Published in final edited form as:

Soc Networks. 2020 October ; 63: 201–212. doi:10.1016/j.socnet.2020.06.003.

The potential of online social networking data for augmenting the study of high-risk personal networks among young men who have sex with men at-risk for HIV

Lindsay E. Young^{1,2}, Arthi Ramachandran^{1,2,4}, L. Phillip Schumm³, Aditya S. Khanna^{1,2}, John A. Schneider^{1,2,3}

¹Chicago Center for HIV Elimination, University of Chicago, Chicago, IL

²Department of Medicine, University of Chicago, Chicago, IL

³Department of Public Health Sciences, University of Chicago, Chicago, IL

⁴Center for Data Science and Public Policy, University of Chicago, Chicago, IL

Abstract

Capturing dynamics in high-risk personal networks is essential for preventing HIV transmission. Online social networking data offer incentive to augment traditional self-reported approaches for network enumeration. To explore what online networks reveal about dynamics among high-risk associates, we examine the relationship between egocentric confidant and sex networks and personal Facebook friendship networks of a cohort of young Black men who have sex with men. Although overlap exists between self-reported and Facebook associates, the stabilities of each were unrelated. Confidants who were also Facebook friends with a respondent were, however, more likely to be retained. Thus, Facebook networks contain stable confidants.

Keywords

network dynamics; personal networks; online social networks; HIV; men who have sex with men

1. Introduction

Despite meaningful strides toward eliminating HIV in the United States, HIV incidence remains disproportionately concentrated at the intersection of sexual and racial minority communities. In 2017, men who have sex with men (MSM) accounted for an estimated 67% (25,748) of new HIV infections. Among MSM with newly diagnosed HIV, non-Hispanic Black MSM (BMSM) accounted for the plurality (38%), followed by Hispanics/Latinos (28%) and non-Hispanic Whites (27%) (Centers for Disease Control and Prevention, 2018).

Corresponding Author: Lindsay Young, University of Chicago, Departments of Medicine, Section of Infectious Disease, 5837 S. Maryland Ave., Chicago IL 60637. Mobile: (610) 331-1930; lyoungl@medicine.bsd.uchicago.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Further, young BMSM (YBMSM) bear a particularly heavy burden, with YBMSM between the ages of 15–29 accounting for 61% of new HIV infections among all BMSM (Centers for Disease Control and Prevention, 2018). Reasons for these race-and age-based disparities are varied and unclear, especially as YBMSM, compared to their White MSM counterparts, tend to demonstrate lower levels of many individual-level behaviors generally assumed to be related to HIV risk, such as number of sex partners, engagement in condomless sex, and frequency of HIV testing (Millet, Flores, Peterson, & Bakeman, 2007).

Inadequacies in individual behavioral explanations alone have given rise to a socio-environmental perspective on HIV risk (Rhodes, Singer, Bourgois, Friedman, & Strathdee, 2005). A socio-environmental perspective accounts for properties of social and sexual networks that either increase or decrease an individual's HIV vulnerability (Millet et al., 2012), for example by focusing on features of sexual partnerships that introduce greater risk like partner concurrency (Morris, Goodreau, & Moody, 2007) and race-and age-based mixing (Adimora & Schoenbach, 2005; Joseph et al., 2011) or the compositional features of social support networks that engender critical protections (Schneider, Michaels, & Bouris, 2012).

To date, most research on the network conditions associated with HIV risk has been limited to network features such as structure (e.g., size, density) or composition (e.g., assortativity) measured at only one time point. However, there is growing awareness that *changes* in networks and their dynamic properties have implications for a variety of health outcomes (Cornwell & Laumann, 2015; Costenbader, Astone, & Latkin, 2006; Schneider et al., 2017). This is particularly true for marginalized social groups susceptible to HIV like YBMSM or injection drug users for whom network turnover can lead to greater HIV risk engagement (Costenbader et al., 2006) as well as larger structural vulnerabilities known to exacerbate HIV risk, such as criminal justice involvement, exposure to violence, and housing instability (Desmond, 2012; Fischer & Beresford, 2014; Stack, 1975).

Measuring changes in the personal sexual and social networks of YBMSM presents practical challenges, as it involves eliciting network contacts, at multiple time points, from members of a population that is generally “hard to reach” (Heckathorn, 1997) in the first place. Further, eliciting contacts through traditional name generators — i.e., the instruments used to identify people (or alters) who are in a focal individual's (or ego's) personal network (Marsden, 2011) — can result in significant under-reporting of relevant contacts due to a variety of response biases like forgetting (Brewer, Garrett, & Kulasingam, 1999), respondent fatigue (Krosnick, 1991), satisficing (Dillman, Eltinge, Groves, & Little, 2002), and information sensitivity (e.g., reluctance to identify individuals as MSM who do not publicly disclose themselves as such). Consequently, the name generator tends to capture close confidants (or strong ties), which are only a small subset of personal network members (Hsieh, 2015; Marsden, 1987).

Alternatively, widespread use of online social networking platforms (e.g., Facebook, Instagram, and Snapchat) among young sexual and racial minorities (Greenwood, Perrin, & Duggan, 2016; Taylor, 2013) has created new sources of near real-time network data with immense potential for providing insights about critical shifts in the social and sexual lives of

YBMSM. This optimism, however, is rooted in an assumption that online social networks like Facebook reflect the character and dynamics of the high-risk personal networks that HIV researchers and front-line staff seek to understand and affect (Boyd & Ellison, 2007; Traud, Mucha, & Porter, 2012).

In this paper, we explore this assumption by addressing the following research question: *“What can digitally-archived online social networking data tell us about the self-reported “offline” networks of YBMSM at-risk for HIV?”* From prior work (Young, Fujimoto, Alon, Zhang, & Schneider, 2019) we know that there are parallels between the “offline” and online social lives of YBMSM members of the house/ballroom community (HBC) — an elective system of queer kinship among African American LGBTQ people that feature anchoring family-like structures called houses and the flamboyant competitive balls that they produce (Arnold & Bailey, 2009). This work demonstrated that members of the HBC use networking platforms like Facebook groups to supplement their communication with house members, to organize events, and to promote aspects of ballroom culture like voguing and competition etiquette. Outside the domain of YBMSM, research on adolescent friendships lends further credence to the notion that offline and online social domains are expected to overlap to some degree (Reich, Subrahmanyam, & Espinoza, 2012).

At the same time, there is also reason to believe that online personal networks will yield unique ties not otherwise represented in self-reported personal networks, as van Woudenberg et al. (2020) found in their comparison of network elicitation approaches. Among a sample of students, they observed a greater number of online ties than they observed in their peer nomination network, suggesting that online networks are capable of capturing additional physical world relationships that would otherwise go unobserved if relying solely on peer nominations. Furthermore, online networks will also inevitably capture relationships that do not exist in the physical world, for example when someone “friends” a co-member of a Facebook group or a contact of one of their Facebook friends, neither of which they know offline.

Ascertaining the degree of overlap between offline and online networks of YBMSM will reveal how much these social contexts intersect. An additional, more advanced step is to verify if there are associations between YBMSMs’ offline network dynamics, for example stability or turnover among their confidants or sex partners, and what can be observed in their Facebook relationships over time. At the macro level, research has shown that the sizes of both offline and online personal networks are often bound by a similar constraint — the cognitive and time-based demands of maintaining relationships (Dunbar, Arnaboldi, Conti, & Passarella, 2015). This leads to a persistent core of network associates who are maintained over time and subsequent layers of increasingly peripheral actors who are more regularly replaced (Arnaboldi, Passarella, Conti, & Dunbar, 2015). Despite what is known about macro-level structural similarities of offline and online personal networks and how those similarities might persist over time, it is unclear whether the churn (or stability) that YBMSM experience in their offline friendships and sex partnerships corresponds to similar degrees of change in their online contacts.

To fill these gaps, we draw from Facebook and self-reported confidant and sexual egocentric network data collected from a cohort of YBMSM to: (1) investigate the bi-directional relationship between macro-level confidant and sexual network stability and Facebook use and turnover; (2) measure the degree of overlap between Facebook network contacts and self-reported confidant and sex partner alters; and, with the results of that matching process (3) determine the association between tie-level Facebook friendship status and the likelihood of retaining a confidant from the previous time point. Findings from these analyses lend valuable insights into the potential of using digitally archived online social networking data to supplement or complement current self-reported network generation techniques with the goals of being able to enumerate a wider range of high-risk network associates and more rapidly and efficiently identify moments of change in high-risk personal networks that can affect HIV vulnerability (Schneider et al., 2017).

