A Mutually Exciting Latent Space Hawkes Process Model for Continuous-time Networks

Abstract

Networks and temporal point processes serve as fundamental building blocks for modeling complex dynamic relational data in various domains. We propose the *latent space Hawkes (LSH)* model, a novel generative model for continuous-time networks of relational events, using a latent space representation for nodes. We model relational events between nodes using mutually exciting Hawkes processes with baseline intensities dependent upon the distances between the nodes in the latent space and sender and receiver specific effects. We propose an alternating minimization algorithm to jointly estimate the latent positions of the nodes and other model parameters. We demonstrate that our proposed LSH model can replicate many features observed in real temporal networks including reciprocity and transitivity, while also achieves superior prediction accuracy and provides more interpretability compared to existing models.

1 INTRODUCTION

Dynamic networks are used to represent time-varying relationships (edges) between a set of nodes. They are useful in a variety of application settings, including messages between users on online social networks and transactions between users on online marketplaces. In such settings, the network typically evolves over time through a set of *timestamped relational events*. Each event is a triplet (u, v, t) denoting that node u initiated an interaction with node v (e.g. u sent a message to v) at timestamp t. We refer to this type of dynamic network as a *continuous-time network* because it is continuously evolving through these relational events.

A topic of much recent interest is identifying latent representations for nodes in networks. These latent representations are often referred to as node embeddings, and node embedding-based approaches for common network analysis tasks including link prediction have gained significant attention in recent years [Tang et al., 2015, Grover and Leskovec, 2016, Goyal and Ferrara, 2018, Cui et al., 2018]. Prior to this surge of interest, latent space models have been used in statistics and mathematical sociology for exploratory analysis of networks [Hoff et al., 2002, Hoff, 2005, 2007, Handcock et al., 2007, Krivitsky et al., 2009]. Latent representations have also been developed for dynamic networks evolving over discrete time steps [Sewell and Chen, 2015] or in continuous time [Nguyen et al., 2018].

Latent space representations can be combined with temporal point processes (TPPs) to form a probabilistic generative model for continuous-time networks, which we consider in this paper. Augmenting the latent representation with a TPP enables one to generate timestamps for the edges between nodes. Yang et al. [2017] proposed the dual latent space (DLS) generative model that combines two types of latent spaces with bivariate Hawkes processes. They found that using two types of latent spaces, one to capture homophily and one to capture reciprocity, provides a richer model that also leads to improved link prediction accuracy. However, much of the interpretability of the latent space, which was the original motivation of the latent space model of Hoff et al. [2002], is lost by using multiple high-dimensional latent spaces. Thus, we consider using a single latent space representation to provide a more interpretable model.

Our main contributions are the following: (1) We propose the latent space Hawkes (LSH) model that uses a single latent space along with sender and receiver effects and both self and reciprocal excitation to model continuous-time networks. (2) We propose an alternating minimization approach to jointly estimate model parameters and latent positions of nodes. (3) We demonstrate that our LSH model is competitive with other models in predictive and generative tasks on 4 real network datasets while providing more interpretable model fits. (4) We apply our LSH model to perform exploratory analysis on a dataset of militarized disputes to reveal network structure between countries.

Submitted to the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022). **To be used for reviewing only**.

2 BACKGROUND

2.1 HAWKES PROCESSES

The Hawkes process model was introduced for temporal point processes by Hawkes [1971]. The defining characteristic of a Hawkes process is that it is self exciting, meaning that each event increases the rate of future events for some period of time. Mutually exciting Hawkes processes allow events from different processes to excite each other in addition to self excitation [Laub et al., 2021]. An m-dimensional mutually exciting Hawkes process is characterized by a conditional intensity function for each dimension i:

$$\lambda_i^*(t) = \lambda_i(t|\mathcal{H}_t) = \mu_i + \sum_{j=1}^{m} \sum_{k:t_k<t} \phi_{ij}(t - t_k), \quad (1)$$

where $\mathcal{H}_t$ denotes the history of the process up to time t, μ_i denotes the baseline rate of events in dimension i, and $\phi_{ij}(\cdot)$ is a kernel function that describes how an event in dimension j influences dimension i.

The most commonly used kernel function is the exponential kernel $\phi(t-t_k) = \alpha\beta e^{-\beta(t-t_k)}$ for $\alpha > 0$ and $\beta > 0$. With each event arrival, the conditional intensity jumps by α. The influence of the arrival then exponentially decays at rate β over time. In practice, both α and β are unknown parameters that need to be estimated from data, which is usually done using maximum likelihood estimation [Laub et al., 2021]. However, it is computationally efficient to make decay β fixed due to the non-convexity of estimating it [Lemonnier and Vayatis, 2014].

An approach that is more general than fixing the value of β is the sum of exponential kernels method [Lemonnier and Vayatis, 2014], which defines $\phi(t - t_k) = \sum_b^B \alpha\beta_b e^{-\beta_b(t-t_k)}$, where B denotes the number of exponential kernels. This method generalizes better as it handles different time scales, which makes the modeling less sensitive to choice of β. We use the sum of exponential kernels decay in this paper.

2.2 LATENT SPACE MODELS

The latent space model (LSM), first proposed by Hoff et al. [2002] is a popular model-based approach for social network analysis. Designed initially for a single static undirected network, the LSM allows the probability of an edge between two nodes to depend on their Euclidean distance in an unobserved or latent space using a logistic regression model. Let A denote the adjacency matrix of a network, with $a_{uv} = 1$ for node pairs (u, v) with an edge and $a_{uv} = 0$ otherwise. By assuming conditional independence between node pairs, the log-likelihood can be written as

$$\log P(A|\eta) = \sum_{u<v} \left[\eta_{uv} a_{uv} - \log(1 + e^{\eta_{uv}})\right],$$

where entry η_{uv} in the matrix η denotes the log odds of an edge being formed between nodes (u, v). η_{uv} is parameterized as follows: $\eta_{uv} = \xi - \|z_u - z_v\|_2$, where z_u denotes the latent position of node u in a d-dimensional latent space, and ξ is an intercept term. Under this parameterization, two nodes with closer latent positions have higher probability of forming an edge.

The latent space model provides a visual and interpretable model-based spatial representation of social relationships. It has been extended by many researchers. Handcock et al. [2007] developed a latent position cluster model to capture transitivity, homophily, and community structure simultaneously. The latent space models were later extended to include node-specific random effects by Krivitsky et al. [2009]. Latent space models have also been extended for more complex network based data structures, including multiple networks [Gollini and Murphy, 2016, Salter-Townshend and McCormick, 2017], discrete-time dynamic networks [Sewell and Chen, 2015, 2016, Friel et al., 2016, Gracious et al., 2021], and multimodal networks [Wang et al., 2019]. We use the latent space model as the building block for our proposed continuous-time LSH model.

2.3 RELATED WORK

Dynamic Network Embeddings One line of related work is focused on node embeddings for dynamic networks. Compared to static network embedding methods, dynamic network embedding methods assign nodes low-dimensional representations that effectively preserve the temporal information. Nguyen et al. [2018] proposed continuous-time dynamic network embeddings (CTDNE), a general framework to learn a time-respecting embedding from continuous-time dynamic networks. Their framework acts as a basis for incorporating temporal dependencies into existing node embedding and deep graph models based on random walks. Many other approaches for dynamic network embedding have also been proposed [Goyal et al., 2018, 2020, Chen et al., 2018, Sankar et al., 2018]. These and other related methods are discussed in a recent survey paper on dynamic network embedding [Xie et al., 2020].

TPP-based Network Models TPP-based network models are generative models for continuous-time dynamic networks that incorporate both a generative process for the nodes (u, v) that form an edge and the time t at which an edge is formed. These timestamped edges or events can be viewed as triplets (u, v, t). Many TPP-based network models utilize a discrete latent variable representation for the nodes [Blundell et al., 2012, DuBois et al., 2013, Miscouridou et al., 2018, Junuthula et al., 2019, Arastuie et al., 2020], dividing them into different blocks or communities.

The most closely related work to this paper is the dual latent space (DLS) model [Yang et al., 2017], which also utilizes

a continuous latent variable representation inspired by the latent space model. The DLS model uses bivariate Hawkes processes to capture the homophily and reciprocity of dynamic networks. They observed that the latent dimensions of users which affect link formation may be different from the latent dimensions of users which affect reciprocity. We discuss shortcomings of the DLS model and its relation to our proposed model in Section 3.2.

3 PROPOSED MODEL

In our model, we employ a latent space to learn hidden node attributes underlying the network and mutually exciting Hawkes processes to capture the temporal dynamics of communication. We model the communications between each pair of nodes as realizations from a bivariate Hawkes process whose conditional intensity function $\lambda_{uv}(t|\mathcal{H}_t)$ includes three components: a baseline rate, a self-exciting term, and a reciprocal term.

Let z_u and z_v denote the latent positions for nodes u and v, respectively. We model baseline rate μ_{uv} as a function of Euclidean distances between z_u and z_v. Gollini and Murphy [2016] showed that squared Euclidean distances are computationally more efficient than Euclidean distances yet resulted in similar latent positions. Thus, we use squared Euclidean distances $||z_u - z_v||_2^2$ in the model for μ_{uv}, similar to DLS [Yang et al., 2017]. We further add sender and receiver node effect terms δ_u, γ_v to the model as in Hoff [2005], Krivitsky et al. [2009], Wang et al. [2019] to capture the degree heterogeneity, namely the tendency of some nodes to send and receive events more than others, respectively.

We use a sum of B exponential kernels in our Hawkes processes. We set $\beta = (\beta_1, \beta_2, \ldots, \beta_B)$ as a set of fixed known decays and $C = (C_1, C_2, \ldots, C_B)$ as a set of scaling parameters for the kernel with $\sum_i^B C_i = 1$. The conditional intensity function can be written as follows:

$$\begin{aligned}\lambda^*_{uv}(t) = \mu_{uv} &+ \sum_{t_{uv}<t} \sum_b^B C_b \alpha_1 \beta_b e^{-\beta_b(t-t_{uv})} \\ &+ \sum_{t_{vu}<t} \sum_b^B C_b \alpha_2 \beta_b e^{-\beta_b(t-t_{vu})}, \quad \forall u \neq v\end{aligned} \tag{2}$$

where the baseline rate μ_{uv} is given by

$$\mu_{uv} = e^{-\theta_1 ||z_u - z_v||_2^2 + \theta_2 + \delta_u + \gamma_v}. \tag{3}$$

3.1 MODEL PARAMETERS

The LSH model has parameters (Z, α_1, α_2, θ_1, θ_2 $\boldsymbol{\delta}$, $\boldsymbol{\gamma}$). Each node has a d-dimensional latent position z_u, a sender propensity parameter δ_u and a receiver propensity parameter γ_u. α_1 and α_2 are the jump size parameters for self-excitation and reciprocal-excitation. Z is a $n \times d$ matrix whose rows are the latent positions vectors z_us for the n nodes in the network, and d is the latent dimension. Each of $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ is a vector of size n. θ_1 and θ_2 are slope and intercept terms, respectively, associated with the baseline rate and latent positions. A positive slope θ_1 provides node pairs closer together in the latent space with a higher probability of forming edges, while a negative slope does the reverse.

Identifiability There are two sets of identifiability problems that need to be discussed. From the observed event times, the Hawkes process parameters can be identified [Ozaki, 1979]. With the baseline intensity parameter μ_{uv} correctly identified, we explore the identifiability of the parameters in the model for μ. The identifiability of parameters in the latent space model has been discussed by Ma et al. [2020] for a single network and by Zhang et al. [2020] for multilayer networks.

Denote 1_n to be the n dimensional vector and $J_n = 1_n 1_n^T$ to be the $n \times n$ matrix whose elements are all 1's. We first note that the magnitude of the parameter θ_1 is not identifiable since it enters the equation for μ as a product with $\|z_u - z_v\|_2^2$. However, the sign of θ_1 is identifiable since $\|z_u - z_v\|_2^2$ is always positive. In the following we set $\theta_1 = 1$ and examine the conditions for identification of other parameters. We have

$$\begin{aligned}\log(\mu_{uv}) &= \theta_2 - \|z_u\|^2 - \|z_v\|^2 + z_u^T z_v + \delta_u + \gamma_v \\ &= \theta_2 + z_u^T z_v + \tilde{\delta}_u + \tilde{\gamma}_v,\end{aligned}$$

where $\tilde{\delta}_u = \delta_u - \|z_u\|^2$ and $\tilde{\gamma}_v = \gamma_v - \|z_v\|^2$. Now let $\tilde{\delta}$ and $\tilde{\gamma}$ denote the n dimensional vectors whose elements are $\tilde{\delta}_i$ and $\tilde{\gamma}_i$ respectively. (All vectors are column vectors.) Writing in matrix form, the above expression is

$$\log(\mu) = \theta_2 J_n + ZZ^T + \tilde{\delta} 1_n^T + 1_n \tilde{\gamma}^T.$$

Theorem 3.1. *Under the following assumptions:*

1. *The latent positions are centered, i.e., $HZ = Z$, where $H = I - \frac{1}{n}11^T$, and*
2. *The total nodal effects sum to 0, i.e., $1_n^T \tilde{\delta} = 0$ and $\tilde{\gamma}^T 1_n = 0$,*

if two sets of parameters $\theta_1, Z, \gamma, \delta$ and $\theta_1', Z', \gamma', \delta'$ lead to the same $\log(\mu)$, then

$$\theta_2 = \theta_2', \ \delta = \delta', \ \gamma = \gamma' \ \text{ and } Z = Z'O,$$

where O is a $d \times d$ orthogonal matrix.

The proof is provided in Appendix A.1. Thus, the remaining model parameters can be correctly identified.

