Generating a heterosexual bipartite network embedded in social network

Asma Azizi1*, Zhuolin Qu2, Bryan Lewis3 and James Mac Hyman4

Abstract
We describe an approach to generate a heterosexual network with a prescribed joint-degree distribution embedded in a prescribed large-scale social contact network. The structure of a sexual network plays an important role in how all sexually transmitted infections (STIs) spread. Generating an ensemble of networks that mimics the real-world is crucial to evaluating robust mitigation strategies for controlling STIs. Most of the current algorithms to generate sexual networks only use sexual activity data, such as the number of partners per month, to generate the sexual network. Real-world sexual networks also depend on biased mixing based on age, location, and social and work activities. We describe an approach to use a broad range of social activity data to generate possible heterosexual networks. We start with a large-scale simulation of thousands of people in a city as they go through their daily activities, including work, school, shopping, and activities at home. We extract a social network from these activities where the nodes are the people, and the edges indicate a social interaction, such as working in the same location. This social network captures the correlations between people of different ages, living in different locations, their economic status, and other demographic factors. We use the social contact network to define a bipartite heterosexual network that is embedded within an extended social network. The resulting sexual network captures the biased mixing inherent in the social network, and models based on this pairing of networks can be used to investigate novel intervention strategies based on the social contacts among infected people. We illustrate the approach in a model for the spread of chlamydia in the heterosexual network representing the young sexually active community in New Orleans.

Keywords: B2K network, Bipartite network, Social contact network, Heterosexual network, Joint degree distribution, Sexually transmitted infections

Introduction
The structure of heterosexual networks plays an important role in the spread of all sexually transmitted infections (STIs), including chlamydia and gonorrhea. These networks are captured in computer simulations by a bipartite graph where the nodes represent the people and the edges are sexual partnerships between nodes of different sexes. Determining what is predictable in STI models requires an algorithm to generate an ensemble of random graphs that resembles real-world sexual activities.
These graphs must account for the distribution for the number of sexual partners people have (their degree distribution) and the number of partners their partners have (the joint-degree, or degree-degree, distribution). The existing algorithms that generate bipartite random graphs preserving degree and joint-degree distributions of the nodes are strictly based on the number of partners people have and not other demographic factors, such as age or location (Newman 2002; Hakimi 1962; Boroojeni et al. 2017; Azizi et al. 2016, 2017, 2018).

The degree and joint-degree distributions are just two of many properties for a heterosexual network that can affect its structure and the validity of an epidemic model. The heterosexual network is also correlated to an underlying social contact network of acquaintances connected by interpersonal relationships. A person’s sexual activity depends on age, race, sociodemographic, and socioeconomic features of the environment that can be captured by a social contact network (Amirkhanian 2014; Adimora and Schoenbach 2005; Ruan et al. 2011; Juher et al. 2017; Morris et al. 1995). In other words, using the extended social network of a person as a source of sexual partner selection when generating a heterosexual network enables the network to capture the bias in heterogeneous mixing based on age, race, economic status, and geographic location (McPherson et al. 2001).

Although it is widely accepted that social contact (non-sexual partners) and heterosexual (sexual partners) networks are related, there are few studies on how a population’s social contact network impacts the spread of heterosexual STIs. The social network can affect the structure of the heterosexual network by providing a pool of sexual partners. A person’s social network also influences their cultural norms regarding STI testing, safe sex practices, and knowledge about the spread and treatment of STIs. As far as we know, there are no other existing mechanistic approaches that construct a heterosexual network embedded within a realistic social network. We will describe a new approach that fills this gap by applying social contact networks to generate the heterosexual network while preserving the joint-degree distribution of data.

Our new network generation approach uses the underlying extended social network of a population to extend these previous algorithms for generating bipartite heterosexual networks with prescribed joint-degree distribution (Boroojeni et al. 2017). Many sexual partnerships are formed from within a person’s social circle, defined by the people they have regular social contact with and the contacts of their contacts (their extended social network). These social circles have been modeled through large-scale simulations of thousands of people in a city as they go through their daily activities. We start with a network that mimics the social activity of the population (Eubank et al. 2010), as generated by a complex social network simulation. We use this simulated data to create an extended social network and then identify a bipartite network of men and women to define our heterosexual network. We then create a virtual heterosexual network as a subgraph of this bipartite social network that captures a prescribed joint-degree distribution.

As a case study, we construct a heterosexual network that is embedded in the social contact network of the New Orleans population and mimics the sexual behavior
obtained from a sexual behavior survey of the young adult African American population in New Orleans (Kissinger 2014; Green et al. 2014).