2. Methods

2.1 Study Design and Sampling

All data collection and study procedures received ethics approval from the Institutional Review Boards of the University of Chicago Biological Science Division and the National Opinion Research Center at the University of Chicago. This work was supported by grants from the National Institutes of Drug Abuse (NIDA) and Mental Health (NIMH). The authors have no competing interests to report.

Data used in this study were collected in 2013–2016 as part of uConnect, a three-wave longitudinal cohort study of YBMSM living in Chicago. Respondents were recruited using Respondent Driven Sampling (RDS) (Heckathorn, 1997). Widely used in public health studies (Goel & Salganik, 2010), RDS provides a mechanism for sampling hard to reach populations (e.g., intravenous drug-users, sex workers, men who have sex with men) together with methods for making statistical inferences about the target population.

To begin, a group of 62 individuals were purposively selected as RDS “seeds,” chosen to have large social networks and ties to a diverse array of people belonging to different subpopulations (Johnston & Sabin, 2010). We recruited seeds from heterogeneous social spaces that YBMSM occupy, including LGBTQ social venues, online networking sites, community-based organizations, and HIV treatment and prevention programs. Potential respondents (including seeds) were deemed eligible if they: 1) self-identified as African American or Black, 2) were assigned male at birth, 3) were 16–29 years of age (inclusive), 4) reported having oral or anal sex with a male partner within the past 24 months, and 5) were willing and able to provide informed consent. Each respondent was given up to six vouchers to recruit others who met the same eligibility criteria. Respondents received \$60 for their participation and \$20 for each recruit who enrolled into the study. This procedure yielded a baseline sample of 618 YBMSM, 525 of which were retained at Wave 2 and 507 at Wave 3. At the time of study enrollment, only one previous study of Black MSM in Chicago (N=212) had been conducted by the investigators. uConnect participants were not directly recruited from this previous study nor were the RDS seeds.

2.2 Data Collection

2.2.1. The Survey Instrument—Interviewer administered survey assessments were conducted at each of three study waves with 9-month intervals between them. A central component of these longitudinal survey assessments was the enumeration and description of the social and sexual networks of YBMSM over time. To this end, we used two separate network generators, a confidant generator and a sex partner generator described in section 2.2.2. We made use of Computer Assisted Personal Interviewing (CAPI) to improve the quality of the network data collected, in particular, to elicit explicit confirmation from respondents of matches between “alters” listed in the separate network sections of the questionnaire and in subsequent waves of the study.

In addition to enumerating participants’ confidant and sexual networks, survey assessments also included modules on a variety of factors known to be associated with HIV vulnerabilities including healthcare engagement, substance use and sexual risk behaviors, structural vulnerabilities (e.g., criminal justice involvement) and other aspects of their social lives including their patronage of LGBTQ social venues (e.g., bars/clubs, gyms, public spaces) and affiliations with LGBTQ cultural groups (e.g., ballroom houses).

2.2.2 Generating Self-reported Personal Confidant and Sex Networks—At Wave 1, the confidant network name generator elicited up to five confidants using the prompt, *“Please list the names of the people with whom you discuss things that are important to you.”* The sex partner network generator first elicited information on up to five most recent sex partners in the past six months, which was followed by an additional prompt about any current primary sex partner not already named. As such, participants were permitted to name a total of six partners (Laumann, Gagnon, Michael, & Michaels, 1996). To capture overlap between the respondent’s confidant and sex partner networks, the respondent was asked to compare the list of confidants mentioned earlier with the list of sex partners that they named. The two lists were displayed side by side on the computer screen and the respondent indicated which names represent the same person. Any matches were recorded.

At Wave 2, the same confidant generator was used as in Wave 1. After Wave 2 confidants were named, the respondent was asked if any of those confidants were also named as a confidant in Wave 1. The list of Wave 1 confidants was displayed next to the newly generated Wave 2 confidants and the respondent indicated matches where applicable. The Wave 2 sex partner generator was similar to that used in Wave 1, except that the prompted time frame was since the last interview instead of the past six months. Again, up to six sex partners could be listed, including a current primary partner if not already named. Respondents were then asked to compare their Wave 2 partners with partners named at Wave 1 in the same manner described above. They were then asked to compare the Wave 2 sex partners with the Wave 2 confidants to confirm any matches across social contexts. It should be noted that in Wave 2, respondents were *not* asked to compare sex partners in Wave 2 with confidants from Wave 1 or vice versa. For the purposes of this analysis, the authors made these comparisons and matches post data collection using names and basic demographic information provided about each alter.

A similar procedure was used in Wave 3, except that the confirmation list was cumulative. For example, respondents were asked whether any of the confidants named in Wave 3 corresponded to a combined list of confidants and sex partners from the current wave and from Wave 2. Similarly, respondents were asked whether the sex partners named at Wave 3 corresponded to a cumulative list of the confidants and sex partners from the current Wave and from Wave 2. Figure 3 visually depicts the within-and between-wave comparison process. Finally, at each wave, name interpreters were asked regarding the first five confidants (e.g., gender, sexual orientation, relationship type, relational closeness) and first five sex network members (e.g., gender, partner type).

2.2.3 Generating Personal Facebook Friendship Networks—Like their self-reported counterparts, digitally archived (online) personal networks are defined as the immediate ties that exist around a focal individual (or ego), much like an individual's Facebook friendship network. Facebook data, including friendships, group memberships, and messaging ties were obtained from consenting respondents at Waves 1 and 2, using a software application that accessed Facebook's application programming interface (API). Since these data were collected, Facebook made changes to its API that have made this method of data collection impossible. However, Facebook data can still be collected using the manual data download feature provided by Facebook, which allows any user to download all or specific portions of their archived data. For the purposes of this analysis, we operationalize a Facebook relationship between i and j as an unweighted friendship connection.

2.3. Statistical Analysis of Network Stability

A Generalized Structural Equation Model (GSEM) was fit to the likelihood of retaining alters from one wave to the next, as well as the likelihood that new alters were named to fill empty slots in the roster, as described in previous work (Schneider et al., 2017). Briefly, this model assumes a single, underlying (latent) variable reflecting a respondent's network stability over the course of the study, which may affect both the likelihood that previous alters are retained and the likelihood that new alters are named to fill the empty slots (both modeled using logistic regression). To identify the model, this latent variable is constrained to have mean 0 and variance 1. Additional covariates include the order in which previous alter(s) were named, as well as both attributes of the respondent (number of alters named at the last timepoint, total size of MSM network) and of his relationship with the alter (relationship type). As in our previous analysis, age of respondent was included as a covariate predicting network stability. This model was fit separately to the Wave 1–3 data for confidants and for sex partners, using maximum likelihood with mean-variance adaptive Gauss-Hermite quadrature and the clustered, robust variance estimator (clustered at the level of the respondent, thereby accounting for potential correlation between alters named at a single wave and between the two time intervals within a respondent) (Rabe-Hesketh & Skrondal, 2004). Empirical Bayes (Rabe-Hesketh & Skrondal, 2004) estimates of confidant and sex partner stability, adjusting for differences due to age, were calculated.

Estimates of confidant and sex partner stability were regressed separately on the following features of respondents' personal Facebook networks at Wave 1: number of Facebook

friends, number of friends with whom the respondent had exchanged personal messages, total number of messages exchanged, and the median and maximum (across a respondent's Facebook friends) proportion of common friends. Each of the covariates was transformed using square roots (to reduce skew) and then standardized (to facilitate comparison of relative magnitude across coefficients). Separate models, each with a single covariate, were fit for each feature.

To examine the correlates of turnover in respondents' Facebook networks from Wave 1 to Wave 2, we fit separate Generalized Linear Models (GLMs) (McCullagh & Nelder, 1989) to the proportion of Facebook ties that were lost (i.e., "unfriended", using logistic regression) and the number of new Facebook ties that were added (using a log link with gamma variance function). Covariates included estimates of confidant and sex partner stability (as described above), the number of Facebook friends at Wave 1, age (in years) and an indicator of whether the respondent was a member of a house/ball community. Robust (i.e., sandwich) variance estimates are reported (Royall, 1986). All analyses were performed using Stata version 15.1 (StataCorp, 2017).

