Integrated molecular and affiliation network analysis: Core-periphery social clustering is associated with HIV transmission patterns

Kayo Fujimoto, PhD[a,†], Dimitrios Paraskevis, PhD[b,†], Jacky C. Kuo, BS[c], Camden J. Hallmark, MPH[d], Jing Zhao, PhD[e], Andre Hochi, BS[a], Lisa M Kuhns, PhD[f], Lu-Yu Hwang, MD[g], Angelos Hatzakis, PhD[b], John A. Schneider, MD, MPH[h]

[a]Department of Health Promotion, The University of Texas Health Science Center at Houston, 7000 Fannin Street, UCT 2514, Houston, TX 77030

[b]Department of Hygiene, Epidemiology, and Medical Statistics, Medical School, National and Kapodistrian University of Athens, Athens, Greece

[c]Department of Biostatistics and Data Science, The University of Texas Health Science Center at Houston, 7000 Fannin Street, Houston, TX 77030

[d]City of Houston Health Department, Houston, TX

[e]Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030

[f]Division of Adolescent Medicine, Ann & Robert H. Lurie Children's Hospital, and Northwestern University, Feinberg School of Medicine, Department of Pediatrics, 225 E. Chicago Avenue, #161, Chicago, IL 60611

[g]Department of Epidemiology, Human Genetics, and Environmental Science, The University of Texas Health Science Center at Houston, 7000 Fannin Street, Houston, TX 77030

[h]Department of Medicine and Public Health Sciences and the Chicago Center for HIV Elimination, University of Chicago, 5837 South Maryland Avenue MC 5065, Chicago, IL 60637

Abstract

This study investigates the two-mode core-periphery structures of venue affiliation networks of younger Black men who have sex with men (YBMSM). We examined the association between

*Corresponding author: Kayo Fujimoto, PhD; Department of Health Promotion and Behavioral Sciences, School of Public Health, The University of Texas Health Science Center at Houston, 7000 Fannin St. 2514, Houston, TX, 77030; Tel: (713) 500-6697; kayo.fujimoto@uth.tmc.edu.
†Fujimoto, K. and Paraskevis, D. share first authorship.

these structures and HIV phylogenetic clusters, defined as members who share highly similar HIV strains that are regarded as a proxy for sexual affiliation networks. Using data from 114 YBMSM who are living with HIV in two large U.S. cities, we found that HIV phylogenetic clustering patterns were associated with social clustering patterns whose members share affiliation with core venues that overlap with those of YBMSM. Distinct HIV transmission patterns were found in each city, a finding that can help to inform tailored venue-based and network intervention strategies.

Keywords

social determinants of health; racial disparity; HIV transmission networks; phylogenetics; venue affiliation network; core-periphery structure; young black men who have sex with men; sexual networks

INTRODUCTION

In the United States (U.S.), younger Black/African American gay, bisexual, same gender-loving, and other men who have sex with men (hereafter YBMSM) bear a disproportionate burden of HIV, with the 25–34 age group as accounting for 47% of estimated HIV incidence among Black MSM in 2018 (CDC, 2020a). The racial disparities in HIV prevalence and incidence among Black MSM have been well described and reflect structural inequities, such as financial and housing instability, incarceration, lack of health insurance, lack of access to HIV prevention and care service, stigma, and racial discrimination. These inequities increase their HIV risk despite their higher level of engagement in preventive behaviors in comparison to non-Black MSM (Millett et al., 2007; Millett et al., 2012).

HIV spreads within networks that are structured by a pattern of direct and indirect linkages to an infectious agent, and these linkages shape transmission pathways among network members (Friedman and Aral, 2001; Friedman et al., 1997). Structural inequities are inextricably linked to racially segregated sexual networks and their intra-racial network features. These features are characterized by racial homophily in sexual partnerships and disassortative mixing patterns (like as mixing with dislike). In these networks, sexually transmitted infections (STIs) may flow from core individuals (who had four or more sex partners) to peripheral ones (who had only one partner) (Laumann and Youm, 1999). Sexual networks that involve YBMSM are extensive and dynamic with a confluence of high HIV prevalence (Hurt et al., 2012; Schneider et al., 2017). Notably, they are characterized by dense and interconnected sexual networks and partnership characteristics with assortativity (like as mixing with like) by race and with dissortativity by age (Amirkhanian, 2014; Chamberlain et al., 2017; Mustanski et al., 2015; Mustanski et al., 2019).

Social venues as *foci* to potentiate sexual affiliation networks

Sexual network density can be substantially increased through attendance at certain venues, which elevates the risk of contracting HIV (Amirkhanian, 2014). Social venues (such as gay bars, clubs, cruising spots, and gyms) where MSM congregate to meet sex partners are considered a "networked space" (Holloway et al., 2014). Such spaces connect multiple anonymous contacts with unnamed or unknown partners, through which "sexual affiliation

networks" (Frost, 2007) are formed and HIV or other STIs can be transmitted. Social venues play a pivotal role in developing social contexts organized around shared activities, referred to as focused activities, or *foci* (Feld, 1981b). *Foci* anchor social interactions through the process of constraining choice to form and maintain social relationships with certain personal characteristics and, thus, bring homogeneous sets of people together (Feld, 1982a). A group of individuals who share a localized social experience within a wider peer network constitutes "local positions" (Frank et al., 2008a). Grounded in the theory of the "duality of persons and groups" (Breiger, 1974) and the notion of *foci* (Feld, 1981a; Feld, 1982b), the concept of the "duality of local positions" (Frank et al., 2008b) refers to the members of a local position and *foci* that represent the prominent place of the "local positions."

Taking into consideration that YBMSM experience social and structural inequities, we posit that social networks organized around shared activities via affiliation with certain social venues constitute "local positions." Using the above-noted concept of "duality of local positions," we presume that YBMSM's social interactions are channeled by "local positions," which systematically constrain their opportunities for identifying their social or sexual partners and their choices in regard to forming and maintaining relationships. Such "focused choices" (Feld, 1982a) lead to homophily through the process of YBMSM's drawing their partners from certain social venues (i.e., *foci*). As such, it is integral to investigate how venue-based relationships can confer HIV risk on its venue affiliation members.

Social clustering-based core-periphery structure

Several existing studies have employed affiliation network analysis to visualize and describe the structural characteristics of affiliation networks between MSM and social venues in relation to HIV risk and acquisition (Brantley et al., 2017; Fujimoto et al., 2015; Holloway et al., 2014; Oster et al., 2013; Young et al., 2017). Venue affiliation patterns are concentrated around a relatively small group of venues, through which densely connected sexual networks are formed that, in turn, may potentiate HIV transmission (Brantley et al., 2017; Oster et al., 2013). This venue-based clustering pattern is represented by a core and/or peripheral structure in venue affiliation networks (Brantley et al., 2017; Holloway et al., 2014). In this core-periphery structure, the core venues are represented by densely connected clusters of several venues through a high level of network member sharing, while the peripheral venues are represented by loosely connected clusters of a larger number of venues with a lower (or zero) level of network member sharing (Holloway et al., 2014). The viral loads among clients have been found to vary between core and peripheral venues (Brantley et al., 2017), indicating that the core versus peripheral status of venues shapes HIV vulnerability.

These prior studies provide useful insights into core-periphery venue clustering patterns as a potential structural force for generating denser sexual affiliation networks through which HIV may be spread. Rarely, however, have studies comprehensively investigated venue affiliation network by analytically decomposing social clustering patterns in terms of the core versus periphery of both social venues and YBMSM. They also have not provided biological support for the putative HIV transmission patterns.

Our study addresses these knowledge gaps by taking the following approaches. First, we employ affiliation network analysis (Everett and Borgatti, 2013) to analytically decompose and classify multiple patterns of core-peripheral social clustering. These clustering patterns are identified by combinations of the core versus periphery status of the social venues and YBMSM. Second, we take a molecular epidemiological approach to generate putative HIV phylogenetic clusters in which members share highly similar HIV strains. Our study postulates that sexual affiliation networks could be proxied by identifying putative HIV phylogenetic clusters, whereby HIV transmission networks are reconstructed based on the commonality of HIV viral sequences.

Two-mode core-periphery social clustering patterns

Our study defines "core YBMSM" as cohesive subgroups of YBMSM who are connected through a high level of venue co-attendance and "peripheral YBMSM" as subgroups of YBMSM who are loosely connected through either low levels of co-attendance or no venue attendance. The "core venues" are cohesive subgroups of popular venues that share a high level of YBMSM clients, and the "peripheral venues" are loosely connected subgroups of marginal venues that share a lower level of YBMSM clients. Based on these four classifications, we specify the following eight social mechanisms of generating social clustering patterns: (1) core YBMSM who co-attend core venues, (2) core YBMSM who co-attend peripheral venues, (3) peripheral YBMSM who co-attend core venues, (4) peripheral YBMSM who co-attend peripheral venues, (5) core YBMSM who co-attend any venue, (6) peripheral YBMSM who co-attend any venue, (7) any YBMSM who co-attend core venues, and (8) any YBMSM who co-attend peripheral venues. Our study will examine which core-periphery social clustering patterns best reflect the HIV phylogenetic clustering patterns.

