

Modeling the dynamism of HIV information diffusion in multiplex networks of homeless youth



Lindsay E. Young^{a,*}, Jerome Mayaud^b, Sze-Chuan Suen^c, Milind Tambe^f, Eric Rice^{d,e}

^a University of Southern California, Anneberg School for Communication and Journalism, United States

^b Spare Labs, Canada

^c University of Southern California Viterbi School of Engineering, Department of Industrial & Systems Engineering, United States

^d University of Southern California Suzanne Dworak-Peck School of Social Work, United States

^e University of Southern California, Center for Artificial Intelligence in Society (CAIS), United States

^f Harvard University, Center for Research in Computation and Society (CRCS), United States

ARTICLE INFO

Keywords:

Information diffusion
Independent cascade model
Linear threshold model
Network dynamics
Multiplexity
Homeless youth
HIV prevention

ABSTRACT

Cascade and threshold models are widely used to predict information diffusion in social networks, yet their characterization of networks as static and monoplex limit their ability to accurately predict how information propagates in dynamic, multiplex social environments. Using data from a peer-led HIV prevention intervention for homeless youth, we determine whether manipulating the baseline social network by (1) adding ties observed at later time points, and (2) accounting for alternative relational contexts improves each model's predictive accuracy. Results show that the addition of new ties improves the performance of both models, while substituting the context of interaction yields only minor improvements.

1. Introduction

For the more than 35,000 unaccompanied youth experiencing homelessness on a given night in the United States (U.S. Department of Housing and Urban Development, 2019), HIV and other sexual health risks are significant concerns (Gangamma et al., 2008; Rew et al., 2005). Whereas the prevalence of HIV among the general population of young people in the United States is about 0.2 % (Sweeney et al., 1995), the rate of infection among homeless youth far exceeds this in many urban areas, with reports ranging from 2% to more than 10 % (Marshall et al., 2010). This disparity can best be explained by differences in behavior and circumstances. Relative to stably housed youth, homeless youth are more likely to engage in behaviors that put them at increased HIV risk, such as condomless sex (Solorio et al., 2008), partner concurrency (Halcón and Lifson, 2004), survival sex (Tyler, 2009), and substance use (Kipke et al., 1997), and are more likely to experience correlates of these behaviors like childhood maltreatment and sexual victimization (Melander and Tyler, 2010).

The idea that peers (or peer networks) play a critical role in these risk behaviors is well established (Ennett et al., 1999; Rice et al., 2012a,b; Rice et al., 2007) and has informed purposeful efforts to (re-)appropriate the influence of peers toward increasing HIV prevention

engagement among homeless youth (Arnold and Rotheram-Borus, 2009; Kennedy et al., 2016; Rice et al., 2012a,b). We refer to these efforts as peer-led network interventions, which have a history of use for HIV prevention (Kelly et al., 1991; Latkin et al., 2003; Schneider et al., 2015; Young et al., 2017, 2013) And, in a community naturally suspicious of adults and external experts (Coates and McKenzie-Mohr, 2010), peer leader approaches that prioritize endogenous sources of knowledge and trust are desirable for linking homeless youth to HIV-related healthcare (Yadav et al., 2018).

However, successful implementation of peer-led HIV prevention interventions for homeless youth requires an understanding of how information is likely to diffuse in their peer networks and which network members are optimally positioned to maximize the spread of information to the greatest number of network members (Yadav et al., 2016). To these ends, algorithmic models of information diffusion, most notably the Independent Cascade Model (ICM) (Goldenberg et al., 2001) and the Linear Threshold Model (LTM) (Granovetter, 1978), offer the opportunity to approximate the structure of the diffusion process and, in turn, the overall reach of the intervention contingent on the set of peer leaders selected to seed the information propagation process. Both models assume that diffusion unfolds iteratively, from activated to unactivated nodes, along observed pathways (or edges) in the network

* Corresponding author at: University of Southern California, Annenberg School for Communication & Journalism, 3630 Watt Way, Los Angeles, CA, 90089, United States.

E-mail address: lindsay.young@usc.edu (L.E. Young).

<https://doi.org/10.1016/j.socnet.2020.06.001>

(Guille et al., 2013). In cascade models, the node activation rule is based on independent edge-wise decisions, while threshold models posit cumulative effects on nodes, which lead to activation once a personal threshold is exceeded (Kempe et al., 2003).

Although the ICM and LTM were initially developed to describe social influence in natural settings where actors physically interact with one another, their most typical applications tend to make assumptions about social network topology that are difficult to generalize to physical diffusion settings (Hu et al., 2017). Specifically, most applications tend to assume that the underlying network is *closed* (i.e., information only spreads among observed network members and along observed edges), is *monoplex* (i.e., consisting of only one mode of social interaction between network members), and is *static* (i.e., the network remains unchanged over time) (Guille et al., 2013; Zhong et al., 2017). Although starting premises like these may be reasonable for modeling the spread of information in well-defined online networks with real-time digital trace data, they seem overly constrained and unrealistic for modeling information propagation in real-world physical social settings that are more porous and dynamic (Guille et al., 2013). The implication is that the ICM and LTM, with their strong assumptions about network topology, are potentially more prone to error and inaccuracy in their predictions of information spread (Butts, 2003).

The limitations of these assumptions come into sharper focus when contemplating the peer networks of homeless youth. First, from a sampling perspective, homeless youth are considered a “hidden” or “hard to reach” population, meaning that 1) there exists no sampling frame, leaving the size and boundaries of the population unknown to researchers, and 2) membership entails privacy concerns due to stigmatized and/or illegal behavior (Heckathorn, 1997), leading to refusals to participate in studies that would identify them as homeless. As such, a complete network of who knows whom among homeless youth is difficult to capture. Whatever snapshot of the network is captured for the purposes of a peer-led intervention will likely not include every homeless youth who could possibly be exposed to the intervention or who could be considered a viable network influencer. The implication is that, with the underlying network artificially bounded as it is, “externalities” (e.g., unobserved nodes and edges) will always influence the information propagation process in this population.

Second, in the real world, individuals often have network members with overlapping roles and, therefore, have the opportunity to interact with their peers in multiple social contexts, for example as neighbors and as members of the same extracurricular club. We call this phenomena *network multiplexity*. Homeless youth are no different; the same peer can play the role of friend, drug partner, and sex partner (De la Haye et al., 2012). Thus, information can feasibly travel between peers in any of these social circles. However, models like the ICM and LTM were not originally developed to combine relational inputs from multiple social contexts (Brummitt et al., 2012; Zhong et al., 2017). Furthermore, network intervention studies rarely attempt to intervene on more than one type of peer relationship, leaving alternative social pathways of diffusion unobserved.

Finally, although the peer networks of homeless youth tend to be smaller than those of stably housed youth (Falci et al., 2011), the relationships they do have, especially those with other homeless youth, are often tenuous and transient (Usborne et al., 2009). Regardless of the degree of social or emotional support that youth receive from their homeless peers, there are a number of situational and behavioral reasons for why their peer relationships lack stability. For example, time spent in homelessness (Falci et al., 2011), mental illness and substance use (Hawkins and Abrams, 2007), transitions in and out of different living situations (e.g., shelters, on the street, friends, transitional housing, etc.) (Sutherland, 2016), and inconsistent or limited access to the Internet and other communication technologies (Barman-Adhikari et al., 2016) are all impediments to maintaining stable relationships for homeless youth. Moreover, in a network intervention setting, the networks of homeless youth may actually expand over time, as these

interventions encourage peer leaders to seek out opportunities to talk with homeless youth they may not otherwise know.

Given these network features, it is evident that models of information diffusion are needed that draw from more realistic assumptions about where and how information flows in networks of “hard to reach” populations like homeless youth. To these ends, research has been directed toward augmenting diffusion model algorithms to account for more realistic network diffusion conditions. For example, efforts have been made to incorporate link prediction techniques into ICM-based influence maximization algorithms (Yadav et al., 2016), to parameterize the effects of multiplex ties (Brummitt et al., 2012; Gui et al., 2014; Yağan and Gligor, 2012; Zhong et al., 2017) and to account for evolving networks (Gayraud et al., 2015; Lahiri et al., 2008).

