
Inferring Polyadic Events With Poisson Tensor Factorization

Aaron Schein

School of Computer Science
University of Massachusetts Amherst
Amherst, MA 01003
aschein@cs.umass.edu

John Paisley

Department of Electrical Engineering
Columbia University
New York, NY 10027
jpaisley@columbia.edu

David M. Blei

Department of Computer Science
Columbia University
New York, NY 10027
david.blei@columbia.edu

Hanna M. Wallach

Microsoft Research
641 Avenue of the Americas
New York, NY 10011
hanna@dirichlet.net

Abstract

We present a Bayesian factorization model for discovering dynamic communities of actors in a social network and for describing action types associated with those communities. The structure of a social network is often hidden in the interactions among its actors. Typically, we observe interactions between pairs of actors (i.e., an edge) and our goal is to uncover their more complex relationships. For example, here we consider a problem in international relations where our goal is to analyze pairwise country interactions to infer communities of countries that are organized into international “events.” We develop a Poisson factorization model for decomposing tensors of time-stamped and typed interactions; and we present a mean-field variational method for efficiently fitting this model to large amounts of data. We study the inferred components from a real-world data set of country–country interactions, and show that they conform to and inform our knowledge of international affairs.

1 Introduction

In this paper, we present a Bayesian tensor factorization model that discovers interpretable latent structure in social networks from dyadic interaction data. Social networks can be viewed either statically or dynamically. When viewed statically, a social network is characterized by a set of actors connected by edges that represent *persistent relationships* such as “friendship” on Facebook and co-authorship in academia. In contrast, when viewed dynamically, a social network is characterized by a set of interactions between actors. Network properties, such as the relationships between those actors, are then determined implicitly via *patterns of interaction*. For example, an email network can be viewed as a set of interactions (i.e., emails) in which patterns in the timing and content of emails imply information about the relationships between actors.

Social interaction data is often limited to (or simply modeled as) *dyadic* interactions, i.e., interactions between only two actors, a sender and a receiver. This assumption is motivated by the sparsity and computational challenges arising from the combinatorial explosion induced by considering interactions between groups of actors. In practice, however, social scientists are often interested in modeling social processes in which dyadic behavior is highly coupled. The modeling challenge in these scenarios is therefore to answer complex questions about the underlying social processes given

only pairwise interaction data. Blockmodeling, in which static blocks (or communities) of actors are inferred based on correlated dyadic behavior, is a common instance of this approach.

The motivation behind this paper is international relations, in which 1) there are substantially different types of interaction between country actors (e.g., praise vs. invade), 2) the relationships between countries change through time (sometimes drastically and abruptly), and 3) bilateral¹ behavior is contingent on multilateral relationships (e.g., treaties). For decades, political scientists have collected records of bilateral interactions between country actors of the form, “Country A did something to Country B.” Given such data, the modeling challenge is to infer and characterize the complex multilateral relationships that give rise to the observed bilateral interaction patterns. Equivalently, in terms more common to the social networks research community, the goal is to infer dynamic and typed communities—*polyadic events*—from patterns in the timing and types of dyadic interactions.

Poisson-based models have been used in international relations since the 1940s [1]. Poisson regression models (e.g., that of King [2]) are highly effective at exploratory and predictive tasks involving interaction counts in international relations. In the social networks research community, Poisson matrix factorization has also been used effectively for community detection in networks (e.g., [3, 4]).

In this paper, we introduce Poisson tensor factorization, a generalization of Poisson matrix factorization, and a variational algorithm that permits inference of parameter values from large data sets. Our model infers K latent components that specify different mixtures over the senders, receivers, action types, and timing of dyadic interactions. We showcase our model as an exploratory analysis tool for international relations scholars by demonstrating that the inferred components correspond to interpretable polyadic events that conform to and inform knowledge of international affairs.

2 Data Representation

Consider a set of observed dyadic interactions $\mathcal{X} = \{x_1, \dots, x_M\}$ where each member is tuple-valued $x_m = (s_m, r_m, a_m, \tau_m)$ encoding the source, receiver, action-type, and time-stamp of the interaction. We assume that senders and receivers come from the same index set of N actors $s_m, r_m \in \{1 \dots N\}$. Action-types similarly come from an index set of A actions $a_m \in \{1 \dots, A\}$. Modeling time discretely with time-steps indexed by $t \in \{1, \dots, T\}$, we obtain counts $y_{ijat} = \#\{x_m : s_m = i, r_m = j, a_m = a, \tau_m \subset t\}$ which are the number of interactions within a discrete time interval t involving a particular sender i , receiver j , and action-type a . These counts form the entries of a four-mode tensor $\mathbf{Y} \in \mathbb{N}^{N \times N \times A \times T}$ that is non-negative and integer-valued.