2.4 Entity resolution of Self-reported and Facebook Alters

Our approach for resolving a respondent's self-reported sex and confidant alters to their Facebook alters (or friends) occurred in three stages. In the first stage, we attempt to match each named sex and confidant alter with a likely Facebook name(s) from the pooled list of study participants' Facebook friends. Matches were based on similarity between an alter's name as provided by a respondent and Facebook names. In the survey data, alter names need not be standardized. Some respondents provide the full name, others just the first, still others provide a designation (e.g. 'mom'). Facebook names also need not be standardized with individuals using first names only, combinations with their middle and last name, or nicknames as their profile name. To account for this, we search for matches in our roster of all Facebook profile names based on various versions of each Facebook alter name and then shortlist all those profiles that match a provided sex or confidant alter name. We search for matches in our larger pool of Facebook profile names, rather than in just the respondent's Facebook friends in order to capture named sex or confidant alters who might not be in their Facebook network.

The Facebook name versions we use, in order of increasing specificity, are: (1) first name only, (2) first name and last name, (3) first name, middle initial, and last name, and (4) first name, middle name and last name. For example, for an alter whose name is "John A. Smith", we shortlist Facebook profiles with the names "John", "John Smith", and "John A. Smith". To account for names that are spelled differently (e.g. John vs. Jon), we also search for Facebook profiles with the same pronunciation as the alter's name using metaphones (Philips, 1990) — i.e., phonetic algorithms developed for indexing words by their English pronunciation.

Then, for each potential match between a self-reported alter and a Facebook alter, a series of scores is computed based on various profile characteristics, including: (1) a name score equal to the normalized levenshtein distance (Levenshtein, 1966) — i.e., the smallest number of insertions, deletions, and substitutions required to change one string into another

— of the reported name and Facebook name, (2) an age score equal to the difference between the ages of the self-reported alter and the Facebook alter, (3) a binary sex score defined as whether or not the sex of a self-reported alter and the Facebook alter match; and (4) a binary city score defined as whether or not the city of a self-reported alter and the Facebook alter match. For each named alter, this process yields a list of self-reported-Facebook alter matches, along with the scores for each of the characteristics.

The second stage of the resolution process ranks these potential matches by quality. In stage 1, a self-reported alter can have more than one possible Facebook username that matches on name alone. For this reason, we categorize each match based on quality. We discard low quality matches: those for whom the listed names are too dissimilar (normalized levenshtein distance > 0.5) or matches where neither the age nor the sex is consistent between the self-reported and Facebook alter. Of the potential Facebook matches, we take only the best one for each alter (for ranking of criteria, see Table 4).

Finally, once a probable match has been determined between a named alter and a Facebook profile, we then determine whether the matched profile is indeed a Facebook friend with the respondent who named him or her. This allows us to distinguish between named alters who are somewhere in the network of Facebook friends among study participants and those who are directly connected to the respondent who named them. Although both circumstances are important to consider when thinking broadly about the relationship between the offline and online networks of a given population, the later circumstance is more germane when determining how personal network stability is related to having multiplex ties that span offline and online domains. Python code for these entity resolution routines will be made available upon request to the corresponding author.

2.5 Statistical analysis of Characteristics of Resolved Entities

Mixed-effects logistic regression models (McCulloch & Neuhaus, 2001) were used to model the probability that an alter was matched to a Facebook profile and, conditional on being matched, was linked to the respondent (i.e., was a Facebook friend). Separate models were fit to confidants and sex partners identified in Waves 1 and 2. Covariates included attributes of the respondent (age, education, sexual orientation, membership in the house/ball community, and an indicator of having spent time in jail), attributes of the alter (gender, MSM), characteristics of the respondent/alter relationship (relationship/partner type, frequency of interaction, and closeness, all as reported by the respondent), and the order in which the alter was named in the name generator process. This analysis allowed us to examine the resolution process itself to identify individual and dyadic characteristics associated with better resolution and potential bias.

A respondent-level random effect was included to account for remaining differences between respondents. The model was fit using maximum likelihood with mean–variance adaptive Gauss–Hermite quadrature (Rabe-Hesketh & Skrondal, 2004). Standard errors were obtained using the clustered form of the robust (i.e., sandwich) variance estimator (Rogers, 1994), thereby ensuring valid inference even if the true covariance structure of the data differs from that specified by the model.

2.6 Statistical Analysis of Confidant and Sex Partner Retention as Associated with Facebook Ties

To the GSEM model of confidant stability described in Section 2.3, we added indicators of whether the previous alter was found in Facebook, and if so, whether he or she was tied to (i.e., a Facebook friend of) the respondent (the latter was set to 0 for those alters not found in Facebook).

3. Results

3.1 Characteristics of YBMSM

Demographic and network stability characteristics of the full YBMSM sample at Wave 1 (N=618) are shown in Table 1. Since our analytic sample is limited to those who provided Facebook data, we summarize each characteristic separately for those who provided Facebook data and those who did not to show characteristics for which the analytic sample differs from those who were excluded. Among the analytic subsample, 90% completed high school, 72% were employed, 30% were members of the house/ballroom community, 51% had never spent time in jail, and 73% had never been homeless in the past year. Those in the analytic subsample were less likely to identify as straight (1% versus 6%), more likely to be HIV positive (40% versus 30%), and slightly older on average (23.1 versus 22.5 years old). No differences were observed in either confidant or partner stability between the analytic subsample and those excluded were observed

3.2 Characteristics of Confidant, Sex, and Facebook Personal Networks

Characteristics of confidant, sex partner, and Facebook personal networks at each wave are shown in Table 2. In total, 599 respondents (of 618) named at least one confidant at Wave 1, 508 respondents (of 525) named at least one confidant at Wave 2, and 488 respondents (of 507) named at least one confidant at Wave 3. At Wave 1 respondents named on average 2.61 confidants (SD=1.17), at Wave 2 they named 2.58 confidants (SD=1.09), and at Wave 3 they named 2.47 confidants (SD=1.07). Efforts to match confidant alters across all respondents yielded 1,414 unique alters named at Wave 1, 1,097 unique alters named at Wave 2, and 1,206 unique alters named at Wave 3.

Regarding sex partners, 599 respondents (of 618) named at least one partner at Wave 1, 498 respondents (of 525) named at least one partner at Wave 2, and 475 respondents (of 507) named at least one partner at Wave 3. Respondents named on average 2.64 partners (SD=1.55) at Wave 1, 2.42 partners (SD=1.41) at Wave 2, and 2.29 partners (SD=1.39) at Wave 3. Matching sex partner alters across all respondents yielded 1,590 unique partners at Wave 1, 1,355 unique partners at Wave 2, and 1,242 unique partners at Wave 3. Moreover, 12% of the confidants named by respondents were also named as sex partners, while 20% of the sex partners named by respondents were also named as confidants. This corresponds to an average of about 1 alter per respondent who was named as both a confidant and sex partner at each wave. Figure 2 shows the distributions of the number of confidants and sex partners named at each wave, both of which remained similar across waves. The distribution of named sex partners shows a bimodal pattern at each wave, with a substantial number of

respondents (about 20%) naming 5 sex partners, which is about 2 times greater than the average.

Finally, personal Facebook network data were successfully collected from a total of 298 respondents at Wave 1 and a total of 391 respondents at Wave 2 (283 respondents provided data at both Waves). At Wave 1, respondents who provided Facebook data had on average 1,139 Facebook friends (Min=14, Max=4,723; SD=936). At Wave 2, respondents who provided Facebook data had on average 1,376 Facebook friends (Min=1, Max=5,860; SD=1,361). Matching Facebook friends across respondents yielded 182,987 and 253,519 unique Facebook friends across all respondents at Wave 1 and Wave 2, respectively.

3.2 Confidant and Sex Network Stability and Facebook Use

Both confidant network stability and sex partner stability were negatively associated with the respondent's number of Facebook friends at Wave 1, though the confidence intervals were fairly wide and narrowly included zero (Table 3). A one SD increase in the square root of the number of Facebook friends was associated with a decrease in confidant stability of -0.11 SDs (95% CI $-0.22, 0.01$) and a decrease in sex partner stability of -0.13 SDs (95% CI $-0.27, 0.01$). There was no evidence of an association between either confidant or sex partner network stability and any of the other measures of Facebook usage metrics.

