Maximizing Awareness about HIV in Social Networks of Homeless Youth with Limited Information

Amulya Yadav[1], **Hau Chan**[2], **Albert Xin Jiang**[2], **Haifeng Xu**[1], **Eric Rice**[1], **Milind Tambe**[1]

[1]Center for Artificial Intelligence in Society, University of Southern California, LA, CA 90089

[2]Trinity University, San Antonio, TX 78212

[1]{amulyaya, haifengx, ericr, tambe}@usc.edu [2]{hchan, xjiang}@trinity.edu

Abstract

This paper presents HEALER, a software agent that recommends sequential intervention plans for use by homeless shelters, who organize these interventions to raise awareness about HIV among homeless youth. HEALER's sequential plans (built using knowledge of social networks of homeless youth) choose intervention participants strategically to maximize influence spread, while reasoning about uncertainties in the network. While previous work presents influence maximizing techniques to choose intervention participants, they do not address two real-world issues: (i) they *completely fail* to scale up to real-world sizes; and (ii) they do not handle deviations in execution of intervention plans. HEALER handles these issues via two major contributions: (i) HEALER casts this influence maximization problem as a POMDP and solves it using a novel planner which scales up to previously unsolvable real-world sizes; and (ii) HEALER allows shelter officials to modify its recommendations, and updates its future plans in a deviation-tolerant manner. HEALER was deployed in the real world in Spring 2016 with considerable success.

1 Introduction

HIV-AIDS kills 2 million people worldwide every year [UN-AIDS, 2012]. Further, HIV has an extremely high incidence among homeless youth, who are at a 10X greater risk of HIV infection than stably housed populations [Council, 2012]. To minimize infection among homeless youth, many homeless shelters organize intervention camps for youth, in order to raise awareness about HIV prevention/treatment practices [Rice *et al.*, 2012c]. These intervention camps consist of day-long educational sessions in which the participants are provided with information about HIV prevention measures.

However, due to financial constraints, the shelters can only organize a limited number of intervention camps. Moreover, in each camp, the shelters can only manage small groups of youth (∼3-4) at a time (as emotional and behavioral problems of youth makes management of bigger groups difficult). Thus, the shelters prefer a series of small sized camps organized sequentially [Rice *et al.*, 2012b]. As a result, the shelter

cannot intervene on the entire target (homeless youth) population. Instead, it tries to maximize the spread of awareness among the target population (via word-of-mouth influence) using the limited resources at its disposal. To achieve this goal, the shelter uses the friendship based social network of the target population to strategically choose the participants of their limited intervention camps. Unfortunately, the shelters' job is further complicated by a lack of complete knowledge about the social network's structure [Rice, 2010]. Some friendships in the network are known with certainty whereas there is uncertainty about other friendships.

Thus, the shelters face an important challenge: they need a sequential plan to choose the participants of their sequentially organized interventions. This plan must address three key points: (i) it must deal with network structure uncertainty; (ii) it must take into account new information uncovered during the interventions, which reduces our uncertainty about the network structure; and (iii) the intervention approach should address the challenge of gathering information about social networks of homeless youth, which usually costs thousands of dollars and many months of time [Rice *et al.*, 2012b].

In this paper, we model the shelters' problem by introducing the *Dynamic Influence Maximization under Uncertainty* (or DIME) problem. Further, we build a new software agent, HEALER (**H**ierarchical **E**nsembling based **A**gent which p**L**ans for **E**ffective **R**eduction in HIV Spread), to provide an end-to-end solution to the DIME problem. HEALER casts the DIME problem as a Partially Observable Markov Decision Process (POMDP) and solves it using HEAL (**H**ierarchical **E**nsembling **A**lgorithm for p**L**anning), a novel POMDP planner which quickly generates high-quality recommendations (of intervention participants) for homeless shelter officials. Our simulations show that HEALER achieves a 100X speed up and 70% improvement in solution quality over PSINET (state-of-the-art in previous work) [Yadav *et al.*, 2015] ; and scales up to larger problem sizes gracefully. Moreover, HEALER quickly gathers information about the homeless youth social network (at low cost) by interacting with youth via a network generation application.

We tested HEALER's performance in the field by conducting three real-world pilot studies, in collaboration with a homeless shelter in Los Angeles (*Safe Place For Youth*[1]).

