Social Network Analysis and Artificial Intelligence:

Methodological Partners in the Study of HIV Prevention and Risk Online

* Research supported by an NIH K99/R00 Career Development Award (K99 HD094648)

Lindsay E. Young, Ph.D.
NIH K99 Postdoctoral Scholar
University of Chicago, Department of Medicine
Chicago Center for HIV Elimination

THE UNIVERSITY OF CHICAGO MEDICINE

CCHE The Chicago Center for HIV Elimination

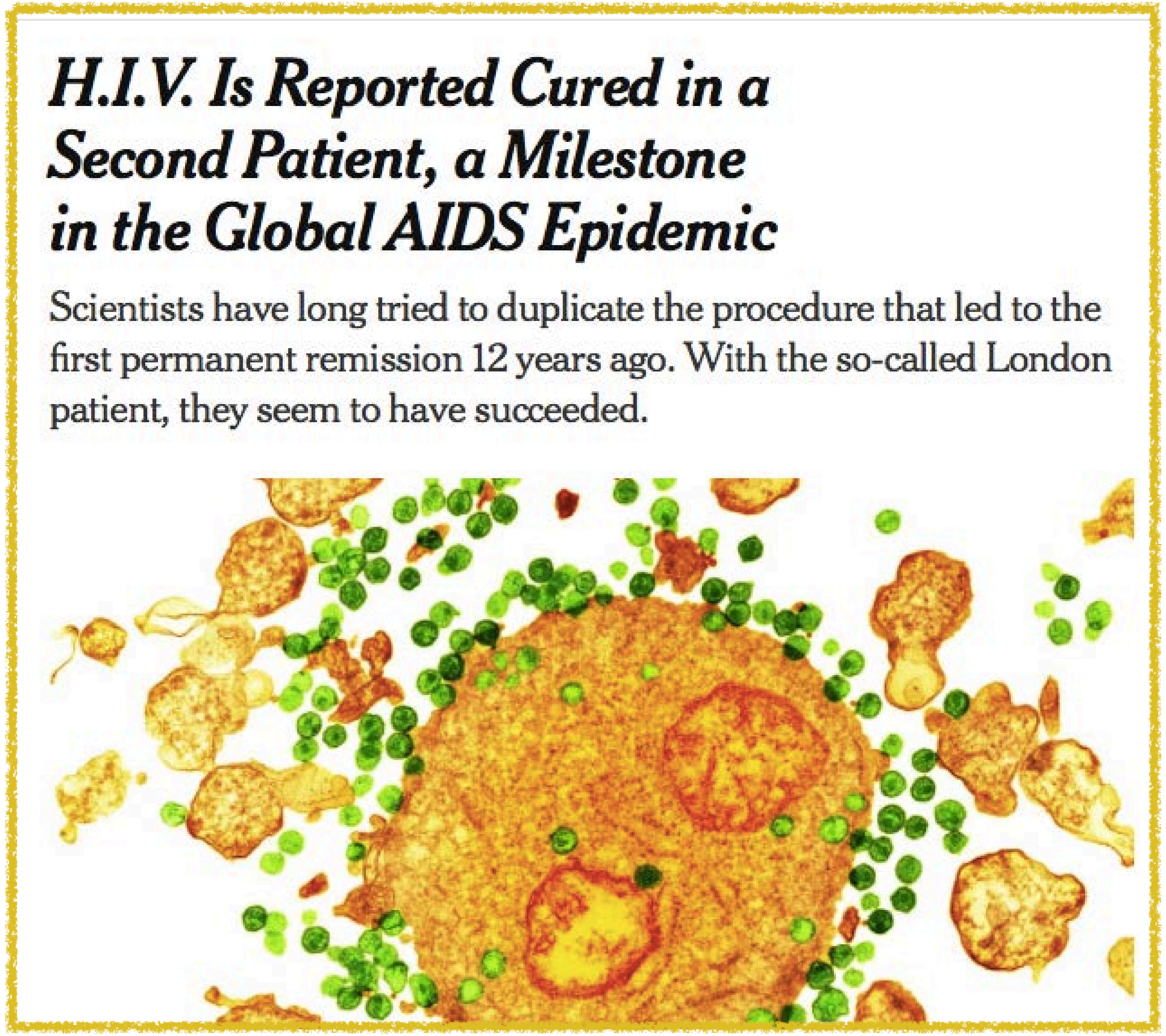
Presentation Overview

- The HIV problem space
- The social network perspective on HIV risk
- Social media: A network data repository
- The methodological toolkit
- Outstanding questions and future directions

The Problem Space

The disproportionate burden of HIV in the United States

On the road to a cure, …


H.I.V. Is Reported Cured in a Second Patient, a Milestone in the Global AIDS Epidemic

Scientists have long tried to duplicate the procedure that led to the first permanent remission 12 years ago. With the so-called London patient, they seem to have succeeded.

…we still have much to address.

HIV incidence in the United States

130,400

- Number of new HIV infections diagnosed in 1984.

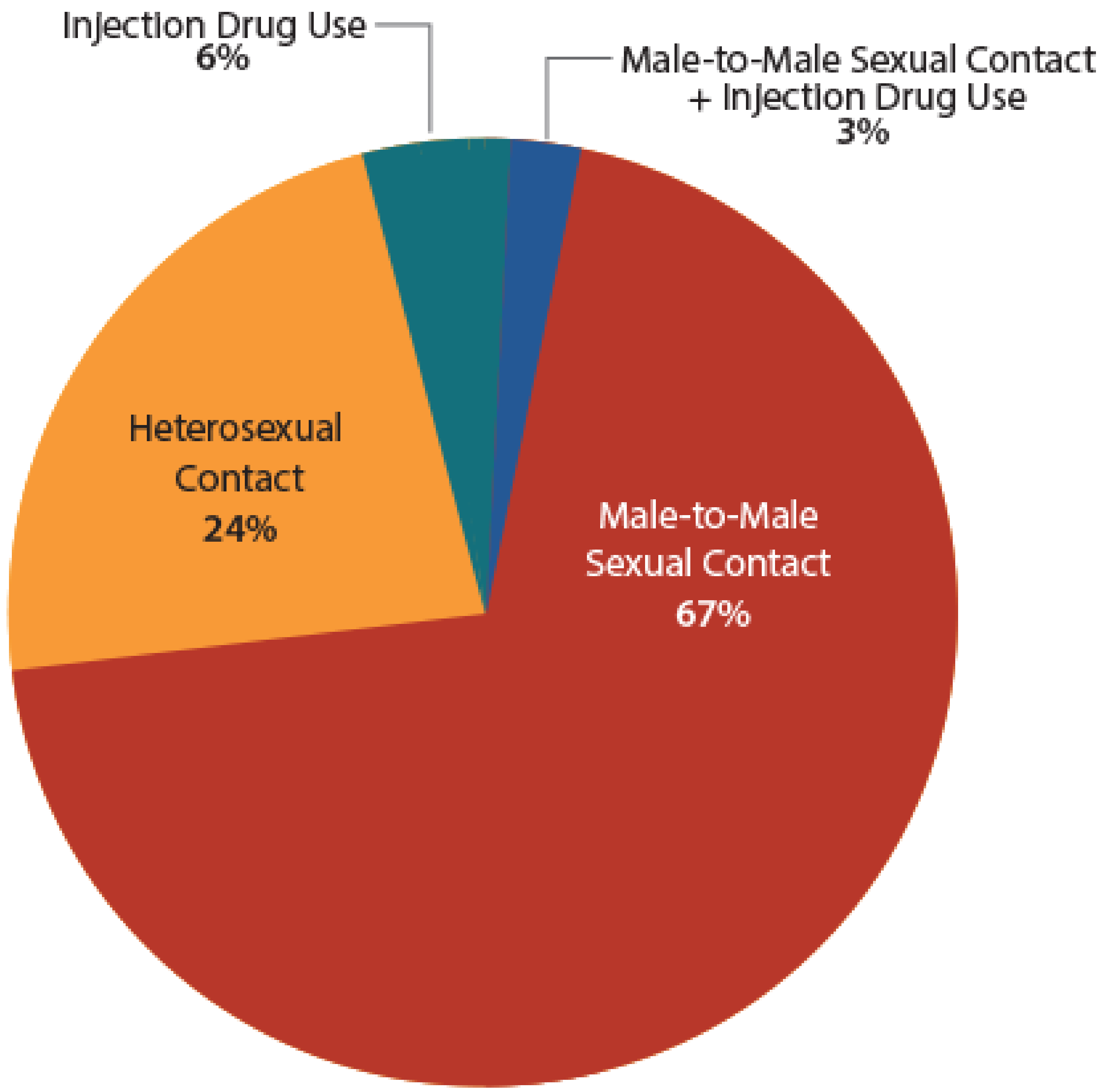
70% reduction

38,739

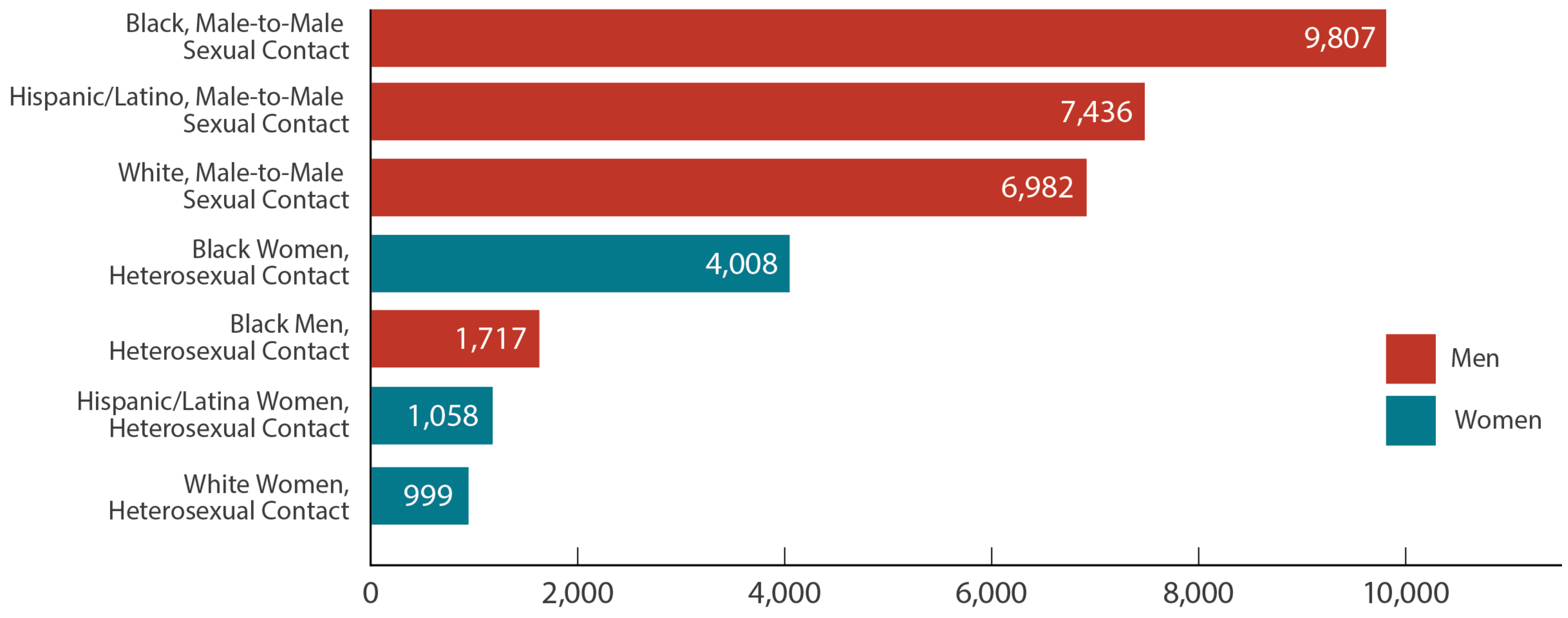
- Number of new HIV infections diagnosed in 2017.

HIV burdens are not proportionate

New HIV Diagnoses by Transmission Category, 2015 (n=39,513)

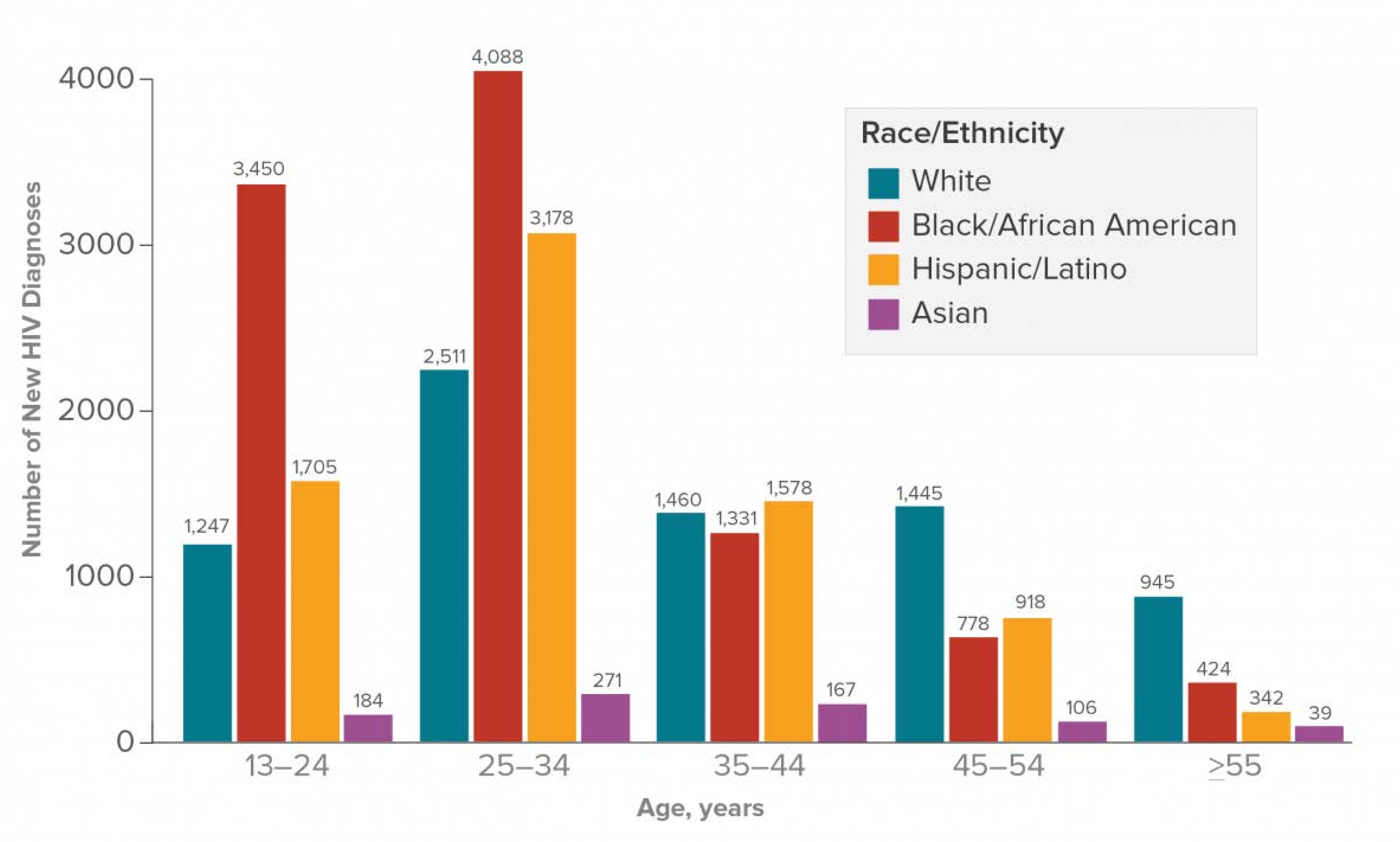


New HIV Diagnoses in the US and Dependent Areas for the Most-Affected Subpopulations, 2017



HIV burdens are not proportionate

New HIV Diagnoses among MSM, by age and race, 2017



- Among HIV diagnosed BMSM, 39% were aged 13-24; 36% were aged 25-34.
- From 2010-2016, HIV diagnoses among YBMSM aged 25-34 increased 40%.

Mechanisms behind disparities are not well understood

- Black MSM report less substance use and fewer partners than White MSM.

- No significant differences by race in reported condomless sex, commercial sex work, or sex with a known HIV+ partner.

