|  |
| --- |
| UCSD Human Research Protections Program **New Biomedical Application**  **RESEARCH PLAN** |
| 1. **PROJECT TITLE** |
| AEHIV 038: Using Social Media to characterize social and sexual networks of men who have sex with men (MSM) |
| **2. PRINCIPAL INVESTIGATOR** |
| Principal Investigator:  Susan Little, MD  Professor of Medicine, Division of Infectious Diseases, UCSD  Nadir Weibel, PhD  Research Scientist, Computer Science and Engineering, UCSD |
| **3. FACILITIES** |
| The study will be conducted at the UCSD Antiviral Research Center (AVRC) located at 220 Dickinson, Suite A, San Diego, CA 92103, and at the UCSD Computer Science and Engineering Department (CSE) located on UCSD’s main campus, building EBU3B, La Jolla, CA 92093  The following facilities are selected San Diego Community and Public Health testing sites that will be associated with recruitment: The Principal Investigator, Dr. Little, will be responsible for the conduct of the study at these sites. As such, Dr. Little with the help of the study staff will be responsible for ensuring proper study materials are available at each site and that the implementation of new protocol amendments and training that may be needed throughout the study is conducted.  ***UCSD Lead the Way Storefront***  3830 Park Blvd, Suite 124  San Diego, CA 92103  ***UCSD AVRC Mobile Testing Van***  ***San Diego LGBT Community Center***  3909 Centre St.  San Diego, CA  ***Selected County Health Services Complex/Public Health Services:***  3851 Rosecrans St.  San Diego, CA 92110  *San Diego County Health and Human Services Night Street Outreach Testing*  *San Diego County Health and Human Services Mobile Test Van*  ***Family Health Centers of San Diego’s Tuesday and Thursday Night Clinic For Gay Men***  3544 30th Street  San Diego, CA 92104 |
| **4. ESTIMATED DURATION OF THE STUDY** |
| 5 years |
| **5. LAY LANGUAGE SUMMARY OR SYNOPSIS (no more than one paragraph)** |
| As of today 36 million people have died of AIDS, and 35.3 million people are living with HIV worldwide [1]. Despite the tremendous speed at which information technology has improved, we still have no fast and accurate methods to identify populations susceptible to HIV infections. This adversely impacts our efforts to implement targeted and therefore more effective HIV prevention. As a solution, we propose a new method of identifying HIV at-risk populations using publicly available social media data as an indicator of HIV risk. Recent research [2] has outlined the feasibility of using Twitter as a broad but real-time monitoring tool for HIV. We propose to take this approach further, identifying and characterizing HIV at-risk populations locally in the San Diego area at a more granular scale in terms of both demographics and communities. Using social network analysis and machine learning techniques, we will combine social media data with the data available from the Primary Infection Research Consortium (PIRC) at the UCSD AntiViral Research Center (AVRC). Our overall aim is to use Twitter to map the local social/sexual network dynamics of at-risk MSM and overlay these with the well-characterized San Diego Primary Infection Cohort (SD PIC) HIV transmission network [3] available at AVRC. Where these networks intersect will provide opportunities to evaluate targeted approaches to HIV prevention interventions. In this project we therefore propose to characterize HIV at-risk population among MSM in San Diego using Twitter data. |
| **6.** **SPECIFIC AIMS** |
| We plan to use publically available data (Phase 1) and confidentially collected data (Phase 2) from Twitter to characterize the structure of social networks of men who have sex with men (MSM) who may be at risk of acquiring or transmitting HIV infection in the San Diego Area.  The specific aims of this study are   1. Evaluate the use of natural language processing applied to publically available Twitter data to identify the social contacts and strength of these contacts in MSM living in California. 2. To infer the social network structure of MSM who may be at risk for HIV infection in California. 3. To infer the sexual network structure of HIV infected MSM in California 4. To compare the MSM social and sexual network structures derived from social network data in California. 5. To compare the sexual network structure of HIV infected MSM derived from social network data to that inferred from well-characterized HIV sequence data in the San Diego Primary Infection Cohort (SD PIC). 6. To assess the utility of social network data to characterize and at-risk social and sexual network to efficiently focus HIV prevention services. |
| **7. BACKGROUND AND SIGNIFICANCE** |
| HIV remains a significant public health problem in the United States today, and this disease disproportionately affects men who have sex with men (MSM). The numbers of new HIV infections in MSM living in the United States are increasing (8% increase between 2001-2012 12% from 2008-2010 [4], and 22% increase among young [13-24 yr old) gay and bisexual men). Similarly, the incidence of HIV in MSM in the United Kingdom has increased between 1990 and 2010 (estimated mean incidence 0.30/100 person-years 1990–1997, 0.45/100 p-y 1998–2010) [5]. The generalized approach to HIV prevention, such as the “safer sex” public health campaign, has not been effective in these populations. In particular, efforts to focus prevention messages and resources to those at greatest risk of acquiring infection are lacking. We believe that social networks can provide the framework to identify appropriate HIV prevention targets. In prior studies, social networks have been shown to influence obesity [6], smoking [7], drug use [8], drinking [9], and sleep [10], and important affective states like happiness [11], loneliness [12], and depression [13]. If we map the social and sexual network dynamics of MSM at greatest risk of acquiring HIV, and associate these networks with available, well-characterized sexual network characteristics of HIV-infected transmission networks, we will have the opportunity to evaluate novel targeted approaches to HIV prevention interventions, such as pre-exposure prophylaxis (PrEP).  