from 1979 to present. Each record contains the basic core elements that identify a unique event – who did what do whom, when, and where – that are extracted automatically from news articles. From this event record data, we induce an interaction network between country actors by counting the total number of articles involving directed actions from one country to another in each discrete time slices. We expect that anomalously high counts of interaction for some pair of countries corresponds to some event of real significance involving that relation, since an anomalously high count means that the media has suddenly redirected its focus. By correlating these increases in media attention, our model finds edges exhibiting anomalously high interaction counts at similar times due to underlying multilateral relations.

In international relations, multilateral relations change (sometimes drastically) over time – however, we assume they are constant within a short enough time window. In our exploration of GDELT,

---

[3]We set the CRP concentration parameter α to 1.0. We set the shape γ_1 and scale γ_2 hyperparameters for the Gamma prior over edge-specific base rates to 2 and 5, respectively. We specified a two-state HMM over activation states and set the Dirichlet prior over the transition matrix to $\rho_{00} = 800, \rho_{01} = 80, \rho_{10} = 200, \rho_{11} = 600$. For the case of $s_g^{(t)} = 0$, we set the emission distribution over $\delta_g^{(t)}$ to a Gamma distribution with shape c_0 and scale d_0 equal to 1000 and 0.0001 (the *spike*). For the case of $s_g^{(t)} = 1$, we set the shape c_1 and scale d_1 to 4 and 1 (the *slab*).
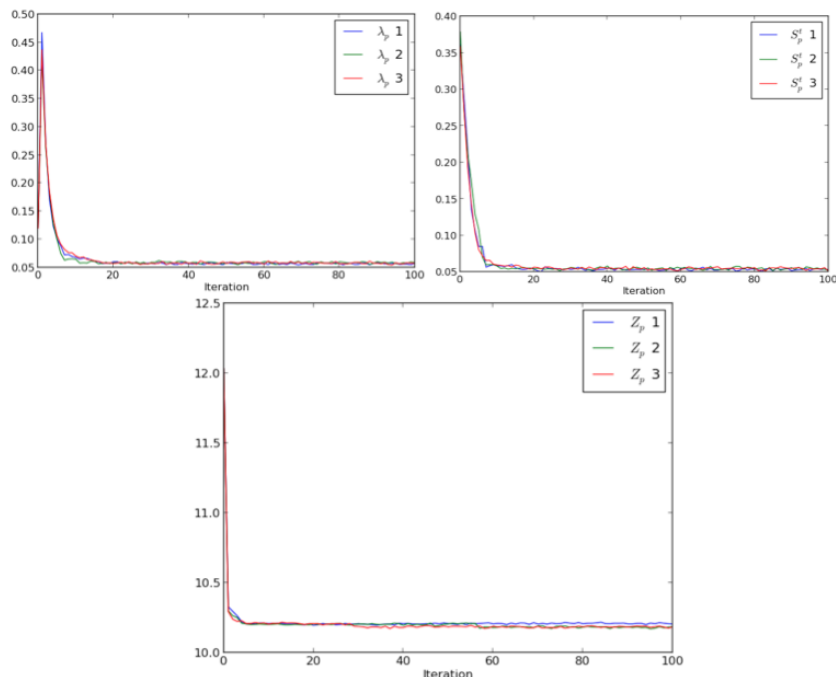
Figure 1: Three plots showing error over sampling iterations. Top left is error on base rates, top right is error on activation states, bottom is variance of information on group assignments. Inference in this experiment was performed conditioned on the true values for all hyper parameters expect for the CRP concentration parameter which was different (0.01 instead of 120).

we ran our model to infer a partition of the most active 1,000 edges independently for each of the years 2000 to 2012. The model was run with two HMM states describing deviation factors, a spike distribution with mean 1 and variance 0.001, and a slab distribution with mean 4 and variance 16. Each run consisted of 3 independent MCMC chains, each with 5,000 sampling iterations. For each year, we identified the single sample across all 3 chains with the highest joint probability and inspected the groups in its inferred partition in decreasing order of size. The groups we found tended either to correspond to multilateral relations that formed temporarily around a major event or ones that are longterm and based around some geopolitical context shared between its edges. Some of the groups corresponded to well-known events about which the authors had prior knowledge. Others, however, required us to learn about some interesting and significant real-world events. We consider this an important validation of the model as an exploratory tool and describe some of these exploratory results in the following section.