Explaining disparities in HIV infection among black and white men who have sex with men: a meta-analysis of HIV risk behaviors

Gregorio A. Millett[a], Stephen A. Flores[a], John L. Peterson[b] and Roger Bakeman[b]

AIDS 2007, **21**:2083–2091

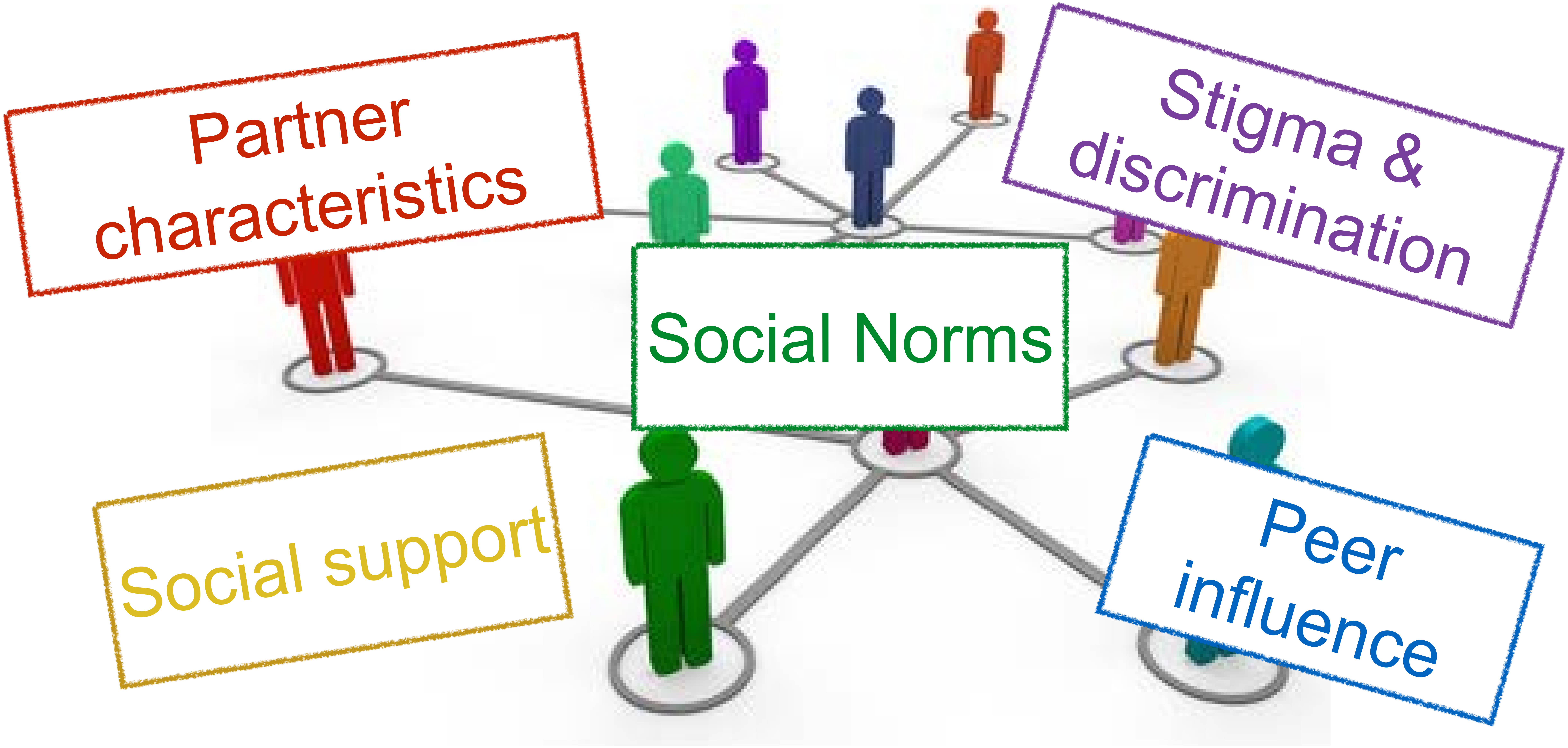
Comparisons of disparities and risks of HIV infection in black and other men who have sex with men in Canada, UK, and USA: a meta-analysis

Gregorio A Millett, John L Peterson, Stephen A Flores, Trevor A Hart, William L Jeffries 4th, Patrick A Wilson, Sean B Rourke, Charles M Heilig, Jonathan Elford, Kevin A Fenton, Robert S Remis

Lancet 2012, 380(983): 341-348

Individual risk behaviors don't adequately explain differences in HIV incidence between White and Black MSM.

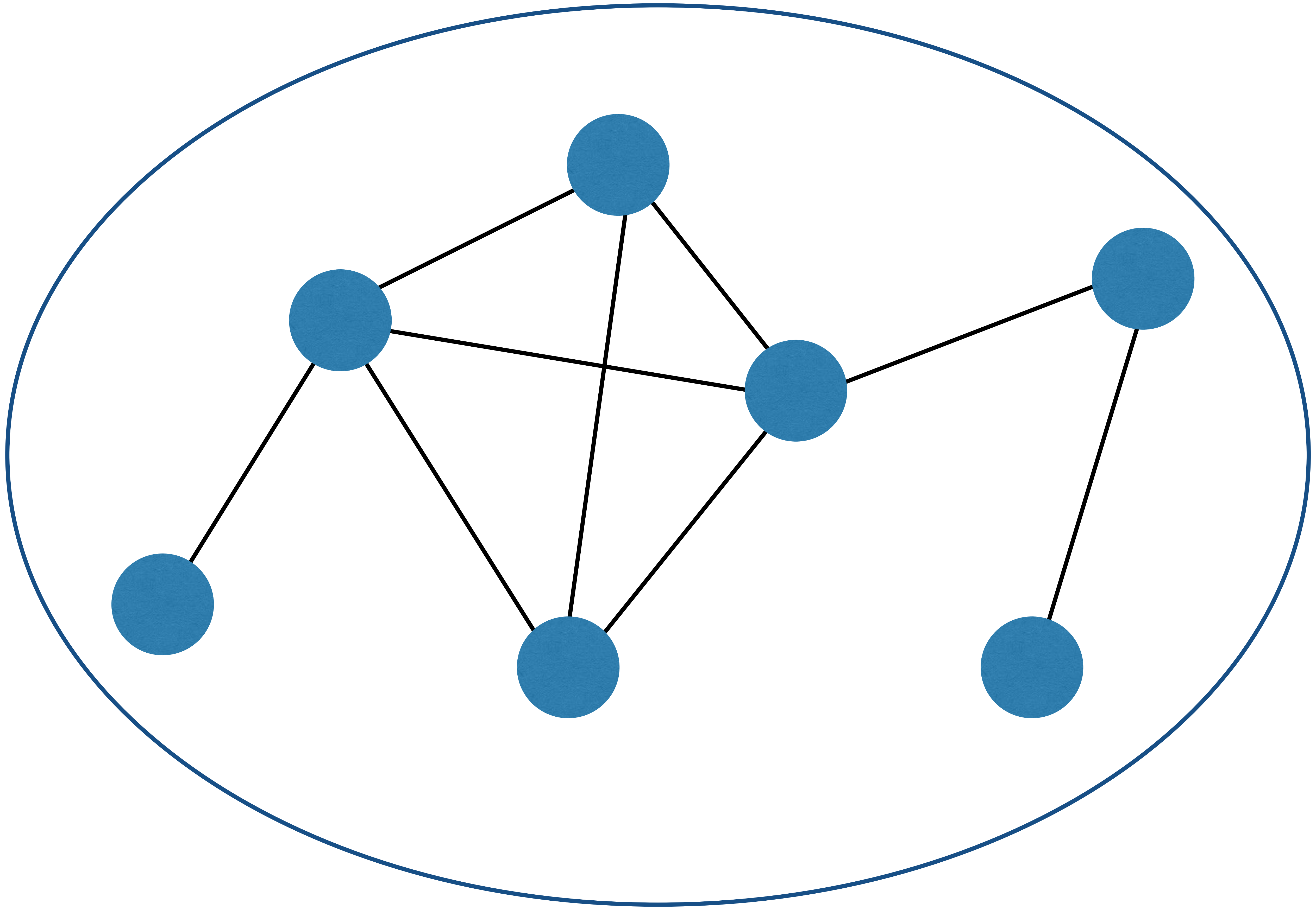
Alternative Explanations



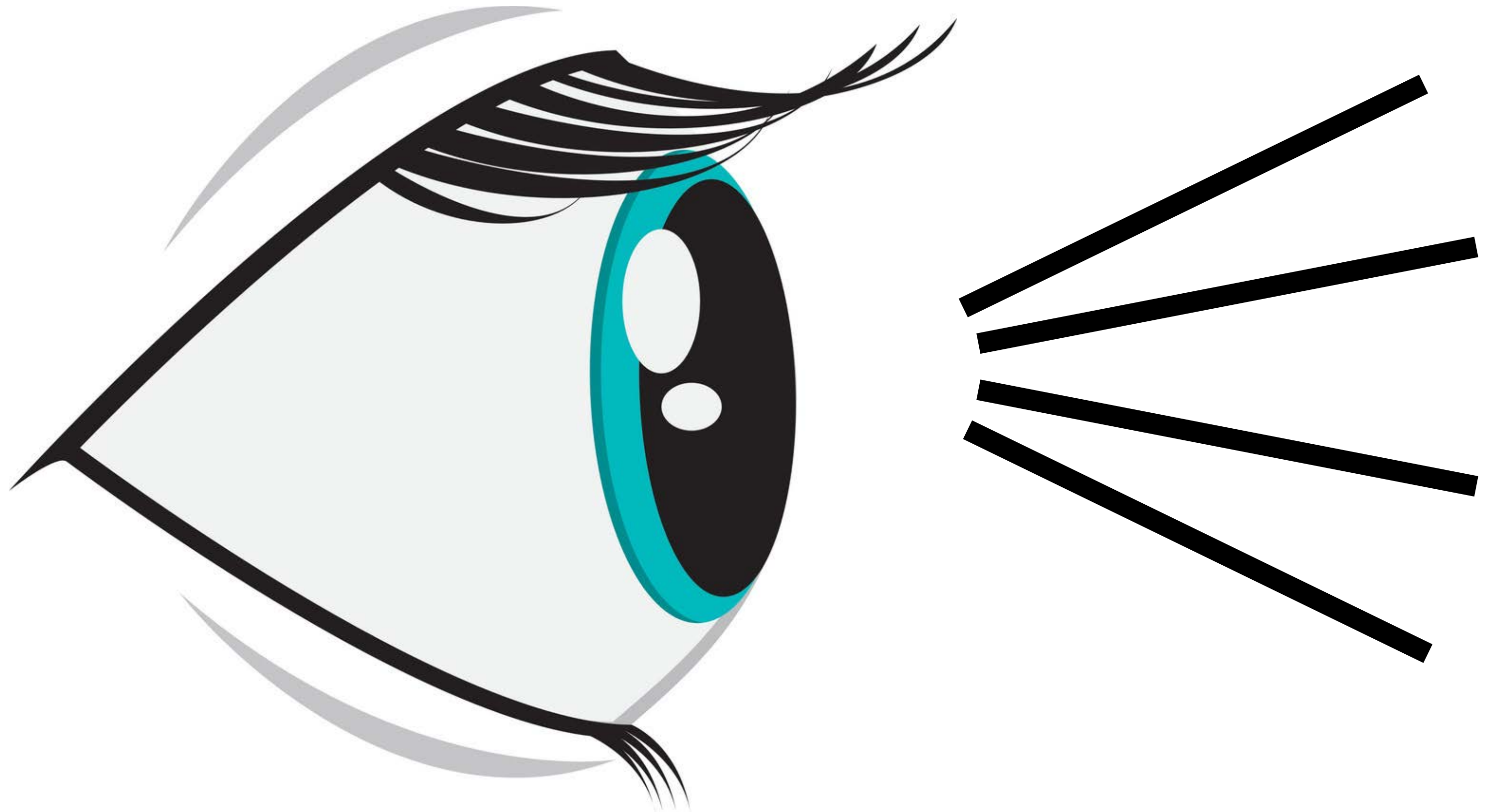
THE SOCIAL NETWORK PERSPECTIVE

Essential Components of a Network

- Nodes = Actors = Vertices
- Ties = Edges = Links
  - Directed (e.g., email)
  - Undirected (e.g., collaboration)
- Network (or graph) structure

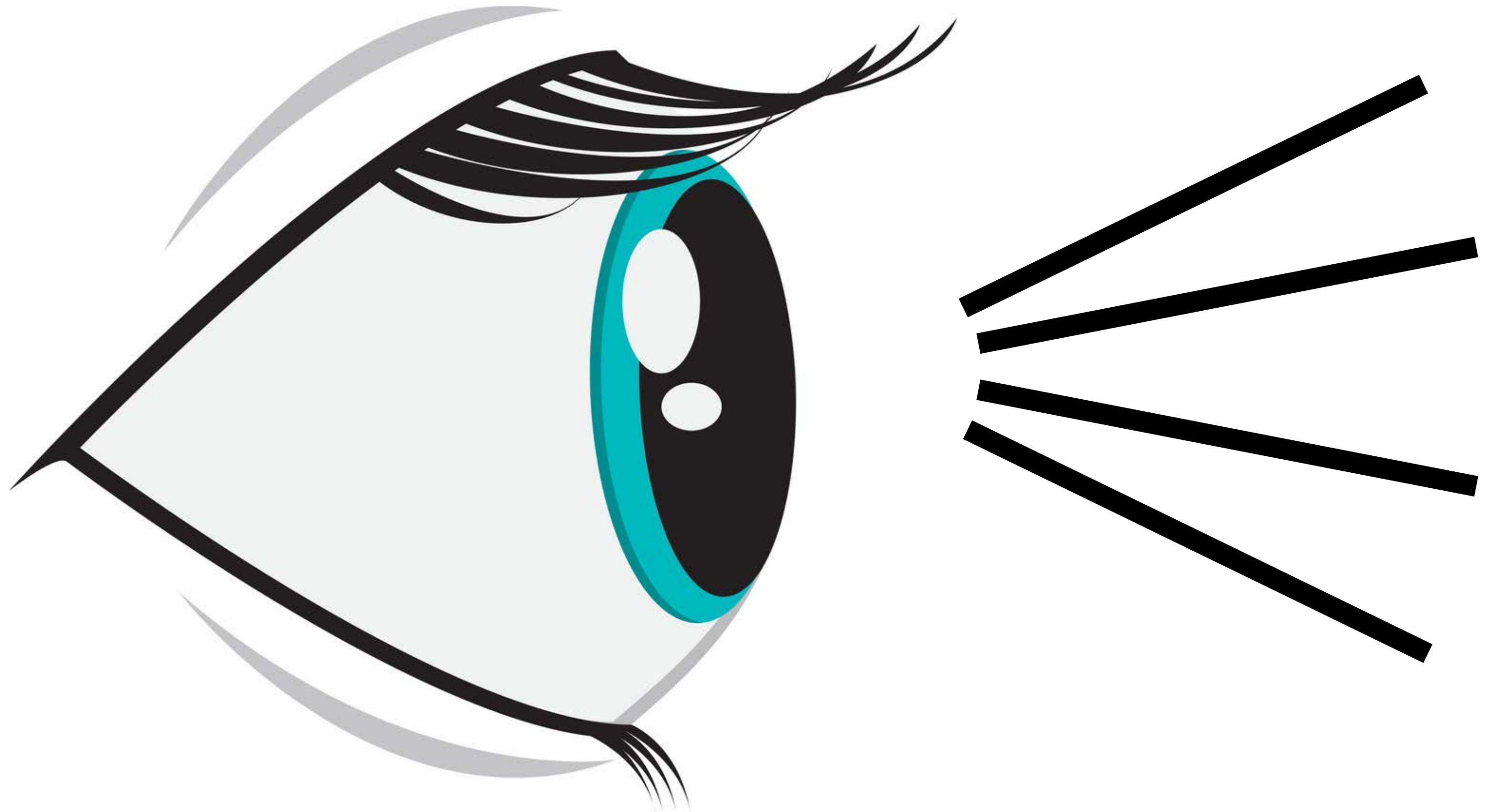


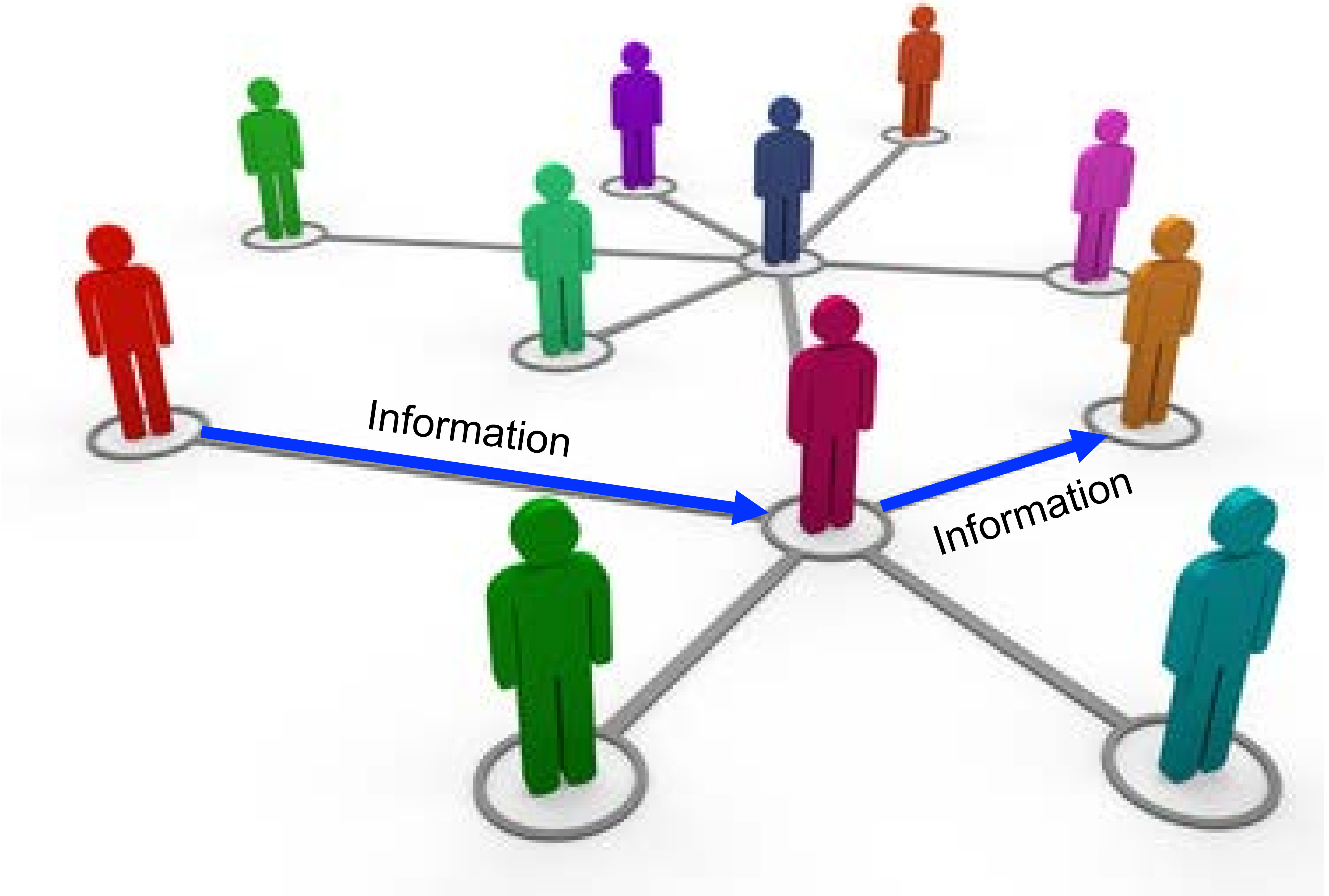
A network perspective on HIV prevention and risk assumes…



