**Unraveling HIV Transmission Networks: A Social Network Analysis Approach**

TNDY-336 Final Project

Kaijie "Yuri" Yu, Ahmad Alasmari, Edward Segura-Tinoco,

Denise Escarcega, Ghardan Alghureid

Project GitHub Link:

Contents

[**Introduction** 1](#_Toc134556939)

[**Literature Review** 1](#_Toc134556940)

[**Network Visualization** 5](#_Toc134556941)

[**Social Network Analysis** 7](#_Toc134556942)

[**Methodological Analysis** 10](#_Toc134556943)

[**Results and Discussion** 13](#_Toc134556944)

[**Reference** 14](#_Toc134556945)

# **Introduction**

In our study, we aim to analyze the HIV transmission network to understand its structure, identify high-risk subgroups, and inform potential interventions. Our research will delve into various aspects, such as centrality measures, community detection, network models, predictive modeling, and identification of key individuals or groups, to provide a comprehensive understanding of the network and its implications on HIV transmission.

The research questions are:

1. What is the five-number summary of the network. How can we interpret them?
2. How do centrality measures (degree, betweenness, eigenvector) differ across node attributes? Are certain subgroups more central in the network?
3. Can community detection algorithms identify subgroups within the network based on node attributes, and if there are subgroups that have a higher risk of HIV transmission? What are the defining characteristics of these subgroups?
4. Can network models, such as Exponential Random Graph Models (ERGMs) helps explain the formation and structure of the HIV transmission network based on node attributes? Which attributes have the strongest impact on network formation and HIV transmission?
5. Can the relationships between node attributes and network properties (e.g., centrality measures) be used to develop predictive models for HIV transmission risk within the network?
6. Can you identify any key individuals or groups of individuals (e.g., opinion leaders, influencers) within the network who could be targeted for interventions to reduce HIV transmission? We can use algorithms like k-shell decomposition or methods like snowball sampling to identify these individuals or groups.
7. Can you identify any structural holes or brokerage positions within the network, and how do these positions relate to node attributes? Individuals occupying these positions may have a unique role in the information or resource flow within the network. We can use methods like Burt's constraint measure to identify structural holes or brokerage positions.

# **Literature Review**

The method for constructing the data set was by weighting the frequency of duplicate interactions that individuals from ID\_1 had with ID\_2. We use this weighted network to follow the theoretical framework of HIV transmission. We believed that there is no randomization to how ties are formed between two individuals who share similar characteristics. We believe under homophily based modeling that individuals who share similar characteristics are inclined to form compatibility through love of the same kind. So, in regard to our HIV model, we are using weights to see how frequent ties were formed and whether new ties were formed that continued transmission of HIV. Some individuals sought multiple sexual partners while others sought a few; however, those who sought a few branched outward and extended a “vine-like” structure of transmission. Our data contained sexual orientation behavior which is linked to certain STIs based on the CDC’s “Sexually Transmitted Disease Surveillance” annual report. In “Fracking and Risky Sexual Behavior” explores an event study through the fracking boom in North Dakota to examine how STIs transmitted and how it affected the sex work market. While this paper looks at a different scope of analysis, it still provides insight on how researchers examine which STIs are attached to certain populations and how pervasive they remain over time. In this case, we use this belief to examine how HIV behaved given one’s sexual orientation. Another paper that examines similar STI behavior and stigmas is “Social Context of Sexual Relationships Among Rural African Americans.” This paper encompasses a similar approach to “Fracking and Risky Sexual Behavior” but focuses on a rural context to examine how transmission behaves under a rural setting for a specific population.

We then utilize papers like “Social Smoking among Intermittent Smokers” to see whether any parallels could be drawn between habitual sexual partners and one-time occurrences for HIV transmission. While some individuals engaged in risky sexual behavior with repeated or unique partners, their behavior did not spillover into influencing the behavior of unique partners. Panel data and informative covariates are needed to determine whether the initial engagement was influential enough to change HIV transmission over time. Our main contributor of inspiration comes from “Social Network Interventions for HIV Transmission Elimination” which focuses on moving the scope of view from the individual for HIV prevention to examining how the HIV transmission network can work towards mitigating HIV transmission. This paper examines how a lot of the pressure to limit the spread of HIV should be shifted from the individual to the network because the network composition itself can determine vulnerability and accessibility. Through segmentation, interventions tailored for specific populations can be reached via social media for example to target LGBT+ folk who are an at-risk population for STIs due to the prevalence of risky sexual behavior.