3.2 RELATION TO DLS MODEL

The most similar model to ours is the dual latent space (DLS) model [Yang et al., 2017]. It uses the following form

for the conditional intensity function[1]:

$$\lambda^*_{uv}(t) = e^{-||z_u - z_v||_2^2 + \theta_2} + \sum_{t_{vu}<t} \sum_{b}^{B} \alpha_2 e^{-||x_u^{(b)} - x_v^{(b)}||_2^2} \beta_b e^{-\beta_b (t - t_{vu})}, \quad \forall u \neq v \tag{4}$$

By comparing the form of the conditional intensity function for DLS (4) with that of our proposed LSH model (2), we identify 3 key differences, each addressing a concern regarding the DLS:

1. The DLS utilizes reciprocal latent spaces $X^{(b)}$ to allow different rates of reciprocity between node pairs. This increase in flexibility of the model comes with a key drawback: the estimated latent positions for a node pair (u, v) and kernel b may result in the jump size $\alpha_2 e^{-||x_u^{(b)} - x_v^{(b)}||_2^2} > 1$, which leads to an unstable process. We were unable to simulate new networks from the DLS model fits to real networks due to the instability as we discuss in Section 5.2.2. In contrast, we use just a single jump size α_2 for all node pairs in our LSH model. While this may be less flexible, it does not lead to instability like the reciprocal latent space.
2. The DLS does not have a self excitation component, unlike our proposed LSH (second term in (2)). The lack of self excitation prevents the DLS from modeling repeated edges from node u to v with no response from v back to u. For example, this setting occurs frequently in militarized conflicts between countries, where one country repeatedly threatens or takes action against another country that does not retaliate.
3. The DLS does not have nodal effects parameters (δ_u and γ_v in (2)). This limits its ability to model nodes with different rates of sending or receiving events.

Furthermore, a primary motivation of the latent space model is to embed the network into a single Euclidean space that can be easily visualized and interpreted. By using a single latent space, our proposed LSH is able to provide a much more interpretable model fit compared to DLS.

4 ESTIMATION PROCEDURE

Our model consists of mutually exciting bivariate Hawkes processes over all pairs of nodes. Using the likelihood theorem of Daley and Vere-Jones [2003], we can write the log-likelihood as

$$\log \mathcal{L} = \sum_{u \neq v} \left\{ \sum_{i=1}^{k} \log(\lambda^*_{uv}(t_i)) - \int_0^{t_k} \lambda^*_{uv}(t) dt \right\}, \tag{5}$$

[1]They include also a periodic kernel in addition to the exponential kernels, which we exclude here for ease of comparison.

where k denotes the total number of events and $\lambda^*_{uv}(t)$ takes on the form in (2). We simplify the log-likelihood and improve the efficiency of the estimation by deriving a recursive form as in Ozaki [1979]. More details and the full log-likelihood derivation for our LSH model are provided in Appendix A.2, resulting in the simplified expression in (9).

Latent space models typically assume that the probability of forming an edge between two nodes is inversely proportional to the distances between the node positions in the latent space. Thus, the observation of an edge between two nodes typically pulls them closer together in the latent space. The presence of the slope parameter θ_1 in the baseline rate μ_{uv} for our LSH model (3) allows us to either pull node pairs with events closer together by constraining $\theta_1 > 0$ or push them further apart by constraining $\theta_1 < 0$. Or, we could leave θ_1 unconstrained—we find that this usually results in the estimate $\hat{\theta}_1 > 0$. We use the L-BFGS-B algorithm [Byrd et al., 1995] to minimize the negative log-likelihood (NLL). The gradients of the log-likelihood can be carried out using the Autograd package [Maclaurin et al., 2015] for automatic differentiation of standard Python functions.

4.1 ALTERNATING MINIMIZATION

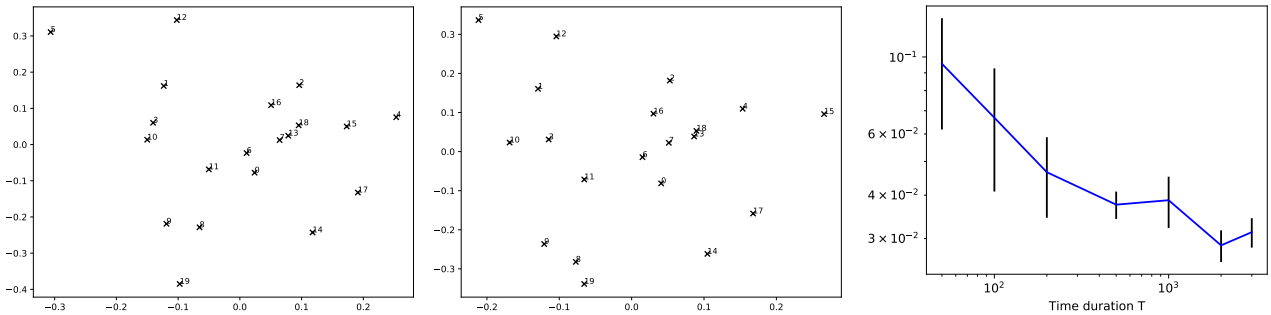
We find that directly minimizing the NLL often leads to convergence to poor local minima; thus, obtaining the maximum likelihood estimate (MLE) for the model parameters is a challenge. Instead, we partition our parameters into two sets: the latent node positions Z and all other parameters $\Theta = (\alpha_1, \alpha_2, \theta_1, \theta_2, \boldsymbol{\delta}, \boldsymbol{\gamma})$. We propose an alternating minimization approach to obtain the MLE. It alternates between optimizing the NLL over the estimated latent positions $\hat{Z}$ while holding all other parameters fixed and optimizing all other parameters $\hat{\Theta}$ while holding the latent positions fixed.

We run each minimization over a fixed number of iterations and then switch to the other minimization. We experiment with different values for the step sizes (s_Θ, s_Z) denoting the number of iterations to run the optimization over Θ and Z, respectively. We find that there is not much difference in the performance for different step sizes for the directions in alternate minimization, and we select the step size of 2 in both directions: $s_\Theta = 2, s_Z = 2$.

We use a multidimensional scaling algorithm as an initialization for the latent space positions Z as in the original latent space model proposed by Hoff et al. [2002]. For all other parameters Θ, we set random values to initialize them. Pseudocode for our alternating minimization estimation procedure is shown in Algorithm 1 in Appendix A.2.

5 EXPERIMENTS

In this section, we perform evaluation tasks for our proposed model on simulated networks and real networks. We use a



(a) (b) (c)

Figure 1: Comparison of (a) actual latent space and (b) estimated latent space (with Procrustes transformation) on a 20 node simulated network with duration $T = 100$. The recovered latent node positions are close to the actual positions. (c) The RMSE over 30 simulated networks (± 2 standard errors) decreases as the duration T increases.

Table 1: Summary statistics of real network datasets

Dataset	Nodes	Events	Time Duration
Reality	65	$2,150$	8 months
Enron	155	$9,646$	15 months
MID	145	$5,088$	23 years
FB-forum	899	$33,720$	5.5 months

sum of $B = 3$ exponential kernels and utilize decays with time scales of an hour, a day, and a week, which is the same as Yang et al. [2017] did in their DLS model. We also fix $C = (1/3, 1/3, 1/3)$ for simplicity[2].

5.1 SIMULATED NETWORKS

We first test our alternating minimization-based estimation procedure on networks simulated from our latent space Hawkes (LSH) model. We simulate networks of 20 nodes in a 2-D latent space using parameters $\theta_1 = 1$, $\theta_2 = -3.2$, $\alpha_1 = 0.01$, and $\alpha_2 = 0.02$. Each dimension of the latent positions as well as sender and receiver effects for nodes are sampled independently from a standard Normal distribution: $z_u, \delta_u, \gamma_u \sim \mathcal{N}(0, 1)$. We increase the time duration T from 50 to $3,000$ and evaluate the estimation accuracy for the latent positions and other parameters. Additional details on the simulation process is provided in Appendix B.1.

A comparison of the actual and estimated latent positions for a simulated network is shown in Figure 1 along with the root mean squared error (RMSE) for estimated latent positions over 30 simulated networks. As expected, the error decreases for increasing time duration T. The error for the other parameters decreases also, as we show in Figure 3 in Appendix B.1. Thus, our proposed alternating minimization procedure appears to accurately estimate latent positions and model parameters.

[2]We also experimented with estimating C but did not find much difference in the results.

5.2 REAL NETWORKS

We perform experiments on several real network datasets, with summary statistics shown in Table 1. Additional details on the datasets are provided in in Appendix B.2. Each dataset consists of a set of relational events where each event is denoted by a sender, a receiver, and a timestamp.

Baselines for Comparisons We compare against three other Hawkes process-based continuous-time network models. The dual latent space (DLS) model [Yang et al., 2017] is the most similar to ours, and we provide a detailed comparison of the DLS model with our proposed LSH model in Section 3.2. We also compare against two recently proposed Hawkes process-based block models: the community Hawkes independent pairs (CHIP) model [Arastuie et al., 2020] and the block Hawkes model (BHM) [Junuthula et al., 2019]. Finally, we compare also against the non-generative continuous-time dynamic network embeddings (CTDNE) [Nguyen et al., 2018] approach. Additional information on these models for comparison along with implementation details are provided in Appendix B.3.

5.2.1 Predictive Accuracy

We first evaluate the predictive ability of our proposed LSH model. We split each dataset into a training set containing the first 70% of events and a test set containing the remaining 30% of events. We estimate model parameters on the training set and evaluate prediction accuracy on the test set. We choose the number of latent dimensions d (for LSH and DLS) and the number of blocks K (for BHM and CHIP) that maximizes the log-likelihood evaluated on the test set.

Test Log-likelihood We use the mean log-likelihood per event on the test set as an evaluation metric for the model's

Table 2: Evaluation metrics for predictive accuracy on real network datasets. Bold entry denotes highest accuracy for each metric on a dataset. The test log-likelihood shows the mean test set log-likelihood per event and the number of latent dimensions d or blocks K that maximize it. The AUC column shows the mean and standard deviation of the AUC across 100 time points for dynamic link prediction. DLS does not scale to the FB-forum data. CTDNE is not generative so test log-likelihood is not applicable.

Dataset	Model	Test log-likelihood	AUC
Reality	LSH	**-3.67** (d=5)	0.945(0.028)
	DLS	-5.64 (d=300)	0.940(0.034)
	BHM	-5.31 (k=50)	**0.957(0.022)**
	CHIP	-4.70 (k=1)	0.937(0.028)
	CTDNE		0.936(0.033)
Enron	LSH	**-5.06** (d=2)	0.938(0.027)
	DLS	-5.29 (d=100)	**0.947(0.017)**
	BHM	-6.35(k=14)	0.839(0.035)
	CHIP	-5.34(k=4)	0.895(0.053)
	CTDNE		0.912(0.035)
FB-forum	LSH	**-7.14** (d=4)	**0.953(0.007)**
	BHM	-11.16(k=57)	0.839(0.017)
	CHIP	-7.65 (k=2)	0.919(0.011)
	CTDNE		0.788(0.028)
MID	LSH	**-3.38** (d=2)	**0.986(0.018)**
	DLS	-4.52 (d=100)	0.977(0.007)
	BHM	-4.97(k=95)	0.971(0.031)
	CHIP	-3.63(k=2)	0.958(0.035)
	CTDNE		0.953(0.018)

predictive ability on future data. As shown in Table 2, our Latent Space Hawkes (LSH) significantly outperforms the other models on all datasets. The test log-likelihood is maximized for the LSH at relatively small latent dimensions d compared to the DLS model. The low-dimensional latent representation using a single latent space makes the LSH fit more interpretable than the high-dimensional DLS representation using multiple latent spaces. Furthermore, these results suggest that the addition of nodal effects and self excitation in the LSH could significantly affect the predictive ability. Finally, we note that the LSH model also achieves better predictive log-likelihood on all datasets compared to the two block models BHM and CHIP.

Dynamic link prediction We further explore the performance of the learned model in a dynamic link prediction task. We use the same experiment set-up as Yang et al. [2017]. We randomly sample 100 time points t_i during the test period. We then compute the probability of a link appearing between each pair of nodes in the $[t_i, t_i + \delta]$ time window. We set δ to be two weeks for the Reality, Enron, and FB-forum datasets and two months for the MID data, which takes place over a longer period of time. For each of these δ intervals, we obtain the Receiver Operating Characteristics (ROC) curve and compute the Area Under the Curve (AUC) measured across all pairs of nodes according to the predicted probabilities given by the model.

The mean AUC values are shown in Table 2 with the value inside the parentheses indicating the standard deviation over these 100 time intervals. The ROC curves and boxplots for the corresponding AUC values are presented in Appendix B.4. We find that our proposed LSH model is competitive at the dynamic link prediction task, achieving highest mean AUC on the FB-forum and MID datasets and second highest on Reality and Enron.

Table 3: Comparison of generative accuracy between models using mean statistic over 15 simulated networks. Bold entry denotes the simulated statistic closest to the actual statistic. While the DLS model is able to replicate the static network statistics from the actual networks, it generates way too many events compared to the actual networks.

		Actual	LSH	DLS
Reality	# events	2,148	**2,236**	9,493
	transitivity	0.29	0.34	**0.32**
	reciprocity	0.80	**0.87**	0.52
	avg coef	0.25	**0.21**	**0.21**
	mean degree	4.86	**4.75**	7.50
Enron	# events	9,646	**10,247**	675,621
	transitivity	0.31	0.43	**0.30**
	reciprocity	0.65	0.57	**0.65**
	avg coef	0.40	0.52	**0.36**
	mean degree	18.46	28.40	**18.43**
MID	# events	5,088	**3,302**	412,890
	transitivity	0.13	0.26	**0.20**
	reciprocity	0.64	0.51	**0.52**
	avg coef	0.25	0.32	**0.29**
	mean degree	6.80	9.45	**9.57**

5.2.2 Generative Accuracy

To evaluate generative accuracy of our proposed LSH model, we simulate networks with our fitted parameters and perform posterior predictive checks (PPCs) using network statistics such as reciprocity and transitivity. While our LSH model has no issues simulating networks, the DLS is problematic due to its model formulation. The jump size for reciprocal excitation depends on distances between nodes in a reciprocal latent space and is further scaled by the parameter α_2 in (4). Since the maximum jump size is not constrained, this results in some node pairs having unstable Hawkes processes so that the simulation does not terminate. To enable us to make comparisons with the DLS model, we stabilize it by fixing the scaling parameter for the jump size $\alpha_2 = 1$.