Meanwhile, others have developed altogether new models of diffusion based on a more relaxed set of graph assumptions. The recent work of Hu et al. (2017) presented the Activation Jump Model (AJM), which describes information diffusion from a multi-agent team perspective. The AJM differs from the leading cascade and threshold models in two key ways: 1) nodes exchange information beyond their immediate social ties (i.e., off edge), and 2) seed nodes (i.e., individuals chosen as peer leaders in a peer-led network intervention) act as a multi-agent team to spread information, where their overall influencing efficacy is a function of both individual and team attributes. Unlike the ICM and LTM, the AJM does not rely on strong tie assumptions, and instead is driven by activating the so-called ‘Breakfast Club’, where individuals from different social contexts form a united team to diffuse information. Using data collected from a pilot study on the spread of HIV prevention awareness in the networks of homeless youth, the AJM was shown to outperform the ICM and LTM in its fit to observed node-level influence data (Hu et al., 2017).

In this paper, we present an alternative approach to address the realities of information diffusion in real-world physical settings. Rather than augment the cascade and threshold algorithms or develop new ones altogether, we approach the problem as a data challenge. Drawing from data collected from a peer-led HIV prevention network intervention for homeless youth, we begin with a replication of Hu et al.’s (2017) study design, comparing the performances of the cascade (ICM), threshold (LTM), and activation jump (AJM) models in predicting node-level influence in a monoplex cross-sectional network among study participants. Here, network edges represent *who knows and communicates with whom* at the onset of the intervention. We consider this the baseline analysis, as it reflects the status quo approach that makes no revisions to the underlying graph. Then, we compare each diffusion model’s performance as the underlying graph is manipulated with observed data in three ways. First, to see how each information diffusion model performs when the assumption of closed networks is relaxed, we introduce observed ties to third party (i.e., non-participant) alters to the baseline monoplex network described above. Second, to address the possibility that another type of relationship serves as the setting for diffusion among homeless youth, we generate two additional networks: one represents their close friendships, and the other their joint participation in youth center programs. Third, to attend to the issue of network dynamism, we combine observed interactions at three time points to create a single meta-network for each of the three relational contexts. Although these data manipulations are incremental, we anticipate that they will yield insights about information diffusion in the real-world physical networks of homeless youth and will help inform the use of these models to plan for and evaluate more impactful social network interventions in similar communities.

2. Methods

2.1. The network intervention

2.1.1. Research design and sampling

Have You Heard? (HYH) is a 12-week behavioral intervention that

focuses on peer-led HIV prevention among homeless youth (Rice et al., 2018). Since homeless youth can be difficult to identify and engage with, the intervention is designed to be delivered in homeless youth drop-in centers, safe havens where homeless youth access food, clothing, and case management services like employment training, housing assistance, and health care referrals (De Rosa et al., 1999). Thus, each deployment of the intervention targets a network of homeless youth who access services from the same drop-in center. All data collection and study procedures implicated in this study received ethics approval from the University of Southern California Institutional Review Board (IRB).

In this study, data are acquired from the deployment of the HYH intervention in a drop-in center located in Los Angeles, California, USA. All youth receiving services at the drop-in center were eligible to participate and were informed of the study as they entered the drop-in center. In total, 72 youth between the ages of 16 and 25 were successfully recruited and enrolled. All participants provided written consent at the time of enrollment, including minors. A waiver of parental consent was obtained from the IRB for participants under the age of 18.

2.1.2. Peer leader selection

Social network data collected from participants were used as the basis for selecting peer leaders. Once enrolled, all participants were asked to indicate which youth at the drop-in center (including both study participants and non-participants) they knew and interacted with. Using an algorithm called DOSIM, which optimizes network-based intervention strategies for health providers by assuming a model of information spread based on a generalization of the Independent Cascade (Yadav et al., 2018), a total of 12 peer leaders were selected from this network. Theoretically, these are the individuals who are best situated in the network to maximize influence in the network.

2.1.3. Intervention design and delivery

The primary intervention training consisted of a 4-hour (one half-day) small group workshop (up to 5 participants) that aimed to educate peer leaders about sexual health risk reduction and promote their personal development. This initial training was then supported by 7 weeks of 30-minute follow-up check-in sessions, which focused on positive reinforcement of peer leaders' successes in engaging peers in HIV prevention conversations, problem-solving strategies to improve future conversations, and setting goals for the week with respect to peer-to-peer conversations about HIV prevention. Because of the transience of this youth population, the check-in schedule was flexible, and peer leaders could check in individually with the facilitator via phone or text. All peer leaders checked in at least once; modal attendance was five sessions (Rice et al., 2018). Peer leaders were compensated \$60 for attending the initial training and an additional \$20 for each follow-up check-in session they attended.

2.1.4. Assessment

All participants including peer leaders were assessed at three time points: a baseline interview; a 1-month follow up administered immediately after peer leaders were trained and deployed; and a final follow-up interview 3 months after baseline. Participants received a \$25 monetary incentive for each assessment they completed. Assessments were computer-based self-administered surveys: baseline (n = 72), 1-month (n = 50), and 3-months (n = 42). Survey items focused on demographics, access to health care, sexual health history, sex behaviors, knowledge about HIV and other sexually transmitted infections, and their drop-in center engagement. Participants were also asked questions about features of their personal networks and their interactions with designated peer leaders.

2.1.5. The outcome measure

Given what peer leaders were trained to do (i.e., to have

conversations about HIV prevention strategies with their peers), our evaluation of the intervention's impact focused on the extent to which HIV-related information was received by study participants from peer leaders. In the absence of a viable method for empirically tracing paths of information propagation as they occur in real-time in a physical-world environment, we draw on the participants' self-reports of having had an HIV-related conversation with a peer leader. Specifically, in the 1-month and 3-month follow-up assessments, participants were asked to indicate whether or not they had a conversation about HIV with a peer leader since the last assessment. In our analysis, we treat this confirmation as an instance of information propagation, although we remain agnostic to whether that confirmation was reported at the 1-month or 3-month follow-up. On a more substantive level, these conversations were expected to translate into shared information about topics like HIV and STD testing (e.g., the importance of testing, where they can go to get tested), condom use, modalities of contraception, and the HIV prevention pill PrEP (e.g., what it is, how it works).

We acknowledge that this measure of information propagation is imperfect, as it only captures primary diffusion moments between peer leaders and study participants, leaving secondary moments of diffusion between study participants unaddressed. That being said, we know from exploratory analysis not shown here that study participants who reported a conversation with a peer leader, subsequently reported having conversations about HIV and STD testing, condoms, and the HIV prevention pill PrEP with more of their close friends (see section 2.2.2 for description of close friendships) than participants who did not report having a conversation with a peer leader. Thus, there is reason to believe that being on the receiving end of the first wave of diffusion is associated with being on the propagating end of a subsequent wave.

2.1.6. Ethical considerations

Although there are no formal guidelines for the ethical conduct of research with homeless youth in particular, we drew on established legal and ethical guidelines for conducting research with adolescents and vulnerable populations in general. Three ethical issues that emerged in this study included questions about research consent, research incentives, and handling adverse events. In the case of consent, all participants provided written consent, including minors for whom we obtained a waiver of parental consent from the IRB. With respect to incentives, the authors' prior experiences working with young people experiencing poverty and/or housing instability informed our belief that reasonable (but not too high to be coercive) monetary incentives would be more effective for increasing retention and would be considered more straightforward to use than gift cards or vouchers. Finally, homeless youth are particularly susceptible to certain adverse events that may occur during the intervention, for example expressions during staff interviews of suicidal intentions or of physical abuse. To ensure participants' safety and well-being, an emergency protocol was developed for handling these events, including reporting procedures and quarterly staff training workshops.

2.2. Network generation

We modeled the propagation of HIV-related information in three distinct social networks among the homeless youth in our sample. In what follows, we describe how we derived these networks using data collected from each participant.