**Data Description**

This data set is a collection of different data sets spanning different years and locations. The purpose of this data set is to establish a source for conducting analysis on the influence of partnership networks on the transmission of sexually transmitted and blood-borne infections. The second purpose of this data source collection is to implement social network study designs of network properties and their impacts. This data set collection consists of the following sources: Colorado Springs Project 90 from the years 1988-1992, Bushwick [Brooklyn, NY] Social Factors and HIV Risks (SFHR) Study from the years 1991-1993, Atlanta Urban Networks Project from the years 1996-1999, Flagstaff Rural Network Study from the years 1996-1998, Atlanta Antiretroviral Adherence Study from the years 1998-2001, Houston Risk Networks Study from the years 1997-1998, Baltimore SHIELD (Self-Help in Eliminating Life-Threatening Diseases) from the years 1997-1999, and Manitoba Chlamydia Study from the years 1997-1998.

Each study contains information on sexual, needle sharing, and/or social networks methods of transmission. The method each study was composed of was intended to allow for a wide range of study types to be performed. Some of the possible analyses include the effect of personal behavior and network connections (via transmission type) on the dynamic nature of disease transmission. Respondents of each study were surveyed to measure the test results of HIV, chlamydia, syphilis, and hepatitis. They also implemented demographic control questions to account for individual characteristics.

Although the dataset contains 8 sources of data, not every source of data surveyed the same questions. Some studies captured questions regarding HIV and others focused on other STDs. Because each study is in a distinct city and state, constructing a social networks model capturing the likelihood of becoming infected with one of the following STIs would be a misspecification and study design error. Each study should compose its own network illustration without assuming the characteristics of one study are applicable to another city’s study. Some studies captured information about sources that acted as vectors for STIs. The data is structured by respondents and who they infected via ID\_1 and ID\_2. The variable ID is used to classify each unique respondent; however, ID\_1 and ID\_2 is used to show which individual in ID\_1 infected another individual in ID\_2. This method is used to show how many times one individual infected another and through which method (needle, sexual, drug, and social). Some studies also capture testing results on STIs and self-reported results from their own recollection of their medical history. Individual characteristic questions were also captured in the survey to account for any variation that could explain how transmission is influenced. Some of the variables that were unique, but had low observations included whether someone was a client of a prostitute (john) and whether the individual was a pimp. Other variables in this category included occupational prestige, which is a subjective score assigned to certain occupations depending on how the public views that occupation regarding its prestige. One of the limitations of this survey is that some of the questions asked did not have the necessary co-questions to properly form other analysis beyond the scope of public health with social networks.

**Data Cleaning:**

Given the nature of real-world data, it often contains errors, inconsistencies, and missing values that must be addressed before conducting an analysis. Our original raw dataset was no exception, comprising 85,890 rows and 155 columns, with numerous issues that required careful cleaning and preprocessing.

Upon examination, we discovered several disorganized categories within the dataset. For instance, the EDUC column contained negative numbers, missing values, and undefined numbers, which are all unsuitable for our analysis. A similar issue extended across various other columns. If we were to remove all Out of Domain (OOD) values (-9) from the dataset, we would essentially be left with no records, demonstrating the pervasiveness of problematic categories in the dataset.

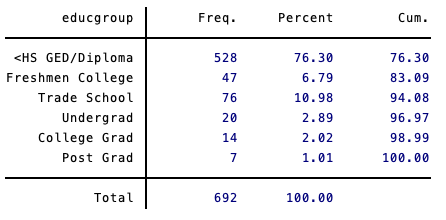
To address these issues and make the dataset manageable for our analysis, we adopted a conservative data cleaning approach. We retained data from the years 1988 to 2003 but chose to keep only a few relevant columns with fewer OOD and missing categories. The columns we retained were ID1, ID2, RACE1, RACE2, SEX1, SEX2, AGE1, AGE2, BEHAV1, BEHAV2, DISABLE1, DISABLE2, UNEMP1, UNEMP2, STREETS1, STREETS2, EDUC1, and EDUC2. This approach allowed us to reduce the dataset size significantly, yielding a final dataset with 4,329 rows and 18 columns, making it more suitable for our analysis.