---

[1]http://www.safeplaceforyouth.org/

(a) Computers at Homeless Shelter

(b) Emergency Resource Shelf at the Homeless Shelter

Figure 1: Facilities at our Collaborating Homeless Shelter

Figure 2: HEALER's Design

These shelters provide food and lodging to ∼55-60 homeless youth (aged 12-25) every day. These pilot studies showed that HEALER was able to outperform other standard baselines used in the real-world. To the best of our knowledge, this pilot study represents the first real-world evaluation of such sequential influence maximization algorithms. Please refer to [Yadav *et al.*, 2017] for details on the pilot studies. Further, please refer to [Yadav *et al.*, 2016a] for a higher level understanding of the problem domain.

2 HEALER's Design

We now explain the high-level design of HEALER. It consists of two major components: (i) a network generation application for gathering information about social networks; and (ii) a DIME Solver, which solves the DIME problem (introduced in Section 4). We first explain HEALER's components and then explain HEALER's design.

Network Generation Application: HEALER gathers information about social ties in the homeless youth social network by interacting with youth via a network generation application. Once a fixed number of homeless youth register in our network application, HEALER parses the contact lists of all the registered homeless youth on social media and generates the social network between these youth. HEALER adds a link between two people, if and only if both people are (i) contacts on social media; and (ii) are registered in its network generation application. Unfortunately, there is *uncertainty* in the generated network as friendship links between people who are only friends in real-life (and not on social media) are not captured by HEALER.

DIME Solver: The DIME solver then takes the approximate social network (generated by HEALER's network generation application) as input and solves the DIME problem using HEAL algorithm (Section 5.1). HEALER provides the solution of this DIME problem as a series of recommendations (of intervention participants) to homeless shelter officials.

HEALER Design: HEALER's design (shown in Figure 2), begins with the network generation application constructing an *uncertain* network (as explained above). HEALER has a *sense-reason-act* cycle; where it repeats the following process for T interventions.

It *reasons* about different long-term plans to solve the DIME problem, it *acts* by providing DIME's solution as a recommendation (of intervention participants) to homeless shelter officials. The officials may choose to not use HEALER's recommendation in selecting their intervention's participants.

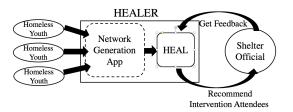This paper is an abridged version of [Yadav et. al., 2015]

Upon the intervention's completion, HEALER *senses* feedback about the conducted intervention from the officials. This feedback includes new observations about the network, e.g., uncertainties in some links may be resolved as intervention participants are interviewed by the shelter officials (explained more in Section 4). HEALER uses this feedback to update and improve its future recommendations.

3 Background

We represent social networks as directed graphs (consisting of *nodes* and *directed edges*) where each *node* represents a person in the social network and a *directed edge* between two nodes A and B (say) represents that node A *considers* node B as his/her friend. *We assume directed-ness of edges as sometimes homeless shelters assess that the influence in a friendship is very much uni-directional; and to account for uni-directional follower links on social media.* Otherwise friendships are encoded as two uni-directional links.

Uncertain Network: The uncertain network is a directed graph $G = (V, E)$ with $|V| = N$ nodes and $|E| = M$ edges. The edge set E consists of two disjoint subsets of edges: E_c (the set of certain edges, i.e., friendships which we are certain about) and E_u (the set of uncertain edges, i.e., friendships which we are uncertain about). Note that uncertainties about friendships exist because HEALER's network generation application misses out on some links between people who are friends in real life, but not on social media. To model the uncertainty about missing edges, every uncertain edge $e \in E_u$ has an existence probability $u(e)$ associated with it, which represents the likelihood of "existence" of that uncertain edge. For example, if there is an uncertain edge (A, B) (i.e., we are unsure whether node B is node A's friend), then $u(A, B) = 0.75$ implies that B is A's friend with a 0.75 chance. In addition, each edge $e \in E$ (both certain and uncertain) has a propagation probability $p(e)$ associated with it. A propagation probability of 0.5 on directed edge (A, B) denotes that if node A is influenced (i.e., has information about HIV prevention), it influences node B (i.e., gives information to node B) with a 0.5 probability in each subsequent time step (our full influence model is defined below). This graph G with all relevant $p(e)$ and $u(e)$ values represents an uncertain network and serves as an input to the DIME problem. Figure 3 shows an uncertain network on 6 nodes (A to F) and 7 edges. The dashed and solid edges represent uncertain (edge numbers 1, 4, 5 and 7) and certain (edge numbers 2, 3 and 6) edges, respectively.