2.2.1. General social network

A General Social Network was generated at each time point by asking participants to name the other youth at the drop-in center that they know and communicate with. Names were generated using free recall responses and could include both study participants and non-participants. Two versions of this network were generated and used in our analysis. The first version is a network comprised of study participants only, which we take as our launch point and naive comparator for

much of the data augmentation experiment. The second version of this network includes ties to third party (i.e., non-participant) alters, which we introduce in the first stage of our experiment.

2.2.2. Close friendship network

The Close Friendship Network is a conglomeration of participants' egocentric networks of people they feel close to. In each survey assessment, participants were asked to name up to 5 close friends, defined as the people in their lives that they feel closest to and talk to most frequently. Participants were permitted to name family or friends they knew from home as well as other people they met on the street, including the other youth in the study. Although participants were permitted to name third party alters as they were for the General Social Network, the incredible variability in who those third party alters were for each participant made it impossible to match the third party alters across egos. For this reason, we opted to restrict each egocentric network to include only observed close friendship ties to other study participants and, then, matched participant alters across egos to create a single sociocentric network.

2.2.3. Program affiliation network

The Program Affiliation Network was generated from program participation data. Self-reports of each participant's engagement in nine drop-in center activities yielded a two-mode network of homeless youth and programs with ties between them representing program participation. Across the nine programs, the number of study participants who attended a program ranged from 7 to 34 with an average of 18.2 study participants per program. Among the 58 (of 72) study participants who took part in at least one program, the number of programs they participated in ranged from 1 to 8 with an average of 2.83 programs per person. Analysis is performed on a transformation of this two-mode network, which converts its two-mode structure into a one-mode dichotomized network of study participants whose ties to one another represent joint participation in at least one drop-in center program.

2.2.4. Temporal aggregations

We account for the temporal aspects of each of the three network contexts by collapsing observed ties over time into three meta-networks. We take this approach, as opposed to treating temporality as dynamic change for the following reason. The way in which ties in the General Social Network and Close Friend Network were ascertained — i.e., through free response and a five-person name generator, respectively — means that any observed changes in these networks could in fact be a product of the inherent biases of these network generation techniques. Name generation techniques that rely on a respondent's self-reports often result in significant under-reporting of relevant contacts (Brewer, 1993), due to recall biases related to forgetting (Brewer et al., 1999), respondent fatigue (Krosnick, 1991), satisficing (Dillman et al., 2002), and alter anonymity, as might be the case when enumerating contacts among homeless youth who may have tenuous relations with one another. Consequently, it may be misleading to classify a missing connection at one time-point as truly absent and, therefore, as an improbable path of information flow in edgewise diffusion models like ICM and LTM. For example, individuals may provide information on salient contacts in a follow-up assessment that were simply forgotten or omitted for other reasons in a prior wave.

For this reason, we examine the effect of combining observed ties at three time points, separated by relatively brief spans of time, in the hope that such an analysis will provide a more complete approximation of network ties that present plausible pathways of diffusion at any given time during the intervention. Specifically, we created two additional meta-networks for each network context, one representing the union of ties observed at baseline and 1-month and the other representing the union of ties at baseline, 1-month, and 3-months. This yielded three networks in total for each network context, where ties in each network represented a general, close friendship, or program affiliation

relationship observed at: (1) baseline only, (2) baseline or 1-month, and (3) baseline or 1-month or 3-months. With these versions of each network, we compare model performance on our temporally aggregated networks to the performance on the non-aggregated baseline network to determine which approach provides greater predictive power.

2.3. Model specifications

2.3.1. Independent cascade model (ICM)

In the ICM, an initial set S of active (or 'seed') nodes v is established and the process of diffusion unfolds in discrete steps according to a randomized rule (Kempe et al., 2003). Nodes u that are active in step t are given a single chance to activate each currently inactive neighbor w according to independent probability p , which can be chosen upon model initialization. Any nodes lacking a path to a seed (i.e. isolated nodes) have a zero probability of being informed. If w has multiple newly activated neighbors, their attempts occur in random order. Whether or not u succeeds in activating w , it cannot attempt to activate w in subsequent rounds. This process runs until no more activations are possible in the network.

2.3.2. Linear threshold model (LTM)

In the LTM, all nodes u are assigned a random threshold θ_w from a uniform distribution $U[0-1]$, and an initial set of seed nodes S is established. A node u is influenced by each of its neighbors v , such that each edge (u, v) is assigned a weight $b_{u,v} = \frac{w}{\text{deg}(v)}$, where we randomly vary w up to the total number of neighbors for each given node. A node u can only become active once a certain fraction of its neighbors is active (i.e. $b_{u,v}\theta_w$). Take this example: if an individual has connections to 10 others in the network, and it is randomly assigned a threshold $\theta_w = 0.2$, that individual's randomly assigned w value must be greater or equal to 2 for it to be activated.

The diffusion process in the LTM unfolds deterministically in discrete steps, such that in step t , all nodes that were active in step $t - 1$ remain active and act to influence their neighbors (Kempe et al., 2003). The process runs until no more activations are possible in the network. Similar to the ICM, nodes without a path to a seed have a zero probability of being informed under the LTM.

2.3.3. Activation jump model (AJM)

In contrast to the ICM and LTM, the AJM features two characteristics that are aligned with real-world information diffusion. First, information flow is not constrained to the edges within a network: nodes can 'jump' outside of their immediate social neighborhood to contact and propagate information to other nodes. Second, information is spread through nodes banding together to act as a multi-agent team, regardless of their social contexts, such that both individual and team attributes influence the diffusion efficacy.

In order to reflect the heterogeneous social disposition of active nodes, the AJM differentially models each influencer's jump position. While we do not replicate the full mathematical details of the AJM here, we refer the reader to Hu et al.'s (2017) paper, which presents the complete generalized and specific forms of the model.

In brief, a seed node's jump activity occurs in two stages. The first 'activation' stage determines the number of other nodes a seed node will attempt to influence during the diffusion period. In this stage, each node in the seed set $v \in S$ is initialized by

drawing an *activation level* A_v from a parameterized distribution ($A = v, S$) with mean $\mu_v = h(S)[a^T x_v]$, where x_v is the node's attribute vector with coefficients a^T . The maximum activation level of the node is represented by $a^T x_v$, which is modulated by $h(S) \leq 1$, a term that describes the activation level of the team as a function of the structural positions of all nodes in S . In the specific form of the model, h captures the negative effect of social homogeneity (or structural equivalence) among the seed nodes on a team's effectiveness (Hu et al., 2017).

Including the h parameter essentially ensures that high-degree nodes are more active in spreading influence, as long as they are part of a structurally diverse team.

In the second ‘jump’ or ‘deployment’ stage, the seed set is deployed in the network and the social influence process unfolds over time. This is largely controlled by a *landing distribution* $L_{v,T}$, a probability distribution expressing to which inactive nodes a particular activated node will jump. The landing probability is a function of the structural attributes of the influencing seed v and particularly targeted node u , such that $L_{v,T} = \frac{1}{d(v,u)}$, where $d(v, u)$ is the path-length distance between the two nodes. The jump stage occurs over the time interval $[0, 1]$, whereby each seed node v draws a series of jump times $t_1^v \dots t_{A_v}^v$ from a uniform distribution over the time interval. At each jump time t_i^v , v selects a target node u drawn from $L_{v,T}$, where T is the set of uninformed nodes at time t . Target node u is successfully influenced with probability p , which can be changed in the AJM much like in the ICM.

Together, the two main stages of the AJM describe how often and to whom an influencing node exchanges information as it moves through the network.

2.3.4. Parameter settings

We set most model parameters equally across all networks, based on standard values used by Hu et al. (2017) for similar experiments. Propagation probabilities p were always set to 0.1 in the ICM and AJM, and edge weights were set to 100 in the LTM. We conducted preliminary sensitivity testing by varying ICM and AJM’s propagation probabilities between 0.1 and 0.9, and LTM’s edge weights between 10 and 1000. We found no significant differences in the magnitude of total influence spread under these various conditions, so adopted Hu et al.’s (2017) original values for the sake of continuity and comparability. Following Hu et al. (2017), an additional small constant, $\emptyset(v, u) = 0.1$, was set in the landing distribution score to account for cases where there is no path between two nodes. This ensures that a ‘jump’ can occur between two nodes even when there is no direct path between them.