Use of molecular epidemiological methods have greatly increased our understanding of HIV transmission dynamics. Using the HIV sequence data derived from routine (standard of care) HIV drug resistance testing, it is possible to infer a partial transmission network. Our group has used these data to extensively characterize the San Diego Primary Infection Cohort (SD PIC) transmission network and identify the network features of persons at greatest risk for HIV transmission within their first year following incident infection. Simulations of these data also demonstrate that targeted antiretroviral therapy (ART) to those with the highest overall risk of transmission (based on an objectively derived “transmission network score”) resulted in a significantly greater reduction in HIV network transmission as compared to random ART2. However, while early and universal ART in HIV infected persons is the cornerstone of effective HIV prevention, these methods do not address the HIV acquisition risk among HIV uninfected persons in the same sexual network.  Just as HIV transmission networks are used to identify high-yield prevention targets for ART interventions, we believe that social network data derived from social media (Twitter) data may be used to characterize the California MSM social network structure, and may help to identify those at greatest for acquiring HIV infection. Twitter is a highly trafficked social media network built on *tweets* and *followers* used around the world. It is used not only to connect with friends/acquaintances online but also to follow real-time events (such as the Arab Spring) and issues users find interesting. The potential to use Twitter as real-time HIV monitoring method was explored using HIV-risk related tweets that contained drug and sex related words in the United States in relation to HIV prevalence data3. These data demonstrated a significant correlation between higher numbers of HIV-risk communications and higher HIV prevalence within the county. This suggests that HIV-risk behavior that is shared by Twitter users online via tweets to their followers may be used to infer regional network characteristics of HIV risk. We propose to investigate the Twitter connections between users who tweet about high-risk behavior to evaluate the use of Twitter data as a real-time tool to characterize and monitor MSM social network connectivity (i.e. the relative strength and number of connections in a social network) in California, and ultimately high-yield targets for HIV prevention. |
| **8. PROGRESS REPORT** |
| In our preliminary work we explored the best way to collect data from Twitter and have decided to the use the “Streaming APIs” provided by Twitter. We preliminary collected publicly available data for three different geographical scopes: World, San Diego city and San Diego county. We receive around 70 tweets per second on the *sample hose* Streaming API for world scope. The *filter hose* Streaming APIs for other two scopes provide around 40 tweets per minute. The Twitter data collected are stored on a NoSQL database called mongoDB. We perform preprocessing on the collected tweets to make it easy to derive inferences later during the research. |
| **9. RESEARCH DESIGN AND METHODS** |
| Twitter is an extremely active online social network built on *tweets* and *followers*. A Twitter user can tweet their own information, can *follow* any other Twitter user and therefore receive notification of their *tweets,* and can *retweet* the tweets posted by others they are following. *Tweets* are textual messages, max. 140 characters long. Tweets shared by a user can be linked to other Twitter users using the @ symbol (identifying a specific user), or can be linked to a specific topic exploiting hashtags (using the # symbol). Tweets can also contain links to articles or pictures on the internet.  The majority of this information is publically available data, and as part of this study we propose to use these data to build a Twitter-based backbone structure of the social and sexual networks of MSM in California (including HIV infected and uninfected persons) and study the relationship between these network structures and characteristics with the HIV transmission network inferred for HIV infected participants of the SD PIC. We call the resulting infrastructure composed by the public network and the SD PIC network, “PIRC-Net”.  Distinct consent processes are proposed for the Phase 1 and Phase 2 components of the study. Phase 3 of the study will not require any additional participant recruitment/enrollment, but will include analyses of data collected from Phase 1 and 2. In order to achieve the proposed analyses, this study is based on three Phases.  ***Phase 1: Backbone MSM-at risk Public Network***  We propose to build a backbone network of tweets likely originating from MSM individuals at risk of acquiring or transmitting HIV based on public available Twitter data from San Diego County. We will exploit the technology identified in our preliminary investigations to retrieve the data. In order to achieve clean and relevant data we will first: use only tweets that have been geolocated to San Diego County, second: filter the available data on the syntactic level, and third: proceed to a semantic filtering. All Twitter data gathered during phase 1 is publically available data and will be gathered without the consent or knowledge of the Twitter users.  First: Tweets that have been sent when the user’s phone or computer GPS services are turned on are “geolocated” to the physical location of that device. We plan to only use tweets that are geolocated to San Diego County.  Second: The syntactic filtering will be based on three basic preprocessing methods executed on the tweets:   1. Conversion to lowercase letters 2. Removal of punctuations 3. Removal of stop words   Third: The semantic filtering has the goal to identify all the *HIV risk tweets* by checking if a given tweet contains HIV-risk words found in a static dictionary of HIV-risk terms that our research team will define as part of this project. We plan to bucket the HIV-risk terms in 5 categories: Drugs Bucket, Homosexual Terms Bucket, Sex behavior Bucket, Sex venues Bucket and Sexually Transmitted Infections Bucket. Each tweet will then be assigned to a ranking for the five categories based on how closely it aligns to each dictionary.  