3.3 Facebook turnover from Wave 1 to Wave 2

Table 4 shows results from generalized linear regression models fit to the turnover in Facebook ties from Waves 1 to 2 among the 201 respondents with usable Facebook data at both time points. The likelihood of a tie being lost (corresponding to the respondent "unfriending" the other party) decreased as the number of Facebook friends increased (OR 0.71, 95% CI 0.63, 0.79), and was also lower among older respondents (OR 0.96, 95% CI 0.93, 1.00), although this association was slightly above the $p < 0.05$ benchmark for statistical significance. At the same time, those with more Facebook friends at Wave 1 added more ties at Wave 2 (on average), while older respondents added fewer (-8% per year of age, 95% CI $-1\%, -14\%$). Neither the likelihood of losing Facebook ties nor the number of new ties added was associated with network stability in the self-reported networks (confidant or sex partner). Finally, respondents who were members of a house/ball community were both more likely to lose ties (OR 2.02, 95% CI 1.61, 2.53) and to add new ties ($+130\%$, 95% CI $+51\%, +249\%$).

3.4 Entity Resolution Results

Across all three Waves of survey data collection, respondents named a total of 5,836 unique confidants and sex partners. Of these, 658 were named by respondents who did not provide Facebook data, leaving 5,178 alters who could be potentially matched on Facebook. Several alters were identified merely by relationship type (e.g., "friend"), by an anonymous designation (e.g., "partner 1"), or by initials (e.g., "AJ"). These were in nearly all cases impossible to match and constituted 431 of the 5,178 remaining alters. An additional 1,948 were unlikely to be matched given the limited amount of identifying information provided (e.g., only first name was given). Thus, a priori we expected at most approximately 2,700 alters to be matched (2,799 were identified with complete names and the respondent gave a

Facebook profile). However, we did not exclude any alters in our resolution analysis in order to capture instances where, for example, the alter matched a profile who wasn't in the respondent's friend network or the first name alone was unique and easy to match.

Conditioned on being found in the pooled list of Facebook friends, 780 alters were successfully matched to Facebook profiles in the friend list of the respondent who named them, including 376 sex partners and 458 confidants (54 alters were both sex partners and confidants). Table 5 shows matches categorized by the quality of the match. When there were multiple potential matches, we used the ranking previously described to select the best among them. Note that a profile is considered to have matched with the sex of the named alter if the profile had the sex specified and it was the same as that of the alter (similarly for age and location). This set of selection criteria has an implicit ranking of first prioritizing similar names, then age, sex and location.

At the respondent level, about 5% ($n=21$) of the respondents who provided Facebook data at either Wave 1 or Wave 2 had no alters match to a Facebook profile anywhere in the pooled list of respondents' friends. When including only matches to Facebook profiles connected to the respondent, 20% ($n=84$) of the respondents had none of their confidant or sex partners match a Facebook profile. Sex partners were less likely to be matched than confidants; 185 respondents had none of their sex partners matched and 156 respondents had none of their confidants matched.

Figure 3 (top) shows the distribution of the number of matches separately for the unique confidant and sex partner alters named across all three waves. A majority of the respondents (81%) had at least one of their named confidants matched to a Facebook profile. A similar percentage (84%) of the respondents had at least one of their named sex partners matched to a Facebook profile. Confidants were more likely to be matched (47%) than sex partners (37%). Figure 3 (middle) shows the fraction of each participant's named alters that were matched to a profile somewhere in the pooled list of respondents' Facebook friends. The distributions for both confidants and sex partners show that, although some participants had no alters matched to someone in the pooled list of Facebook friends, a majority of the participants had somewhere between 0.10 to 0.50 of their named alters matched to a profile in the pooled list of Facebook friends. Figure 3 (bottom) shows the fraction of those matches who were in the respondent's Facebook network, with obvious modes at 0 (reflecting respondents with no matches), 1 (reflecting respondents for whom all alters were matched to someone in their personal friendship network), and approximately 0.5 (with a range of 0.25 – 0.75) for the remaining majority of the respondents.

Multilevel factors associated with successfully locating self-reported alters in Facebook and, subsequently, linking them to the respondent who named them are shown in Table 6. Regarding confidants, there was little evidence that the likelihood of confidants being found in Facebook was associated with characteristics of the confidant or the relationship with the respondent. However, several respondent-level characteristics including age, education, sexual orientation, and time spent in jail were associated with the likelihood of their confidants being found. Confidants named by older YBMSM were more likely to be found in Facebook, while confidants named by those who completed high school (relative to those

who had not), who identified as Straight (relative to Gay), and who had spent time in jail were less likely to be found. Conditional on being found in Facebook, confidants named by respondents who identify as Straight were also less likely to be directly linked to the them, while those named by respondents who experienced time in jail were more likely to be directly linked to the respondent. Finally, confidants described as friends were more likely to be found and linked to the respondent, although these effects were not statistically significant at the 0.05 level.

In contrast, the likelihood of sex partners being found in Facebook was associated with both partner and relationship characteristics. Less recent partners (i.e., those named further down on the roster), casual partners (relative to main partners), and partners named in the second wave (relative to the first wave) were less likely to be found in Facebook. Age was the only respondent characteristic with a statistically significant association; partners named by older respondents were more likely to be found. None of the factors was associated with being directly linked to a respondent. The estimated variance of the random intercepts in both models (0.40 and 0.46) indicates that after accounting for both alter/tie and respondent characteristics, considerable respondent-level variability remained in the likelihood of sex partners being found, and when found, being linked to the respondent.

3.5 Retention of Self-reported Ties and Facebook Friendship Ties

The model of the likelihood of retaining previous confidants and naming new confidants (when possible) during the next wave (Schneider et al., 2017) was featured in previous work (Schneider et al., 2017) and extended here to include covariates representing characteristics of their Facebook friendship. Results of this extended model are shown in Table 7. Effects were estimated using data from Waves 1–3 for the 298 respondents who provided Facebook friendship data at Wave 1. Estimates are similar to those reported in our previous paper. Confidants that were Facebook friends were also more likely to be retained (OR 1.46; 95% CI 1.03, 2.08), adjusting for the other variables in the model. In contrast, there was no evidence that retention of sex partners was associated with being Facebook friends (not shown).

4. Discussion

Personal networks and their dynamics play a critical role in HIV-related outcomes for young racial and sexual minorities and other vulnerable populations (Costenbader et al., 2006; Schneider et al., 2017). Capturing these networks over time, however, present certain challenges. As young sexual and racial minorities increasingly turn to online social networking environments to find community and interact with peers, an opportunity arises to leverage the near real-time network data to better understand social dynamics in this community and to engage a greater number and, potentially, wider variety of at-risk network contacts in sexual health interventions, like network-based testing programs (Centers for Disease Control and Prevention, 2017).

A critical first step to these ends, is to investigate what online social networks like Facebook can even tell us about the high-risk “offline” personal network dynamics among YBMSM. In this paper, we provide a high-resolution window into the relationship between their self-

reported social and sexual networks and their digitally archived Facebook network. Findings reveal that although overlap exists between self-reported and digitally archived network associates, network stability in each context is essentially unrelated. Thus, the online social networks studied here reflect distinct relationship dynamics and might serve better as a complement to offline network data collection rather than a substitute.

There are several aspects of our findings that deserve further consideration and have implications for how we think about the empirical relationship between online and offline sources of network data. In the first phase of our analysis, we established that the relationship between confidant and sex partner stability (or turnover) and generalized metrics of Facebook use and relational stability are tenuous. First, we learned that macroscopic Facebook metrics like total number of friends, total number of messages, and proportion of friends in common are unrelated to stability in one's offline contacts (Table 2). This suggests that there is little one could infer about the stability of one's high-risk network contacts from course-grained measures of Facebook use and network structure. Likewise, we also learned that latent measures of confidant and sex partner stability tell us little about turnover in Facebook network contacts (Table 3).

Two types of distinctions might shed light on why our analysis revealed such tenuous associations between the dynamics in each domain. First, reasons for tie persistence and turnover may simply differ in offline and online domains, leading to orthogonal patterns of change in each social environment. Studies of offline network dynamics show that ties tend to persist if they are with alters who provide social support, are in frequent communication, or are kin (Morgan, Neal, & Carder, 1997; Roberts & Dunbar, 2015; Wellman, Wong, Tindall, & Nazer, 1997), while tie turnover tends to be more related to life events (e.g., transitioning from school into the labor force, marriage, residential mobility) (Bidart & Lavenu, 2005; Roberts & Dunbar, 2015; Viry, 2012).