**Materials and method**

People often find their sexual partners within their extended social network, the individuals they contact each day at work, school, or other social activities. There are sophisticated simulations of these social networks, such as EpiSims that is based on the Transportation Analysis and Simulation System (TRANSIMS) developed at Los Alamos National Laboratory (Barrett et al. 2005) or Simfrastructure (Eubank et al. 2010; Eubank 2008). These social networks can be used to produce a sexual network, which is more realistic than basing partnerships on just the sexual activity of different individuals.

The social contact network is a graph where the nodes are synthetic people, labeled by their demographics (sex, age, income, location, etc.), and the edges between the nodes represent contacts determined in which each synthetic person is deemed to have made contact with a subset of other synthetic people through some Activity types. Each edge of the network is labeled with one of these activity locations and is weighted by the time spent on these contacts per day. For example edge \((i,j)\) labeled by the activity \(A = \text{Work}\) and weighted by \(T_{ij}^{\text{W}}\) means two persons \(i\) and \(j\) have a contact for \(T_{ij}^{\text{W}}\) fraction of their total time spent at work. We base our algorithm on a social contact network, called SocNet, generated by Eubank et al. (2010) with activity at different locations (e.g., home, work, school, shopping, or other activity).

We introduce an algorithm that embeds a heterosexual network within a social network and matches the sexually active population’s joint-degree distribution. The heterosexual network preserves the bipartite joint degree \((BJD)\) distribution matrix that represents the correlations between then number of partners a person has and the number of partners their partners have (Boroojeni et al. 2017). The algorithm has three stages:

(i) Generate an extended social contact network, **ESocNet**: The original social contact network is a simple graph, whose nodes are synthetic people, and neighboring nodes are their social contacts during a typical day. We assume that most sexual partnerships come from a person’s social contacts or the social contacts of their social contacts, e.g., the neighbors of the neighbors of a node. We extend the social contact network to create a new network, the extended social network, **ESocNet**, where some of the neighbors (social contacts) of an individual’s neighbors in this network are added to his/her social contacts.

(ii) Generate a reduced social bipartite network, **BSocNet**: The **ESocNet** includes all the individuals in the region being modeled. Our sexual network is based on individuals within a prescribed age range. In this step, we remove all nodes where the associated individuals are outside this age range. The extended social network is a simple graph where nodes have some neighbors that are of the same sex. We identify the embedded bipartite subgraph of this network by removing all edges between individuals of the same sex. The resulting bipartite graph is a social network where male nodes are only connected to female nodes and vice versa.

Finally, we removed monogamous couples from the simulation since they are not part of the STI transmission network and assume that siblings are not sexual partners.
This was achieved by removing all edges between individuals living in the same household, which is the edge labeled activity H for home. This is an approximation since there are some non-monogamous couples living in the same household, even in the high-risk young adult African American population being modeled. We call this reduced social bipartite network as **BSocNet**.

(iii) Generate an embedded heterosexual bipartite network, **SexNet**: We then use the **BSocNet** to define a heterosexual network of sexual partnerships, the **SexNet**, with a prescribed bipartite joint degree matrix $BJD$ based on survey data (Boroojeni et al. 2017). That is, we preserve the correlations between the number of partners a person has and the distribution for the number of partners their partners have. We assume that most of a person’s sexual partners are neighbors in the **BSocNet** and a few of the partners are randomly selected from elsewhere in the population where they might have met through social media or at any other event.

**Generate an extended social contact network (ESocNet)**

In the first stage of our algorithm, we create an extended social contact network, **ESocNet**, so that an individual’s social contacts include some of the contacts of their contacts. That is, **ESocNet** will add potential sexual partners by including some of the social contacts of an individual’s social contacts.

Consider two people (nodes) $i$ and $j$ who are not currently connected, but have $k_A(i,j) > 0$ common social contacts within activity $A$. We define $p_A^{ij}$ as the probability that they will meet through a single contact. Therefore, the probability that they will meet and be connected in the **ESocNet** after $k_A(i,j)$ contacts is $1 - (1 - p_A^{ij})^{k_A(i,j)}$.

The probability $p_A^{ij}$ is a function of the time that $i$ and $j$ spend in an activity $A$ in **SocNet**. From the data in **SocNet**, we can define $\tau_A^k$ as the average fraction of time person $k$ spends with each social contact, when engaged in activity $A$:

$$\tau_A^k = \frac{\sum_{l \in N_A(k)} T_A^k l}{|N_A(k)|},$$

(1)

where $T_A^k l$ is the fraction of time two contacts $k$ and $l$ spend together in activity $A$, and $N_A(k)$ is set of all social contacts for person $k$ through an activity location $A$. We then define $p_A^{ij} = \tau_A^i \tau_A^j$. Figure 1 describes a schematic of this algorithm for a simple network.