We plan to then analyze the collected HIV-risk data through two asynchronous processes with the goal of identifying missing/duplicate information. The first process checks if each of the collected tweets is a reply to another tweet or if it is a *retweet*. In either case, related tweets will be identified and acquired using the Twitter’s *status* API. Once all related tweets are acquired, a second process will identify related Twitter *user handles* and fetches the publicly available user information using the corresponding Twitter *users* API. All personal identifiable information contained in the tweets (namely tweet authors, followers, and followees) will be redacted and automatically assigned a unique confidential identifier. Confidential identifiers will allow us to characterize the social connections across the network, without the need to maintain private identifiable information  The large amount of data collected through this method will allow us to construct a large social graph. We will use the Neo4J graph database to store the data, which will enable us to characterize and study the various social structures of the Backbone MSM-at risk Public Network.  ***Phase 2: Validation of Backbone HIV Network***  Although the semantic filtering described above will allow us to filter data in such a way to retrieve data from MSM at risk individuals, in this Phase we aim to validate our data enriching the backbone with a layer of demographic data, Twitter user-information, and published data retrieved from selected SD PIC participants.  Men 18 years and older will be recruited from our Early Test (HIV uninfected) and our PIRC (HIV infected) programs (UCSD projects #140585 and #140093, respectively) to participate in the proposed study. Inclusion criteria will include age greater than 18, male gender and sex with men in the past 5 years, and an active Twitter account with at least one tweet in the prior 2 weeks. After appropriate informed consent (see below), , participants will allow access to their private Twitter accounts. We will build a simple Twitter application that will allow enrolled users to easily give us access through the application to key information from their Twitter posts. Our developed Twitter application will generate a user-based authentication token that will allow us to access their posts with a “user context”, i.e. seeing everything participants sees, this will include social connection as well as geo endpoints (see <http://support.twitter.com/articles/14016-about-public-and-protected-tweets> <https://dev.twitter.com/oauth/application-only>).  In other words this approach will allow us not only to have more accurate demographic and location data (through any geographic data they share), but also access to their private tweets and their usage of Twitter and “see” what they “see” through Twitter. Study participation will involve collection of participants’ tweets, re-tweets, geographic location, and your friends’ User IDs (known as your friend list) as part of participants’ basic information. This includes the information Twitter users choose to make public, as well as the information that is always publicly available. Demographic and geographic data provided by each participant will help us understand characteristics of individual users, while the private tweets and the “user context” will allow us to establish connections not found when relying solely on the publicly available data. The demographic information as well as keyword data will be generated from their pre-existing demographic and HIV risk behavior information that participants have previously provided as part of their study participation in either Early Test or PIRC. (UCSD projects #140585 and #140093, respectively).  **PIRC-NET application installation and data collection**  Subjects will be asked to install a free Twitter application in their Twitter account, PIRC-Net. PIRC-Net will not be creating any new data and all existing data remains the sole property of the profile owner, which adheres to Twitter’s policy on data ownership. The application itself will be owned and maintained by the co-PIs, and it will be active for as long as the co-PIs have permission from the IRB to keep it active. The PIRC-NET application will be deployed on our servers, and its sole role will be to link the recruited participants’ twitter profiles to our study. The installation process will therefore be limited to associating the subjects Twitter account through the generation of an access token. The access token will be stored securely on our server, and nothing will be installed locally on the computer or the smartphone of the participants. Our PIRC-NET application will be listed in the “Apps” list of the Twitter account’s settings and participants will be able to revoke access to their account at all times by clicking the “revoke” button associated with the PIRC-NET application. This will also include temporary access revocation with the goal of short periods of “private” activity, if participants decide they don’t want to share these particular data. The “revoke access” functionality can be deactivated, by selecting the Twitter “Undo Revoke Access” button for the PIRC-NET application.  As part of the installation phase, participants will be presented with information about the study:  *“This application is part of a scientific study carried out at the University of California San Diego (UCSD), investigating the structure of social networks to inform HIV prevention. If you use this application, you consent to participate in this study. Your demographic information, as well as information about your social network (friends) and your interaction with them will be collected, anonymized and stored on a secure server at UCSD. Your data will always remain confidential and may be studied by researchers or used in scientific publications. If you do not consent to participate in this study, please do not use this application. For more information or detailed description of the study, please go to http://pircnet.ucsd.edu or contact the researchers by email at pircnet@ucsd.edu.”