For AJM, we calculated the value for h (the term describing the negative effect of social homogeneity (or structural equivalence) on a team’s effectiveness, which ranges from 0 to 1) separately for each of the three networks and the three time periods. This is because h is required as a specific input when running the AJM model for each network. In the authors’ publicly available code, h is calculated in a module outside the main AJM model, by mapping the exact overlaps between the social networks of different Peer Leaders. This provides a mathematical representation of structural equivalency in the network, ranging from 0 (low equivalency) to 1 (high equivalency). In our experiments, h ranged between 0.83 (for the General Social Network at baseline) and 0.94 (for the Close Friend Network at baseline), which is very similar to the h values reported by Hu et al. (2017).

Code for ICM, LTM, and AJM models is available in open-source format on Github, at <https://github.com/jeromemayaud/SocialNetworkInformationSpread>.

2.4. Measure of model performance

A common method for determining the performance of diffusion models is to compare the magnitude of total influence spread in simulations to empirical observations – for instance, by minimizing root-mean square error (RMSE) as a function of actual spread or recapitulating cascade sizes (Goyal et al., 2011). While such volume-based metrics are unproblematic when one is agnostic about who in the network is actually influenced, many applications require influence maximization within specific subsets rather than the entire population (Hu et al., 2017). We therefore performed a finer-grained analysis by evaluating and comparing AJM, ICM and LTM predictions on node-level influence using analysis methods commonly employed in the field of machine learning.

Using the peer leaders from the field experiments as seed nodes, we generated diffusion outcomes according to the AJM, ICM and LTM. Each model was treated as a binary classifier that outputs the predicted probability of each node becoming influenced. We ran each model 100 times per simulation to trace out Receiver Operating Characteristic (ROC) curves, which plot a classifier’s True Positive Rate (TPR) against its False Positive Rate (FPR). Along the curve, each point corresponds to a predictive threshold such that all nodes with a probability of being informed above (below) the threshold are classified as influenced (not influenced). This methodology has previously been used to evaluate the performance of information diffusion models (Goyal et al., 2011; Hu et al., 2017; Wang et al., 2013) and has been shown to be more appropriate for binary classification tasks than Precision-Recall curves (Provost et al., 1998). The integrated area under each ROC curve (AUROC, or AUC) provides an indication of classification performance (Fawcett, 2006). An AUC value of 1 represents a perfect classifier and a value of 0.5 represents a purely random classifier.

We are also interested in realistically capturing the network that produces an accurate prediction of diffusion. Unfortunately, there are no quantitative metrics that will allow us to directly compare the accuracy of our network model dynamics with reality without a considerable amount of additional data (such as on who spoke to whom about the information, and at what time, their prior relationship, and whether the information was understood). However, the AUC may also serve as an indirect measure of this, as a more realistic model of diffusion may be reasonably assumed to have higher predictive power (and therefore higher AUC).

In order to capture the inherent variability between simulations, we ran 10 simulations per network and per model, for which we calculated individual ROC curves using the *scikitlearn* package in Python (v.3.4). From these individual curves, we derived an average ROC by interpolating all curves over the full range of unique FPR values and calculating the mean of all the TPRs for each FPR value. We also used the *scikitlearn* package to calculate mean AUCs and associated standard deviations. Each mean ROC and AUC value we present in our analysis is thus calculated from a total of 1000 individual runs (10 simulations x 100 runs).

3. Results

3.1. Network descriptives

We first describe the structural features of each temporal combination of each social network context, including network size (i.e., number of network members excluding isolates), number of observed edges, and density (i.e., the proportion of possible edges that are actually observed in the network) (Table 1). At baseline, the General Social Network is comprised of 67 study participant network members with 201 observed edges among them, corresponding to a network density of 0.045. As new ties observed at 1-month and 3-months are added to the baseline network, network size remains about the same ($n = 68$) while the number of observed edges among network members increases to 385, corresponding to a network density of 0.085.

In the baseline Close Friend Network observed among study participants, there are 63 study participant network members, with 144 observed edges among them. This corresponds to an average of 2.3 close friend ties per participant and an overall network density of 0.037. When new close friend ties captured at 1-month and 3-months are both added to the baseline network, the number of observed edges increases to 230, which corresponds to an average of 3.4 close friendship ties per participant and an overall network density of 0.052.

Finally, the baseline Program Affiliation Network has 58 study participant network members who are connected to one another via 1962 observed joint-program participation ties, which corresponds to an average of 33.8 joint-program participation ties per participant and a network density of 0.593. As new program affiliation ties are added to

Table 1 Network-level descriptive statistics for each temporal aggregation of the General Social Network, Close Friend Network, and Program Affiliation Network.

Metrics	General Social Network			Close Friend Network			Program Affiliation Network		
	Baseline only	Baseline + 1 M	Baseline + 1 M + 3M	Baseline only	Baseline + 1 M	Baseline + 1 M + 3M	Baseline only	Baseline + 1 M	Baseline + 1 M + 3M
Network size (excludes isolates)	67	68	68	63	67	67	58	67	68
Number of observed edges	201	289	385	144	198	230	1962	2672	2963
Density	0.045	0.063	0.085	0.037	0.045	0.052	0.593	0.604	0.650

the baseline network, the number of observed joint-program participation ties increases to 2,963, corresponding to an average of 43.6 joint-program participation ties per participant and a network density of 0.650.

Given these structural dynamics, the main feature that distinguishes the General, Close Friend, and Program Affiliation networks (irrespective of their temporal layering) is density. Specifically, with a baseline density of 0.593, the Program Affiliation Network is about 13 and 16 times denser than the baseline General and Close Friend networks, respectively.

We also note that new relationships are being formed throughout the intervention period, as evidenced by the increase in observed edges when new relationships observed at 1-month and 3-month follow-ups are added to those that already exist at baseline. The contribution of these newly added ties to the baseline structure of each network is shown in Fig. 1, with the newly added ties colored in blue.

Finally, the degree of overlap and distinction between each pair of meta-networks is shown in Table 2. Although the Close Friend Network is the sparsest (or least dense) of the three networks, it shares the most dyads in common with the other networks. Specifically, 60 percent and 72 percent of the ties observed in the Close Friend meta-network are also found in the General and Program Affiliation meta-networks, respectively. In contrast, the Program Affiliation meta-network is the most distinct as it has the greatest number of unique ties relative to the General and Close Friend meta-networks.

3.2. Breakdown of model performance

We evaluate the overall performance of each model (ICM, LTM and AJM) using Receiver Operating Characteristic (ROC) curves and the corresponding Area Under Curve (AUC) metric. Here, we unpack these performance metrics across each phase of the data augmentation experiment.

3.2.1. Diffusion in the Baseline General Social Network (with and without third party alters)

We begin by replicating Hu et al.’s (2017) initial study design, comparing the performances of the cascade (ICM), threshold (LTM), and activation jump (AJM) models in predicting node-level influence in a monoplex cross-sectional network among study participants. Here, the network we model is the baseline General Social Network, which represents who knows and communicates with whom among study participants at the onset of the intervention. As is shown in Fig. 2 (left side), the performance of ICM, LTM, and AJM are all below 0.5, suggesting each model performs worse than random under these conditions. However, when we introduce observed ties to non-participant homeless youth to the participant only baseline network (Fig. 2, right side), the performance of AJM improves considerably, achieving better than random results.

Understanding why AJM improves with the addition of third party ties and not ICM and LTM requires some unpacking. Because we lacked information about whether non-participants received information from a peer leader, we chose to leave their information status unmodeled. In the face of that, AJM’s off-edge jumping mechanism may have improved its ability to predict which nodes received information precisely because it was better able to “work around” ties (and, therefore, alters) that presented ambiguity.

