

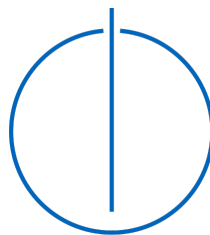
DEPARTMENT OF INFORMATICS

TECHNICAL UNIVERSITY MUNICH

Guided Research Project Report

**A Variational Inference Framework for
Motif Detection with Temporal Point
Processes**

Tobias Schmidt



DEPARTMENT OF INFORMATICS

TECHNICAL UNIVERSITY MUNICH

Guided Research Project Report

**A Variational Inference Framework for
Motif Detection with Temporal Point
Processes**

**Ein Variational Inference Framework für die
Erkennung von Motiven mit Temporal Point
Processes**

Author:	Tobias Schmidt
Supervisor:	Prof. Dr. Stephan Günnemann
Advisor:	Simon Geisler & Marin Biloš
Submission Date:	May 1st 2022

I confirm that this guided research project report is my own work and I have documented all sources and material used.

Munich, May 1st 2022

Tobias Schmidt

Contents

1	Introduction	1
2	Notations and Definition	3
2.1	Motifs in Time Series	4
2.2	Temporal Point Process	4
2.3	Problem Definition	6
3	Related Work	7
3.1	SeqNMF	7
3.2	PP-Seq	9
4	Our Approach	11
4.1	Generative Model	11
4.2	Variational Inference	12
5	Results	15
6	Conclusion	20
	List of Figures	21
	Bibliography	22

1 Introduction

Data mining has been a central part of computer science for over a decade, with pattern recognition and knowledge discovery being key value drivers. The goal of these two is to find or learn about information and phenomena, to generate hypotheses, conclusions, and predictions, and to explore properties and hypotheses that are difficult or impossible to discover with traditional information extraction methods or with the human eye.

A key task of pattern recognition is motif discovery, which has gained more and more traction in the research community over the last decade and in particular over the last years. The term motif originally stems from the field of DNA sequencing and analysis, where the goal is to find specific segments of DNA giving rise to certain regulatory elements of the gene expression mechanism. These segments of DNA that have similar or equal shapes for the same regulatory element are called motifs and play an important role in uncovering the mechanisms that regulate gene expression. Hence, many motif-finding algorithms have been developed for discrete sequences such as DNA [DD07].

The notion of motifs has quickly been picked up in other fields of study and been generalized to capture recurring patterns from a variety of different data sources such as time series, network, and graph data, as well as image and video data. For the latter, motifs can for example be overlaid objects such as text, symbols, or drawings [Her+19] but also whole scenes of historical buildings or certain social situations [YKS16]. In the graph domain, on the other hand, motifs are considered to be repeated sub-graphs or frequently occurring patterns of connectivity [Yu+20][Xia+19]. Equally well studied are motifs for time series data. There, motifs are frequently occurring sub-sequences of one or multiple time series [Mue14][TL17].

Applications for motif discovery algorithms are as widespread as applications of pattern recognition itself. Since the existence of motifs typically hints at important underlying dynamics of the system under study there are multiple ways of applying motif discovery algorithms. For example, they can be used as a stepping stone for other pattern recognition tasks such as knowledge and rule discovery, anytime classification, near-duplicate detection, or summarization. But they can also be of value simply

as a means of qualitative data analysis by discovering previously unknown patterns to be inspected and interpreted by subject matter experts. An example of this from the domain of neuroscience is the study of the repetitive structure of brain activity recording via EEG or similar techniques, which allowed researchers to find hierarchical patterns of brain functions [Mue13][Mue+09].

Surprisingly, in the aforementioned fields of study that quickly picked up on the notion of motifs, a prominent field in the domain of temporal data is missing from the list. Namely the study of temporal point data and processes, which concerns itself with temporal event (timestamped) data. Even though research in this area of study has picked up speed in recent years, leveraging the advances in other areas of pattern recognition and particularly deep learning [Shc+21], the task of motif discovery has barely been studied to the best knowledge of the authors. This is particularly surprising given both the number of domains in which temporal point data is generated, the quantity of data generated in these domains, and the amplitude of challenges and applications this gives rise to [YXL19].

Especially, given the nature of multidimensional temporal point data, which is inherently more sparse, less structured, and harder to interpret visually, the problem of finding motifs in temporal point data should be considered particularly relevant. Typical examples of temporal point data can be found in telecommunications, seismology, social networks, and neuroscience. In any one of these, being able to discover motifs would have the potential to be a huge advancement for the field. For example, extracting behavioral motifs from Twitter users could help better understand their interactions with one another and be used to track down bots or other malicious behavior. In the domain of neuroscience, where advances and widespread of high-density microelectrode arrays for neural signal recordings [Pau+22] have given rise to unseen numbers of high-resolution recordings of neuronal spiking data both in vivo and in vitro, finding motifs in the firing behavior of neurons could provide us with a new understanding of the brain.

Luckily, there have been first advances that try to tackle this problem. Mackevicius et al. developed a novel and simple technique - called *seqNMF* - to discover repeated temporal patterns for event data, demonstrating their algorithm on large-scale neural recordings [Mac+19]. However, their algorithm based on non-negative matrix factorization requires discretizing the data and is based on a suboptimal least-squares reconstruction loss. Because of that, the motif prototypes their technique finds are prone to noise.

A follow-up paper from Williams et al. proposes a radically different approach to overcome the shortcomings of the previous one. They introduce *PP-Seq*, a Bayesian

point process model, more precisely a marked Neyman-Scott process, with which they model the occurrences of motifs and the responses to such an occurrence in the observed temporal data. To fit the model and assign motifs to each individual spike, they use a variation of Gibbs sampling and probabilistic inference [Wil+20]. Their work is superior to Mackevicius et al. in a few key factors: a) They are able to quantify uncertainty both in the assignments of motifs as well as key parameters of the model, and b) they allow the model to operate and find motifs on the raw data, providing assignments to motifs or background noise for each individual time point and c) they improve the scalability by accommodating for the inhering sparsity of the data at hand. Unfortunately, their model lacks the expressiveness to find any motifs more complex than a simple Gaussian impulse response. Additionally, their choice of optimization framework requires closed-form solutions for the conditional likelihoods of an observed event occurring given the presence of a motif at a particular point in time. Having to derive such a closed-form solution makes extending the model to other impulse responses a tedious process at best and for particular choices of impulse responses or simply more complex ones impossible at worst.