*  By proceeding with the installation of the application they will consent for the application to collect data as described below. As part of the installation process we will also provide the participant agreeing to participate in the study with a link to the consent form.  After consenting, subjects will be asked to authenticate the application to access Twitter information about the subject. We will first confirm that the participant has the required amount of Twitter activity as required per the inclusion criteria, and then we will proceed to collect initial information about the participant, and then collect the time and type of interactions with entities published on the participant’s Twitter account. The initial information about the participant will include their basic profile information (Twitter name, time zone, location, website). By using the list of followers and followees retrieved from the participant’s account we will then collect Twitter’s interactions from published tweets, and re-tweets. The PIs will have access to only the consented users’ sent protected tweets and recipients’ user IDs. Any private tweets that are not generated by the study subject will not be available to researchers. Private tweets shall be designated as private in the collected research data.  For Phase 2 participants, PIRC-Net will collect the same public available information as in Phase 1, but will give us a direct access to geo endpoints and social collection (see above). Permissions requests will be directly embedded in the developed PIRC-Net application and will be shown to the participant before enrolling the patient in the study. Once consented, we will match the participant’s ID with the PIRCs he/she participated in. We will use the participant assigned ID number, and the participant’s name will never be linked to the collected data. We will link the Twitter data with the list of involved PIRCs and the date the single participants where actually involved in the PIRC.  Please note that this research project is exceptionally *non-intrusive* to participants. Subjects do one thing and one thing only. Once they have consented and authenticated, subjects do not need to take any more actions. Instead of surveying Twitter users, we will collect data directly from Twitter about participant network connectivity. We will not be retaining any non-anonymized personal information in this study.  All data collected will be linked to a unique ID, which will be the only way of tracking a participant within our database. Of note, not only study participants will have unique IDs, but also any third parties identified through private tweets will also be linked immediately to a unique ID. Additionally, the content of the tweets will be redacted to replace any direct reference to another user through the @ sign with the unique ID assigned to that user. We will redact URLs directly integrated in the tweets and, after classifying them, we will apply specific “Name Entity Recognition” Natural Language Processing filters to attempt to remove Personally Identifiable Information (PII), such as the MITRE approach (<http://mist-deid.sourceforge.net>). Additionally in an effort to limit content reconstruction and linkage to the original users who posted the message, we will remove stop words and punctuation from tweets and uniform them to a lowercase form, and apply stem words reduction techniques. Finally, all collected geographic information based on latitude and longitude will be abstracted at their 5-digit ZIP code level.  Despite all of our effort, given the heterogeneous way that PII can be expressed in tweets, note that there is no guarantee that all PII information will be conceived. Should a participant decide to remove his or her data while removing the app, he or she can request it directly as part of the application de-installation process. The ID is the key to link and get access to the stored data, and once access is revoked, this ID will not be linked to the PIRC-NET app anymore. The application will be notified of this revoked access and the ID will be communicated to the backend server application, which will automatically remove all information related to that ID.  Once the IRB approval has ended we will automatically remove the PIRC-Net application from all Twitter profiles, by invalidating all active access tokens on Twitter. This will be transparent for the user who will only notice the application not being listed anymore. We will issue this request on-demand by calling a specific functionality when the Co-PI and the IRB decide that this study came to an end. Participants can also manually revoke access to the PIRC-NET application, or check that upon termination of the study the App does not have access to their profile anymore, by clicking on the “revoke” button next to the PIRC-NET application in their Twitter account settings.  ***Phase 3: Networks analysis and comparison***  We will analyze both the raw backbone public network, as well as the SD PIC network and investigate how well they correlate with our well-described SD PIC HIV transmission network. Like many other social networks, the SD PIC network has “scale-free” network topology, which we would expect to see in the backbone public network as well as the SD PIC network. We will also compare the two networks and investigate the value added by the research participants’ feedback and ultimately combine the two into a comprehensive Twitter-based network of at-risk individuals. The resulting network will be overlaid with our SD PIC HIV transmission network. Where these two networks intersect: the at-risk Twitter-based network with the known HIV transmission network, will identify targets for prevention interventions such as PrEP (although it is outside of the scope of this project to design or implement any prevention intervention, we believe that the scope and the innovative nature of our proposed project will drive much interest and potential opportunities for initiating a new wave of data-driven prevention strategies). We can use network modeling to identify the overall network impact of targeted PrEP to high risk individuals, and if this will have a disproportionate impact of HIV transmission. |