In comparison, reasons for tie persistence and turnover in online personal networks are less clear. The relative low stakes in forming and maintaining an online connection with someone intimates that online ties will simply increase over time with little turnover. However, Dunbar and colleagues (2015) have shown that online personal networks experience the same size constraints that affect offline personal networks – i.e., there is a limited number of social relationships that an individual can actively manage. This might compel some individuals to strategically “prune” online contacts by “unfriending” or “unfollowing” those with whom they have had little or no recent interaction, or those for whom the interaction costs are too high (Carpenter & Tong, 2017). So, because offline ties are simultaneously easier to maintain and easier to prune, they may be subject to vastly different and less predictable rhythms of change than offline ties.

Another explanation for the lack of association between YBMSMs' offline and online network dynamics pertains to fundamental differences in YBMSMs' life experiences and how they manage their social spheres compared to young adults more generally. YBMSMs' intersectionality — i.e., having multiple subordinate identities like being a racial and sexual minority (Purdie-Vaughns & Eibach, 2008) — means that they must cope with different types of discrimination in multiple social contexts (e.g., family, church, school, work). These

struggles may trigger them to strategically manage their online networks by creating alternative accounts that allow them to compartmentalize their social circles (Haimson, Brubaker, Dombrowski, & Hayes, 2015). Although our analysis only included Facebook friendships from primary Facebook accounts, many respondents self-reported having multiple alternative Facebook profiles. Thus, a stronger relationship between offline and online personal network dynamics may have emerged had we been able to account for Facebook friendships stemming from their alternative accounts.

This leads us to conclude that more research is needed to better understand mechanisms of online network maintenance and turnover among YBMSM, especially as they relate to HIV risk and prevention. For example, we still lack an empirical understanding of the extent to which newly added or purged online contacts reflect critical moments or life experiences. For example, among YBMSM, it would be worth knowing when newly added Facebook friends result from episodic increases in socializing behavior, such as their patronage of gay bars or their use of online dating sites, rather than from banal decisions to “friend” mutual friends of current Facebook friends. Similarly, it would be salient to know when the sudden pruning of online network associates reflects something more than the casual “spring cleaning” of one’s contact list, for example a break up with a sexual partner or a purposive split from a particular social scene. Toward closing this knowledge gap, a study that tracks changes in the personal online networks of YBMSM and uses ecological momentary assessments to ascertain in real-time what motivated each change would provide immense insight.

In the second phase of our analysis, we demonstrated a viable method of resolving identities across network contexts that could be applied to entity resolution challenges in practices like contact tracing. That said, we learned from this process that our ability to locate offline alters in Facebook depends on a combination of factors aside from the accuracy of the identifying information provided about each alter. First, the type of relationship clearly matters. Confidants were more likely to be found on Facebook and matched directly to a respondent than sex partners. Second, matching hinges on characteristics of respondents, alters and the relationships between them. With respect to confidants, older respondents were more likely to have confidants found in Facebook, while this likelihood decreased for respondents who were more educated, identified as Straight, and had spent time in jail. Sex partners, on the other hand, were more likely to be successfully located in Facebook if they were more recent partners, main partners, and named by younger respondents. More research is required to fully grasp the mechanisms that undergird these effects as it is difficult to ascertain whether these characteristics affect a respondent’s privacy concerns and, hence, their willingness to provide accurate identifying information for their confidants or partners or whether these particular respondents tend not to have confidants and partners who are active on Facebook.

The entity resolution analysis also illuminated a practical issue. The inconsistent and sometimes illusive ways with which individuals identify alters underscores the need for a more standardized identification system for enumerating network associates. Social network HIV testing strategies could benefit from using more standardized memory aids like Facebook friendship lists to elicit more names of at-risk associates from clients who have

been recently tested. Relying on the Facebook username as the primary trait on which to base the entity resolution algorithm would reduce much of the ambiguity one faces when pooling various identifiers and demographic details across multiple data sources. Further, by loading Facebook or other SNS profile usernames into a CAPI system, one can still obtain additional information about each contact using traditional interview techniques (Schneider, Schumm, Fraser, Yeldani, & Liao, 2018).

With the results of the matching process, the final stage of the analysis took a more fine-grained approach to assessing the relationship between offline and online relationships. In our model of confidant network stability (Table 7), we learned that confidants who are also Facebook friends with a respondent are more likely to be retained at a subsequent Wave, thus supporting conclusions from other work that friendships spanning social contexts, including contexts in offline and online domains, are more durable (Harper, Hamill, & Gilbert, 2013; Mollenhorst, Volker, & Flap, 2014). The key insight is when a tie between two individuals is more multiplex, there are more shared foci of activity that bring them together, which makes it more difficult for the tie to decay (Burt, 2000; Feld, 1997). That said, the health implications of relational multiplexity remains underexplored. Research has linked multiplex relationships (e.g., a sex and drug partner) to engagement in higher risk drug use and sex behaviors (Felsher & Koku, 2018; Lakon, Ennett, & Norton, 2006) and an increased chance in contracting HIV (Rudolph, Crawford, Latkin, & Lewis, 2017). However, the degree to which relational multiplexity serves a preventive function for those at-risk for HIV demands further attention.

This study has several limitations worth mentioning. The first set of limitations pertains to how offline and online relationships were captured and operationalized. To elicit confidants, we used the “important matters” name generator used in the General Social Survey. Despite widespread use, this and other single name generators have been criticized for failing to address the full definition of social support, including emotional and instrumental aid, and companionship (Marin & Hampton, 2007). Relatedly, a limit was imposed on the number of confidants and partners that could be named. As such, we likely underestimated the size of respondents’ offline networks and, consequently, their degrees of overlap with online contacts.

Regarding Facebook ties, an unweighted friendship tie is not the only way to operationalize a relationship on Facebook. Connections can also be ascertained from Facebook group co-affiliations (Young, Fujimoto, & Schneider, 2019) or direct messaging (Wilson, Sala, Puttaswamy, & Zhao, 2012). Wilson et al. (2012) argue that interactions like direct messaging are a more accurate representation of meaningful peer relationships than online friendship status. Thus, had we operationalized a Facebook relationship in this way, perhaps we would have detected stronger relationships with offline tie dynamics.

Another set of limitations relates to the entity resolution algorithm. Although we develop a viable algorithm for resolving alter identities across offline and online domains, we did not assess its accuracy by training it and testing its predictive capabilities on a dataset with an available gold-standard. This is largely due to the fact that no such dataset exists. As such, our matches were likely to include false positives. Additionally, our entity resolution

algorithm was ill equipped to handle pseudonyms as Facebook usernames. For example, we know members of ballroom houses typically adopt their house name as their last name. For example, a member of the House of Mizrahi named Jackson Smith may adopt the name Jackson Mizrahi as his “play” name and use an iteration of that (e.g., “J Mizrahi”) as his Facebook username. For this reason, false negatives in the matching process were also likely.

Despite these limitations, results of this exploratory study enhance our understanding of the degree to which Facebook reflects and complements the dynamic properties of high-risk personal networks of YBMSM and direct our attention to avenues for future research. It is clear that effective social network strategies for HIV testing require innovative, cost-effective methods for collecting information about high risk network associates over time. Although, perhaps not to be considered supplements for offline network data, digital social network data like Facebook, Instagram and Snapchat offer ways to complement those data to gain access to information about domains of personal network associates and their dynamics that could be leveraged by public health services to more rapidly and effectively intervene where testing is needed. Further exploration of the feasibility of implementing SNS-enhanced network testing programs is, therefore, an important next step.

Acknowledgements

This work was supported in part by the NIH (R01DA033875, R01MH100021 and R01DA039934) and conducted under the auspices of the uConnect Study Team. We would like to thank Matthew Ferreira for his contributions to the data analysis and our partners at the National Opinion Research Center (NORC) at the University of Chicago for their invaluable support. We also thank study participants for contributing to the network cohort study.