Therefore we present a new framework for discovering motifs in temporal point data, extending on and generalizing the work of Williams et al. and Mackevicius et al., overcoming the shortcomings of the previously presented techniques. Our key contributions are as follows:

- We are the first to formally formulate the task of motif discovery for temporal point data
- We propose an extension to the temporal point process model of Williams et al. that allows arbitrarily complex motifs
- We propose a variational inference framework that can handle temporal point process models for the choices of the latent, the observed, and the variational distribution and use it to find a distribution over the occurrences of motifs and the motif impulse responses
- We propose a flexible parameterization of a temporal point process be used as the variational distribution where we can both efficiently sample from and efficiently calculate the joint probability for a given sequence

2 Notations and Definition

The central issue for which we propose a solution is the problem of finding motifs in temporal point data. To formalize this problem and lay the theoretical foundations for the comparison and discussion of related work as well as for our approach, we will introduce some definitions and notes on the notation used in the following chapters.

2.1 Motifs in Time Series

In general motifs in time series can be thought of as similar segments within a single or multiple given time series that repeat themselves. As described by Mueen, motifs can be of interest for mainly two different reasons[Mue14]:

1. because the different instances of reoccurring sub-sequences are significantly similar to each other, such that they could not have occurred by random chance
2. because the sub-sequence or sufficiently similar sub-sequences occurs in such a frequency that this as well can not be a result of random chance

In both these cases, the existence of these sub-sequences means that there exists a prototype motif that carries significantly relevant information about the underlying process that generated the time series. Formally, this results in the definition of similarity-based motifs and support-based motifs[Mue14]:

Definition 2.1.1 *Given a time series and its length, **similarity-based motifs** are the repeated segments in order of their similarities among the repeated occurrences that have at least a minimal similarity to one another.*

Definition 2.1.2 *Given a time series and its length, **support-based motifs** are the segments that have the most number of repetitions that have at least a minimal similarity to one another.*

The convention in most motif discovery algorithms for time series data is to find the top-k motifs. Meaning the top-k most similar motifs or top-k most frequently occurring motifs. For more details on the variations and differences of motif finding algorithms for time series data, we refer the reader to the survey papers from Torkamani and Lohweg and Mueen.

2.2 Temporal Point Process

Temporal Point Processes (TPPs) concern themselves with modeling a specific type of temporal data, where the focus in contrast to time series data is not necessarily on the

values of quantities at certain times, but on the time events occur. Therefore one might also talk about temporal point data as event data or timestamped data. Additionally, while time series data is typically considered to be dense since it usually provides quantities over equally spaced time points, temporal point data is most often sparse. Temporal point data, however, can also come with quantities associated with every time point, which are then called marks. Formally, we define temporal point data as marked event sequences:

Definition 2.2.1 A *marked event sequence* is an ordered sequences $X = \{(t_1, n_1), \dots, (t_S, n_S)\}$ of arrival times $0 < t_1 < \dots < t_S < T$ defined over some time interval $[0, T]$, where each arrival time t_i can have a mark $n_i \in \mathcal{N}$ associated with it.

Generally, having marks is optional and marks can be of arbitrary form, e.g. categorical marks, continuous marks, graphs, or any other data structure.

Another construct crucial for the understanding of the temporal point process is the concept of history, which will allow us to introduce some more formal definitions regarding temporal point processes:

Definition 2.2.2 The *history* of a marked event $x_i \in X$ is $\mathcal{H}(t_i) = \{x_j | t_j < t_i\}$

Definition 2.2.3 A *temporal point process (TPP)* is a stochastic process that defines a probability distribution over a variable length discrete event sequence, which can be uniquely described by either the conditional probability density $p^*(t)$, the conditional intensity $\lambda^*(t)$, the cumulative distribution function $F^*(t)$ or the survival function $S^*(t)$.

Definition 2.2.4 The *conditional intensity function* $\lambda^*(t) = \lambda(t|\mathcal{H}(t))$ is the conditional probability of an event happening in the interval $[t, t + \delta t)$ given that no other event as occurred before in the interval $[t_{i-1}, t)$. The conditional intensity function can be interpreted as the instantaneous firing rate of events per time unit.

Definition 2.2.5 The *conditional density* $p^*(t) = p(t|\mathcal{H}(t))$ gives us the probability of the next event happening in the time interval $[t, t + \delta t)$ given the history $\mathcal{H}(t)$:

Definition 2.2.6 The *cumulative intensity function* is defined as $\Lambda^*(t) = \Lambda(t|\mathcal{H}(t)) = \int_0^t \lambda^*(u) du$ and can intuitively be interpreted as the expected number of events that happen in the interval $[0, t)$

Definition 2.2.7 The *cumulative distribution function* $F^*(t) = F(t|\mathcal{H}(t)) = \int_{t_{i-1}}^t p^*(u) du$ defines the probability of the next event happening before time t

Definition 2.2.8 The *survival function* is defined as $S^*(t) = S(t|\mathcal{H}(t)) = 1 - F^*(t)$ is the complementary cumulative distribution function which means it is the probability of the next event happening after time t .

Given all of these definitions, we can express the likelihood of a given event sequence being generated by a given TPP both in terms of the probability density function and in terms of the intensity function.

Definition 2.2.9 The *likelihood* of a sequence of events $\{t_1, \dots, t_S\}$ defined over the interval $[0, T]$ is:

$$p(\{t_1, \dots, t_S\}) = \left(\prod_{i=1}^S p^*(t_i) \right) \cdot S^*(T) \quad (2.1)$$

$$p(\{t_1, \dots, t_S\}) = \left(\prod_{i=1}^S \lambda^*(t_i) \right) \cdot \exp(-\Lambda^*(T)) \quad (2.2)$$

2.3 Problem Definition

The setting in which we will look into the problem of motif discovery is as follows: We are given a marked event sequence $X = \{(t_1, n_1), \dots, (t_S, n_S)\}$, later referred to as the observed events, defined over the interval $[0, T]$ where the marks $n \in \{1, \dots, N\}$ are categorical and represent N different sources that can generate events. These could for example be a number of different biological neurons from which we record action potential spikes or a number of Twitter users that post tweets at particular times. Note that for each individual biological neuron or Twitter user there exists a corresponding TPP defining the behavior in terms of the arrival times of action potential spikes or tweets respectively.