HIV phylogenetic clustering as a proxy for sexual affiliation networks

This study proposes a methodological approach to combining affiliation network analysis into molecular phylogenetic analysis by postulating that sexual networks can be proxied by putative HIV phylogenetic clusters whose members share highly similar HIV strains. Molecular phylogenetic analysis enables the identification of genetic HIV transmission clustering patterns by reconstructing HIV transmission networks based on the commonality of HIV viral sequences (Brooks et al., 2013; Sacks-Davis et al., 2012). Membership in the same phylogenetic cluster implies that the viral genetic distances of these cluster members are small, indicating they may have been infected from a closely related source in an overlapping time period (Hassan et al., 2017). Several existing studies have provided greater insight into HIV transmission routes by comparing phylogenetic clustering with social clustering generated through social/affiliation network analysis or contact tracing of sexual/drug-using partners (Chan et al., 2015; Dennis et al., 2013; Fujimoto et al., 2017a; Kostaki et al., 2018; Lee et al., 2009; Morgan et al., 2017; Smith et al., 2009; Wertheim et al., 2017b). Our study integrates affiliation network analysis with molecular phylogenetic analysis to create a powerful methodological tool and uses this tool to identify the social clustering patterns that may potentiate sexual affiliation networks in relation to HIV transmission among the YBMSM who reside in Houston, TX, and Chicago, IL.

DATA AND METHODS

Study design and recruitment

The study data were collected for "iMAN: integrated Molecular and Affiliation Networks" project. iMAN participants consist of sub-samples from a parent study, "YMAP: Young Men's Affiliation Project." YMAP was a longitudinal network study of young MSM in Houston ($N = 378$) and Chicago ($N = 377$) who were recruited between 2014 and 2016, using a respondent-driven sampling method (Heckathorn, 2002), and who were followed up on between 2015 and 2017. Eligibility criteria for YMAP participants were: (a) between the ages of 16 and 29, (b) male sex assigned at birth and current male identification, (c) reported sex (oral or anal) with another man in the past year, (d) frequented at least one social or preventive venue in the past year, and (e) willing to provide biological samples (Fujimoto et al., 2018a). A sample from the iMAN project was collected during the parent YMAP study, primarily for baseline data collection, with the aim of providing molecular-level validation of venue-based affiliation networks in relation to HIV transmission among HIV-seropositive predominantly YBMSM.

iMAN participants consisted of 190 HIV-seropositive younger Black MSM, including multi-race and regardless of Hispanicity, who participated in the YMAP study ($N = 106$ for Houston and $N = 84$ for Chicago). These included six newly infected YMAP participants at follow-up and men who have received antiretroviral drugs in the past, regardless of whether they were antiretroviral drug resistant. Among the 190 iMAN participants, 76 (40%) ($N = 35$ for Houston, and $N = 41$ for Chicago) did not have HIV sequence data due to lack of plasma specimens, insufficient viral load, or missing venue affiliation data. Consequently, our final study sample consisted of 114 HIV-seropositive YBMSM with sequence and survey data ($N = 71$ for Houston, and $N = 43$ for Chicago). We obtained assent/consent from all iMAN participants, and parental consent was waived for minors (under 18 years of age). All participating institutions received approval from the respective institutional review boards (IRB #HSCSPH120830).

Two-mode venue affiliation data

Social and health affiliation data were collected by asking participants whether they had attended a list of 76 social/entertainment venues, such as bars/clubs (36 Houston venues and 40 Chicago venues), in the past 12 months. These venue lists were compiled and selected from formative research as well as through input from community advisory boards for each city. We constructed a two-mode affiliation matrix, **V**, with rows that index the first mode of a set of actors (YBMSM) and columns that index the second mode of a set of venues, with each entry as recording the actor's being either affiliated or not with each venue (coded as 1 for affiliation and 0 otherwise, for each venue type and city). The two-mode matrix, **V**, was then projected, yielding two one-mode symmetric valued matrices: (1) an actor-by-actor projected matrix (**VV'**), whose off-diagonal entries count the number of venues jointly affiliated with each pair of actors and whose diagonal entries count the total number of venues with which each actor is affiliated; and (2) a venue-by-venue projected matrix (**V'V**), whose off-diagonal entries count the number of co-occurring actors with each pair of venues and whose diagonal entries count the total number of actors that occurred at each venue.

These two projections were used to classify core versus peripheral subgroups of YBMSM and venues.

Affiliation network analysis

We took a dual-projection approach to estimate the two-mode core-periphery structure (Everett and Borgatti, 2013) to classify core versus peripheral subgroups of YBMSM and venues. This model uses the mathematical technique of singular value decomposition (Trefethen and Bau III, 1997) to recover the original two-mode matrix, **V**, from its two projections (**VV'** and **V'V**) by finding the eigenvalues (i.e., singular values) and eigenvectors (i.e., singular vectors) associated with both projected matrices to reduce potential loss of structural data through projections (Boyd et al., 2010; Everett and Borgatti, 2013).

We employed the standard continuous core-periphery model implemented in UCINET 6 (Borgatti and Everett, 1999), as it is well designed to valued symmetric data (Everett and Borgatti, 2013). In this model, a "coreness" score is computed for each actor as a core/periphery measure, and the levels of coreness (or closeness to the core of each actor) are estimated. In our application, the continuous core-periphery model was constructed such that the strength of venue co-affiliation between two actors is a function of the closeness of each actor to the core, or, equivalently, the strength of actor co-occurrence between two venues is a function of the closeness of each venue to the core. We chose correlation as the concentration measure for both projections (i.e., computing the eigenvectors of **VV'** and **V'V** separately) based on which optimization procedure was employed to evaluate the goodness of fit between the data matrix and an ideal core-periphery structure and to successively refine the model. This procedure generates the recommended partition that separates the partition of core from the partition of periphery for each projection. These identified core-periphery partitions were then combined and mapped back onto the original two-mode data matrix, **V**, by applying these partitions to the rows and columns of **V**. The resulting two-mode master matrix, **P,** represents a matrix of the two-mode block model of the original matrix, **V**.

Table 1 presents a two-mode block model that represents the core-periphery partitions of YBMSM (row) and venues (column) in a two-mode master matrix, **P**.

Each two-mode block of **A** and **B** represents core YBMSM who affiliate with core venues (**A**) and with peripheral venues (**B**), respectively. Similarly, each two-mode block of **C** and **D** represents peripheral YBMSM who affiliate with core venues (**C**) and with peripheral venues (**D**), respectively. Each two-mode row marginal block of (**A** + **B**) and (**C** + **D**) indicates, regardless of the core-periphery status of venues, that these venues were being affiliated by core YBMSM (**A** + **B**) and by peripheral YBMSM (**C** + **D**). Each two-mode column marginal block of (**A** + **C**) and (**B** + **D**) indicates, regardless of the core versus peripheral subgroups of YBMSM, that they affiliated with core venues (**A** + **C**) and peripheral venues (**B** + **D**).

Each of these two-mode blocks was then projected to construct a corresponding actor-by-actor valued matrix of **AA'**, **BB'**, **CC'**, **DD'**, (**A** + **B**) (**A** + **B**)', (**C** + **D**) (**C** + **D**)', (**A** + **C**) (**A** + **C**)', and (**B** + **D**) (**B** + **D**)', which is presented in Table 2.

In each actor-by-actor projected matrix, off-diagonal entries count the number of core and/or peripheral venues shared by core and/or peripheral YBMSM. We converted the corresponding valued projected matrix into binary by recording the presence (coded as 1) or absence (coded as 0) of co-attendance ties based on a threshold of the median number of shared venues for each pair of actors (minimum = 1). These resulting binarized co-attendance matrices were defined as core-periphery social clustering matrices and were used to statistically assess the association with an actor-by-actor phylogenetic matrix, **G**, that is described below.

HIV sequence data and phylogenetic analysis

Phylogenetics is the basic method for reconstructing within- or inter-species viral genealogy and estimating transmission history based on analysis of DNA or protein sequences (Crandall, 1995; Yokoyama and Gojobori, 1987). Molecular phylogenetics uses nucleotide or amino acid sequences to infer viral genealogy. With the early identification of HIV, it became evident that high genetic diversity provides a hallmark of the virus. Phylogenetic analysis has been applied to investigate epidemiological linkage among people to characterize HIV genetic heterogeneity (i.e.; classification into groups, genotypes, subgenotypes, recombinant forms; (Foley et al., 2016), for investigating the patterns of HIV dispersal across or within different geographic areas, exploring associations between genetic diversity biological properties, and determining clinical characteristics or associations between transmission links and social network-based ties among individuals (Han et al., 2020; Kostaki et al., 2018; Wertheim et al., 2017a).

In our study, HIV sequences were determined through reverse transcription polymerase chain reaction (RT-PCR). HIV RNA was isolated from 0.5 mL of plasma samples. RNA was reverse transcribed to single copy DNA (cDNA). DNA sequences were then aligned in MEGA v7. Patterns of HIV phylogenetic clustering among the study population were investigated by phylogenetic analysis performed on the 122 sequences, along with a random set of subtype B sequences available on the public HIV sequence database ($N = 2,258$) as a reference. A set of the sequences most closely related to the study population also was included as a reference. Phylogenetic trees were inferred, using the maximum-likelihood method with bootstrap evaluation, as implemented in RAxML v8.0.20 (Stamatakis, 2014), using the GTR model and gamma (Γ) distribution. A more detailed description of the HIV sequencing procedures and phylogenetic analysis is provided in the online supplemental material.