3.2.2. Diffusion in alternative relational contexts

Next, to determine whether there are alternative relational contexts in which diffusion seems more plausible among study participants, we introduced two new social networks — the Close Friend Network and the Program Affiliation Network. As is shown in Fig. 3 (left column), when modeling node-level influence in the baseline Close Friend Network (Fig. 3b) and Program Affiliation Network (Fig. 3c), all the models perform no better or worse than random, achieving only a marginal

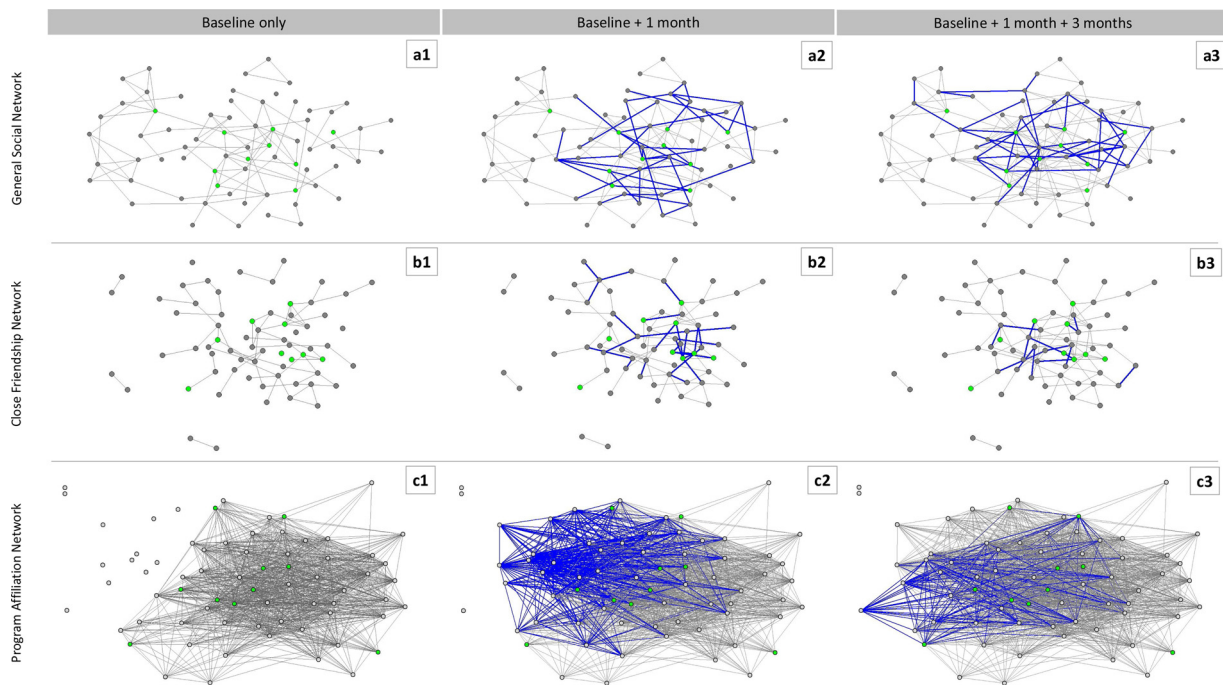


Fig. 1. Graph visualizations of the: General Social Network (top row), Close Friend Network (middle row), and Program Affiliation Network (bottom row). Each row contains three visualizations representing ties observed at three moments in time: baselines ties only (left column), baseline ties + newly formed ties at 1-month (middle column), and baseline + 1-month + newly formed ties observed at 3-months (right column). The newly formed ties at 1-month and 3-month are shown in blue. All networks at each time point show observed ties between homeless youth who were either study participants (shown in gray) or peer leaders (shown in green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Table 2

Number and percentage (reported in parentheses) of overlapping and unique dyads in the temporally aggregated General Social, Close Friend, and Program Affiliation meta-networks when compared to every other meta-network. Note: All comparisons are reported with respect to the meta-network specified in the first column (e.g., first row shows number and proportion of overlapping and unique dyads in the General Social Network when compared to the Close Friend and Program Affiliation networks.).

Main network for comparison	Number of matching node pairs with main network			Number of unique node pairs compared with main network		
	GSN	CFN	PAN	GSN	CFN	PAN
General Social (GSN)		136 (35 %)	256 (66 %)	249 (65 %)	129 (34 %)	
Close Friend (CFN)	136 (59 %)		166 (72 %)	94 (41 %)		64 (28 %)
Program Affiliation (PAN)	256 (9%)	166 (6%)		2707 (91 %)	2797 (94 %)	

improvement over their ability to predict influence in the baseline General Social Network (Fig. 3a). This suggests that neither alternative relational context presents plausible pathways of influence when captured just at baseline.

3.2.3. Diffusion in multiple relational contexts with temporal aggregations

Finally, we examine what happens as additional information on social networks is incorporated through time (i.e. from baseline to follow-up at 3 months). As is shown in Fig. 3 (middle and right columns), the performance of both ICM and LTM improve in the General (Fig. 3a) and Close Friend (Fig. 3b) networks. This suggests that the inclusion of temporal data may be providing an additional socially valuable mechanism for information to diffuse through the networks. However, ICM and LTM do not improve with temporal aggregation on the Program Affiliation Network (Fig. 3c).

By contrast, there is no evidence that AJM improves with additional

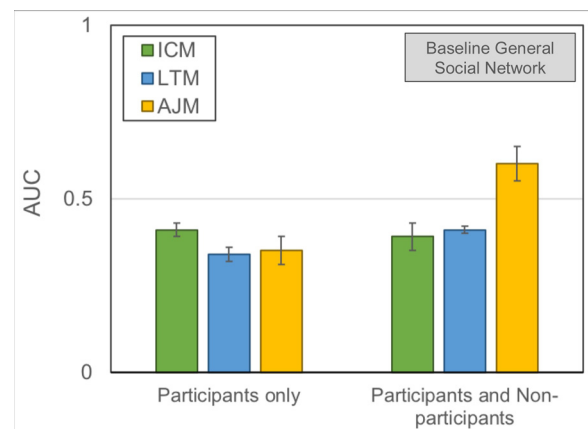


Fig. 2. Mean area under ROC curve (AUC) values for simulations using the three diffusion models (ICM, LTM and AJM) on the baseline General Social Network comprised of study participants only and the baseline General Social Network comprised of both study participants and non-participants. An AUC value of 1 represents a perfect classifier and a value of 0.5 represents a purely random classifier. The error bars represent the distance between one standard deviation above and below the mean.

time information on any of the three networks. This is perhaps to be expected: since AJM relies on predicting random jumps to nodes to infer diffusion on incomplete networks, making a network more complete may paradoxically result in AJM’s generated jumps being less useful (and maybe even counter-productive). The AJM performs less well than the ICM and LTM when fed with General and Close Friend network data but performs on par when fed with Program Affiliation network data. Since the jumping mechanism captures ‘off-edge’ propagation, AJM’s improved performance with the Program Affiliation Network may suggest that the direct affiliation ties registered in this network are not plausible diffusion pathways. This allows AJM to jump

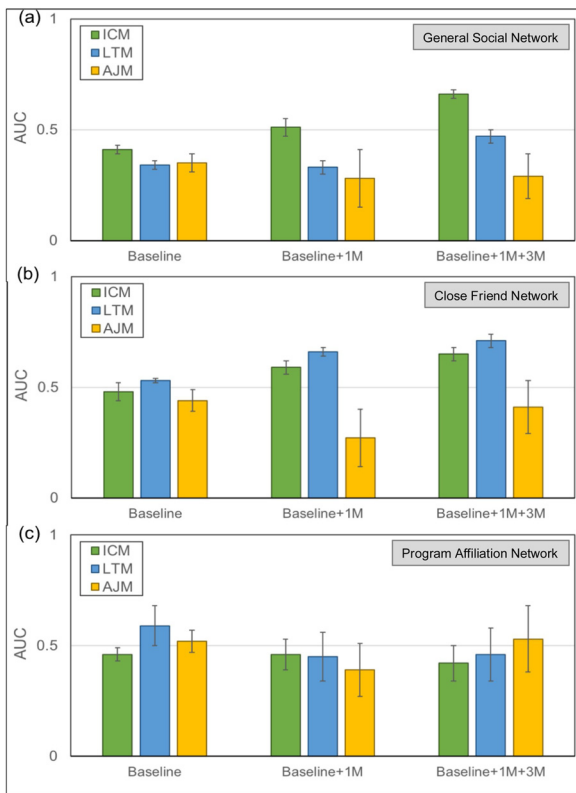


Fig. 3. Mean area under ROC curve (AUC) values for simulations using the three models (ICM, LTM and AJM) at different intervention time periods, grouped by network type. An AUC value of 1 represents a perfect classifier and a value of 0.5 represents a purely random classifier. The error bars represent the distance between one standard deviation above and below the mean.

more effectively between disconnected nodes in the network.