1. Actors and their actions are viewed as interdependent rather than independent units.

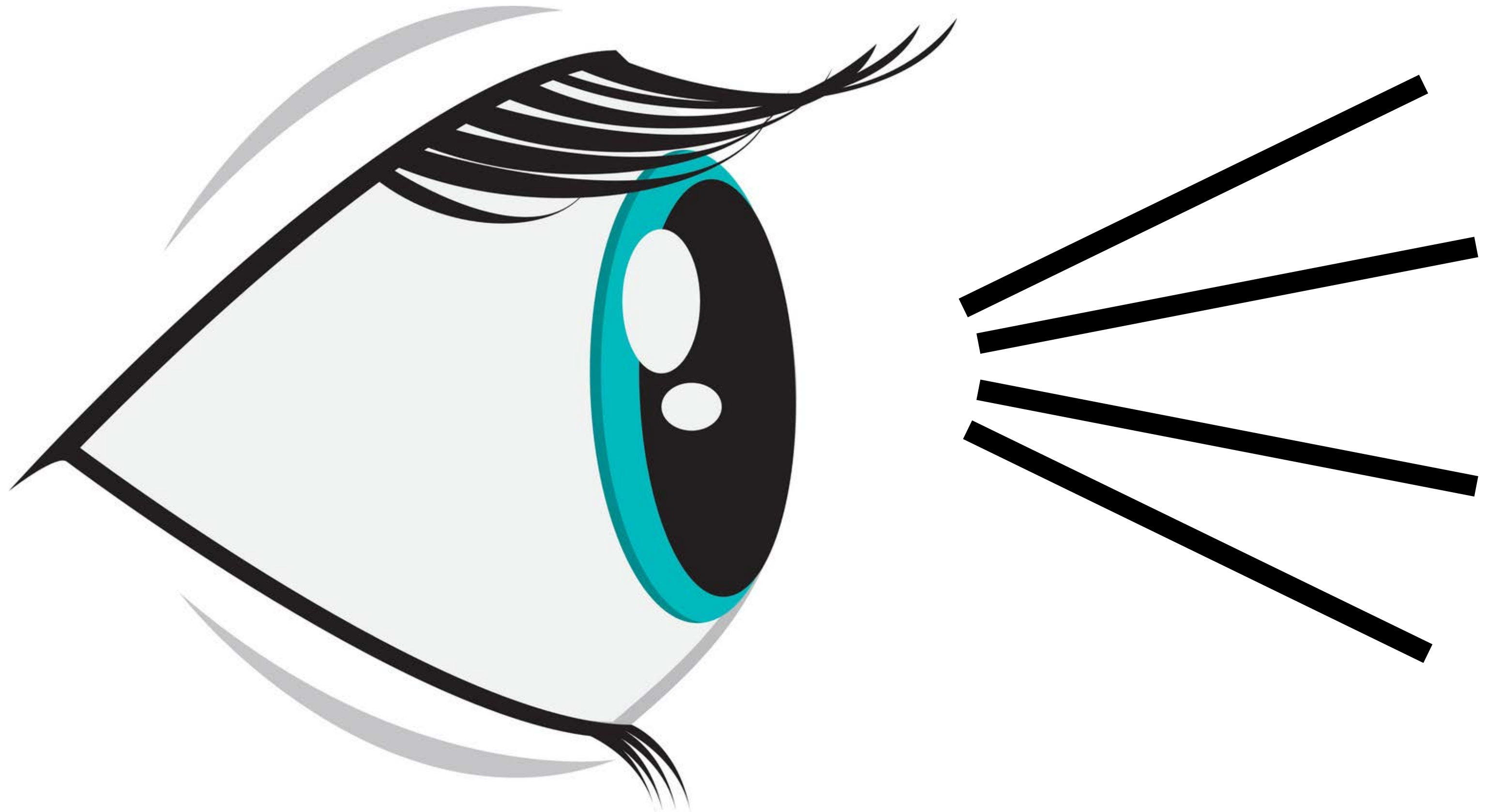
A network perspective on HIV prevention and risk assumes…

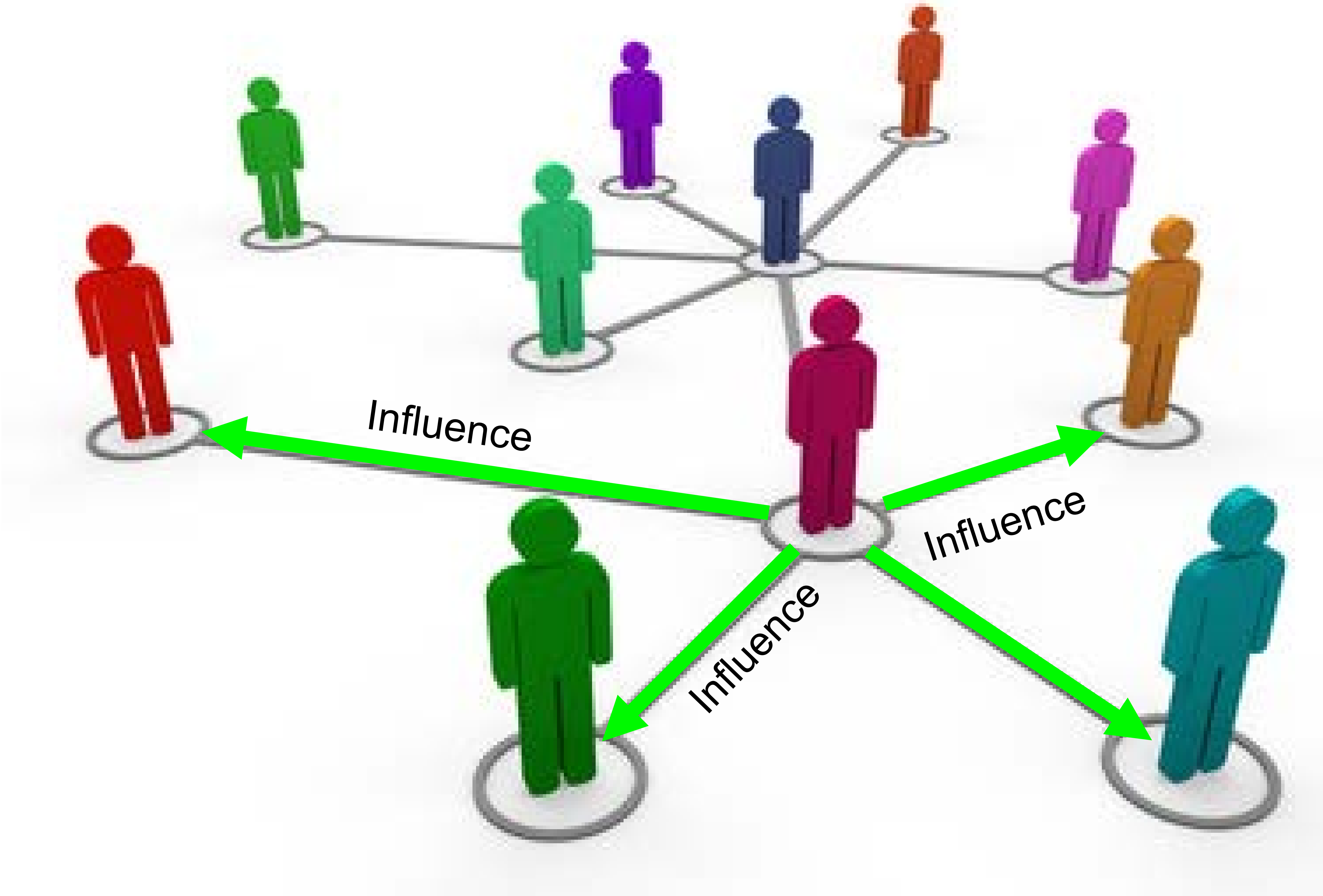




2. Ties between actors are channels through which resources flow.

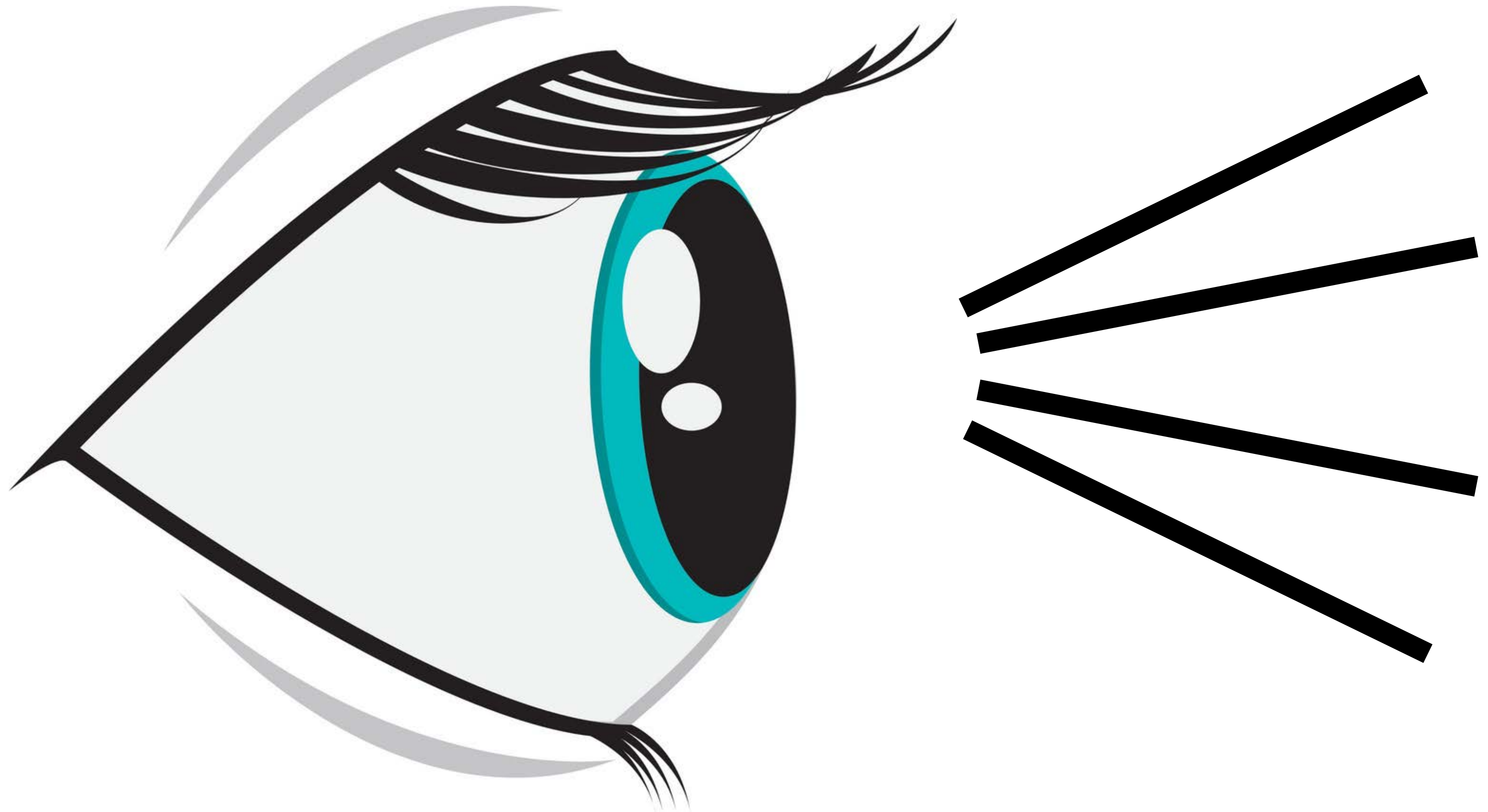
A network perspective on HIV prevention and risk assumes…





2. Ties between actors are channels through which resources flow.

A network perspective on HIV prevention and risk assumes…



3. Network structure provides opportunities for or constraints on individual actions.

A network perspective on HIV risk allows us to ask…

How are the…

- people you know and interact with…
  - things that you talk about…
    - attributes of your networks…

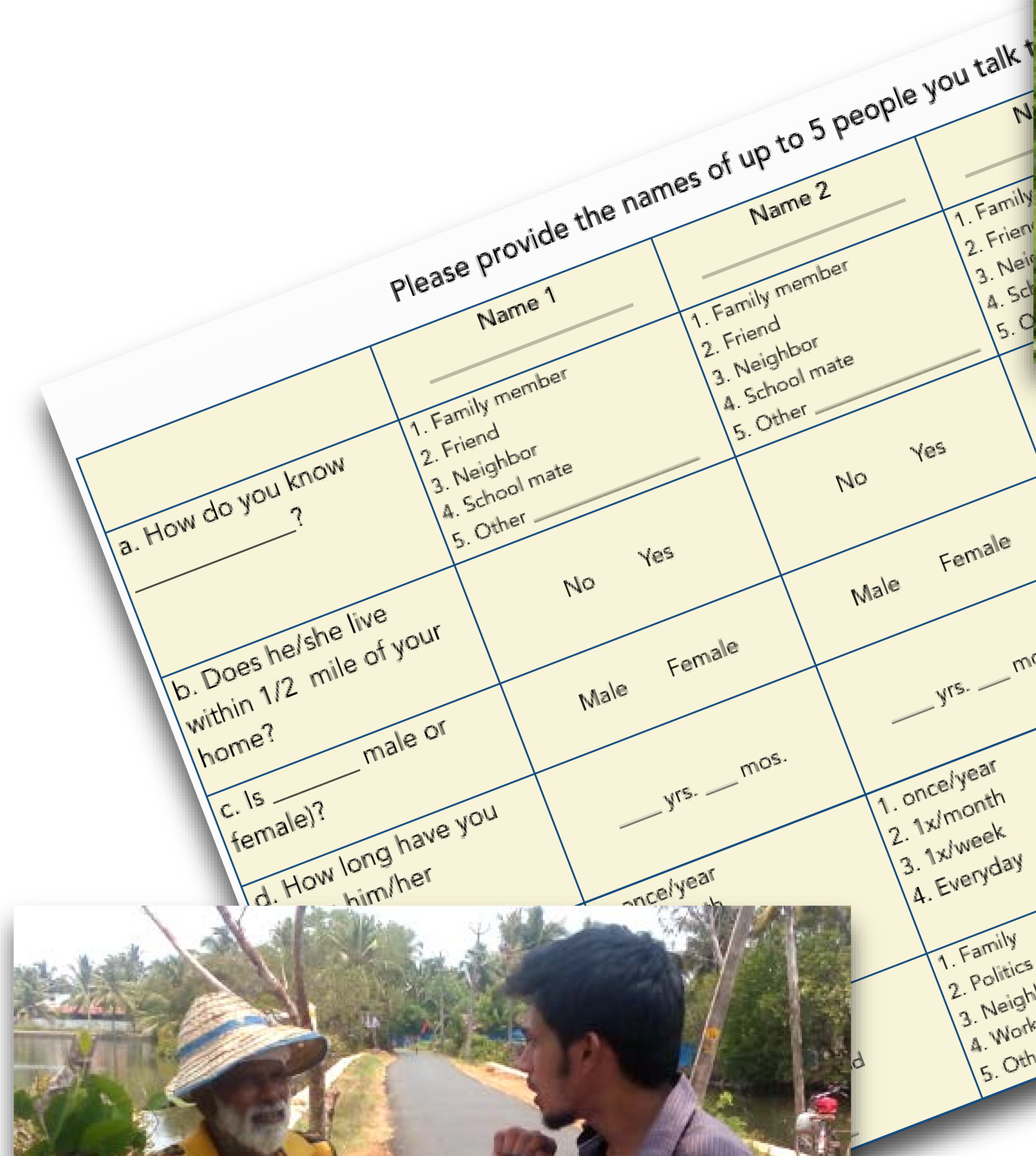
…related to your HIV-related knowledge, attitudes, and behaviors?

