Final Project Report: The Impact of of Sexual Education Policies on Disease Transmission in Sexual Networks

Austin van Loon, Jessica Hinman, and Yuguan Xing

I. Introduction

In the United States, the topic of sexual education courses and programs in the public school system is a hotly debated issue. While evidence exists about individual-level effects of sexual education, no work (to our knowledge) has demonstrated the impact of sexual education on the structure of sexual networks. While such an impact is difficult to asses, sexual education isn't just an individual-level intervention with individual-level consequences; it is a school-level policy which may have impacts on the social structure of the school. The effects on the structure of the sexual network may be positive, negative, non-existent, or contingent on other attributes of the school. For policy-makers to make well-informed policy decisions, it is important that we understand all of the consequences of the policy at every level of analysis. Here we develop a methodology which utilizes empirical data in conjunction with an extension of the configuration model that allows us to answer previously un-approachable questions such as the consequences of sexual education on the structure of schools' sexual networks under various conditions. We find that under certain conditions, sexual education makes sexual networks less "infectable", whereas under other conditions it makes it more "infectable" when compared to schools with no sexual education. Important moderating variables include school size and sex ratio.

II. Related Work

Impact of sexual education on individual behavior

Kirby et al. (2007) accumulated evidence from many different studies examining the impact of different sexual education programs on individual sexual behavior and outcomes, such as unwanted pregnancy and STI infection [1]. All studies that were included in the report were RCTs which randomly assigned individuals, groups, or schools to a sexual education program or a control condition. Overall, the team found that comprehensive sexual education programs are largely effective at reducing the mean number of sexual partners adolescents have, as well as increasing condom usage. However, comparing means by condition only scratches the surface of how these programs may affect the sexual environment of the school in which they are implemented. From classic research in network science (e.g. Watts and Strogatz 1998; Granovetter 1973), we know that the density of a network is not the only property that is important for the diffusion of simple contagions; the structure of the sexual network formed by the aggregation of these behaviors can have a huge impact on diffusion dynamics, even holding the density constant.

Disease spread through sexual networks

Wylie et al. utilized routinely collected case information on chlamydia and gonorrhea infections in the province of Manitoba to evaluate transmission patterns through sexual networks [2,3]. Upon diagnosing a new case, demographic information on the case individual as well as their reported sexual contacts are entered into a computerized database. Public health nurses then follow up with the index cases contacts in order to perform further diagnostic testing, provide resources on prevention and treatment, and collect information about further contacts where appropriate. Using this information, researchers were able to reconstruct sexual contact networks and map them to the geography of Manitoba. In doing so, they identified two secondary structures of interest which they termed radial and linear components.

The identification of two different component structures allowed the researchers to determine underlying demographic and transmission differences as well. Individuals connected to radial components tended to cluster geographically, whereas those in linear network structures were often geographically distant from one another. Furthermore, endemic rates of gonorrhea were only observed within large linear components, suggesting the possibility that the structure and interactions within those components may be necessary to maintain its persistence. These data provide observational support for the hypothesis that information on network-level data provides a more complete picture of STI transmission than can be obtained from individual-level data alone, and suggests compelling targets for interventions in key structural nodes or components. While Wylie et al., as early adopters of the use of network analysis to evaluate sexual disease transmission patterns, advanced our understanding of how the structure of sexual contacts impacts disease spread, they employed a very rudimentary modeling strategy.

Measurement of the "infectability" of sexual networks

Bearman et al. introduce their findings about the structure of a sexual network within a high school in the midwestern region, which they refer to as "Jefferson High" [4]. This work is distinct in that the structure of the network was not simulated using statistics from egocentric surveys. Instead, the researchers managed to acquire data from adolescents on not only the number of sexual partners that they have, but also who those partners are. As a result, they produced an empirical network describing sexual interactions among high school-aged participants. They demonstrate that 1) the tendency of people to choose others that are similar to them as sexual partners, and 2) the tendency of people to avoid having sexual relationships with people who are reachable within several hops explains the unusual spanning-tree structure of the network, which was not easily explainable with extant models. The authors emphasized six different metrics concerning the structure of the network: density at maximum reach, network centralization, mean geodesic length, maximum geodesic length, skew of reach distribution, and number of cycles. They claim that these network characteristics are particularly relevant to the spread of sexually transmitted diseases, and thus that they represent the key areas of focus when evaluating the fitness of models to empirical data.