3.2.4. Diagnostics

To illustrate the dynamics of the false and true positives contributing to the AUCs, we present the individual and average ROC curves for model runs on the General Social Network, focusing on ICM and AJM to highlight the models with the largest contrast (Fig. 4). The curves were calculated using the *scikitlearn* package in Python, which automatically determines the most optimal thresholds for each ROC, and sometimes results in step-like behavior. The average curves do not cross (0,0) and (1,1) on the graphs because they are calculated by taking the mean across the full range of TPR values for each corresponding FPR value. While the ROC curves from individual runs exhibit

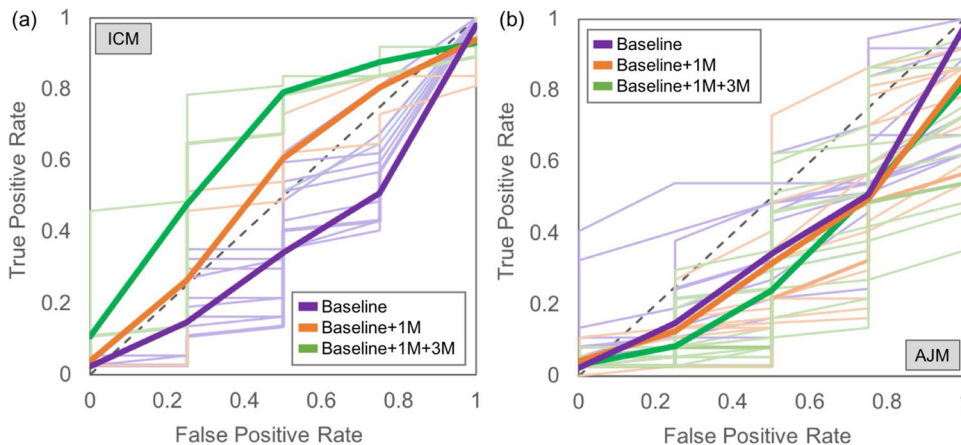


Fig. 4. Receiver Operating Characteristic (ROC) curves (false positive rate vs true positive rate), for two models run on the General Social Network at the three time points: (a) Independent Cascade Model (ICM), and (b) Activation Jump Model (AJM). Curves for individual runs are shown as thin lines and average curves are shown as thick lines.

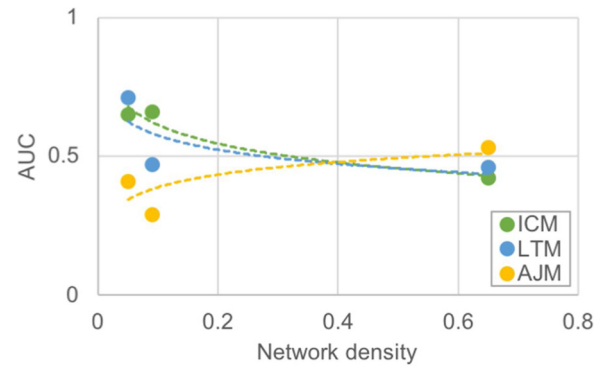


Fig. 5. Relationships between network density and Area Under Curve (AUC), when the three models (ICM, LTM and AJM) were run on the General Social, Close Friend, and Program Affiliation networks at Baseline + 1 M + 3 M.

considerable noisiness, the average curves for ICM (Fig. 4a) show that adding information over time improves model performance. The curves for AJM (Fig. 4b) display even more noisiness, owing to the probabilistic nature of the jumping mechanism, but do not show improvement over time, reflecting the AUC trends in Fig. 3.

Finally, we examine whether denser networks result in better model performance in general. We plot the relationships between network density and AUC when the three models were run on the three networks at Baseline + 1 M + 3 M (Fig. 5). We find effectively no relationship between network density and model performance. Similar lack of correlation was found for the Baseline and Baseline + 1 M data (not shown).

4. Discussion and conclusions

Cascade and threshold models remain widely used in predicting information propagation and influence maximization analyses due to their versatility and simple node-to-node edgewise dynamics. However, the assumptions about network topology built into most applications of these models — i.e., that networks are closed, monoplex, and static — may limit their ability to capture complexities in real-world diffusion processes, such as changes in relationships over time and uncertainty in network structure or node properties. The implication is that when cascade and threshold models are applied with these naïve assumptions about network topology, they are potentially more prone to error and inaccuracy in their predictions of information spread (Butts, 2003; Hu et al., 2017) in real-world settings.

While attempts have been made to address such deficiencies through algorithmic improvements like allowing for random jumps in the propagation process when the network structure is not completely known (Hu et al., 2017) or layering networks to capture effects of

different relational contexts (Gómez et al., 2013; Zhong et al., 2017), perhaps augmentation of the data itself can help improve model performance without any adjustment to the predictive algorithms. Investigation of this approach may not only improve modeled outcomes, but it may potentially inform data collection efforts to direct attention to areas where additional information is most needed.

In this paper, we tested this data-driven approach for handling network dynamism and relational multiplexity — two motifs we believe to be both common and consequential in many real-world networks — in the spread of HIV-related information in the networks of homeless youth. Specifically, our goal was to see whether adding additional ties observed at later time points and accounting for alternative social contexts in which information might spread yielded improvements in model performance.

Although the mean AUC values for most model runs were close to or less than 0.5, suggesting predictive performance no better than random, some noteworthy patterns do emerge. Our analysis confirms that existing models struggle when operating in real-world network conditions, particularly with the challenge of unobserved edges. We found that the Independent Cascade Model (ICM), Linear Threshold Model (LTM), and the recently developed Activation Jump Model (AJM) perform poorly when modeling diffusion among General Social ties, Close Friend ties, and Program Affiliation ties at baseline, despite AJM's design to address unobserved links in the data. However, when we additionally accounted for edges observed at one- and three-month follow-ups, both ICM and LTM performed better, indicating that missing edges may have driven their initial poor performance. Meanwhile, AJM performance declined as newly formed ties were added, as expected — its method of random jumps to account for unobserved links would become less accurate as gaps in each of these networks are filled in with edges from later time points. Fig. 4 demonstrates that this improved performance was not due to simply increasing network density; there is little relationship between density and performance when holding temporal information constant. This suggests that, unlike adding random links to increase network density, the additional one- and three-month relational information provided information pertinent to the information spread process.

While it initially may seem odd that including links formed in the future would improve prediction performance, this is not so strange on second glance: links observed within a three-month window may provide a good proxy for social proximity between individuals. It is reasonable to assume that the relationships confirmed at a given time point do not completely capture all the links between individuals in the dataset, as survey respondents may neglect to include all social ties in their responses or may speak to others in the community with whom they would not consider themselves to have strong ties. The dataset also omits third party individuals, who were named by multiple respondents as homeless youth they know and talk to but who were not enrolled in the study. Thus, our network likely has both missing links and nodes. Knowledge that two respondents will confirm having a social connection within the next three months may be an indicator that they are socially closer than if we knew they would not report any contact. While this is a loose proxy at best, with this temporal information the cascade and threshold models do seem to perform better than just random jumps (AJM).

Our findings also underscore the importance of establishing which types of relationships represent the most plausible pathways of influence, especially if edgewise propagation is assumed. Consider how ICM and LTM perform when modeling diffusion in the Close Friendship Network (Fig. 2). Although each model starts off at baseline performing no better or worse than random, as close friend ties observed at later time points are added to the network their performances improve, suggesting that these are indeed viable pathways of information propagation. Compare that to ICM's and LTM's performance in the Program Affiliation Network. Both models underperform, particularly as additional affiliation ties are added to the network. From this, we

conclude that program affiliation ties are not probable pathways of diffusion for homeless youth, and as more of these improbable ties are added to the network, the edgewise diffusion models perform worse. Alternatively, in the face of a network that presents implausible edgewise pathways of influence, a model like AJM which allows for off-edge diffusion may be better equipped by default to predict who becomes informed.