3 Poisson Tensor Factorization

Poisson tensor factorization (PTF) generalizes Poisson matrix factorization (PMF) which is a specific case of non-negative matrix factorization (NNMF). In NNMF, an observed non-negative matrix $\mathbf{Y} \in \mathbb{R}^{D \times V}$ is assumed to be generated from the product of latent matrices $\Theta \in \mathbb{R}_+^{D \times K}$ and $\Phi \in \mathbb{R}_+^{K \times V}$ known as the excitation and template matrices or weight and factor matrices, respectively. PMF assumes a Poisson likelihood, so that entries of the observed matrix are counts $y_{nd} \in \mathbb{N}$ and are assumed to be generated $y_{nd} \sim \text{Poisson}(\sum_{k=1}^K \theta_{dk} \phi_{kv})$. In PTF, we observe an M -mode tensor $\mathbf{Y} \in \mathbb{N}^{\mathcal{D}_1 \times \dots \times \mathcal{D}_M}$, where \mathcal{D}_m is the dimensionality of mode m . We assume that the observed tensor was generated from the multiway product of M latent matrices $\Theta^1, \dots, \Theta^M$ where a given latent matrix $\Theta^m \in \mathbb{R}_+^{\mathcal{D}_m \times K}$ is the dimensionality of mode m by the number of latent components. For the purposes of exposition, we will assume from here on that \mathbf{Y} is a four-mode tensor, as is the case in our particular application, but note the generality of the model and inference algorithm for arbitrary sized tensors. We then assume a Poisson likelihood such that a given count in the observed tensor $y_{ijat} \in \mathbf{Y}$ is drawn $y_{ijat} \sim \text{Poisson}(\theta_{ik}^1 \theta_{jk}^2 \theta_{ak}^3 \theta_{tk}^4)$.

To complete the model, we impose prior distributions over the entries of the latent matrices. In PMF, it is typical to impose a Gamma prior over the entries of the excitation matrix so that

¹The term *bilateral* is standard in the international relations literature while the term *dyadic* is standard in the social networks literature. For the purpose of this paper, we consider the terms *bilateral*, *dyadic*, and *pairwise* to be interchangeable and similarly for *multilateral* and *polyadic*.

$\theta_{dk} \sim \text{Gamma}(a, b)^2$. Cemgil et al. [5] introduced to the machine learning community PMF with Gamma priors over both the entries of the excitation and template matrices and derived both MCMC and variational inference algorithms for the model. We follow Cemgil and others ([6], [7]) in imposing Gamma priors over the entries of all M latent matrices. For our case, in which \mathbf{Y} is a four-mode tensor containing the counts of typed dyadic interactions within discrete time intervals, the corresponding PTF model can be described as:

$$\begin{aligned} \theta_{ik}^s &\sim \text{Gamma}(a, b) \\ \theta_{jk}^r &\sim \text{Gamma}(a, b) \\ \psi_{ak} &\sim \text{Gamma}(c, d) \\ \delta_{tk} &\sim \text{Gamma}(e, f) \\ y_{ijat} &\sim \text{Poisson} \left(\sum_{k=1}^K \theta_{ik}^s \theta_{jk}^r \psi_{ak} \delta_{tk} \right) \end{aligned} \quad (1)$$

We use $\Theta^s = \{\theta_{ik}^s\}$, $\Theta^r = \{\theta_{jk}^r\}$, $\Psi = \{\psi_{ak}\}$, $\Delta = \{\delta_{tk}\}$ to refer to the four latent matrices (one per mode of \mathbf{Y}) where the four modes correspond to the space of possible senders, receivers, action-types, and time-steps of a dyadic interaction, respectively. We assume that senders and receivers come from the same set of actors. Thus for a set of N actors, $\Theta^s, \Theta^r \in \mathbb{R}_+^{N \times K}$. Similarly, for a set of A possible action-types, $\Psi \in \mathbb{R}_+^{A \times K}$ and for T time-steps, $\Delta \in \mathbb{R}_+^{T \times K}$. In matrix factorization, the K -length rows of the excitation matrix are interpreted as per-item weights on the K latent components while the D -length columns of the template matrix are interpreted as per-component templates (or ‘topics’) of the covariate space (e.g., the space of unique word types). In the case of PTF, it will be interesting to consider both rows and columns of these latent matrices with no latent matrix having a particular interpretation beyond embedding a mode into latent space. For example, a row $\vec{\theta}_i$ of Θ^s versus a column $\vec{\theta}_k$ may be interesting when asking “which component is sender i most active in?” versus “which sender is most active in component k ?”.