| **10. HUMAN SUBJECTS** |
| In Phase 2 we intend to recruit human subjects from Early Test and PIRC studies (UCSD projects #140585 and #140093, respectively)who have been interested in participating in additional research studies age ≥ 18 years. Data analyses for Phase 1 and 3 will not require recruitment of any human subjects, however we will be using publically available Twitter data from users’ accounts without their consent and knowledge of the research study.  ***Phase 1:*** The first Phase of the study involves using data available in the public domain. For this component of the research, the researchers will filter the publicly available Twitter data in an attempt to isolate users in California who are men who have sex with men and at risk for acquiring or transmitting HIV. The Twitter user information will be made confidential for use in the research.  ***Phase 2:*** Over a period of 5 years we will enroll 2000 total subjects. We will recruit via email Early Test and PIRC study patients who have communicated an interest in participating in future research studies who are at least ≥ 18 years old. Existing participants in our Early Test (HIV uninfected) and our PIRC (HIV infected) programs (UCSD projects #140585 and #140093, respectively) will be sent an email to participate in the proposed study. For a copy of the email, please see Appendix D.  This research study poses no greater than minimal risk to the subject. Study participation will involve collection of HIV risk behavior data that had been previously provided as part of Early Test or PIRC, as well as participants’ tweets, re-tweets, geographic location, and your friends’ User IDs (known as your friend list) as part of participants’ basic information. This includes the information Twitter users choose to make public, as well as the information that is always publicly available.  ***Phase 3***: The data collected from Phase 1 (no informed consent [IC] required) and II participants (IC required) will be used for secondary analyses to evaluate the two groups separately and then together – with the plan to use network models to better understand where there may be opportunities for future targeted prevention interventions.  **Inclusion Criteria (Phase 2)**   1. Male >18 years of age 2. Report of sex with sex with other men within the last 5 years. 3. Must have a personal Twitter account and tweeted at least once during the preceding two weeks. 4. Must have the capacity to provide informed consent. 5. Subject must be able to read and understand English.   **Exclusion Criteria (Phase 2)**   1. Refusal to participate. 2. Unable to understand informed consent. |
| **11. RECRUITMENT AND PROCEDURES PREPARATORY TO RESEARCH** |
| HIV uninfected participants will be recruited from the PIRC HIV screening program (UCSD project #101414 [Early Test]) and HIV infected participants will be recruited from the PIRC (UCSD project #140093).  When participants are enrolled in either Early Test or PIRC research studies, they are asked whether or not they would be willing to be contacted for any future research studies. We have participants’ email addresses and will email each participant for recruitment. There are approximately 8,000 participants who would be eligible for this email recruitment (i.e. agreed to be contacted for future research and also have an email address). We plan to send this recruitment email to a random 1,000 participants as a first “wave” to estimate approximate recruitment, and hope for a total of 2,000 participants. We will continue to email in 1,000 participant waves until this goal has been met. Please see appendix D for a copy of the email. If recruitment flyers are developed, they will be submitted for IRB approval prior to posting. |
| **12. INFORMED CONSENT** |
| This protocol and the requested waiver of informed consent document and any subsequent modifications will be reviewed and approved by the UCSD Institutional Review Board responsible for oversight of the study.  The proposed study will be collecting Personally Identifiable Information (PII).  ***For Phase 1:*** We propose a waiver of informed consent for all Phase 1 (PII) as this will be collected as publically available data. Regardless, we will redact all personal identifiable information contained in tweets and twitter accounts as well as automatically assign each user a unique confidential identifier. Confidential identifiers will allow us to characterize connections across the network without the need to maintain private identifiable information. Because the PII we will be collecting in Phase 1 are within records already in the public domain, our collection will therefore not adversely affect the privacy rights and welfare of these individuals.  ***For Phase 2 & 3:*** We propose a waiver of documented informed consent for all Phase 2 & 3 PII as the proposed study presents no more than minimal risk of harm to subjects and involves no procedures for which written consent is normally required outside of the research context. In addition, the consent form would be an additional document linking the participant and the research and therefore would increase the risk for a breach of confidentiality.  ***Moreover, for Phase 2:***   * Subject privacy rights are not violated by waiving documented consent * The project could not practically be conducted without a waiver * The privacy risks are reasonable relative to the anticipated benefits of research * An adequate plan to protect identifiers from improper use and disclosure is included in the research proposal   Although we request a waiver of documented consent, we will track enrolled participants and document their acknowledgment to our Twitter application terms and their understanding of the research with an explicit “I agree to participate” statement that every Phase 2 enrolled participants will have to select prior to be allow our Twitter application to collect their data. Information on user’s acknowledgment of the consent and terms of participation will be stored confidentially and securely on our server and linked to the uniquely generated user ID. Only when participants will accept the application’s terms and conditions they will be enrolled in the study. An additional statement regarding the research will be posted on a linked web site and can be downloaded by the subject.  We will protect the privacy and confidentiality of both the main subjects and their social community, as well as the confidentiality of any information they provide. We will collect their information and ensure confidentiality by classifying users with a key unrelated to their name or profile (e.g. A-Z) and link the data and the key to the collected information. |
| **13. ALTERNATIVES TO STUDY PARTICIPATION** |
| The alternative to phase 2 participation for PIRC NET app users is to not participate. Twitter API users will be included in phase 1 of this study without their consent and knowledge, as their information is publically available, and therefore will not have the opportunity to opt for an alternative to study participation. |
| **14. POTENTIAL RISKS** |
| There is a minimal risk of loss of confidentiality and subsequent risk of social reputation and standing. No individually identifiable data will be shared in the publication or dissemination of study results. No participant will be provided feedback regarding qualitative or quantitative statements about their personal data collected. These data will be stored on a secure server as outlined in Section 15, with permissions to access PII information limited to a small number of critical study personnel. There is additionally potential loss of confidentiality risk to non-participants who are identified as parts of a study participant’s Twitter network. |
| **15. RISK MANAGEMENT PROCEDURES AND ADEQUACY OF RESOURCES** |
| There are no health risks associated with this study.  Loss of confidentiality has to be taken into consideration, although this would likely not be serious for the subjects since the primary data collected has already been consciously and actively made public by the enrolled subject on her social network. By joining the online application, subjects grant access to personal and identifying information associated with their social network accounts. However, we will only store and analyze published information pertaining to their social network in a confidential way with unique and confidential patient identifiers.  Twitter *user handles* are the publicly available user information. All personal identifiable information contained in the tweets (namely the Twitter *user handles* and user information about tweet authors, followers, and followees) will be redacted immediately and automatically assigned a unique confidential identifier. Confidential identifiers will allow us to characterize the social connections across the network, without the need to maintain private identifiable information. We will also immediately assign confidential identifiers to any third parties identified through private tweets.  Collecting posted information within the social network is key to effectively study the participants’ social network. Since this research presents minimal risks, it will not adversely affect the rights and welfare of the online community of the enrolled subject. In the event that a study participant engages in any private communication (ie NOT a part of the public data domain) with a third party, we will only be able to monitor and store the study participant’s data. In other words, we will only have access to one side of any private conversation: the consented participant’s side. We will not have access to any third party private information. Twitter handle will be immediately turned into a unique ID and the content of the tweets will be redacted to replace any direct reference to another user with the unique ID assigned to that user. Additionally, retweets from a third party who has not consented to participate in this study will not be available for researchers. Moreover, after classifying the tweets we will filter them to attempt to remove PII. Given the heterogeneous way that PII can be expressed in tweets, note that there is no guarantee that all PII information will be conceived. Stored confidential data will be accessible only via password-protected secure login. Strong passwords will be chosen to protect confidentiality. Only the PIs and designated research assistants will be allowed to access the data, and no identifying information will ever leave the secure environment. The risk to users is unlikely to be significantly different from the risk they already bear by having their data stored on the publicly available Twitter servers.  Should unanticipated events involving risks to subjects or others occur, the reporting of unanticipated problems and adverse events will occur in a timely manner. The unanticipated event will be reported to the IRB and appropriate institutional officials within 2 weeks of the investigator becoming aware of the event. |
| **16. PRIVACY AND CONFIDENTIALITY CONSIDERATIONS INCLUDING DATA ACCESS AND MANAGEMENT** |