**Generate a reduced social bipartite network, BSocNet**

In our heterosexual network, we only consider the sexually active population within a prescribed age range $\alpha = [\alpha_1, \alpha_2]$. That is, we trim the **ESocNet** by removing all people with ages outside this range to not including any edges (sexual contacts) with people outside this range. We then remove all edges between people of the same sex to create a bipartite heterosexual social network. To avoid including siblings as potential sexual partners and because men and women in our young high-risk population cohort are less likely to live together as sexual partners in the same household, we eliminate the sexual partnerships between individuals living in the same household by removing all edges labeled activity H as home, to define the **BSocNet**.
Generate an embedded heterosexual network (SexNet)

The Soc2Sex algorithm uses BSocNet to generates a heterosexual network, SexNet, that mimics the heterogeneous mixing of the real population. We assume that we have an estimate for the distribution for the number of partners of men and women (the degree distributions for their associated nodes) and the joint-degree distribution for the number of partners that their partners have (Boroojeni et al. 2017).

An edge, $ij$, between two persons $i$ and $j$ in SexNet represents a sexual partnership. The degree of a person $i$, is defined by the number of his/her sexual partners. The degree distribution $\{d_k\}$ defines the number of people with degree $k$. The joint-degree distribution $(k, j)$ is the number of partnerships between a man with degree $j$ and a woman with degree $k$. This distribution can be represented by the Bipartite Joint Degree or BJD matrix:

$$BJD_{SexNet} = \begin{pmatrix}
    e_{11} & e_{12} & e_{13} & \cdots & e_{1m} \\
    e_{21} & e_{22} & e_{23} & \cdots & e_{2m} \\
    \vdots \\
    e_{w1} & e_{w2} & e_{w3} & \cdots & e_{wm}
\end{pmatrix},$$

where, $w$ is the maximum degree in women nodes, and $m$ is the maximum degree in men nodes, each element $e_{ij}$ is the number of edges between women with $i$ partners and men with $j$ partners.

The degree distribution of the number of women nodes, $d^w_k$, and men nodes, $d^m_k$, with $k$ partners can be obtained from $BJD_{SexNet}$:

$$d^w_k = \frac{\sum_{j=1}^{m} e_{kj}}{k} , \text{ and } d^m_k = \frac{\sum_{i=1}^{w} e_{ik}}{k} . \quad (2)$$
Though the heterosexual network SexNet is a subgraph of BSocNet, we also consider some sexual partners that are within a person’s extended social circle. That is, for the general case, SexNet is partially embedded in BSocNet.

The Soc2Sex algorithm first generates an initial heterosexual network that closely agrees with the desired BJD matrix and is partially embedded in BSocNet. Usually, this network satisfies the desired BJD, but can fail when the average degree (number of partners people have) becomes large. When this happens, a second fix-up algorithm, based on rewiring the network, is used to repair any discrepancies so that the final SexNet has the desired BJD matrix.

Generating the bipartite network

The Soc2Sex algorithm starts with the SocNet, the BJD matrix corresponding to SexNet, and the fraction $p \in [0, 1]$ of partners that are chosen randomly from the extended social contacts in the BSocNet. The remaining fraction, $(1 - p)$, of a person’s sexual partners are randomly selected from elsewhere in the population. These partnerships might have formed by meeting through social media or a social event not captured by the original SocNet. The Soc2Sex algorithm then generates a heterosexual network that is a partial subgraph of BSocNet and has a joint-degree distribution given by the BJD matrix. Note that $p$ is approximately the percentage of SexNet that is a subgraph of BSocNet.

The algorithm starts with an empty set of nodes SexNet and then builds a network guided by the BJD matrix. The nodes with the smallest degree have the least flexibility, so we start building SexNet by randomly selecting a man node of BSocNet with the highest (social) degree and assign its desired sexual degree to be column size of BJD matrix. This is represented by stubs, or unconnected edges, associated with this node.

We repeat the following process until all edges in SexNet, which is equal to the summation of elements in BJD matrix, are placed: at any step, we select a node with the highest stub in SexNet. Then we generate a uniform random number. If this random number was less than $p$ then we use the pool of the node’s social contact in BSocNet to find him/her a partner with proper degree defined by BJD. If the random number was bigger than $p$, we find the partner with proper degree defined by BJD from people other than their social contact but with smallest distance in their social contact. After finding such partner, we reduce the node’s new partner’s stub by one. If we find all the partners for all nodes in SexNet, we add a new node to SexNet from highly socially active nodes in BSocNet. Then, we assign this new added node a desired degree and stub equal to the current maximum degree frequency of SexNet.