Influence Model We use a variant of the independent cascade model [Yan *et al.*, 2011]. In the standard independent
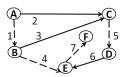
Figure 3: Uncertain Network

cascade model, all nodes that get influenced at round t get a **single** chance to influence their un-influenced neighbors at time $t + 1$. If they fail to spread influence in this **single** chance, they don't spread influence to their neighbors in future rounds. Our model is different in that we assume that nodes get **multiple** chances to influence their un-influenced neighbors. If they succeed in influencing a neighbor at a given time step t', they stop influencing that neighbor for all future time steps. Otherwise, if they fail in step t', they try to influence again in the next round. This variant of independent cascade has been shown to empirically provide a better approximation to real influence spread than the standard independent cascade model [Cointet and Roth, 2007; Yan *et al.*, 2011]. Further, we assume that nodes that get influenced at a certain time step remain influenced for all future time steps.

4 DIME Problem

We now provide an informal problem statement for DIME. Please refer to [Yadav *et al.*, 2016b] for a more formal problem statement and some hardness results for DIME.

Given the *uncertain network* as input, HEALER runs for T *rounds* (corresponding to the number of interventions organized by the homeless shelter). In each round, HEALER chooses K *nodes* (youth) as intervention participants. These participants are assumed to be influenced post-intervention with certainty. Upon influencing the chosen nodes, HEALER 'observes' the true state of the *uncertain edges* (friendships) out-going from the selected nodes. This translates to asking intervention participants about their 1-hop social circles, which is within the homeless shelter's capabilities [Rice *et al.*, 2012a].

After each round, influence spreads in the network according to our influence model for a single time step, before we begin the next round. HEALER only knows that explicitly chosen nodes (our intervention participants in all past rounds) are influenced. Informally then, given an uncertain network $G_0 = (V, E)$ and integers T and K (as defined above), the DIME problem (solved by HEALER) involves finding an online policy Π_T^* for choosing *exactly* K nodes for T successive rounds (interventions) which maximizes influence spread in the network at the end of T rounds.

5 HEAL: DIME's Problem Solver

We model DIME as a POMDP [Puterman, 2009] because of two reasons. First, POMDPs are a good fit for DIME as (i) we conduct several interventions sequentially, similar to sequential POMDP actions; and (ii) we have *partial observability* (similar to POMDPs) due to uncertainties in network structure and influence status of nodes. Second, POMDP solvers

have recently shown great promise in generating near-optimal policies efficiently [Silver and Veness, 2010]. We now explain how we map DIME onto a POMDP.

States. A POMDP state in our problem represents the state of the nodes (that is, whether they are influenced or not), along with the state of the uncertain edges (that is, whether they exist or not). This information is represented as a set of two binary tuples: W and F. Intuitively, $W_i = 1$ denotes that node i is influenced and $W_i = 0$ otherwise. Moreover, $F_i = 1$ denotes that the ith uncertain edge exists in reality, and $F_i = 0$ otherwise.

Actions. Every choice of a subset of K nodes is a POMDP action. For example, in Figure 3, one possible action is $\{A, B\}$ (when $K = 2$).

Observations. Previous studies [Rice *et al.*, 2012b] show that in order to learn more about the social network structure, intervention participants can be asked about their one-hop social circles by homeless shelter officials during the intervention. Therefore, upon taking a POMDP action, we "*observe*" the ground reality of the uncertain edges outgoing from the nodes chosen in that action. For example, by taking action $\{B, C\}$ in Figure 3, the values of F_4 and F_5 (i.e., the F-values of uncertain edges in the 1-hop social circle of nodes B and C) would be observed. In other words, we would find out whether the uncertain edges 4 and 5 actually exist in the network or not.

Rewards. The reward attained by taking an action a in a POMDP state s is the number of nodes that are newly influenced in the network by taking that action.

Please refer to [Yadav *et al.*, 2016b] for the full POMDP model. We solve this POMDP using a novel algorithm (described in Section 5.1) to find the optimal policy Π_T^* for the DIME problem.

5.1 HEALER's DIME Solver: HEAL

HEAL solves this *original POMDP* using a novel *hierarchical ensembling heuristic*: it creates ensembles of imperfect (and smaller) POMDPs at *two* different layers, in a hierarchical manner (see Figure 4). HEAL's *top layer* creates an ensemble of smaller sized *intermediate POMDPs* by subdividing the original *uncertain network* into several smaller sized *partitioned networks* by using graph partitioning techniques [LaSalle and Karypis, 2013], which generates partitions that minimize the edges going across partitions (while ensuring that partitions have similar sizes). Since these partitions are "almost" disconnected, we solve each partition separately. Each of these partitioned networks is then mapped onto a POMDP, and these *intermediate POMDPs* form our *top layer* ensemble of POMDP solvers.