Constructing the phylogenetic matrix.—HIV sequences from the study population were defined as "clustered" if they fulfilled two criteria: (a) fell within highly supported clusters that receive >75% bootstrap support (phylogenetic criterion) and (b) consisted of sequences from only the study population. Our criteria for clustering were based on a combination of phylogenetic and geographic criteria, rather than a genetic distance threshold, as we were interested not only in those with direct links but also in all sets of HIV sequences aggregated in a nonrandom manner linked to their epidemiology (Hassan et al., 2017). Based on these criteria, we constructed a binary actor-by-actor phylogenetic matrix, **G**, whose off-diagonal entries record the presence (coded as 1) or absence (coded as 0) of a

phylogenetic link between each pair of actors. The presence of a phylogenetic link indicates that the pair of actors share similar HIV strains and, thus, they are likely to have a common transmission network or closely related source of HIV infection. A more detailed description of the HIV sequencing procedures is provided in the online supplemental material.

Visualization of core-periphery social and phylogenetic clustering

We visualized core-periphery social clustering patterns that are mapped onto phylogenetic links, using NetworkX in Python (Hagberg et al., 2005).

Assessment of similarity of social and phylogenetic clustering

We assessed the similarity of ties based on a phylogenetic matrix (**G**) with each type of blocked core-periphery social clustering matrices by computing the Jaccard similarity coefficient and Hamming distance. The Jaccard coefficient has been employed to measure network stability (Snijders et al., 2010) and to measure the association between the one-mode and the two-mode networks (Fujimoto et al., 2018b) in actor-based models for network dynamics. The Jaccard similarity coefficient is defined as the following:

$$J = \frac{N_{11}}{N_{01} + N_{10} + N_{11}},$$

where N_{11} is the number of ties present at both the social clustering and phylogenetic matrices, N_{01} is the number of ties present at only the social clustering matrix, and N_{10} is the number of ties present at only the phylogenetic matrix. As a sensitivity analysis, we computed the Jaccard similarity coefficient using two different thresholds: the first quartile (*Q1*) and the third quartile (*Q3*) of the numbers of shared venues, and the second quartile (or the median) threshold number (*Q2*) which is described above. As an alternative to the Jaccard similarity coefficient, we computed the Hamming distance by counting the number of cell entries for one matrix that would need to be changed to make it identical to the other matrix. We statistically tested these similarities based on quadratic assignment procedures (QAP).

We also took a regression approach to modeling a phylogenetic clustering matrix by employing the double semi-partialing MR-QAP procedure (Dekker et al., 2007) and the QAP logistic regression model (LR-QAP). The latter model is more appropriate for fitting a binary phylogenetic clustering matrix than is the former one (i.e., a linear probability model on a binary dependent network). For these models, we selected multiple types of core-periphery social clustering as independent matrices based on the significant results of the Jaccard coefficients and the Hamming distance. We conducted both univariate and multivariate analyses by regressing a phylogenetic clustering matrix on these social clustering matrices. All analyses were conducted using UCINET 6.

RESULTS

Sample characteristics

Table 3 presents the distributions of our study samples.

A majority of individuals identified as gay (86%) experienced housing instability in the past 12 months (71%) and had an HIV viral load >10,000 copies/ml (72%). We employed HIV viral load as an indicator of HIV activity and used a clinically relevant cutoff level of 10,000 copies/ml to capture the increased level of transmissibility (Quinn et al., 2000). Included individuals were 24 years of age, on average; had an average of four sex partners during the past six months; and attended an average of five social venues in the past 12 months. There was no statistically significant difference in these variables between the Houston and Chicago samples.

Affiliation network analysis

Table 4 presents the descriptive statistics of the coreness scores and concentration measure (correlation) from the continuous core-periphery model based on the two projections for each city. As we noted previously, the first projection matrix records the strength of venue co-affiliation between two YBMSM (**VV'**). The second projection matrix records the strength of YBMSM co-occurrence between two venues (**V'V**).

For Houston, among 71 YBMSM (**VV'**), the mean coreness score was 0.10 ($SD = 0.07$, $min = 0$, $max = 0.27$). We identified 38 core subgroups of YBMSM ($density = 0.27$) with a concentration score of 0.84 and 33 periphery subgroups of YBMSM ($density = 0.02$). Among 36 venues (**V'V**), the mean coreness score was 0.11 ($SD = 0.13$, $min = 0$, $max = 0.50$). We identified nine core venues ($density = 0.35$) with a concentration score of 0.86 and 27 periphery venues ($density = 0.01$). A majority of core venues were bars, dance clubs, raves, or circuit parties (90%). Similarly, a majority of peripheral venues were bars, dance clubs, raves, or circuit parties (74%), but these venues included sex-related establishments (11%) and other venues related to retail business (11%).

For Chicago, among 43 YBMSM (**VV'**), the mean coreness score was 0.12 ($SD = 0.09$, $min = 0$, $max = 0.39$). We identified 12 core subgroups of YBMSM ($density = 0.42$) with a concentration score of 0.83 and 31 periphery subgroups of YBMSM ($density = 0.05$). Among 40 venues (**V'V**), the mean coreness score was 0.12 ($SD = 0.10$, $min = 0$, $max = 0.32$). We identified 20 core venues ($density = 0.33$) with a concentration score of 0.88 and 20 periphery venues ($density = 0.01$). The core venues consisted of bars and dance clubs (60%) and public cruising venues, such as parks and beaches (25%), while a majority of peripheral venues were bars, dance clubs, raves, or circuit parties (95%).

Based on the results from the core-peripheral partitions/subgroups of YBMSM and venues (i.e., two-mode block matrices) that constitute a two-mode master matrix **P** (refer to Table 1) for Houston, core venues were attended, on average, by 21 members of the core YBMSM subgroup (i.e., two-mode block of **A**) that represents high variance ($SD = 9.79$; $min = 12$, $max = 38$). These core venues were attended, on average, by four members of the peripheral YBMSM subgroup (i.e., two-mode block of **C**; $SD = 3.77$; $min = 0$, $max = 10$). Peripheral venues were attended, on average, by three members of the core YBMSM subgroup (i.e., two-mode block of **B**) ($SD = 2.76$; $min = 0$, $max = 10$) and one member of the peripheral YBMSM subgroup (i.e., two-mode block of **D**; m$ean = 0.6$, $SD = 1.08$; $min = 0$, $max = 5$).

For Chicago, core venues were attended, on average, by seven members of the core YBMSM subgroup (i.e., two-mode block of **A**) (*SD* = 1.90; *min* = 4, *max* = 11) and five members of the peripheral YBMSM subgroup (i.e., two-mode block of **C**) (*SD* = 3.80; *min* = 0, *max* = 13). Peripheral venues were attended, on average, by few members of the core YBMSM subgroup (i.e., two-mode block of **B**) (*mean* = 0.9, *SD* = 1.29; *min* = 0, *max* = 5) and few members of the peripheral YBMSM subgroup (i.e., two-mode block of **D**) (*mean* = 0.9, *SD* = 1.23; *min* = 0, *max* = 4).

In regard to the results from an actor-by-actor projected valued matrix that corresponds to each two-mode block matrix based on a two-mode block matrix **P** (Table 2) for Houston, members of the core YBMSM subgroup co-attended a median of three core venues (**AA'**), a median of 0 peripheral venues (**BB'**), and a median of three combined core and peripheral venues ((**A** + **B**) (**A** + **B**)'). Members of the peripheral YBMSM subgroup co-attended a median of 0 core venues (**CC'**), a median of 0 peripheral venues (**DD'**), and a median of 0 combined core and peripheral venues ((**C** + **D**) (**C** + **D**)'). Combined core and peripheral YBMSM subgroups affiliated with a median of 1 core venue ((**A** + **C**) (**A** + **C**)') and a median of 0 peripheral venues ((**B** + **D**) (**B** + **D**)').

For Chicago, members of the core YBMSM subgroup co-attended a median of seven core venues (**AA'**), a median of 0 peripheral venues (**BB'**), and a median of seven combined core and peripheral venues ((**A** + **B**) (**A** + **B**)'). Members of the peripheral YBMSM subgroup co-attended a median of 0 core venues (**CC'**), a median of 0 peripheral venues (**DD'**), and a median of 0 combined core and peripheral venues ((**C** + **D**) (**C** + **D**)'). Combined core and peripheral YBMSM subgroups affiliated with a median of one core venue ((**A** + **C**) (**A** + **C**)') and a median of 0 peripheral venues ((**B** + **D**) (**B** + **D**)').

Phylogenetic analysis

Houston.—We identified 12 phylogenetic clusters, consisting of nine dyads (two sequences) and three triads (three sequences). In relation to the results from the core-periphery analysis, 15 members of the core YBMSM subgroups and 12 members of the peripheral YBMSM subgroups belonged to one of these phylogenetic clusters.