The motifs we are concerned with are mainly patterns across the N different sources of events but also within a single source. Since our goal will be to find motifs within X we define the prototype of any possible motif via a TPP itself:

Definition 2.3.1 The *prototype* of a *motif* m for temporal point data is given by a marked TPP defined over an interval $[0, T_m]$ with categorical marks representing the different sources of events that participate in the motif. For the existence of a motif in a given event sequence X to be considered, the categorical marks of the motif prototype TPP and the marks in X must be equal. Additionally we constrain $T_m < T$.

Definition 2.3.2 An *instance* of a *motif* m with start point τ_m in a given event sequence X can potentially be any subset $X_m = \{x_1^{(m)}, \dots, x_{s_m}^{(m)}\} \subseteq X$ for which the likelihood $p(X_m') > \epsilon_m$

of this sub-sequence of events surpasses a given threshold ϵ_m . Note that since the TPP of the motif m is defined over the interval $[0, T_m]$ we require to have an offset τ_m which determines the position of the motif within the time interval $[0, T]$ of the original event sequence X . Hence to evaluate the likelihood of X_m being an instance of motif m we define the auxiliary sequence $X'_m = \{(t_1^{(m)} - \tau_m, n_1^{(m)}), \dots, (t_{S_m}^{(m)} - \tau_m, n_{S_m}^{(m)})\}$. Additionally we require that each $x_i \in X$ can only be assigned to at most one motif instance X_m .

With these definitions of motifs and their instances we can formally define our problem statement: Given a marked event sequence X , a number of motif types R , and assignment threshold ϵ_m , we want to find the prototypes of these R motifs and all of their instances in X , such that the total likelihood of all motif assignments is maximized.

3 Related Work

3.1 SeqNMF

Mackevicius et al. propose the framework SeqNMF which uses a variation of non-negative matrix factorization to find repeated temporal sequences. Their approach is straightforward and easy to understand. They start with the same setting formulated in the previous section 2.3:

A sequence of marked events $\{(t_1, n_1), \dots, (t_S, n_S)\}$ for which they discretize the arrival times and create a matrix $X \in \mathbb{R}^{N \times T}$, where N is the number of event sources, T the number of discretized time steps and $X_{i,j}$ is the normalized number of events of the same source within the j -th time step.

Their idea is to approximate X via the sum of K rank-one matrices, where each of them is the outer product of two non-negative vectors $\mathbf{w}_k \in \mathbb{R}_+^N$ and $\mathbf{h}_k \in \mathbb{R}_+^T$. This intuitively means that \mathbf{h}_k is encoding the time and strength at which the k -th pattern is found in X and \mathbf{w}_k is encoding the k -th pattern itself. Note, that with this approach it is not possible to find sequential patterns, i.e. motifs, since the pattern encoded by \mathbf{w}_k can not extend over multiple time steps. Therefore convolutional non-negative matrix factorization (convNMF) is presented which proposes to encode sequential patterns via the non-negative matrices $W_k \in \mathbb{R}_+^{N \times L}$ instead of the vector \mathbf{w}_k . L here can be interpreted as the maximal length of a motif. Additionally, instead of an outer product a convolutional operation $W_k \circledast h_k$ is used to reconstruct the data matrix X . The full term for convNMF is therefore $\tilde{X} = \sum_{k=1}^K W_k \circledast h_k$. Fitting this model via a least-squares

reconstruction loss, however, has some serious drawbacks. There are mainly three types of errors that can occur.

1. **Motif separation error:** A sequential pattern that should be a single motif is split up into two motifs. The corresponding encoding of motif occurrence via h_1 and h_2 will be approximately equal except for a temporal offset w.r.t. to the motif start time.
2. **Motif duplication error:** One or multiple motif prototypes W_k are proportionally equal to each other, with instances of them found at the same location. Essentially reconstructing the same instance, but spreading the strength/amplitude across multiple instantiations.
3. **Instance duplication error:** One or multiple motif prototypes W_k are equal to each other, with instances of them found at different locations.

Mackevicius et al. propose solutions for each of these errors in their seqNMF framework by introducing penalty terms to the reconstruction loss. The penalty term is designed to reduce the correlation between the motif occurrences $H = \{h_1, \dots, h_K\} \in \mathbb{R}^{T \times K}$ as well as the correlation between motifs. Their framework can therefore be summed up in the following formula:

$$(W^*, H^*) = \arg \min_{W, H} \left(\|\tilde{X} - X\|_F^2 + \lambda \left\| (W \circledast X) S H^T \right\|_{i \neq j}^1 \right) \quad (3.1)$$

where S is a smoothing matrix with $S_{i,j} = 1$ if $|i - j| < L$ and otherwise $S_{i,j} = 0$. For further details on the derivation of the penalty term the reader is referred to the paper from Mackevicius et al.

Interestingly enough Mackevicius et al. also proposed a variation of their factorization strategy by introducing additional orthogonality constraints on either W or H , resulting in what they call "event-based" and "parts-based" factorization. Where event-based factorization gives rise to fewer more complex motifs with either similarity to instances of the same motif and part-based factorization gives rise to more frequently occurring less complex motifs. Coincidentally these approximately correspond to the notion of similarity and support-based motifs studied in the literature on time series motifs.

Important to note for seqNMF is that even though it was shown to be robust to noise when it comes to identifying motifs, the motif prototypes found in real-world studies contain a significant amount of noise. This should not be a surprise given that the model does not explicitly handle noise in any form. Nonetheless, it is a significant

downside to this approach. Especially when motifs are not simply used for visual inspection but instead to be used for downstream tasks.

However, the most important flaw of this approach w.r.t. to the proposed problem statement of the section 2.3 is, that it only partially solves it. It can find the approximate times of motif occurrences and even discretized approximations of the motif prototypes. But a) these are only discretized and b) it doesn't provide a solution to the motif assignment part of the problem.

3.2 PP-Seq

Williams et al. identified various shortcomings of seqNMF and proposed their own solution to the problem of motif discovery. In doing so, to the best knowledge of the authors, they are the first and only to propose a technique that finds possible solutions to the problem of motif discovery for temporal point data as we formalized it in section 2.3.

Their approach is inspired by the Neyman-Scott process, which they use to generalize the framework of seqNMF to continuous time. They operate in the same setting as seqNMF that is described in 2.3. But in contrast to seqNMF, they use TPPs to model motifs by formulating a latent variable model.