To keep or remove an edge, we have to calculate the degree of nodes attached to it for each possible edge in the SocNet, thus, the full set of experiments run in $O(|E|P^mP^w)$ time, where $|E|$ is the number of edges in SocNet, $P^m$ number of its men nodes and $P^w$ number of its women nodes. This method is feasible if the average degree $\frac{2|E|}{P^mP^w}$ of the network is not high.

Table 1 shows the average CPU time for generating a SexNet with between 400 nodes to 12,800 nodes that are embedded within the BSocNet. These networks all maintained the same average degree, 1.5, and men-women ratio, 0.35:0.65, as our survey data.

For completeness, we provide pseudo-code for our Python scripts in Algorithms 1 and 2. Table 2 is the table of symbols used in these Algorithms.
Table 1 Average CPU time in seconds and 95% confidence intervals (CI) for generating 50 random SexNet on a MacBook Pro laptop computer. Notice that the CPU time scales almost linearly with the network size for a fixed average degree

| Network size | 400 | 800 | 1600 | 3200 | 6400 | 12,800 |
|--------------|-----|-----|------|------|------|--------|
| CPU time     | 120 | 23.6| 43.2 | 86.7 | 225  | 497    |
| 95% CI       | [11.5, 12.4] | [23.3, 23.8] | [42.3, 44.2] | [86, 88] | [219, 230] | [454, 540] |

Algorithm 1: Extracting sexual network SexNet, from a given social network BSSocNet

function SexNet ← Soc2Sex (BSSocNet, BJD, p)

Input: Revised social network BSSocNet, bipartite joint-degree matrix BJD, p ∈ [0, 1] average fraction of sexual partners selected from social contacts.

Output: Sexual network SexNet.