Chicago.—We identified six phylogenetic clusters, consisting of three dyads, one triad, one tetrad, and one pentad. In relation to the results from the core-periphery analysis, four members of the core YBMSM subgroups and 14 members of the peripheral YBMSM subgroups belonged to one of these phylogenetic clusters.

Visualizing social and phylogenetic clustering patterns

Figure 1 presents a visualization of the affiliation networks between YBMSM and venues by core-versus-periphery subgroups of YBMSM and venues (red for core and blue for periphery) in relation to the phylogenetic connections (purple ties). In this graph, affiliation ties that connect members of phylogenetic clusters are indicated by bold lines (bold orange ties to core venues, and bold blue ties to peripheral venues).

Houston.—A majority of phylogenetic clusters (seven out of 12 clusters) consisted of within-cluster members with a mixture of core and peripheral subgroups, representing disassortative mixing in the degree of co-attendance at any venues with their peers. In these mixed clusters, member(s) of the core subgroup tended to affiliate with both core and periphery venues and were phylogenetically connected to the member(s) of the periphery subgroup who attended only a few marginal venues or no venues.

Chicago.—A majority (four out of six) of phylogenetic clusters consisted of within-cluster members of either core or periphery subgroups, representing assortative mixing in the degree of co-attendance at any venues with their peers. Phylogenetic clusters that consisted only of members of the periphery subgroup tended to affiliate primarily with core venues. In phylogenetic clusters whose members were all core subgroups (one dyad) or a mixture of core and peripheral subgroups (one dyad and one tetrad), all members of the core subgroup attended core venues exclusively, with the exception of one member in a tetrad, who affiliated with both core and peripheral venues.

Testing associations between social clustering and phylogenetic clustering

The following results were based on each of the block matrices of the binarized actor-by-actor projected core-periphery social clustering (Table 2) and the corresponding blocks of a binary actor-by-actor phylogenetic matrix **G**, which are described above.

Network similarity measures.—In Houston, the tie profiles of the two social clustering matrices $(\mathbf{A} + \mathbf{C}) (\mathbf{A} + \mathbf{C})'$ and $(\mathbf{B} + \mathbf{D}) (\mathbf{B} + \mathbf{D})'$ showed significant similarities to those of the corresponding phylogenetic clustering matrices. More specifically, phylogenetic connections were significantly associated with having shared attendance of at least one core venues (*Jaccard coefficient* = 0.01, $p < 0.05$ and Hamming distance = 0.59, $p < 0.01$) with a threshold of 1 for $(\mathbf{A} + \mathbf{C}) (\mathbf{A} + \mathbf{C})'$ and as having shared attendance of at least one periphery venue (*Jaccard coefficient* = 0.02, $p < 0.05$ and Hamming distance = 0.09, $p < 0.05$) with a threshold of 1 for $(\mathbf{B} + \mathbf{D}) (\mathbf{B} + \mathbf{D})'$). In Chicago, the results of the Jaccard similarity coefficients and the Hamming distance indicate a non-significant similarity between the two tie profiles of the phylogenetics and any of the social clustering block matrices.

Sensitivity analysis using different threshold values.—In Houston, for $(\mathbf{A} + \mathbf{C}) (\mathbf{A} + \mathbf{C})'$, the first quartile ($Q1 = 0$) was set to the minimum threshold number of 1 that is identical with the afore-mentioned finding from the median threshold number ($Q2 = 1$). For the third quartile ($Q3 = 2$), the Jaccard coefficient was not significant. However, when we restrict to the social clustering matrix of **AA'** (i.e., core YBMSM connecting to other core YBMSM via jointly affiliating with core venues) using the threshold number of $Q3 = 2$ based on $(\mathbf{A} + \mathbf{C}) (\mathbf{A} + \mathbf{C})'$, the result became significant (*Jaccard coefficient* = 0.01; $p < 0.05$). The equivalent sensitivity analysis for $(\mathbf{B} + \mathbf{D}) (\mathbf{B} + \mathbf{D})'$ was not conducted, as all quartiles were zero ($Q1 = Q2 = Q3 = 0$) that are identical with the afore-mentioned finding from the median threshold number.

Regression analysis.—In Houston, for both MR-QAP and LR-QAP, the univariate model including each $(\mathbf{A} + \mathbf{C}) (\mathbf{A} + \mathbf{C})'$ and $(\mathbf{B} + \mathbf{D}) (\mathbf{B} + \mathbf{D})'$ social clustering block matrix

respectively indicated significant association with phylogenetic clustering. The MR-QAP coefficients were 0.01 ($p < 0.05$) for $(\mathbf{A} + \mathbf{C})$ $(\mathbf{A} + \mathbf{C})$', and 0.02 ($p < 0.05$) for $(\mathbf{B} + \mathbf{D})$ $(\mathbf{B} + \mathbf{D})$'. The LR-QAP coefficients were 1.71 ($p < 0.05$) for $(\mathbf{A} + \mathbf{C})$ $(\mathbf{A} + \mathbf{C})$' and 1.45 ($p < 0.05$) for $(\mathbf{B} + \mathbf{D})$ $(\mathbf{B} + \mathbf{D})$'. Then, we fit multivariate models with including the $(\mathbf{A} + \mathbf{C})$ $(\mathbf{A} + \mathbf{C})$' block first, and then adding the $(\mathbf{B} + \mathbf{D})$ $(\mathbf{B} + \mathbf{D})$' block matrix to this univariate model. The result indicates that the coefficient for $(\mathbf{A} + \mathbf{C})$ $(\mathbf{A} + \mathbf{C})$' was 0.01 ($p < 0.05$) for the MR-QAP, and 1.57 ($p < 0.05$) for the LR-QAP. For $(\mathbf{B} + \mathbf{D})$ $(\mathbf{B} + \mathbf{D})$', the coefficient was 0.02 ($p < 0.05$) for the MR-QAP and 1.21 ($p < 0.05$) for the LR-QAP.

DISCUSSION

The social lives of YBMSM are spatially organized by the social and structural inequities that surround these individuals, which may shape their venue-affiliation patterns and potentiate dense sexual networks and may increase HIV vulnerability and acquisition. From the perspective of the social determinants of health, intersectional stigma is inextricably linked to structural inequities and social risks through a venue-affiliation network. YBMSM occupy multiple and intersecting stigmatized identities related to race, sexuality, socioeconomic status, and cultural expectations that reflect the macro-level social-structural inequities that they experience (Fields et al., 2016). These intersectional forms of stigma may shape social risks (e.g., being arrested or being discriminated against due to sexuality or low socioeconomic status) when YBMSM engage in sexual relationships at certain venues. These risks influence their decisions about where to seek sex partners (Parker et al., 2017) and where they should avoid congregating (Fujimoto et al., 2020). Consequently, social venues serve as a place where YBMSM are connected to sexual network members and which may increase their vulnerability to HIV.

Our study indicates that the two-mode core-periphery problem could be framed through the notions of intersectional stigma, social risk, and the social production of space (Parker et al., 2017). We posit that social venues serve as common points that may potentiate HIV transmission (or multiple anonymous contacts that lead to transmission), and we identified social clustering patterns that are associated with sexual affiliation networks in relation to HIV transmission. Further, our study examined whether individuals who were socially connected through (a) affiliation with the same popular/core venues with greater overlap of YBMSM or (b) affiliation with marginal/peripheral venues with little to no overlap of YBMSM, which was associated with sharing similar phylogenetics or genetic markers of HIV viral strains.

In Houston, our findings indicated that sexual affiliation networks were potentiated through affiliation with a cohesive subgroup of popular social venues or through affiliation with loosely connected clusters of marginal social venues in Houston. Such sexual affiliation networks may comprise of a mixture of core and peripheral subgroups of YBMSM. Our findings are consistent with previous studies that reported core-periphery venue-clustering patterns in relation to HIV infection (Brantley et al., 2017; Oster et al., 2013). Further, our results indicate that these sexual affiliation networks are potentiated through affiliating with a cohesive subgroup of popular social venues with higher density (based on our sensitivity analysis) by the members of core subgroups of YBMSM.

In Chicago, however, our findings indicate that phylogenetic clustering was not significantly associated with any of the social clustering patterns. These non-significant results, however, could be partially explained by the sparseness of phylogenetic ties among members of core subgroups of YBMSM. Our data indicate that only four out of the total 12 members of core subgroups shared highly similar HIV strains with their peers. Under such low density, network similarity measures, especially for Hamming distance that treats a joint absence of ties as similarity, would present little variation among the tie profiles, which may cause difficulty in discerning network similarity (Hanneman and Riddle, 2005). Despite these non-significant results, the network visualization (Figure 1) indicates that sexual affiliation networks might be potentiated primarily among members of core subgroups of YBMSM with a majority of them (three out of four members) affiliating exclusively with popular social venues, as well as among members of peripheral subgroups of YBMSM who tend to affiliate with core social venues.