We simulate 15 networks from the fitted model on each real dataset, with the exception of DLS, which does not scale to the FB-forum data. We then perform PPCs on the number of events generated, and 4 static network statistics: transitivity, reciprocity, average clustering coefficient, and mean degree. A comparison between the actual statistics and mean simulated statistics is shown in Table 3. We present a comparison between LSH and DLS since they are both based on the latent space model. Additional results on generative accuracy, including plots comparing the actual statistics with the distribution of the simulated statistics over all models, are provided in Appendix B.5.

The DLS model generates significantly more events than exist in the actual network, ranging from roughly a 4x increase (Reality) to an 80x increase (MID). We believe that this is due to the reciprocal latent space used in the DLS model. Even though we stabilized the model by setting $\alpha_2 = 1$, some nodes are likely still extremely close in the reciprocal latent space, causing too many events to be generated. Conversely, the DLS model performs quite well at replicating the static network statistics, and in many cases, even better than our proposed LSH. We believe that this is partially due to the much higher latent dimension d that maximizes the test log-likelihood for DLS. The LSH could potentially achieve better generative accuracy using higher d as well.

6 CASE STUDY

We now present a case study demonstrating our proposed LSH model being used for exploratory analysis on a real continuous-time network: the Militarized Interstate Disputes (MID) incident network. Timestamped edges in this network correspond to individual incidents within disputes between countries. Incidents include threats, displays, and uses of force initiated by one country towards another.

Incidents in the MID network are indicative of negative relationships between countries. As a result, one might expect the network to be disassortative. On the other hand, incidents frequently occur between countries that are geographically close, particularly if they share a boundary, which suggests that the network may also have an assortative structure. Thus, we conduct exploratory analysis of this network using two different parameterizations of our model. We fix the latent dimension to be $d = 2$ in both models so that we can visualize the latent positions of the countries.

We first consider a *positive slope* model by constraining $\theta_1 > 0$ in (3) so that two countries with lots of incidents between them are pulled closer together in the latent space, as is typically the case for assortative networks. In this parameterization, countries that engage in lots of incidents are likely to appear centrally in the latent space. We next consider a *negative slope* model by constraining $\theta_1 < 0$ in (3) so that two countries with lots of incidents between them are pushed further apart in the latent space. Under this parameterization, countries that engage in lots of incidents are likely to appear on the periphery of the latent space.

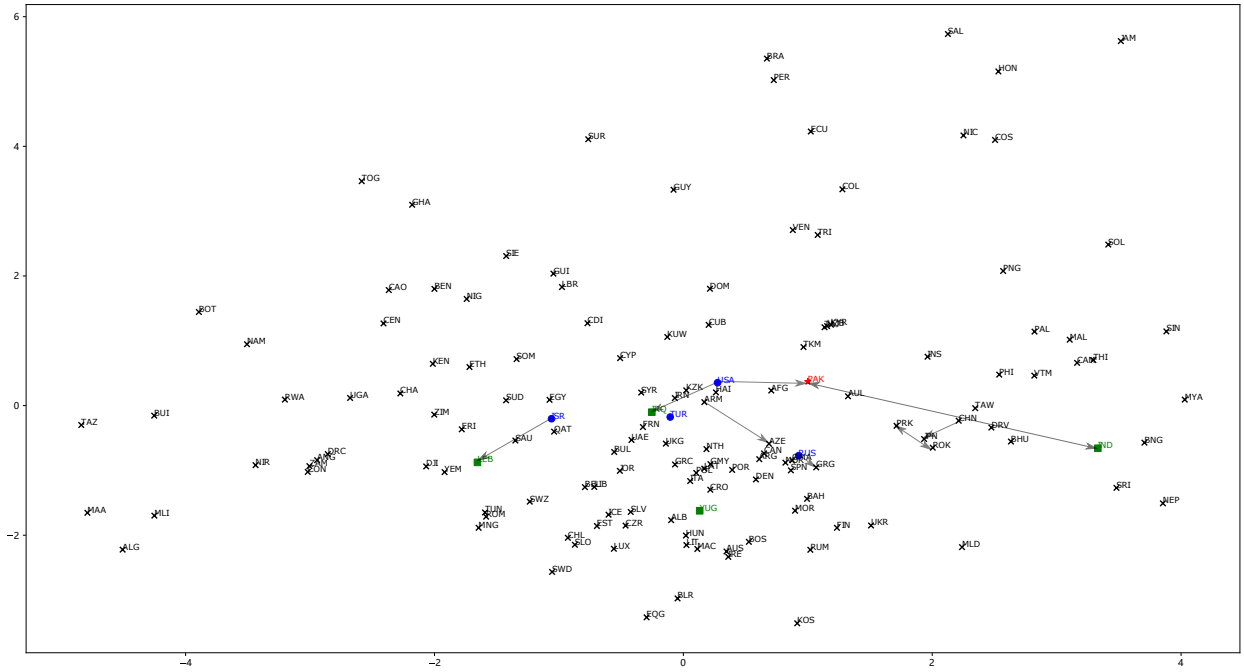
Findings and Discussion We show the 2-D latent space plot with both positive and negative slope terms in Figure 2. We first consider the latent positions from the positive slope model. Notice that the most active nodes tend to appear centrally, and the node pairs with the most frequent incidents, e.g. (USA, IRQ) denoting the conflict between the USA and Iraq, tend to be placed close together. Additionally, countries that are geographically close do mostly appear close together in the latent space. This can be seen from Figure 13 in Appendix C, where nodes are colored by continent. The estimated parameters are $\theta_1 = 1.3, \theta_2 = -8.3, \alpha_1 = 0.81, \alpha_2 = 0.13$. The high value for α_1 compared to α_2 indicates the importance of self excitation in addition to reciprocal excitation.

Next, we consider the negative slope model. From examining the latent positions, we find that the most active nodes tend to appear on the periphery of the latent space, which is reasonable because the model attempts to push nodes with many incidents far apart. For example, the USA and Iraq are on opposite sides of the latent space. The estimated parameters for this model are $\theta_1 = -0.11, \theta_2 = -13, \alpha_1 = 0.86, \alpha_2 = 0.14$. While the θ parameters used for modeling the baseline intensity have changed significantly, the α parameters modeling self and reciprocal excitation are very similar to the positive slope model.

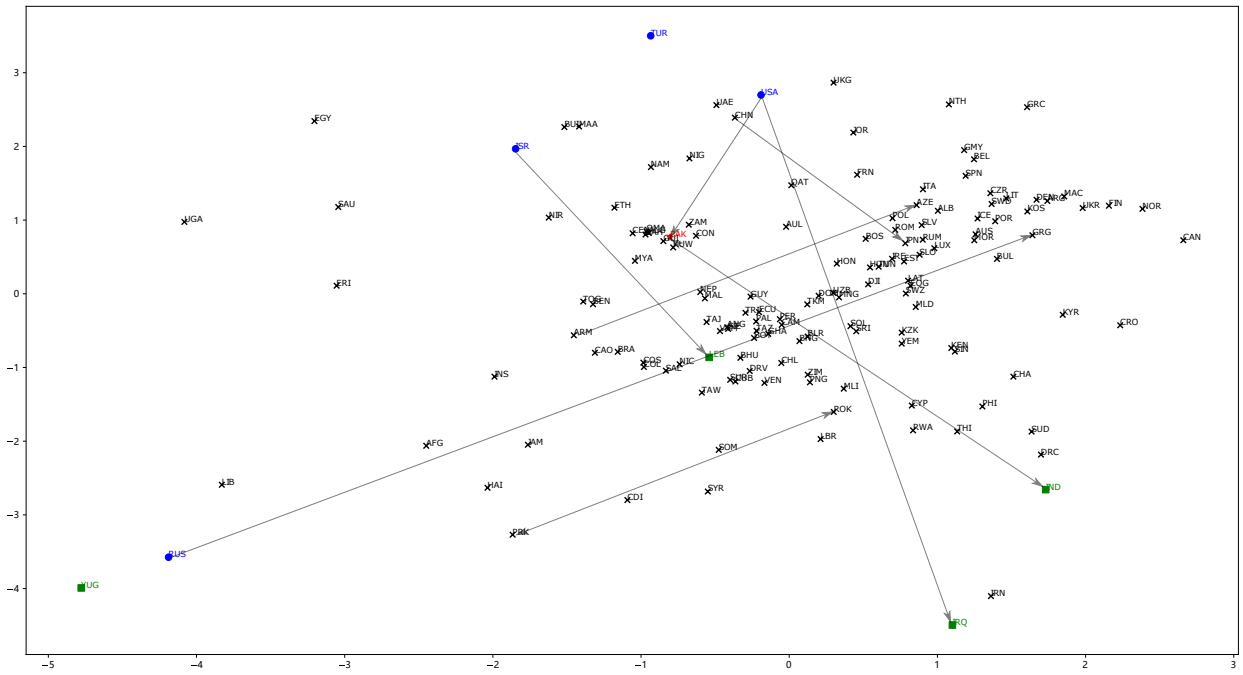
Additional results are presented in Appendix C. We note that this case study is intended to be exploratory rather than confirmatory. We caution readers from jumping to conclusions about countries from our results.

7 CONCLUSION

We proposed the latent space Hawkes (LSH) model for continuous-time networks of relational events, which models interactions between each pair of nodes as realizations from a mutually exciting Hawkes processes whose intensity functions include a baseline rate along with both self and reciprocal excitation terms. The LSH model makes use of a single latent space along with sender and receiver effects to provide a more interpretable fit while remaining competitive in accuracy compared to the dual latent space (DLS) model. We performed an exploratory analysis of militarized disputes between countries using the LSH, where the latent space was quite informative of the dispute network structure. We also found that self excitation was stronger than reciprocal excitation in this network, demonstrating the importance of self excitation, which is not present in the DLS model. We hope this paper inspires future work combining continuous latent space representations with TPPs, which have not gotten as much attention as block model-based TPPs.



(a) Estimated latent positions from model with positive slope



(b) Estimated latent positions from model with negative slope

Figure 2: 2-D latent space plots for LSH model fit to MID data. Edges are shown for the 10 most frequently occurring incidents. The most active countries that initiate and receive the 5 most incidents are shown in blue and green, respectively. Pakistan (PAK) is among the top 5 initiators and receivers and is shown in red. (a) The model with positive slope places countries with lots of conflicts close together. The most active countries tend to appear centrally in this latent space. (b) The model with negative slope places countries with lots of conflicts far apart. The most active countries tend to appear on the periphery of this latent space.

References

Makan Arastuie, Subhadeep Paul, and Kevin Xu. Chip: A hawkes process model for continuous-time networks with scalable and consistent estimation. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Charles Blundell, Jeff Beck, and Katherine A. Heller. Modelling reciprocating relationships with Hawkes processes. In *Advances in Neural Information Processing Systems 25*, pages 2600–2608, 2012.

Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16 (5):1190–1208, 1995.

Xumin Chen, Peng Cui, Lingling Yi, and Shiqiang Yang. Scalable optimization for embedding highly-dynamic and recency-sensitive data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 130–138, 2018.

Peng Cui, Xiao Wang, Jian Pei, and Wenwu Zhu. A survey on network embedding. *IEEE Transactions on Knowledge and Data Engineering*, 31(5):833–852, 2018.

Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes. Volume I: Elementary theory and methods*, volume 2. Springer New York, 2003.

CSIRO's Data61. Stellargraph machine learning library. `https://github.com/stellargraph/stellargraph`, 2018.

Christopher DuBois, Carter T. Butts, and Padhraic Smyth. Stochastic blockmodeling of relational event dynamics. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics*, pages 238–246, 2013.

Nathan Eagle and Alex Sandy Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268, 2006.

Nial Friel, Riccardo Rastelli, Jason Wyse, and Adrian E Raftery. Interlocking directorates in irish companies using a latent space model for bipartite networks. *Proceedings of the National Academy of Sciences*, 113(24):6629–6634, 2016.

Isabella Gollini and Thomas Brendan Murphy. Joint modeling of multiple network views. *Journal of Computational and Graphical Statistics*, 25(1):246–265, 2016.

John C Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.

Palash Goyal and Emilio Ferrara. Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems*, 151:78–94, 2018.

Palash Goyal, Nitin Kamra, Xinran He, and Yan Liu. Dyngem: Deep embedding method for dynamic graphs. *arXiv preprint arXiv:1805.11273*, 2018.

Palash Goyal, Sujit Rokka Chhetri, and Arquimedes Canedo. dyngraph2vec: Capturing network dynamics using dynamic graph representation learning. *Knowledge-Based Systems*, 187:104816, 2020.

Tony Gracious, Shubham Gupta, Arun Kanthali, Rui M Castro, and Ambedkar Dukkipati. Neural latent space model for dynamic networks and temporal knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4054–4062, 2021.

Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.

Mark S Handcock, Adrian E Raftery, and Jeremy M Tantrum. Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):301–354, 2007.

Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

Peter D Hoff. Bilinear mixed-effects models for dyadic data. *Journal of the american Statistical association*, 100(469): 286–295, 2005.

Peter D Hoff. Modeling homophily and stochastic equivalence in symmetric relational data. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*, pages 657–664, 2007.

Peter D Hoff, Adrian E Raftery, and Mark S Handcock. Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97(460):1090–1098, 2002.

Ruthwik Junuthula, Maysam Haghdan, Kevin S Xu, and Vijay Devabhaktuni. The block point process model for continuous-time event-based dynamic networks. In *The World Wide Web Conference*, pages 829–839, 2019.

Bryan Klimt and Yiming Yang. The enron corpus: A new dataset for email classification research. In *European Conference on Machine Learning*, pages 217–226. Springer, 2004.

Pavel N Krivitsky, Mark S Handcock, Adrian E Raftery, and Peter D Hoff. Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social networks*, 31(3): 204–213, 2009.