By the Superposition Theorem [8], the Poisson likelihood in equation 1 can be equivalently represented as the sum of K independent Poisson random variables $y_{ijat} = \sum_{k=1}^K z_{ijatk}$ where $z_{ijatk} \sim \text{Poisson} \left(\theta_{ik}^s \theta_{jk}^r \psi_{ak} \delta_{tk} \right)$. This property is practically important for deriving the variational inference algorithm but also highlights a unique property of Poisson factorization models in that discrete observations arise from independent latent *sources* that are also discrete. It can be shown, as in Cemgil et al. [5], that the likelihood, as it appears in equation 1, results from marginalizing out the latent sources, $\vec{z}_{ijat} \equiv z_{ijat1}, \dots, z_{ijatK}$:

$$P(y_{ijat} | \theta_{ik}^s, \theta_{jk}^r, \psi_{ak}, \delta_{tk}) = \sum_{\vec{z}_{ijat}} P(y_{ijat}, \vec{z}_{ijat} | \theta_{ik}^s, \theta_{jk}^r, \psi_{ak}, \delta_{tk}) \quad (2)$$

When performing inference in Poisson factorization models, we can choose whether to explicitly represent the sources as latent variables or marginalize them out. We will see in section 4 that explicitly representing the sources is useful for deriving mean-field variational inference. The joint probability of counts and sources can be factorized in the following way (where we use the notation $P(x|-)$ to mean the probability of x given values of all other variables):

$$P(y_{ijat}, \vec{z}_{ijat} | -) = P(\vec{z}_{ijat} | y_{ijat}, -) P(y_{ijat} | -) \quad (3)$$

Here the term $P(y_{ijat} | -)$ is the marginal likelihood shown in equation 1 while the term $P(\vec{z}_{ijat} | y_{ijat}, -) \equiv P(z_{ijat1}, \dots, z_{ijatK} | y_{ijat}, -)$ is the joint probability of independent Poisson random variables conditioned on their sum, equivalent to a single Multinomial random variable.

4 Variational Inference

At inference time, we “reverse” the generative process to form the posterior distribution of our latent variables given our observations, $P(\Theta^s, \Theta^r, \Psi, \Delta | \mathbf{Y})$. Since no closed-form expression exists for the posterior distribution of this model, we approximate it using mean-field variational inference. Variational methods optimize the parameters of an instrumental distribution $q(\Theta^s, \Theta^r, \Psi, \Delta)$

²We use the *rate* parameterization of the Gamma distribution throughout.

to minimize the KL-divergence between it and the true posterior. Mean-field variational methods, select an instrumental distribution that factorizes over all latent variables. As shown in Cemgil et al. [5], if we explicitly represent the sources as latent variables so that the instrumental distribution factorizes over them, closed-form expressions for the variational updates can be easily derived from known properties about conditionally conjugate models [9]:

$$q(\mathbf{Z}, \Theta^s, \Theta^r, \Psi, \Delta) = \left[\prod_{i=1}^N \prod_{j=1}^N \prod_{a=1}^A \prod_{t=1}^T q(\vec{z}_{ijat}) \right] \prod_{k=1}^K \left[\prod_{i=1}^N q(\theta_{ik}^s) \right] \left[\prod_{j=1}^N q(\theta_{jk}^r) \right] \left[\prod_{a=1}^A q(\psi_{ak}) \right] \left[\prod_{t=1}^T q(\delta_{tk}) \right] \quad (4)$$

We choose an instrumental distribution that takes the same form over each latent variable as in the true generative process:

$$\begin{aligned} q(\vec{z}_{ijat}) &= \text{Multinomial}(y_{ijat}, \vec{\phi}_{ijat}) \\ q(\theta_{ik}^s) &= \text{Gamma}(\alpha_{ik}^s, \beta_{ik}^s) \\ q(\theta_{jk}^r) &= \text{Gamma}(\alpha_{jk}^r, \beta_{jk}^r) \\ q(\psi_{ak}) &= \text{Gamma}(\gamma_{ak}, \chi_{ak}) \\ q(\delta_{tk}) &= \text{Gamma}(\rho_{tk}, \nu_{tk}) \end{aligned} \quad (5)$$

The algorithm then works by finding values of the variational parameters which maximize the following tight lower bound on the true log joint probability:

$$\mathcal{B} = \mathbb{E}_q[\ln P(\mathbf{Y}, \Theta^s, \Theta^r, \Psi, \Delta)] + H(q) \quad (6)$$

where $H(q)$ is the entropy of the instrumental distribution. It can be shown that maximizing this lower bound is equivalent to minimizing the KL-divergence from the instrumental distribution to the true posterior.