As such, results from our particular case study suggest that there is a hierarchy of information that is most optimal for predicting diffusion. If available, predictive models should ideally make use of information about social links between individuals through time. The biggest gains from such temporal aggregation occur when initial network densities are low. However, if temporal data are not available, the use of situational network data such as PAN may result in predictions of diffusion that are on par with, if not slightly better than predictions made using the other more explicitly social networks at baseline (i.e., the General and Close Friend networks). This may be attributable to the fact that at baseline the Program Affiliation Network is more densely connected and by chance accounts for many of the unobserved General and Close Friend ties at baseline that were eventually confirmed at one- and three-month follow-ups.

From a practical perspective, this is a noteworthy finding. Collecting social network data in the field, as is done for most health behavior interventions, is costly in terms of time and resources and is often prone to the contingencies of imperfect data capture. Program affiliation networks, on the other hand, can be built from participation records that drop-in centers routinely maintain. As such, these data can typically be collected more rapidly and cheaply than traditional relational data. Similar future studies may opt to rely on collecting situational data to inform their selection of peer leaders as well as their approach to the network intervention itself.

We recognize that this analysis has many limitations, and much future work remains to be done. First, we assume that it is reasonable to combine edges from multiple cross-sections of time into single meta-networks without information on when these edges were actually established (or confirmed by the respondent). In future work, we hope to develop algorithms that can integrate link prediction to account for unobserved addition and dissolution of edges and nodes over time, to extend existing dynamic network models (Yadav et al., 2018). Explanatory models like Stochastic Actor-oriented Models (SAOMs) (e.g., as implemented in RSiena) also hold great promise, as they could be used to identify a wider array of propagation mechanisms that include both structural features of networks (and their dynamics) as well as external influences like the attributes of network actors. Second, although we examine three different relationships that exist among the homeless youth in our sample, we do not account for true network multiplexity in our models. In future work, the effects of multiplex relationships on information diffusion could be modeled by introducing simple edge weights to represent the number of different types of relationships between two actors or by explicitly modeling each relationship type as a unique layer in a multiplex network diffusion model (Zhong et al., 2017). Finally, by exclusively sampling youth from a single drop-in center environment, we risk over-representing the role of center-based networks in the information diffusion process. Although this was inevitable given that the target audience for the *Have You Heard?* intervention was drop-in center patrons, we acknowledge that our approach ignores relationships that exist outside the center that may play a critical role in the diffusion of HIV-related information.

In general, we believe that our findings call for greater emphasis on using real-world physical network data to test and verify assumptions built into predictive models of diffusion. This process not only identifies the difficulty of modeling realistic network dynamics, but also emphasizes the challenge inherent in modeling noisy, incomplete network data that needs to be met before such models can be confidently used for informing policy. We hope that with greater usage of real-world data, we can identify more heterogeneous mechanisms of diffusion and

develop a more sophisticated and informed understanding of the factors that will increase the effectiveness of peer-led social network interventions in vulnerable communities at-risk for HIV and other health risk outcomes.

Acknowledgements

This work was supported in part by grants from the California HIV/AIDS Research Center and Schmidt Futures. We would like to acknowledge Bryan Wilder and Lily Hu for their time and effort in developing the Python code used in this analysis. We would also like to thank *My Friends Place*, the drop-in center where the “Have you Heard?” intervention was implemented, and their youth patrons who participated in the study.