Their model - called PPSeq - consists of two TPPs: One to model the occurrence, type, and amplitude/strength of motifs and a second one to model the arrival time and source of observed events. They do this by formally defining both processes via conditional intensity functions. The first one, modeling the latent events, i.e. motifs, is defined over the marked event sequence $\{z_1, \dots, z_K\}$ where each event $z_k = (\tau_k, r_k, A_k)$ defines the arrival time $\tau_k \in [0, T]$, the motif type $r_k \in \{1, \dots, R\}$ and the amplitude $A_k \in \mathbb{R}_+ / \{0\}$. Each of these properties is assumed to be independent of each other and follows a homogeneous Poisson process, a categorical distribution, and a gamma distribution respectively. This results in the following intensity function:

$$\lambda(z_k) = \psi \cdot \text{Categorical}(r_k | \pi) \cdot \text{Gamma}(A_k | \alpha, \beta) \quad (3.2)$$

where $\psi \in \mathbb{R}_+ / \{0\}$ is the constant intensity of the homogeneous Poisson process, $\pi \in \mathbb{R}_+^R / \{0\}$ the parameterization of the categorical distribution and $\alpha \in \mathbb{R}_+ / \{0\}$, $\beta \in \mathbb{R}_+ / \{0\}$ the parameterization of the gamma distribution.

For observed events $X = \{(t_1, n_1), \dots, (t_s, n_s)\}$, defined through their arrival time $t_s \in [0, T]$ and source $n_s \in \{1, \dots, N\}$, the intensity function is defined condition on a

sequence of latent events $\{z_k\}_{k=1}^K$. Additionally and in contrast to seqNMF, the intensity function also contains a term modelling the background noise:

$$\lambda_n(t|\{z_k\}_{k=1}^K) = \lambda_n^\varnothing + \sum_{k=1}^K g_n(t, z_k) \quad (3.3)$$

Here, λ^\varnothing is the background noise rate which means the background noise is assumed to follow the distribution of a homogeneous Poisson process and $g_n(t, z_k)$ is the Gaussian impulse response. Note, that the intensity function and thereby the Gaussian impulse response as well as the background noise rate are defined as dependent on the source type n . This simply means that each source can generate observed events according to its own intensity function. The whole sequence of observed events $\{x_s\}_{s=1}^S$ can therefore be viewed as the union of background and motif events generated by N different temporal point processes.

The Gaussian impulse responses are further defined based on the properties of a given latent event and proportionally follow a Gaussian pdf

$$g_n(t, z_k) = A_k \cdot a_{n,r_k} \cdot \mathcal{N}(t|\tau_k + b_{n,r_k}, c_{n,r_k}) \quad (3.4)$$

where $a_{n,r_k} \in \mathbb{R}_+^{N \times R}$ is a weighting or scaling factor based on the source type and motif type, $b_{n,r_k} \in \mathbb{R}_+^{N \times R}$ is the impulse latency or offset and $c_{n,r_k} \in \mathbb{R}_+^{N \times R}$ is the temporal spread or width of the impulse.

Fitting this model to a given sequence of observed events and inferring the motifs is a highly non-trivial task. To tackle this challenge Williams et al. developed a collapsed Gibbs sampling routine that allows them to do efficient parameter inference in a fully Bayesian fashion. However, to do so they 1) require to place conjugate priors on all model parameters 2) introduce an iterative sampling algorithm that assigns and re-assigns individual observed events to a dynamically growing list of partitions representing background events or motif events 3) derive the conditional probabilities for the partition assignments by marginalizing out time, type and amplitude of latent events and 4) derive the posterior and predictive posterior probability to infer the latent event sequence. Additionally, they propose even more improvements to their Gibbs sampling technique in order to make their approach more computationally efficient. Having tackled all of these individual challenges, they were able to successfully demonstrate their algorithm and prove its usefulness in finding simple motifs. Unfortunately, the keyword of the previous statement is "simple".

That PPSeq is only able to find these simple motifs is due to the way $g_n(t, z_k)$ is defined in the equation 3.4. In theory, the framework inspired by Neyman-Scott processes would

allow an arbitrary choice for $g_n(t, z_k)$. But since Williams et al. restricts it to be a proportional to single Gaussian, the expressiveness of a motif w.r.t to the intensity of any single given source n is limited to have intensity increase originating only at a single point in time. For an even slightly more complex motif, PPSeq would fail to capture it properly. For example, a motif with which one might be interested to model e.g. a biological neuron firing repeatedly directly after stimulus onset for a fraction of a second, stop firing for a second, and then firing again for another fraction of a second. PPSeq would not be able to find a single motif like this but only two more simpler motifs that each only cover one of the segments covering the fraction of a second firing impulse.

Crucial to note here is that this design choice of a relatively simple impulse response $g_n(t, z_k)$ is essential for the Gibbs sampling technique to be feasible. Additionally, any change to the generative model a practitioner interested in using this framework would like to make would require the corresponding re-derivation of all conditional and marginalized probabilities of the partition assignments, the posterior as well as the predictive posterior. Apart from the fact, that few practitioners from application fields that should benefit from this model are equipped with the skill set to do so, this might not even be possible depending on the changes made to the generative model. This illustrates clearly the need for a more general framework that can provide essentially the same benefits and solutions while offering higher flexibility in the parameterization of the generative model.

4 Our Approach

In the following, we will present an alternative framework for the discovery of motifs in temporal event sequences for which we developed a more general optimization and inference strategy that opens up the possibility to model almost arbitrarily complex motifs.

4.1 Generative Model

We propose a generalized version of the latent variable model introduced by Williams et al., keeping the latent event model as is, with the latent event sequence $\{(\tau_k, r_k, A_k)\}_{k=1}^K$ defining the arrival times, types and amplitudes of motifs. We model them correspondingly with a TPP specified via the intensity function $\lambda(z_k)$:

$$\lambda(z_k) = \psi \cdot \text{Categorical}(r_k|\pi) \cdot \text{Gamma}(A_k|\alpha, \beta) \quad (4.1)$$

