

Constructing Disease Network and Temporal Progression Model via Context-Sensitive Hawkes Process

Edward Choi, Nan Du, Robert Chen, Le Song and Jimeng Sun

School of Computational Science and Engineering

Georgia Institute of Technology

Atlanta, USA

Email: {mp2893, dunan, rchen87}@gatech.edu, {lsong, jsun}@cc.gatech.edu

Abstract—Modeling disease relationships and temporal progression are two key problems in health analytics, which have not been studied together due to data and technical challenges. Thanks to the increasing adoption of Electronic Health Records (EHR), rich patient information is being collected over time. Using EHR data as input, we propose a multivariate context-sensitive Hawkes process or *cHawk*, which simultaneously infers the disease relationship network and models temporal progression of patients. Besides learning disease network and temporal progression model, *cHawk* is able to predict when a specific patient might have other related diseases in future given the patient history, which in turn can have many potential applications in predictive health analytics, public health policy development and customized patient care. Extensive experiments on real EHR data demonstrate that *cHawk* not only can uncover meaningful disease relations and model accurate temporal progression of patients, but also has significantly better predictive performance compared to several baseline models. Scalability aspect is addressed with optimization techniques that can speed up the optimization process by more than three orders of magnitude.

I. INTRODUCTION

Applying automatic computational/statistical approaches to medical fields has attracted much attention from the communities of both academia and industry in the era of Big Data [1]. Such popularity has been spurred by the introduction of Electronic Health Records (EHR). EHR contains temporal event sequences such as admission time, discharge time, sex, ethnicity, age, weight, diagnoses, procedures, and medications. Recently, there have been a number of studies that tried to utilize such data [2]–[7] for different purposes, such as disease progression modeling [7], phenotyping [3], [5] and mortality modeling [6].

There are two key problems in health analytics that are particularly challenging, namely,

- **disease relation discovery:** What are the temporal relationships between diseases?
- **temporal progression model:** How do different diseases progress over time for each individual patient?

In order to model temporal relations among diseases for a diverse patient population, we propose context-sensitive Hawkes Process model *cHawk*. The classical Hawkes Process [8], a point process to model a finite set of temporal events, considers the relations between all past events and

the current event, namely, how past events will affect the chance of the current event happening. In our setting, Hawkes Process can capture the fact that a person who recently visited a hospital for hypertension has a higher chance of having heart failure than a person who has never suffered from hypertension. However, the classical Hawkes Process does not consider the context of each patient’s specific diagnosis. For instance, in our setting, the relation between hypertension and heart failure is applied to all people regardless of their physical differences. This is far from being realistic, since a person’s age, weight and many other factors might affect the chance of having hypertension or heart failure. It is plausible to think that heart failure is more likely to follow hypertension for an obese person than an average person. Therefore, *cHawk* identifies disease relations and temporal progression while still capturing the personal physical differences of patients using multivariate context-sensitive Hawkes process.

The contributions of the paper include the following :

- We propose *cHawk*, a context-sensitive Hawkes Process to simultaneously model disease relationship network and temporal progression using EHR data.
- *cHawk* captures both global interacting relations among diseases and how the characteristic of individual patients affect the occurrence of diseases.
- *cHawk* generates sparse and interpretable models through regularization.
- We discover clinically meaningful disease relationship network by applying *cHawk* on real EHR datasets and demonstrate accurate risk predictions for individual patients using the proposed model.
- We propose various optimization techniques that improve the training speed of Hawkes Process by more than three orders of magnitude.

The rest of the paper is organized as follows: After we survey the related work in section II, we briefly review the general Hawkes Process in section III. We then describe in section IV, our proposed Context-Sensitive Hawkes Process and parameter learning, including regularization and optimization techniques. In section V, we conduct extensive qualitative and quantitative evaluations on a real-world dataset, MIMIC II. We conclude in section VI with future research directions.

II. RELATED WORK

Disease Relationship Discovery Studies have been conducted on disease relations discovery [9]–[12] through analyzing how past diseases can affect the occurrence of current diseases. Most of the studies, however, has limitations in dealing with time dimension. Beck and Pauker [9] used Markov Decision Process to model medical prognosis, but they assume time to be discrete instead of continuous as stored in EHR. Leiva-Murillo et al. [10] applied continuous-time Hidden Markov Model (HMM) to capture disease relations. However, besides computational expensive modeling since their continuous-time HMM is a first-order model, their model only capture the influence from recent events. Savova et al. [11] used natural language processing algorithms to extract temporal relations between disease occurrences. Their work, however, uses free-text which cannot capture the exact duration between the events. A recent study by Zhao et al. [12] tries to learn the triggering kernels of the Hawkes Process in order to study the disease relations, and also proposes a metric termed ‘Individual Physique’ to represent a person’s natural fitness. The major difference to our work is that we utilize concrete features of patients that change over time (e.g. weight, age) rather than represent the natural fitness of a person as a single constant value. This will facilitate personalized medical care for patients by simply using their physical information. Moreover, our method uncovers the general latent disease network by appropriate regularizations, which produces clinically meaningful sparse structures as verified in the experiments.

Temporal Progression Model A number of recent studies tried to model temporal aspect of patients and their diseases [7], [13]–[16]. Most of the studies, however, focus on predicting the progression of a specific disease. Tangri et al. [13] uses Cox proportional hazards regression methods to predict the progression stage of chronic kidney disease to kidney failure. Ito et al. [15] and Zhou et al. [14] focuses on modeling the progression of Alzheimer’s disease, respectively using meta-analysis and biomarkers. Liu et al. [16] captured the functional and structural degeneration in the glaucoma process by using the 2-dimensional continuous-time Hidden Markov Model. Wang et al. [7] proposed a more general approach to model the progression of an arbitrary disease. They used unsupervised learning to analyze the comorbidities of chronic obstructive pulmonary disease (COPD) patients and predicts the progression of COPD. Since they model the progression of COPD through various stages along with its comorbidities, the performance of their work depends on the prior knowledge the target disease and its comorbidities.

Point Process Hawkes Process, which is the basis of our proposed method, has been attracting increased attention from sociology, and social media and network analysis. For instance, Alexey et al. [17], Mohler [18], Porter and White [19] and Zammit-Mangion [20] apply Hawkes Process to model the behaviors and conflicts among gangs and even to detect terrorist attacks; Halpin and Boeck [21], and Masuda et al. [22] use Hawkes Process to capture the interactions within e-mail networks and daily conversations; Hassen and Sharda [23], and Yang and Zha [24] construct different and mixture Hawkes Process to study the information diffusion patterns; More recently, Xu et al. [25] try to predict the online advertisement conversion rate by using mutual-exciting Hawkes Process,

and Farajtabar et al. [26] even propose to formulate the user engagement promotion as a social activity maximization problem based on multivariate Hawkes Process. Although there were many applications of Hawkes Process, however, they essentially were not designed for learning networks such as disease relations.

Network Diffusion Modeling Uncovering the relation between diseases shares similarities with network diffusion modeling, which has been actively studied recently [27]–[32]. Just as a patient experiencing multiple diseases forms a cascade, a tweet being retweeted by Twitter users forms a cascade. CONNIE [27] and NETINF [28] respectively use convex programming and submodular optimization to infer the network connectivity with fixed transmission rates. NETRATE [29] and KernelCascade [30] employ a survival analysis approach for learning probabilistic transmission rates. More recently, MoNET [31] and TopicCascade [32] respectively used the features of nodes and the features of events to infer the transmission rates. TopicCascade is similar to *cHawk* in that it assumes the topic of a meme affects its diffusion process in a network, similar to the patient characteristic affecting his/her infection likelihood. It also assumes, however, that the topic of a meme does not change over time. This assumption does not apply to patient modeling, as patient features do change over time. While there are certain similarities between *cHawk* and network diffusion models, there are two big differences: 1. Network diffusion models are mainly interested in uncovering the hidden network structure while *cHawk* performs disease relations, patient context, and risk prediction. 2. The transmission process in network diffusion models is influenced only by the most recent event. While this is a valid assumption in most network diffusion processes (e.g. a user retweeting a tweet, usually do not care how the tweet was previously retweeted to have reached him), diseases work very differently. A patient’s disease could easily be related to another disease he had before the most recent disease, which is captured by *cHawk*.