Social Media

A network data repository

Challenges of capturing networks “in the field”

- Resource intensive
- Participant fatigue
- Reporting biases
- Missing data



Social media present opportunities for behavioral research and intervention among young MSM

- Young LGBTQ adults have been found to use social media more than their heterosexual counterparts (Taylor, 2013; Harris, 2007).
- Social media allows them to explore their sexual and gender identities and to find and build community, which can be harder to do "offline"

But which platform is best?

But which platform is best?

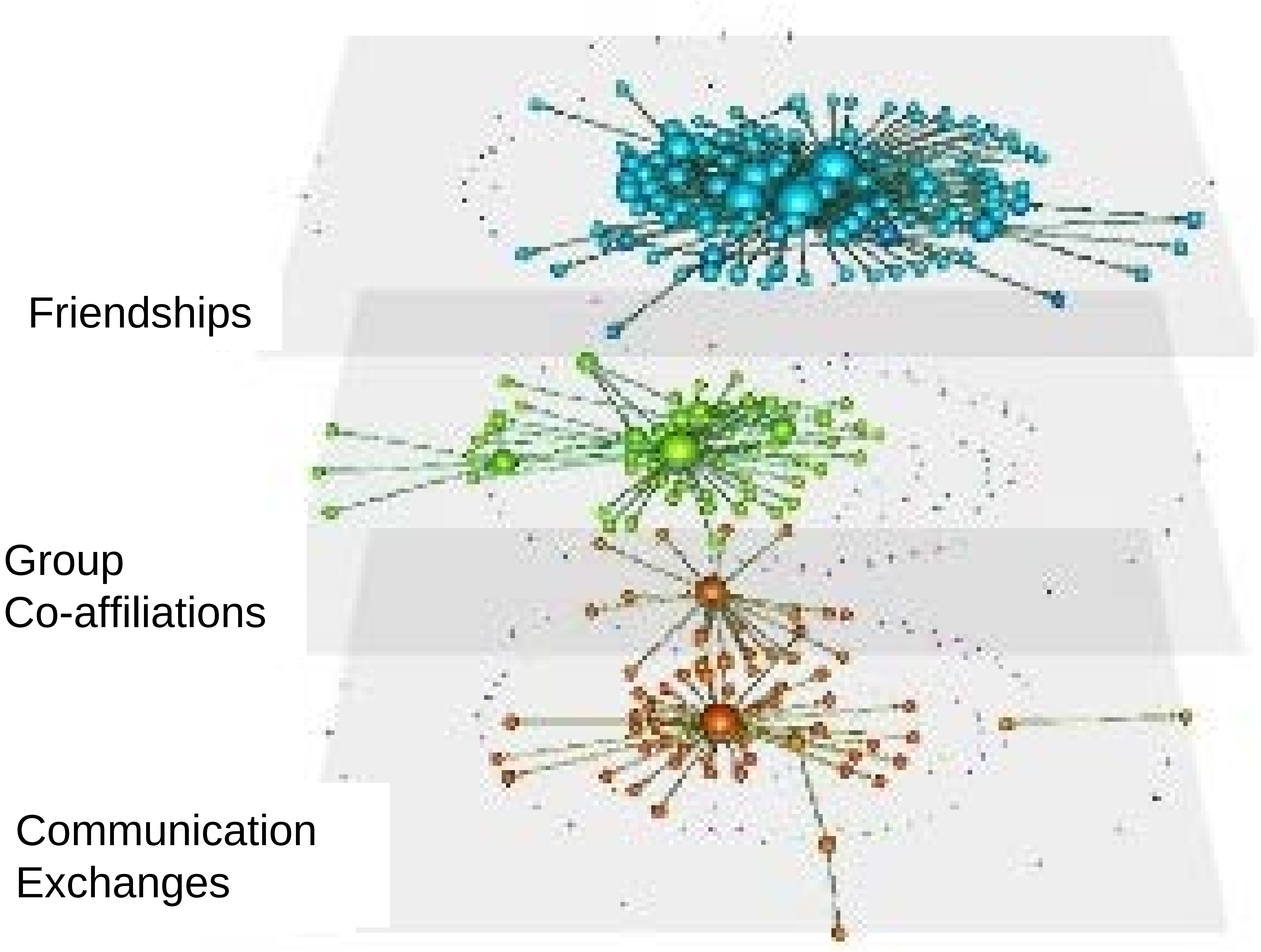
Considerations:

- Capture: Does a majority of the targeted population use the platform with enough frequency?
- Purpose: What types of content are generated and shared among users?
- Feasibility: How easy/hard is it to obtain the data you need?

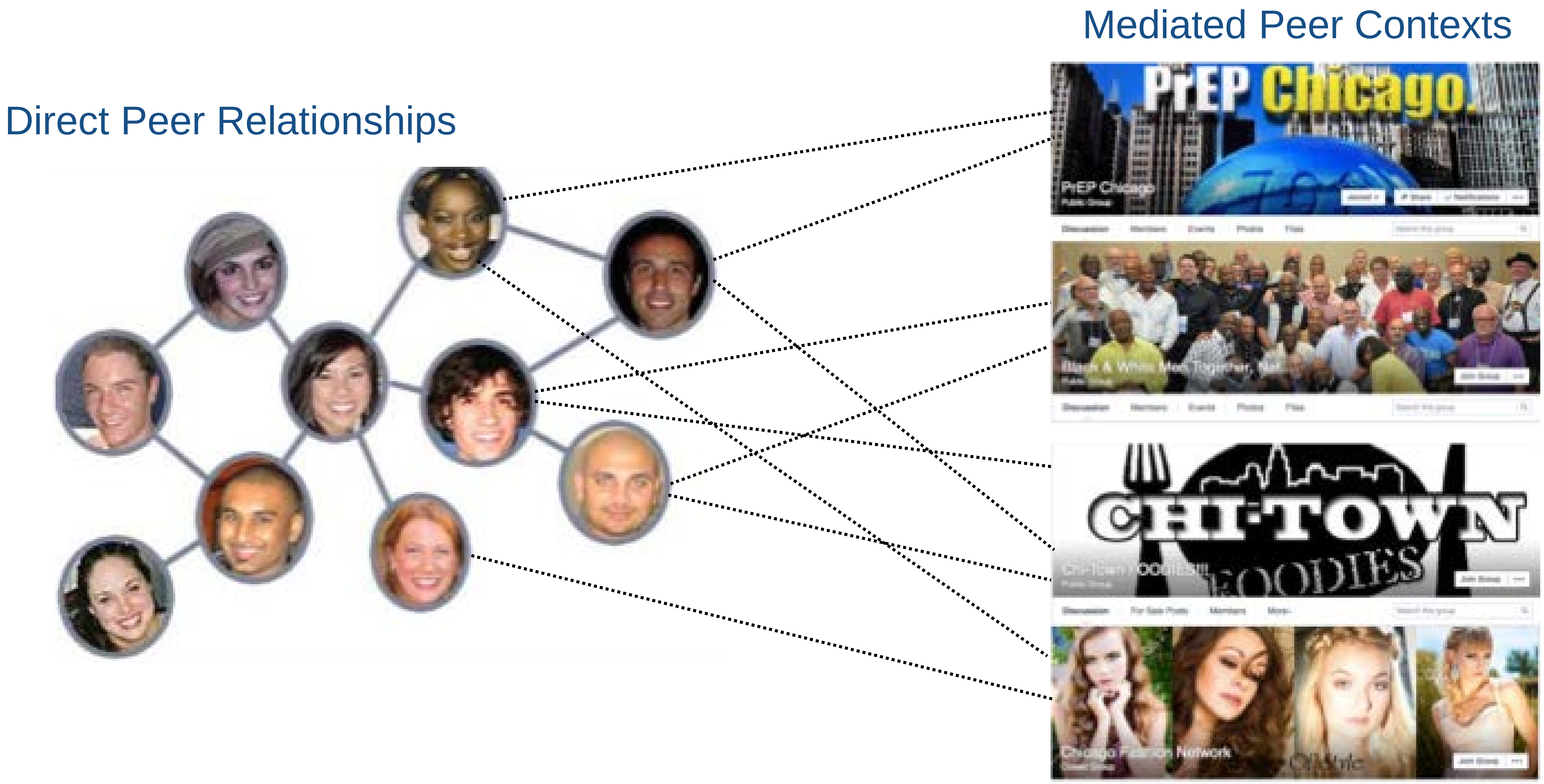
Facebook: A multidimensional social space

- As a social network, Facebook can be examined as observed relationships across multiple dimensions (or layers):

  - Person-to-person friendships
  - Person-to-group affiliations
  - Person-to-person (or person-to-public) communication

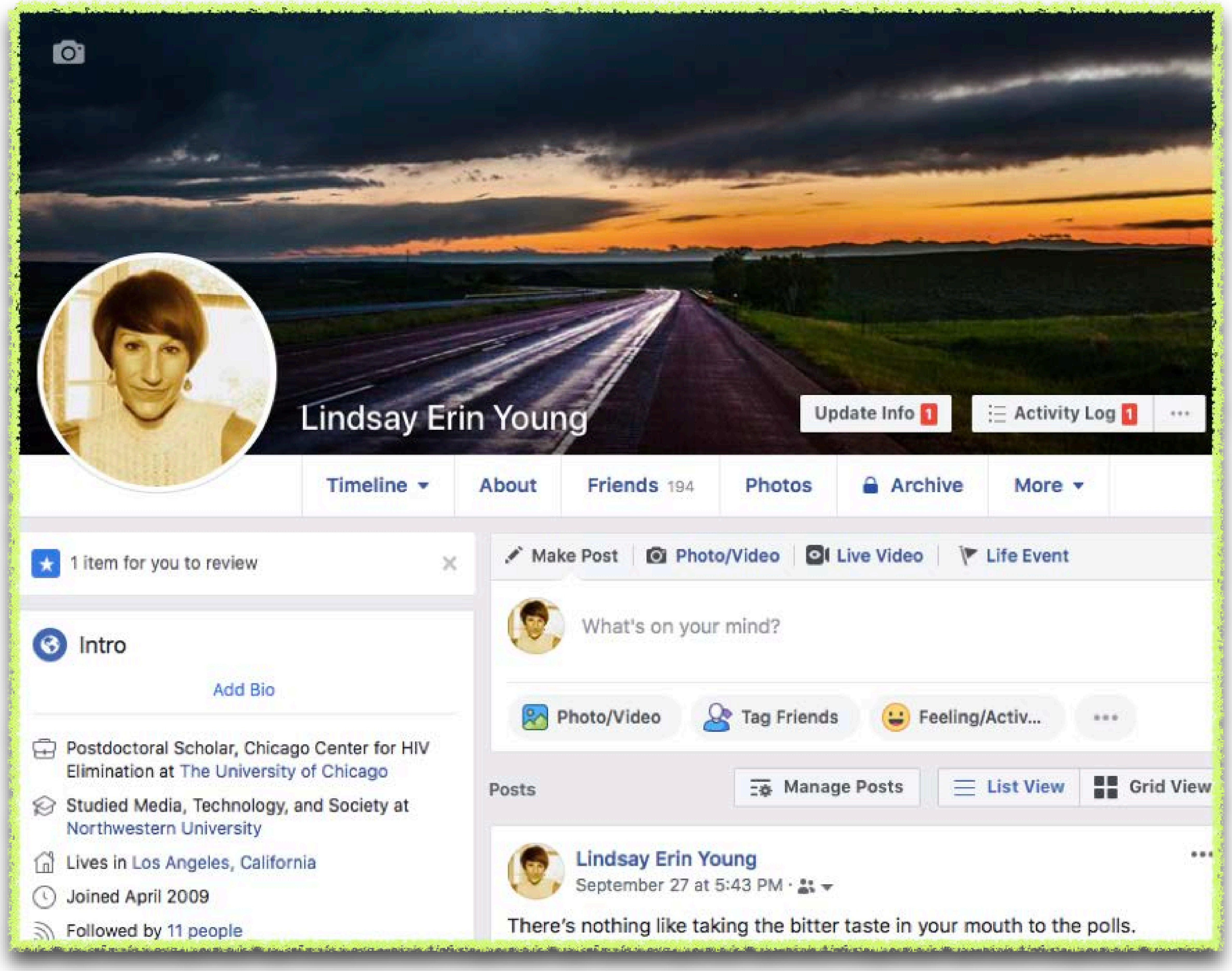


The structural dimension of Facebook



The communicative dimension of Facebook

- What one talks about in these contexts reveals clues about personal and shared interests, norms, and propensity to engage in HIV prevention and risk behaviors.



The Computational Toolkit

Research Illustrations

Specific Aims

1. Examine the relationship between observed Facebook friendship and group affiliation ties and HIV prevention/care engagement.
2. Characterize the semantic features of Facebook communication and explore their associations with HIV prevention/care engagement.
3. Predict HIV prevention/care engagement as a function of Facebook engagement patterns.

Data Sources

Study Population:

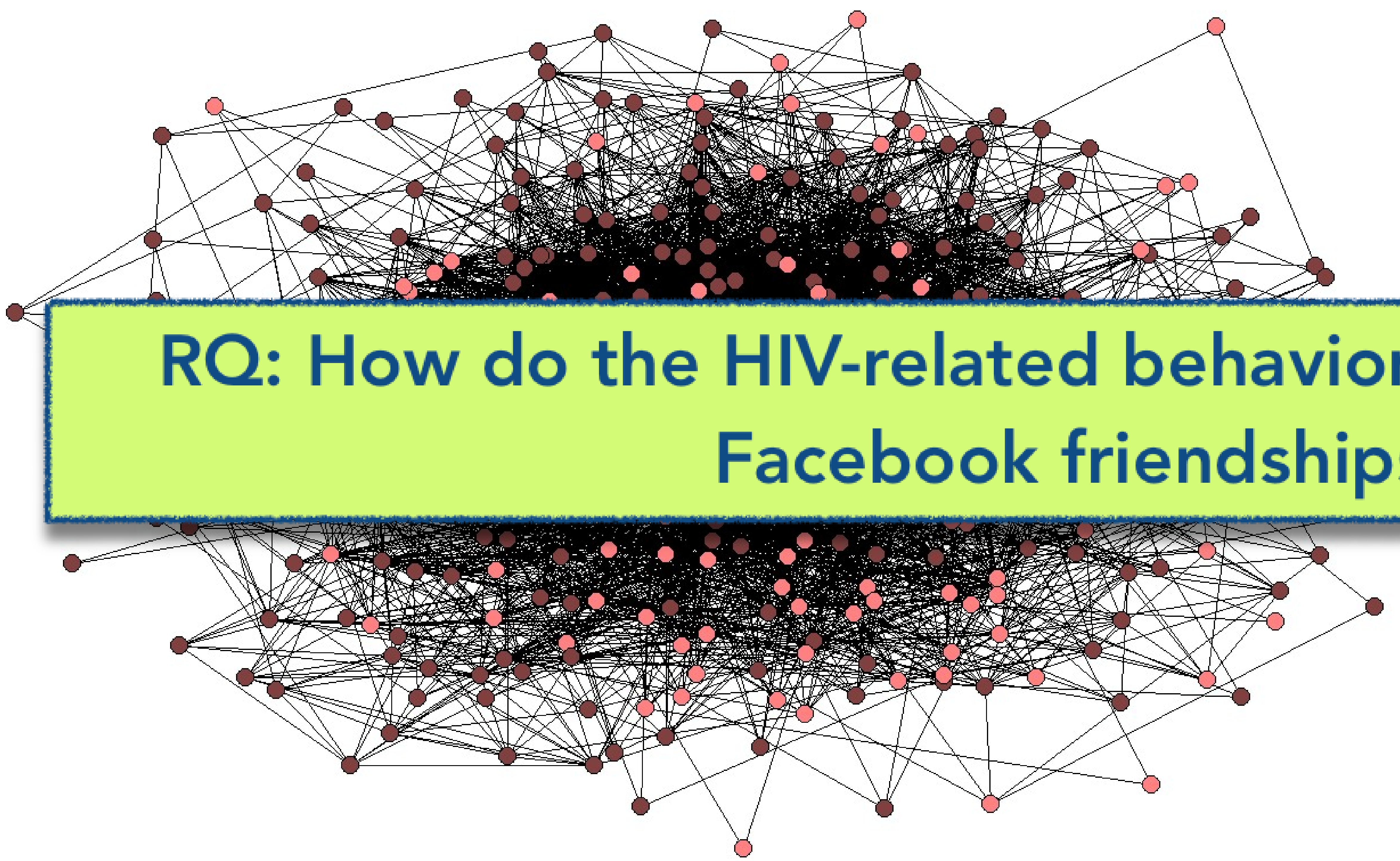
- 423 YBMSM (aged 18-35) living in Chicago