| The AVRC research staff has undergone the CITI Biomedical Human Research, and Good Clinical Practice (GCP) training along with the HIPAA training. We will obtain a Certificate of Confidentiality for this study from The U.S. Department of Health and Human Services (DHHS). Even with the Certificate of Confidentiality, if the study staff learns of possible child abuse and/or neglect or a risk of harm to the participant or others, we will be required to tell the proper authorities. We may need to report information about known or reasonably suspected incidents of abuse or neglect of a child, dependent adult or elder including physical, sexual, emotional, and financial abuse or neglect.  The research staff will protect Patient Protected Health Information (PHI) or other PII of any individual in general, obtained from as part of the University or Healthcare or other work-related records, for whatever purpose, as private and confidential, and will make every effort to safeguard such information from unauthorized access or dissemination. Steps in place to protect this information are outlined below (Data Security).  ***Consent Process***  For confidentiality purposes, interested clients will be instructed online to complete the requested waiver of informed consent the consent process in a private location and using a secure/private internet connection.  ***Data Security***  Any data collected as part of this study that is stored at the AVRC or CSE and/or is transferred via the internet will follow our data security process as outlined below.  With the fast-developing technology, dependable and comprehensive data security measures are key components to defy the perceived threats of Internet hackers and accidental disclosure of confidential information. In the following we provide a summary of the key features pertinent to this project.   * A confidential participant identification number is used for all data collection, recording, and submission to the project database. * Data that contain any participant identifiers (e.g., name or contact information) other than the unique identifier are password protected and accessible only to staff members whose job requires knowledge of such data. * Research staffs are instructed not to disseminate any participant identifiers in any communications with, or data submissions to, any other AVRC collaborators. Any data transfer over the Internet uses encryption. * Data transfer and all Web-based utilities use secure access (user and server authentication, 128-bit SSL encryption). This type of encryption is the same as is used for Web-based transactions that involve credit cards or Web banking. * The server infrastructure will be setup to minimize any intrusion and potential loss of data. In particular hard drives will be encrypted, firewalls will follow a DROP default policy, only allowing access to the services that need to be exposed on the network (such as the web server); HTTPS and SSL security will be set to the strongest possible standard (an “A” rating provided by the Qualisys SSL Lab report), including disabling SSL3 access, as well as 1-step SSL certificate download; application services will be separated using secure virtualization software such as Docker; a dedicated Apparmor profiles tailored to the specific services and requirements of our application will be defined on the server.   ***Study Closure***  Once the IRB approval has ended we will automatically remove the PIRC-Net application from all Twitter profiles, by invalidating all active access tokens on Twitter. This will be transparent for the user who will only notice the application not being listed anymore. We will issue this request on-demand by calling a specific functionality when the Co-PI and the IRB decide that this study came to an end. Participants can also manually revoke access to the PIRC-NET application, or check that upon termination of the study the App does not have access to their profile anymore, by clicking on the “revoke” button next to the PIRC-NET application in their Twitter account settings. Data collected during the study will be kept anonymized on our secure servers. We will exploit the security measures explained above to ensure protection of the collected data.  ***Biological Samples***  No biological samples will be collected as part of this study. The PIRC-Net app will collect data as described to characterize the participant’s Twitter network. We envision this data to be available upon direct request to the PIs to investigators working in the field of HIV prevention. In order to support external collaborations, we will require interested investigators to obtain appropriate IRB approvals from their respective institutions with explicit provisions to prevent any attempt for re-identification of the participants in the study or other 3rd parties. Furthermore, we will require a written Data Use Agreement signed between the UCSD Contracts and Grants Office and the respective contracts/legal department of the partnering institution to ensure proper legal measures are in place before any data is exchanged (a model of such DUA was already developed at UCSD in partnership with the legal office of UCOP). Lastly, all data will be deidentified and encrypted prior to transmission to the collaborator. We will investigate means for watermarking the data to traceback potential leaks. At the same time, upon receipt of the data, the collaborator is fully responsible for its safety. |
| **17. POTENTIAL BENEFITS** |
| There is no guarantee that participants will directly benefit from being in this study or upon completion of the study. However, what is learned from this study may help other people who are infected with or at risk for infection with HIV. |
| **18. RISK/BENEFIT RATIO** |
| The overall risk for this project is very low. There are no health risks associated with this study, and no sensitive personal data will be published without explicit consent. In addition, participation in this study will help researchers understand social networks of MSM-at risk individuals.  All evaluation forms, reports, and other records will be identified by a coded number only to maintain subject confidentiality. All records will be kept in a locked file cabinet. All computer entry and networking programs will be done with coded numbers only. Clinical information will not be released without written permission of the subject, the sponsor, or the UCSD IRB.  With minimal, managed risks and potentially far-reaching results, the benefits associated with this project outweigh the risks  Participants will be told of any new information learned during the course of the study, which might cause then to change their minds about staying in the study. Results from Twitter network analyses will not be shared with participants. |