We optimize the bound using coordinate ascent. Closed-form update equations for coordinate ascent with Poisson matrix factorization have been presented previously [5, 7, 6]. The updates for Poisson tensor factorization follow directly from these:

$$\begin{aligned} \vec{\phi}_{ijatk} &:= \frac{\exp(\mathbb{E}_q[\ln \theta_{ik}^s \theta_{jk}^r \psi_{ak} \delta_{tk}])}{\sum_{k=1}^K \exp(\mathbb{E}_q[\ln \theta_{ik}^s \theta_{jk}^r \psi_{ak} \delta_{tk}])} \\ \alpha_{ik}^s &:= a + \sum_{j=1}^N \sum_{a=1}^A \sum_{t=1}^T \mathbb{E}_q[z_{ijatk}] & \beta_{ik}^s &:= b + \sum_{j=1}^N \sum_{a=1}^A \sum_{t=1}^T \mathbb{E}_q[\theta_{jk}^r \psi_{ak} \delta_{tk}] \\ \alpha_{jk}^r &:= a + \sum_{i=1}^N \sum_{a=1}^A \sum_{t=1}^T \mathbb{E}_q[z_{ijatk}] & \beta_{jk}^r &:= b + \sum_{i=1}^N \sum_{a=1}^A \sum_{t=1}^T \mathbb{E}_q[\theta_{ik}^s \psi_{ak} \delta_{tk}] \\ \gamma_{ak} &:= c + \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \mathbb{E}_q[z_{ijatk}] & \chi_{ak} &:= d + \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \mathbb{E}_q[\theta_{ik}^s \theta_{jk}^r \delta_{tk}] \\ \rho_{tk} &:= e + \sum_{i=1}^N \sum_{j=1}^N \sum_{a=1}^A \mathbb{E}_q[z_{ijatk}] & \nu_{tk} &:= f + \sum_{i=1}^N \sum_{j=1}^N \sum_{a=1}^A \mathbb{E}_q[\theta_{ik}^s \theta_{jk}^r \psi_{ak}] \end{aligned} \quad (7)$$

Since all the latent variables are independent under the instrumental distribution, the expectations of products factorize into products of expectations, e.g.:

$$\mathbb{E}_q[\theta_{ik}^s \theta_{jk}^r \psi_{ak}] = \mathbb{E}_q[\theta_{ik}^s] \mathbb{E}_q[\theta_{jk}^r] \mathbb{E}_q[\psi_{ak}] \quad (8)$$

Similarly:

$$\mathbb{E}_q[\ln \theta_{ik}^s \theta_{jk}^r \psi_{ak} \delta_{tk}] = \mathbb{E}_q[\ln \theta_{ik}^s] + \mathbb{E}_q[\ln \theta_{jk}^r] + \mathbb{E}_q[\ln \psi_{ak}] + \mathbb{E}_q[\ln \delta_{tk}] \quad (9)$$

Therefore, the expectations $\mathbb{E}_q[\theta_{ik}^s]$ and $\mathbb{E}_q[\ln \theta_{ik}^s]$ for each Gamma variable as well as $\mathbb{E}_q[z_{ijatk}]$ are the only sufficient statistics needed to implement the algorithm. We provide the form of the Gamma expectations for θ_{ik}^s only and note that the other Gamma variables are directly analogous:

$$\begin{aligned}\mathbb{E}_q[\theta_{ik}^s] &= \frac{\alpha_{ik}^s}{\beta_{ik}^s} \\ \mathbb{E}_q[\ln \theta_{ik}^s] &= \psi(\alpha_{ik}^s) - \ln \beta_{ik}^s\end{aligned}\tag{10}$$

where $\psi(\cdot)$ is the digamma function. Finally, the Multinomial expectation needed is:

$$\mathbb{E}_q[z_{ijatk}] = y_{ijat} \phi_{ijatk}\tag{11}$$

This identity is exploited in practice so that we never need to instantiate what would be a prohibitively large tensor containing all of the ϕ_{ijatk} variables.

5 Results & Discussion

5.1 Global Database of Events, Location and Tone

For several decades, political scientists (e.g., Schrodtt [10]) have collected and analyzed public records of dyadic interactions³ between countries of the form “Country A did something to Country B” in order to study patterns of international relations. In recent years, information extraction systems (e.g., TABARI [11]) have been developed to automatically extract large collections of records from digital news archives. The size and scope of these new datasets present both an unprecedented opportunity to understand the complex structure of international affairs as well as a challenge to develop scalable models.