With the cleaned dataset, we constructed an edge list using the ID1 and ID2 columns to represent the nodes and a 'Weight' column to indicate the strength of the relationship between the nodes. The final edge list comprised 1,810 rows and 3 columns.

Simultaneously, we created a separate node attribute data frame to store the attributes of the nodes in our network. This data frame included the columns 'ID', 'RACE', 'SEX', 'AGE', 'BEHAV', 'DISABLE', 'UNEMP', 'STREETS', and 'EDUC', and resulted in 694 rows and 9 columns.

Descriptive Statistics: Node Attributes/Covariates

A picture containing text, screenshot, font, line

Description automatically generated

A picture containing text, screenshot, font, line

Description automatically generatedA picture containing text, screenshot, font, line

Description automatically generated

A picture containing text, screenshot, font, number

Description automatically generatedA picture containing text, screenshot, font, line

Description automatically generated

A picture containing text, screenshot, font, line

Description automatically generatedA picture containing text, screenshot, font, number

Description automatically generated

# **Network Visualization**

***Figure 1***

Node size and color: Weighted out-degree.

Edge size and color: Weight

A picture containing drawing, flower, illustration, art

Description automatically generated

***Figure 2***

Node size and color: Weighted in-degree.

Edge size and color: Weight

A picture containing flower, drawing, illustration, art

Description automatically generated

# **Social Network Analysis**

First, we will examine the five-number summary of the network and interpret its significance in understanding the network properties.

We will then analyze how centrality measures such as degree, betweenness, and eigenvector centrality differ across node attributes, and explore whether certain subgroups are more central in the network.

We will use community detection algorithms to identify subgroups based on node attributes and investigate if any of these subgroups have a higher risk of HIV transmission. We will also examine the defining characteristics of these subgroups.

Next, we will assess whether network models like Exponential Random Graph Models (ERGMs) can help explain the formation and structure of the HIV transmission network based on node attributes, and identify which attributes have the strongest impact. We will explore the possibility of developing predictive models for HIV transmission risk within the network, using the relationships between node attributes and network properties. We will attempt to identify key individuals or groups within the network who could be targeted for interventions to reduce HIV transmission, using algorithms like k-shell decomposition or methods like snowball sampling.

Finally, we will look for structural holes or brokerage positions within the network and examine how these positions relate to node attributes. Individuals occupying these positions may play a unique role in the flow of information or resources within the network, and we will use methods like Burt's constraint measure to identify these positions.

**Five-Number Summary of The Network**

|  |  |  |
| --- | --- | --- |
| **Term** | **Value** | **Interpretation of value** |
| Size | 694 | There have 694 nodes in this network |
| Density | 0.0038 | The proportion of observed ties in a network to the maximum number of possible ties. 0.0038 is the interconnection of this network, which is not very high. |
| Components | 311 | The network split into 311 subgroups. |
| Diameter | 20 | It’s the longest of the shortest paths across all pairs of nodes. Diameter 20 means it takes 20 steps to connect the two nodes that are furthest in the network. |
| Clustering coefficient | 0.2144 | It’s the proportion of closed triangles to the total number of open and closed triangles. 0. 2144 means a moderate level of clustering in this network. |

**Centrality Analysis:**

Centrality measures are used to identify the most important nodes in a network. Degree centrality measures the number of direct connections a node has with other nodes in the network. The betweenness centrality measures the number of shortest paths that pass through a node, indicating its importance as a "bridge" between different parts of the network. The eigenvector centrality measures a node's influence based on the importance of the other nodes it is connected to.

Looking at the centrality measures provided for RACE, SEX, BEHAV, and AGE, we can see that different groups within each variable have different levels of centrality. For example, in the RACE variable, group 2 has the highest mean degree centrality and eigenvector centrality, indicating that this group has the most direct connections and the most influence within the network. Group 4 has the highest mean betweenness centrality, indicating that nodes in this group are important "bridges" between different parts of the network.

Similarly, in the SEX variable, group 1 has higher mean degree centrality and betweenness centrality than group 0, indicating that nodes in group 1 have more direct connections and are more important as "bridges" between different parts of the network.

In the BEHAV variable, group 2 has the highest mean degree centrality and eigenvector centrality, indicating that nodes in this group have the most direct connections and the most influence within the network. Group 2 also has the highest mean betweenness centrality, indicating that nodes in this group are important "bridges" between different parts of the network.

