Event-Driven Continuous Time Bayesian Networks: An Application in Modeling Progression out of Poverty through Integrated Social Services

Debarun Bhattacharjya[†], **Karthikeyan Shanmugam**[†], **Tian Gao**[†], **Nicholas Mattei**[§], and **Kush R. Varshney**[†]

[†] IBM Research, IBM T.J. Watson Research Center, Yorktown Heights, NY, USA
[§] Tulane University, Department of Computer Science, New Orleans, LA, USA

Abstract

We introduce a novel event-driven continuous time Bayesian network (ECTBN) representation to model situations where a system's state variables could be influenced by occurrences of events of various types. In this way, the model parameters and graphical structure capture not only potential "causal" dynamics of system evolution but also the influence of event occurrences that may be interventions. Our model is applicable in numerous domains, including health care, politics, and finance. We propose a greedy search procedure for structure learning based on the BIC score for a special class of ECTBNs; this is asymptotically consistent and also effective for limited data. We demonstrate the representation's power by applying it to model paths out of poverty for clients of CityLink Center, a non-profit integrated social service provider in Cincinnati, USA. The ECTBN captures the effect of classes/counseling sessions on an individual's life outcome areas such as education, transportation, employment and financial education.

1 Introduction

Real-world decision situations often involve uncertainties that interact with each other in a way that requires modeling complex dynamic inter-dependencies. The uncertain variables of interest of a system can be modeled as state variables whose evolution is captured by some dynamic process. A continuous-time Bayesian network (CTBN) [Nodelman *et al.*, 2002] representation handles modeling joint trajectories of a system's state variables where transitions are recorded at irregular time intervals. This relies on homogeneous Markov processes to model state transitions. In a CTBN, the distribution of time to next transition and the probability matrix of transitioning to a future state depend on the current state. While CTBNs offer a simple way to model the dynamics of irregular state transitions, they are only suitable when a dynamical system is observed in an isolated context. There are many real life scenarios that have the following added complexity: external occurrences of events could influence the manner in which the system evolves. In this paper, we consider modeling situations where occurrences of various types of *events* influence evolution of a set of *state variables*.

Many applications have external events that affect the evolution of a system: *Health*: A diabetic patient's blood glucose level and mental well-being are influenced by events such as insulin intake, meals and physical activity. *Finance*: The stock prices for a set of companies in an industry may be affected by natural events such as disasters or political events such as trade deals. *Social Impact*: Integrated social services such as counseling sessions and classes have an impact on a person's level of education, employment, and well-being.

Event datasets are sequences of labels on a time line where each label indicates the type of event. For example, time stamps of medication, exercise, and meals would indicate events that could be relevant for a patient's health outcomes as monitored by tracking vital measures. To capture the influence of events on state variables, we introduce **Event-Driven Continuous Time Bayesian Networks (ECTBNs)**, where in addition to state variables driving transitions of other state variables, a time stamped event history involving various types of events could influence the time to transition as well as the probability of transition of state variables.

Including events in the scope of the model requires a fundamental non-Markov extension to CTBNs. Such a model cannot be reduced to an expanded CTBN with proxy state variables for events. This is because the intensity function that determines the time to next transition in a CTBN only depends on the current configuration of parent state variables; it does not depend on *when* the configuration of these state variables attained their current configuration. However, when event sequences influence the intensity functions of state transitions, their previous times of occurrence could matter in general.

Contribution. We introduce and provide learning algorithms for a novel, interpretable yet analytically sophisticated graphical model that captures joint dynamics between events, modeled as a multivariate point process, and state variables, modeled as Markov processes. We apply our model to a real-world dataset of social service clients in the USA. This is work in partnership with CityLink Center – a non-profit organization in Cincinnati, USA that provides a suite of services to help adults in poverty meet their goals in areas such as education, employment and transportation.

2 Related Work

Continuous time processes model the dynamics of events occurring irregularly on a common timeline. Conditional inten-

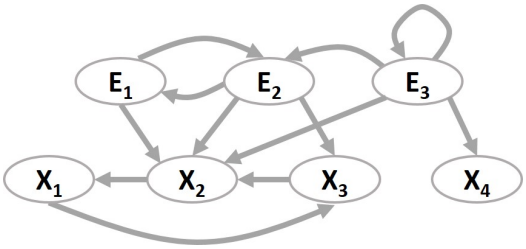

Figure 1: Illustrative ECTBN graph representing a dynamic process involving 4 state variables and 3 event labels.

sity functions capture the instantaneous rate of occurrence of a specific event given the history of other events. There has been a lot of work on various parametric models for learning conditional intensity functions for event streams. Notable amongst them are Hawkes processes [Hawkes, 1971], Poisson networks [Rajaram *et al.*, 2005], Poisson cascades [Simma and Jordan, 2010], piecewise-constant conditional intensity models [Gunawardana *et al.*, 2011], etc.