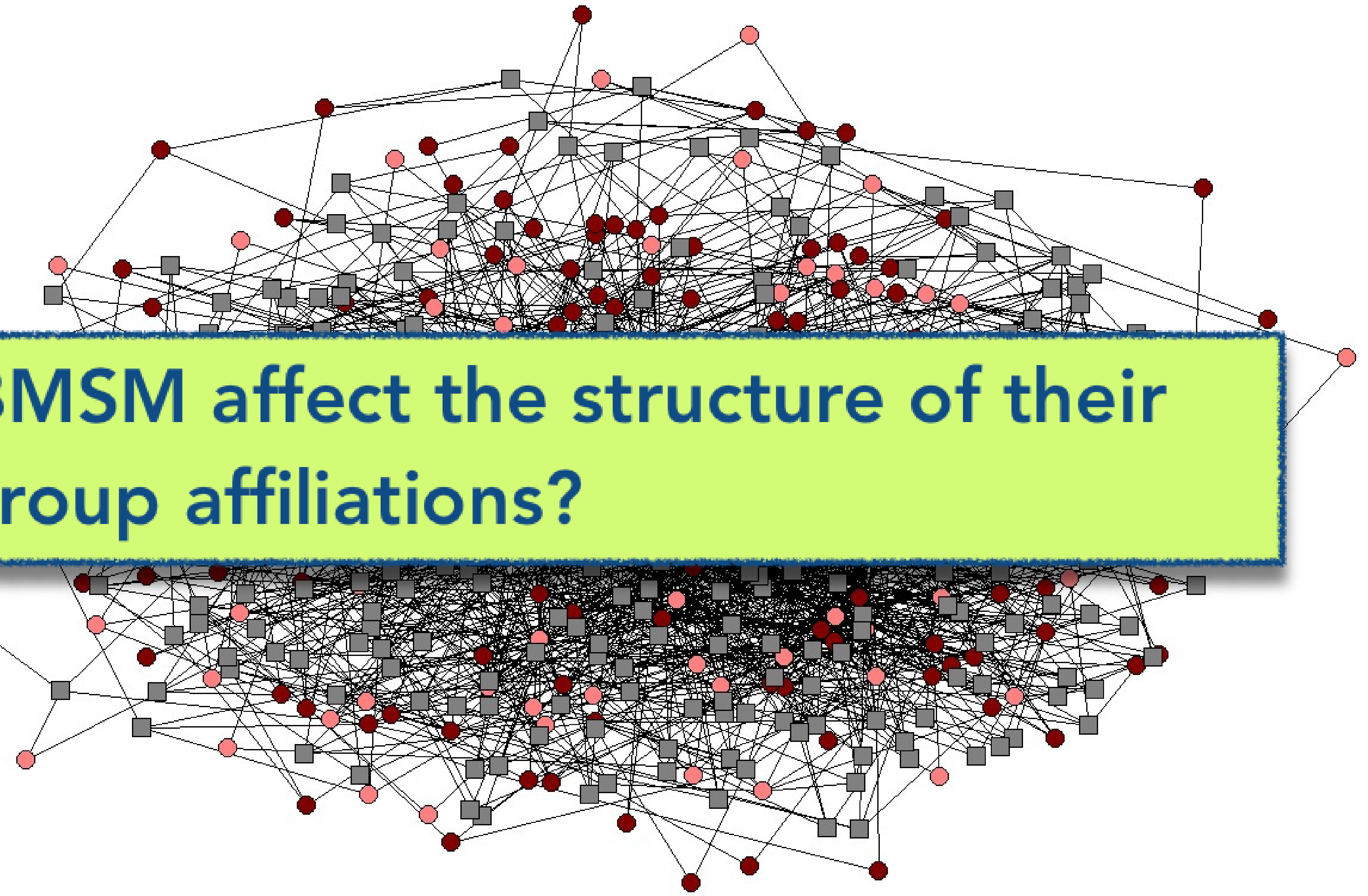
Available Data:

- *Facebook networks and communication content*
  - Facebook friendships
  - Facebook group affiliations
  - Facebook timeline posts
- *Prevention and care engagement behaviors*
  - retention in care (HIV, PrEP, or Primary)
  - recent STI/HIV testing
- *HIV-related risk factors*
  - sexual risk behaviors (e.g., condomless sex, sex drug use, exchange sex)
  - structural vulnerabilities (e.g., housing instability, criminal justice involvement)

Aim 1

Facebook network structure & HIV-related behaviors





RQ: How do the HIV-related behaviors of YBMSM affect the structure of their Facebook friendships and group affiliations?

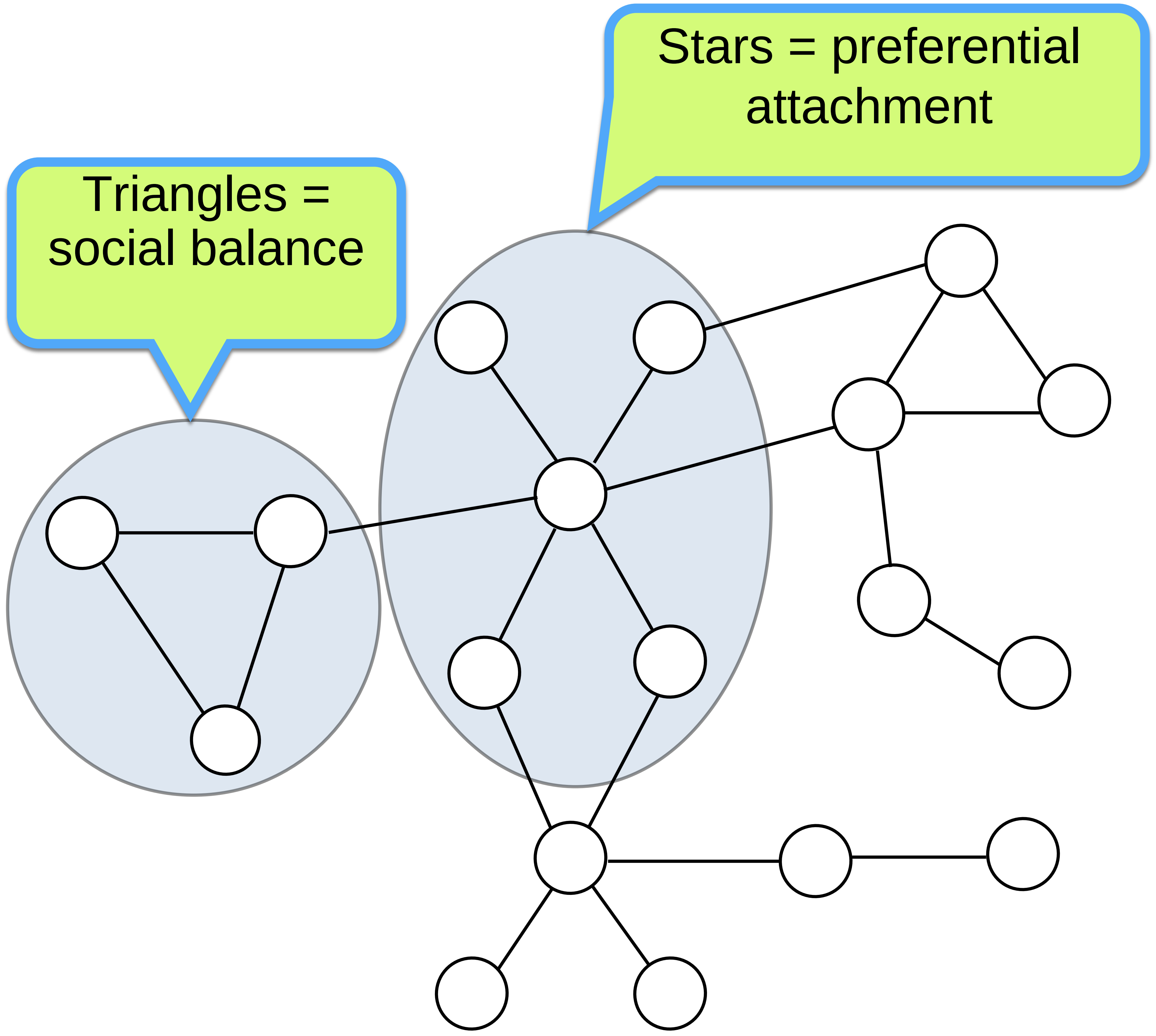
Facebook friendship network

Facebook group affiliation network

Aim 1

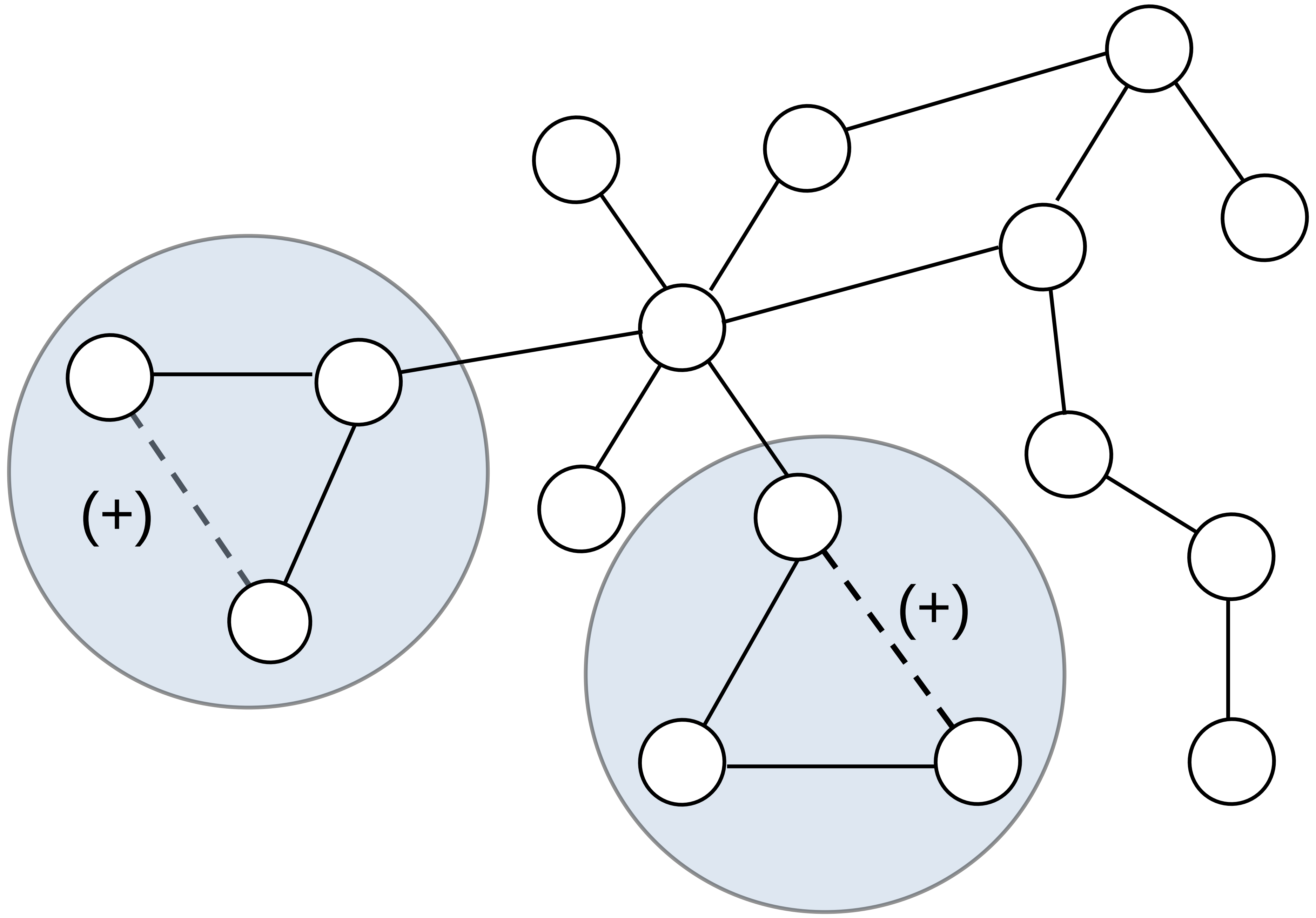
Exponential random graph models (ERGMs)

- EGRMs are statistical models for network structure that help us understand how and why network ties emerge.
- They model the likelihood of local configurations (or network motifs) — i.e., they determine whether these configurations occur more (or less) often than would be expected by chance alone.
- Local configurations represent distinct social processes, for example social balance or preferential attachment.



Exponential random graph models (ERGMs)

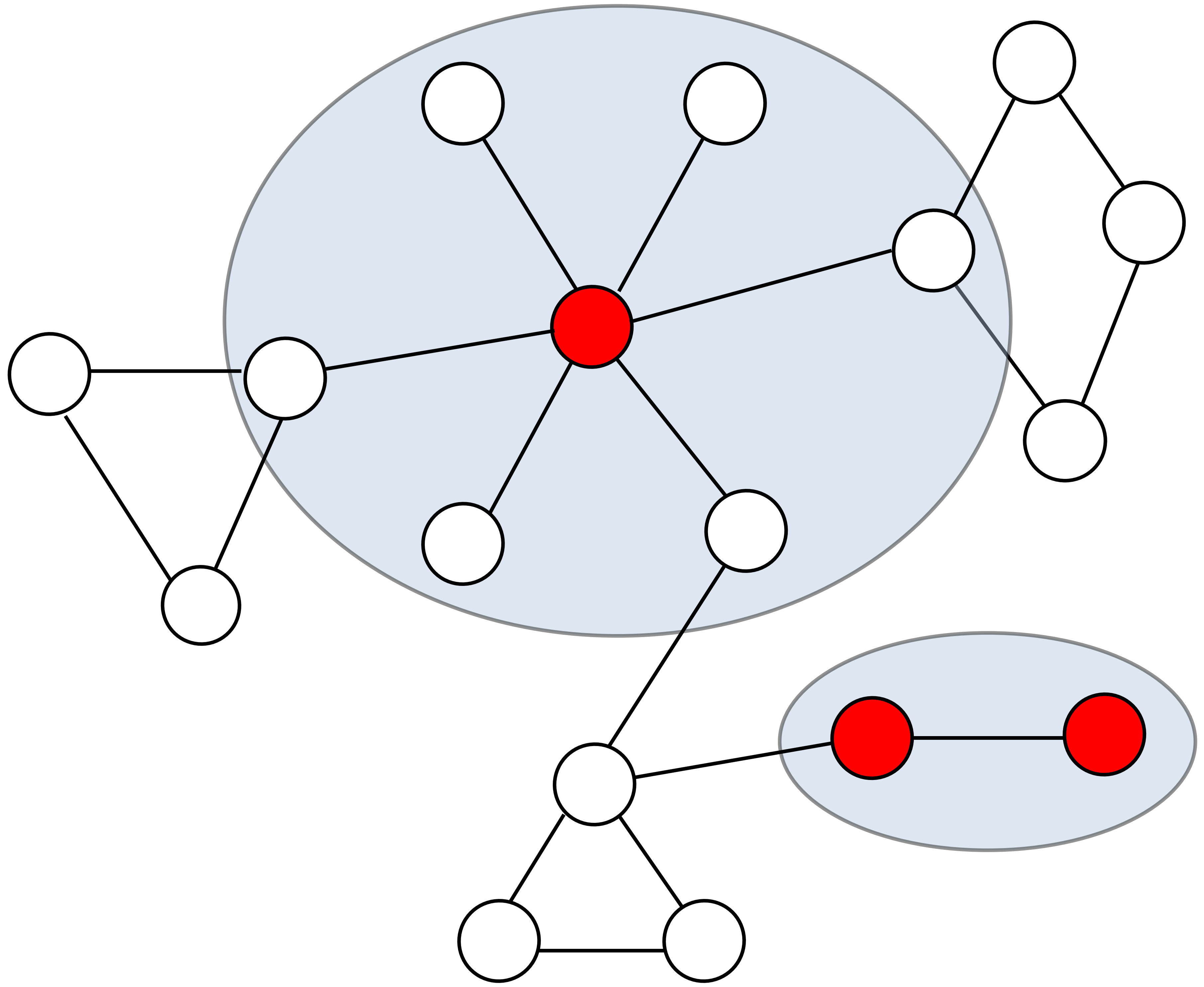
- These configurations can be *endogenous* — i.e., emerging from the connections actors make in response to other ties in their social environment.
  - E.g., when YBMSM are more likely to become Facebook friends with the friends of their friends



Aim 1

Exponential random graph models (ERGMs)

- Configurations can also be *exogenous* — i.e., emerging in response to properties that exist outside the network, like the attributes of network actors.
  - E.g., when marijuana smokers tend to be popular; or when marijuana smokers tend to be Facebook friends with one another



Aim 1

ERGMs “under the hood”

A simple heuristic:

- Calculate number of each configuration in the observed data.
- Simulate a sample of random graphs with similar structure to the observed network (like network rewiring).
- Calculate number of each configuration for each sampled graph.
- Compare the distribution mean for each configuration (what we expect by chance) to the observed counts

The General Form:

$$P(Y = y) = \frac{\exp(\theta' g(y))}{k(\theta)}$$

- Y is a network realization
- y is the observed network
- g(y) is a vector of model statistics for network y,
- Θ is the vector of coefficients for those statistics,
- k(Θ) is a normalizing factor.

Aim 1

Facebook network structure & HIV-related behaviors

Model 1: ERGM of Facebook friendships among YBMSM

- RQ: How do the HIV-related attributes of YBMSM bring structure to their Facebook friendships?