Finally, in the AGE variable, we can see that different groups have different levels of centrality, with groups 23 and 24 having the highest mean degree centrality and eigenvector centrality, and group 18 having the highest mean betweenness centrality. This indicates that nodes in groups 23 and 24 have the most direct connections and the most influence within the network, while nodes in group 18 are important "bridges" between different parts of the network.

**Community Detection:**

* Community 119 has a majority of individuals from race category 2 (92.73%) and is predominantly of sex category 0 (67.27%). This community exhibits behavior category 0 (72.73%) and has a higher proportion of individuals in age category 37 (16.36%) and unemployment category 1 (70.91%).
* Community 104 is similar to community 119 in terms of race, sex, and behavior. However, it has a higher proportion of disabled individuals (90.91%) and a slightly older population with the majority in age category 45 (18.18%).
* Community 100 has a more diverse racial composition (category 4 at 55.56%) and a higher prevalence of behavior category 0 (88.89%). Most individuals in this community are not unemployed (55.56%) and have education category 12 (55.56%).
* Community 96 has a diverse racial composition (category 4 at 56.52%) and a high proportion of individuals in behavior category 0 (95.65%). The majority of the community is not unemployed (65.22%) and has education category 12 (34.78%).
* Community 98 is characterized by equal representation of sex category 0 and 1 (both at 50%). Most individuals are from race category 4 (50%) and have behavior category 2 (60%).
* Community 10 has a higher proportion of individuals from race category 4 (59.26%) and disability category 0 (96.30%). The community is relatively younger, with most individuals in age category 30 (11.11%).
* Community 109 has a majority of individuals from race category 4 (80%) and behavior category 0 (90%). The community has an even sex distribution (50% for category 0) and a relatively older population in age category 38 (20%).

**Exponential Random Graph Models (ERGMs):**

The full ERGM model provided includes node attributes **RACE, SEX, BEHAV, AGE, DISABLE, UNEMP, STREETS, and EDUC.** From the summary output, we can assess the impact of these attributes on network formation and HIV transmission.

**Statistically significant attributes:**

**RACE (Estimate: 0.577145, p-value: < 1e-04) \*:** This positive coefficient indicates that nodes with the same race are more likely to form connections in the network. In the context of HIV transmission, this suggests that individuals from the same racial group may have a higher probability of being connected within the network, influencing the transmission dynamics.

**UNEMP (Estimate: 0.125353, p-value: 0.00833):** This positive coefficient implies that nodes with the same unemployment status are more likely to form connections. This means that unemployed individuals are more likely to be connected with other unemployed individuals, and employed individuals are more likely to be connected with other employed individuals. This attribute might affect the HIV transmission network structure by creating clusters of individuals with similar employment statuses.

**STREETS (Estimate: 0.122385, p-value: 0.03190):** This positive coefficient suggests that nodes with the same streets' status (living on the streets or not) are more likely to form connections. This can impact the network structure by creating groups of individuals who either live on the streets or do not, which may influence the patterns of HIV transmission.

**Non-significant attributes:**

The following attributes have non-significant p-values, meaning that their impact on network formation and HIV transmission is not statistically significant in this model: SEX, BEHAV, AGE, DISABLE, and EDUC. It's important to note that non-significance doesn't necessarily mean these factors have no impact on the network. It could be due to limited data, model specification, or other reasons.

In conclusion, the results show that network models like ERGMs can help explain the formation and structure of the HIV transmission network based on node attributes. The attributes with the strongest impact on network formation and HIV transmission in this specific model are RACE, UNEMP, and STREETS. These factors help us understand the connections between individuals in the network, which can be important for public health interventions, resource allocation, and targeted prevention strategies.

# **Methodological Analysis**

**Predictive models for HIV transmission risk within the network:**

First, let us understand the key elements involved in this process:

Node attributes: In your dataset, the node attributes include **RACE, SEX, AGE, BEHAV, DISABLE, UNEMP, STREETS, and EDUC.** These represent individual characteristics of each person in the network.

**Network properties:** These are quantitative measures that describe the overall structure of the network or the position of individual nodes within the network. Examples include degree centrality (the number of connections a node has), betweenness centrality (how often a node lies on the shortest path between other nodes), and closeness centrality (how close a node is to all other nodes in the network).

**HIV transmission risk:** In order to develop a predictive model, we need an outcome variable that represents the risk of HIV transmission within the network. This could be a binary variable indicating whether a transmission event occurred or some proxy for transmission risk, such as the number of high-risk contacts or the presence of serodiscordant partnerships.