The model for the observed events $X = \{(t_1, n_1), \dots, (t_S, n_S)\}$, however, we generalize in the following way: Similar to PPSeq we choose a TPP specified via a conditional intensity function that depends on the latent event sequence. But in contrast to PPSeq, we a) allow the response to a motif to have arbitrary form and b) use this flexibility to reformulate the background noise as a special case of a motif. Since the background noise has the unique property of inducing a "global" response we restrict the latent event model to always generate exactly one occurrence of the "background motif". This design decision comes with the benefit that we can leave the latent event model as is, since defining an event to always be generated means the likelihood of the latent event sequence is not affected. Only when sampling from the latent event model, we need to concatenate a latent event corresponding to the background motif to the latent event sequences. This simplifies the conditional intensity of the observed event TPP to be:

$$\lambda\left(x \middle| \{z_k\}_{k=1}^K\right) = \text{Categorical}(n|\pi_{\text{obs}}) \cdot \sum_{k=1}^K \lambda_n\left(t \middle| \{z_k\}_{k=1}^K\right) \quad (4.2)$$

$$\lambda_n\left(t \middle| \{z_k\}_{k=1}^K\right) = \sum_{k=1}^K \lambda_n(t|z_k) \quad (4.3)$$

Note that the change in notation should now also reflect that $\lambda_n(t|z_k)$ can be chosen to be an arbitrary intensity function that can be interpreted as the intensity function of a motif instance. For example, in the special case of the background motif, we define this to be a constant intensity function to model the assumption of the background noise following a homogeneous Poisson process (HPP). In most cases, however, the intensity function of motifs should be chosen such that they are restricted to local effect around or directly after the arrival time τ_k . It is therefore recommended to design a function that is zero everywhere, except within the interval $[\tau_k, \tau_k + T_k]$, where T_k is the maximal length of the motif. To keep it simple for now, we will define $\lambda_n(t|z_k) = A_k \cdot a_{n,r_k} \cdot \mathcal{N}(t|\tau_k + b_{n,r_k}, c_{n,r_k})$ to be equal to the simple Gaussian impulse response we have already seen for PPSeq.

4.2 Variational Inference

To tackle the challenge of fitting the model to the data we draw the connection between the proposed generative model inspired by the Neyman-Scott process and the class of latent variable models (LVMs).

LVMs were developed to model complex distributions by define a two step process of generating data. First by sampling a so called latent variable $\tilde{z} \sim p_\theta(\tilde{z})$ from a simple distribution and then sampling the observed variable $\tilde{x} \sim p_\theta(\tilde{x}|\tilde{z})$. Given these two distributions and using the framework of variational inference (VI) both the model parameters θ of the prior and conditional distribution as well as the parameters ϕ of the so called variational distribution $q_\phi(\tilde{Z}) \approx p(\tilde{z}|\tilde{x})$ approximating the posterior can be found.

We point out that we can view the generative model introduced in the previous section 4.1 as an LVM by setting $\tilde{z} = \{z_k\}_{k=1}^K$ and $\tilde{x} = \{x_s\}_{s=1}^S$. The only requirement being that we can calculate the quantities $p_\theta(\{z_k\}_{k=1}^K)$ and $p_\theta(\{x_s\}_{s=1}^S|\{z_k\}_{k=1}^K)$. For reasons that will become apparent in section ?? we formulate these quantities via intensity functions according to the relation of intensity functions and likelihood of event sequences in TPPs described in eq. 2.2:

$$p_\theta(\{z_k\}_{k=1}^K) = \left(\prod_{k=1}^K \lambda(z_k) \right) \cdot \exp \left(- \int_0^T \lambda(\tau) d\tau \right) \quad (4.4)$$

$$p_\theta(\{x_s\}_{s=1}^S|\{z_k\}_{k=1}^K) = \left(\prod_{i=1}^S \lambda(x_i|\{z_k\}) \right) \cdot \exp \left(- \sum_{n=1}^N \int_0^T \lambda_n(t) dt \right) \quad (4.5)$$

Given the prior distribution (latent event model), the conditional likelihood (observed event model) and the variational distribution $q_\phi(\{z_k\}_{k=1}^K)$, which we discuss in detail later, we can use variational inference to estimate the maximum likelihood estimate for θ and ϕ . In doing so we can find the motif prototypes and motif occurrences specified through $p_\theta(\{x_s\}_{s=1}^S|\{z_k\}_{k=1}^K)$ and $q_\phi(\{z_k\}_{k=1}^K) \approx p(\{z_k\}_{k=1}^K|\{x_s\}_{s=1}^S)$ respectively. Note, because variational inference uses gradient-based optimization, the only restriction to the motif intensity functions $\lambda_n(t|z_k)$ from eq. 4.3 we impose is that it is differentiable.

Given the re-framing of the optimization problem in terms of variational inference, we also formulate the respective loss function, namely the ELBO:

$$\mathcal{L}(\theta, \phi) = \mathbb{E}_{\{z_k\}_{k=1}^K \sim q_\phi(\{z_k\}_{k=1}^K)} \left[\log p_\theta(\{x_s\}_{s=1}^S, \{z_k\}_{k=1}^K) - \log q_\phi(\{z_k\}_{k=1}^K) \right] \quad (4.6)$$

However, we still need to discuss a key ingredient for our framework to work, namely the variational distribution $q_\phi(\{z_k\}_{k=1}^K)$, which needs to fulfill the following three properties to be a feasible choice:

- It should model a marked event sequence, requiring it to be formulated as a TPP

- We need to be able to efficiently sample and evaluate the likelihood of a given latent event sequence.
- Both the sampling and likelihood evaluation must be differentiable so we can backpropagate through it.

Luckily, Shchur et al. propose a framework called TriPP that draws the connection between recent developments in the field of normalizing flows, triangular maps, and temporal point processes with which we can achieve all of the above-mentioned properties [Shc+20]. They show that all TPPs densities can be expressed via the proper choice of an increasing triangular map F , with $F = \Lambda^*(t)$. And since increasing triangular maps provide a bijective mapping, they are able to demonstrate that the inverse map F^{-1} can be used to sample from any TPP. They propose to sample from a TPP specified through such an increasing triangular map F by first sampling from an HPP and then applying the inverse map F^{-1} to each sample. As a post-processing step, they then propose to discard all sampled time points $t_i > T$. This introduces a discontinuity in the sampling, which would make the sampling non-differentiable. However, they fix this by proposing a differentiable relaxation, which assigns a smooth weighting factor to each sample, indicating whether it is within the interval $[0, T]$ or not.

Since they already propose a parameterization for an inhomogeneous point (IPP) process based on splines we use this as one possible parameterization of the variational distribution.