/* From the highest degree men nodes in BSSocNet, randomly select one node to add to SexNet */
SexNet = ∅, SexNet = u = max\_degree\(u\)(k ∈ BSSocNet.n) ;
/* The desired degree of the first selected node in SexNet is the column size of BJD */
\(d_{SexNet}(u) := BJD.col, \text{ stub}(u) := d_{SexNet}(u) ;
/* E is total number of edges in SexNet to be filled */
E := \sum \sum BJD(i, j) ;
while |SexNet| ≤ E do
    /* NF includes all the nodes in SexNet who are still looking for partners */
    NF := {k ∈ SexNet n if stub(k) > 0} ;
    while |NF| ≥ 1 do
        /* Start with the highest degree nodes in NF */
        u = max\_degree\(u\)(k ∈ NF) ;
        /* Find partner for u via Algorithm 2 → partner v of degree d’ */
        \(d’, v) = FP(u, BSSocNet, SexNet, BJD, NF, p) ;
        if \((d’, v) = False\) then
            /* If could not find partner for u remove it from NF */
            NF.remove(u) ;
        else
            /* If find the partner v, add edge between u and v and reduce both their stubs by one */
            Add edge \(u, v)\) in SexNet, \(\text{stub}(u) ← \text{stub}(u) - 1, \text{stub}(v) ← \text{stub}(v) - 1 ;
            /* If u is the only one left, remove it from the candidates in SexNet */
            if stub(u) = 0 \(\text{stub}(v) = 0\) then
                \(d_{SexNet}.remove(d_{SexNet}(u)) \) \(d_{SexNet}.remove(d_{SexNet}(v)) ;
            end
            /* Update the corresponding entry in BJD */
            if u is woman then
                \(BJD(d, d’) ← BJD(d, d’) - 1 ;
            else
                \(BJD(d’, d) ← BJD(d’, d) - 1 ;
            end
        end
    end
/* From the nodes in BSSocNet but not SexNet, choose one node with highest social degree, add to SexNet */
SexNet = u = max\_degree\(u\)(k ∈ BSSocNet.n − SexNet.n) ;
/* Define its degree and stub equal to maximum value in degree frequencies of SexNet */
\(d_{SexNet}(u) := \text{max}(d_{SexNet}), \text{ stub}(u) := d_{SexNet}(u) ;
end
return SexNet ;
The initial algorithm will generate a network with the prescribed degree distribution. We have observed that the resulting network always had the desired BJD matrix for all of the sparse heterosexual networks we have generated in this project. However, there are some situations, where the desired network is not sparse, the algorithm can fail to exactly produce a network with the desired BJD matrix for the joint-degree distribution. When this happens, a second algorithm is used to rewire the network so that it will exactly match the desired bipartite joint-degree distribution for the number of partners that a person’s partners have.

```
Algorithm 2: Finding partner with proper degree for a given node (FP)

function (d', v) = FP(u, SexNet, SocNet, BJD, NF, p)

Input: Node u in the sexual network SexNet, social network SocNet, BJD matrix, set NF of nodes who need partners, p ∈ [0, 1] fraction of sexual partnerships from social network.

Output: (d', v): Node v with its desired degree d'.

\[ d = d_{SexNet}(u) \]

/* Depending on sex of node and its degree, take the dth row or column of BJD */
if u is woman then \( R := BJD(d, .) \), else \( R := BJD(., d) \);
for range(\( R \)) do
  if \( R \neq 0 \) then
    /* Choose one nonzero entry of \( R \) with index \( d' \), which is the degree of partner to be found */
    w = \( R_{\{d' \neq 0\}} \), \( d' = R_{\text{index}(w)} \);
    if \( w \leq p \) then
      /* Prepare for choosing from the social contacts: */
      /* K1 - set of available nodes in SexNet with degree \( d' \), who are also social contacts: */
      /* K2 - set of nodes that are not in SexNet but the social contacts of \( u \) with social degree \( \geq d' \) */
      \( K1 := \{ k \in NF : k \in SocNet.N(u) \land SexNet.N(u) = d_{SexNet}(k) = d' \} \);
      \( K2 := \{ k \in SocNet.N(u) \land SexNet.n - SexNet.n : d_{SexNet}(k) \geq d' \} \);
    else
      /* Prepare for choosing outside the social contacts: */
      /* K1 - set of available nodes in SexNet with sexual degree \( d' \), who are also social contacts */
      /* K2 - set of nodes that are not SexNet and not social contact of \( u \) with sexual degree \( \geq d' \) */
      \( K1 := \{ k \in NF : k \notin SocNet.N(u) \cup SexNet.N(u) \land d_{SexNet}(k) = d' \} \);
      \( K2 := \{ k \in SocNet.n - SexNet.n : d_{SexNet}(k) \geq d' \} \);
    end
  if \( K1 \neq \emptyset \) then
    /* If \( K1 \) is not empty, select the closest node in SexNet */
    \( w \in K1 \cdot \text{sample}(w : \text{dist}(\text{SexNet}, u, w) = \text{min}(\text{dist}(\text{SexNet}, u, k) \text{ for } k \in K1)) \);
    Break;
  else if \( K2 \neq \emptyset \) then
    /* If \( K2 \) is not empty, select the closest node in SexNet, define as sex degree and add to \( d' \) */
    \( w \in K2 \cdot \text{sample}(w : \text{dist}(\text{SexNet}, u, w) = \text{min}(\text{dist}(\text{SexNet}, u, k) \text{ for } k \in K2)) \);
    \( d_{SexNet}(w) = d' \), \( \text{index}(w) = d' \);
    Break;
  else /* If both don’t work, move on to a different degree */
    \( R(d') := 0 \);
  end
  else /* If all elements of \( R \) are 0, we fail to find a proper partner for \( w \) */
    \( (d', v) = \text{False} \);
    Break;
end
Return (d', v);
```
Rewiring SexNet for a given BJD matrix

The rewiring algorithm corrects any mismatch between the joint-degree distribution of generated SexNet and the desired BJD. We define the joint-degree distribution of the generated SexNet as \( \tilde{BJD} \) and the mismatch error matrix \( E = \text{BJD} - \tilde{BJD} \). If the matrix \( E \) has nonzero elements, then the network is rewired to eliminate the error. There are three possible cases:

(i) If entry \( E(i,j), (i,j > 1) \), is a positive value \( k \), it means that SexNet needs \( k \) more edges between degree \( i \) women and degree \( j \) men. To create these edges, we iterate the following process \( k \) times:

(a) First, identify a woman, \( w_i \), in SexNet, where \( d_{\text{SexNet}}(w_i) = i \), and has as a partner, \( m_1 \), with degree-1, i.e. \( d_{\text{SexNet}}(m_1) = 1 \).

(b) Next, identify another man, \( m_j \in \text{SocNet}.N(w_i) - \text{SexNet}.N(w_i) \), where \( d_{\text{SexNet}}(m_j) = j \) and \( m_j \) has a degree-1 partner, \( w_i \). That is \( d_{\text{SexNet}}(w_i) = 1 \).

(c) Finally, we rewire the network by removing the edges \( (w_i,m_j) \) and \( (m_j,w_i) \) and add edge \( (w_i,m_j) \), as illustrated in the Rewiring (1) of the Fig. 2.