When the conditional intensity of an event depends only on the history of a set of parent events on the timeline, this can be represented using graphical event models (GEMs) [Didelez, 2008; Meek, 2014; Gunawardana and Meek, 2016]. These are different from time series graphs [Eichler, 1999; Dahlhaus, 2000] and dynamic Bayesian networks [Dean and Kanazawa, 1989; Murphy, 2002] as they represent continuous time processes. Additionally, continuous-time Bayesian networks (CTBNs) [Nodelman *et al.*, 2002, 2003], represents joint trajectories of discrete variables, as opposed to models of event streams in continuous time. In this work we introduce a model that can be viewed as a novel combination of joint trajectories in CTBNs and the effect of event arrivals. CTBNs are useful in diverse applications, including reliability analysis [Boudali and Dugan, 2006], cardiogenic heart failure [Gatti *et al.*, 2012], cybersecurity [Xu and Shelton, 2008] and gene network reconstruction [Acerbi *et al.*, 2014].

There is past work on dynamic modeling for social services that examines the use of Markov decision process-based planning to help navigate the welfare-to-work initiatives established in the United States during the Clinton presidency [Gehlbach *et al.*, 2006; Yi *et al.*, 2008; Dekhtyar *et al.*, 2009]. There has also been work on partially observed Markov decision processes to plan interventions to help homeless youths [Yadav *et al.*, 2016a,b]. Recent work by Kube *et al.* [2018] on allocating interventions to the homeless in the United States focuses on causal discovery, similar to our work. However, the setting is a static one instead of the dynamic events setting considered here. The development economics literature contains data-driven contributions that mathematically describe the dynamics of poverty in the developing world, but does not consider analysis for directly improving operations [Carter and May, 2001; Naschold, 2013]

3 Model Description

We introduce the ECTBN model, a representation that captures processes involving state variables and event occurrences, combining elements of CTBNs and GEMs. We start with a more general formulation and then specify the case that we use for learning and experimental investigation.

Consider a set of discrete state variables $\mathcal{X} = \{X_i\}_{i=1}^{I}$. Let $\mathrm{Val}(X_i)$ be the domain of variable X_i. The states of these variables are assumed to be known at all times between initial time $t_0 = 0$ to the end time T. Data about each variable is of the form of state transitions, $D_{X_i} = (t_k, x_k)_{k=0}^{N_i}$ where the state at time t_0 is the initial state and $x_{k+1} \neq x_k \; \forall k, x_k \in \mathrm{Val}(X_i)$. Data for all state variables taken together is denoted $D_{\mathcal{X}} = \bigcup_{X \in \mathcal{X}} D_X$.

We assume there is also data about events occurring over time, $D_{\mathcal{E}} = (t_k, e_k)_{k=1}^{N_E}$, where t_k are time stamps and e_k belong to an event label set $\mathcal{E} = \{E_j\}_{j=1}^{J}$. All the data taken together is $D = D_{\mathcal{X}} \cup D_{\mathcal{E}}$. We use $h(\cdot)$ to denote historical occurrences of events. Let $h_{\mathcal{B}}(t) = \{(t_k, e_k) \in D_{\mathcal{B}} : t_k < t\}$ represent the history of events in the set $\mathcal{B} \subset \mathcal{E}$ until time t.

Definition 1. *An Event-Driven Continuous Time Bayesian Network $\mathcal{N}$ includes:*