- YBMSM attributes: HIV prevention and risk behaviors, HIV status, structural vulnerabilities

Configurations of Interest

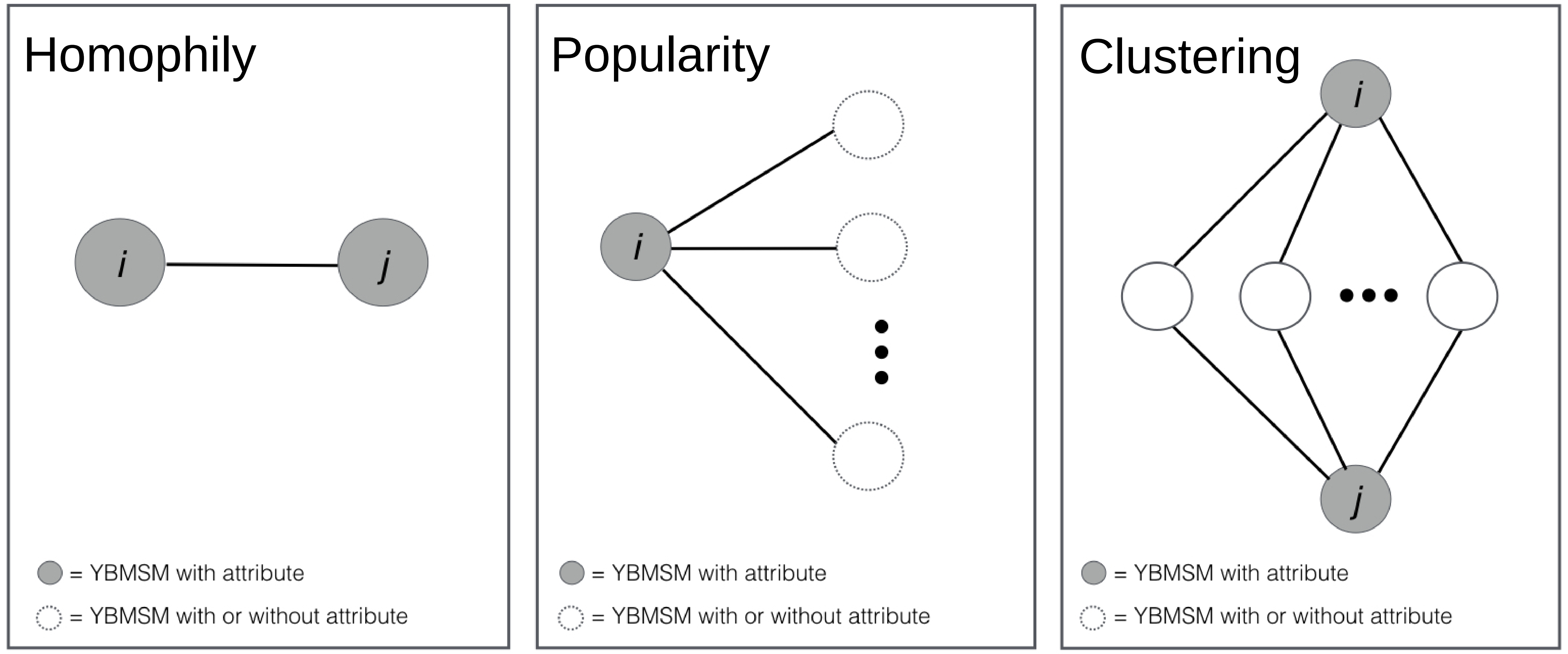


Figure 3. Interaction Edge **Figure 4. Alternating K-star, actor centered** **Figure 5. Alternating K-two path**

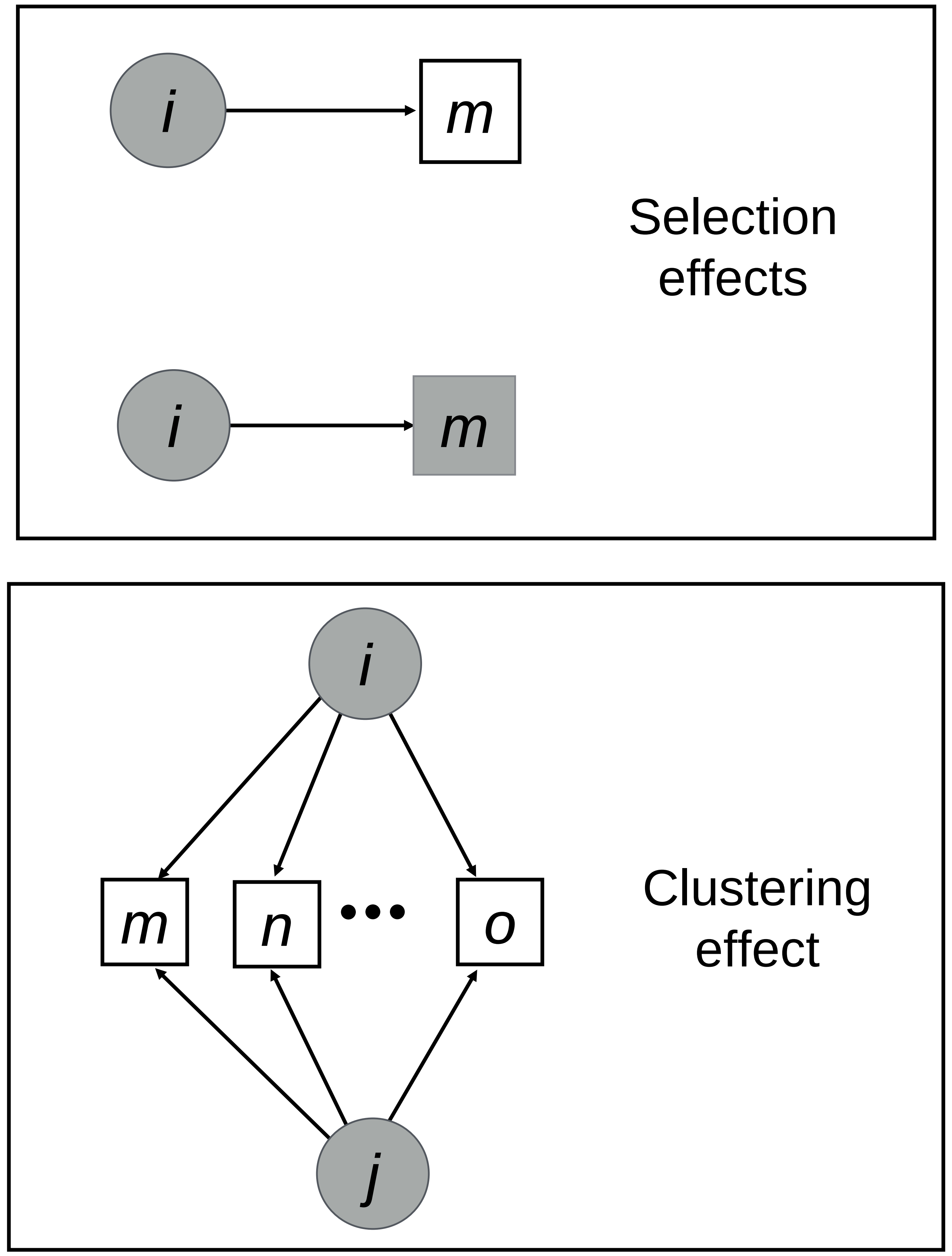
Aim 1

Facebook network structure & HIV-related behaviors

Model 2: ERGM of Facebook group affiliations among YBMSM

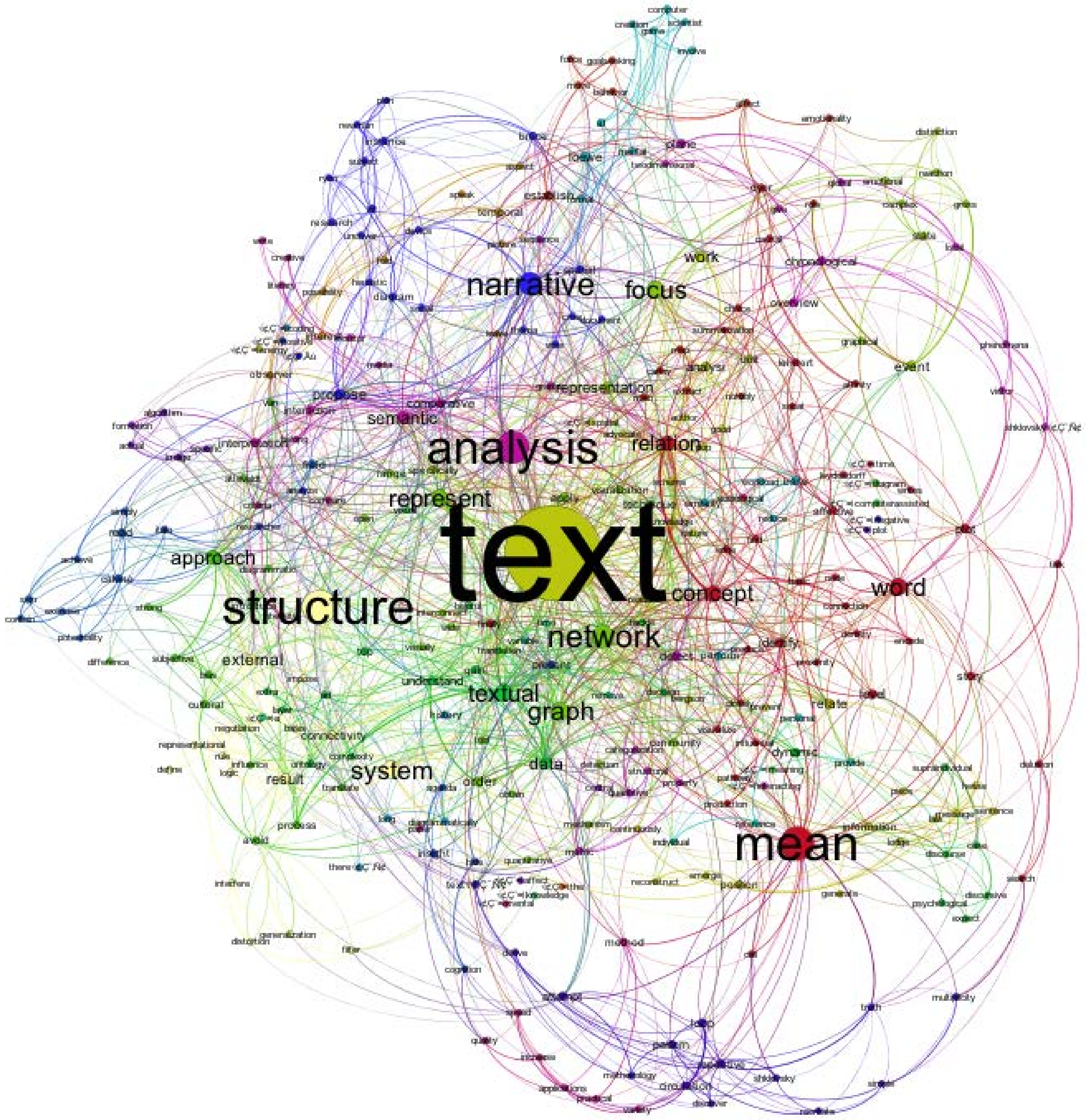
- RQ: How do the HIV-related attributes of YBMSM bring structure to their Facebook group affiliations?
- YBMSM attributes: sex behaviors, prevention behaviors, HIV-related communication traits
- Group attributes: group focus (e.g., LGBTQ identity, sex/sexuality, general chat, recreational interests)

Configurations of Interest



Aim 2

Facebook communication & HIV-related behaviors

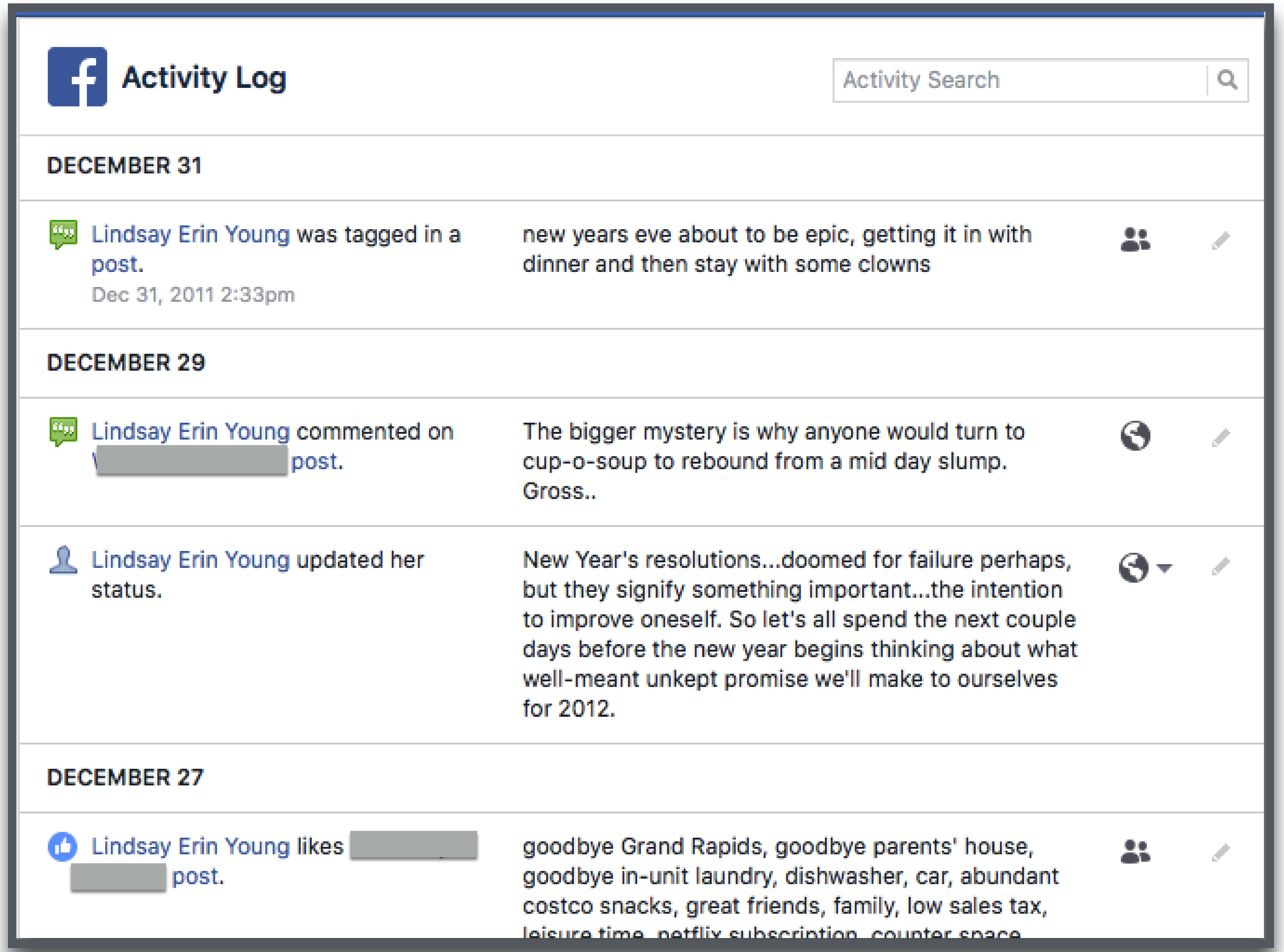


- RQ1: To what extent do YBMSM talk about HIV-related issues on Facebook?
- RQ2: What is the underlying structure of that discourse?
- RQ3: To what extent are the features of an individual's communication related to their HIV-related behaviors?

Aim 2

Facebook communication & HIV-related behaviors

Method 1:
Automated textual analysis



Aim 2

Facebook communication & HIV-related behaviors

Method 1: Automated textual analysis

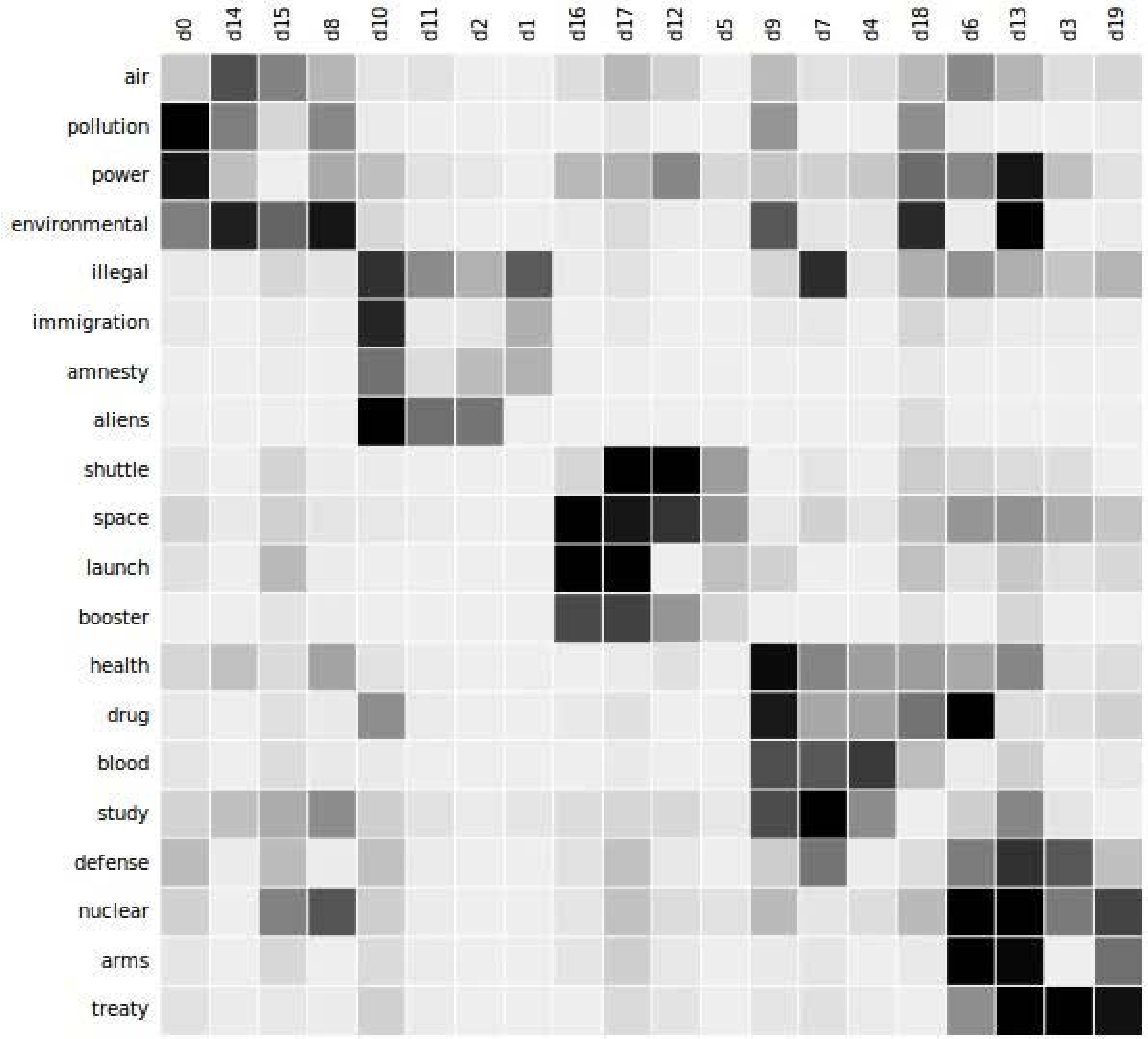
- Build a search term dictionary of words and phrases related to sexual activity, substance use, sexual health, and other factors associated with HIV.
- Extract key terms used in each participant’s corpus of timeline posts.
- Classify individuals into more meaningful categories based on their use (and non-use) of key terms using Latent Semantic Analysis.