(ii) If element \( E(i,j), (i,j > 1) \), is a negative value \( k' \), it means that SexNet have extra \( k' \) edges between degree \( i \) women and degree \( j \) men. To remove these edges, we iterate following process \( k' \) times:

(a) First, identify a woman, \( w_i \), in SexNet, where \( d_{\text{SexNet}}(w_i) = i \), which has a degree-\( j \) partner like \( m_j \), that is \( d_{\text{SexNet}}(m_j) = j \). The nodes \( w_i \) and \( m_j \) are selected so that they have at least one social contact with opposite sex that is not their sexual partner.

(b) Next, identify another man, \( m_1 \in \text{SocNet}.N(w_i) - \text{SexNet}.N(w_i) \), and a woman, \( w_1 \in \text{SocNet}.N(m_j) - \text{SexNet}.N(m_j) \).
(c) Finally, rewire the network by removing the edge \((w_i, m_j)\) and add edges \((w_i, m_1)\) and \((w_1, m_j)\), as illustrated in the Rewiring (2) of the Fig. 2.

(iii) In the previous steps, we pushed back nonzero elements in \(E\) to its first row and column, which causes new nonzero elements in the first row and column. To remove these nonzero values, we have to add or remove small components. For example, if the element \((1,1)\) of \(E\) is a positive value \(k\), it means that we need a small component of a degree 1 woman whose partners are all degree 1 men. Therefore, we simply make this component from the people who are not currently in \(\text{SexNet}_n\). If the element \((1,j)\) of \(E\) is a negative value \(k\), it means we have to remove a small component of a degree 1 man whose partners are all degree 1 women. Therefore, we simply look for such a component and remove it from \(\text{SexNet}\).

We have found that this algorithm almost always converges to the desired BJD. However, there are rare cases when the desired rewiring nodes may not exist, and algorithm stalls with \(E = 0\). In test simulations, we could construct some sparse social networks where this algorithm could not exactly match the desired BJD. These cases were rare and, when they did occur, the rewiring step still improved the approximation to the desired BJD of \(\text{SexNet}\). In the numerical simulations, all of the generated New Orleans heterosexual network had exactly the desired BJD based on survey data.

In the next section, we apply our algorithm to generate and analyze several random \(\text{SexNet}\) corresponding to sexual activity of adolescent and young adult sexually active African Americans reside in New Orleans. First, we explain the inputs of our approach: New Orleans social network and joint-degree distribution- \(\text{BJD}\) matrix- corresponding to its sexual network. Then we use our algorithm to generate and analyze a bunch of \(\text{SexNets}\) for a subpopulation of people in New Orleans.

**Simulations**

We analyze an ensemble of sexual networks with a prescribed joint-degree distribution representing sexual activity of young adult African Americans in New Orleans.

**The New Orleans social activity data and SocNet**

The \(\text{SocNet}\) is based on the synthetic data generated by Simfrastructure (Eubank et al. 2010; Eubank 2008) for 130,000 synthetic people residing in New Orleans. Simfrastructure is a high-performance, service-oriented, agent-based simulation system,
representing and analyzing interdependent infrastructures. The data for the network includes information for each individual, identified by their PID (personal identifier), and includes their age, gender, household, and other demographic information. The contact information for each PID is encoded in the contact file in Table 3.

The Simfrastructure data was used to generate the original SocNet, which was then used to generate ESocNet and BSocNet as described in the previous section.

The BJD matrix for New Orleans heterosexual activity
An ongoing community-based pilot study was conducted among sexually active African Americans ages 15—25 in New Orleans (Kissinger 2014), to assess the effectiveness of prevention and intervention programs for chlamydia. Socio-demographic information including age, race, educational level and, sexual behavior—number and age of heterosexual partners in the past two months—and history of their STI test results were collected from 202 men and 414 women participants. Meanwhile, their partners’ information has been collected by asking questions referring to the status of each relationship such as the partner’s age and the possibility that their partner(s) have intercourse with others. The survey results were used to construct the BJD matrix of a heterosexual network of individuals in New Orleans. For a population \( P = 15,000 \) sexually active young adult men and their women partners residing in New Orleans, we have

\[
\text{BJD}_{15000} = \begin{pmatrix}
1663 & 1588 & 1225 & 896 & 645 & 469 & 342 & 252 & 186 & 138 & 105 & 90 & 72 & 57 & 41 & 34 & 23 & 14 & 13 & 14 \\
474 & 452 & 350 & 255 & 185 & 133 & 97 & 73 & 52 & 39 & 31 & 26 & 20 & 17 & 11 & 9 & 8 & 7 & 3 & 4 & 4 \\
198 & 188 & 145 & 107 & 77 & 57 & 40 & 29 & 22 & 16 & 12 & 9 & 6 & 5 & 3 & 2 & 4 & 1 & 1 & 2 \\
68 & 63 & 49 & 35 & 26 & 18 & 15 & 10 & 8 & 6 & 4 & 3 & 3 & 1 & 2 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
18 & 18 & 13 & 9 & 6 & 6 & 3 & 4 & 1 & 1 & 2 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
3 & 3 & 3 & 2 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]

The dimension of this BJD matrix is \( 6 \times 21 \), that is, the maximum number of partners women have is 6 and for men is 21.