Our findings indicate strong differences between cities in the composition of members within the sexual affiliation networks. For Houston, sexual affiliation networks were characterized by a disassortative mixing pattern with respect to the level of venue co-attendance with peers. As such, there is a tendency for core subgroup members to affiliate with both core and peripheral venues and for peripheral subgroup members to affiliate with very few or no venues. In Chicago, in contrast, sexual affiliation networks were characterized by an assortative mixing pattern; i.e., they consist of members from either core or peripheral subgroups, and there is a tendency for both core and peripheral subgroup members to affiliate primarily with core venues.

These results may indicate potential differences in venue affiliation patterns in relation to the ongoing HIV transmission between the two cities. In Houston, core and peripheral venues serve equally as common points of potential HIV transmission possibly via sexual affiliation networks. In this scenario, these members could then bridge HIV to others who have low levels of co-attendance with peers or outside of venue settings. In Chicago, in contrast, core venues may have a potential for playing an important role in connecting anonymous, multiple contacts that potentiate sexual affiliation networks where HIV transmission could be confined based upon crude visualization.

These differences between cities could be partially explained by the selection of social venues in each city. For both cities, core and peripheral venues are represented mainly by venues that are associated with alcohol (e.g., bars, dance clubs) or a party-type environment (e.g., raves, circuit parties). Chicago's core venues, however, also include public cruising spots, such as parks and beaches, for finding sex partners, which Houston's core venues do not include. This difference may contribute to differences in HIV transmission in these two cities, but future research to examine the role that specific venue types play in transmitting disease is needed.

The differences between Houston and Chicago also could be explained by the geographic concentration patterns among these venues (Fujimoto et al., 2017b) and certain environmental or cultural factors that surround local MSM communities (Fujimoto et al., 2020) that may contribute to the local dynamics of venue affiliation patterns. In Houston,

social venues are geographically concentrated within a gay enclave area (i.e., Montrose) with high HIV prevalence rates and with more racially/ethnically diverse YMSM. This may encourage YBMSM to affiliate with both popular and marginal venues, and, once their sexual affiliation networks are formed or maintained, their memberships then expand to others outside of the venue settings. Future research that investigates social mechanisms using longitudinal data is warranted.

In Chicago, by contrast, social venues are concentrated not only in major areas of a gay enclave on the north side of the city (i.e., Boystown) but also in other areas of the mid-east side of the central business district (i.e., the Loop). The city also has small pockets of concentration on the south side, with high and middle levels of HIV prevalence rates. These citywide segregations of gay enclaves reflect the longstanding racism among MSM and institutional racism, leading to a potentially more independent YBMSM community. This sociocultural environment may influence YBMSM's decision to avoid certain venues in a main gay enclave (Fujimoto et al., 2020) as well as their decision to attend popular gay venues within their own community that are frequented mainly by members of racial/ethnic minority groups, where same-sex sexual relationships are supported and venue-related risks are perceived to be low (Parker et al., 2017).

MSM who are living with HIV tend to be clustered tightly in a geographic space, and neighborhoods affect sexual expression, racial/ethnic identity, and health outcomes (Egan et al., 2011). Our study suggests that co-attendance with popular venues in Houston, in particular, is a robust social clustering pattern that is associated with the putative HIV phylogenetic clustering pattern. This indicates that sexual affiliation networks are organized around these popular venues, and co-attendance of these venues may be contingent on local characteristics in the MSM community.

Our study, iMAN, has multiple strengths. First, iMAN synthesized affiliation network analysis with HIV phylogenetic analysis, which allows for the assessment of how much of the ongoing transmission is associated with venue-based social clustering patterns, and is validated with biological/molecular support. Our refined analysis goes beyond a previous study that explored overlaps between genetic and social clusters, in which social venues are defined in general classifications, such as "sauna' and "internet" (Lee et al., 2009).

Second, our study shifted the analytic focus from one-mode social/sexual networks formed by direct contacts to the extended boundary of venue affiliation networks through which contacts with unnamed or unknown partners can be connected. Affiliation data have been found to be more robust to sampling biases when compared to data based on the sampling of direct sexual contact tracing (Frost, 2007). As such, these affiliation networks offer more reliable data to approximate sexual networks. To illustrate this, our supplemental analysis showed that phylogenetic links for Houston participants overlapped with only two one-mode direct social ties (11%) in the main network component that was identified in social networks of combined peer referral, social and sexual ties. Further, some of the non-overlapped phylogenetic links connected three separate components into the main component. In contrast, for Chicago participants, there was no overlap of phylogenetic links

with one-mode social ties, and none of the non-overlapped phylogenetic links connect any components to the main component.

Third, as a methodological strength, our study employed the dual-projection approach, as compared to the alternative direct approach (Borgatti and Everett, 1997; Borgatti and Halgin, 2011), to model a two-mode core-periphery structure (Everett and Borgatti, 2013) that provides insight into the social mechanisms that generate core-periphery social clustering patterns. This direct approach considers all venues (or all YBMSM) to have equal weight when evaluating whether YBMSM (or venues) are in the core or periphery subgroup and does not consider the popularity of venues that attract more YBMSM (indicated by high co-occurrence of YBMSM) when determining the core subgroup of YBMSM.

There are several limitations of this study. First, due to the cross-sectional nature of our data, our study was limited to assessing association, not causality, and did not empirically identify transmission pathways. Second, due primarily to a lack of sufficient biological specimens or an insufficient viral load, it may be that selection bias occurred. Specifically, our samples came from participants who were not being treated for HIV (if they were being treated, their viral load should be near zero, assuming that they do not have a drug-resistant type). This potential limitation, however, could increase the significance of our study, as those with a high viral load are more contagious. Further, iMAN, as a community-based sample, does not rely on patients who are seeking clinical care, which comprises the majority of phylogenetic analyses in the U.S. Thus, our results are expected to better inform social clustering patterns associated with ongoing HIV transmission.

Third, our study may suffer from a lack of generalizability to other urban U.S. contexts, considering the differences in particular political, cultural, and historical factors that surround different regions of the U.S. that may result in different patterns of venue networks in gay communities (Fujimoto et al., 2017b; Fujimoto et al., 2017d). We studied only Houston and Chicago, and, as such, there may be innate differences in these cities and regions as compared to others that could contribute to the research findings. Nevertheless, our results may provide a roadmap for applications to other northern and southern urban areas at the national level. For instance, our results from Chicago could be applicable to other larger, segregated Midwest cities, while the ones from Houston could apply to other large Southern cities. With the advancement of graph-based deep learning methodologies and the utilities that apply these methods to social network research related to HIV and MSM populations (Xiang et al., 2021; Xiang et al., 2019), there is a great potential for our data to be integrated with transfer learning techniques (Devlin et al., 2018; Ganin and Lempitsky, 2015). Such techniques enable the re-use of a pre-trained model for other cities (i.e., as an extended task), providing a data science approach to generalize knowledge to different settings.

Despite the limitations of our study, its interdisciplinary nature enabled us to chart new territory in terms of accounting for underlying social determinants of HIV transmission. The results of this study provide new directions for developing and implementing venue-based network interventions that are informed by molecular and network science. Our findings aid in focusing interventions in and around social venues for those most at risk for HIV specific

to each city. Interventions may include contact tracing, information dissemination, HIV testing, and/or pre-exposure prophylaxis, or PrEP (a daily pill that prevents HIV acquisition among people who are HIV-negative).

Our study suggests that effective implementation strategies would involve targeting major social venues in Houston gay enclaves as well as less-established venues with a combination of network-based interventions. These network-based interventions may require working through networks of individuals rather than through venues, given the unique venue-affiliation structure. An effective solution to better reach this community in Houston may be through peer-driven approaches, such as the use of a social network strategy (CDC, 2020b) or induction interventions (Valente, 2012), whereby peers identify people in their social network for HIV prevention interventions.

An additional approach may be an alteration intervention, whereby the network itself is modified (Valente, 2012). Alteration could occur through the introduction of a drop-in center that partners with specific social venues to provide HIV prevention services. This could serve two complementary purposes. First, it could alter the pattern of venue affiliation, which, in turn, could restructure the patterns of sexual affiliation networks to include more users of prevention services. Second, it could increase network proximity of specific venues that partner with the drop-in center and, thus, drive more contacts with these venues. Such specificity in focus will be important as we continue to move incidence downward in ending the HIV epidemic.

To continue harnessing insights from phylogenetic approaches, it is imperative to consider ethical issues around the use of molecular data collected through research and public health surveillance. A growing number of molecular epidemiological studies have identified and characterized emerging local transmission clusters in the U.S. metropolitan areas using molecular surveillance data (Fujimoto et al., 2021; Guilamo-Ramos et al., 2020; Ragonnet-Cronin et al., 2021). Although these studies reported the notable vulnerability of racial/ethnic minority populations to HIV viral transmission dynamics, there are growing concerns among researchers and community stakeholders. These concerns include the possibility of further stigmatization of communities or unintended identification of individuals (Mehta et al., 2019).

Perhaps the most frequently cited forewarning is that data may be used to support HIV criminalization or immigration penalties, and/or that surveillance data are re-purposed from clinical to surveillance or research uses without explicit consent (Molldrem & Smith, 2020). Current methods of HIV phylogenetic analysis cannot establish the directionality of transmission between network members within putatively identified (Rose et al., 2019). Nonetheless, there is still unease that next-generation sequencing will allow researchers to correctly identify the direction of transmission between individuals (Zhang et al., 2021), which could then increase the risks of criminalization or violations of privacy (Mehta, Schairer, & Little, 2019).