An evaluation of sexual network properties in a random sample of 2,810 Swedish adults by Liljeros et al. emphasized the importance of the distribution of the number of sexual partners [5]. The authors demonstrated that this distribution was scale-free rather than single-scale, which is compatible with a preferential attachment process. Scale-free networks are characterized by distributions which follow the power law, in which a few key nodes are highly connected but the majority of nodes exist on the periphery. These characteristics make them resistant to random failures, but susceptible to strategic attacks on the highly connected nodes. The authors note this as a crucial element that could be utilized in order to target highly connected individuals with sexual education efforts, and thereby decrease the susceptibility of the network as a whole to sexually transmitted infections.

Chakrabarti et al. took a broader view in an attempt to create a more generalized framework for evaluating disease spread through networks [6]. The authors proposed an epidemic threshold value, consisting of the inverse of the largest eigenvalue of a network's adjacency matrix. If the fraction of a contagion's birth rate over its death rate is below a network's epidemic threshold, it will be unable to propagate in the network. They argued that their proposed model is both general, in that it can be utilized across a wide variety of network structures, and precise, in that it improves upon the accuracy of other network-based epidemic models. While this necessitates a drastic reduction in the complexity of biological infection mechanisms in order to maintain generalizability and tractability, it adds immense value in allowing for the calculation of a measure of susceptibility that is an intrinsic property of the network itself. Particularly with respect to group-based interventions such as sexual education programs, the epidemic threshold condition proposed by Chakrabarti et al. provides a precise measure for evaluating the extent to which the intervention may have modified the underlying susceptibility of the network as a whole.

III. Data

The ideal data for studying the impact of sexual education on the structure of sexual networks would be data from a large randomized control trial where many schools are randomly assigned to either teach or withhold sexual education from their students, after which the structure of the sexual network of each school would be collected. This kind of data is unavailable due to not only the ethical concern of withholding sexual education from entire schools of children, but also because of concerns of feasibility. The next best kind of data would be observational data; a collection of the sexual networks of many randomly sampled schools along with many school-level covariates. To our knowledge, this data doesn't exist either, perhaps due to the immense cost and difficulty of pursuing such a data collection project. The best information we have about sexual behavior in schools with which we can seriously reason about the effect of sexual education on the structure of sexual networks is at the individual level.

For the purposes of this project, we have gained access to the restricted-use data from The National Longitudinal Study of Adolescent to Adult Health, or Add Health, project. Add Health was initiated in 1994 when it enrolled a nationally representative sample of adolescents in grades 7 through 12. It has continued to follow up with the initial student cohort, as well as their families and social groups, to the present day, with Wave V of the data collection process rolling out as of 2016. We utilized the data collected as part of Wave I of this study, which allows us to retain the maximum available sample size which was attenuated gradually over subsequent waves due to attrition. For each participant, we have access to their total number of sexual partners, whether their schools are forced by state law to teach various kinds of sexual education, and estimates of how often they use contraceptives when they do have sex.

In cleaning and preparing the data for analysis, we divided the participants into three "sexual education regimes": (1) individuals whose schools are required by state law to teach both "HIV prevention" as well as "STD prevention", (2) individuals whose school were not required by state law to teach either "HIV prevention" or "STD prevention", and (3) individuals whose schools were required to teach either "HIV prevention" or "STD prevention" but not both. We discarded individuals in sexual education regime (3) to provide the clearest distinction between exposure categories. We then examined the distribution of number of sexual partners amongst individuals in sexual education regimes (1) and (2). We refer to individuals in sexual education regime (1) as individuals who received sex ed (for convenience) and those in sexual education regime (2) as individuals who did not receive sex ed. It is worth noting explicitly that we do not have information about whether these respondents are in the same school, county, or even state, as we do not have access to this information under our current agreement with the owners of the data.

For establishing "number of sexual partners", we used individuals' response to the question "With how many people, in total, including romantic relationship partners, have you ever had a sexual relationship?". Self-report measures such as this are subject to biases including social desirability bias and differential respondent recall, but also represent the best data available regarding individual sexual behavior. Due to the nature of the data and concerns about potential re-identification, we are not allowed to share certain aspects of the data, including exact counts for cross-sections which contain less than a certain number of individuals. Due to these legal limitations, we have shared only proportions of individuals within sexual education regimes who have a certain number of sexual partners, and restricted our analyses to individuals with 10 or fewer sexual partners.

IV. Descriptive Statistics

While we cannot share the exact number of men and women that fall under each sexual education regime (as this in combination with the proportional information below might allow someone to recover potentially identifiable cross-tabular information), the table below shows that although the population sizes for individuals with and without sexual education are unequal, both are of a reasonable size for our analysis. As we look at the probability density functions reported in Figure 1 for both men and women in these different regimes, we see distinct differences in the curves, despite the fact that these differences don't seem drastic. Keep in mind that the area under each curve sums to one, so any differences in the curve at any one point must be made up somewhere else. Interestingly, sexual education appears to have different effects on men and women.