III. HAWKES PROCESS

Hawkes Process is one type of point processes for modeling temporal event sequences such as diagnosis events of a patient in EHR. The intuition behind Hawkes Process is *self-excitation*, meaning that the past occurrences of events make the future event more probable. More formally, given an observation window of event sequence $\mathcal{T} := \{t_1, \dots, t_n\}$ where t_i is the i -th occurrence of the event, the event occurrence time can be modeled as a continuous random variable X . We denote the conditional density of the next event time t_i given all the past events \mathcal{H}_{t_i} as $f(t_i|\mathcal{H}_{t_i}) := f(t_i|t_1, \dots, t_{i-1})$. Given that no events have happened since the last event up to time t , the probability that a new event will occur just within the short interval $[t, t + d\tau)$ is given by

$$\Pr(X \in [t, t + d\tau) | X > t, \mathcal{H}_t) = \frac{f(t|\mathcal{H}_t)d\tau}{S(t|\mathcal{H}_t)} = \lambda(t|\mathcal{H}_t)d\tau,$$

where $\lambda(t|\mathcal{H}_t)$ is the intensity function of the process indicating the risk that an event will occur at time t , and $S(t|\mathcal{H}_t) = 1 - F(t|\mathcal{H}_t) = \exp\left(-\int_0^t \lambda(\tau|\mathcal{H}_\tau)d\tau\right)$ is the survival function showing the probability that no events happen up to time t . For the ease of notation, we will denote $f(t) := f(t|\mathcal{H}_t)$, $\lambda(t) := \lambda(t|\mathcal{H}_t)$ and $S(t) := S(t|\mathcal{H}_t)$, which implicitly

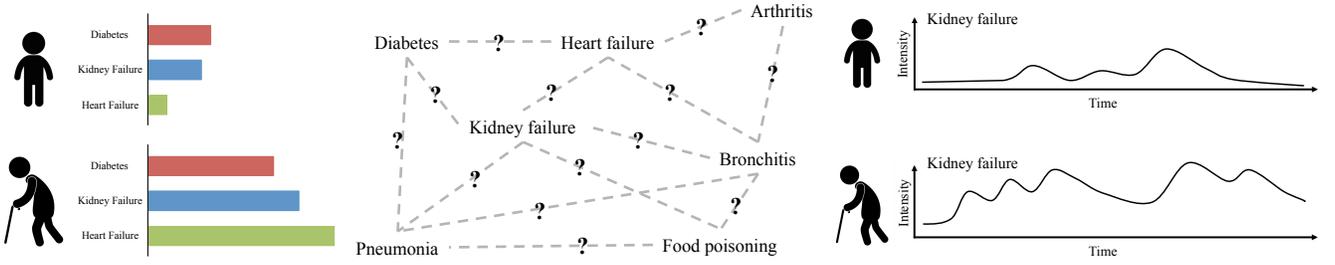


Fig. 1. From left to right, Disease risks, Disease relations, Temporal dynamics: The left figure depicts the different levels of risks of two patients being afflicted by the same set of diseases: diabetes, kidney failure and heart failure. The middle figure depicts the latent relations among diseases. The last figure shows how the strength of a disease can change over time differently for each individual.

assume the dependency of the history \mathcal{H}_t . Hawkes Process has the following relations between $f(t)$, $\lambda(t)$ and $S(t)$:

$$f(t) = \lambda(t) \exp\left(-\int_0^t \lambda(\tau) d\tau\right)$$

$$\lambda(t) = -\frac{d}{dt} \log S(t)$$

and we assume $S(0) = 1$.

By the chain rule, the joint likelihood of observing a single sequence of events $\mathcal{T} = \{t_1, \dots, t_n\}$ within the time window starting from $t = 0$ and ending at $T \geq t_n$, can be given as follows:

$$\mathcal{L}(\mathcal{T}) = \prod_{t_i \in \mathcal{T}} \lambda(t_i) \cdot \exp\left(-\int_0^T \lambda(\tau) d\tau\right), \quad (1)$$

From the joint likelihood of Equation 1, we can see that a general temporal point process can be uniquely determined by its conditional intensity function $\lambda(t)$. For one-dimensional Hawkes Process, its conditional intensity function is given by

$$\lambda(t) = \mu + \alpha \sum_{t_i < t} g(t - t_i),$$

where μ is the base intensity rate capturing the spontaneous rate to generate new events, $g(t - t_i)$ is the triggering kernel quantifying the influence of a past event on the occurrence of the new event as a function of the duration between the past and current time, and α measures the amount of influence from past events on the current event. One dimensional Hawkes Process is also called a self-exciting process. The larger the value of α , the more influence past events will have on the current event. The limitation of one-dimensional Hawkes Process is that it can only model a single type of event. In our setting, we want to capture the interacting processes of different event types (or diseases, more specifically). This is when the multi-dimensional Hawkes Process comes into play.

Multi-dimensional Hawkes Process models not only the self-excitation of a single type of event but also captures the mutual excitations among different types of events. Formally, we denote $\mathcal{T} = \{(t_i, d_i)\}_{i=1}^n$ an event sequence of time t_i associated with event type d_i . The conditional intensity function for each event type d is thus given by

$$\lambda_d(t) = \mu_d + \sum_{t_i < t} \alpha_{d, d_i} g(t - t_i) \quad (2)$$

TABLE I. POPULAR DISEASES OF TWO PATIENT GROUPS

Age 20-40, Weight 60-80kg
HIV disease
Fracture of vertebral column with spinal cord injury
Chronic liver disease and cirrhosis
Poisoning by psychotropic agents
Septicemia
Age 60-80, Weight 100-150kg
Acute myocardial infarction
Septicemia
Other forms of chronic ischemic heart disease
Cardiac dysrhythmias
Diseases of pancreas

where μ_d is the base intensity rate of event type d , and α_{d, d_i} is the strength of influence event type d_i has over event type d . Then the log-likelihood of observing \mathcal{T} is the following :

$$\ell(\mathcal{T}) = \sum_{d=1}^D \left\{ \sum_{(t_i, d_i=d) \in \mathcal{T}} \log \lambda_d(t_i) - \int_0^T \lambda_d(\tau) d\tau \right\}, \quad (3)$$

where D is the total number of diseases.

So far, at the first glance, the multi-dimensional Hawkes Process seems able to model the inter-relations among diseases. However, the direct application to our setting will incur two major issues. First, the model is constructed for each disease, which ignores the physical difference of the patients. As a result, we cannot make any patient-specific risk predictions. Second, because the number of unknown parameters $\{\alpha_{d_j, d_i}\}_{i, j}$ grows quadratically as the number of diseases increases, even on moderate size of EHR data, the model will incur huge computation cost and is often overfitted.

IV. CONTEXT-SENSITIVE HAWKES PROCESS

From EHR data, there are three modeling insights, namely, 1) patient's context-sensitive disease risk, 2) the relationship between various diseases as to how they influence the occurrence of one another, and 3) the temporal dynamics of diseases. Figure 1 illustrates the three insights in details. To capture these insights from EHR, we propose Context-sensitive Hawkes Processes *cHawk*.