Experimenting with this parameterization, however, appeared to have some drawbacks, which we demonstrate in more detail in chapter 4.2. This motivates the design of an alternative parameterization of the variational distribution TPP for which we proposed to use an intensity function constructed from multiple Gaussians:

$$\lambda_r(\tau) = \sum_{k=1}^K \mathcal{N}(\tau | \mu_k, \sigma_k) \quad (4.7)$$

where τ is the arrival time of a latent event (motif), r the motif type, and μ, σ define the means and standard deviations of the K Gaussians. Framing the variational distribution in this way is particularly well suited for the task of motif discovery since it allows for an intuitive interpretation of the model parameters, where each μ_k defines the most likely time point for the k -th instance of a motif from type r and σ_k can be interpreted as the certainty or uncertainty we have in the exact starting time point of the motif.

The issue with this parameterization, however, is that the TriPP framework can not be straightforwardly applied to it, since the corresponding cumulative intensity function

$\Lambda_r(t)$ does not have a tractable inverse. This is because of the sum of Gaussians in the intensity function. To circumvent this, we propose a simple variation of the sampling strategy of TriPP. From the superposition principle of temporal point processes we know that when an intensity function is composed of sums of multiple non-negative components, each of these components defines its own valid intensity function and thereby TPP. Additionally, sampling from all of these TPPs based on the individual components and taking the union of these samples is equivalent to sampling from the original TPP in the first place. This means we can use the sampling technique of TriPP to first sample latent events for each possible motif occurrences independently via their intensity function $\lambda_{r,k}(\tau) = \mathcal{N}(\tau|\mu_k, \sigma_k)$ and then take the union of these samples to get a latent event-sequences for all possible occurrences of the motif of type r . This is possible since the inverse of the CDF from a single Gaussian is both differentiable and can be approximated efficiently.

5 Results

In the following chapter, we want to show first experimental results to demonstrate that the framework we propose is worth developing and researching further.

To do so we mostly work with synthetic data sampled from the generative model itself. The question we want to answer with these types of experiments is whether the optimization framework we proposed to simultaneously find both the motif prototypes as well as their occurrences is able to achieve this.

Additionally, we want to identify any room for improvement that exists in the proposed framework. An example of the synthetic data we generated can be seen in figure 5.1, which shows an easier sequence of observed events that exhibit only one type of motif. However, we also test our framework on synthetic data with three different motifs, each occurring three times.

Afterwards, we will also test our data on the Zebra Finch HVC data [Mac+19], the same benchmark data set Williams et al. use to compare PPSeq and SeqNMF. The Zebra Finch HVC data set contains the recordings of 75 premotor neurons in the higher vocal center of the Zebra Finch for 23 seconds in which a total of 3336 action potential spikes are recorded.

5 Results

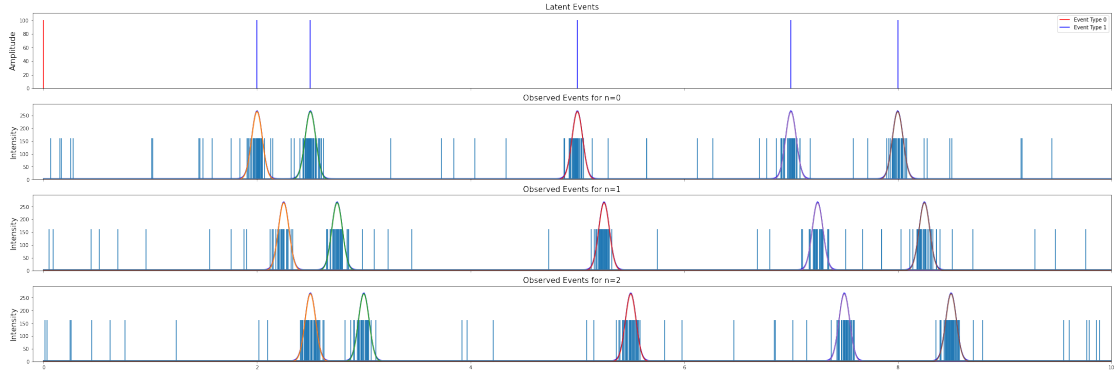


Figure 5.1: Synthetic data sampled from our generative model with one type of motif that occurs five times with some additional background noise added to the data.

To begin with, we want to illustrate the reasons for designing an alternative to the spline-based IPP which we first proposed as a possible parameterization of the variational distribution. The setting in which we first tested the spline-based IPP is the following: The parameters of the variational distribution were initialized at random while all the parameters of the observed and latent event model were equal to the ones used to generate the synthetic test data set. Then only the parameters of the spline-based variational distribution were optimized over via the ELBO from eq. 4.6 until convergence.

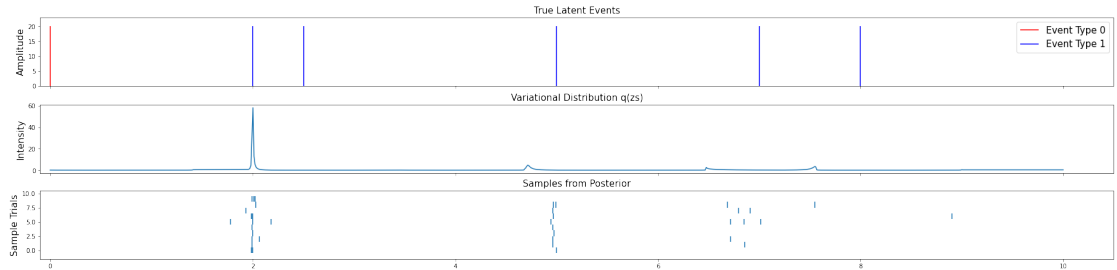


Figure 5.2: Depicted are the true latent event arrival times of the synthetic data from 5.1, as well as the intensity of the spline based IPP and 10 different trials in which we sampled from this distribution

However, as can be seen in figure 5.2, even in this simple setting the spline-based IPP was not able to capture all the latent events. Furthermore, the IPP often concentrated a lot of the intensity only on a single occurrence of a motif. Additionally, if two motif

occurrences were sufficiently close to one another, sampling from the fitted IPP would often lead to latent events distributed in between these two true motif occurrence time points. Furthermore, as can be seen with the third motif occurrence at $\tau_3 = 5$, the peak of the intensity is shifted to the left, making it harder to use the intensity function to interpret the occurrences of motifs visually. This behavior could also be reproduced when directly fitting to the true latent event arrival times via maximum likelihood estimation. We, therefore, argue that the spline-based IPP is not well suited to the task of motif discovery within our framework.