Patrick J. Laub, Young Lee, and Thomas Taimre. *The Elements of Hawkes Processes*. Springer Nature, 2021. ISBN 3030846393.

Remi Lemonnier and Nicolas Vayatis. Nonparametric markovian learning of triggering kernels for mutually exciting and mutually inhibiting multivariate hawkes processes. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 161–176. Springer, 2014.

Zhuang Ma, Zongming Ma, and Hongsong Yuan. Universal latent space model fitting for large networks with edge covariates. *J. Mach. Learn. Res.*, 21:4–1, 2020.

Dougal Maclaurin, David Duvenaud, and Ryan P Adams. Autograd: Effortless gradients in numpy. In *ICML 2015 AutoML Workshop*, volume 238, page 5, 2015.

Xenia Miscouridou, Francois Caron, and Yee Whye Teh. Modelling sparsity, heterogeneity, reciprocity and community structure in temporal interaction data. In *Advances in Neural Information Processing Systems*, volume 31, pages 2343–2352, 2018.

Giang Hoang Nguyen, John Boaz Lee, Ryan A Rossi, Nesreen K Ahmed, Eunyee Koh, and Sungchul Kim. Continuous-time dynamic network embeddings. In *Companion Proceedings of the The Web Conference 2018*, pages 969–976, 2018.

Tohru Ozaki. Maximum likelihood estimation of hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155, 1979.

Glenn Palmer, Roseanne W McManus, Vito D'Orazio, Michael R Kenwick, Mikaela Karstens, Chase Bloch, Nick Dietrich, Kayla Kahn, Kellan Ritter, and Michael J Soules. The mid5 dataset, 2011–2014: Procedures, coding rules, and description. *Conflict Management and Peace Science*, page 0738894221995743, 2021.

Ryan A. Rossi and Nesreen K. Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI*, 2015. URL `https://networkrepository.com`.

Michael Salter-Townshend and Tyler H McCormick. Latent space models for multiview network data. *The annals of applied statistics*, 11(3):1217, 2017.

Aravind Sankar, Yanhong Wu, Liang Gou, Wei Zhang, and Hao Yang. Dynamic graph representation learning via self-attention networks. *arXiv preprint arXiv:1812.09430*, 2018.

Daniel K Sewell and Yuguo Chen. Latent space models for dynamic networks. *Journal of the American Statistical Association*, 110(512):1646–1657, 2015.

Daniel K Sewell and Yuguo Chen. Latent space models for dynamic networks with weighted edges. *Social Networks*, 44:105–116, 2016.

Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In *Proceedings of the 24th international conference on world wide web*, pages 1067–1077, 2015.

Selena Shuo Wang, Subhadeep Paul, and Paul De Boeck. Joint latent space model for social networks with multivariate attributes. *arXiv preprint arXiv:1910.12128*, 2019.

Yu Xie, Chunyi Li, Bin Yu, Chen Zhang, and Zhouhua Tang. A survey on dynamic network embedding. *arXiv preprint arXiv:2006.08093*, 2020.

Jiasen Yang, Vinayak A Rao, and Jennifer Neville. Decoupling homophily and reciprocity with latent space network models. In *Conference on Uncertainty in Artificial Intelligence*, 2017.

Xuefei Zhang, Songkai Xue, and Ji Zhu. A flexible latent space model for multilayer networks. In *International Conference on Machine Learning*, pages 11288–11297. PMLR, 2020.

A ADDITIONAL MODEL AND ESTIMATION DETAILS

A.1 PROOF OF THEOREM 3.1

Suppose two sets of parameters lead to the same $\log(\mu)$,

$$\theta_1 J_n + ZZ^T + \tilde{\delta} 1_n^T + 1_n \tilde{\eta}^T = \theta_1' J_n + Z' Z'^T + \tilde{\delta}' 1_n^T + 1_n \tilde{\eta}'^T$$

Left multiplying both sides by 1_n^T and noting that (i) $1_n^T Z = 1_n^T H Z = 1_n^T (I_n - \frac{1}{n} 1_n 1_n^T) Z = 0$ by assumption 1 and (ii) $1_n^T \tilde{\delta} = 0$ by assumption 2, we get

$$\begin{aligned} &1_n^T \theta_1 J_n + 1_n^T 1_n \tilde{\eta}^T = 1_n^T \theta_1' J_n + 1_n^T 1_n \tilde{\eta}'^T \\ &\Rightarrow n(\theta_1 - \theta_1') 1_n^T + n(\tilde{\eta}^T - \tilde{\eta}'^T) = 0 \end{aligned} \tag{6}$$

Now right multiplying by 1_n we get

$$n^2((\theta_1 - \theta_1') = 0$$

which implies $\theta_1 = \theta_1'$. With this, using Equation 6, we have

$$n(\tilde{\eta}^T - \tilde{\eta}'^T) = 0 \Rightarrow \tilde{\eta} = \tilde{\eta}'.$$

Finally, right multiplying by 1_n from the beginning and then using the above result $\theta_1 = \theta_1'$ we have

$$\begin{aligned} &\theta_1 J_n 1_n + \tilde{\delta} 1_n^T 1_n = \theta_1' J_n 1_n + \tilde{\delta}' 1_n^T 1_n \\ &\Rightarrow n(\tilde{\delta} - \tilde{\delta}') = 0, \quad \Rightarrow \tilde{\delta} = \tilde{\delta}'. \end{aligned}$$

In light of the above results we then conclude

$$ZZ^T = Z' Z'^T.$$

which means $Z = Z'O$ where $O_{d \times d}$ is an orthogonal matrix such that $OO^T = I$.

A.2 FULL LOG-LIKELIHOOD EXPRESSION

Let $\Lambda_{uv}(t_k^{uv}) = \int_0^{t_k} \lambda_{uv}^*(t) dt$. We can break the time interval $[0, t_k]$ to $[0, t_1], [t_1, t_2], ..., [t_{k-1}, t_k]$, resulting in

$$\begin{aligned}
\Lambda_{uv}(t_k^{uv}) &= \int_0^{t_k^{uv}} \lambda_{uv}^*(t) dt \\
&= \int_0^{t_1^{uv}} \lambda_{uv}^*(t) dt + \sum_{i=1}^{k-1} \int_{t_i^{uv}}^{t_{i+1}^{uv}} \lambda_{uv}^*(t) dt \\
&= \int_0^{t_1^{uv}} \mu_{uv} dt + \sum_{i=1}^{k-1} \int_{t_i^{uv}}^{t_{i+1}^{uv}} [\mu_{uv} + \sum_{j: t_j^{uv} < t_i^{uv}} \sum_b^B C_b \alpha_1 \beta_b e^{-\beta_b (t - t_j^{uv})} + \sum_{j: t_j^{vu} < t_i^{uv}} \sum_b^B C_b \alpha_2 \beta_b e^{-\beta_b (t - t_j^{vu})}] dt \\
&= \mu_{uv} t_k^{uv} + \sum_{i=1}^{k-1} \int_{t_i^{uv}}^{t_{i+1}^{uv}} \sum_b^B [\sum_{j: t_j^{uv} < t_i^{uv}} C_b \alpha_1 \beta_b e^{-\beta_b (t - t_j^{uv})} + \sum_{j: t_j^{vu} < t_i^{uv}} C_b \alpha_2 \beta_b e^{-\beta_b (t - t_j^{vu})}] dt \\
&= \mu_{uv} t_k^{uv} + \sum_{i=1}^{k-1} \sum_{j: t_j^{uv} < t_i^{uv}} \sum_b^B \int_{t_i^{uv}}^{t_{i+1}^{uv}} [C_b \alpha_1 \beta_b e^{-\beta_b (t - t_j^{uv})}] dt + \sum_{i=1}^{k-1} \sum_{j: t_j^{vu} < t_i^{uv}} \sum_b^B \int_{t_i^{uv}}^{t_{i+1}^{uv}} [C_b \alpha_2 \beta_b e^{-\beta_b (t - t_j^{vu})}] dt \\
&= \mu_{uv} t_k^{uv} - \sum_{i=1}^{k-1} \sum_b^B \sum_{j: t_j^{uv} < t_i^{uv}} C_b \alpha_1 [e^{-\beta_b (t_{i+1}^{uv} - t_j^{uv})} - e^{-\beta_b (t_i^{uv} - t_j^{uv})}] dt \\
&\quad - \sum_{i=1}^{k-1} \sum_b^B \sum_{j: t_j^{vu} < t_i^{uv}} C_b \alpha_2 [e^{-\beta_b (t_{i+1}^{uv} - t_j^{vu})} - e^{-\beta_b (t_i^{uv} - t_j^{vu})}] dt
\end{aligned} \tag{7}$$

Algorithm 1 Alternating minimization estimation algorithm

Input: Relational event triplets (u, v, t), latent dimension d, kernel decays β, kernel scale parameters C
Initialize $\hat{Z}$ using multidimensional scaling
Initialize $\hat{\Theta}$ randomly
repeat
  $\hat{\Theta} \leftarrow s_\theta$ steps of L-BFGS-B over $\hat{\Theta}$ on log-likelihood (9) while keeping $\hat{Z}$ fixed
  $\hat{Z} \leftarrow s_Z$ steps of L-BFGS-B over $\hat{Z}$ on log-likelihood (9) while keeping $\hat{\Theta}$ fixed
until log-likelihood (9) converges
Return: Estimated model parameters $(\hat{Z}, \hat{\Theta})$

By extending the summation, many of terms can cancel out, and then we can simplify $\Lambda_{uv}(t_k^{uv})$ as follows:

$$\Lambda_{uv}(t_k^{uv}) = \mu_{uv} t_k^{uv} - \sum_b^B C_b \alpha_1 \sum_{j:t_j^{uv} < t_k^{uv}} [e^{-\beta_b (t_k^{uv} - t_j^{uv})} - 1] - \sum_b^B C_b \alpha_2 \sum_{j:t_j^{vu} < t_k^{uv}} [e^{-\beta_b (t_k^{uv} - t_j^{vu})} - 1]. \tag{8}$$

Substituting equation (8) into the (5) gives the following simplified expression for the log-likelihood.

$$\begin{aligned} l = \sum_{u \neq v} \{ \sum_{i=1}^k log[\mu_{uv} + \sum_{j:t_j^{uv} < t_i^{uv}} \sum_b^B C_b \alpha_1 \beta_b e^{-\beta_b (t_i^{uv} - t_j^{uv})} + \sum_{j:t_j^{vu} < t_i^{uv}} \sum_b^B C_b \alpha_2 \beta_b e^{-\beta_b (t_i^{uv} - t_j^{vu})}] \\ - \mu_{uv} t_k^{uv} + \sum_b^B C_b \alpha_1 \sum_{j:t_j^{uv} < t_k^{uv}} [e^{-\beta_b (t_k^{uv} - t_j^{uv})} - 1] - \sum_b^B C_b \alpha_2 \sum_{j:t_j^{vu} < t_k^{uv}} [e^{-\beta_b (t_k^{uv} - t_j^{vu})} - 1] \}. \end{aligned} \tag{9}$$

We use the log-likelihood expression in (9) as the objective function for our alternating minimization algorithm shown in Algorithm 1.

Algorithm 2 Generative procedure for Latent space Hawkes network

1: **Input:** Number of nodes n, time duration T, latent dimension d, kernel decays β, kernel scaling parameters C, model parameters $(\theta_1, \theta_2, \alpha_1, \alpha_2)$
2: Sample latent positions Z and sender and receiver effects δ, γ from Normal distributions: $Z \leftarrow \mathcal{N}(0, \sigma_z^2 I_{nd}), \delta \leftarrow \mathcal{N}(0, \sigma_\delta^2 I_n), \gamma \leftarrow \mathcal{N}(0, \sigma_\gamma^2 I_n)$
3: Set $Z \leftarrow Z\sqrt{|\theta_1|}, \theta_1 \leftarrow \theta_1/\sqrt{|\theta_1|}$ to remove identifiability issues
4: **for all** node pairs $u \neq v$ **do**
5: $\quad \mu_{uv} \leftarrow e^{-\theta_1||z_u - z_v||_2^2 + \theta_2 + \delta_u + \gamma_v}$
6: $\quad \mathcal{H}(u, v) \leftarrow$ Ogata's thinning algorithm $(\mu_{uv}, \alpha_1, \alpha_2, T, \beta, C)$
7: **end for**
8: **Return:** Simulated network $\mathcal{H}$ containing events $\mathcal{H}(u, v)$ for all $u \neq v$

B ADDITIONAL EXPERIMENT RESULTS

B.1 SIMULATED NETWORKS

The generative process for the simulated networks is shown in Algorithm 2. As discussed in Theorem 3.1, the latent positions Z are only identifiable up to a rotation. Furthermore, the slope parameter θ_1 is not identifiable, so we absorb it into the latent positions by setting

$$\hat{Z} \leftarrow \hat{Z}\sqrt{|\hat{\theta}_1|} \qquad \text{and} \qquad \hat{\theta}_1 \leftarrow \frac{\hat{\theta}_1}{\sqrt{|\hat{\theta}_1|}}.$$

Then, to compare the estimated and actual latent positions, we apply a Procrustes transform [Gower, 1975] to the estimated latent positions $\hat{Z}$.

The root mean squared error (RMSE) for the latent positions, each of the model parameters, and baseline intensities μ_{uv} for all node pairs is shown in Figure 3. As the time duration T increases, more events are generated. Notice that the errors are all decreasing with increasing T as one would expect.