Data gathered as part of research are protected by certificates of confidentiality (Dawson et al., 2020), however, surveillance data remain vulnerable to state and local statutes or

regulations governing release (NASTAD, 2018). For our proposed integrated molecular and affiliation network methodologies to provide conceptual enrichment in our scientific knowledge of racial/ethnic disparity in HIV infection and network science, we must emphasize the need and importance of a surveillance-research-community partnership to address such ethical issues. Only then, can we improve the health of communities most impacted and effectively move towards the shared goal of HIV elimination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement:

REFERENCES

Amirkhanian YA (2014). Social networks, sexual networks and HIV risk in men who have sex with men. Current HIV/Aids Reports, 11(1), 81–92. 10.1007/s11904-013-0194-4. [PubMed: 24384832]

Borgatti SP & Everett MG (1997). Network analysis of 2-mode data. Social Networks, 19(3), 243–269. 10.1016/S0378-8733(96)00301-2.

Borgatti SP & Everett MG (1999). Models of core/periphery structures. Social Networks 21(4), 375–395. 10.1016/S0378-8733(99)00019-2.

Borgatti SP & Halgin D (2011). Analyzing affiliation networks. In Carrington P & Scott J (Eds.), Sage handbook of social network analysis. Sage Publications.

Boyd JP, Fitzgerald WJ, Mahutga MC, & Smith DA (2010). Computing continuous core/periphery structures for social relations data with MINRES/SVD. Social Networks, 32(2), 125–137. 10.1016/j.socnet.2009.09.003.

Brantley M, Schumacher C, Fields EL, Perin J, Safi AG, Ellen JM, Muvva R, Chaulk P, & Jennings JM (2017). The network structure of sex partner meeting places reported by HIV-infected MSM: Opportunities for HIV targeted control. Social Science & Medicine, 182, 20–29. 10.1016/j.socscimed.2017.04.006. [PubMed: 28411524]

Breiger RL (1974). Duality of persons and groups. Social Forces, 53(2), 181–190.

Brooks JI, Niznick N, Ofner M, Merks H, & Angel JB (2013). Local phylogenetic analysis identifies distinct trends in transmitted HIV drug resistance: Implications for public health interventions. BMC Infectious Diseases, 13, 509. 10.1186/1471-2334-13-509. [PubMed: 24171696]

Centers for Disease Control and Prevention. (2020, 5). Estimated HIV incidence and prevalence in the United States, 2014–2018. https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-25-1.pdf.

Centers for Disease Control and Prevention. (2020, 7). Social network strategy for HIV testing recruitment. https://www.cdc.gov/hiv/effective-interventions/diagnose/social-network-strategy?Sort=Title%3A%3Aasc&Intervention%20Name=Social%20Network%20Strategy.

Chamberlain N, Mena LA, Geter A, & Crosby RA (2017). Is sex with older male partners associated with higher sexual risk behavior among young black MSM? AIDS and Behavior, 21(8), 1–7. 10.1007/s10461-017-1699-4.

Chan PA, Hogan JW, Huang A, DeLong A, Salemi M, Mayer KH, & Kantor R (2015). Phylogenetic investigation of a statewide HIV-1 epidemic reveals ongoing and active transmission networks among men who have sex with men. Journal of Acquired Immune Deficiency Syndromes, 70(4), 428. 10.1097/QAI.0000000000000786. [PubMed: 26258569]

Crandall KA (1995). Intraspecific phylogenetics: Support for dental transmission of human immunodeficiency virus. Journal of Virology, 69(4), 2351–2356. 10.1128/JVI.69.4.2351-2356.1995. [PubMed: 7884881]

Dekker D, Krackhardt D, Snijders TAB, 2007. Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. Psychometrika 72, 563–581. https://doi.org/doi:10.1007/s11336-007-9016-1. [PubMed: 20084106]

Dennis AM, Murillo W, Hernandez A, Guardado ME, Nieto AI, Lorenzana de Rivera I, Eron JJ, & Paz-Bailey G (2013). Social network based recruitment successfully reveals HIV-1 transmission networks among high-risk individuals in El Salvador. Journal of Acquired Immune Deficiency Syndromes, 63(1), 135–141. 10.1097/QAI.0b013e318288b246. [PubMed: 23364512]

Devlin J, Chang M-W, Lee K, & Toutanova K (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv. arXiv:1810.04805.

Egan JE, Frye V, Krutz SP, Latkin CA, Chen M, Tobin K, Yang C, & Koblin BA (2011). Migration, neighborhoods, and networks: Approaches to understanding how urban environmental conditions affect syndemic adverse health outcomes among gay, bisexual and other men who have sex with men. AIDS Behavior, 15, S35–S50. https://doi.org/doi:10.1007/s10461-011-9902-5. [PubMed: 21369730]

Everett MG & Borgatti SP (2013). The dual projection approach for 2-mode networks. Social Networks, 35(2), 204–210. 10.1016/j.socnet.2012.05.004.

Feld SL (1981). The focused organization of organizational ties. American Journal of Sociology, 86(5), 1015–1035.

Feld SL (1982). Social structural determinants of similarity among associates. American Sociological Review, 47(6), 797–801.

Fields E, Morgan A, & Sanders RA (2016). The intersection of sociocultural factors and health-related behavior in lesbian, gay, bisexual, and transgender youth: Experiences among young black gay males as an example. Pediatric Clinics of North America, 63(6), 1091–1106. 10.1016/j.pcl.2016.07.009. [PubMed: 27865335]

Foley BT, Leitner T, Paraskevis D, & Peeters M (2016). Primate immunodeficiency virus classification and nomenclature. Infection, Genetics and Evolution, 46, 150–158. 10.1016/j.meegid.2016.10.018.

Frank KA, Chandra M, Schiller KS, Riegle-Crumb C, Mueller AS, Crosnoe R, & Pearson J (2008). The social dynamics of mathematics coursetaking in high schools. American Journal of Sociology, 113(6), 1645–1696. 10.1086/587153

Friedman SR & Aral S (2001). Social networks, risk-potential networks, health, and disease. Journal of Urban Health, 78(3), 411–418. 10.1093/jurban/78.3.411. [PubMed: 11564845]

Friedman SR, Neaigus A, Jose B, Curtis R, Goldstein M, Ildefonso G, Rothenberg RB, & Des Jarlais DC (1997). Sociometric risk networks and risk for HIV infection. American Journal of Public Health, 87(8), 1289–1296. 10.2105/ajph.87.8.1289. [PubMed: 9279263]

Frost SDW (2007). Using sexual affiliation networks to describe the sexual structure of a population. Sexually Transmitted Infections, 83, i37–i42. 10.1136/sti.2006.023580. [PubMed: 17664363]

Fujimoto K, Bahl J, Wertheim JO, Del Vecchio N, Hicks JT, Damodaran L, Hallmark CJ, Lavingia R, Mora R, Carr M, Yang B, Schneider JA, Hwang L-Y, McNeese M, 2021. Methodological synthesis of Bayesian phylodynamics, HIV-TRACE, and GEE: HIV-1 transmission epidemiology in a racially/ethnically diverse Southern US context. Scientific reports, 11(1), 3325. 10.1038/s41598-021-82673-8. [PubMed: 33558579]

Fujimoto K, Cao M, Kuhns LM, Li DH, & Schneider JA (2018). Statistical adjustment of network degree in respondent-driven sampling estimators: Venue attendance as a proxy for network size among young men who have sex with men. Social Networks, 54, 118–131. 10.1016/j.socnet.2018.01.003. [PubMed: 29910531]

Fujimoto K, Coghill LM, Weier CA, Hwang L-Y, Kim JY, Schneider JA, Metzker ML, & Brown JM (2017). Lack of support for socially connected HIV-1 transmission among young adult black men

who have sex with men. AIDS Research and Human Retroviruses, 33(9), 935–940. 10.1089/AID.2016.0228. [PubMed: 28398775]

Fujimoto K, Turner R, Kuhns LM, Kim JY, Zhao J, & Schneider JA (2017). Network centrality and geographical concentration of social and service venues that serve young men who have sex with men. AIDS and Behavior, 21(12), 3578–3589. 10.1007/s10461-017-1711-z. [PubMed: 28220310]

Fujimoto K, Wang P, Kuhns LM, Ross MW, Williams ML, Garofalo R, Klovdahl AS, Laumann EO, & Schneider JA (2017). Multiplex competition, collaboration, and funding networks among social and health organizations: Towards organization-based HIV interventions for young men who have sex with men. Medical Care, 55(2), 102–110. 10.1097/MLR.0000000000000595. [PubMed: 27676400]

Fujimoto K, Wang P, Li DH, Kuhns LM, Amith M, & Schneider JA (2020). Collective avoidance of social and health venues and HIV racial inequities: Network modeling of venue avoidance on venue affiliation, social networks, and HIV risk. Health Education & Behavior, 47(2), 202–212. 10.1177/1090198119876240. [PubMed: 32090656]

Fujimoto K, Wang P, Ross MW, & Williams ML (2015). Venue-mediated weak ties in multiplex HIV risk transmission networks among drug-using male sex workers and associates. American Journal of Public Health, 105(6), 1128–35. 10.2105/AJPH.2014.302474. [PubMed: 25880956]

Ganin Y & Lempitsky V (2015). Unsupervised domain adaptation by backpropagation, International conference on machine learning. Proceedings of Machine Learning Research, 37, 1180–1189.