4.3   Exploratory Results

For an exploratory tool to be useful, it must be capable of discovering significant and previously unknown patterns in data. For such a tool to be trusted by practitioners though, it must also discover "obvious" patterns, that are consistent with prior knowledge [11]. We formulated mini-hypotheses about certain "obvious" groups we expected to see based on our knowledge of major international events, and we confirmed that our model inferred groups corresponding to those events. For example, after running inference on the 2003 data, we observed that several of the largest groups inferred by the model corresponded to relations involved in the invasion of Iraq (confirming our mini-hypothesis that this major event should be represented).

We show 3 of the largest 5 groups for each of the 2004 and 2005 runs in Figures 2-4 and 5-7. Some of these groups correspond to multilateral relations that formed temporarily around major international events while others correspond to longterm multilateral relations formed around shared geopolitical contexts. One of the groups we display – the Central Asian republics – is an example of a longterm multilateral relation: though we only display the group inferred from the 2004 run,

5

Figure 2: Central Asian republics: This group contains all of the directed edges between the former Soviet-bloc Central Asian republics: Kyrgyzstan, Tajikistan, Uzbekistan, Kazakhstan. Our model inferred a very similar group across all the years. Indeed, this is an example of a multilateral relationship that exists due to a longterm shared geopolitical context.
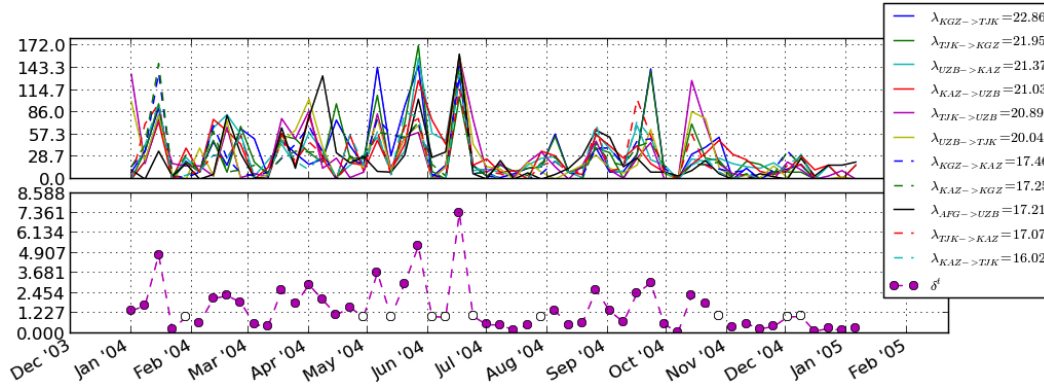

Figure 3: 2004 enlargement of the European Union (as it is called on Wikipedia): Several European countries simultaneously acceded to the EU in May 2004 including Poland, Slovakia, Latvia, and Estonia. This group contains directed edges between those countries and EU, many of which exhibit a burst in activity during and surrounding that time.
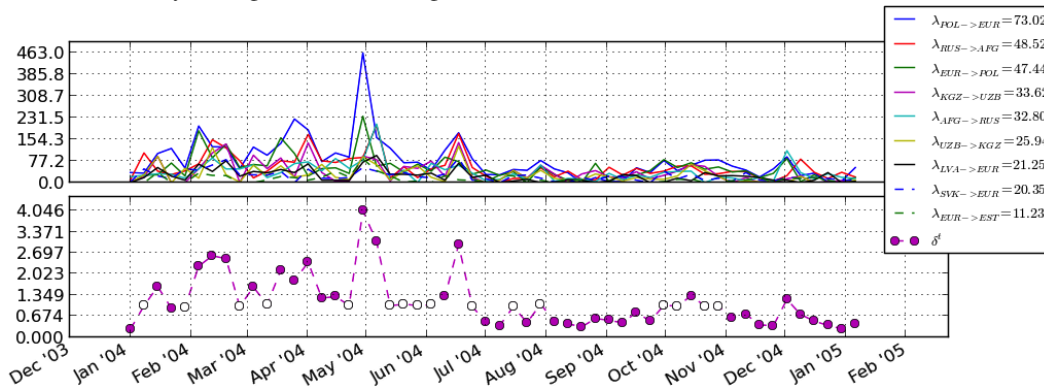
Figure 4: 2004 Indian Ocean tsunami: A single event in December clearly ties these edges together – this is the tsunami that hit Southeast Asia in 2004 and resulted in a massive humanitarian crisis and international response. The edges involve Pacific countries close to the disaster: China, Japan, Australia, Philippines.
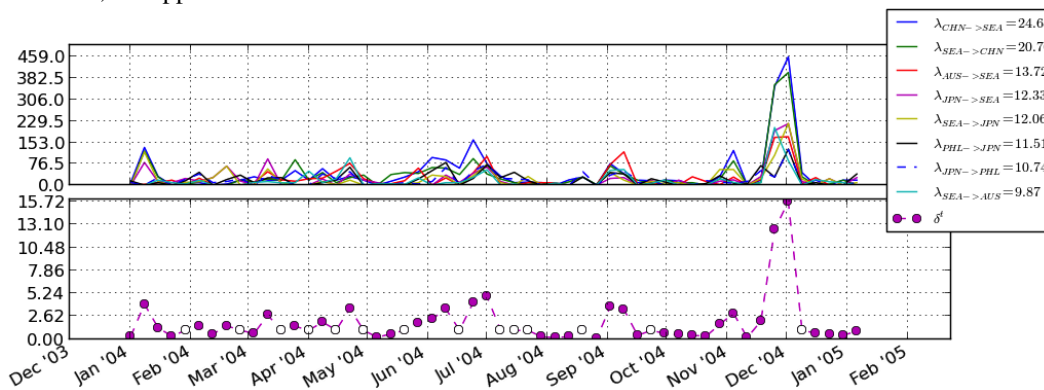