B.2 REAL NETWORK DATASET DESCRIPTIONS

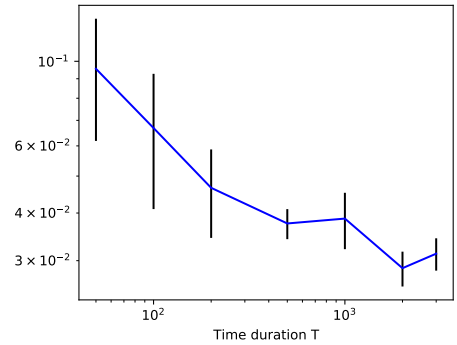
We perform experiments on several real network datasets:

- Reality mining dataset Eagle and Pentland [2006] is derived from the Reality Commons project from 14 September 2004 to 5 May 2005. The dataset we use contains 65 students and 2150 communications (Denote as Reality).
- Enron email dataset Klimt and Yang [2004] consists communications among 155 Enron individuals. It contains 9,646 email message spanning a period of 453 days. This dataset is tested in Yang's paper for their DLS model (Denote as Enron)
- The Militarized Interstate Disputes (MID) dataset [Palmer et al., 2021] consists of 145 countries with 5088 Disputes among them with a period of 8380 days (Denote as MID).
- Facebook-forum dataset Rossi and Ahmed [2015] consists of 899 students posted 33,720 broadcast messages in the forum over 165 days (Denote as FB-forum).

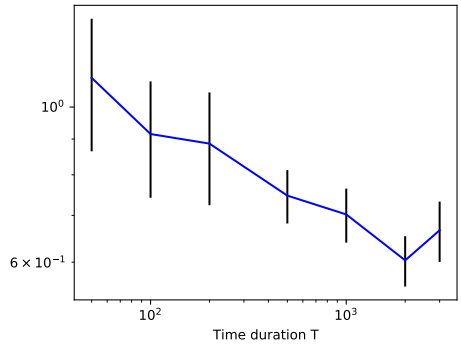
For the Reality, FB-forum, and MID datasets, we re-scale the timestamps so that they are in all $[0, 1000]$ in the same manner as Arastuie et al. [2020]. For Enron, we keep the same scale as Yang et al. [2017] to make a fair comparison against the DLS model.

B.3 DESCRIPTIONS OF OTHER MODELS

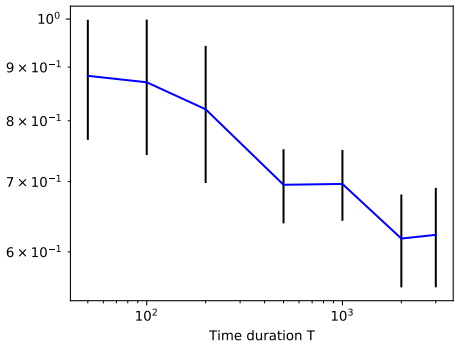
Dual Latent Space (DLS) Model We use the implementation at `https://github.com/jiaseny/lspp` for the DLS model. We provide a detailed comparison of the DLS model with our proposed LSH model in Section 3.2. It adopts a bivariate Hawkes process and latent space-based approach to capture homophily and reciprocity of continuous-time dynamic



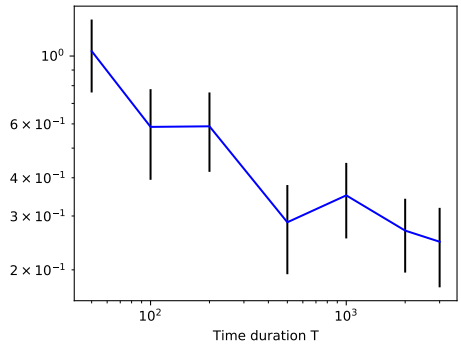
(a) latent positions z



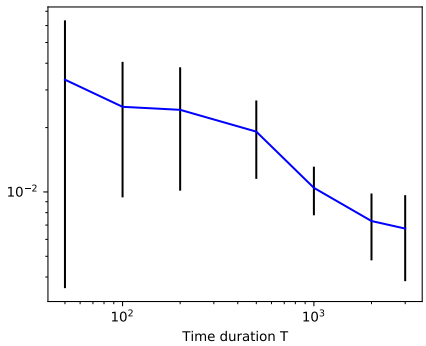
(b) sender effects δ



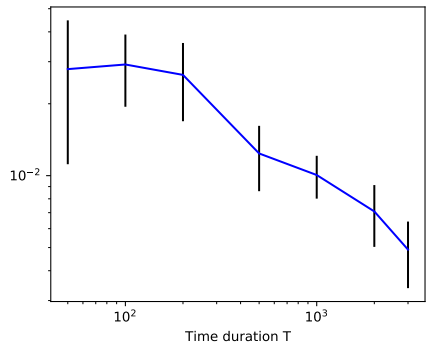
(c) receiver effects γ



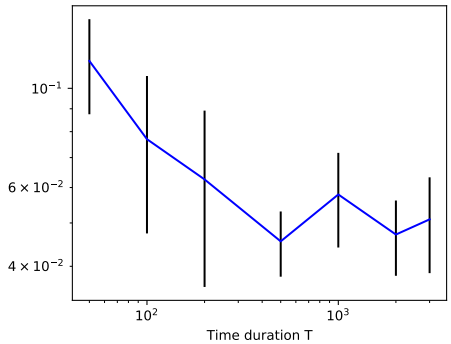
(d) intercept parameter θ_1



(e) jump size for self-exciting term α_1



(f) jump size for reciprocal term α_2



(g) base line intensity μ

Figure 3: Root mean square error over 30 simulated networks ($\pm$ 2 standard errors) for different parameters

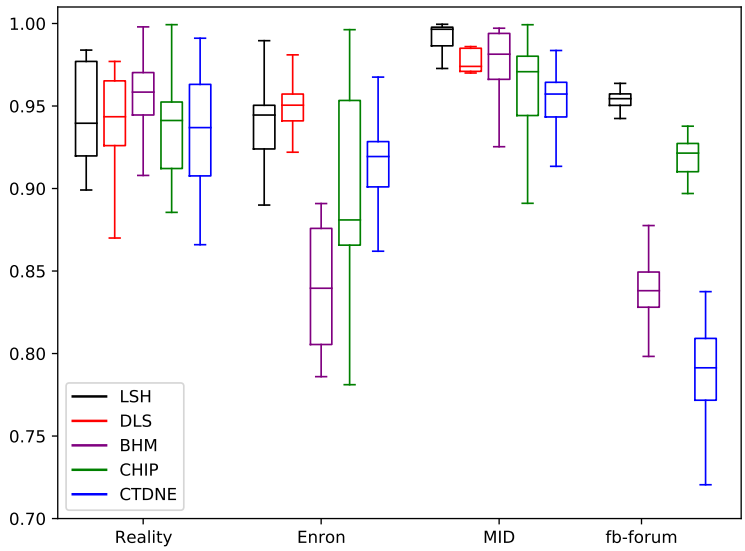


Figure 4: AUC values for dynamic link prediction over 100 random time intervals

networks. Unlike our proposed LSH model, the DLS model does not have the self-excitation term. Their reciprocal terms are parameterized by multiple latent spaces associated with different decay values.

Hawkes Process-based Block Models We use the implementation at `https://github.com/IdeasLabUT/CHIP-Network-Model` for the Community Hawkes Independent Pairs (CHIP) model [Arastuie et al., 2020] and the Block Hawkes Model (BHM) [Junuthula et al., 2019]. CHIP is a univariate Hawkes process network model with block structure where each node pair is independent of all others. The BHM is also a univariate Hawkes process network model with block structure; however, an event between a node pair equally excites all node pairs in the same block pair.

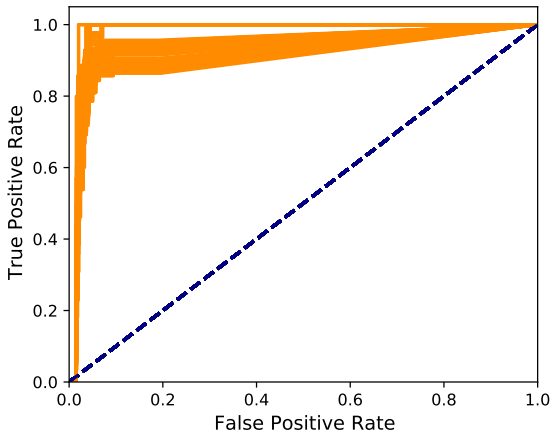
Continuous-time Dynamic Network Embeddings (CTDNE) We used the same hyperparameters ($d = 128, R = 10, L = 80, \omega = 10$) as mentioned in Nguyen et al. [2018]. We used the implementation from the StellarGraph package [Data61, 2018]. We used timestamps up to t_i (beginning of test window) to create both temporal walks and the classifier's positive and negative examples. Edges feature vector is computed using weighted-L2 operation. To create walks, we test using directed/undirected graphs, also varied the neighbor selection distribution between biased(exponential)/unbiased. Best results are reported for each dataset.

B.4 DYNAMIC LINK PREDICTION

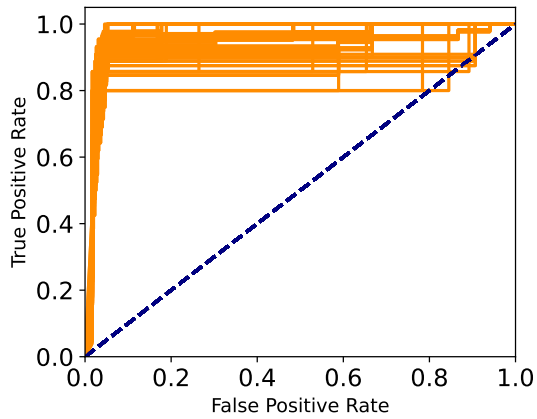
In section 5.2.1, we perform dynamic link prediction experiments. In this section, we demonstrate more results of the AUC values and the ROC curves over 100 time intervals for different models and different real world datasets.

Figure 4 shows the boxplot of the AUC values for dynamic link prediction over 100 random time intervals. The DLS model does not scale to the fb-forum dataset. The boxplot indicates that our proposed model achieves the best in MID and fb-forum dataset and is competitive in Reality and Enron datasets. Moreover, in the fb-forum, MID, and Enron datasets, the dispersion in AUC values from our model is less as evidenced by the low width of the boxes. In all datasets, the bulk of the distribution of AUC values for our LSH model is above 90% indicating a superior performance in the dynamic link prediction task.

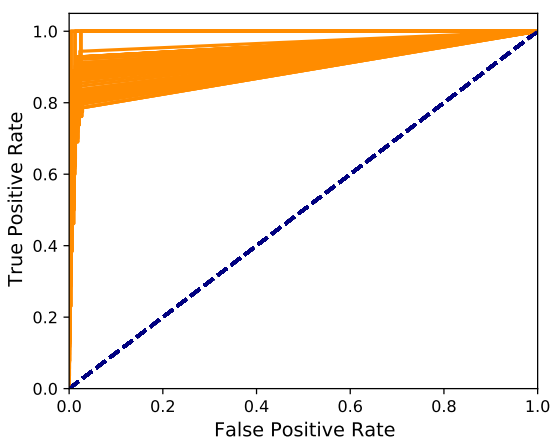
Figure 5, 6, 7, and 8 demonstrate the corresponding ROC curves for different models and real world datasets. The curves matches what we observed in Table 2 that our proposed model obviously outperforms other models on fb-forum and is competitive on Reality, Enron, and MID.



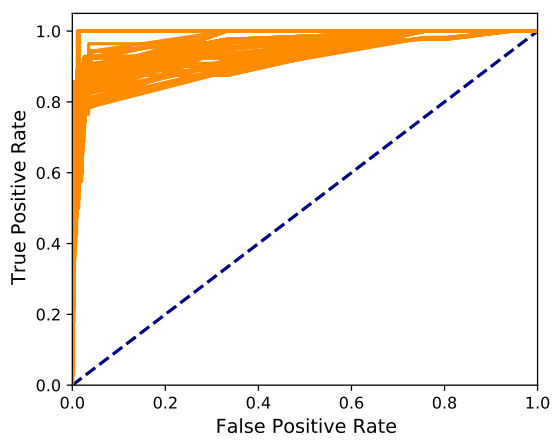
(a) ROC curve of LSH model for Reality Mining



(b) ROC curve of DLS model for Reality Mining

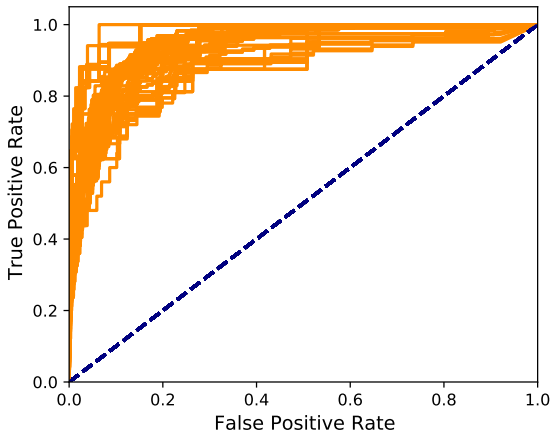


(c) ROC curve of CHIP model for Reality Mining

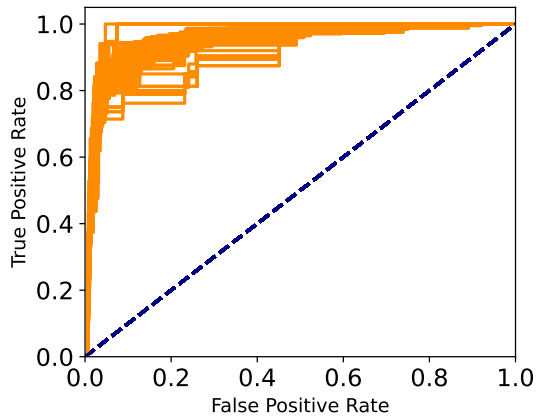


(d) ROC curve of BHM model for Reality Mining

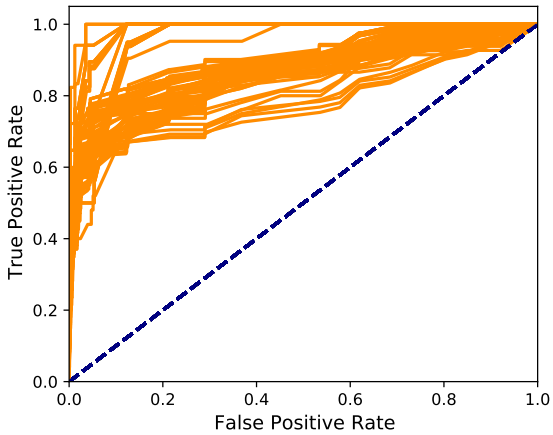
Figure 5: Dynamic link prediction on 100 random time intervals on Reality Mining