To make this even more apparent, we refer to figure 5.3. In this figure we demonstrate the results of our framework fully at work using the Gaussian-based IPP to model the variational distribution for the same synthetic data the spline-based IPP failed to fit. In addition to using the Gaussian-based IPP and in contrast to the previous experiment, in this one not just the parameters of the variational distribution are re-initialized and fitted by maximizing the ELBO, but also the parameters of the generative model are.

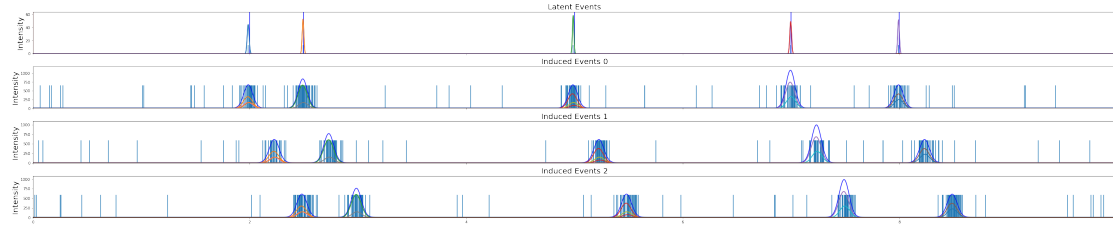


Figure 5.3: Gaussian-based variational distribution fitted to the single motif type data set

Given the perfect fit to the data, we are also able to demonstrate the assignment of every single observed event to either the background or the motif.

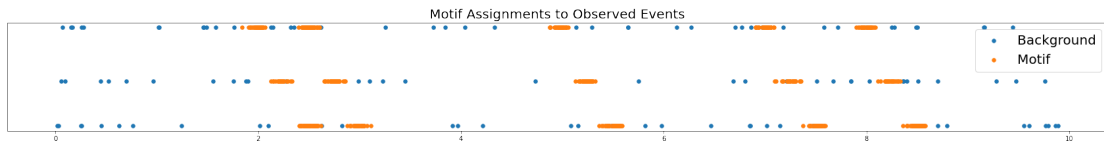


Figure 5.4: Motif assignments inferred from the Gaussian-based variational distribution fitted to the single motif type data

Going one step further we also show how our framework can find motifs in a more complex setting, where the synthetically generated data exhibits three different motifs with three occurrences each. This is depicted in figure 5.5.

However, we must point out, that our approach was not able to find a perfect fit to the true underlying model as it did in the earlier experiment. Even though the motif prototypes found closely resemble the true motif prototypes, they don't capture them perfectly. Additionally, our approach missed two motif instances completely.

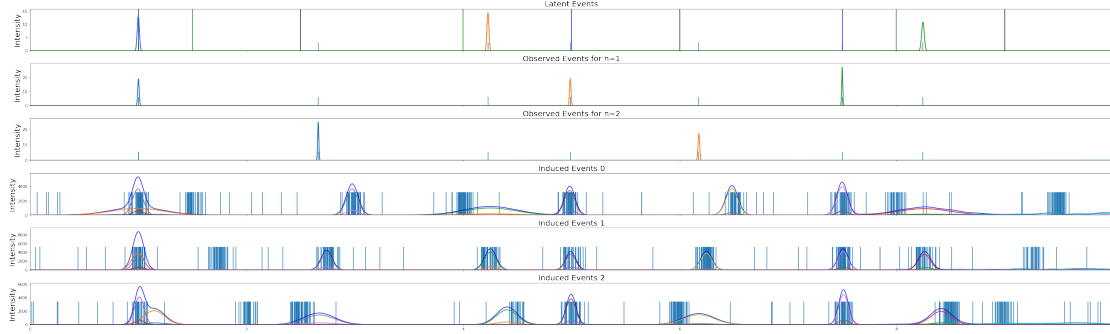


Figure 5.5: Gaussian-based variational distribution fitted to the multiple motif type data set

What can be seen from this result are the pitfalls that come from the fact that we rely on a gradient-based optimization technique, namely local minima. This can lead to a motif prototype not being properly refined, to capture the true motif prototype, because an occurrence of this motif is confidently predicted at a point that is a) far away from the true occurrence time and/or b) where it confidently captures part of this motif prototype in the data.

Finally, we present our results on the Zebra Finch HVC data set to demonstrate the degree to which our approach can find motifs in complex real-world event sequences. In figures 5.6 we mainly depict the data set itself and want to point out two things: 1) We can find various motifs via visual inspection, which is only possible because the order in which it is plotted is already informed by the motifs found with PPSeq and 2) our approach finds two motif prototypes and multiple occurrences of them that fit the motifs found via visual inspection and hence the ones PPSeq also found.

Because of the size and complexity of the data set at hand, we additionally refer to figures 5.7 and 5.8 that clearly depict the motif occurrences and motif prototypes respectively.



Figure 5.6: The Zebra Finch HVC data [Mac+19] overlayed with normalized intensities of the observed event model conditioned on motif occurrences found by our algorithm



Figure 5.7: The intensity function of the Gaussian-based variational distribution fitted via variational inference to the Zebra Finch HVC data set [Mac+19] depicting 12 motif occurrences found by our algorithm

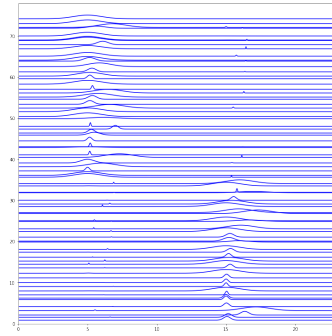


Figure 5.8: The two motif prototypes for the Zebra Finch HVC data set [Mac+19] found by our algorithm, centered around $t = 5$ and $t = 15$ respectively

6 Conclusion

We formally formulate the task of motif discovery for temporal point data and propose an algorithm based on variational inference for which we demonstrate its capability to find motif prototypes and their occurrences both in synthetically generated data as well as real-world data. With our results, we demonstrate the validity of our approach and show the potential it has.

However, our results also show that for our approach to be used in real-world settings it requires further development and research. The main research direction to be considered should be the development of a more suitable regularization to be added to the loss function to better guide the gradient-based optimization. Inspiration for this could be taken from the correlation-based penalty terms proposed by Mackevicius et al. for SeqNMF, which successfully steer their gradient-based optimization away from various kinds of errors and towards more meaningful motifs. Additionally, considering something similar to the orthogonality constraints they introduce could also provide our framework with the notions of "event-based" and "parts-based" motifs. As an alternative to a correlation-based penalty, one could also look into penalties based on entropy-centric metrics like mutual information gain.