In the bottom layer, each *intermediate POMDP* is solved using TASP (**T**ree **A**ggregation for **S**equential **P**lanning), our novel POMDP planner, which subdivides the POMDP into another ensemble of smaller sized *sampled POMDPs*. Each member of this *bottom layer* ensemble is created by randomly sampling uncertain edges of the partitioned network to get a sampled network having no uncertain edges, and this sampled network is then mapped onto a *sampled POMDP*. Each sampled POMDP is solved using a novel tree search algorithm, which avoids the exponential branching factor seen
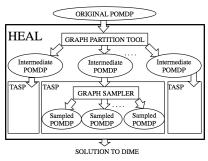
Figure 4: Hierarchical decomposition in HEAL

(a) Solution quality

(b) Scale up in K

Figure 5: Experiments show improvement over previous work

in PSINET [Yadav *et al.*, 2015]. Finally, the solutions of POMDPs in both the *bottom* and *top layer* ensembles are aggregated using novel techniques to get the solution for HEAL's original POMDP.

These heuristics enable scale up to real-world sizes (*at the expense of sacrificing performance guarantees*), as instead of solving one huge problem, we now solve several smaller problems. However, these heuristics perform very well in practice (as we show in Section 6). We now explain HEAL's top layer, and refer the reader to [Yadav *et al.*, 2016b] for details of HEAL's bottom layer and for a more complete understanding of the HEAL algorithm.

Top layer: Using Graph Partitioning In HEAL's top layer, we use METIS [LaSalle and Karypis, 2013], a state-of-the-art graph partitioning technique, to subdivide our original uncertain network into different partitioned networks. These partitioned networks form the ensemble of *intermediate POMDPs* (in Figure 4) in HEAL. Then, TASP (our bottom layer solver) is invoked on each intermediate POMDP independently, and their solutions are aggregated to get the final DIME solution. We try two different partitioning/aggregation techniques, which leads to two variants of HEAL:

K Partition Variant (HEAL): Given the *uncertain* network G and the parameters K and T as input, we first partition the uncertain network into K partitions. In each round from 1 to T, we invoke the bottom layer TASP algorithm [Yadav *et al.*, 2016b] to select 1 node from each of the K clusters. These singly selected nodes from the K clusters give us an action of K nodes, which is given to shelter officials to execute. Based on the *observation* (about uncertain edges) that officials get while executing the action, we update the partition networks (which are input to the *intermediate POMDPs*) by either replacing the *observed* uncertain edges with certain edges (if the edge was *observed* to exist in reality) or removing the uncertain edge altogether (if the edge was *observed* to *not exist* in reality).

T Partition Variant (HEAL-T): Given the *uncertain* network G and parameters K and T as input, we first partition the uncertain network into T partitions and TASP picks K nodes from the ith partition ($i \in [1,T]$) in the ith round.

6 Simulation Results

Figure 5a shows the influence spread of different algorithms on four real world networks of homeless youth. The x-axis shows the four networks and the y-axis shows the in-
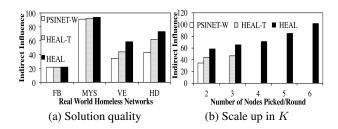
fluence spread achieved ($K = 2$, $T = 5$) by HEAL, HEAL-T and PSINET-W [Yadav *et al.*, 2015]. This figure shows that (i) HEAL outperforms all other algorithms on every network; (ii) *it achieves ~70% improvement over PSINET-W in VE and HD networks*; (iii) it achieves ~25% improvement over HEAL-T. The difference between HEAL and other algorithms is not significant in the FB and MYS networks, as PSINET is already influencing almost all nodes in these two relatively small networks.

Figure 5b shows influence spread achieved by HEAL, HEAL-T and PSINET-W on the VE and HD networks respectively ($T = 5$), as we scale up values of K, i.e., number of nodes picked per round. The x-axis shows increasing K values, and the y-axis shows the influence spread. This figure show that (i) PSINET-W and HEAL-T fail to scale up – they cannot handle more than $K = 2$ and $K = 3$ respectively (thereby not fulfilling real world demands); (ii) HEAL outperforms all other algorithms. Note that we do not compare with Greedy (the well known influence maximization algorithm [Golovin and Krause, 2011]) as DIME's influence function is *not adaptive submodular* which leads to loss of its performance guarantees. Please refer to a recent paper [Wilder *et al.*, 2017] for a more detailed analysis on Greedy.