Men	1536	No sex ed	803
Women	1275	Sex ed	2008

Table 1. Basic descriptive statistics.

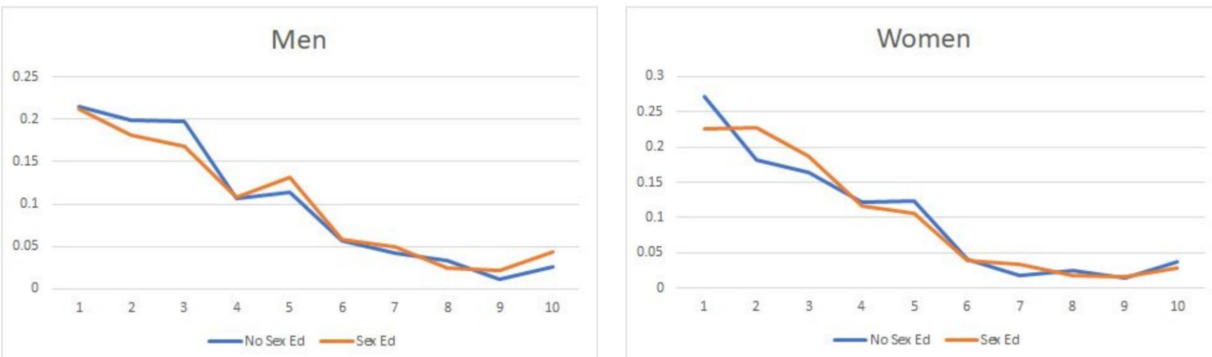

Figure 1. PDF of number of sexual partners by sexual education regime.

V. Model[1]

We use a probabilistic bipartite extension of the configuration model in order to test whether these different distributions generate networks with different structural characteristics. In our algorithm, each generated network is populated with "male" and "female" nodes according to a size and "sex ratio" parameter passed to the algorithm. Each node is assigned a number of "spokes" with probability respective to the appropriate PDF. We then randomly match spokes between male and female nodes together (like most research in this area we assume a fully heterosexual network) to create randomly generated networks which reflect the different degree distributions amongst men and women who were exposed to different sexual education regimes. Since men and women will not always have the same number of total spokes, we delete all spokes that are unmatched when either all male nodes or all female nodes have no available spokes.

The model is only valid under a number of undesirable assumptions. If we assume that individuals' degree is only a function of measured network-level characteristics (here we incorporate whether all individuals simulated into the network were exposed to sexual education) and individual characteristics which are uncorrelated with the treatment

[1] Code used to generate network available at https://github.com/yuguanx/CS224W_project

variable, then this allows us to test the effect of those network-level characteristics on the structure of the network. However, insofar as there are individual-level characteristics that are correlated with the treatment or interpersonal characteristics which affect individuals' degree, these may bias our results. Further, our algorithm does not take into account possible higher-order relational differences (e.g. motif prevalence) that may be caused by the treatment. It seems that the algorithm we develop here could be further proliferated and built upon to relax these assumptions, but we save this for future research.

After we simulate a network, we collect four measures about the structure of the largest component of the network. The first is the epidemic threshold, measured as $\frac{1}{\lambda_1}$, where λ_1 is the largest eigenvalue of the adjacency matrix representing the generated network. Our second measure of structure is the mean geodesic distance, measured as $\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}$, where n is equal to the number of nodes in the network and d_{ij} is the length of the shortest possible path between nodes i and j. Third, we measure the Freeman eigenvector centralization of the network, measured as $\frac{1}{n-1}\sum_{i=1}^{n}(e_{max} - e_i)$, where n is the number of nodes, e_i is the eigenvector centrality of node i, and e_{max} is the highest eigenvector centrality of any node. Lastly, we measure the GINI-based eigenvector centralization of the simulated networks, measured as $\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|e_i - e_j|}{2n\sum_{i=1}^{n} e_i}$, where e_i is a node's eigenvector centrality and n is the number of nodes in the network [7]. Each of these is some approximate measure of how prone the network is to an outbreak of an infection, though all approximate this in different ways. For each experiment, we simulate 6000 networks (3000 per "condition") and compare the distributions of these values with a simple independent-sample t-test.