To develop a predictive model for HIV transmission risk within the network, you will need to follow these steps:

**Step 1: Calculate network properties.**

Compute network properties, such as centrality measures, for each node in the network. This can be done using network analysis tools and packages in R or Python, such as the igraph or networkx libraries.

**Step 2: Explore relationships between node attributes and network properties.**

Examine the association between the node attributes and network properties by conducting correlation analyses, cross-tabulations, or regression models. This will help you identify which attributes are related to the network properties that may be relevant for HIV transmission risk.

**Step 3: Collect data on HIV transmission events or a proxy for transmission risk.**

As mentioned earlier, you need an outcome variable representing the risk of HIV transmission within the network. Collect data on actual transmission events or some proxy for transmission risk and link it to the nodes in the network.

**Step 4: Develop predictive models.**

Use various modeling approaches to predict the risk of HIV transmission based on the relationships between node attributes, network properties, and the outcome variable. This could involve logistic regression, machine learning algorithms (e.g., decision trees, random forests, or neural networks), or network-based models (e.g., agent-based models or stochastic network models).

**Step 5: Evaluate model performance.**

Evaluate the performance of the predictive models using appropriate metrics, such as accuracy, precision, recall, F1 score, or area under the ROC curve. This will help you determine how well the models can predict HIV transmission risk within the network based on node attributes and network properties.

**Finally**, it is possible to use relationships between node attributes and network properties to develop predictive models for HIV transmission risk within the network. By following the steps outlined above and using appropriate data and modeling techniques, you can build and evaluate models that predict the risk of HIV transmission based on individual characteristics and network structure.

**Key individuals or groups of individuals use algorithms like k-shell decomposition or methods like snowball sampling.**

Can we identify any key individuals or groups of individuals (e.g., opinion leaders, influencers) within the network who could be targeted for interventions to reduce HIV transmission? We can use algorithms like **k-shell decomposition** or methods like **snowball sampling** to identify these individuals or groups.

**K-shell decomposition:**

K-shell decomposition is a way for identifying influential nodes in a network based on their connectivity. The process involves iterative removal of nodes with the lowest degree (**number of connections**) until the network is empty. The nodes are then classified into shells based on the order in which they were removed.

In the chosen network, the k-shell decomposition method has identified the following 35 individuals as potentially influential:

**343 344 364 369 370 377 439 440 513 515 516 517 518 519 520 521 523 524 525 526 527 529 533 538 539 540 549 550 557 558 560 563 565 573 577**

These individuals belong to the higher k-shells in the network, indicating that they are more connected and possibly more influential than others. Targeting these individuals for HIV interventions might lead to more effective dissemination of information and behavior change within the network**.**

**Snowball sampling:**

Snowball sampling is a method used to discover influential nodes in a network by recursively exploring the neighborhood of a set of seed nodes. Starting with a small set of seed nodes, their neighbors are added to the sample, followed by the neighbors of the neighbors, and so on until a desired sample size is reached or no more nodes can be added.

In the chosen network, the snowball sampling method has identified the following 38 individuals as potentially influential:

**1 18 33 84 121 585 3 15 19 25 28 43 98 127 131 151 234 241 10 26 52 96 106 114 73 17 31 34 93 132 133 135 136 193 251 310 599 226**

These individuals are not only central in the network but also well-connected to other well-connected individuals. By targeting these individuals for HIV interventions, the information and behavior change might spread more effectively throughout the network, reaching both centrally connected individuals and those who may be influential within their local communities.

**Illustration:**

Consider a network with nodes representing individuals and edges representing relationships between them. In this network, nodes with high k-shell values are those that are well-connected to others and tend to form the core of the network. On the other hand, nodes identified through snowball sampling are those that, while potentially not as central, still have connections to well-connected individuals.

By targeting both sets of individuals for HIV interventions, the information and behavior change can spread more effectively throughout the network. For example, imagine a public health campaign that aims to promote safe sex practices or provide information on HIV testing services. By engaging the influential individuals identified through k-shell decomposition and snowball sampling, the campaign can reach a larger portion of the network, ultimately leading to a greater impact on reducing HIV transmission.