References

- Adimora AA, & Schoenbach VJ (2005). Social Context, Sexual Networks, and Racial Disparities in Rates of Sexually Transmitted Infections. *The Journal of Infectious Diseases*, 191(Supplement_1), S115–S122. doi:10.1086/425280 [PubMed: 15627221]
- Arnaboldi V, Passarella A, Conti M, & Dunbar RIM (2015). Online social networks: human cognitive constraints in Facebook and Twitter personal graphs: Elsevier.
- Arnold E, & Bailey MM (2009). Constructing Home and Family: How the Ballroom Community Supports African American GLBTQ Youth in the Face of HIV/AIDS. *J Gay Lesbian Soc Serv*, 21(2–3), 171–188. doi:10.1080/10538720902772006 [PubMed: 23136464]
- Bidart C, & Lavenu D (2005). Evolutions of personal networks and life events. *Social Networks*, 27(4), 359–376.
- Boyd DM, & Ellison NB (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1).
- Brewer DD, Garrett SB, & Kulasingam S (1999). Forgetting as a cause of incomplete reporting of sexual and drug injection partners. *Sexually transmitted diseases*, 26(3), 166–176. [PubMed: 10100775]
- Burt RS (2000). Decay functions. *Social Networks*, 22(1), 1–28.
- Carpenter CJ, & Tong ST (2017). Relational distancing and termination between online friends: an application of the investment model. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 6925–6935.
- Centers for Disease Control and Prevention. (2017). Social Network Strategy for HIV Testing Recruitment. Retrieved from <https://effectiveinterventions.cdc.gov/en/HighImpactPrevention/PublicHealthStrategies/SocialNetworkStrategy.aspx>

- Centers for Disease Control and Prevention. (2018). HIV Surveillance Report, 2017 (Vol. 29). Retrieved from <http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html>.
- Cornwell B, & Laumann EO (2015). The health benefits of network growth: New evidence from a national survey of older adults. *Social science & medicine*, 125, 94–106. [PubMed: 24128674]
- Costenbader EC, Astone NM, & Latkin CA (2006). The dynamics of injection drug users' personal networks and HIV risk behaviors. *Addiction*, 101(7), 1003–1013. [PubMed: 16771892]
- Desmond M (2012). Disposable ties and the urban poor. *American Journal of Sociology*, 117(5), 1295–1335.
- Dillman DA, Eltinge JL, Groves RM, & Little RJA (2002). Survey nonresponse in design, data collection, and analysis. *Survey nonresponse*, 3–26.
- Dunbar RIM, Arnaboldi V, Conti M, & Passarella A (2015). The structure of online social networks mirrors those in the offline world. *Social Networks*, 43, 39–47. doi:10.1016/j.socnet.2015.04.005
- Feld SL (1997). Structural embeddedness and stability of interpersonal relations. *Social Networks*, 19(1), 91–95.
- Felsher M, & Koku E (2018). Explaining HIV risk multiplexity: a social network analysis. *AIDS and Behavior*, 22(11), 3500–3507. [PubMed: 29680933]
- Fischer CS, & Beresford L (2014). Changes in support networks in late middle age: The extension of gender and educational differences. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 70(1), 123–131. [PubMed: 24898029]
- Goel S, & Salganik MJ (2010). Assessing respondent-driven sampling. *Proc Natl Acad Sci U S A*, 107(15), 6743–6747. doi:10.1073/pnas.1000261107 [PubMed: 20351258]
- Greenwood S, Perrin A, & Duggan M (2016). Social Media Update 2016. Retrieved from Pew Research Center: <http://www.pewinternet.org/2016/11/11/social-media-update-2016/>
- Haimson OL, Brubaker JR, Dombrowski L, & Hayes GR (2015). Disclosure, stress, and support during gender transition on Facebook. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 1176–1190.
- Harper R, Hamill L, & Gilbert N (2013). Modelling Digital Habitus: The relationship between the internet and the density and duration of friendship ties.
- Heckathorn DD (1997). Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. *Social Problems*, 44(2), 174 Retrieved from <http://socpro.oxfordjournals.org/content/44/2/174.abstract>
- Hsieh YP (2015). Check the phone book: Testing information and communication technology (ICT) recall aids for personal network surveys. *Social Networks*, 41, 101–112.
- Johnston LG, & Sabin K (2010). Sampling hard-to-reach populations with respondent driven sampling. *Methodological innovations online*, 5(2), 38–48.
- Joseph HA, Marks G, Belcher L, Millett GA, Stueve A, Bingham TA, & Lauby J (2011). Older partner selection, sexual risk behaviour and unrecognised HIV infection among black and Latino men who have sex with men. *Sex Transm Infect*, 87(5), 442–447. doi:10.1136/sextrans-2011-050010 [PubMed: 21705378]
- Krosnick JA (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, 5(3), 213–236.
- Lakon CM, Ennett ST, & Norton EC (2006). Mechanisms through which drug, sex partner, and friendship network characteristics relate to risky needle use among high risk youth and young adults. *Social science & medicine*, 63(9), 2489–2499. doi:10.1016/j.socscimed.2006.06.015 [PubMed: 16875770]
- Laumann EO, Gagnon JH, Michael RT, & Michaels S (1996). National Health and Social Life Survey, 1992: Inter-university Consortium for Political and Social Research.
- Levenshtein VI (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics doklady*, 10(8), 707–710.
- Marin A, & Hampton KN (2007). Simplifying the personal network name generator: Alternatives to traditional multiple and single name generators. *Field methods*, 19(2), 163–193.
- Marsden PV (1987). Core Discussion Networks of Americans. *American Sociological Review*, 52(1), 122–131. doi:10.2307/2095397

- Marsden PV (2011). Survey methods for network data. *The Sage handbook of social network analysis*, 25, 370–388.
- McCullagh P, & Nelder JA (1989). *Generalized linear models* (Vol. 37): CRC press.
- McCulloch CE, & Neuhaus JM (2001). *Generalized linear mixed models*: Wiley Online Library.
- 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 behaviors. *AIDS*, 21(15), 2083–2091. doi:10.1097/QAD.0b013e3282e9a64b [PubMed: 17885299]
- Millett GA, Peterson JL, Flores SA, Hart TA, Jeffries W. L. t., Wilson PA, ... 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. doi:10.1016/S0140-6736(12)60899-X [PubMed: 22819656]
- Mollenhorst G, Volker B, & Flap H (2014). Changes in personal relationships: How social contexts affect the emergence and discontinuation of relationships. *Social Networks*, 37, 65–80. doi:10.1016/j.socnet.2013.12.003
- Morgan DL, Neal MB, & Carder P (1997). The stability of core and peripheral networks over time. *Social Networks*, 19(1), 9–25.
- Morris M, Goodreau S, & Moody J (2007). *Sexual networks, concurrency, and STD/HIV Sexually Transmitted Diseases*. New York: McGraw-Hill, 109–126.
- Philips L (1990). Hanging on the metaphone. *Computer Language*, 7(12 (December)).
- Purdie-Vaughns V, & Eibach RP (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles*, 59(5–6), 377–391.
- Rabe-Hesketh S, & Skrondal A (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*: Chapman and Hall/CRC.
- Reich SM, Subrahmanyam K, & Espinoza G (2012). Friending, IMing, and hanging out face-to-face: overlap in adolescents' online and offline social networks. *Developmental psychology*, 48(2), 356. [PubMed: 22369341]
- Rhodes T, Singer M, Bourgois P, Friedman SR, & Strathdee SA (2005). The social structural production of HIV risk among injecting drug users. *Soc Sci Med*, 61(5), 1026–1044. doi:10.1016/j.socscimed.2004.12.024 [PubMed: 15955404]
- Roberts SGB, & Dunbar RIM (2015). Managing relationship decay. *Human Nature*, 26(4), 426–450. [PubMed: 26489745]
- Rogers W (1994). Regression standard errors in clustered samples. *Stata technical bulletin*, 3(13).
- Royall RM (1986). Model robust confidence intervals using maximum likelihood estimators. *International Statistical Review/Revue Internationale de Statistique*, 221–226.
- Rudolph AE, Crawford ND, Latkin C, & Lewis CF (2017). Multiplex Relationships and HIV: Implications for Network-Based Interventions. *AIDS and Behavior*, 21(4), 1219–1227. doi:10.1007/s10461-016-1454-2 [PubMed: 27272657]
- Schneider JA, Cornwell B, Jonas A, Lancki N, Behler R, Skaathun B, ... Laumann E (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(1), 1–29. doi:10.1017/nws.2016.27
- Schneider JA, Michaels S, & Bouris A (2012). Family network proportion and HIV risk among black men who have sex with men. *Journal of Acquired Immune Deficiency Syndrom*, 15(61), 627–635. doi:10.1097/QAI.0b013e318270d3cb
- Schneider JA, Schumm LP, Fraser M, Yeldani V, & Liao C (2018). A Gold Standard for Entity Resolution within Sexually Transmitted Infection Networks. *Scientific Reports*, 8(8776). doi:DOI:10.1038/s41598-018-26794-7
- Stack CB (1975). *All our kin: Strategies for survival in a black community*: Basic Books.
- StataCorp. (2017). *Stata Release: 15. Statistical Software* College Station, TX: StatCorp, LLC.
- Taylor P (2013). *A survey of LGBT Americans: attitudes, experiences and values in changing times*: Pew Research Center.
- Traud AL, Mucha PJ, & Porter MA (2012). Social structure of Facebook networks. *Physica A: Statistical Mechanics and its Applications*, 391(16), 4165–4180. doi:10.1016/j.physa.2011.12.021