VI. Results

Baseline model

As a first test of our model, we simulate networks of size 300 (approximately the size of the sexual network analyzed by Bearman et al) with 50/50 sex ratio. As you can see in the "Baseline" entry in Table 2, we find mixed evidence for how sexual education impacts the "infectability" of the school's sexual network. Specifically, we find that networks generated from the degree distribution of individuals exposed to sexual education had a lower GINI-based centralization and a lower mean geodesic distance than networks generated from the degree distribution of individuals not exposed to sexual education. Having a lower centralization is usually assumed to mean a less "infectible" network, while a lower mean geodesic usually means a more "infectible" network. These results are under a very specific, stylized model. To better understand the effect of sexual education on the structure of sexual networks, we perform various experiments where we alter various parameters of our model.

Investigating the role of short cycles, mean degree, and the tail

To understand the impact of short cycles on the structure of these networks, we forced all networks to have no cycles of length 4 or shorter. Bearman et al found that this simple principle in sexual networks accounts for many other otherwise quizzical structural features. To implement this in our algorithm, after each pair of spokes are matched, we test the network to see if there are any cycles of length 4 or shorter. If there are, the recently added edge is removed and paired with other spokes. This is computationally inefficient but extends the algorithm in a general way (any function that returns a Boolean value can replace the function which checks for short cycles). As you can see from the "Drop short cycles" entry in Table 2, we see similar results to the baseline model. Since short cycles don't seem to drastically affect our results, we run the rest of our experiments allowing for short cycles.

Experiment	Epidemic Threshold	GINI Centralization	Freeman Centralization	Mean Geodesic
Baseline		-***		-**
Drop short cycles		-***		-***
Match mean degrees	+*	-*		
Drop tail	-***	-***	-**	-***
Large school	+***	-***		+***
Small school	-***	-***	-***	-***
55% men, 45% women	+***			+***
45% men, 55% women	-***	-***	-***	-***

Table 2. Results of experiments on baseline model.
NOTE: a "+" means that networks simulated with the degree distributions associated with sexual education had a significantly higher value for the respective measure than those simulated with the degree distributions associated with no sexual education.
* $p<0.05$ ** $p<0.01$ *** $p<0.001$

As a further investigation into our data and algorithm, we forced each generated pair of networks (one with the information corresponding to individuals with sexual education, one with information corresponding to individuals without sexual education) to have the same number of edges by dropping random edges from the network with more edges until the number of edges in the two networks was equal. Doing this, we find that sexual education increases the epidemic threshold of networks and decreases the GINI centralization of the network. This suggests to us that while sexual education may create more dense networks, net of such effects sexual education creates less infectible networks. These are very preliminary results, however, as we should be matching the mean degree through more sophisticated means. For instance, future research should consider altering the degree distributions so as to equalize the means of the two distributions while minimizing the earth mover's' distance (EMD) between the original distributions and the transformed distributions.

When collecting our information from Add Health concerning the degree distributions of individuals in different sexual education regimes, we were legally bound to only provide information concerning individuals who have only had 10 or less sexual partners. In one experiment we "cut the tail" off of our distribution and only consider information from individuals with 9 or less sexual partners. As is reported in the "Drop tail" entry of Table 2, under this alteration to the model we find that sexual education decreases all of the measured variables in social networks, again giving us mixed results concerning whether or not sexual education decreases the infectability of a network. Instead of taking these results too seriously in and of themselves, we take the observed difference in these results from the results achieved with the baseline model as suggesting to us that more rigorous testing, such as considering all of the degree distribution as well as excluding individuals with lower degrees (e.g. only considering individuals with degree less than 5), may be necessary for future research.

Exploring size, sex ratio, and gender-specific education

To investigate the conditions under which sexual education may make sexual networks more or less infectable, we model larger networks (n=1500) and smaller networks (n=150). We find that in larger networks, sexual education leads to a significantly larger epidemic threshold and mean geodesic distance, but a significantly lower GINI centralization. In small schools, we find more mixed results; sexual education decreases both Freeman and GINI centralization but also decreases mean geodesic distance and the epidemic threshold. Thus, school size may be an important moderator on the impact of sexual education on the structure of sexual networks.

We further explore generating networks with different sex ratios. Remember that nodes representing men and women are drawn from their respective empirical distributions, and are not arbitrarily assigned genders. In networks generated with 55% male nodes and 45% female nodes, we find that networks generated from the degree distribution associated with individuals who were exposed to sexual education had a higher epidemic threshold as well as a higher mean geodesic distance. In a network with 55% female nodes and 45% male nodes, we find a decrease in all measured variables. Sex ratio also appears to an important moderating variable for assessing the effect of sexual education on the structure of sexual networks.