After having solved the learning dynamics issues our framework still exhibits, the methods we presented for modeling complex and sample efficient TPPs through which we can backpropagate can similarly be applied to the modeling of more complex motifs prototypes.

As a final remark, we would explicitly like to point out the potential of our framework to be generalized to even broader use cases. In this report, we've constrained the use case to cover single-trial recordings. In practice, however, studies usually cover multiple trials on the same or different subjects. Applying a mean-field assumption, our framework can be adapted to find shared motifs across these trials and subjects.

List of Figures

5.1	Synthetic data sampled from our generative model with one type of motif that occurs five times with some additional background noise added to the data.	16
5.2	Depicted are the true latent event arrival times of the synthetic data from 5.1, as well as the intensity of the spline based IPP and 10 different trials in which we sampled from this distribution	16
5.3	Gaussian-based variational distribution fitted to the single motif type data set	17
5.4	Motif assignments inferred from the Gaussian-based variational distribution fitted to the single motif type data	17
5.5	Gaussian-based variational distribution fitted to the multiple motif type data set	18
5.6	The Zebra Finch HVC data [Mac+19] overlayed with normalized intensities of the observed event model conditioned on motif occurrences found by our algorithm	19
5.7	The intensity function of the Gaussian-based variational distribution fitted via variational inference to the Zebra Finch HVC data set [Mac+19] depicting 12 motif occurrences found by our algorithm	19
5.8	The two motif prototypes for the Zebra Finch HVC data set [Mac+19] found by our algorithm, centered around $t = 5$ and $t = 15$ respectively	19

Bibliography

- [DD07] M. K. Das and H.-K. Dai. “A survey of DNA motif finding algorithms.” eng. In: *BMC bioinformatics* 8 Suppl 7.Suppl 7 (Nov. 2007), S21. issn: 1471-2105 (Electronic). doi: 10.1186/1471-2105-8-S7-S21.
- [Her+19] A. Hertz, S. Fogel, R. Hanocka, R. Giryes, and D. Cohen-Or. “Blind visual motif removal from a single image.” In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2019-June. IEEE Computer Society, 2019, pp. 6851–6860. isbn: 978-172813293-8. doi: 10.1109/CVPR.2019.00702.
- [Mac+19] E. L. Mackevicius, A. H. Bahle, A. H. Williams, S. Gu, N. I. Denisenko, M. S. Goldman, and M. S. Fee. “Unsupervised discovery of temporal sequences in high-dimensional datasets, with applications to neuroscience.” eng. In: *eLife* 8 (Feb. 2019). issn: 2050-084X (Electronic). doi: 10.7554/eLife.38471.
- [Mue+09] A. Mueen, E. Keogh, Q. Zhu, S. Cash, and B. Westover. “Exact Discovery of Time Series Motifs.” eng. In: *Proceedings of the SIAM International Conference on Data Mining* 2009 (2009), pp. 473–484. issn: 2167-0102 (Print). doi: 10.1137/1.9781611972795.41.
- [Mue13] A. Mueen. “Enumeration of Time Series Motifs of All Lengths.” In: *2013 IEEE 13th International Conference on Data Mining*. 2013, pp. 547–556. doi: 10.1109/ICDM.2013.27.
- [Mue14] A. Mueen. “Time series motif discovery: dimensions and applications.” In: *WIREs Data Mining and Knowledge Discovery* 4.2 (2014), pp. 152–159. doi: <https://doi.org/10.1002/widm.1119>.
- [Pau+22] A. C. Paulk, Y. Kfir, A. R. Khanna, M. L. Mustroph, E. M. Trautmann, D. J. Soper, S. D. Stavisky, M. Welkenhuysen, B. Dutta, K. V. Shenoy, L. R. Hochberg, R. M. Richardson, Z. M. Williams, and S. S. Cash. “Large-scale neural recordings with single neuron resolution using Neuropixels probes in human cortex.” In: *Nature Neuroscience* 25.2 (2022), pp. 252–263. issn: 1546-1726. doi: 10.1038/s41593-021-00997-0.

- [Shc+20] O. Shchur, N. Gao, M. Biloš, and S. Günnemann. “Fast and Flexible Temporal Point Processes with Triangular Maps.” In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. NIPS’20. Red Hook, NY, USA: Curran Associates Inc., 2020. ISBN: 9781713829546.
- [Shc+21] O. Shchur, A. C. Türkmen, T. Januschowski, and S. Günnemann. “Neural Temporal Point Processes: A Review.” In: *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. Ed. by Z.-H. Zhou. International Joint Conferences on Artificial Intelligence Organization, 2021, pp. 4585–4593.
- [TL17] S. Torkamani and V. Lohweg. “Survey on time series motif discovery.” In: *WIREs Data Mining and Knowledge Discovery* 7.2 (2017), e1199. DOI: <https://doi.org/10.1002/widm.1199>.
- [Wil+20] A. H. Williams, A. Degleris, Y. Wang, and S. W. Linderman. “Point Process Models for Sequence Detection in High-Dimensional Neural Spike Trains.” In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. NIPS’20. Red Hook, NY, USA: Curran Associates Inc., 2020. ISBN: 9781713829546.
- [Xia+19] F. Xia, H. Wei, S. Yu, D. Zhang, and B. Xu. “A Survey of Measures for Network Motifs.” In: *IEEE Access* 7 (2019), pp. 106576–106587. DOI: 10.1109/ACCESS.2019.2926752.
- [YKS16] R. Yonetani, K. M. Kitani, and Y. Sato. “Visual Motif Discovery via First-Person Vision.” In: *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II*. Ed. by B. Leibe, J. Matas, N. Sebe, and M. Welling. Vol. 9906. Lecture Notes in Computer Science. Springer, 2016, pp. 187–203. DOI: 10.1007/978-3-319-46475-6_12.
- [Yu+20] S. Yu, Y. Feng, D. Zhang, H. D. Bedru, B. Xu, and F. Xia. “Motif discovery in networks: A survey.” In: *Computer Science Review* 37 (2020), p. 100267. ISSN: 1574-0137. DOI: <https://doi.org/10.1016/j.cosrev.2020.100267>.
- [YXL19] J. Yan, H. Xu, and L. Li. “Modeling and Applications for Temporal Point Processes.” In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD ’19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 3227–3228. ISBN: 9781450362016. DOI: 10.1145/3292500.3332298.