Aim 2

Facebook communication & HIV-related behaviors

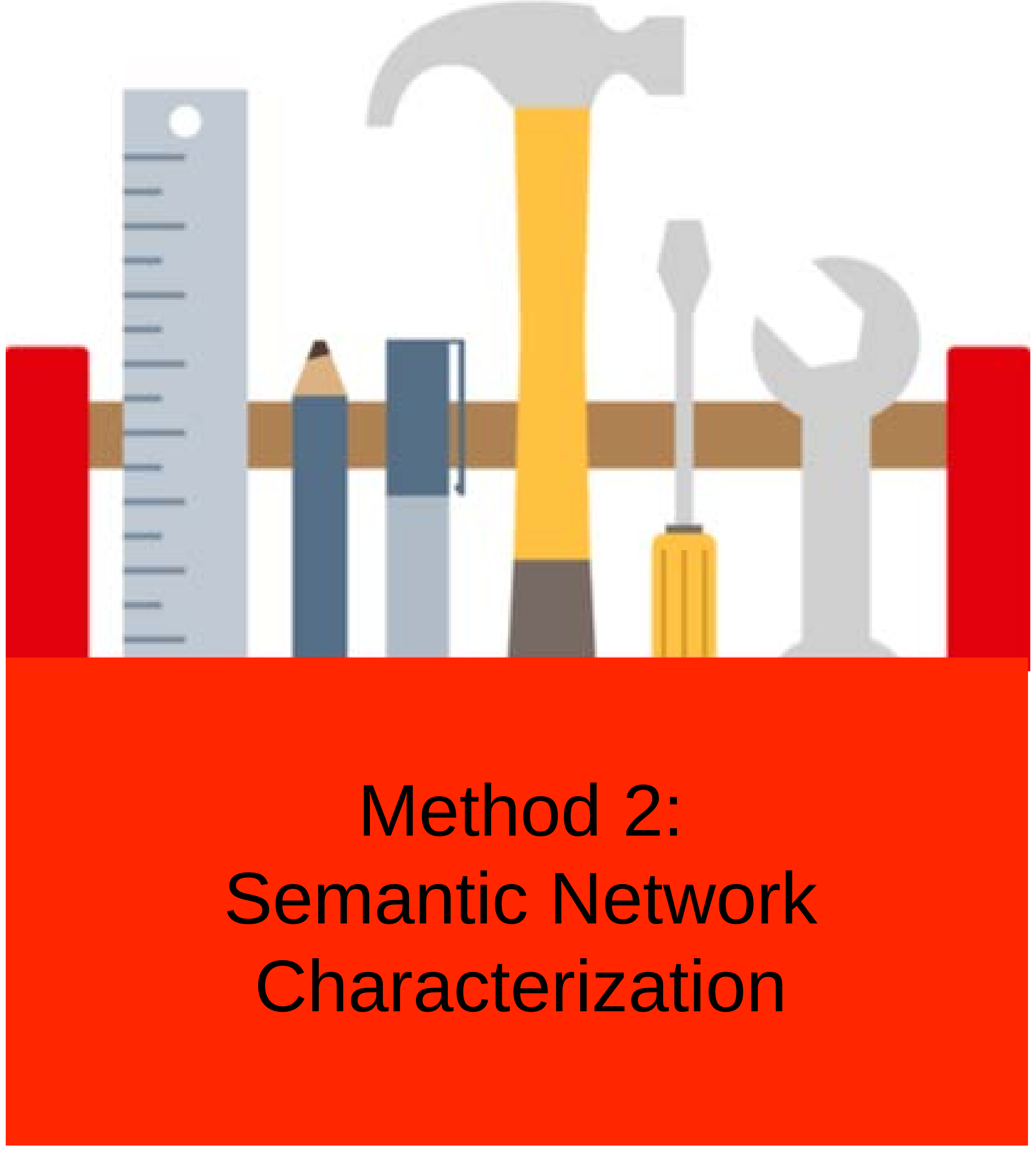
Latent Semantic Analysis (LSA) Output

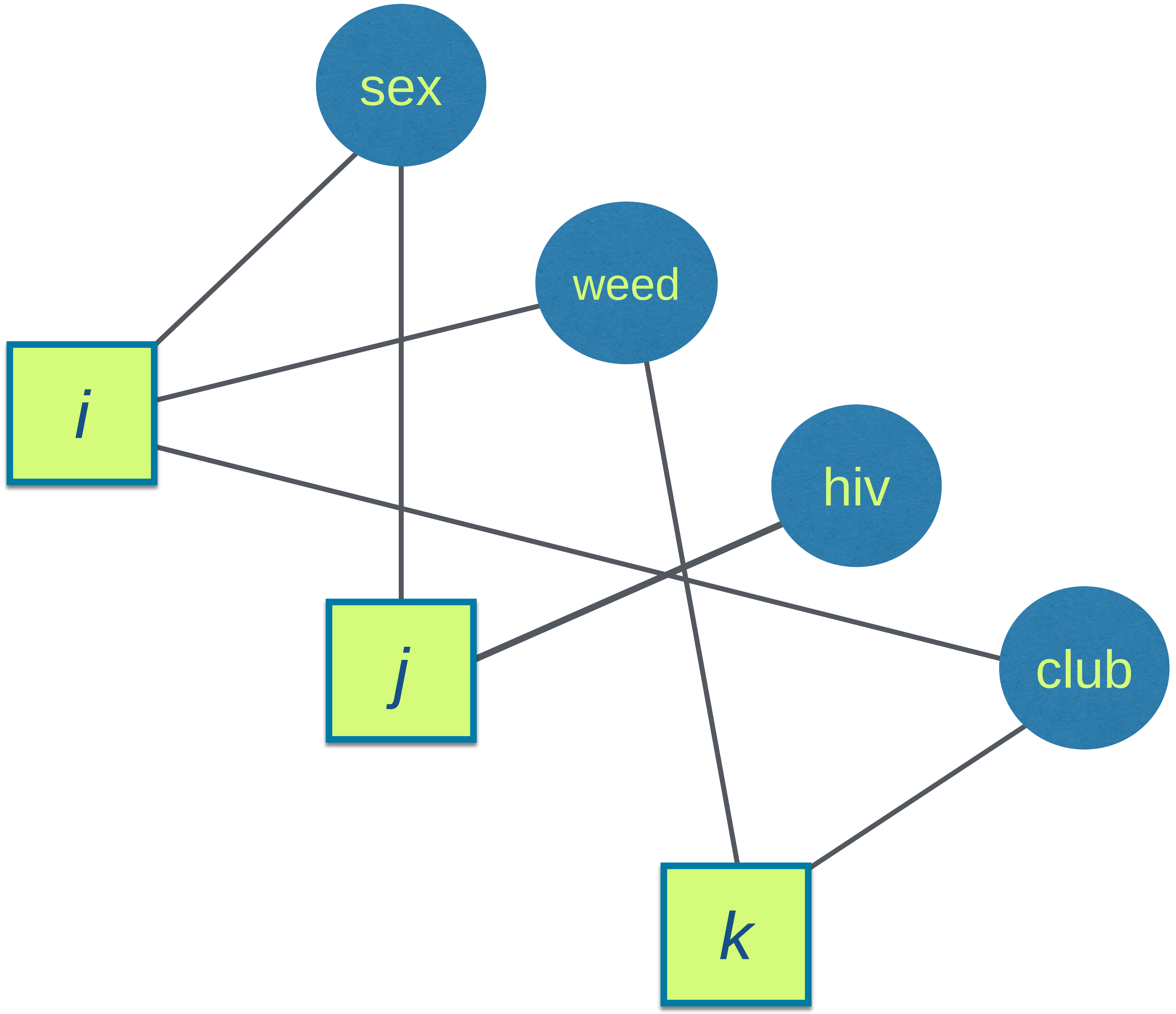
Terms	Documents													
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14
abnormalities	0	0	0	0	0	0	0	1	0	1	0	0	0	0
age	1	0	0	0	0	0	0	0	0	0	0	1	0	0
behavior	0	0	0	0	1	1	0	0	0	0	0	0	0	0
blood	0	0	0	0	0	0	0	1	0	0	1	0	0	0
close	0	0	0	0	0	0	1	0	0	0	1	0	0	0
culture	1	1	0	0	0	0	0	1	1	0	0	0	0	0
depressed	1	0	1	1	1	0	0	0	0	0	0	0	0	0
discharge	1	1	0	0	0	1	0	0	0	0	0	0	0	0
disease	0	0	0	0	0	0	0	0	1	0	1	0	0	0
fast	0	0	0	0	0	0	0	0	0	1	0	1	1	1
generation	0	0	0	0	0	0	0	0	1	0	0	0	1	0
oestrogen	0	0	1	1	0	0	0	0	0	0	0	0	0	0
patients	1	1	0	1	0	0	0	1	0	0	0	0	0	0
pressure	0	0	0	0	0	0	0	0	0	0	1	0	0	1
rats	0	0	0	0	0	0	0	0	0	0	0	0	1	1
respect	0	0	0	0	0	0	0	1	0	0	0	1	0	0
rise	0	0	0	1	0	0	0	0	0	0	0	0	0	1
study	1	0	1	0	0	0	0	0	1	0	0	0	0	0



Aim 2

Facebook communication & HIV-related behaviors



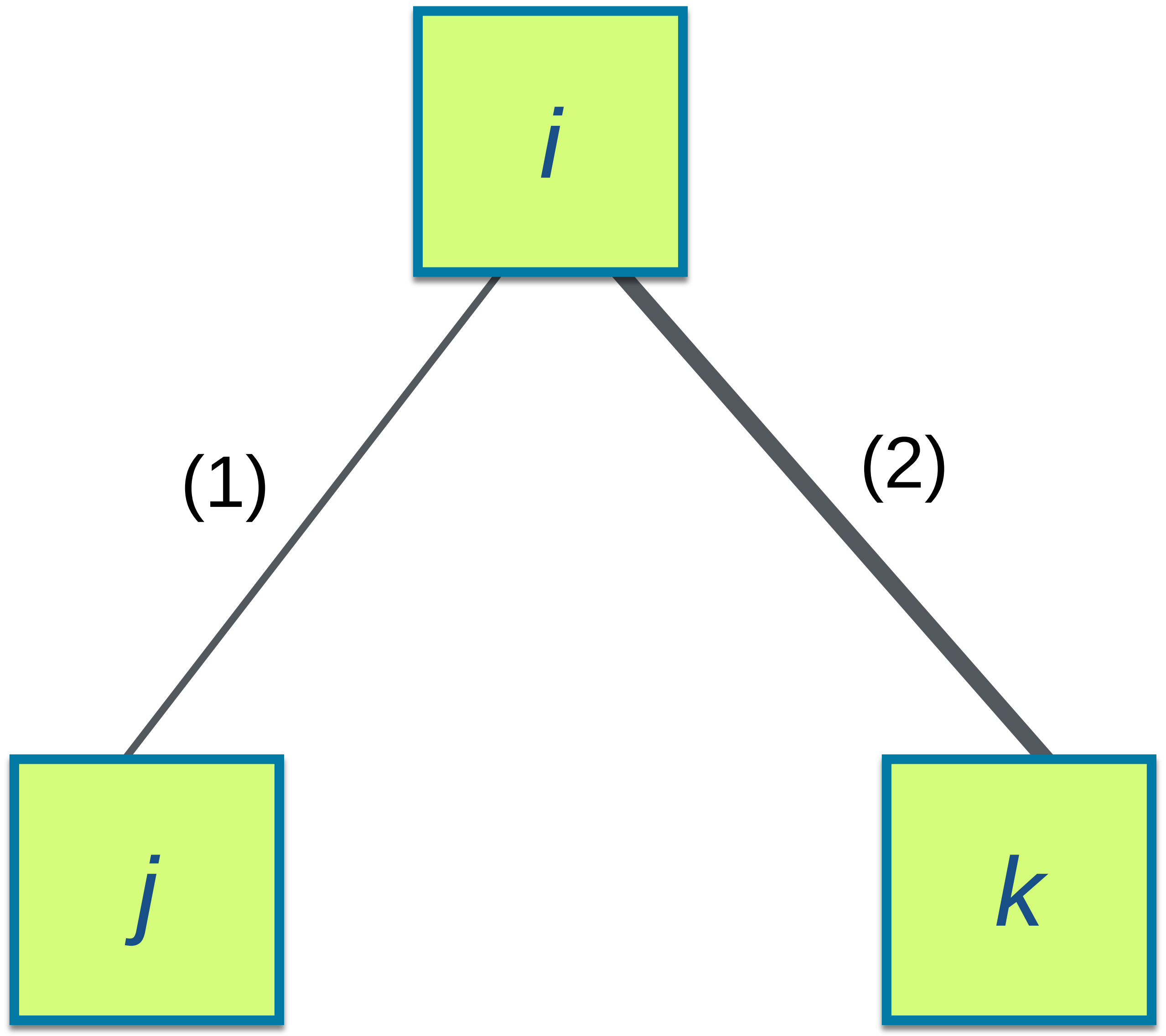


Two-mode network: ties between individuals and the keywords they use

Aim 2

Facebook communication & HIV-related behaviors

Method 2: Semantic Network Characterization

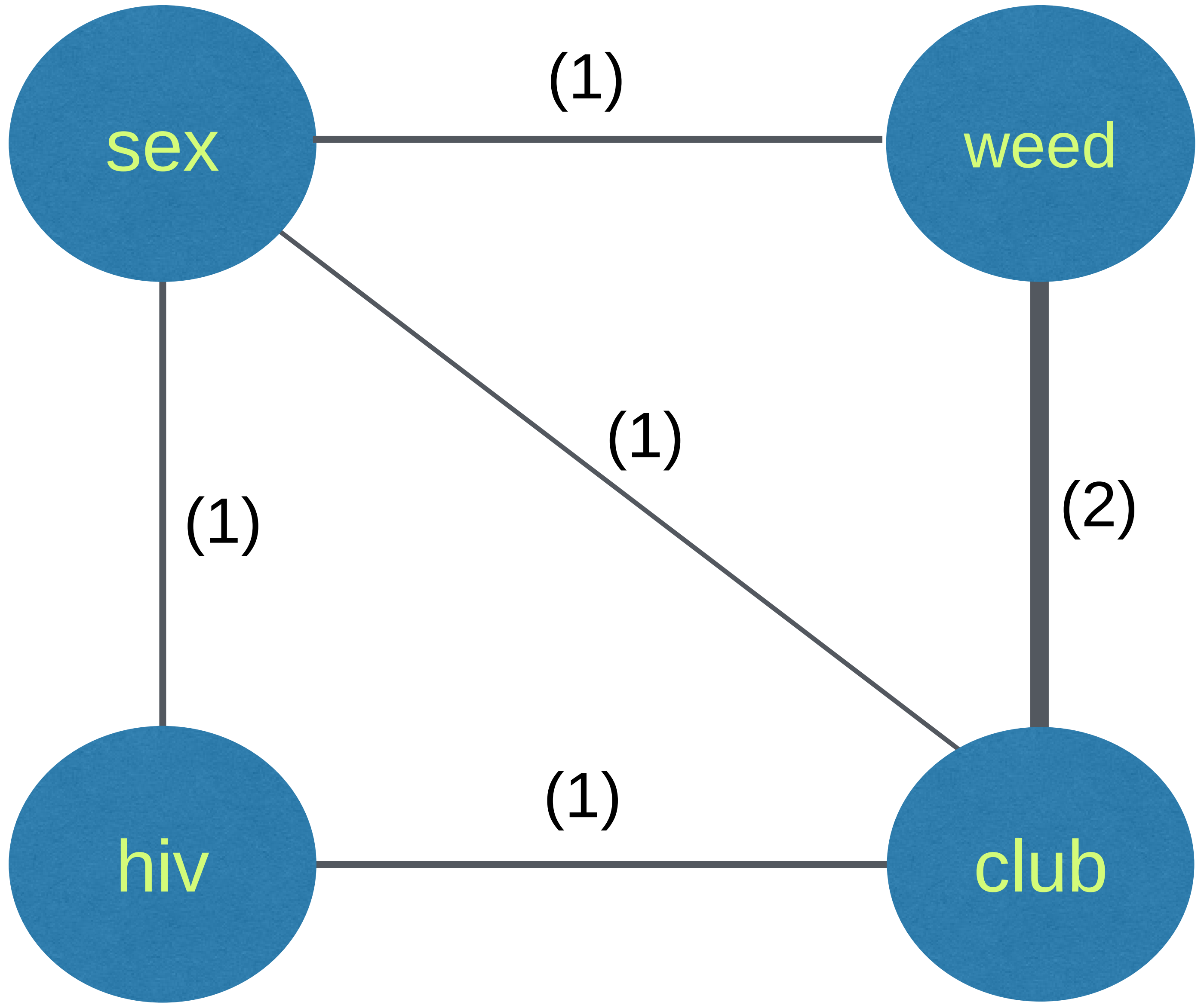


One-mode network: weighted ties between individuals represent joint use of keywords.