6

Figure 5: Tulip Revolution: In April 2005, a popular uprising in Kyrgyzstan, later called the "Tulip Revolution", overthrew the government of President Askar Akayev. This group contains edges involving Kyrgyzstan, the most active ones being those with the Russia and the United States. While Russia has always maintained an interest in the neighboring Central Asian republics, the United States had a particular stake in maintaining its access to a central military installation near the capital, Bishkek, that was critical for servicing its ongoing operations in Afghanistan.
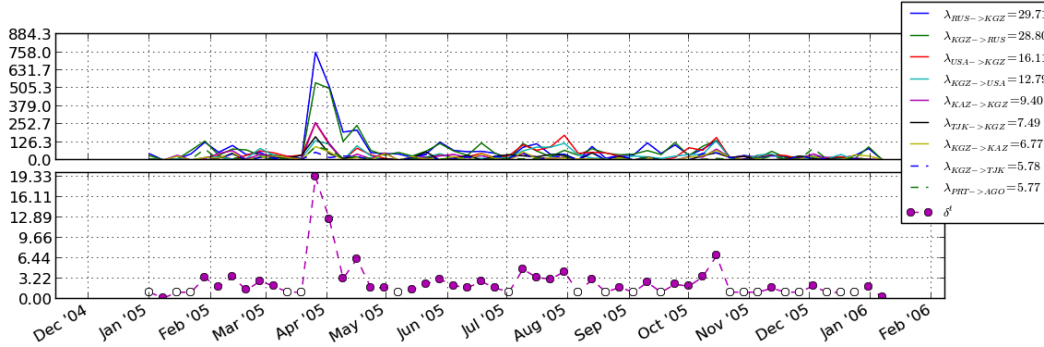