- *A directed (possibly cyclic) graph $\mathcal{G}$ where $\mathbf{U}_E \subset \mathcal{E}$ are the parents of event label E and $\mathbf{U}_X \subset \mathcal{X} \bigcup \mathcal{E} \setminus X$ are the parents of state variable $X \in \mathcal{X}$. We decompose the latter into: parents that are state variables $\mathbf{U}_{X(\mathcal{X})} \subseteq \mathcal{X} \setminus X$ and parents that are event labels $\mathbf{U}_{X(\mathcal{E})} \subseteq \mathcal{E}$.*
- *An initial distribution $\mathbf{P}^0_{\mathcal{X}}$ over state variables.*
- *Conditional intensity matrices for every $X \in \mathcal{X}$, $Q_{X|\mathbf{u}_{X(\mathcal{X})}, h_{\mathbf{U}_{X(\mathcal{E})}}(t)}$, which model state transitions. This depends on the current state $\mathbf{u}_{X(\mathcal{X})}$ of the parents $\mathbf{U}_{X(\mathcal{X})}$ at time t and history of labels in $\mathbf{U}_{X(\mathcal{E})}$ till time t, denoted $h_{\mathbf{U}_{X(\mathcal{E})}}(t)$. A matrix $Q(\cdot)$ is equivalent to considering waiting times $q_{x|\mathbf{u}_{X(\mathcal{X})}, h_{\mathbf{U}_{X(\mathcal{E})}}(t)}$ in state $X = x$ before transitioning to some other state $x' \neq x$, as well as the probabilities of transitioning from state x to state x' at time t, $\theta_{xx'|\mathbf{u}_{X(\mathcal{X})}, h_{\mathbf{U}_{X(\mathcal{E})}}(t)}$.*
- *Conditional intensity rates for every event label $E \in \mathcal{E}$, $\lambda_{E|h_{\mathbf{U}_E}(t)}$, which model event arrivals. This depends on the history of event labels in the parent set $\mathbf{U}_E$ at time t, denoted $h_{\mathbf{U}_E}(t)$.*

Figure 1 shows an illustrative ECTBN graph with 4 state variables and 3 event labels. Note that there may be cycles and even self-loops for an event label because its occurrence rate could depend on its own history. State variables could have event labels as parents but not vice versa. Our motivation here is to study situations where events probabilistically influence the uncertainties in a system but not the other way around.

It should be evident from the complex inhomogeneous history-dependence in Definition 1 that it is impractical in an ECTBN to consider all possible histories for modeling the influence of events on state variables; one cannot learn arbitrary dependencies with finite data as it would be difficult to generalize for learning. We simplify the historical dependence by making an assumption that results in an important special case, motivated by Bhattacharjya *et al.* [2018]:

Assumption 2. *Consider a set $\mathcal{W}$ of time windows for every edge from event label E directed into state variable X in graph $\mathcal{G}$, each denoted $w_{E,X}$. Assume that the rates and probabilities associated with state variable transitions depend only on whether a parent event label $E \in \mathbf{U}_{X(\mathcal{E})}$ occurred at least once in some recent time window $w_{E,X}$.*

Figure 2: Categorization of the six outcome areas into levels for the CityLink dataset.

The above assumption is the *proximal* or *recency* assumption, which captures the view that recent events matter more than older ones and simplifies parent conditions to be binary for each event label parent. Specifically, if $\mathbf{u}_{X(\mathcal{E})}$ denotes a vector of indicators, one each for whether an event label in $\mathbf{U}_{X(\mathcal{E})}$ occurs or not, then Assumption 2 simplifies the dependence of $q_{(\cdot)}$ and $\theta_{(\cdot)}$ parameters as follows:

$$\theta_{xx'|\mathbf{u}_{X(\mathcal{X})},h_{\mathbf{U}_{X(\mathcal{E})}}(t)} = \theta_{xx'|\mathbf{u}_{X(\mathcal{X})},\mathbf{u}_{X(\mathcal{E})}}$$
$$q_{x|\mathbf{u}_{X(\mathcal{X})},h_{\mathbf{U}_{X(\mathcal{E})}}(t)} = q_{x|\mathbf{u}_{X(\mathcal{X})},\mathbf{u}_{X(\mathcal{E})}} \quad (1)$$

The number of parameters can now be ascertained for any state variable. As an example, for the ECTBN in Figure 1, if state variable X_3 can take 3 values in its domain $\mathrm{Val}(X_3)$, then state variable X_2 has $2^3 * 3 = 24$ parental conditions $(\mathbf{u}_{X(\mathcal{X})}, \mathbf{u}_{X(\mathcal{E})})$ since X_2 has 3 event labels as parents, $\mathbf{U}_{X_2(\mathcal{E})} = \{E_1, E_2, E_3\}$, along with 1 state variable parent $\mathbf{U}_{X_2(\mathcal{X})} = \{X_3\}$.