A. Context-Sensitive Hawkes Process

We first denote by $\mathcal{T}^i = \{(t_j^i, d_j^i, \mathbf{f}_j^i)\}_{j=1}^n$ the sequence of clinical visiting events of patient i , where t_j^i is the time of visit j of patient i , d_j^i is the type of disease of patient i at visit j , and \mathbf{f}_j^i is the associated set of physical features of patient i at

visit j , such as age, height, weight, blood pressure, etc. Each \mathcal{T}^i is referred to as a *cascade* in the sense that for patient i , his (or her) current disease might trigger other related symptoms and diseases in the future. Overall EHR data are modeled as a collection \mathcal{C} of *i.i.d.* cascades $\{\mathcal{T}^1, \dots, \mathcal{T}^{|\mathcal{C}|}\}$, one from each patient.

The general multi-dimensional Hawkes Process assumes that the spontaneous intensity rate μ_d and the mutual-excitation rate $\alpha_{d,d'}$ are the same for all cascades. Since different patients have different sets of physical features, such as age, weight and height, it would be unrealistic to assume that the same μ_d and $\alpha_{dd'}$ can be applied to all patients. To illustrate the patient heterogeneities, Table I lists popular diseases among two different groups of patients in the MIMIC II dataset (a publicly available EHR dataset that we used in our experiments). It can be clearly seen that physical differences between young, average weight patients and old, overweight patients do play a role in the chance of disease occurrences.

To incorporate such patient contexts, we introduce a feature vector \mathbf{f}_j^i for patient i at visit j , which can be parameterized differently based on what information are available in EHR data. For example, patient features can include discrete values such as ethnicity and gender, as well as real values such as age and weight. In our experiments, for simplicity all features are converted to binary values. In particular, real value such as weight are mapped to 6 binary variables indicating from very low to very high, as can be seen from Figure 3¹. We now modify the conditional intensity function for the d th disease given patient i as follows.

$$\lambda_d^i(t) = \underbrace{\boldsymbol{\mu}_d^\top \mathbf{f}_j^i}_{\text{patient context}} + \sum_{t_j^i < t} \underbrace{\alpha_{d,d_j^i}}_{\text{disease network}} \underbrace{g(t - t_j^i)}_{\text{temporal dynamics}} \quad (4)$$

Note that what used to be $\lambda_d(t)$ in multi-dimensional Hawkes Process is now $\lambda_d^i(t)$, which is the conditional intensity function of disease type d given patient i . Essentially, we are learning intensity functions for individual patients. As shown in Equation 4, the model consists of the following three key components:

- **Patient context:** We formulate the spontaneous occurrence strength $\boldsymbol{\mu}_d$ as a linear combination of patient-specific, time-variant features \mathbf{f}_j^i . As a result, the conditional intensity function of each disease can now be adaptable to different patients, so that we can capture the heterogeneous evolving process with respect to each specific patient.
- **Disease network:** We will learn mutual excitation $\{\alpha_{d_j, d_i}\}$ variables for any pair of diseases to construct the disease relationship network. Although one might want to learn a patient specific disease network, the reality is that the available information from a single patient is often too limited to reliably learn all the parameters. As a result, we choose to learn a global disease relation network for all patients.
- **Temporal dynamics:** The triggering kernel $g(t)$ controls the aspect of temporal dynamics of diseases. Without loss of generality, in this work, we use the exponential decay kernel $g(t) = \lambda e^{-\lambda t}$, which is commonly used in many fields

for simplicity. We have also experimented with Rayleigh kernel, another widely used decay kernel. But such approach yielded a slightly inferior performance. Another option is to use nonparametric density kernels, which is more suitable for a larger dataset. We plan to explore this option in the future when dealing with a bigger dataset.

B. Parameter Estimation

By Equation 3, the log-likelihood $\ell(\mathcal{T}^i)$ of observing a single cascade $\mathcal{T}^i \in \mathcal{C}$ for patient i is given by

$$\ell(\mathcal{T}^i) = \sum_{d=1}^D \left\{ \sum_{(t_j^i, d_j^i=d, \mathbf{f}_j^i) \in \mathcal{T}^i} \left(\log \lambda_d^i(t_j^i) - \int_{t_{j-1}^i}^{t_j^i} \lambda_d^i(\tau) d\tau \right) - \int_{t_{n,d}^i}^T \lambda_d^i(\tau) d\tau \right\},$$

where $t_{n,d}^i$ is the last event on the dimension d . Then, the joint log-likelihood of observing all the cascades $\mathcal{C} = \{\mathcal{T}^1, \dots, \mathcal{T}^{|\mathcal{C}|}\}$ is simply derived as

$$\ell(\mathcal{C}|\mathbf{A}; \{\boldsymbol{\mu}_d\}_{d=1}^D) = \sum_{\mathcal{T}^i \in \mathcal{C}} \ell(\mathcal{T}^i),$$

where $\mathbf{A} = \{\alpha_{d_j, d_i}\}_{d_i, d_j}$ is the D -by- D matrix and D is the total number of diseases. A desirable characteristic of this log-likelihood function is that it is concave in the arguments \mathbf{A} and $\{\boldsymbol{\mu}_d\}$, which will allow us to find the global maximum solution efficiently using various convex optimization tools. Moreover, we want to induce a sparse network structure from the diseases and avoid overfitting. If the mutual excitation rates $\alpha_{d_j, d_i} = 0$, then there is no edge (or direct transmission) from the disease d_i to d_j . For this purpose, we impose L_1 type of regularization on the parameters $\{\alpha_{d_j, d_i}\}$ so that we can obtain a sparse disease network structure. As a consequence, the sparse disease network structure is reflected in the non-zero patterns of the final matrix \mathbf{A} . Similarly, we impose L_2 regularization on the parameters $\{\boldsymbol{\mu}_d\}$ so that we can obtain robust estimates of the parameters over patient features. Finally, we have the following optimization problem:

$$\begin{aligned} \min & \left\{ -\ell(\mathcal{C}|\mathbf{A}; \{\boldsymbol{\mu}_d\}_{d=1}^D) + \lambda_1 \|\mathbf{A}\|_1 + \frac{\lambda_2}{2} \sum_{d=1}^D \|\boldsymbol{\mu}_d\|_2^2 \right\} \\ \text{subject to} & \mathbf{A} \geq 0, \{\boldsymbol{\mu}_d\}_{d=1}^D \geq 0 \end{aligned} \quad (5)$$

After we learned the network structure with the L_1 regularization, we then refit the nonzero parameters $\{\alpha_{d_j, d_i}\}$ to achieve better estimations of those parameters without L_1 regularization.