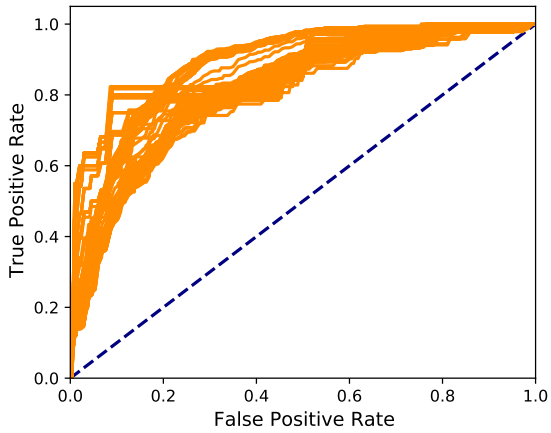
(a) ROC curve of LSH model for Enron



(b) ROC curve of DLS model for Enron

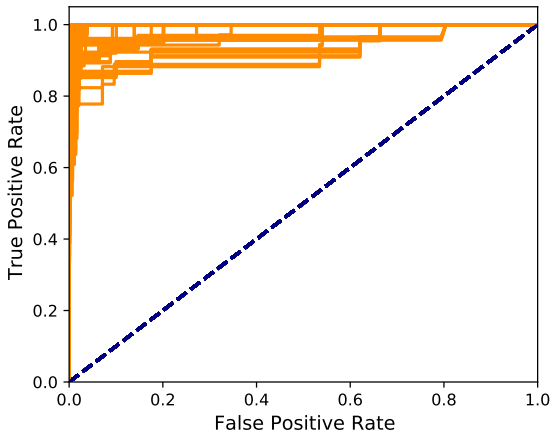


(c) ROC curve of CHIP model for Enron

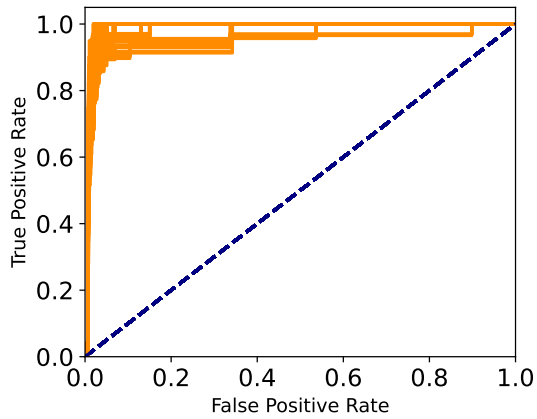


(d) ROC curve of BHM model for Enron

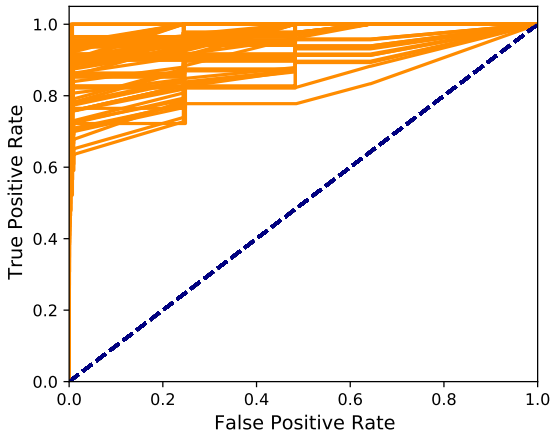
Figure 6: Dynamic link prediction on 100 random time intervals on Enron



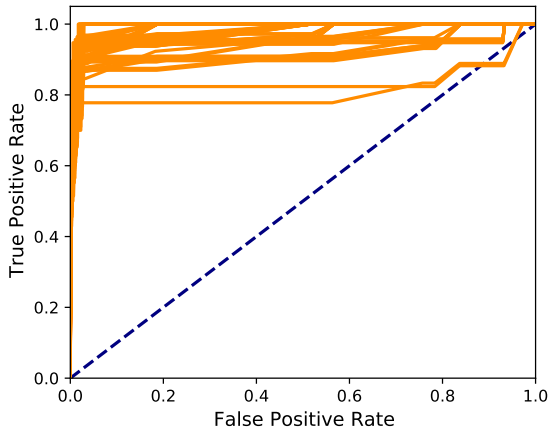
(a) ROC curve of LSH model for MID



(b) ROC curve of DLS model for MID

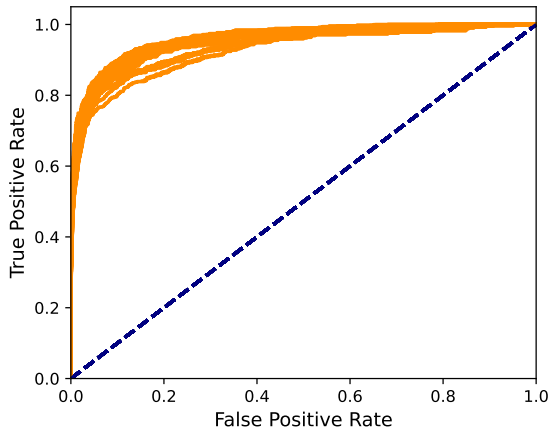


(c) ROC curve of CHIP model for MID

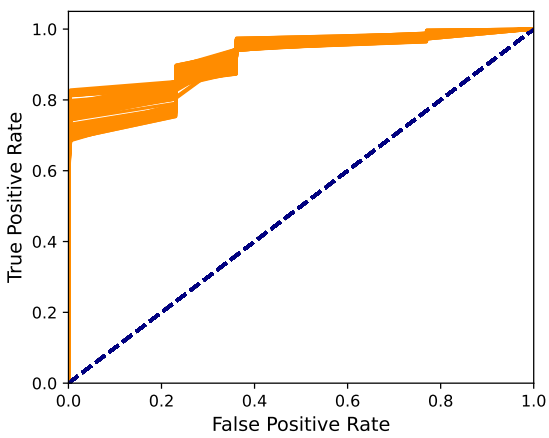


(d) ROC curve of BHM model for MID

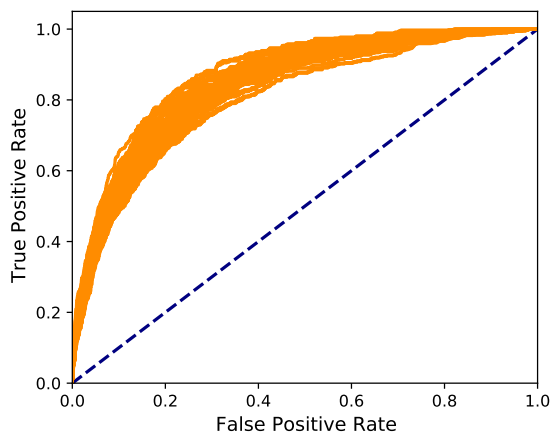
Figure 7: Dynamic link prediction on 100 random time intervals on MID



(a) ROC curve of LSH model for fb-forum



(b) ROC curve of CHIP model for fb-forum



(c) ROC curve of BHM model for fb-forum

Figure 8: Dynamic link prediction on 100 random time intervals on fb-forum

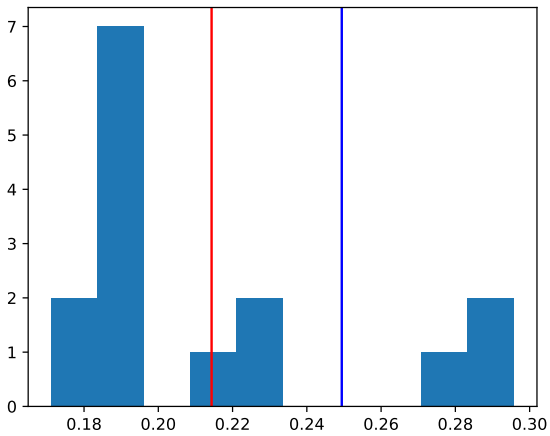
Table 4: Comparison of generative accuracy between models using mean statistic over 15 simulated networks. A portion of this table is shown in the main body as Table 3. Simulations from the BHM fits to Enron and MID were unstable, and DLS does not scale to the fb-forum data.

		Actual	LSH	CHIP	BHM	DLS
Reality	# events	2148	2236	2174	1382	9493
	transitivity	0.29	0.34	0.09	0.38	0.32
	reciprocity	0.80	0.87	0.08	0.78	0.52
	avg coef	0.25	0.21	0.09	0.17	0.21
	mean degree	4.86	4.75	10.88	4.51	7.50
Enron	# events	9646	10247	9510		675621
	transitivity	0.31	0.43	0.20		0.30
	reciprocity	0.65	0.57	0.21		0.65
	avg coef	0.40	0.52	0.14		0.36
	mean degree	18.46	28.40	25.43		18.43
MID	# events	5088	3302	3914		412890
	transitivity	0.13	0.26	0.03		0.20
	reciprocity	0.64	0.51	0.04		0.52
	avg coef	0.25	0.32	0.03		0.29
	mean degree	6.80	9.45	8.91		9.57
fb-forum	# events	33412	59889	56937	61168	
	transitivity	0.04	0.15	0.03	0.07	
	reciprocity	0.01	0.06	0.01	0.06	
	avg coef	0.03	0.21	0.03	0.05	
	mean degree	16.18	49.38	40.24	134.68	

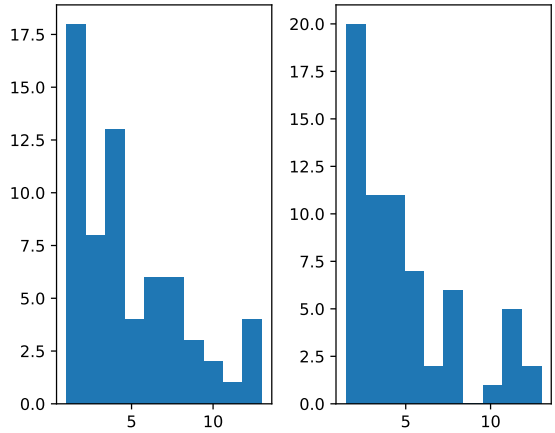
B.5 POSTERIOR PREDICTIVE CHECKS

As we discussed in Section 5.2.2, we simulate 15 networks from the fitted model and demonstrate the PPCs on the statistics in Table 4. We show the corresponding histogram of the PPCs for Reality Mining in Figure 9, 10, 11, and 12 for different models. The actual value of the statistics is plotted as the blue vertical line. The red vertical line indicates the mean value of the statistics observed over the 15 simulated networks. Each figure consists of five subplots. The first, third, fourth, and fifth subplots give the histograms of average clustering coefficient, counts of events, reciprocity, and transitivity observed over the 15 simulated networks. The second subplot in each figure shows the histogram of the degree distribution. The left is the actual degree distribution, and the right is the degree distribution of the mean degree for each node of 15 simulated networks.

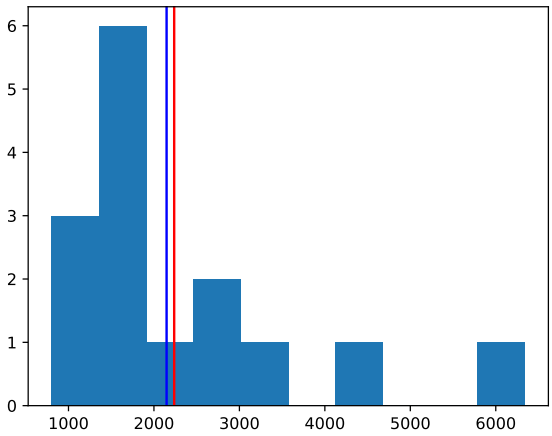
In general, the LSH performs quite well on generating average clustering coefficients, number of events, reciprocity, transitivity, and degree distribution that are similar to what is observed in the corresponding real dataset. The DLS performs well on the average clustering coefficients, the transitivity, and the degree distribution but simulates orders of magnitude more events. The DLS model also fails to simulate the high reciprocity observed in the data. The BHM captures the reciprocity and transitivity very well but replicates fewer events and lower average clustering coefficients than the actual statistics. The CHIP replicates quite well on the number of events; but fails to capture the average clustering coefficients, reciprocity, and transitivity, except in the fb-forum dataset which has very low values for these statistics. The CHIP model, being based on stochastic block model, struggles at replicating properties that typically originates from dependence. Therefore in terms of generative ability, the proposed LSH is capable of generating the most realistic features among all models considered. The other latent space based model DLS is also successful in replicating many of these features, but simulates an unrealistically high number of events making the model not suitable as a generative model.