- van Woudenberg TJ, Bevelander KE, Burk WJ, Smit CR, Buijs L, & Buijzen M (2020). Comparing the measurement of different social networks: Peer nominations, online communication, and proximity data. *Network Science*, 1–17.
- Viry G (2012). Residential mobility and the spatial dispersion of personal networks: Effects on social support. *Social Networks*, 34(1), 59–72.
- Wellman B, Wong R. Y. I., Tindall D, & Nazer N (1997). A decade of network change: Turnover, persistence and stability in personal communities. *Social Networks*, 19(1), 27–50.
- Wilson C, Sala A, Puttaswamy KPN, & Zhao BY (2012). Beyond social graphs: User interactions in online social networks and their implications. *ACM Transactions on the Web (TWEB)*, 6(4), 1–31.
- Young LE, Fujimoto K, Alon L, Zhang L, & Schneider J (2019). The multiplex social environments of young Black men who have sex with men: How online and offline social structures impact HIV prevention and sex behavior engagement. *Journal of Social Structure*, 20(3), 70–95. [PubMed: 32855626]
- Young LE, Fujimoto K, & Schneider JA (2019). Facebook group affiliation ties, group topics, and HIV behavioral characteristics among young Black men who have sex with men: Potential for public health intervention. *SSM-Population Health*, 9.

Highlights

- Capturing dynamics in high-risk personal networks is essential for HIV prevention.
- Online networks can help augment self-reported network enumeration approaches.
- Overlap exists between self-reported and Facebook network associates.
- Confidants who are also Facebook friends with a respondent tend to be retained.
- The overall stabilities of each network context are essentially unrelated.

Figure 1.

Comparison of alter lists within and between data collection waves. Solid lines show within wave comparisons and dashed lines show between wave comparisons.

Figure 2.

Degree distributions of respondents' confidant and sex partner networks at each wave. N represents the total number of respondents who provided confidant or sex partner alter data at each wave.

Figure 3.

Distribution of matched alters. (Top) Distribution of the number of uniquely named confidant (left) and sex partner (right) alters that were matched. (Middle) Accounting for only the respondents who also provided Facebook profile information, the distribution of the fraction of their named alters (left confidant, right sex partner) who were matched to any profile on Facebook. (Bottom) Among those alters who were matched to a profile on Facebook, fraction who were friends with their naming respondent on Facebook.

Table 1.

Characteristics of the Wave 1 Sample of Young Black Men who have Sex with Men (YBMSM) living in Chicago (N=618), Stratified by Whether the Respondent Provided Facebook Data

Categorical Characteristics	Respondent Provided Facebook Data				p-value ^I
	No (N=320)		Yes (N=298)		
	N	%	N	%	
Educational Attainment					p = 0.32
Less than high school	47	14.69	29	9.73	
High school diploma	157	49.06	158	53.02	
Some college (no degree)	64	20.00	66	22.15	
Associate's degree	32	10.00	24	8.05	
Bachelor's degree or more	20	6.25	21	7.05	
Unemployed					p = 0.07
No	209	65.31	215	72.15	
Yes	11	34.69	83	27.85	
Sexual Orientation					p = 0.04
Gay	208	65.00	202	67.79	
Straight	18	5.62	4	1.30	
Bisexual	85	26.56	82	27.52	
Other	9	2.81	10	3.36	
House/Ballroom member					p = 0.08
No	21	63.21	207	69.93	
Yes	17	36.79	89	30.07	
Jail (ever)					
No	181	56.74	151	50.67	p = 0.13
Yes	138	43.26	147	49.33	
Homelessness (past 12 mos.)					
No	242	76.1	219	73.49	p = 0.46
Yes	76	23.90	79	26.51	
HIV Positive					p = 0.01
No	225	70.31	180	60.4	
Yes	95	29.69	118	39.6	
Continuous Characteristics	Mean	SD	Mean	SD	p-value
Age (16–29)	22.46	3.19	23.07	3.04	p = 0.01
Confidant Network Stability ² (–3.09 – 3.00)	–0.08	0.98	0.03	1.00	p = 0.36
Partner Network Stability ² (–4.20 – 1.36)	–1.14	1.36	–1.23	1.21	p = 0.56

¹ p-values derive from Chi-square tests for categorical variables and t-tests for continuous variables.

² Estimated among a sample of 385 respondents who named at least one confidant and one sex partner at study Waves 1 and 2 (275 respondents provided Facebook data and 110 did not).

Table 2.
Characteristics of Confidant, Sex Partner, and Facebook Personal Networks at each Wave

Network	Wave	Total No. of respondents at each wave	No. of respondents who provided network data	Minimum No. Alters	Maximum No. Alters	Mean No. Alters	SD	Total No. unique alters across all respondents ¹
Confidant Network	1	618	599	1	5	2.61	1.17	1,414
	2	525	508	1	5	2.58	1.09	1,097
	3	507	488	1	5	2.47	1.07	1,206
Sex Partner Network	1	618	599	1	6	2.64	1.55	1,590
	2	525	498	1	6	2.42	1.41	1,355
	3	507	475	1	6	2.29	1.39	1,242
Facebook Network	1	618	298	14	4,723	1,139	936	182,987
	2	525	391	1	5,860	1,376	1,361	253,519

¹ Estimates of number of unique confidant and sex partner alternates are subject to false negatives. Matches of alters between participants are obtained from prior work (Schneider et al., 2017).

Table 3.

Linear regressions of confidant and partner stability on individual Facebook metrics calculated at Wave1¹

	Confidant stability ²			Partner stability ³		
	Estimate	95% CI	p-value	Estimate	95% CI	p-value
<i>Facebook metric</i>						
No. friends (n=276)	-0.11	(-0.22, 0.01)	0.084	-0.13	(-0.27, 0.01)	0.076
No. friends with messages (n=276)	-0.03	(-0.15, 0.09)	0.675	-0.04	(-0.19, 0.10)	0.557
Total messages ⁴ (n=267)	-0.03	(-0.15, 0.09)	0.641	-0.07	(-0.2, 0.07)	0.328
Med. prop. common friends ⁵ (n=275)	-0.01	(-0.13, 0.11)	0.920	0.01	(-0.13, 0.16)	0.851
Max. prop. common friends (n=276)	0.02	(-0.10, 0.14)	0.741	0.03	(-0.11, 0.17)	0.692

¹ Individual linear regressions, each with a single Facebook metric as a covariate. All covariates are transformed using square roots to reduce skewness and then standardized.

² Estimated from SEM fit to confidant data from Waves 1, 2 and 3.

³ Estimated from SEM fit to partner data from Waves 1, 2 and 3.

⁴ Nine respondents with extreme values (more than 2,500 total messages) were excluded.

⁵ One respondent with an extreme value (50% or more of Facebook friends with 2/5 or more common friends) was excluded.

Table 4.