To aggregate**,** the k-shell decomposition and snowball sampling methods help identify key individuals or groups within the network who could be targeted for interventions to reduce HIV transmission. By engaging these individuals in interventions, public health campaigns can potentially achieve more effective dissemination of information, promote behavior change, and ultimately reduce the risk of HIV transmission within the network.

**Identify the structural holes or brokerage positions within the network by using Burt's constraint measure.**

The following individuals have been identified as occupying structural holes or brokerage positions:

These individuals may have unique roles in the information or resource flow within the network due to their brokerage positions. They can potentially connect diverse groups or act as bridges between otherwise disconnected parts of the network.

To explore how these positions relate to node attributes, you can look for patterns or associations between the attributes and the constraint scores. **For example:**

**Are individuals of a specific race or sex more likely to occupy brokerage positions?**

**Is there a relationship between age or education and brokerage positions?**

**Do individuals with certain behaviors (e.g., drug use or sexual behaviors) tend to occupy structural holes?**

By examining these relationships, you can gain insights into the factors that may influence an individual's likelihood of occupying a brokerage position within the network. This information can be useful for designing targeted interventions to reduce HIV transmission, as individuals in these positions may have a broader reach or unique influence on others within the network.

**Here is a table of the top 5 individuals with the lowest constraint scores:**

|  |  |
| --- | --- |
| **Node ID** | **Constraint Score** |
| **52** | **0.06542113** |
| **43** | **0.07971723** |
| **133** | **0.08156873** |
| **135** | **0.08374918** |
| **346** | **0.08576506** |

**These individuals (with Node IDs 52, 43, 133, 135, and 346)** are the ones occupying key positions in the network, as they have the lowest constraint scores. They are likely in a position to control the flow of information or resources within the network. By targeting these individuals with interventions, we can potentially have a larger impact on reducing HIV transmission within the network.

# **Results and Discussion**

A number of factors play a crucial role in the spread of HIV. Gender, race, and sexual behavior are among the most important factors that should be considered when developing intervention policies. It is important to tailor these policies to specific community needs, as not all communities look or behave the same. To effectively target individuals likely to occupy brokerage positions in the network, policymakers should consider factors such as race, sexual behavior, and employment status. In addition, our research highlights the importance of addressing key underlying socioeconomic issues that play a role in determining transmission networks. Factors such as race, unemployment status, and housing status are all key factors in determining the structure of these networks, and it is imperative that policymakers address these issues to effectively combat the spread of HIV.

Our research is limited in that many important attributes were not available to us for analysis. Further research is needed to explore the role of certain occupations, such as sex workers and drug dealers, as well as the role of ethnicity in transmission networks. Another major limitation of our research is that our data comes from multiple other studies, and therefore many different cities and with different testing and surveying methods. Future research should be done with a more cohesive dataset.

Overall, the findings of our research have important implications for policymakers seeking to combat the spread of HIV. It is apparent that HIV transmission is a systemic issue wherein communities that share certain attributes are at higher risk. By considering the specific needs of different communities, targeting individuals likely to occupy brokerage positions in the network, and addressing key underlying socioeconomic issues, policymakers can develop more effective intervention policies and strategies to prevent the spread of HIV and improve the health outcomes of those living with the virus.

# **Reference**

Skvoretz, J., & Blau, P. M. (1995). Structural contexts of opportunities. *Contemporary Sociology*, *24*(6), 805. <https://doi.org/10.2307/2076711>

Pagkas-Bather, J., Young, L. E., Chen, Y.-T., & Schneider, J. A. (2020). Social network interventions for HIV transmission elimination. *Current HIV/AIDS Reports*, *17*(5), 450–457. <https://doi.org/10.1007/s11904-020-00524-z>

Shiffman, S., Li, X., Dunbar, M. S., Ferguson, S. G., Tindle, H. A., & Scholl, S. M. (2015). Social smoking among intermittent smokers. *Drug and Alcohol Dependence*, *154*, 184–191.

ADIMORA, A. A., SCHOENBACH, V. J., MARTINSON, F. E., DONALDSON, K. H., FULLILOVE, R. E., & ARAL, S. O. (2001). Social context of sexual relationships among rural African Americans. *Sex Transm Dis*, *28*(2), 69–76.

Cunningham, S., DeAngelo, G., & Smith, B. (2020). Fracking and risky sexual activity. *Journal of Health Economics*, *72*, 102322. <https://doi.org/10.1016/j.jhealeco.2020.102322>