Guilamo-Ramos V, Thimm-Kaiser M, Benzekri A, Chacón G, López OR, Scaccabarrozzi L, Rios E, 2020. The invisible US Hispanic/Latino HIV crisis: addressing gaps in the national response. American Journal of Public Health, 110, 27–31. 10.2105/AJPH.2019.305309. [PubMed: 31725313]

Hagberg A, Schult D, & Swart P (2005). Networkx: Python software for the analysis of networks. Mathematical Modeling and Analysis, Los Alamos National Laboratory.

Han X, Zhao B, An M, Zhong P, & Shang H (2020). Molecular network-based intervention brings us closer to ending the HIV pandemic. Frontiers of Medicine, 14(2), 136–148. 10.1007/s11684-020-0756-y. [PubMed: 32206964]

Hanneman RA and Riddle M (2005). Introduction to Social Network Methods (free introductory textbook on social network analysis). University of California Riverside. https://faculty.ucr.edu/^hanneman/.

Hassan AS, Pybus OG, Sanders EJ, Albert J, & Esbjörnsson J (2017). Defining HIV-1 transmission clusters based on sequence data. AIDS, 31(9), 1211–1222. 10.1097/QAD.0000000000001470. [PubMed: 28353537]

Heckathorn DD (2002). Respondent driven sampling II: Deriving valid population estimates from chainreferral samples of hidden populations. Social Problems, 49(1), 11–34.

Holloway IW, Rice E, Kipke MD (2014). Venue-based network analysis to inform HIV prevention efforts among young gay, bisexual, and other men who have sex with men. Preventive Science, 15(3), 419–427. 10.1007/s11121-014-0462-6.

Hurt CB, Beagle S, Leone PA, Sugarbaker A, Pike E, Kuruc J, Foust EM, Eron JJ Jr, Cohen MS, & Hightow-Weidman LB (2012). Investigating a sexual network of black men who have sex with men: Implications for transmission and prevention of HIV infection in the United States. Journal of Acquired Immune Deficiency Syndromes, 61(4), 515–521. 10.1097/QAI.0b013e31827076a4. [PubMed: 22972020]

Kostaki E-G, Nikolopoulos GK, Pavlitina E, Williams L, Magiorkinis G, Schneider J, Skaathun B, Morgan E, Psichogiou M, Daikos GL, Sypsa V, Smyrnov P, et al. (2018). Molecular analysis of Human Immunodeficiency Virus Type 1 (HIV-1)-infected individuals in a network-based intervention (Transmission Reduction Intervention Project): Phylogenetics identify HIV-1-infected individuals with social links. The Journal of Infectious Diseases, 218(5), 707–715. 10.1093/infdis/jiy239. [PubMed: 29697829]

Laumann EO & Youm Y (1999). Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: A network explanation. Sexually Transmitted Diseases, 26(5), 250–261. 10.1097/00007435-199905000-00003. [PubMed: 10333277]

Lee SS, Tam DKP, Tan Y, Mak WL, Wong KH, Chen JHK, & Yam WC (2009). An exploratory study on the social and genotypic clustering of HIV infection in men having sex with men. AIDS, 23(13), 1755–1764. 10.1097/QAD.0b013e32832dc025. [PubMed: 19609202]

Mehta SR, Schairer C, & Little S (2019). Ethical issues in HIV phylogenetics and molecular epidemiology. Current Opinion in HIV and AIDS, 14(3), 221–226. 10.1097/COH.0000000000000538. [PubMed: 30946143]

Millett GA, Flores SA, Peterson JL, & Bakeman R (2007). Explaining disparities in HIV infection among Black and White men who have sex with men: A meta-analysis of HIV risk behavior. AIDS, 21(15), 2083–2091. 10.1097/QAD.0b013e3282e9a64b. [PubMed: 17885299]

Millett GA, Peterson JL, Flores SA, Hart TA, Jeffries WL 4th, Wilson PA, Rourke SB, Heilig CM, Elford J, Fenton KA, & Remis RS (2012). Comparisons of disparities and risks of HIV infection in black and other men who have sex with men in Canada, UK, and USA: A meta-analysis. Lancet, 380(9839), 341–348. https://doi.org/S0140-6736(12)60899-X. [PubMed: 22819656]

Molldrem S, & Smith AK (2020). Reassessing the ethics of molecular HIV surveillance in the era of cluster detection and response: Toward HIV data justice. The American Journal of Bioethics, 20(10), 10–23. 10.1080/15265161.2020.1806373. [PubMed: 32945756]

Morgan E, Nyaku AN, Richard T, & Schneider JA (2017). Determinants of HIV phylogenetic clustering in Chicago among young black men who have sex with men from the uConnect cohort. Journal of Acquired Immune Deficiency Syndromes, 75(3), 265–270. 10.1097/QAI.0000000000001379. [PubMed: 28328553]

Mustanski B, Birkett M, Kuhns LM, Latkin CA, & Muth SQ (2015). The role of geographic and network factors in racial disparities in HIV among young men who have sex with men: An egocentric network study. AIDS and Behavior, 19(6), 1037–1047. 10.1007/s10461-014-0955-0. [PubMed: 25430501]

Mustanski B, Morgan E, D'Aquila R, Birkett M, Janulis P, & Newcomb ME (2019). Individual and network factors associated with racial disparities in HIV among young men who have sex with men: Results from the RADAR cohort study. Journal of Acquired Immune Deficiency Syndromes, 80(1), 24–30. 10.1097/QAI.0000000000001886. [PubMed: 30365451]

NASTAD. (2018). HIV Data Privacy and Confidentiality: Legal & Ethical considerations for Health Department Data Sharing. Accessed March 28, 2021. https://www.nastad.org/sites/default/files/Uploads/2018/nastad-hiv-data-privacy-06062018.pdf.

Oster AM, Wejnert C, Mena LA, Elmore K, Fisher H, & Heffelfinger JD (2013). Network analysis among HIV-infected young black men who have sex with men demonstrates high connectedness around few venues. Sexually Transmitted Diseases, 40(3), 206–212. 10.1097/OLQ.0b013e3182840373. [PubMed: 23403601]

Parker CM, Garcia J, Philbin MM, Wilson PA, Parker RG, & Hirsch JS (2017). Social risk, stigma and space: key concepts for understanding HIV vulnerability among black men who have sex with men in New York City. Culture, Health & Sexuality, 19(3), 323–337. 10.1080/13691058.2016.1216604.

Ragonnet-Cronin M, Benbow N, Hayford C, Poortinga K, Ma F, Forgione LA, Sheng Z, Hu YW, Torian LV, Wertheim JO, 2021. Sorting by race/ethnicity across HIV genetic transmission networks in three major metropolitan areas in the United States. AIDS Research and Human Retroviruses. 10.1089/AID.2020.0145.

Rose R, Hall M, Redd AD, Lamers S, Barbier AE, Porcella SF, Hudelson SE, Piwowar-Manning E, McCauley M, Gamble T, 2019. Phylogenetic methods inconsistently predict the direction of HIV transmission among heterosexual pairs in the HPTN 052 cohort. The Journal of Infectious Diseases, 220, 1406–1413. 10.1093/infdis/jiy734. [PubMed: 30590741]

Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, Meehan MO, Lutalo T, & Gray RH (2000). Viral load and heterosexual transmission of Human Immunodeficiency Virus Type 1. New England Journal of Medicine, 342(13), 921–929. 10.1056/NEJM200003303421303.

Sacks-Davis R, Daraganova D, Aitken C, Higgs P, Tracy L, Bowden S, Jenkinson R, Rolls D, Pattison P, Robins G, Grebely J, Barry A, et al. (2012). Hepatitis C virus phylogenetic clustering is associated with the social-injecting network in a cohort of people who inject drugs. PLoS ONE 7, e47335. 10.1371/journal.pone.0047335. [PubMed: 23110068]

Schneider JA, Cornwell B, Jonas A, Lancki N, Behler R, Skaathun B, Young L, Morgan E, Michaels S, Duvoisin R, Khanna AS, Friedman S, et al. (2017). Network dynamics of HIV risk and prevention in a population-based cohort of young Black men who have sex with men. Network Science, 5(3), 1–29. 10.1017/nws.2016.27.

Smith DM, May S, Tweeten S, Drumright L, Pacold ME, Pond SLK, Pesano RL, Lie YS, Richman DD, Frost SD, Woelk CH, & Little SJ (2009). A public health model for the molecular surveillance of HIV transmission in San Diego, California. AIDS 23(2), 225–232. 10.1097/QAD.0b013e32831d2a81. [PubMed: 19098493]

Snijders TAB, van de Bunt GG, & Steglich CEG (2010). Introduction to stochastic actor-based models for network dynamics. Social Networks, 32(1), 44–60. 10.1016/j.socnet.2009.02.004.