Complicating the Model: Condom Usage

Difference in the degree distribution of individual students is not the only likely impact of sexual education on the spread of STIs in schools. Perhaps what is even more important is the fact that sexual education (or at least so-called “comprehensive” sexual education programs) teaches students to use contraceptives such as condoms when having sex. To proliferate our simple model outlined above, we consider simulating condom usage in these generated networks. In order to do so, we calculate on average[2] how often individuals of each degree and of each demographic quadrant (sexual education/no sexual education; female/male) use condoms when having sex.

Weight assigned to ties representing sexual activity using condoms	Epidemic Threshold	GINI Centralization	Freeman Centralization	Mean Geodesic
0.01	+***	-***	-***	
0.02	+***	-***	-***	
0.03	+***	-***	-***	
0.04	+***	-***	-***	
0.05	+***	-***	-***	+**
0.25	+***	-***	-***	
0.50	+***	-***	-***	
0.75	+***	-***	-***	
0.90	+***	-***	-***	

Table 3. Results of experiments simulating condom usage.
NOTE: a “+” means that networks simulated with the degree distributions associated with sexual education had a significantly higher value for the respective measure than those simulated with the degree distributions associated with no sexual education.
* p<0.05 ** p<0.01 *** p<0.001

In this extension of the model, when a node is generating spokes, they produce “regular” spokes and “protected” spokes. Let’s say an individual node is drawn from the males with no sexual education degree distribution and it is determined he will have degree 4. This is also associated with an expected percentage of ties to be protected. If it were 25%, for instance, then they would generate 4 spokes, each of which has an independent probability of 25% of being a protected spoke. When we match spokes to each other, we only match protected spokes to other protected spokes (forming “protected ties”), match regular spokes only to other regular spokes (forming “regular ties”), and don’t allow both a tie formed by two protected spokes and a tie formed by two regular spokes between the same node. Protected ties are down weighted to some parameter for which we test various values. This weight represents how much less likely it is to spread an STI through sexual intercourse while using a condom in comparison to sexual intercourse without a condom.

[2] We use the average instead of the exact distribution so that we do not provide potentially identifiable cross-tabular information

As you can see in all of the models presented in Table 3, whether we try realistic values for the weight of protected ties (epidemiological work suggests that condoms, when used correctly, reduce the risk of spreading an STI by an average of 98%-99%, corresponding to a weight of 0.01-0.02) or conservative values, we find that the differences in condom usage across sexual education regime, in conjunction with the different degree distributions associated with individuals in the sexual education regimes, are far less infectable; they have higher epidemic thresholds, lower GINI and Freeman centralizations, and, in one case, a higher mean geodesic distance.

VII. Conclusions

Sexual education policy is hotly contested in the United States. We have strong evidence concerning its effects on individual behavior, but here we present (to our knowledge) the first evidence of its contentious and complicated impact on the structure of sexual networks. In order to examine this question, we developed a probabilistic and bipartite extension of the configuration model which we use in conjunction with nationally representative survey data. While of course these results are only preliminary, certain results are surprising and interesting.

First, we find that sexual education makes larger schools less infectable, but has mixed effects in smaller schools. Second, even in some of these smaller settings, if we force the mean degree of the networks to be the same, sexual education makes networks less infectable. Third, we find that sexual education has a much more beneficial impact in schools were there are significantly more men than women as opposed to schools where there are more women than men. Lastly, we find that the difference in the pattern of condom use between individuals exposed to sexual education as opposed to those who were not leads to sexual networks that are much less infectable.

References

1. Kirby, J., *Emerging Answers: Research Findings on Programs to Reduce Teen Pregnancy and Sexually Transmitted Diseases*. 2007, National Campaign to Prevent Teen and Unplanned Pregnancy 2007-11-11.
2. Wylie, J.L. and A. Jolly, *Patterns of chlamydia and gonorrhea infection in sexual networks in Manitoba, Canada.* Sex Transm Dis, 2001. **28**(1): p. 14-24.
3. Wylie, J.L., T. Cabral, and A.M. Jolly, *Identification of networks of sexually transmitted infection: a molecular, geographic, and social network analysis.* J Infect Dis, 2005. **191**(6): p. 899-906.
4. Peter S. Bearman, James Moody, and Katherine Stovel, *Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks.* American Journal of Sociology, 2004. **110**(1): p. 44-91.
5. Liljeros, F., et al., *The web of human sexual contacts.* Nature, 2001. **411**: p. 907.
6. Chakrabarti, D., et al., *Epidemic thresholds in real networks*. 2008.
7. Badham, J.M., *Commentary: Measuring the shape of degree distributions.* Network Science, 2013. **1**(2): p. 213-225.