Generalized linear regression models predicting turnover in Facebook ties from Wave 1 to Wave 2 (n = 201 respondents)¹

	Proportion Facebook ties lost ²			Number new Facebook ties ³		
	OR	95% CI	p-value	$\exp(\hat{\beta})$	95% CI	p-value
Confidant stability ⁴	1.03	(0.93, 1.15)	0.577	1.10	(0.89, 1.35)	0.397
Partner stability ⁵	1.06	(0.96, 1.16)	0.252	1.12	(0.94, 1.33)	0.202
No. FB friends (W1)	0.71	(0.63, 0.79)	< 0.001	1.52	(1.22, 1.89)	< 0.001
Age (per 1 year)	0.96	(0.93, 1.00)	0.051	0.92	(0.86, 0.99)	0.028
House/ball member	2.02	(1.61, 2.53)	< 0.001	2.30	(1.51, 3.49)	< 0.001

¹Estimated among 201 respondents with Facebook data at Waves 1 and 2.

²Logistic regression with binomial denominator equal to the number of Wave 1 Facebook friends.

³Generalized linear model with log link and gamma variance function.

⁴Estimated from SEM fit to confidant data from Waves 1, 2 and 3.

⁵Estimated from SEM fit to partner data from Waves 1, 2 and 3.

Table 5.

Criteria used for ranking potential alter-Facebook matches and the number of resulting matches.

Levenshtein distance of alter name and profile name	Number of possible matches	Criteria			Number of alter-Facebook profile matches	Number of alter-Facebook profile matches where Facebook profile is friends with respondent
		Difference in Age	Same sex?	Same city?		
0	1	-	-	-	44	22
0.5	1	5 years	Yes	Yes	5	4
0.5	1	5 years	Yes	No	5	1
0.5	1	5 years	No	Yes	4	3
0.5	1	5 years	No	No	39	9
0.5	1	> 5 years	Yes	No	2	0
0.5	> 1	5 years	Yes	Yes	380	191
0.5	> 1	5 years	Yes	No	350	126
0.5	> 1	5 years	No	Yes	113	80
0.5	> 1	5 years	No	No	513	307
0.5	> 1	> 5 years	Yes	Yes	21	12
0.5	> 1	> 5 years	Yes	No	35	4
0.5	> 1	> 5 years	No	Yes	24	21

Table 6.

Logistic regression models of the likelihood of locating confidants and sex partners in Facebook, and conditionally, on the likelihood of being a Facebook friend with the respondent¹

A. Confidants	Found in Facebook (n =1,012)			Facebook Friend (n =687)		
	OR	95% CI	p-value	OR	95% CI	p-value
Alter/Tie Characteristics						
Order named (1–5)	0.95	(0.84, 1.07)	0.409	0.97	(0.84, 1.12)	0.658
Gender (ref. Male)						
Female	1.12	(0.64, 1.97)	0.697	1.52	(0.73, 3.14)	0.260
Transgender	1.31	(0.44, 3.91)	0.630	1.07	(0.31, 3.72)	0.921
Relationship type (ref. Family)						
Partner	1.44	(0.89, 2.34)	0.141	0.94	(0.54, 1.66)	0.844
Friend	1.41	(0.99, 2.01)	0.054	1.44	(0.96, 2.16)	0.075
Other	0.96	(0.58, 1.62)	0.890	0.98	(0.52, 1.85)	0.950
Freq. of interaction (1–5)	0.97	(0.82, 1.14)	0.689	1.04	(0.87, 1.24)	0.683
Closeness (1–4)	0.90	(0.62, 1.30)	0.567	0.88	(0.60, 1.29)	0.510
Alter is MSM	0.99	(0.56, 1.78)	0.986	0.79	(0.39, 1.60)	0.505
Wave 2 (vs. Wave 1)	0.75	(0.53, 1.06)	0.102	0.56	(0.37, 0.84)	0.002
Respondent Characteristics						
Age (per 1 year)	1.06	(1.01, 1.12)	0.032	0.98	(0.93, 1.04)	0.603
Education (ref. less than high school)						
High School	0.55	(0.31, 0.98)	0.044	0.66	(0.34, 1.28)	0.215
More than high school	0.61	(0.33, 1.13)	0.144	0.95	(0.47, 1.91)	0.892
Sexual Orientation (ref. Gay)						
Straight	0.32	(0.11, 0.99)	0.049	0.14	(0.02, 0.94)	0.043
Bisexual	1.05	(0.75, 1.46)	0.791	0.87	(0.59, 1.29)	0.497
Other	1.20	(0.60, 2.39)	0.602	0.67	(0.32, 1.41)	0.293
House/ball member	0.84	(0.60, 1.17)	0.292	0.95	(0.67, 1.35)	0.794
Jail (ever)	0.63	(0.46, 0.87)	0.004	1.46	(1.06, 2.02)	0.020
Variance of random effect	0.21	(0.05, 0.94)	0.058	0.00		
B. Sex partners						
	Found in Facebook (n =1,537)			Facebook Friend (n = 771)		
	OR	95% CI	p-value	OR	95% CI	p-value
Alter/Tie Characteristics						
Order named (1–6)	0.81	(0.73, 0.89)	< 0.001	1.01	(0.88, 1.15)	0.907
Partner type (ref. Main)						
Casual	0.66	(0.52, 0.85)	0.001	1.01	(0.71, 1.44)	0.965
Exchange	0.59	(0.31, 1.13)	0.111	0.74	(0.27, 2.03)	0.555
Wave 2 (vs. Wave 1)	0.73	(0.59, 0.90)	0.004	0.79	(0.58, 1.08)	0.146
Respondent Characteristics						
Age (per 1 year)	0.95	(0.90, 0.99)	0.016	0.96	(0.90, 1.02)	0.216

Education (ref. less than high school)						
High School	0.67	(0.45, 1.01)	0.053	0.95	(0.52, 1.74)	0.866
More than high school	0.77	(0.49, 1.20)	0.253	1.38	(0.72, 2.65)	0.338
Sexual Orientation (ref. Gay)						
Straight	0.94	(0.28, 3.17)	0.922	0.31	(0.06, 1.54)	0.152
Bisexual	0.98	(0.72, 1.33)	0.898	0.92	(0.62, 1.37)	0.672
Other	1.35	(0.58, 3.13)	0.491	0.62	(0.27, 1.43)	0.266
House/ball member	1.08	(0.80, 1.46)	0.620	1.23	(0.6, 1.82)	0.288
Jail (ever)	0.98	(0.75, 1.29)	0.894	1.01	(0.70, 1.45)	0.974
Variance of random effect	0.40	(0.24, 0.68)	< 0.001	0.46	(0.17, 1.28)	0.005

[/] Fit to confidant and sex partner data from Waves 1 and 2 (n = 295 respondents overall, 271 respondents with at least one confidant found in Facebook, and 269 respondents with at least one sex partner found in Facebook).

Table 7.

Association between Facebook friendship status and the likelihood of retaining confidants from the previous timepoint^I

A. Probability of retaining previous confidant (n = 1,276 alters)			
	Odds ratio	95% CI	p-value
Relationship attributes (last timepoint)			
Order named (1–5)	0.57	(0.49, 0.66)	< 0.001
Also sex partner	1.38	(0.92, 2.07)	0.120
Found in Facebook	1.25	(0.85, 1.84)	0.257
Facebook friend	1.46	(1.03, 2.08)	0.034
Respondent attributes			
No. of confidants (last timepoint)	1.10	(0.94, 1.29)	0.212
MSM network size (0–20)	1.01	(0.99, 1.04)	0.250
Confidant network stability (latent)	2.33	(1.70, 3.18)	< 0.001
9–18 months (vs. 0–9)	0.57	(0.45, 0.73)	< 0.001
B. Probability of reporting new confidant (n = 1,934 possible slots)			
Respondent attributes			
No. of confidants (last timepoint)	0.89	(0.79, 0.99)	0.029
MSM network size (0–20)	1.00	(0.98, 1.01)	0.729
Confidant network stability (latent)	0.72	(0.49, 1.07)	0.108
9–18 months (vs. 0–9)	0.82	(0.67, 1.00)	0.047
C. Regression of confidant network stability on baseline characteristics (n = 275 respondents)			
	Estimate	95% CI	p-value
Age (per 1 year)	0.08	(0.01, 0.15)	0.027

^I Structural equation model in which a single, respondent-specific latent variable (network stability) is assumed to affect both the likelihood of retaining confidants from the previous wave and the likelihood of naming new confidants. Estimated among 298 respondents who provided Facebook friendship data at Wave 1, using their self-reported network changes from Waves 1 to 2 and Waves 2 to 3.