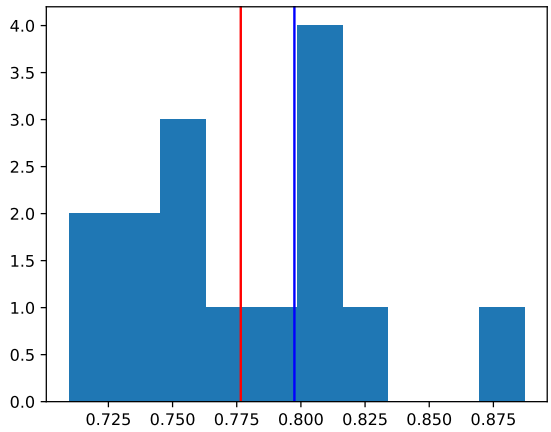
(a) Histogram of average clustering coefficients



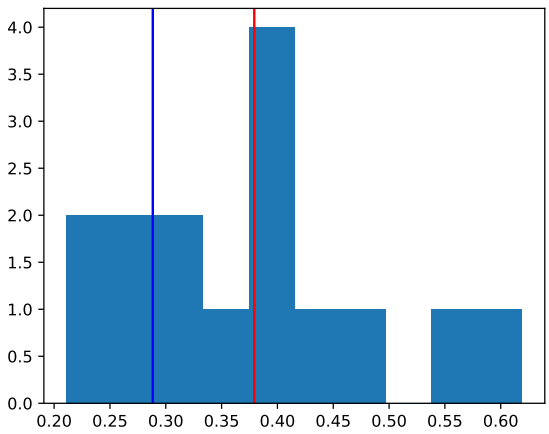
(b) Histogram of degree distribution for real network(left) and mean of 15 simulated networks (right)



(c) Histogram of number of counts

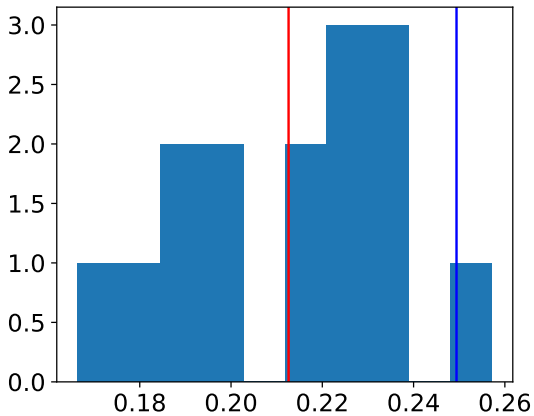


(d) Histogram of reciprocity

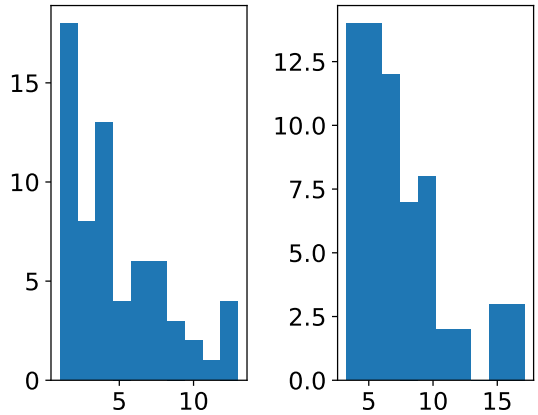


(e) Histogram of transitivity

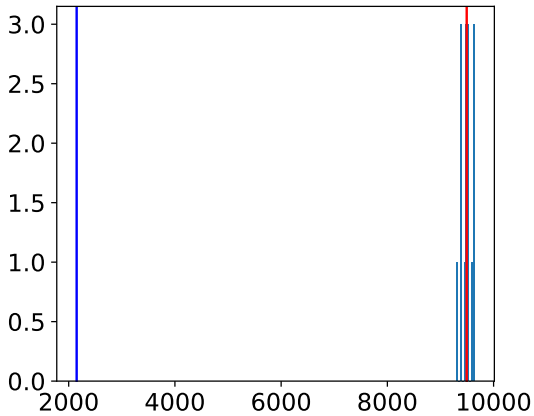
Figure 9: Histogram of 15 simulations from the LSH model fitted to Reality mining (blue line: actual value; red line: mean of 15 simulated value).



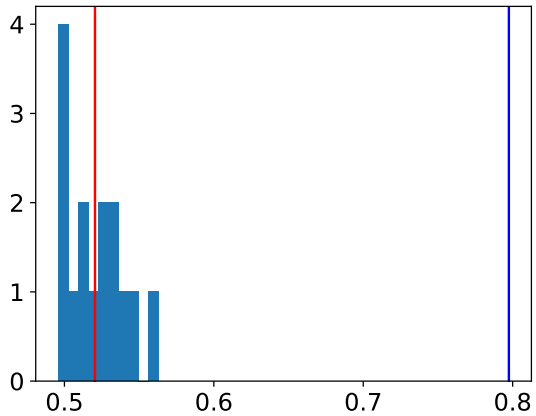
(a) Histogram of average clustering coefficients



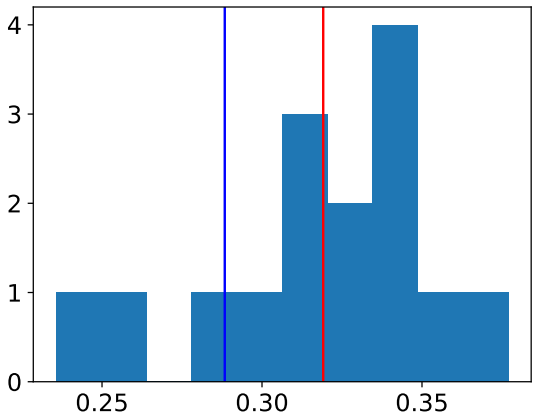
(b) Histogram of degree distribution for real network(left) and mean of 15 simulated networks (right)



(c) Histogram of number of counts

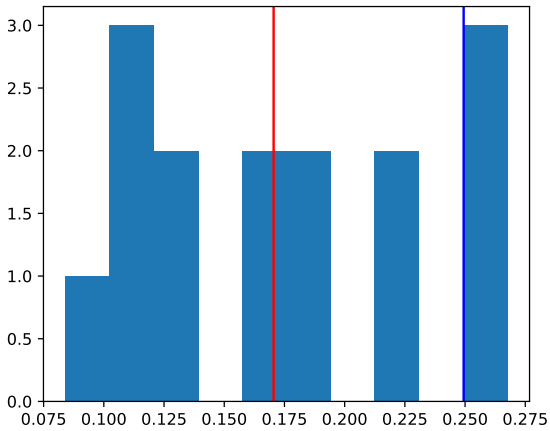


(d) Histogram of reciprocity

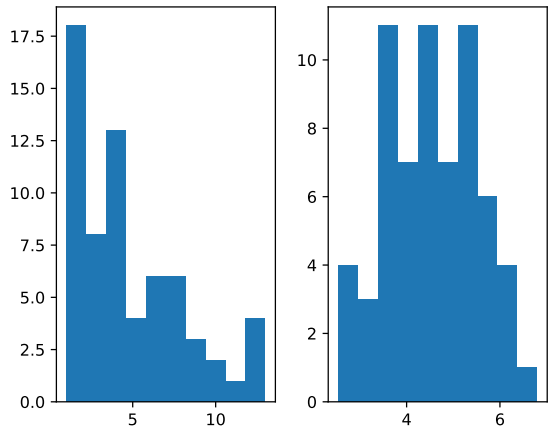


(e) Histogram of transitivity

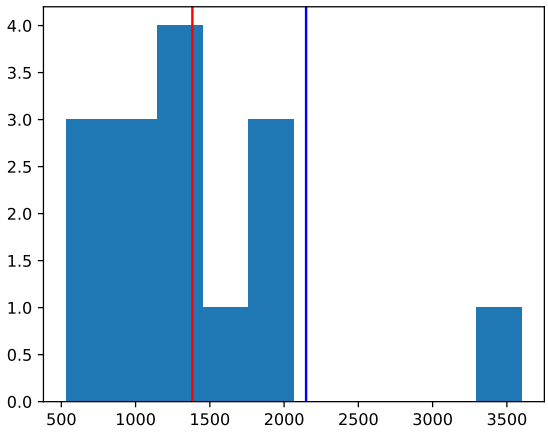
Figure 10: Histogram of 15 simulations from the DLS model fitted to Reality mining (blue line: actual value; red line: mean of 15 simulated value).



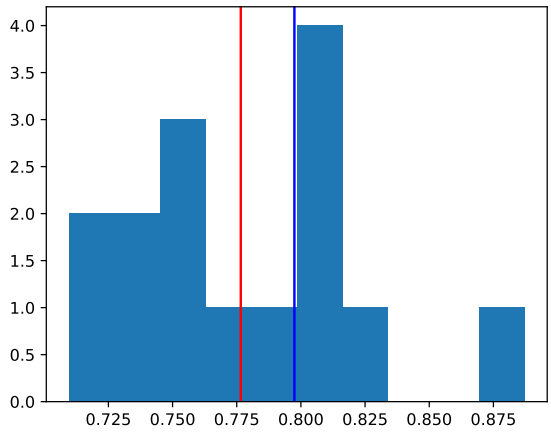
(a) Histogram of average clustering coefficients



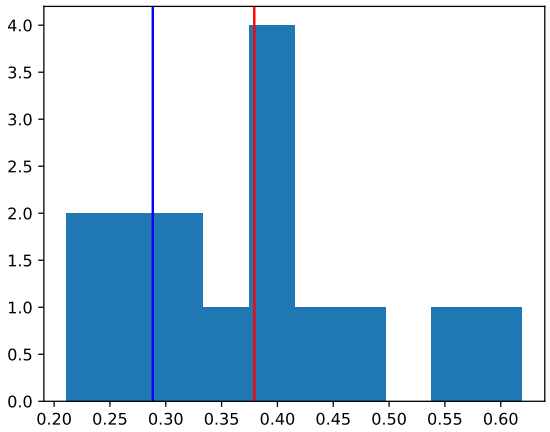
(b) Histogram of degree distribution for real network(left) and mean of 15 simulated networks (right)



(c) Histogram of number of counts

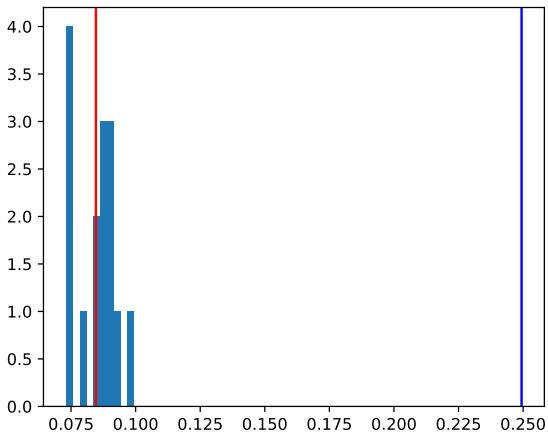


(d) Histogram of reciprocity

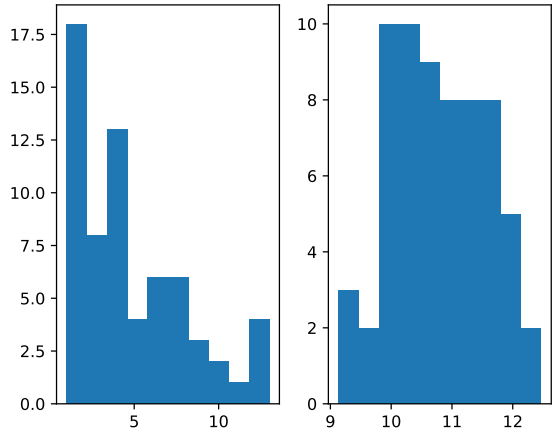


(e) Histogram of transitivity

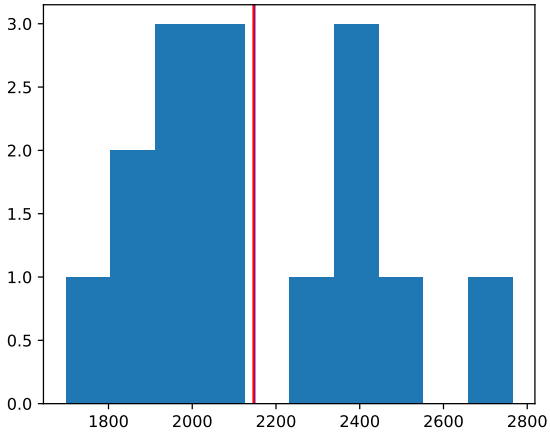
Figure 11: Histogram of 15 simulations from the BHM model fitted to Reality mining (blue line: actual value; red line: mean of 15 simulated value).



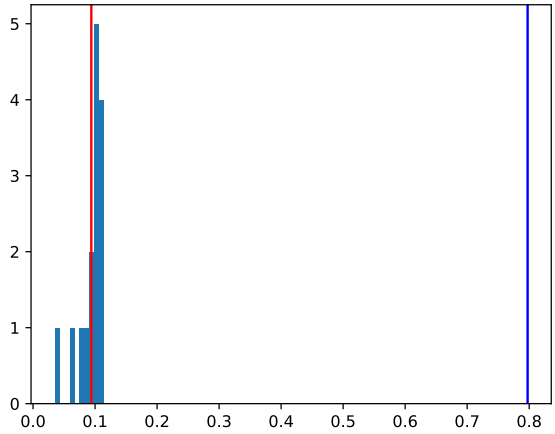
(a) Histogram of average clustering coefficients



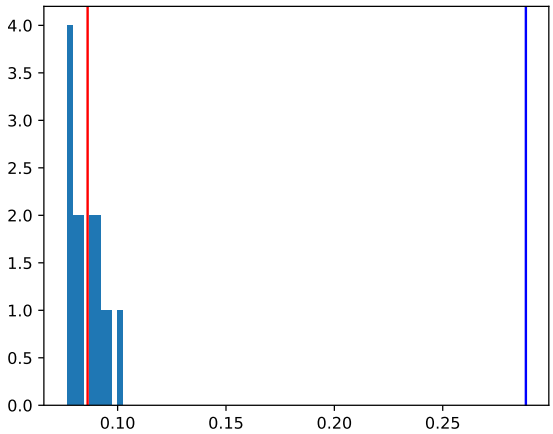
(b) Histogram of degree distribution for real network(left) and mean of 15 simulated networks (right)



(c) Histogram of number of counts



(d) Histogram of reciprocity



(e) Histogram of transitivity

Figure 12: Histogram of 15 simulations from the CHIP model fitted to Reality mining (blue line: actual value; red line: mean of 15 simulated value).