Figure 6: 2005 terrorist attacks in London: In July 2005, suicide bombers attacked London's public transportation system. Though the terrorists were home-grown, the attacks garnered an international response particularly from Coalition countries in the War on Terror. Most of the edges in this group involve Great Britain and exhibit a particular spike of activity during the time of the bombings.
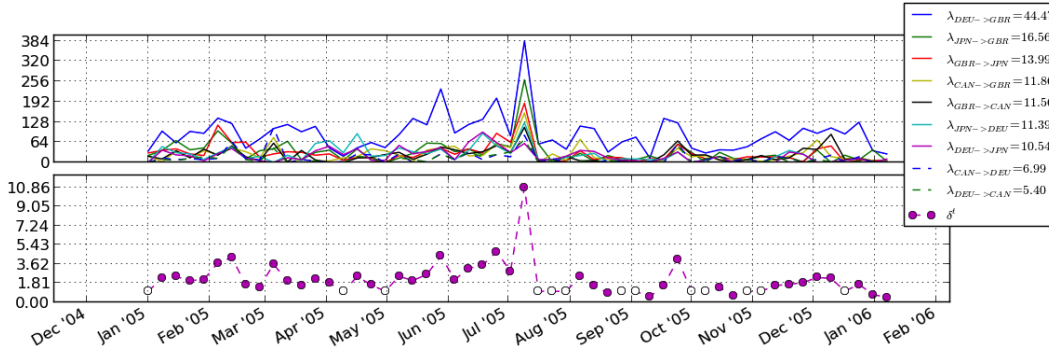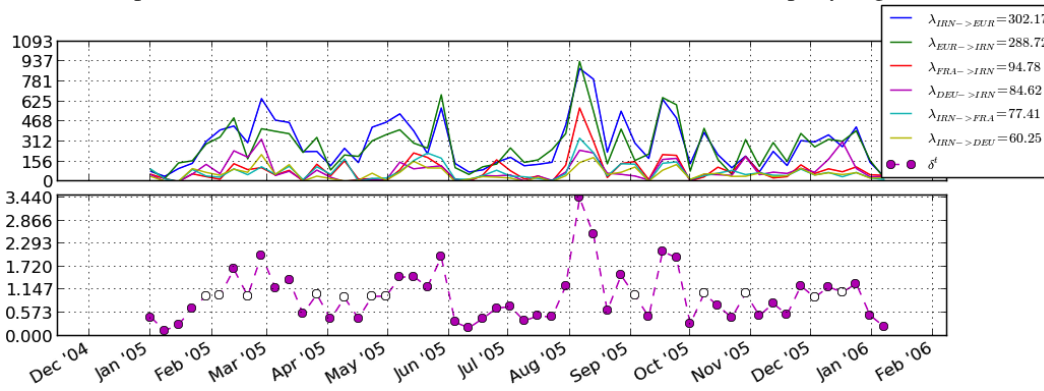

Figure 7: Negotiations with Iran: Throughout 2005, Iran was involved in negotiations with several European countries (mainly France and Germany) over its controversial nuclear program. This is a classic example of a multilateral relation in international relations: a multiparty negotiation.

7

our model inferred a very similar group across all of the years. We didn't know about some of the groups – this is what an exploratory tool is for. In each the figures on the following two pages, we plot the time series for all of the edges in a single group in the top subplot. In the bottom subplot, we plot the posterior mean of that group's deviation factor $\delta_g^{(t)}$, at each time slice. The deviation factors at each time slice are colored white or red, denoting the state of the HMM at for the corresponding group at that time slice, corresponding to the spike distribution ($s_g^{(t)} = 0$) and the slab distribution ($s_g^{(t)} = 1$), respectively.

Acknowledgments

5 Conclusions

We introduced a new non-parametric Bayesian latent variable model that discovers latent structure in complex social processes by capturing correlated deviations from normal behavior. We demonstrated that our model recovers the true parameters of simulated data and presented the results from running our model on a subset of the Global Database of Events, Language, and Tone, an exciting data set that is new to the machine learning community. Our findings from the country interaction data are consistent with established theory in international relations. Based only on the observed interactions between pairs of countries, our model discovers underlying multilateral relations that explain correlated deviations from normal behavior and are consistent with our prior knowledge of international affairs.

References

[1] Nicholas A Heard, David J Weston, Kiriaki Platanioti, and David J Hand. Bayesian anomaly detection methods for social networks. *The Annals of Applied Statistics*, 4(2):645–662, 2010.

[2] Jon Kleinberg. Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, 7(4):373–397, 2003.

[3] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

[4] Charles Kemp, Joshua B Tenenbaum, Thomas L Griffiths, Takeshi Yamada, and Naonori Ueda. Learning systems of concepts with an infinite relational model. In *AAAI*, volume 3, page 5, 2006.

[5] Edoardo M Airoldi, David M Blei, Stephen E Fienberg, and Eric P Xing. Mixed membership stochastic blockmodels. *The Journal of Machine Learning Research*, 9:1981–2014, 2008.

[6] Kalev Leetaru and Philip A. Schrodt. Gdelt: Global data on events, location and tone, 1979-2012. 2013.

[7] Jim Pitman. Combinatorial stochastic processes. Technical report, Technical Report 621, Dept. Statistics, UC Berkeley, 2002. Lecture notes for St. Flour course, 2002.

[8] Radford M Neal. Slice sampling. *Annals of statistics*, pages 705–741, 2003.

[9] Radford M Neal. Markov chain sampling methods for dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2):249–265, 2000.

[10] Kevin P Murphy. *Machine Learning: A Probabilistic Perspective*, chapter 17, pages 617–618. MIT Press, 2012.

[11] Yanchuan Sim, Brice Acree, Justin H Gross, and Noah A Smith. Measuring ideological proportions in political speeches. In *Proceedings of EMNLP*, 2013.