SexNet analysis
Using the social network and BJD matrix provided in the previous subsections and approach described in Section “Materials and method”, we generated 150 SexNets of
15000 people for \( p = 0.2, 0.4, 0.6, 0.8 \) and 1, where \( p \) is the proportion of sexual partners that are selected from social friends. That is, 30 of the SexNets are 20%, 30 are 40%, 30 are 60%, 30 are 80%, and the rest 30 are 100% subgraph of BSocNet. We then compared some descriptive metrics of this ensemble of random networks that were not imposed when generating the networks, including the size of giant components and bi-components, number of connected components, and average redundancy coefficient.

First, we evaluated and compared the size of the giant component and bi-component (the first and second biggest connected components of the network) for each group of the networks. Figure 3 shows the box plot of these sizes: there is an increment in the size of giant components when people select most of their sexual partners from their social contacts. Because in that case, sexually active people are tighter together within the social contact network. But there is not a significant difference in the size of giant bi-component.

The number of connected components, \( N_c \), is another measure characterizing network toughness. This measure can be, not necessarily, correlated to the component’s size of the network. Figure 4 displays descriptive statistics for \( N_c \) in each network group. Note that data distributions are approximately symmetrical, and metrics of \( N_c \) are similar across groups, but, they change by changing the source of partner selection- changing \( p \).

Redundancy coefficients are the measure of the degree to which nodes in a bipartite graph tend to cluster together:

**Definition 1** For a bipartite network, redundancy for a node is the ratio of its overlap to its maximum possible overlap according to its degree. The overlap of a node is the number of pairs of neighbors that have mutual neighbors themselves, other than that node (Latapy et al. 2008). For a typical node \( v \), the redundancy coefficient of \( v \) is defined as

\[
Rc(v) = \frac{|\{u, w \subseteq N(v), \exists v' \neq v. tuv' \in E, wv' \in E\}|}{|N(v)||N(v)|-1},
\]

where, \( N(v) \) is the set of all neighbors of node \( v \), and \( E \) is the set of all edges in the network.

We compare this measure for the networks in Fig. 5: each data point \( Rc(k) \) for degree \( k \) is obtained by averaging redundancy coefficient over the group of people with \( k \) partners. In most of the networks, \( Rc(k) \) decreases with \( k \) (Newman 2010). Redundancy coefficient \( Rc \) is affected by social network BSocNet: when people select more sexual partners from their social contacts the value for \( Rc \) increases, which is because of stage one of the algorithm- Generate an extended social network. In that stage, by connecting the social contacts of a person in BSocNet, we increase its clustering coefficient. Therefore, because increasing \( p \) SexNet becomes a stronger sub-graph of BSocNet, it inherits more properties from BSocNet, that is, by increasing \( p \) \( Rc \) of the SexNet, which is correlated to clustering coefficient of BSocNet, increases.
Discussion

We described a new algorithm to generate an ensemble of heterosexual networks based on heterosexual behavior surveys for the young adult African American population in New Orleans. The prescribed degree and joint-degree distribution represented the heterosexual network embedded within a social network that captures the biased mixing of the population based on age, physical location, and social activities.

We generated an ensemble of different heterosexual networks with the same $BJD$. When the networks had a more percentage of partners selected from their extended social contacts, then we observed a tighter distribution in the number of connected components and the size of giant component and bi-components. When more partners are chosen from the extended social network, instead of randomly selected from the population, then the size of giant component increases, and following that the
number of connected components decreases, which is because of reducing the mixing in generating sexual network: when people select their sexual partners from their social contacts they stand in a tight group within social network. In fact, when being subgraph of the extended social network become stronger, the candidate set of sexual partners for each person that is set of social contacts decreases and becomes local (this set includes close contacts and contacts of contacts) compared with when this set is the whole population.

As $p$ increases, then more partners are chosen from a person’s extended social network. This also increases the network clustering coefficients (where more partners of your partner’s partner are also one of your partners). The redundancy coefficients for networks increases as the dependence of sexual network on social one rises when $p$ increases, which is because of the high clustering coefficient of the social network due to the first stage of the algorithm, generate an extended social contact network. In that stage, we made some new contacts between the contacts of each individual, which causes the increment in the clustering coefficient of the social contact network. On the other hand, when more partners are chosen from the social contact network, more properties of the social network such as the clustering coefficient become inherited by SexNet. Thus, increasing $p$, we observe increment in the redundancy coefficient of SexNet.