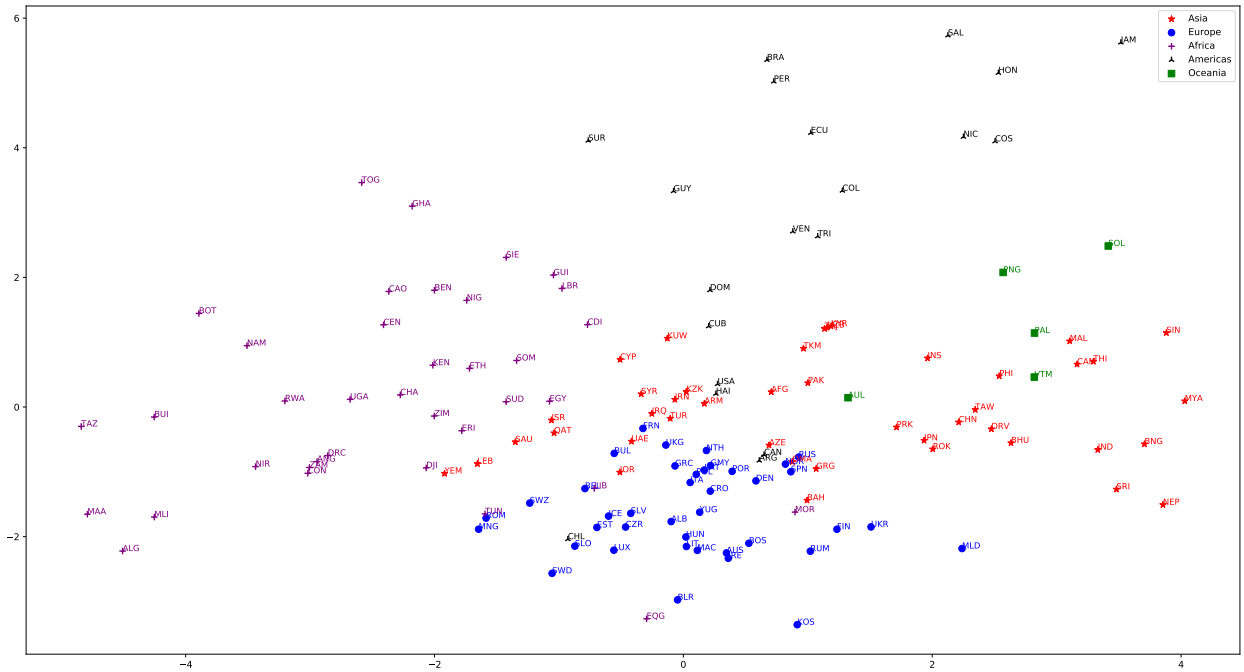


Figure 13: 2-D latent space plot for MID data with positive slope and countries colored by continent.

C ADDITIONAL CASE STUDY RESULTS

Figure 13 shows an alternate view of the positive slope latent space from Figure 2a, with nodes colored by continent. Notice that the latent positions of the countries tend to reflect their geographic proximities, with countries in the same continent mostly appearing together.

In Table 5, we show pairs of countries that have the top 50 number of incidents by one country towards another. In Figure 2, we draw the edges to represent the top 10 frequently occurring incidence between pairs of countries in table 5. The edge with two-way arrows indicate they both serve as initiator and receiver to the other (e.g. PRK and ROK both has disputes to the other with the number of incidences 109 and 95 respectively). Table 6 shows the estimated sender and receiver effects parameters from the LSH model with positive slope. To compare with the highlighted countries in 2, countries that initiate more incidents tend to have higher sender effect and countries that receive more incidents tend to have higher receiver effect, e.g. USA has sender effect 1.84, and Yugoslavia (YUG) has receiver effect 2.57.

Table 5: Pairs of countries that have the top 50 highest incidents

Initiator	Receiver	# of incidents	Initiator	Receiver	# of incidents
ISR	LEB	588	CHN	TAW	38
PAK	IND	174	YUG	ALB	37
USA	IRQ	174	SYR	TUR	34
USA	PAK	140	ALB	YUG	31
ARM	AZE	128	JPN	CHN	31
IND	PAK	116	USA	PRK	29
CHN	JPN	110	IRQ	USA	29
PRK	ROK	109	UKG	YUG	29
ROK	PRK	95	KUW	IRQ	28
RUS	GRG	88	LEB	ISR	28
TUR	GRC	85	FRN	YUG	27
UKG	IRQ	83	AZE	ARM	26
TUR	IRQ	80	RUS	JPN	26
GRC	TUR	73	PHI	CHN	26
SYR	LEB	61	NTH	YUG	26
PAK	AFG	61	ITA	YUG	26
TUR	SYR	54	CAM	THI	25
TUR	CYP	53	TAW	CHN	25
IRN	IRQ	53	SUD	UGA	25
THI	CAM	52	GMY	YUG	25
AFG	PAK	51	TUR	YUG	25
ISR	SYR	50	BEL	YUG	24
RUS	AFG	44	UKR	RUS	24
IND	BNG	44	GRG	RUS	24
USA	YUG	39	GRC	YUG	24

Table 6: Nodal effect parameters estimated by the model with positive slope

Country	send	receive	Country	send	receive	Country	send	receive	Country	send	receive
AFG	-0.77	1.97	DRV	-0.66	-0.57	LUX	-0.09	-0.97	SPN	0.13	-0.71
ALB	-0.62	-0.82	ECU	0.86	-1.4	MAA	-0.19	0.04	SRI	-0.24	-0.33
ALG	0.73	-0.09	EGY	-0.18	0	MAC	-0.33	-0.37	SUD	0.73	1.07
ANG	0.59	-0.28	EQG	0.37	0.15	MAL	0.03	-1.97	SUR	1.95	0.06
ARG	-0.43	-1.04	ERI	0.51	-1.21	MLD	1.5	1.76	SWD	0.63	0.32
ARM	-0.58	-0.65	EST	-0.75	-0.73	MLI	0.41	0.3	SWZ	-0.33	0.24
AUL	0.37	-0.57	ETH	1.11	0.76	MNG	-0.06	-1.6	SYR	-0.78	-0.38
AUS	-1.7	-1.09	FIN	-0.5	0.18	MOR	-0.9	-1.22	TAJ	0.32	-0.2
AZE	-0.46	-0.36	FRN	0.44	-0.22	MYA	-0.86	1.6	TAW	-1.02	-0.33
BAH	-0.25	-0.82	GHA	0.17	0.53	NAM	-0.27	1.29	TAZ	-0.15	-1.86
BEL	-0.46	-0.2	GMY	-0.04	-0.85	NEP	0.44	0.07	THI	-0.29	0.74
BEN	-0.46	-1.22	GRC	0.12	-0.67	NIC	1.5	-0.33	TKM	-0.66	-1.04
BHU	-1.22	-1	GRG	-1.14	0.4	NIG	0.76	-0.32	TOG	1.64	1.43
BLR	-0.27	-0.56	GUI	0.61	-0.25	NIR	-1.8	-0.22	TRI	-0.76	-0.06
BNG	0	1.11	GUY	0.94	0.79	NOR	0.07	-0.56	TUN	-0.45	-0.41
BOS	-0.32	-1.46	HAI	-2.55	0.54	NTH	0.12	-0.88	TUR	1.56	0.26
BOT	-0.39	0.71	HON	0.5	0.25	OMA	-0.53	-0.77	UAE	-0.15	-0.82
BRA	-0.17	-0.82	HUN	0.22	-0.75	PAK	-1.47	0.36	UGA	0.14	1.54
BUI	0.17	2.69	ICE	-1.26	-0.56	PAL	-0.07	-0.56	UKG	0.17	-1.07
BUL	-0.48	-0.14	IND	1.14	0.84	PER	0.95	1.28	UKR	0.17	0.79
CAM	-1.21	-1.26	INS	0.43	0.92	PHI	0.01	1.13	USA	1.84	0.03
CAN	0.61	-0.92	IRE	-2.29	-0.61	PNG	1.79	-0.58	UZB	0.43	0.32
CAO	0.94	-0.04	IRN	1.07	0.82	POL	-0.43	-0.53	VEN	1.51	-2.07
CDI	0.46	0.71	IRQ	-1.27	2.41	POR	-0.28	-0.57	VTM	0.24	0.43
CEN	-0.11	-1.01	ISR	0.84	-0.25	PRK	0.28	0.82	YEM	0.37	0.4
CHA	-1.06	0.63	ITA	-0.34	-0.91	QAT	0.14	-0.36	YUG	2.14	2.57
CHL	-0.91	0.16	JAM	1.55	1.46	ROK	-0.07	0.61	ZAM	-1.27	0.26
CHN	2.31	1.38	JOR	-0.95	-0.98	ROM	-0.35	-0.77	ZIM	-0.4	-0.73
COL	0.4	1.69	JPN	0.34	1.61	RUM	-0.18	-0.19			
CON	-0.98	0.18	KEN	0.27	0.42	RUS	2.45	1.48			
COS	1.05	-0.23	KOS	1.64	0.67	RWA	0.09	2.33			
CRO	-0.18	-0.88	KUW	-0.35	-0.01	SAL	0.77	-0.14			
CUB	-0.45	-0.61	KYR	0	0.55	SAU	-0.23	0.08			
CYP	-1.14	0.25	KZK	-0.88	-1.45	SIE	0.8	0.04			
CZR	-0.67	-0.98	LAT	-0.93	-0.81	SIN	-1.35	0.71			
DEN	0.12	-1.06	LBR	1.12	0.43	SLO	-0.17	-0.3			
DJI	-0.94	0.18	LEB	-0.45	-0.67	SLV	-1.14	-0.31			
DOM	1.4	0.02	LIB	-0.27	2.4	SOL	-0.38	1.22			
DRC	0.63	1.85	LIT	-0.9	-0.77	SOM	-0.52	0.04			

Table 7: Nodal effect parameters estimated by the model with negative slope

Country	send	receive	Country	send	receive	Country	send	receive	Country	send	receive
AFG	0.01	1.96	ECU	0.73	1.68	MAC	-0.85	-1.81	SUR	-0.79	-0.11
ALB	0.40	-0.82	EGY	-1.45	-1.38	MAL	1.02	0.65	SWD	-0.53	-0.48
ALG	-0.65	-0.41	EQG	-1.17	-2.40	MLD	0.51	-0.41	SWZ	0.49	0.77
ANG	1.17	-0.42	ERI	0.58	-0.86	MLI	1.02	1.33	SYR	-0.92	0.93
ARG	-0.83	-0.03	EST	0.30	1.17	MNG	0.91	0.71	TAJ	1.07	-0.23
ARM	0.36	0.46	ETH	-0.80	0.95	MOR	-1.19	-0.11	TAW	-1.31	1.17
AUL	0.73	1.25	FIN	-2.14	-0.87	MYA	0.16	1.45	TAZ	0.15	-0.32
AUS	-1.17	-0.16	FRN	1.86	0.10	NAM	-1.59	-0.41	THI	0.90	-1.54
AZE	1.04	0.48	GHA	1.09	-0.77	NEP	0.02	1.07	TKM	0.96	0.91
BAH	-0.10	0.47	GMY	0.11	-0.06	NIC	-1.27	0.02	TOG	-0.89	-0.32
BEL	-0.15	-0.08	GRC	-0.99	-1.49	NIG	0.19	0.48	TRI	0.05	1.18
BEN	-1.08	-0.43	GRG	0.03	-0.07	NIR	0.01	-0.97	TUN	-0.02	0.35
BHU	1.74	-0.35	GUI	-0.06	0.43	NOR	-0.34	-0.67	TUR	0.83	-0.41
BLR	-0.04	-1.42	GUY	-0.40	1.43	NTH	-0.09	-0.76	UAE	0.54	-2.02
BNG	0.43	-0.83	HAI	-0.10	0.78	OMA	0.06	0.52	UGA	-0.15	-0.98
BOS	-0.28	0.52	HON	-0.65	0.69	PAK	0.66	1.27	UKG	0.65	-0.34
BOT	0.32	0.27	HUN	1.91	-1.30	PAL	-0.43	0.55	UKR	-0.37	-1.59
BRA	-0.03	-1.18	ICE	-0.21	0.00	PER	0.19	1.23	USA	1.91	0.94
BUI	-0.75	-0.88	IND	-0.50	0.25	PHI	-0.87	0.97	UZB	1.18	-0.77
BUL	0.30	0.14	INS	1.92	-1.14	PNG	0.37	-0.64	VEN	1.84	-0.89
CAM	0.05	0.36	IRE	-0.38	0.70	POL	1.51	-0.47	VTM	-0.64	1.09
CAN	-0.04	-0.49	IRN	0.15	-0.29	POR	0.87	0.02	YEM	-1.73	1.43
CAO	-0.41	-0.31	IRQ	-0.77	0.78	PRK	0.27	0.12	YUG	-1.24	-0.97
CDI	0.31	0.03	ISR	1.02	-0.20	QAT	-0.06	0.97	ZAM	-0.70	0.27
CEN	0.20	-1.42	ITA	0.98	0.28	ROK	0.33	0.99	ZIM	1.54	0.16
CHA	1.08	1.10	JAM	-2.48	-1.44	ROM	-0.25	-0.66			
CHL	-0.01	0.00	JOR	0.01	-1.87	RUM	-0.90	0.45			
CHN	1.13	0.76	JPN	0.65	1.84	RUS	0.06	-0.84			
COL	-1.16	0.47	KEN	-2.00	1.36	RWA	1.12	1.07			
CON	-0.23	0.12	KOS	-1.04	-1.66	SAL	-0.62	-0.25			
COS	-1.54	-0.54	KUW	0.55	0.78	SAU	-0.07	-1.38			
CRO	0.06	0.62	KYR	-0.11	-0.67	SIE	-2.29	1.37			
CUB	0.82	0.17	KZK	0.80	-1.67	SIN	-1.37	-0.18			
CYP	-0.05	-0.53	LAT	1.33	0.37	SLO	-0.69	1.21			
CZR	-0.63	-0.14	LBR	0.53	0.71	SLV	0.24	0.25			
DEN	0.08	0.24	LEB	0.25	-0.62	SOL	0.19	-0.60			
DJI	-0.63	0.09	LIB	-0.44	1.18	SOM	-1.07	0.70			
DOM	-0.44	0.59	LIT	-0.72	0.61	SPN	1.00	-0.57			
DRC	1.51	-0.77	LUX	0.81	-0.02	SRI	-0.81	-0.83			
DRV	1.02	0.57	MAA	-2.24	-2.03	SUD	1.76	1.16			