References

- Arnold, E.M., Rotheram-Borus, M.J., 2009. Comparisons of prevention programs for homeless youth. *Prev. Sci.* 10 (1), 76–86.
- Barman-Adhikari, A., Bowen, E., Bender, K., Brown, S., Rice, E., 2016. A social capital approach to identifying correlates of perceived social support among homeless youth. *Child Youth Care Forum* 45 (5), 691–708.
- Brewer, D.D., 1993. Patterns in the recall of persons in a student community. *Soc. Networks* 15 (4), 335–359.
- Brewer, D.D., Garrett, S.B., Kulasingam, S., 1999. Forgetting as a cause of incomplete reporting of sexual and drug injection partners. *Sex. Transm. Dis.* 26 (3), 166–176.
- Brummitt, C.D., Lee, K.-M., Goh, K.I., 2012. Multiplexity-facilitated cascades in networks. *Phys. Rev. E* 85 (4), 045102-1–045102-4.
- Butts, C.T., 2003. Network inference, error, and informant (in) accuracy: a Bayesian approach. *Soc. Networks* 25 (2), 103–140.
- Coates, J., McKenzie-Mohr, S., 2010. Out of the frying Pan, into the fire: trauma in the lives of homeless youth prior to and during homelessness special issue on homelessness in Canada. *J. Sociol. Soc. Welf.* 37, 65–96.
- De la Haye, K., Green Jr, H.D., Kennedy, D.P., Zhou, A., Golinelli, D., Wenzel, S.L., Tucker, J.S., 2012. Who is supporting homeless youth? Predictors of support in personal networks. *J. Res. Adolesc.* 22 (4), 604–616.
- De Rosa, C.J., Montgomery, S.B., Kipke, M.D., Iverson, E., Ma, J.L., Unger, J.B., 1999. Service utilization among homeless and runaway youth in Los Angeles, California: rates and reasons. *J. Adolesc. Health* 24 (3), 190–200. [https://doi.org/10.1016/S1054-139X\(98\)00081-0](https://doi.org/10.1016/S1054-139X(98)00081-0).
- Dillman, D.A., Eltinge, J.L., Groves, R.M., Little, R.J.A., 2002. Survey nonresponse in design, data collection, and analysis. *Survey Nonresponse*. pp. 3–26.
- Ennett, S.T., Bailey, S.L., Federman, E.B., 1999. Social network characteristics associated with risky behaviors among runaway and homeless youth. *J. Health Soc. Behav.* 40 (1), 63–78. <https://doi.org/10.2307/2676379>.
- Falci, C.D., Whitbeck, L.B., Hoyt, D.R., Rose, T., 2011. Predictors of change in self-reported social networks among homeless young people. *J. Res. Adolesc.* 21 (4), 827–841.
- Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recognit. Lett.* 27 (8), 861–874.
- Gangamma, R., Slesnick, N., Toviss, P., Serovich, J., 2008. Comparison of HIV risks among gay, lesbian, bisexual and heterosexual homeless youth. *J. Youth Adolesc.* 37 (4), 456–464.
- Gayraud, N.T.H., Pitoura, E., Tsaparas, P., 2015. Diffusion maximization in evolving social networks. *Proceedings of the 2015 ACM Conference on Online Social Networks* 125–135.
- Goldenberg, J., Libai, B., Muller, E., 2001. Talk of the network: a complex systems look at the underlying process of word-of-mouth. *Mark. Lett.* 12 (3), 211–223.
- Gómez, S., Díaz-Guilera, A., Gómez-Gardeñes, J., Pérez-Vicente, C.J., Moreno, Y., Arenas, A., 2013. Diffusion dynamics on multiplex networks. *Phys. Rev. Lett.* 110 (2), 028701. <https://doi.org/10.1103/PhysRevLett.110.028701>.
- Goyal, A., Bonchi, F., Lakshmanan, L.V.S., 2011. A data-based approach to social influence maximization. *Proc. VLDB Endowment* 5 (1), 73–84.
- Granovetter, M., 1978. Threshold models of collective behavior. *Am. J. Sociol.* 83 (6), 1420–1443. <https://doi.org/10.1086/226707>.
- Gui, H., Sun, Y., Han, J., Brova, G., 2014. Modeling topic diffusion in multi-relational bibliographic information networks. *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management* 649–658.
- Guille, A., Hacid, H., Favre, C., Zighed, D.A., 2013. Information diffusion in online social networks: a survey. *Acm Sigmod Rec.* 42 (2), 17–28.
- Halcón, L.L., Lifson, A.R., 2004. Prevalence and predictors of sexual risks among homeless youth. *J. Youth Adolesc.* 33 (1), 71–80.
- Hawkins, R.L., Abrams, C., 2007. Disappearing acts: the social networks of formerly homeless individuals with co-occurring disorders. *Soc. Sci. Med.* 65 (10), 2031–2042.
- Heckathorn, D.D., 1997. Respondent-driven sampling: a new approach to the study of hidden populations. *Soc. Probl.* 44 (2), 174. Retrieved from. <http://socpro.oxfordjournals.org/content/44/2/174.abstract>.
- Hu, L., Wilder, B., Yadav, A., Rice, E., Tambe, M., 2017. Activating the “breakfast club”: Modeling Influence Spread in Natural-World Social Networks. *arXiv preprint arXiv:1710.00364*.
- Kelly, J.A., St Lawrence, J.S., Diaz, Y.E., Stevenson, L.Y., Hauth, A.C., Brasfield, T.L., et al., 1991. HIV risk behavior reduction following intervention with key opinion leaders of population: an experimental analysis. *Am. J. Public Health* 81 (2), 168–171.
- Kempe, D., Kleinberg, J., Tardos, É., 2003. Maximizing the spread of influence through a social network. Paper Presented at the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- Kennedy, D.P., Hunter, S.B., Osilla, K.C., Maksabedian, E., Golinelli, D., Tucker, J.S., 2016. A computer-assisted motivational social network intervention to reduce alcohol, drug and HIV risk behaviors among Housing First residents. *Addict. Sci. Clin. Pract.* 11 (4), 1–13.
- Kipke, M.D., Montgomery, S.B., Simon, T.R., Unger, J.B., Johnson, C.J., 1997. Homeless youth: drug use patterns and HIV risk profiles according to peer group affiliation. *AIDS Behav.* 1 (4), 247–259.
- Krosnick, J.A., 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Appl. Cogn. Psychol.* 5 (3), 213–236.
- Lahiri, M., Maiya, A.S., Sulo, R., Wolf, T.Y.B., 2008. The impact of structural changes on predictions of diffusion in networks. In *ICDMW’08. IEEE International Conference on Data Mining Workshops* 939–948.
- Latkin, C.A., Sherman, S., Knowlton, A., 2003. HIV prevention among drug users: outcome of a network-oriented peer outreach intervention. *Health Psychol.* 22 (4), 332–339.
- Marshall, B.D.L., Shannon, K., Kerr, T., Zhang, R., Wood, E., 2010. Survival sex work and increased HIV risk among sexual minority street-involved youth. *J. Acquir. Immune Defic. Syndr.* 53 (5), 661–664. <https://doi.org/10.1097/QAI.0b013e3181c300d7>.
- Melander, L.A., Tyler, K.A., 2010. The effect of early maltreatment, victimization, and partner violence on HIV risk behavior among homeless young adults. *J. Adolesc. Health* 47 (6), 575–581. <https://doi.org/10.1016/j.jadohealth.2010.04.010>.
- Provost, F.J., Fawcett, T., Kohavi, R., 1998. The case against accuracy estimation for comparing induction algorithms. *ICML* 98, 445–453.
- Rew, L., Whittaker, T.A., Taylor-Seehafer, M.A., Smith, L.R., 2005. Sexual health risks and protective resources in gay, lesbian, bisexual, and heterosexual homeless youth. *J. Spec. Pediatr. Nurs.* 10 (1), 11–19.
- Rice, E., Milburn, N.G., Rotheram-Borus, M.J., 2007. Pro-social and problematic social network influences on HIV/AIDS risk behaviours among newly homeless youth in Los Angeles. *AIDS Care* 19 (5), 697–704.
- Rice, E., Barman-Adhikari, A., Milburn, N.G., Monro, W., 2012a. Position-specific HIV risk in a large network of homeless youths. *Am. J. Public Health* 102 (1), 141–147.
- Rice, E., Tulbert, E., Cederbaum, J., Barman Adhikari, A., Milburn, N.G., 2012b. Mobilizing homeless youth for HIV prevention: a social network analysis of the acceptability of a face-to-face and online social networking intervention. *Health Educ. Res.* 27 (2), 226–236.
- Rice, E., Yoshioka-Maxwell, A., Petering, R., Onasch-Vera, L., Craddock, J., Tambe, M., et al., 2018. Piloting the use of artificial intelligence to enhance HIV prevention interventions for youth experiencing homelessness. *J. Soc. Social Work Res.* 9 (4), 551–573.
- Schneider, J.A., Zhou, A.N., Laumann, E.O., 2015. A new HIV prevention network approach: sociometric peer change agent selection. *Soc. Sci. Med.* 125, 192–202.
- Solorio, M.R., Rosenthal, D., Milburn, N.G., Weiss, R.E., Batterham, P.J., Gandara, M., Rotheram-Borus, M.J., 2008. Predictors of sexual risk behaviors among newly homeless youth: a longitudinal study. *J. Adolesc. Health* 42 (4), 401–409. <https://doi.org/10.1016/j.jadohealth.2007.09.023>.
- Sutherland, S.P., 2016. *Transitional Planning and Homelessness of Youths Emancipated From the Child Welfare System* [Doctoral Dissertation]; Walden University.
- Sweeney, P., Lindegren, M.L., Buehler, J.W., Onorato, I.M., Janssen, R.S., 1995. Teenagers at risk of human immunodeficiency virus type 1 infection: results from seroprevalence surveys in the United States. *Arch. Pediatr. Adolesc. Med.* 149 (5), 521–528.
- Tyler, K.A., 2009. Risk factors for trading sex among homeless young adults. *Arch. Sex. Behav.* 38 (2), 290–297.
- U.S. Department of Housing and Urban Development, 2019. HUD 2019 Continuum of Care (CoC) Homeless Assistance Programs Homeless Populations and Subpopulations. Retrieved from. https://files.hudexchange.info/reports/published/CoC_PopSub_NatlTerrDC_2019.pdf.
- Usborne, E., Lydon, J.E., Taylor, D.M., 2009. Goals and social relationships: windows into the motivation and well-being of “street kids” 1. *J. Appl. Soc. Psychol.* 39 (5), 1057–1082.
- Wang, Y., Shen, H.-W., Liu, S., Cheng, X.-Q., 2013. Learning User-specific Latent Influence and Susceptibility From Information Cascades. *arXiv preprint arXiv:1310.3911*.
- Yadav, A., Chan, H., Xin Jiang, A., Xu, H., Rice, E., Tambe, M., 2016. Using social networks to aid homeless shelters: dynamic influence maximization under uncertainty. *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems* 740–748.
- Yadav, A., Wilder, B., Rice, E., Petering, R., Craddock, J., Yoshioka-Maxwell, A., et al., 2018. Bridging the gap between theory and practice in influence maximization: raising awareness about HIV among homeless youth. *International Joint Conference on Artificial Intelligence (IJCAI)* 5399–5403.
- Yağın, O., Gligor, V., 2012. Analysis of complex contagions in random multiplex networks. *Phys. Rev. E* 86 (3), 036103-1–036103-10.
- Young, S.D., Cumberland, W.G., Lee, S.-J., Jaganath, D., Szekeres, G., Coates, T., 2013. Social networking technologies as an emerging tool for HIV prevention: a cluster randomized trial. *Ann. Intern. Med.* 159 (5), 318–324.
- Young, L., E., Schumm, P., Alon, L., Bouris, A., Ferreira, M., Hill, B., et al., 2017. PrEP Chicago: A randomized controlled peer change agent intervention to promote the adoption of pre-exposure prophylaxis for HIV prevention among young Black men who have sex with men. *Clin. Trials* 15 (1), 44–52. <https://doi.org/10.1177/1740774517730012>.
- Zhong, Y.D., Srivastava, V., Leonard, N.E., 2017. On the linear threshold model for diffusion of innovations in multiplex social networks. *Decision and Control (CDC), 2017 IEEE 56th Annual Conference* 2593–2598.