C. Optimization

Although the optimization problem of Equation 5 has simple non-negativity constraints, all the parameters of \mathbf{A} are tangled together, which makes the direct optimization inefficient. By carefully investigating the structure of Equation 5, we can observe that the negative log-likelihood is readily separable for each dimension (or disease) d . Therefore, we can decompose the original optimization of Equation 5 into D independent convex optimization subproblems where D is the

¹We also tried using real-valued features directly, with interpolation applied between visits. But such an approach exhibited inferior performance.

total number of diseases. Then, given a particular disease d , the objective function Equation 5 can be evaluated as

$$-\sum_{\mathcal{T}^i \in \mathcal{C}} \left\{ \sum_{(t_j^i, d_j^i=d, \mathbf{f}_j^i) \in \mathcal{T}^i} \left(\log(\lambda_d^i(t_j^i)) - \int_{t_{j-1}^i}^{t_j^i} \lambda_d^i(\tau) d\tau \right) - \int_{t_{n,d}^i}^T \lambda_d^i(\tau) d\tau \right\} + \lambda_1 \|\mathbf{A}_d\|_1 + \frac{\lambda_2}{2} \|\boldsymbol{\mu}_d\|_2^2. \quad (6)$$

where $t_{n,d}^i$ is the last occurrence time of disease type d in i th cascade. The respective gradients of $\ell(\mathcal{C}|\mathbf{A}_d; \boldsymbol{\mu}_d)$ with respect to α_{d,d_k} and $\boldsymbol{\mu}_d$ can be derived as

$$\begin{aligned} \frac{\partial \ell(\mathcal{C}|\mathbf{A}_d; \boldsymbol{\mu}_d)}{\partial \alpha_{d,d_k}} &= - \sum_{\mathcal{T}^i \in \mathcal{C}} \left\{ \sum_{(t_j^i, d_j^i=d, \mathbf{f}_j^i) \in \mathcal{T}^i} \left(\frac{\sum_{t_k^i < t_j^i} g(t_j^i - t_k^i)}{\lambda_d^i(t_j^i)} - G(T - t_j^i) \right) \right\} + \lambda_1. \\ \frac{\partial \ell(\mathcal{C}|\mathbf{A}_d; \boldsymbol{\mu}_d)}{\partial \boldsymbol{\mu}_d} &= - \sum_{\mathcal{T}^i \in \mathcal{C}} \left\{ \sum_{(t_j^i, d_j^i=d, \mathbf{f}_j^i) \in \mathcal{T}^i} \left(\frac{\mathbf{f}_j^i}{\lambda_d^i(t_j^i)} - \boldsymbol{\mu}_d^\top \mathbf{f}_j^i (t_j^i - t_{j-1}^i) \right) - \boldsymbol{\mu}_d^\top \mathbf{f}_{n,d}^i (T - t_{n,d}^i) \right\} + \lambda_2 \boldsymbol{\mu}_d. \end{aligned} \quad (7)$$

where $G(t) = 1 - e^{-\lambda t}$ is the integral of the decay kernel $g(t)$. Therefore, we can optimize all these independent convex subproblems associated with each dimension d in parallel with projected gradient descent (PGD) (or more advanced projected-quasi-newton method).

D. Speed-up

In addition to decomposing Equation 5, we also used OpenMP² to utilize the power of our multi-core processor. The log-likelihood of Equation 6 can be separately calculated for each patient, and the gradient for each dimension can also be also calculated in a parallel fashion.

For additional speed-up, we profiled the optimization process to find the bottleneck. Over 95% of the time was spent calculating the intensity functions Equation 4 in the likelihood function Equation 6 and the gradient function Equation 7. It is worth noting that, at each optimization step, the values of the intensity functions in the likelihood function and the gradient function are the same. This led us to calculate the intensity functions once and share the results. We also linearized the iterative structure of the intensity function. As can be seen from Equation 4, all past events need to be processed. However, by unfolding the iteration and summing the decay kernel values of the same α variable, we can express the intensity function as follows:

$$\lambda_d^i(t) = \boldsymbol{\mu}_d^\top \mathbf{f}_j^i + \sum_{d_t \in \mathcal{D}_t} \alpha_{d,d_t} g(\cdot)$$

where \mathcal{D}_t is the set of all distinct diseases that occurred before t that influence disease d , and $g(\cdot)$ is the corresponding aggregate decay kernel³. Then we calculate the coefficients

Algorithm 1: Learning *cHawk* Model

```

1 for  $d = 1$  to  $D$  in Parallel do
2   Initialize  $\mathbf{A}_d$ : and  $\boldsymbol{\mu}_d$  randomly;
3   Pre-calculate the decay kernel and the integral*;
4   repeat
5     Pre-calculate all intensity functions*;
6     Project  $\mathbf{A}_d$ : onto  $\mathbf{A}_d \geq 0$  and  $\boldsymbol{\mu}_d \geq 0$ ;
7     Evaluate the gradients* (Eq.7);
8     Update  $\mathbf{A}_d$ : and  $\boldsymbol{\mu}_d$  using the gradients;
9     Evaluate the objective function* (Eq.6);
10  until Change in the objective function  $< \epsilon$ ;
11 end

```

$g(\cdot)$ at the beginning of the optimization and reuse them, as their values do not change during the optimization process. This significantly improves the optimization speed because a typical patient experiences limited kinds of diseases even if he/she makes many clinical visits. This strategy becomes more effective as the number of events in a cascade increases. The same strategy is also applied to the integral decay kernels in the likelihood function and the α gradient function. This pre-calculation requires on average 15 times additional memory space, but reduces the optimization time by an order of magnitude as will be shown in section V.E. Algorithm 1 summarizes the key steps. * indicates that OpenMP was used to parallelize the calculations.

V. EXPERIMENTS

A. Dataset

Our experiments used the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC II) clinical database [33]. MIMIC II is a collection of de-identified clinical visit records of Intensive Care Unit patients between 2001 and 2008 from a single tertiary teaching hospital. Although private information of the patients such as birth dates are de-identified, the duration between each patient's birth and his/her clinical visits are preserved, so that we can apply point processing algorithms. At each visit, a patient is diagnosed with the ICD-9 code system. A patient could receive more than one diagnosis at a single visit, one of which is assigned as the primary diagnosis. MIMIC II also includes information regarding the patient such as gender, birth date, weight, medication and various lab test results.

We construct the appropriate patient cohort for our experiments. We first filter out patients who have visited the hospital less than two times. As we are interested in how diseases affect one another over time, we need patients who have visited the hospital at least twice. Then we converted 5-digit ICD9 code to 3-digit ICD9 code. A typical ICD9 code consists of three primary digits and two supplementary digits (e.g. 493.01 is *extrinsic asthma with status asthmaticus*, while 493 is *asthma*). Such conversion reduces the number of dimensions from 14,000 to 1,000, yet still captures reasonably detailed diagnostic information. We further filter out non-primary diagnoses at each visit. Even though a patient is given multiple diagnoses at the visit, there must be a sequence in which those diseases occurred. EHR, however fails to capture that information, which makes it improbable in the

²<http://openmp.org>

³For example, even if a patient made multiple clinical visits for the same disease, \mathcal{D}_t would still have only one element.