We studied the metrics of networks that may affect the spread of an STI at the population level. Metrics such as the size of giant components and redundancy coefficients provide information about connectivity among the individuals. In our future work, when studying the spread of chlamydia on heterosexual networks, we will measure their impact on the prevalence of chlamydia over SexNets generated using different $p$ values.

Our bipartite networks are being used to study the spread of chlamydia in situations where the probability of infection spread via homosexual activities is negligible. The approach can be extended to model the spread of STIs with mixed heterosexual, bisexual, and homosexual activities such as the spread HIV/AIDS. This can be accomplished by modifying the second stage of the algorithm to include the fractions of the population within these different sexual groupings.
There are still unanswered questions for proving the existence of a heterosexual network with a prescribed joint-degree distribution embedded within a prescribed social network. That is, there are no explicit criteria to guarantee that a heterosexual network with a particular joint-degree distribution can be embedded within a particular social network or not.

We are currently simulating a stochastic agent-based network model on SexNet for the spread of chlamydia and comparing different intervention strategies to control the spread of STIs that are implemented in public health, such as screening, partner treatment, rescreening, and peer referrals (Qu et al. 2020; Azizi et al. 2020). These simulations will use the underlying social contact network to improve the current intervention models by considering the impact of counseling and behavioral changes such as increasing condom use or social contact notification.

Abbreviations
STI: Sexually transmitted infection; SocNet: Social contact network; BJD: Bipartite joint degree; ESocNet: Extended social contact network; BSocNet: Social bipartite network; SexNet: Sexual network; PID: Personal identifier; FID: Friend (social contact) identifier; NC: Number of connected component; RC: Redundancy coefficient.

Acknowledgements
Not applicable.

Authors’ contributions
AA contributed to the design of the algorithm and data analysis, undertook numerical simulations and visualisation, and interpreted results. ZQ and BL reviewed the model design and interpretation. JH contributed to the study, reviewed the model, data analysis and results interpretations, and oversaw and coordinated the investigation. AA wrote the first draft of the article. All authors read and approved the final manuscript.

Funding
This work was supported by the endowment for the Evelyn and John G. Phillips Distinguished Chair in Mathematics at Tulane University and grants from the National Institutes of Health National Institute of Child Health and Human Development (R01HD086794), MIDAS Coordination Center, MIDASUGP2020-2 from National Institute of General Medical Science (3U24GM132013-02S2) and Office of Adolescent Health (TP2AH000013) and the National Institute of General Medical Sciences program for Models of Infectious Disease Agent Study (U01GM097658).

Availability of data and materials
The datasets used during the current study are available from the corresponding author on reasonable request.

Competing interests
The authors declare that they have no competing interests.

Author details
1 Department of Mathematics, University of California, Irvine, CA 92697, USA. 2 Department of Mathematics, The University of Texas at San Antonio, San Antonio, TX 78249, USA. 3 Biocomplexity Institute and Initiative, University of Virginia, Charlottesville, VA 22904, USA. 4 Department of Mathematics, Tulane University, New Orleans, LA 70118, USA.

Received: 28 September 2020   Accepted: 23 December 2020
Published online: 12 April 2021

References
Adimora AA, Schoenbach VJ (2005) Social context, sexual networks, and racial disparities in rates of sexually transmitted infections. J Infect Dis 191(Supplement-1):S115–S122
Amirkhanian YA (2014) Social networks, sexual networks and HIV risk in men who have sex with men. Curr HIV/AIDS Rep 11(1):81–92
Azizi A, Xue L, Hyman JM (2016) A multi-risk model for understanding the spread of chlamydia. In: Mathematical and statistical modeling for emerging and re-emerging infectious diseases. Springer, pp 249–268
Azizi A, Rios-Soto K, MuBayi A, Hyman JM (2017) A risk-based model for predicting the impact of using condoms on the spread of sexually transmitted infections. Infect Dis Model 2:100–112
Azizi A, Dewar J, Hyman JM (2018) Using an agent-based sexual-network model to guide mitigation efforts for controlling chlamydia. bioRxiv, page 233239
Azizi A, Dewar J, Qu Z, Hyman JM (2020) Using an agent-based sexual-network model to analyze the impact of mitigation efforts for controlling chlamydia. arXiv preprint arXiv:2008.05882
Barrett CL, Eubank SG, Smith JP (2005) If smallpox strikes Portland. Sci Am 292(3):S4–61