Stamatakis A (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), 1312–1313. 10.1093/bioinformatics/btu033. [PubMed: 24451623]

Trefethen LN & Bau III D (1997). Numerical Linear Algebra. SIAM: Society for Industrial and Applied Mathematics.

Valente TW (2012). Network interventions. Science, 337(6090), 49–53. 10.1126/science.1217330. [PubMed: 22767921]

Wertheim JO, Pond SLK, Forgione LA, Mehta SR, Murrell B, Shah S, Smith DM, Scheffler K, & Torian LV (2017). Social and genetic networks of HIV-1 transmission in New York City. PLoS Pathogens, 13(1), e1006000. 10.1371/journal.ppat.1006000. [PubMed: 28068413]

Xiang Y, Fujimoto K, Li F, Wang Q, Del Vecchio N, Schneider J, Zhi D, & Tao C (2021). Identifying influential neighbors in social networks and venue affiliations among young MSM: A data science approach to predict HIV infection. AIDS, 35(Suppl 1), S65–S73. 10.1097/QAD.0000000000002784. [PubMed: 33306549]

Xiang Y, Fujimoto K, Schneider J, Jia Y, Zhi D, Tao C (2019). Network context matters: graph convolutional network model over social networks improves the detection of unknown HIV infections among young men who have sex with men. Journal of the American Medical Informatics Association, 26(11), 1263–1271. 10.1093/jamia/ocz070. [PubMed: 31197365]

Yokoyama S & Gojobori T (1987). Molecular evolution and phylogeny of the human AIDS viruses LAV, HTLV-III, and ARV. Journal of Molecular Evolution, 24(4), 330–336. 10.1007/BF02134131. [PubMed: 3110425]

Young LE, Michaels S, Jonas A, Khanna AS, Skaathun B, Morgan E, Schneider JA, Team, & uConnect Study Team. (2017). Sex behaviors as social cues motivating social venue patronage among young black men who have sex with men. AIDS and Behavior, 21(10), 2924–2934. 10.1007/s10461-017-1679-8. [PubMed: 28097618]

Zhang Y, Wymant C, Laeyendecker O, Grabowski MK, Hall M, Hudelson S, … Hosseinipour MC (2021). Evaluation of Phylogenetic Methods for Inferring the Direction of Human Immunodeficiency Virus (HIV) Transmission: HIV Prevention Trials Network (HPTN) 052. Clinical Infectious Diseases, 72(1), 30–37. 10.1093/cid/ciz1247. [PubMed: 31922537]

Fujimoto et al. Page 22

Highlights

- Affiliation networks are integrated with molecular data to examine HIV transmission.

- Core-peripheral structure in venue affiliation is associated with HIV transmission.

- Popular social venues potentiate sexual affiliation network of minority young MSM.

- Molecular and affiliation network approach informs effective network interventions.

- A surveillance-research-community partnership is important to address ethical issues.

Soc Networks. Author manuscript; available in PMC 2023 January 01.

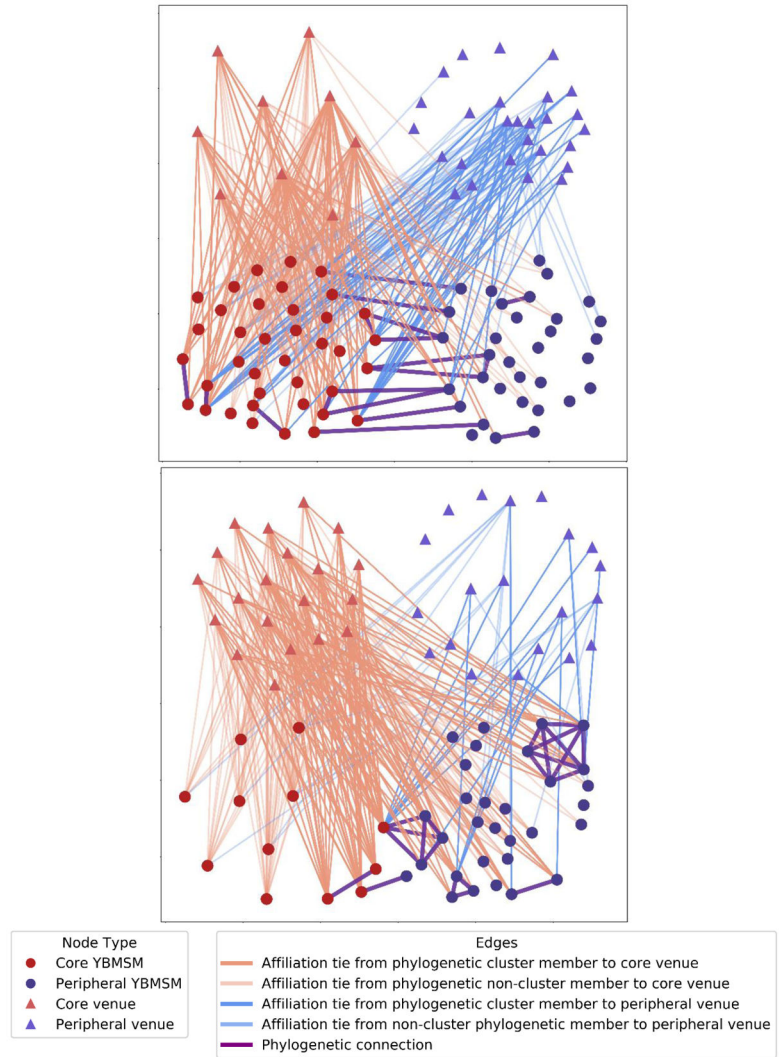Houston   Chicago

Figure 1. Affiliation networks between YBMSM and venues by core versus peripheral subgroups of YBMSM and venues in relation to phylogenetic connections.

Note: Affiliation ties that connect members of phylogenetic clusters are indicated by a bold line.

Table 1.

Core-peripheral partitions/subgroups of YBMSM (row) and venues (column) in a two-mode master matrix, **P**.

		Venues		
		Core	**Periphery**	
YBMSM	Core	A	B	A + B
	Periphery	C	D	C + D
		A + C	B + D	

Table 2.

Actor-by-actor projected valued matrix based on a two-mode block matrix, **P**

		Venues		
		Core	**Periphery**	
YBMSM	Core	AA'	BB'	(A + B) (A + B)'
	Periphery	CC'	DD'	(C + D) (C + D)'
		(A + C) (A + C)'	(B + D) (B + D)'	

Table 3.

Descriptive statistics: frequency (%) and mean (SD; mix, max) of iMAN sample, Houston (*N* = 71) and Chicago (*N* = 43), 2014–2016.

Variable	Total (*N* = 114)	Houston (*N* = 71)	Chicago (*N* = 43)
Age (in years)	24.2 (2.8; 19.3, 29.9)	24.4 (2.5; 19.3, 29.9)	24.0 (2.8; 19.3, 29.9)
Sexual identity			
Gay-identified	96 (85.7%)	60 (85.7%)	36 (85.7%)
Bisexual	12 (10.7%)	7 (10.0%)	5 (11.9%)
Straight	1 (0.9%)	1 (1.4%)	0 (0%)
Other	3 (2.7%)	2 (2.9%)	1 (2.4%)
Education			
High school or less	55 (49.1%)	35 (50.0%)	22 52.4%)
More than high school	57 (50.9%)	35 (50.0%)	20 (47.6%)
Past 12-month housing instability	33 (29.5%)	18 (25.7%)	15 (35.7%)
Number of sex partners during past 6 months	4.4 (7.3; 0, 60)	4.4 (7.6; 0, 60)	4.2 (6.8; 0, 40)
Viral load level			
<400, suppressed	4 (3.9%)	4 (6.2%)	0 (0.0%)
400–9,999	25 (24.3%)	17 (26.7%)	8 (21.1%)
10,000–49,999	49 (47.6%)	31 (47.7%)	18 (47.4%)
50,000+	25 (24.3%)	13 (20.0%)	12 (31.6%)
Number of social venues attended	5.3 (4.7; 0, 24)	4.4 (4.1; 0, 20)	6.6 (5.4, 0, 24)

Note: Missing values include two for sexual identity, education, housing instability, number of sex partners, and 11 for viral load level. Percentage computation excludes missing values.

Table 4.

Descriptive statistics: average coreness scores (SD; min, max) and concentration scores (correlation) from continuous core-periphery model of iMAN sample, Houston ($N = 71$) and Chicago ($N = 43$), 2014–2016.

	Houston		Chicago	
Variable	**YBMSM (VV')**	**Venues (V'V)**	**YBMSM (VV')**	**Venues (V'V)**
Sample size	71	36	43	40
Average coreness score	0.10 (0.07; 0, 0.27)	0.11 (0.13; 0, 0.50)	0.12 (0.09; 0, 0.39)	0.12 (0.10; 0, 0.32)
Core size	38	9	12	20
Core density	0.27	0.35	0.42	0.33
Concentration score	0.84	0.86	0.83	0.88
Periphery size	33	27	31	20
Periphery density	0.02	0.01	0.05	0.01