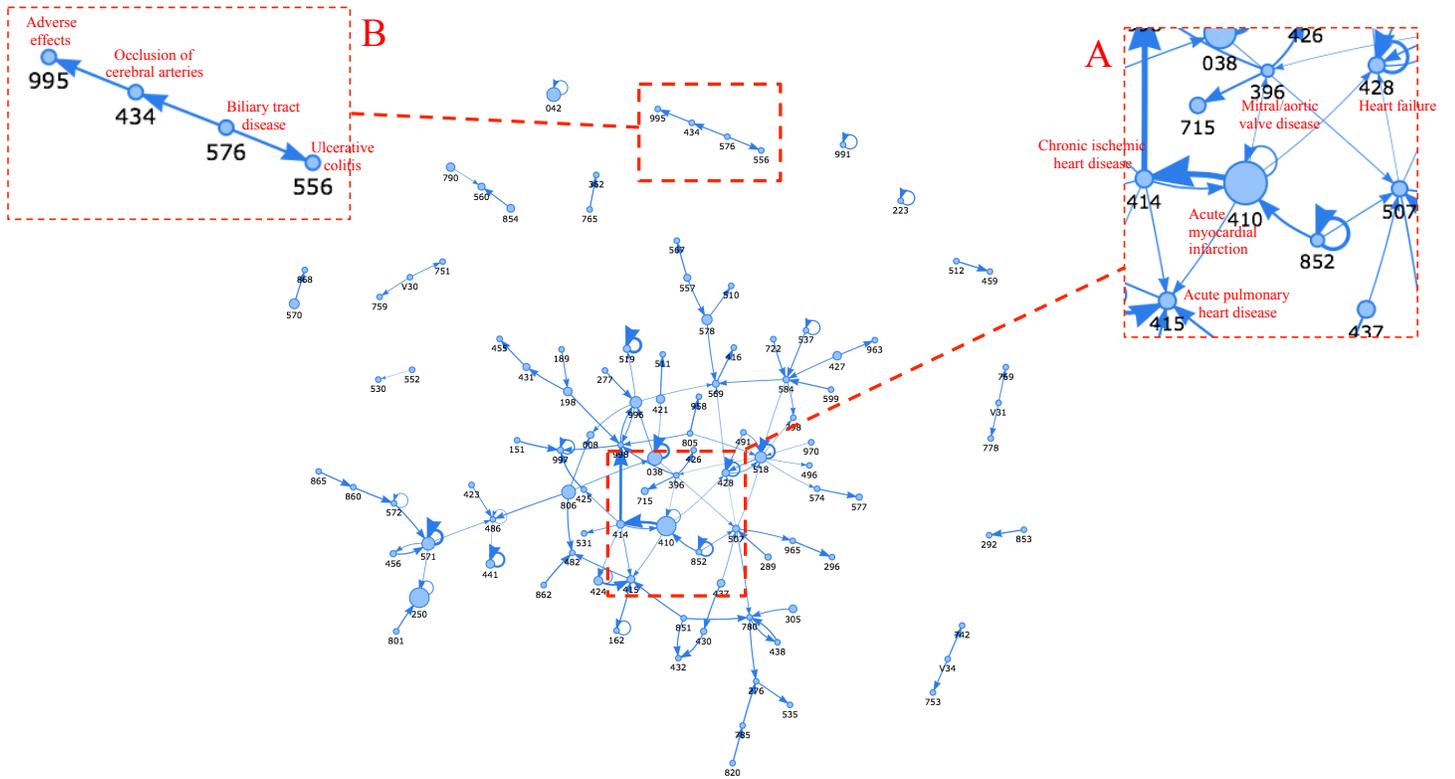


Fig. 2. Disease network of a 65kg, 25-year-old, built with *cHawk*. Each node represents a type of disease, under which the number is its ICD9 code. The size of each node represents the strength of its spontaneous occurrence (for a person of age 25, weight 65kg). Edges between the nodes represent the direction of influence. Thicker edges mean stronger influence. Diseases that have no connection or very weak connections with other diseases were filtered out in the generation process for succinct representation of information

perspective of point processes. We plan to address this issue of simultaneous diagnoses in the future. Finally, we only take the patients who had their weight checked on every visit, so that we can apply *cHawk*. Weight was chosen among other features because it is known to affect many diseases and is the least sparse physical feature in MIMIC II. Finally we are left with 593 patients, 186 disease types and total 1,269 visits. We chose weeks as the unit of time for convenience, and all 593 cascades are normalized so that the first clinical visit occurs at time $t = 0$. The observation length T for each cascade is set to the time of the last event t_n . The features we used, age and weight, are shown in the axes of Figure 3. We tried to include more features such as blood pressure. However, due to the sparse nature of EHR data, weight and age were the most suitable candidates.

B. Hyper Parameter Setting

cHawk uses three hyper parameters: Regularization parameters λ_1, λ_2 in Equation 5 and the exponential decay kernel parameter λ in $g(t) = \lambda e^{-\lambda t}$. For λ_1 and λ_2 we tested values of 0, 10, 100 and 1000. For the exponential decay kernel parameter we tested 0.2, 0.4, 0.6, 0.8, 1.0. After iterations of rigorous experiments using a machine equipped with an Intel Xeon E5-2630 (24 cores) and 132GB memory, we chose $\lambda_1 = \lambda_2 = 10$ and 0.2 for the decay kernel parameter. The criteria for choosing the optimal value was the model’s predictive performance, which will be discussed in section V.D.

C. Disease Relation, Context Sensitivity and Temporal Dynamics

In this section, we present the disease network to show the relations between diseases, explain how change of context (patient features) affects disease occurrence, and also show how occurrence intensities of diseases change over time.

1) *Disease Network*: Figure 2 is the disease network constructed using *cHawk*, specifically for a patient of age 25, weight 65kg. The qualitative interpretation of the networks was provided by a current medical student who also has broad experience in medical data mining. The network was confirmed to be clinically meaningful with connections that represent possible real-world scenarios where certain diseases may precede others. In Figure 2, we provided two specific examples where relations between nodes are explained.

In Figure 2, subgraph A, the largest node (410) corresponds to acute myocardial infarction. Most of the nodes that this one leads to are considered possible events that may sequentially happen after a real-life acute myocardial infarction event. The edge with the largest influence connects *acute myocardial infarction* to other forms of *chronic ischemic heart disease*, which is a relationship that is commonly seen in real life. Furthermore, the node for *acute myocardial infarction* has an edge that points to itself. This is also clinically significant because some patients may experience successive episodes of re-infarction after the first myocardial infarction event [34]. The *heart failure* node connected to *acute myocardial*

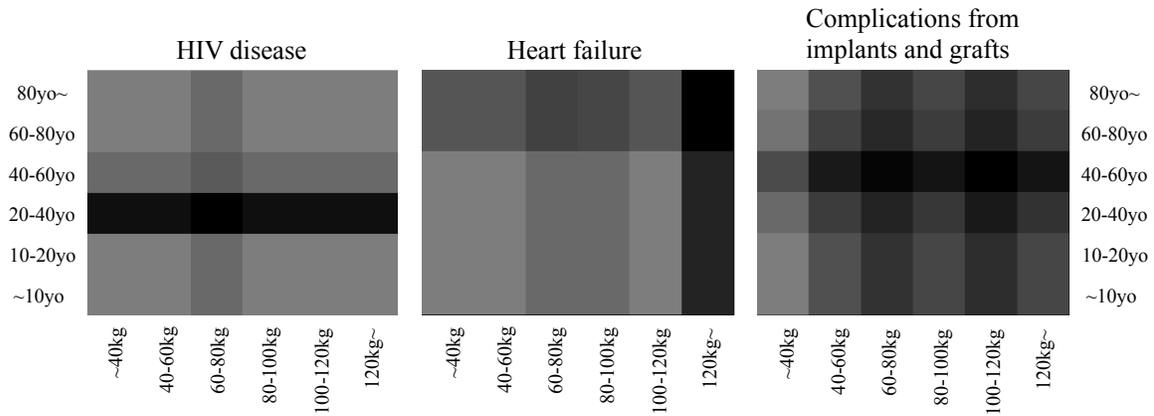


Fig. 3. Heat map of three different diseases: HIV, heart failure, complications from implants and grafts. The Y-axis corresponds to ages: < 10, [10, 20), [20, 40), [40, 60), [60, 80), and > 80. The X-axis consists of weights: <40kg, [40kg, 60kg), [60kg, 80kg), [80kg, 100kg), [100kg,120kg), and >120kg. The darker regions represent stronger activity of the disease.

infarction represents the fact that heart failure may follow myocardial infarction in about 29% of patients [35]. The *acute pulmonary heart disease* node is connected to the *acute myocardial infarction* node, which is clinically meaningful, as cardiogenic pulmonary edema can occur as a complication of systolic heart failure, due to pulmonary capillary pressure increases caused by an impaired ability of the left ventricle to pump blood to systemic circulation [36, Chapter 11]. The *diseases of the mitral and aortic valves* node may represent complications of myocardial infarction such as mitral valve prolapse [36, Chapter 11]. While the separate nodes connected to *acute myocardial infarction* are all clinically meaningful, they all collectively represent an important, interconnected network that effectively captures the highly complex nature of patients in severe cardiovascular condition.