Our data of country interactions comes from the Global Database of Events, Location, and Tone (GDEL T) [12]. GDEL T is unparalleled in its scope and size, containing over a quarter billion interactions dating back to 1979. Interactions are coded for a number of attributes including detailed information about senders, receivers, and action-types. GDEL T employs the CAMEO hierarchical coding scheme [13] for both actors and action-types. For actors, the top level represents their country of origin while lower levels represent successively finer distinctions (e.g., sector, government branch). For action-type, the top level consists of 20 basic classes that are loosely ranked based on sentiment ranging from 01 *Make Public Statement* to 20 *Use Unconventional Mass Violence*.

In all of our experiments, we considered the top level for both actors and action-types as index sets, leading to $N = 219$ unique country actors and $A = 20$ unique high-level action-types. Interactions in GDEL T are time-stamped with per-day granularity. We experimented with 1-day, 3-day, and 7-day intervals as time-steps for a single year window ($T = 365$, $T = 121$, and $T = 52$, respectively). Our observed tensor, with e.g., 7-day time-steps, was therefore $\mathbf{Y} \in \mathbb{N}^{219 \times 219 \times 20 \times 52}$.

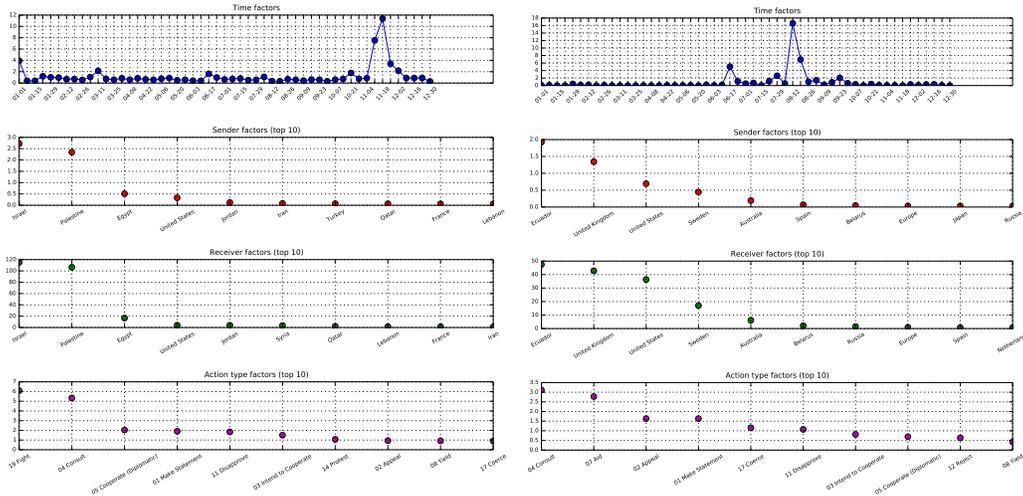
5.2 Exploratory results

We provide example output of the model in Figure 5.2 where we display two of the latent components inferred for a dataset of interactions occurring in 2012. A component k is an index into the column space of the four latent matrices associating a vector of factors over senders $\vec{\theta}_k^s$, receivers $\vec{\theta}_k^r$, action-types $\vec{\psi}_k$, and time-steps $\vec{\delta}_k$. For a given component, we display all of its time-step factors chronologically, while for sender, receiver, and action-type factors we display the top ten (by magnitude) in decreasing order. In examining a component, we asked whether the association of its top senders, receivers, action-types, and time-steps corresponded to a major polyadic event that would have received significant media coverage. In many cases, the component’s correspondence to a major polyadic event was clear given our background knowledge. For example, the first subplot of Figure 5.2 shows a component in which the top action-type is *Fight*, the top actors are Israel, Palestine, and other countries directly involved in that ongoing conflict (e.g., Egypt, United States) and the time factors show a sharp spike in November 2012, during which time Operation Pillar

³These interactions are referred to as *events* in the political science community with the datasets referred to as *political events data*. We will refer to them as interactions to distinguish the discrete dyadic observations from the latent persistent events we are trying to extract.