Aside from the fact that the proximal assumption greatly simplifies notations, which helps with the exposition, it is often suitable in practice due to the nature of real-world causal influences; note that the influence itself may last for a long while, depending on the domain, which could be modeled using large time windows. Furthermore, Bhattacharjya *et al.* [2018] showed that the simplification often results in better performance as it can prevent overfitting using more general models such as the piece-wise constant intensity family [Gunawardana *et al.*, 2011; Parikh *et al.*, 2012].

4 Application to Tracking Life Outcomes

We apply the ECTBN model to study the effect of a set of services which are events , e.g., enrollment and attendance in a class or counseling session, on an individual's life outcome areas which are state variables, e.g., educational attainment or depression levels, in an integrated social services context. The data used for this section comes from our partnership with the CityLink Center in Cincinnati, Ohio, USA – a city-wide initiative launched in 2013 by a group of social service agencies and churches who recognized the need for a systemic approach to poverty.

CityLink recognizes that different realms of clients' lives are interrelated, so simply maximizing one realm without consideration of another realm leads to sub-optimal outcomes. CityLink's case management team works with clients in a holistic manner to help time and sequence the services for clients to best achieve and sustain their goals. This leads to integrated longitudinal data as clients are evaluated across multiple realms of their lives and client engagement stretches from one service to the next, providing a longer period of engagement and client data. CityLink utilizes this data to drive continuous improvement and determine how to best support clients in achieving their goals.

4.1 About the Data

We use data about a subset of CityLink's clients. Specifically, we use the approximately 1400 clients who have had more than 15 total interactions with CityLink. We consider 6 outcome areas that are tracked through CityLink's data: education, employment, financial education, transportation, anxiety, and depression. These are dimensions of an individual's progress in attaining a self-sustainable way out of poverty. Figure 2 summarizes the chosen levels for outcome areas,[1] each of which is modeled as a state variable. While it is typical for an individual's outcome area levels to increase, it is possible for an individual to regress as well. For some analysis, we also consider a higher level mapping where each outcome area has at most three levels.

We consider 11 types of services provided by CityLink and its partners, which are treated as events: 6 of them are group classes/sessions and 5 are one-on-one. The services include group industrial training, group classes on education, employment, financial education, transportation and wellness, as well as one-on-one sessions on employment, wellness, and financial education.

4.2 Learning and Analysis

We adopt the following learning procedure on the CityLink data, conducted separately for each state variable (outcome area) X. First, we configure a hyper-parameter setting for windows in $\mathcal{W}^c$ associated with incoming edges into X by uniformly randomly choosing a window from the list $\{15, 30, 60, 90, 180\}$ days for each event label. We repeat this procedure 100 times to build various window hyper-parameter configurations. Using 5-fold cross validation, we determine the optimal hyper-parameter setting by maximizing the average BIC score across folds.Finally, this optimal hyper-parameter setting is used to learn the optimal graph and parameters for X using all the training data.

Structure. Figure 3 presents the learned graphical structure and windows for the CityLink data. This graph was learned using a higher level mapping than shown in Figure 2 and

[1]These levels are a simplified version of CityLink's.

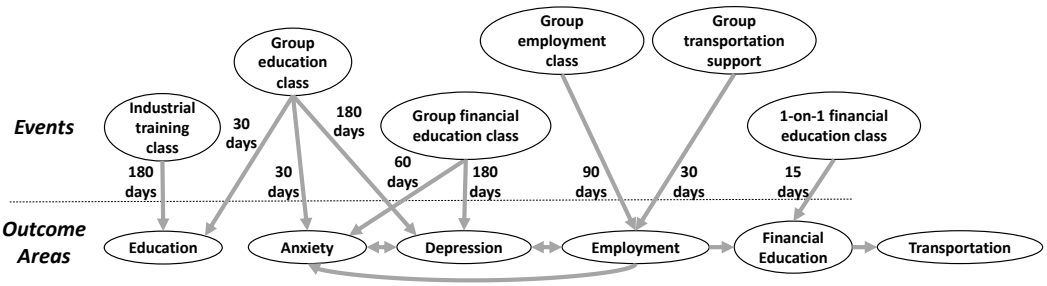

Figure 3: Learned ECTBN graph from CityLink data.