Another set of disease relations worth mentioning is shown in Figure 2, subgraph B. The root of this subgraph is *biliary tract disease*. One node that connects from this node is *ulcerative colitis*. This relationship is a fundamentally important one because there is a known medical association between primary sclerosing cholangitis (which is a type of biliary disease) and ulcerative colitis [37]. About 80% of patients with primary sclerosing cholangitis (PSC) have inflammatory bowel disease including ulcerative colitis (UC). There is vast evidence in the medical literature for shared etiology for PSC and UC. Most notably, genetic markers in the HLA class II genes confer risk to both diseases [38]. The *biliary disease* node also points to *occlusion of cerebral arteries*. One explanation for this is that both biliary disease (such as gallstones or cholangitis) and occlusion of cerebral arteries (a possible result of thromboembolism) may result due to side effects of oral contraceptives such as drospirenone/ethinylestradiol [39], [40] [36, Chapter 19]. The node for occlusion of cerebral arteries points to the node *certain adverse effects not classified elsewhere*. This is an umbrella term that could refer to a number of adverse effects such as side effects of diseases. This particular node makes sense to be linked to cerebral artery occlusion, because ischemic blood loss that results from the occlusion may result in a number of musculoskeletal, spatiovisual, or cognitive deficits.

2) *Context-Sensitivity of Diseases*: As we have optimized μ vectors of different disease, we can now plug in specific age and weight to analyze how diseases behave under different contexts. Figure 3 is the heat map of three different diseases, *Human immunodeficiency virus disease*(042), *Heart failure*(428) and *complications from implants and grafts*(996) calculated using a range of values for ages and weights. The heat maps are based on μ variables of Equation (4).

It can be seen that diseases behave quite differently in different contexts. For example, HIV disease acts especially strongly for people between the age of 20 and 40, which in fact correctly reflects reality⁴. This characteristic could be due to the fact that younger people are more sexually active, rendering them more susceptible to sexually transmitted diseases than older people. It is interesting that people who have average weight are more vulnerable to HIV disease. This could be interpreted that people who are physically healthy are more likely to be exposed to sexual activity than people who are not. *Heart failure*, on the other hand, is heavily affected by a person's weight. It is also correlated with higher age, but the correlation is stronger with obesity. This is consistent with the common knowledge that obese people are more susceptible to heart diseases⁵. *Complications from implants and grafts* seem to be ubiquitous to various groups of people except for patients with a very light weight. This is equivalent to saying that this disease can generally occur to anyone except for very young children who weigh under 40kg. This makes sense as young children are less likely to receive implants or grafts such as cardiac devices, vascular devices, prosthetic joints or organ transplants.

3) *Temporal Dynamics*: Here we present two groups of disease intensities, one for a young, average weight person (age 21, weight 67kg), and another for an old, overweight person (age 71, 122kg). Each person suffers different sets of diseases throughout 4 months as shown in Figure 4. The younger, average weight patient has suffered *septicemia*(038), *other diseases of lung*(518) and *heart failure*(428). The older, overweight patient has suffered *heart failure*(428), *complications from implants and grafts*(996) and *other forms of chronic*

⁴<http://www.cdc.gov/hiv/risk/age/olderamericans/>

⁵<http://www.cdc.gov/healthyweight/effects/>

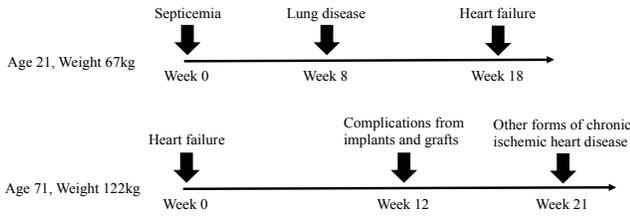


Fig. 4. Disease history of two people with different physical features

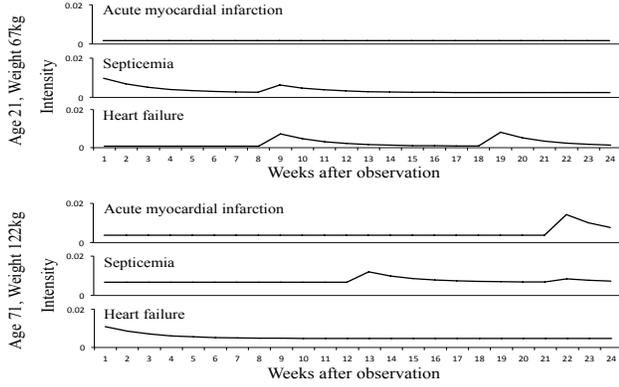


Fig. 5. Disease intensities of two people with different physical features. We can see that intensity trajectories differ for different patients even though the diseases being plotted are the same.

ischemic heart disease(414). Based on these disease records, we plotted the intensities of three well-known common diseases for each patient using Equation 4.

It can be seen from Figure 5 that the intensity of a disease not only depends on what types of diseases the patient has during the observation and how much time has passed after each disease, but also on the physical features of the patient. For example, the intensity of acute myocardial infarction for the younger patient maintains its default strength since the diseases he suffered have little influence on acute myocardial infarction. The intensity of septicemia, on the other hand, spikes after he suffers septicemia and lung disease, as they both influence the occurrence of septicemia. You can see, however, that after a period of time, the intensity of septicemia drops down to its default strength. The intensity of heart failure maintains very weak default strength except for the two times the patient suffers lung disease and heart failure. Now when we study the intensity of heart failure for the older, overweight patient, after being influenced by the initial instance of heart disease, the intensity maintains relatively higher strength throughout the whole observation compared to the younger, average weight patient. This is due to the difference of physical features of the two patients. A higher heart failure intensity for the older, overweight patient is consistent with what we have presented in section V.C.2, where we have shown that heart failure acts more strongly when combined with obesity.

D. Disease Prediction

Next we provide quantitative evaluation of *cHawk* by performing disease risk prediction. Given a disease history of a patient, we try to predict the most likely disease he/she will have in a certain future time window by calculating the

conditional cumulative distribution of each disease.

We first analyze the influence of the patient history length. We divided the patients into two groups: patients with less than 3 visits and patient with at least 3 visits. Then we tried predicting the disease that occurs in the next three months window by picking out 10 diseases with the highest conditional cumulative probability. If one of them was correctly predicted, we consider it an accurate prediction. We performed 10-fold cross validation with *cHawk* to obtain accuracy 0.554375 for the former group and 0.649206 for the latter group. Clearly, it is easier to predict future diseases when provided with longer disease history.

Figure 6 is the result of 10-fold cross validation of risk prediction by various methods including Hawkes Process and other well-known methods. We tried predicting diseases that occurred in four different three months windows, while varying the number of predictions p . If one of p predictions is correct, we consider it an accurate prediction. We focus on positive examples for two reasons: 1. Negative instances are difficult to deal with since absence of visit does not mean diseases didn't occur. 2. False negatives are far more critical in risk prediction than false positives. For *cHawk* and Hawkes Process, we choose p diseases with the highest conditional cumulative probability.

Poisson Process is a stochastic process which does not consider the influence of past events, as it assumes the number of events occurring in a given time interval follows Poisson distribution. In this experiment, we used homogeneous Poisson Process which is equivalent to removing the α variables from the intensity function (2) of Hawkes Process. For Poisson Process, we also use conditional cumulative probability for prediction.

Linear regression models were trained to predict the time of the next occurrence of the disease, given all past diseases and the most recent age, weight information. All past diseases were aggregated into a 186-dimension vector, with possible multiple entries being 1. Due to the sparsity of our data, we were able to train models for approximately 120 diseases out of 186. Given a test sample, we ran 120 models to predict the onset time of each disease, and we picked all diseases that occurred in the target time window. The average number of the selected diseases for the different time windows are 48, 14, 11 and 11. This is the reason we put a single dot for linear regression, as a type of baseline.