of Defense [14] occurred. In the cases where our background knowledge was insufficient to map a component to a polyadic event, a simple Google query containing the top sender, receiver and time step (e.g., “israel palestine november 2012”) almost always returned a sensible answer. The second subplot of Figure 5.2 shows an example of such a component where we Googled “ecuador UK august 2012” to see if any major event spiking around August 2012 could explain the component’s top actors being Ecuador, United Kingdom, United States, Sweden, and Australia. The top hit from this query was a Wikipedia article [15] about Julian Assange, a journalist from **Australia**, wanted by the **United States** for his role in the Wikileaks scandal and by **Sweden** for alleged sexual misconduct. Assange sought refuge in **June 2012** at **Ecuador**’s embassy in the **United Kingdom** and was granted political asylum in **August 2012**. We see that the component’s time factors (with spikes in June and August), its actor factors, and action-type factors provide an excellent summary for this complex ordeal. We consider this a perfect example of the type of structure we sought originally to discover: a time-dependent complex relationship between multiple different countries (a polyadic event). That some components conformed to our prior knowledge while others prompted us to “converse” with the data, gives us confidence in the model as an exploratory tool.

Figure 1: Two of the components inferred for 2012 data with 7-day time-intervals and $K = 50$. *Left:* Israeli-Palestinian hostilities renew in November 2012 during Operation Pillar of Defense. *Right:* Julian Assange is sheltered at Ecuador’s embassy in United Kingdom, June-August.



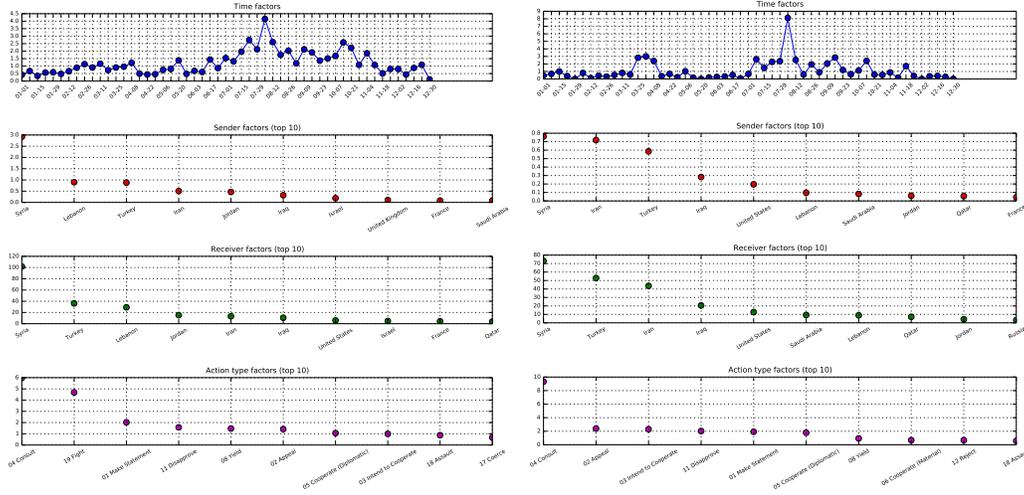
5.3 Sparsity & choice of K

As with all parametric models, the choice of K matters. We expect that the proportion of observations explained by a particular latent component is inversely proportional to the number of latent components: the more components, the more each component specializes on a particular subset of observations. If we are using the model for prediction, we should choose the K which maximizes the model’s ability to generalize to new observations (e.g., minimizes test set error). If we are using the model as an exploratory tool, we seek an intuitive understanding for how choices of K affect the results, treating K as a knob to be controlled by a human analyst.

We observed that as we increased K , the components we inferred went from corresponding to more general polyadic events that unfolded over longer periods of time, to more specific ones exhibiting sparse time factors around specific peaks. We give an example of this in Figure 5.3 in which we show a component from a run on 2012 data with $K = 50$ and a component from a run with $K = 100$. The two components clearly explain overlapping subsets of the data: interactions involving Syria and regional neighbors that peak around August 2012. The time factors in the first component ($K = 50$) show a gradual buildup to the August peak followed by an equally gradual decline in activity. This contrasts with the second component ($K = 100$) whose time factors are sparse around the August peak. After some Googling, we find that the first component corresponds

generally to the Syrian civil war which began in late 2011, entered a tenuous cease-fire in April 2012 and renewed in July 2012 with major battles over Damascus and Aleppo in August 2012 [16]. We see that the top two actors in the first component, other than Syria, are its Eastern neighbors, Turkey and Lebanon, who we expect to be the two countries most generally affected by the war (fighting was concentrated in Aleppo which borders Turkey and Damascus which borders Lebanon). We see in the second component, whose time factors are highly peaked around early August, that the top two actors, other than Syria, are Iran and Turkey, with Lebanon lower on the ranking. We find, again after Googling, that a major defection of Syrian military leaders into Turkey on August 7th coincided with the capture of Iranian agents by the Syrian rebels and the revelation that Iran was providing material support to the Syrian regime [17]. We conclude that although the general Syria civil war component from the $K = 50$ run accounts in part for the August interactions between Syria, Turkey and Iran, the component in the $K = 100$ run specializes in those particular interactions.