Outcome Area	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
Education	group edu class	group edu class	group edu class; indus. training	indus. training	indus. training	N/A
Employment	group emp class; group transp. class	group emp class; group transp. class	group emp class; group transp. class	-	-	N/A
Financial Education	1-on-1 fin-ed; group fin-ed	1-on-1 fin-ed; group fin-ed	1-on-1 fin-ed	1-on-1 fin-ed	N/A	N/A

Table 1: Event label parents during ECTBN state variable transition analysis for three outcome areas. New state variables are created with the following states: current level, next higher level and other level, showcasing important events for specific level increments.

Outcome Area	ECTBN	CTBN	CTBN-EV
Anxiety	**-5530**	-7282	-5639
Depression	**-5079**	-6715	-5122
Education	**-2100**	-2776	-2105
Employment	**-7486**	-7877	-7877
Financial Education	**-1400**	-1795	-1564
Transportation	**-1481**	**-1481**	**-1481**

Table 2: Log likelihood for the models on the CityLink data.

with a slightly reduced weight for the penalty term in the BIC score due to limited data. There are a number of interesting results that can be gleaned from the graph that may affect the way CityLink operates. To start with, group education classes have a direct and lasting effect on the Anxiety and Depression outcome areas, as do group financial education classes. Industrial training classes are deemed to have a longer duration of effect (180 days) on the Education outcome area than the other group education classes (30 days). Similarly, one-on-one financial education classes have more impact on the Financial Education outcome area than group financial education classes. The case of the Employment outcome area is an interesting result due to CityLink's integrated approach. For CityLink, group transportation classes are embedded into employment classes which immediately precede an individual's move to transitional employment. For this reason, the model picks transportation classes as influential.

The advantage of the ECTBN formalism is that it allows one to see the inter-related effects of not only the events but also the state variables. Employment has a direct effect on Anxiety, Depression, and Financial Education. It is interesting to see how critical Anxiety, Depression, and Employment are for the clients, and reinforces the importance of taking a holistic approach to case management.

State Variable Transition Analysis. We also conducted a study to better identify influential events that affect transitions from a particular outcome area level to the next level, which was of interest to CityLink. We did this by creating additional state variables to track when the level of an outcome area increased; this new state variable has three states – the current level (not the maximum level), the next higher level and some other level of the outcome area under consideration. An ECTBN is learned for each new state variable while considering other outcome areas and events.

Table 1 summarizes the ECTBN event parents for three outcome areas determined from this transition analysis, enabling us to foreground local effects that are not obvious from Figure 3. Selecting a few of these additional insights: (1) core education classes are important for transitions at lower levels of education whereas industrial training is important for transitions at higher levels; (2) the impact of group employment classes is particularly felt on low to mid levels of employment transitions; and (3) group financial education classes affect lower level transitions whereas the one-on-one classes are influential throughout the progression. For this particular analysis, all windows were set to 180 days during learning.

ECTBN vs. CTBN(s). We also checked how the ECTBN fit the data as compared to the following two baselines: CTBN: This is a regular CTBN that only considers state variables and does not see the event occurrences; and CTBN-EV (event variables): This is a CTBN where new state variables are introduced, one each for the $|\mathcal{E}|$ event labels. These are binary state variables which are active when the corresponding event is the most recent one observed and inactive otherwise.

Table 2 compares the log likelihood on the data across models, demonstrating that the ECTBN performs better than the baselines in this application. The Transportation outcome area is the exception, where no parent events were chosen in the optimal graph, though there is a parent state variable.

5 Conclusions

We have proposed a novel, interpretable graphical model that captures joint dynamics involving both event occurrences and state variables. As part of this research, we are applying AI methods for social good by modeling and analyzing the integrated social services delivery process. We aim to draw data-driven insights that could help CityLink Center's team be more effective in their counseling and operations, and help guide their clients out of poverty. Our modeling work reveals the most impactful events on state variables and probabilistically quantifies their effect on how long it takes for individuals to make progress along various outcome areas.

Acknowledgements

The authors thank Johnmark Oudersluys and the rest of the CityLink Center team for their partnership in this research.

References

E. Acerbi, T. Zelante, V. Narang, and F. Stella. Gene network inference using continuous time Bayesian networks: a comparative study and application to Th17 cell differentiation. *BMC Bioinform.*, 15(1):387, 2014.

D. Bhattacharjya, D. Subramanian, and T. Gao. Proximal graphical event models. In *Proceedings of Neural Information Processing Systems (NIPS)*, 2018.

H. Boudali and J. B. Dugan. A continuous-time Bayesian network reliability modeling, and analysis framework. *IEEE T. Reliab.*, 55(1):86–97, 2006.