For multinomial logistic regression, we used the same features as linear regression. Unlike other methods logistic regression calculates the conditional probability $p(y|x)$ of the next disease being y given patient records x . This makes logistic regression simple and powerful, but it can only predict what the next disease will be, losing all time-related information. Again, p diseases with the highest probability were chosen. Although logistic regression cannot be directly compared with other temporal models, we plot its performance in the first figure so that it can serve as a reference.

It is readily visible from Figure 6, that *cHawk* outperforms other well-known methods except for the first three month window. In that particular window, *cHawk* was having trouble correctly predicting diseases concerned with newborn infants, especially *Other and ill-defined conditions originating in the*

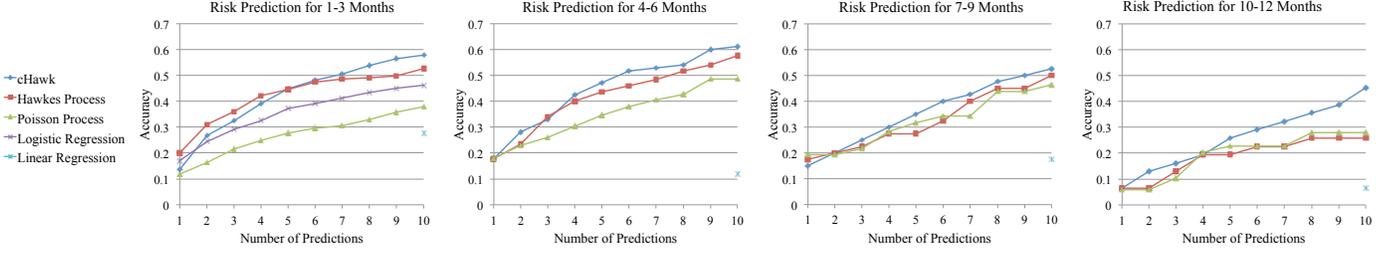


Fig. 6. Prediction performance comparison for different time windows. We used $\lambda_1 = 10, \lambda_2 = 10$ for *cHawk* and Hawkes Process. Regularization was ineffective for Poisson Process. For the exponential decay parameter, we used 0.2 for *cHawk*, Hawkes Process and Poisson Process. We used Python Scikit-Learn for Logistic Regression and Linear Regression.

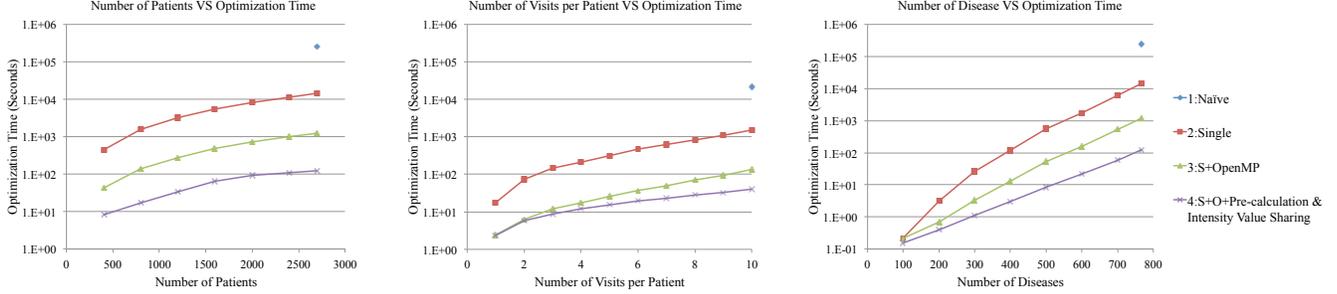


Fig. 7. Log-scaled optimization time for different number of patients, visits per patient, and diseases. No regularization was used for μ and α . Exponential decay parameter was set to 0.5. ϵ in Algorithm 1 was set to 10^{-7} . L-BFGS provided by Dlib C++ library was used.

perinatal period(779). *cHawk*, however, was predicting *Disorders relating to short gestation and low birthweight*(765) instead, which is very similar to 779. In all other situations, *cHawk* exhibits superior performance. Also it is worth noting that *cHawk* is particularly robust in making a longer prediction, which can be attributed to its taking into consideration patient feature. We can also see that after the first six months, the predictive power starts to decrease. This corresponds to our intuition that making a longer prediction is naturally more difficult.

E. Scalability

In this section, we address the scalability issue by describing the relation between the size of the data and the optimization time, and how additional methods can help speed up the optimization process.

Figure 7 depicts the log-scaled optimization time for different numbers of patients, individual visits and diseases respectively⁶. We tested four different implementations: naive implementation, dimension decomposition, dimension decomposition with OpenMP, and the combination of all techniques mentioned in section IV.C and IV.D. For the naive implementation, we only measured the most extreme cases due to its slow speed. Theoretically, optimization of Hawkes Process takes $O(d^2v^2n)$ where d is the number of diseases, v the number of individual visits, and n the number of patients. In reality, however, optimization is heavily influenced by the distribution of diseases among patients, the number of occurrences of each disease, and the distribution of number of visits among patients. Also, the number of patients, diseases and visits are

all tied to one another(*e.g.* If you increase number of patients, the number of diseases also increases).

In the leftmost figure, we can check the significantly improved speed of the most optimized fourth implementation compared to other implementations, especially, the naive implementation. The benefit of the pre-calculation technique is readily visible in the center figure. As the number of visits increases, the gap between the third and fourth implementation also increases. The rightmost figure displays the relation between the number of diseases and the optimization time. The near-exponential increase in optimization time is due to the fact that the number of patients explosively increases when the number of disease increases. The most realistic behavior is captured by the leftmost figure, since the number of patients is the most frequently used measure to represent the size of EHR. In all three figures, it can be seen that the proposed optimization techniques reduce the optimization time by more than three orders of magnitude compared to the naive implementation.

Finally, as we ran optimization on only a single machine, there is room for further speed-up if more machines could be harnessed.

VI. CONCLUSIONS

In this paper we proposed *cHawk* to capture the three aspects of EHR, namely disease relations, context-sensitivity of diseases and temporal dynamics of diseases. We showed a detailed derivation of the model and also presented an optimization algorithm. In the experiments, we applied *cHawk* to MIMIC II, a real-world EHR comprised of ICU patients, to build disease networks, context-sensitive heat maps and show

⁶We used unfiltered MIMIC2 to display the scalability of our method

temporal dynamics of disease intensities. Risk predictions were performed for quantitative evaluation, which showed that *cHawk* is able to predict future diseases more accurately than the original Hawkes Process and other traditional methods. We also addressed the scalability issue by presenting various optimization methods and experimenting on data with varying number of patients, visits, and diseases.

In the future, we plan to apply *cHawk* to several EHR datasets of varying patient demographics and diverse sets of diagnoses. Such application will uncover different sets of meaningful disease relations. We also plan to use a larger dataset with numerous patient features, so that we can identify useful relations between various patient features and the risk of disease occurrence.