7 Conclusion

In this paper, we develop HEALER, a new software agent for selecting participants of sequentially deployed interventions (which are organized by homeless shelters to spread awareness about HIV among homeless youth). HEALER casts the problem as a POMDP and uses a novel algorithm (HEAL) to achieve a 100X speedup over state-of-the-art algorithms while outperforming them by 70% in terms of solution quality. More than that, it runs when previous algorithms can't scale up. Also, HEALER saves homeless shelters' thousands of dollars and many months of time by generating uncertain networks at low cost using its network generation application. HEALER was tested in a real-world pilot study with 60 homeless youth from Safe Place for Youth (a homeless shelter in Los Angeles), and was shown to outperform other baselines in the field.

Acknowledgements

References

[Cointet and Roth, 2007] Jean-Philippe Cointet and Camille Roth. How Realistic Should Knowledge Diffusion Models Be? *Journal of Artificial Societies and Social Simulation*, 10(3):5, 2007.

[Council, 2012] National HCH Council. HIV/AIDS among Persons Experiencing Homelessness: Risk Factors, Predictors of Testing, and Promising Testing Strategies, December 2012.

[Golovin and Krause, 2011] Daniel Golovin and Andreas Krause. Adaptive Submodularity: Theory and Applications in Active Learning and Stochastic Optimization. *Journal of Artificial Intelligence Research*, 42:427–486, 2011.

[LaSalle and Karypis, 2013] Dominique LaSalle and George Karypis. Multi-threaded Graph Partitioning. In *Parallel & Distributed Processing (IPDPS), 2013 IEEE 27th International Symposium on*, pages 225–236. IEEE, 2013.

[Puterman, 2009] Martin L Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, 2009.

[Rice *et al.*, 2012a] Eric Rice, Anamika Barman-Adhikari, Norweeta G Milburn, and William Monro. Position-specific HIV risk in a Large Network of Homeless Youths. *American journal of public health*, 102(1):141–147, 2012.

[Rice *et al.*, 2012b] Eric Rice, Anthony Fulginiti, Hailey Winetrobe, Jorge Montoya, Aaron Plant, and Timothy Kordic. Sexuality and Homelessness in Los Angeles public schools. *American Journal of Public Health*, 102, 2012.

[Rice *et al.*, 2012c] Eric Rice, Eve Tulbert, Julie Cederbaum, Anamika Barman Adhikari, and Norweeta G Milburn. Mobilizing Homeless Youth for HIV Prevention: a Social Network Analysis of the Acceptability of a face-to-face and Online Social Networking Intervention. *Health education research*, 27(2):226, 2012.

[Rice, 2010] Eric Rice. The Positive Role of Social Networks and Social Networking Technology in the Condom-using Behaviors of Homeless Young People. *Public health reports*, 125(4):588, 2010.

[Silver and Veness, 2010] David Silver and Joel Veness. Monte-Carlo Planning in large POMDPs. In *Advances in Neural Information Processing Systems*, pages 2164–2172, 2010.

[UNAIDS, 2012] UNAIDS. Report on the Global AIDS Epidemic, March 2012.

[Wilder *et al.*, 2017] Bryan Wilder, Amulya Yadav, Nicole Immorlica, Eric Rice, and Milind Tambe. Uncharted but not Uninfluenced: Influence Maximization with an uncertain network. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2017.

[Yadav *et al.*, 2015] A Yadav, L Marcolino, E Rice, R Petering, H Winetrobe, H Rhoades, M Tambe, and H Carmichael. Preventing HIV Spread in Homeless Populations Using PSINET. In *Proceedings of the Twenty-Seventh Conference on Innovative Applications of Artificial Intelligence (IAAI-15)*, 2015.

[Yadav *et al.*, 2016a] A Yadav, L Marcolino, E Rice, R Petering, H Winetrobe, H Rhoades, M Tambe, and H Carmichael. PSINET: Assisting HIV Prevention Amongst Homeless Youth by Planning Ahead. In *AI Magazine*, 2016.

[Yadav *et al.*, 2016b] Amulya Yadav, Hau Chan, Abert Jiang, Haifeng Xu, Eric Rice, and Milind Tambe. Using Social Networks to Aid Homeless Shelters: Dynamic Influence Maximization under Uncertainty - An Extended Version. In *Proceedings of the Fifteenth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016)*, 2016.

[Yadav *et al.*, 2017] Amulya Yadav, Bryan Wilder, Robin Petering, Eric Rice, and Milind Tambe. Influence maximization in the field: The arduous journey from emerging to deployed application. In *In Proceedings of the International Conference on Autonomous Agents and Multi-agent Systems (AAMAS), 2017*, 2017.

[Yan *et al.*, 2011] Qiuling Yan, Shaosong Guo, and Dongqing Yang. Influence Maximizing and Local Influenced Community Detection based on Multiple Spread Model. In *Advanced Data Mining and Applications*, pages 82–95. Springer, 2011.