| **19. EXPENSE TO PARTICIPANT** |
| There is no cost to subjects for online data collection, or proposed analyses, which are part of this study. The proposed web-application is provided at no cost to participants. |
| **20. COMPENSATION FOR PARTICIPATION** |
| Subjects will receive a $5.00 Amazon gift card as compensation for their time and inconvenience. Once subject installs the PIRC-Net app, subjects will be emailed a link to redeem the online gift card. If subjects keep the app for 6 months, an additional $5.00 Amazon gift card will be emailed at that time.  Usage charges or data plan charges incurred while on Twitter will not be reimbursed nor will subjects be provided compensation for said charges while on Twitter. |
| **21. PRIVILEGES/CERTIFICATIONS/LICENSES AND RESEARCH TEAM RESPONSIBILITIES** |
| Drs. Susan Little has full physician privileges at the UCSD Medical Center and VA Medical Center. She will oversee recruitment and be involved in study analyses.  Dr. Nadir Weibel is a Research Assistant Professor in the Computer Science and Engineering Department and will be responsible for the technical development and analysis of the study.  Dr. Nella Green is a postdoctoral fellow in the Department of Medicine. She will be involved in all Phases of research.  Helene Le, CPhT and Joseph Lencioni, MABMH will serve as the regulatory contacts for this study.  Melissa Davis will serve as the administrative contact for this study.  The PI, co-PIs, and staff at the AVRC have completed the required UCSD research training to include CITI Human Subjects and GCP training along with the UCSD IRB HIPAA tutorial. |
| **22. BIBLIOGRAPHY** |
| [1] WHO website on HIV/AIDS. <http://www.who.int/gho/hiv/en/>  [2] Sean D. Young, Caitlin Rivers, Bryan Lewis, Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes, Preventive Medicine, Volume 63, June 2014, Pages 112-115.  [3] Little, Susan J., Sergei L. Kosakovsky Pond, Christy M. Anderson, Jason A. Young, Joel O. Wertheim, Sanjay R. Mehta, Susanne May, and Davey M. Smith. "Using HIV Networks to Inform Real Time Prevention Interventions." PloS one 9, no. 6 (2014): e98443.  [4] CDC website. <http://www.cdc.gov/hiv/risk/gender/msm/facts/index.html>  [5] Phillips, A. N., Cambiano, V., Nakagawa, F., Brown, A. E., Lampe, F., Rodger, A., Delpech, V. C. (2013). Increased HIV incidence in men who have sex with men despite high levels of ART-induced viral suppression: analysis of an extensively documented epidemic. *PLoS One, 8*(2), e55312. doi: 10.1371/journal.pone.0055312  [6] Christakis NA, Fowler JH. The Spread of Obesity in a Large Social Network Over 32 Years, *New England Journal of Medicine* 357 (4): 370–379 (26 July 2007)  [7] Christakis NA, Fowler JH. The Collective Dynamics of Smoking in a Large Social Network, *New England Journal of Medicine* 358 (21): 2249–58 (22 May 2008)  [8] Mednick SC, Fowler JH, Christakis NA. The Spread of Sleep Loss Influences Drug Use in Adolescent Social Networks, *PLoS ONE* 5 (3): e9775 (March 2010)  [9] Rosenquist JN, Murabito J, Fowler JH, Christakis NA. The Spread of Alcohol Consumption Behavior in a Large Social Network, *Annals of Internal Medicine* 152 (7): 426–433 (6 April 2010)  [10] Mednick SC, Fowler JH, Christakis NA. The Spread of Sleep Loss Influences Drug Use in Adolescent Social Networks, *PLoS ONE* 5 (3): e9775 (March 2010)  [11] Fowler JH, Christakis NA. The Dynamic Spread of Happiness in a Large Social Network, *British Medical Journal* 337: a2338 (4 December 2008)  [12] Cacioppo JT, Fowler JH, Christakis NA. Alone in the Crowd: The Structure and Spread of Loneliness in a Large Social Network, *Journal of Personality and Social Psychology* 96 (6): 977–991 (December 2009)  [13] Rosenquist JN, Fowler JH, Christakis NA. Social Network Determinants of Depression, *Molecular Psychiatry* 16 (3): 273–281 (March 2011) |
| **23. FUNDING SUPPORT FOR THIS STUDY** |
| This trial is supported by the grant—award number 1 R24 AI106039-01. Parts of this study are also supported by a UCSD CFAR developmental grant (Phase I, II III) and the UCSD’s Frontiers of Innovation Scholar Program (graduate students support).  The financial contacts for this study are Fernando Mares (619-543-8178) at the AVRC and David Bareno (858-534-3569) in CSE. |
| **24. BIOLOGICAL MATERIALS TRANSFER AGREEMENT** |
| Not Applicable |
| **25. INVESTIGATIONAL DRUG FACT SHEET AND IND/IDE HOLDER** |
| Not Applicable |
| **26. IMPACT ON STAFF** |
| The study personnel assigned to this study are funded through extramural funding. |
| **27. CONFLICT OF INTEREST** |
| No conflicts of interest to report. |
| **28. SUPPLEMENTAL INSTRUCTIONS FOR CANCER-RELATED STUDIES** |
| Not Applicable |
| **29. OTHER APPROVALS/REGULATED MATERIALS** |
| Not Applicable |
| **30. PROCEDURES FOR SURROGATE CONSENT AND/OR DECISIONAL CAPACITY ASSESSMENT** |
| Not applicable.  In order for a subject to be eligible for this study they must not be cognitively impaired and must be able to communicate effectively with the study staff; therefore, the subjects enrolling/participating in this study will have the ability to:  1. Understanding, i.e., the ability to comprehend the disclosed information about the nature and purpose of the study, the procedures involved, as well as the risks and benefits of participating versus not participating;  2. Appreciation, i.e., the ability to appreciate the significance of the disclosed information and the potential risks and benefits for their own situation and condition;  3. Reasoning, i.e., the ability to engage in a reasoning process about the risks and benefits of participating versus alternative, and  4. The ability to express a choice about whether or not to participate.  If for any reason, the study staff finds that the subject does not understand, appreciate, have reasoning ability and/or cannot express his/her choice to participant in the study, the subject will not be enrolled and provided with the options that may be available to them. |