REFERENCES

- [1] L. Ohno-Machado, "Nih's big data to knowledge initiative and the advancement of biomedical informatics," *Journal of the American Medical Informatics Association*, vol. 21, no. 2, pp. 193–193, 2014.
- [2] F. Wang, P. Zhang, B. Qian, X. Wang, and I. Davidson, "Clinical risk prediction with multilinear sparse logistic regression," in *KDD*, 2014, pp. 145–154.
- [3] J. Zhou, F. Wang, J. Hu, and J. Ye, "From micro to macro: Data driven phenotyping by densification of longitudinal electronic medical records," in *KDD*, 2014, pp. 135–144.
- [4] Y. Park and J. Ghosh, "Ludia: An aggregate-constrained low-rank reconstruction algorithm to leverage publicly released health data," in *KDD*, 2014, pp. 55–64.
- [5] J. C. Ho, J. Ghosh, and J. Sun, "Marble: High-throughput phenotyping from electronic health records via sparse nonnegative tensor factorization," in *KDD*, 2014, pp. 115–124.
- [6] M. Ghassemi, T. Naumann, F. Doshi-Velez, N. Brimmer, R. Joshi, A. Rumshisky, and P. Szolovits, "Unfolding physiological state: Mortality modelling in intensive care units," in *KDD*, 2014, pp. 75–84.
- [7] X. Wang, D. Sontag, and F. Wang, "Unsupervised learning of disease progression models," in *KDD*, 2014, pp. 85–94.
- [8] A. G. Hawkes and D. Oakes, "A cluster process representation of a self-exciting process," *Journal of Applied Probability*, pp. 493–503, 1974.
- [9] J. R. Beck and S. G. Pauker, "The markov process in medical prognosis," *Medical Decision Making*, vol. 3, no. 4, pp. 419–458, 1983.
- [10] J. Leiva-Murillo, A. Artés-Rodríguez, and E. Baca-García, "Visualization and prediction of disease interactions with continuous-time hidden markov models," in *NIPS 2011 Workshop on Personalized Medicine*, 2011.
- [11] G. Savova, S. Bethard, W. Styler, J. Martin, M. Palmer, J. Masanz, and W. Ward, "Towards temporal relation discovery from the clinical narrative," in *AMIA*, 2009, pp. 568–572.
- [12] Y. Zhao, X. Qi, Z. Liu, Y. Zhang, and T. Zheng, "Mining medical records with a klipi multi-dimensional hawkes model," in *KDD 2014 Workshop on Health Informatics*, 2014.
- [13] N. Tangri, L. A. Stevens, J. Griffith, H. Tighiouart, O. Djurdjev, D. Naimark, A. Levin, and A. S. Levey, "A predictive model for progression of chronic kidney disease to kidney failure," *Jama*, vol. 305, no. 15, pp. 1553–1559, 2011.
- [14] J. Zhou, J. Liu, V. A. Narayan, and J. Ye, "Modeling disease progression via fused sparse group lasso," in *KDD*, 2012, pp. 1095–1103.
- [15] K. Ito, S. Ahadiéh, B. Corrigan, J. French, T. Fullerton, T. Tensfeldt, A. D. W. Group *et al.*, "Disease progression meta-analysis model in alzheimer's disease," *Alzheimer's & Dementia*, vol. 6, no. 1, pp. 39–53, 2010.
- [16] Y.-Y. Liu, H. Ishikawa, M. Chen, G. Wollstein, J. S. Schuman, and J. M. Rehg, "Longitudinal modeling of glaucoma progression using 2-dimensional continuous-time hidden markov model," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*, 2013, pp. 444–451.
- [17] A. Stomakhin, M. B. Short, and A. L. Bertozzi, "Reconstruction of missing data in social networks based on temporal patterns of interactions," *Inverse Problems*, vol. 27, 2011.
- [18] M. G., "Modeling and estimation of multi-source clustering in crime and security data," *The Annals of Applied Statistics*, vol. 7, pp. 1525–1539, 2013.
- [19] P. M.D. and W. G., "Self-exciting hurdle models for terrorist activity," *Annals of Applied Statistics*, vol. 6, pp. 106–124, 2012.
- [20] A. Zammit-Mangion, M. Dewar, V. Kadiramanathan, and G. Sanguinetti, "Point process modelling of the afghan war diary," *PNAS*, vol. 109, pp. 12414–12419, 2012.
- [21] P. F. Halpin and P. D. Boeck, "Modelling dyadic interaction with hawkes processes," *Psychometrika*, vol. 78, pp. 793–814, 2013.
- [22] M. N., T. T., S. N., and Y. K., "Self-exciting point process modeling of conversation event sequences," *Temporal Networks*, pp. 245–264, 2013.
- [23] H. Z. A. and S. R., "Modeling brand post popularity dynamics in online social networks," *Decision Support Systems*, vol. 65, pp. 59–68, 2014.
- [24] Y. S.H. and Z. H., "Mixture of mutually exciting processes for viral diffusion," in *ICML*, 2013, pp. 1–9.
- [25] X. L., D. J.A., and W. A.B., "Path to purchase: A mutually exciting point process model for online advertising and conversion," *Management Science*, vol. 60, pp. 1392–1412, 2014.
- [26] M. Farajtabar, N. Du, M. Gomez-Rodriguez, I. Valera, H. Zha, and L. Song, "Shaping social activity by incentivizing users," in *NIPS*, 2014, pp. 2474–2482.
- [27] S. Myers and J. Leskovec, "On the convexity of latent social network inference," in *NIPS*, 2010, pp. 1741–1749.
- [28] M. Gomez Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," in *KDD*. ACM, 2010, pp. 1019–1028.
- [29] M. Gomez Rodriguez, D. Balduzzi, B. Schölkopf, G. T. Scheffer *et al.*, "Uncovering the temporal dynamics of diffusion networks," in *ICML*, 2011, pp. 561–568.
- [30] N. Du, L. Song, M. Yuan, and A. J. Smola, "Learning networks of heterogeneous influence," in *NIPS*, 2012, pp. 2789–2797.
- [31] L. Wang, S. Ermon, and J. E. Hopcroft, "Feature-enhanced probabilistic models for diffusion network inference," in *ECML PKDD*, 2012, pp. 499–514.
- [32] N. Du, L. Song, H. Woo, and H. Zha, "Uncover topic-sensitive information diffusion networks," in *AISTATS*, 2013, pp. 229–237.
- [33] M. Saeed *et al.*, "Multiparameter intelligent monitoring in intensive care II (MIMIC-II): a public-access intensive care unit database," *Critical care medicine*, vol. 39, no. 5, p. 952, 2011.
- [34] S. J. Kernis, K. J. Harjai, G. W. Stone, L. L. Grines, J. A. Boura, M. W. Yerkey, W. O'Neill, and C. L. Grines, "The incidence, predictors, and outcomes of early reinfarction after primary angioplasty for acute myocardial infarction," *Journal of the American College of Cardiology*, vol. 42, no. 7, pp. 1173–1177, 2003.
- [35] P. S. Jhund and J. J. McMurray, "Heart failure after acute myocardial infarction a lost battle in the war on heart failure?" *Circulation*, vol. 118, no. 20, pp. 2019–2021, 2008.
- [36] E. F. Goljan, *Rapid Review Pathology*, 12nd ed. Elsevier Saunders, 2014.
- [37] R. Chapman, Z. Varghese, R. Gaul, G. Patel, N. Kokinon, and S. Sherlock, "Association of primary sclerosing cholangitis with hla-b8," *Gut*, vol. 24, no. 1, pp. 38–41, 1983.
- [38] T. Karlsen, K. Boberg, M. Vatn, A. Bergquist, J. Hampe, E. Schruppf, E. Thorsby, S. Schreiber, and B. Lie, "Different hla class ii associations in ulcerative colitis patients with and without primary sclerosing cholangitis," *Genes and immunity*, vol. 8, no. 3, pp. 275–278, 2007.
- [39] N. Gronich, I. Lavi, and G. Rennert, "Higher risk of venous thrombosis associated with drospirenone-containing oral contraceptives: a population-based cohort study," *Canadian Medical Association Journal*, vol. 183, no. 18, pp. E1319–E1325, 2011.
- [40] M. A. Goldstein, *The MassGeneral Hospital for Children Adolescent Medicine Handbook*. Springer Science & Business Media, 2010.