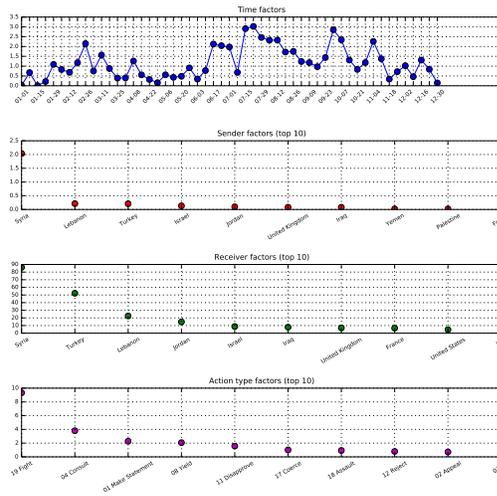
Figure 2: Corresponding components from separate runs on 2012 data with different settings for K . *Left:* $K = 50$. Time factors show a gradual buildup up to the peak of the Syrian civil war in 2012. *Right:* $K = 100$. Time factors are sparse around two specific peaks.



Upon further inspection, we also find that a different component from the run with $K = 100$, shown in Figure 5.3, accounts for the gradual buildup of interactions between Syria and its Eastern neighbors. We see that the general Syrian civil war component inferred in the $K = 50$ run appears to be a mixture of the two more specialized components inferred in the $K = 100$ run. We see that its top actors and top action-types, are a mixture of the top actors and action-types in the two $K = 100$ components, and its time factors are a mixture of the August spike in the first $K = 100$ component and the gradual buildup in the second. Intuitively, it makes sense that with twice as many latent components, the model infers components that are twice as specialized.

It is a well-known property of the Gamma distribution that its asymmetry and spike at zero make it a very suitable prior for sparse variables. Following Gopalan et al. [6], we imposed Gamma priors that encouraged the model towards sparse factors. We found that the sparse representation makes the components highly interpretable and that the direct relationship between the number of latent components and the sparsity of their factors

Figure 3: $K = 100$. A different component corresponding to the Syrian civil war.



allow an analyst to intuitively configure the hyperparameters for exploratory tasks, with larger K leading to more specific and temporally-localized polyadic events.

We have made public⁴ sample results from experiments on 2003-2012 data that include all of the components inferred from the 2012 runs discussed here.

6 Related and Future Work

This work is closely related to Schein et al. [18] who present a Poisson-based generative model for clustering time-series of dyadic interactions in GDELT for the task of discovering multilateral relations. Though the task is similar, the model presented there is different in several important respects including that it is a clustering (as opposed to admixture) model and it does not model the action-type of interactions. Dubois & Smyth [19] modeled senders, receivers and action-types of dyadic interactions using a generative model. Their model conditions on the number of total interactions and imposes Dirichlet priors over the factors in each component. Hoff [20] has also recently presented general work on Tucker tensor decompositions that includes an example analysis on GDELT data.

Our preliminary findings suggest several promising next steps. First, a hierarchical model over sender and receiver factors would allow the model to share statistical strength between the subsets of interactions in which an actor is a sender or receiver. This makes intuitive sense, as we expect the activity of an actor as a sender and receiver to be highly correlated. Hierarchical priors of this kind have been employed effectively in Poisson matrix factorization for recommendation (e.g., Gopalan et al. [6]). Second, there are several non-parametric extensions of the current model that are worth investigating particularly in light of recent work on the Negative-Binomial Process (e.g., Zhou & Carin [21], Broderick et al. [22]). It has been shown (as in Zhou et al. [23]) that the Negative-Binomial Process is equivalent to Poisson matrix factorization with a Gamma process prior imposed over the excitations. Lastly, the current model assumes column exchangeability for time-steps which, although practical, violates common sense. Typically, the way to induce column dependency is with a Gaussian prior. Paisley et al. [24] use a logistic normal prior to induce dependencies between weights drawn from a Gamma process. Adams et al. [25] also derived inference for a Poisson process model with Gaussian process intensities. Another way to induce column dependence is to impose a Hidden Markov Model over the intensities (as in Schein et al. [18]), though doing so makes inference more challenging.