M. R. Carter and J. May. One kind of freedom: Poverty dynamics in post-apartheid South Africa. *World Development*, 29(12):1987–2006, December 2001.

R. Dahlhaus. Graphical interaction models for multivariate time series. *Metrika*, 51:157–172, 2000.

T. Dean and K. Kanazawa. A model for reasoning about persistence and causation. *Computational Intelligence*, 5:142–150, 1989.

A. Dekhtyar, J. Goldsmith, B. Goldstein, K. K. Mathias, and C. Isenhour. Planning for success: The interdisciplinary approach to building Bayesian models. *Int. J. Approx. Reason.*, 50(3):416–428, March 2009.

V. Didelez. Graphical models for marked point processes based on local independence. *J. R. Stat. Soc. B*, 70(1):245–264, 2008.

M. Eichler. *Graphical Models in Time Series Analysis*. PhD thesis, University of Heidelberg, Germany, 1999.

E. Gatti, D. Luciani, and F. Stella. A continuous time Bayesian network model for cardiogenic heart failure. *Flex. Serv. Manuf. J.*, 24(4):496–515, 2012.

K. R. Gehlbach, B. Laracuente, C. Isenhour, J. Goldsmith, B. Goldstein, and M. Truszczyński. A benchmark model for decision-theoretic planning with constraints. In *Proc. Bayesian Model. Appl. Workshop*, pages 68–73, Cambridge, MA, USA, July 2006.

A. Gunawardana and C. Meek. Universal models of multivariate temporal point processes. In *Proc. Int. Conf. Artif. Intell. Stat.*, pages 556–563, 2016.

A. Gunawardana, C. Meek, and P. Xu. A model for temporal dependencies in event streams. In *Adv. Neur. Inf. Process. Syst.*, pages 1962–1970, 2011.

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

A. R. Kube, S. Das, and P. J. Fowler. Allocating interventions based on counterfactual predictions: A case study on homelessness services. In *Proc. Workshop Mechan. Des. Social Good*, Ithaca, NY, USA, June 2018.

C. Meek. Toward learning graphical and causal process models. In *Proc. UAI Workshop Causal Inference: Learning and Prediction*, pages 43–48, 2014.

K. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of California Berkeley, USA, 2002.

F. Naschold. Welfare dynamics in Pakistan and Ethiopia – does the estimation method matter? *J. Develop. Stud.*, 49(7):936–954, July 2013.

U. Nodelman, C. R. Shelton, and D. Koller. Continuous time Bayesian networks. In *Proc. Conf. Uncertainty Artif. Intell.*, pages 378–378, 2002.

U. Nodelman, C. R. Shelton, and D. Koller. Learning continuous time Bayesian networks. In *Proc. Conf. Uncertainty Artif. Intell.*, pages 451–458, 2003.

A. P. Parikh, A. Gunawardana, and C. Meek. Conjoint modeling of temporal dependencies in event streams. In *Proceedings of Uncertainty in Artificial Intelligence Workshop on Bayesian Modeling Applications*, August 2012.

S. Rajaram, T. Graepel, and R. Herbrich. Poisson-networks: A model for structured point processes. In *Proc. Int. Workshop Artif. Intell. Stat.*, pages 277–284, 2005.

A. Simma and M. I. Jordan. Modeling events with cascades of Poisson processes. In *Proc. Conf. Uncertainty Artif. Intell.*, pages 546–555, 2010.

J. Xu and C. R. Shelton. Continuous time Bayesian networks for host level network intrusion detection. In *Joint Eur. Conf. Mach. Learn. Knowl. Disc. Databases*, pages 613–627, 2008.

A. Yadav, H. Chan, A. Jiang, E. Rice, E. Kamar, B. Grosz, and M. Tambe. POMDPs for assisting homeless shelters - computational and deployment challenges. In *Proc. IDEAS Workshop Int. Conf. Auton. Agents Multiagent Syst.*, Singapore, May 2016.

A. Yadav, E. Kamar, B. Grosz, and M. Tambe. HEALER: POMDP planning for scheduling interventions among homeless youth. In *Proc. Int. Conf. Auton. Agents Multiagent Syst.*, pages 1504–1506, Singapore, May 2016.

L. Yi, R. Finkel, and J. Goldsmith. Planning for welfare to work. In *Proc. Int. FLAIRS Conf.*, pages 696–701, Coconut Grove, FL, USA, May 2008.