7 Summary

In this paper, we presented Poisson tensor factorization which discovers interpretable polyadic events from dyadic interaction data. Recent work in political science has yielded large datasets of country-country interactions that motivate the development of scalable models for learning patterns of country interaction. Our mean-field variational algorithm allows for efficient inference on such scales. We showcased the model as an exploratory analysis tool for international relations researchers by fitting it to GDELT data and showing that its latent components corresponded to interpretable polyadic events. We further demonstrated how the model's sparse representation makes it easy to use and interpret. Finally, we outlined promising future directions for this line of research including modeling temporal dependence and Bayesian non-parametric extensions.

References

- [1] Lewis F Richardson. The distribution of wars in time. *Journal of the Royal Statistical Society*, pages 242–250, 1944.
- [2] Gary King. Statistical models for political science event counts: Bias in conventional procedures and evidence for the exponential poisson regression model. *American Journal of Political Science*, pages 838–863, 1988.
- [3] Prem Gopalan, Sean Gerrish, Michael Freedman, David M Blei, and David M Mimno. Scalable inference of overlapping communities. In *Advances in Neural Information Processing Systems*, pages 2249–2257, 2012.

⁴https://github.com/aschein/PTF_GDELT

- [4] Prem K Gopalan and David M Blei. Efficient discovery of overlapping communities in massive networks. *Proceedings of the National Academy of Sciences*, 110(36):14534–14539, 2013.
- [5] Ali Taylan Cemgil. Bayesian inference for nonnegative matrix factorisation models. *Computational Intelligence and Neuroscience*, 2009, 2009.
- [6] Prem Gopalan, Jake M Hofman, and David M Blei. Scalable recommendation with poisson factorization. *arXiv preprint arXiv:1311.1704*, 2013.
- [7] Dawen Liang, John Paisley, and Daniel PW Ellis. Codebook-based scalable music tagging with poisson matrix factorization.
- [8] John Frank Charles Kingman. *Poisson processes*, volume 3. Oxford university press, 1992.
- [9] Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley. Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347, 2013.
- [10] Philip A Schrodtt. Event data in foreign policy analysis.
- [11] Philip A Schrodtt. Automated coding of international event data using sparse parsing techniques. In *annual meeting of the International Studies Association, Chicago*, 2001.
- [12] Kalev Leetaru and Philip A Schrodtt. Gdelt: Global data on events, location, and tone, 1979–2012. In *of: Paper presented at the ISA Annual Convention*, volume 2, page 4, 2013.
- [13] Deborah J Gerner, Philip A Schrodtt, Omür Yilmaz, and Rajaa Abu-Jabr. Conflict and mediation event observations (cameo): A new event data framework for the analysis of foreign policy interactions. *International Studies Association, New Orleans*, 2002.
- [14] Wikipedia. Operation pillar of defense — wikipedia, the free encyclopedia, 2014. [Online; accessed 30-October-2014].
- [15] Wikipedia. Embassy of ecuador, london — wikipedia, the free encyclopedia, 2014. [Online; accessed 30-October-2014].
- [16] Wikipedia. Syrian civil war — wikipedia, the free encyclopedia, 2014. [Online; accessed 30-October-2014].
- [17] Wikipedia. Timeline of the syrian civil war (mayaugust 2012) — wikipedia, the free encyclopedia, 2014. [Online; accessed 30-October-2014].
- [18] Aaron Schein, Juston Moore, and Hanna Wallach. Inferring multilateral relations from dynamic pairwise interactions. *NIPS Workshop on Frontiers of Network Analysis*, 2013.
- [19] Christopher DuBois and Padhraic Smyth. Modeling relational events via latent classes. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 803–812. ACM, 2010.
- [20] Peter David Hoff. Equivariant and scale-free tucker decomposition models. *arXiv preprint arXiv:1312.6397*, 2013.
- [21] Mingyuan Zhou and Lawrence Carin. Negative binomial process count and mixture modeling. 2013.
- [22] Tamara Broderick, Lester Mackey, John Paisley, and Michael I Jordan. Combinatorial clustering and the beta negative binomial process. 2011.
- [23] Mingyuan Zhou, Lauren Hannah, David Dunson, and Lawrence Carin. Beta-negative binomial process and poisson factor analysis. *arXiv preprint arXiv:1112.3605*, 2011.
- [24] John Paisley, Chong Wang, and David Blei. The discrete infinite logistic normal distribution. *arXiv preprint arXiv:1103.4789*, 2011.
- [25] Ryan Prescott Adams, Iain Murray, and David JC MacKay. Tractable nonparametric bayesian inference in poisson processes with gaussian process intensities. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 9–16. ACM, 2009.