Aim 2

Facebook communication & HIV-related behaviors

Method 2:
Semantic Network Characterization

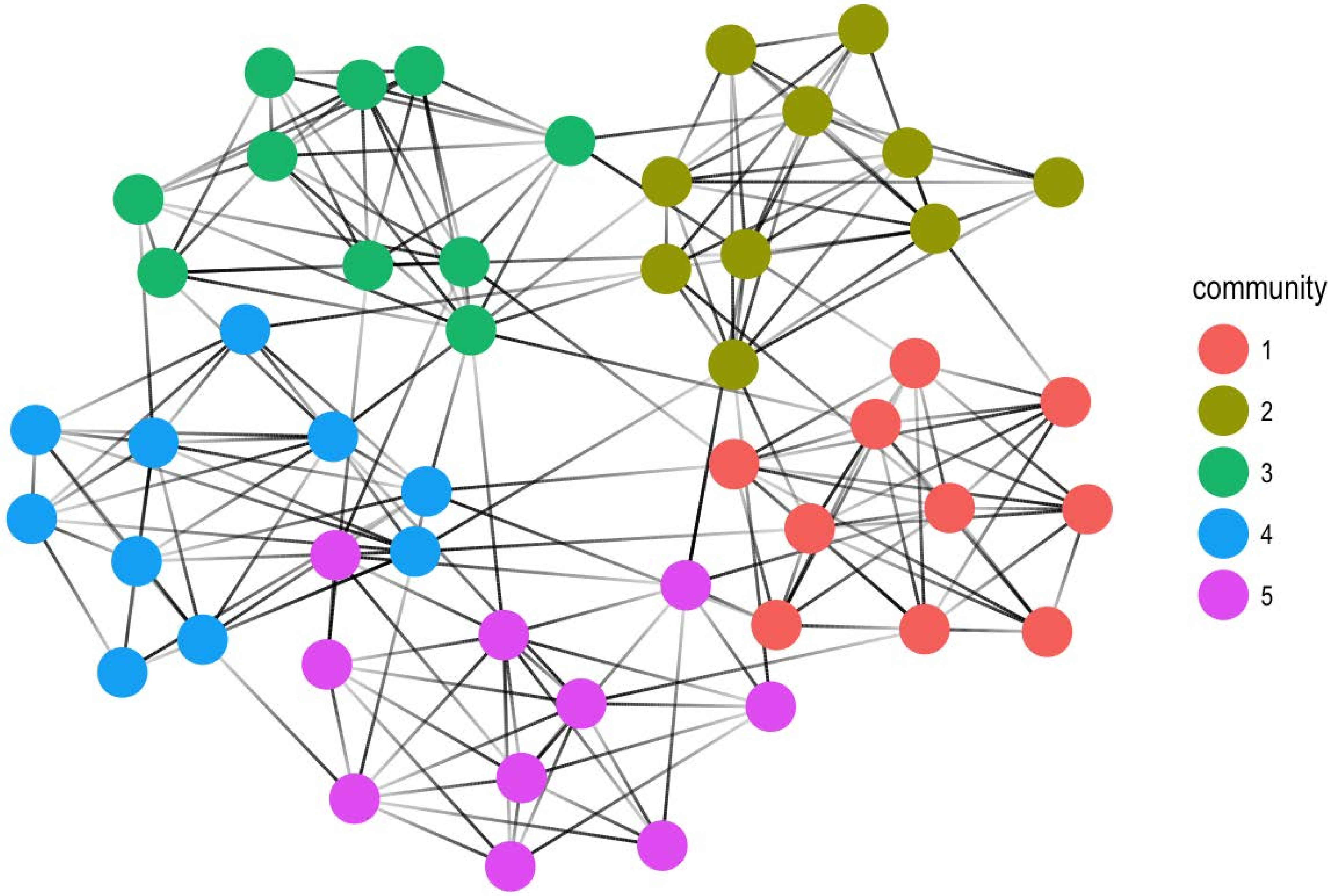


One-mode network: weighted ties between keywords represent keyword co-occurrence

Aim 2

Facebook communication & HIV-related behaviors

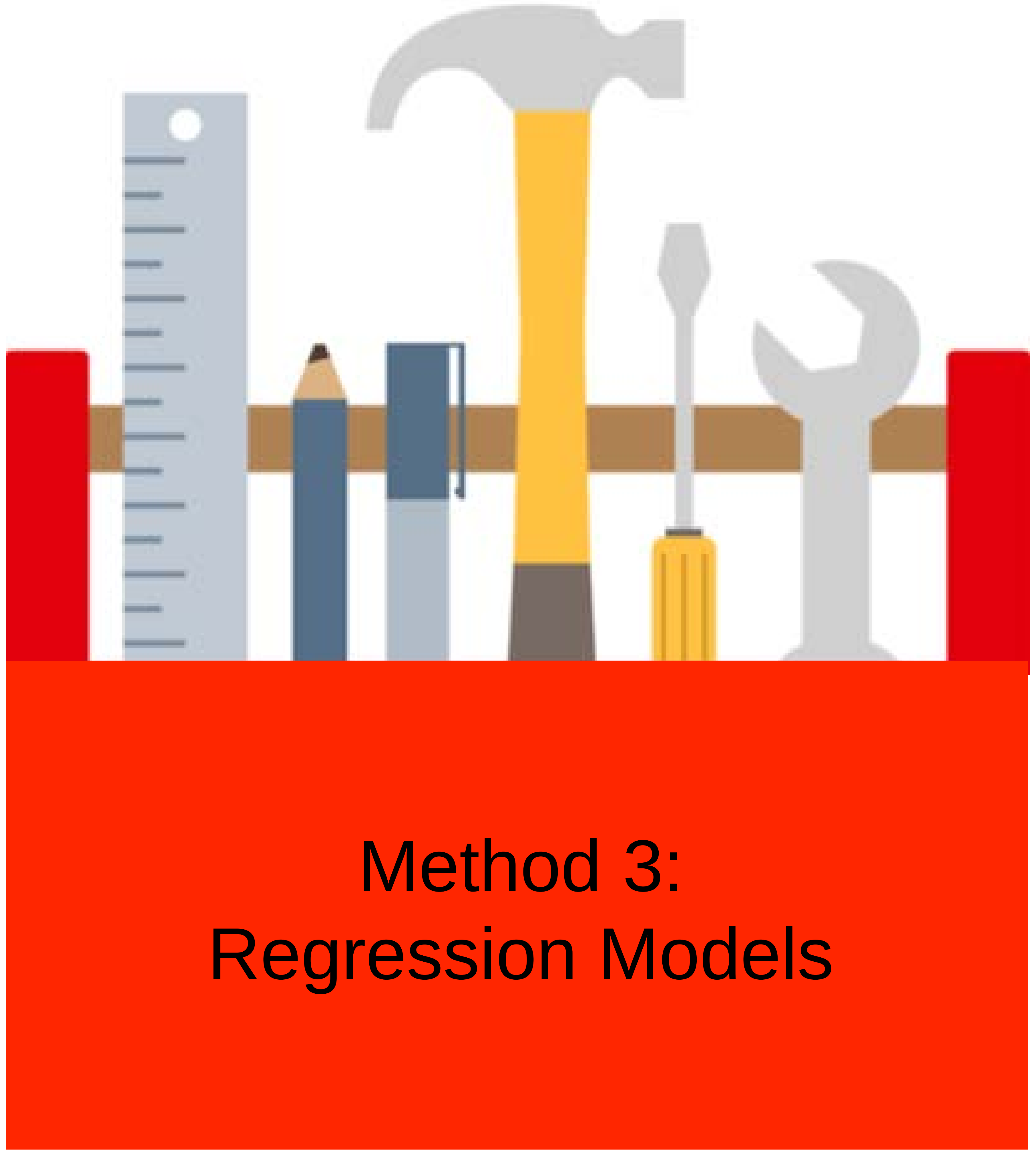
Method 2:
Semantic Network Characterization



Community detection among individuals or among words

Aim 2

Facebook communication & HIV-related behaviors



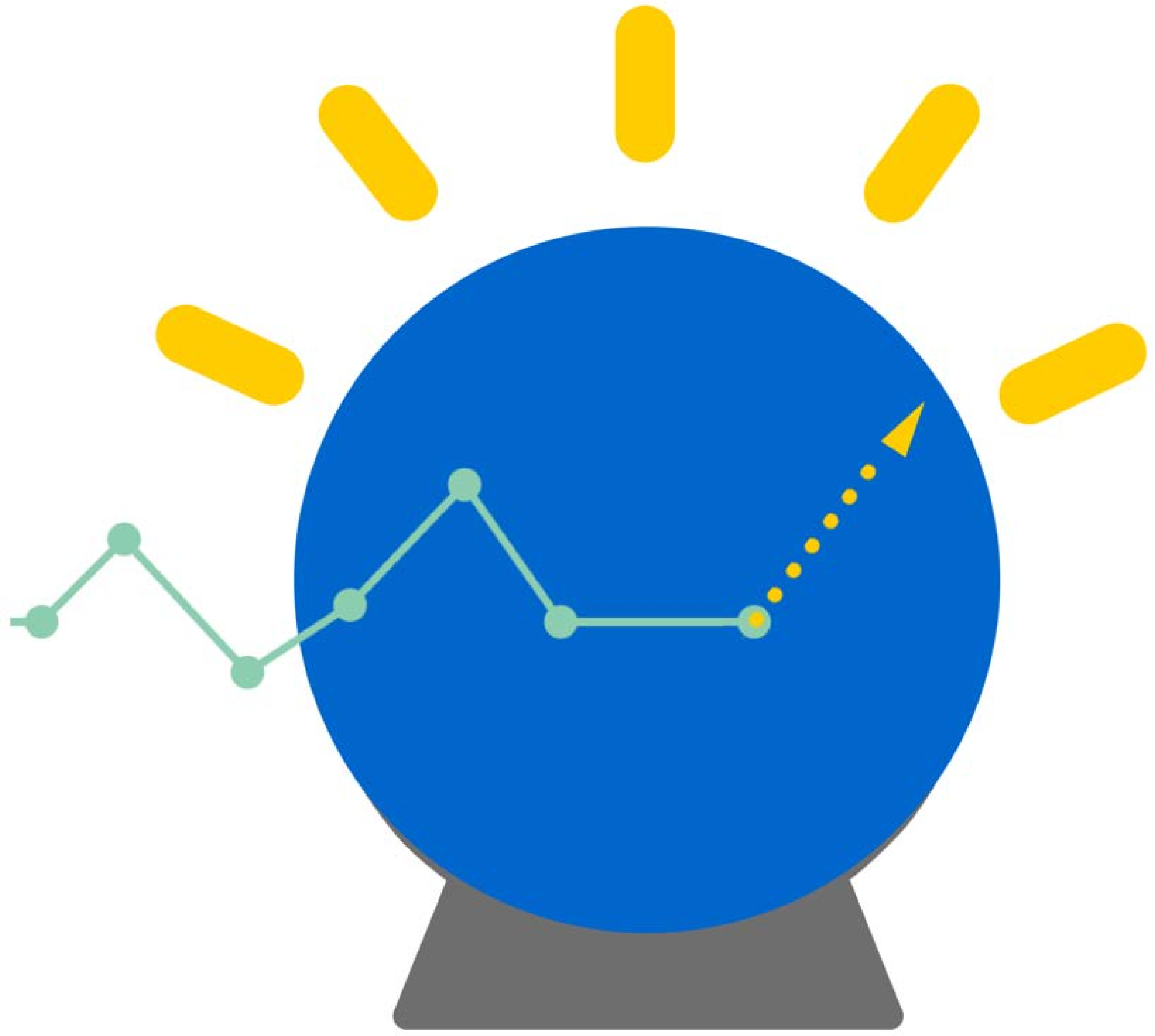
RQ: Is there an association between the way YBMSM communicate and their HIV prevention and care engagement?

Aim 3

Building a predictive model of HIV prevention/care engagement

Digitally-archived social media data

+



Predictive Analytics

=

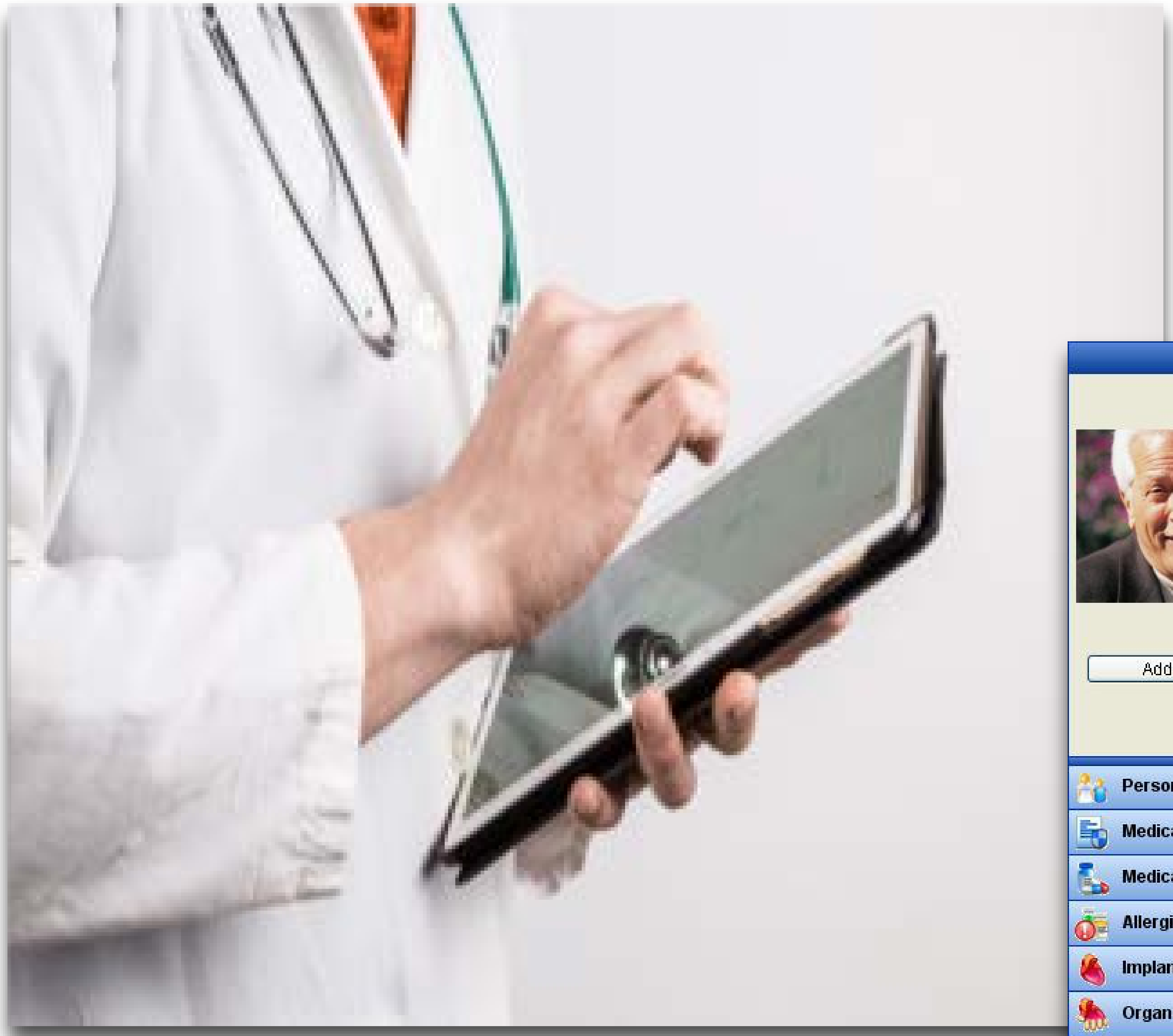
Tailored interventions for high risk YBMSM

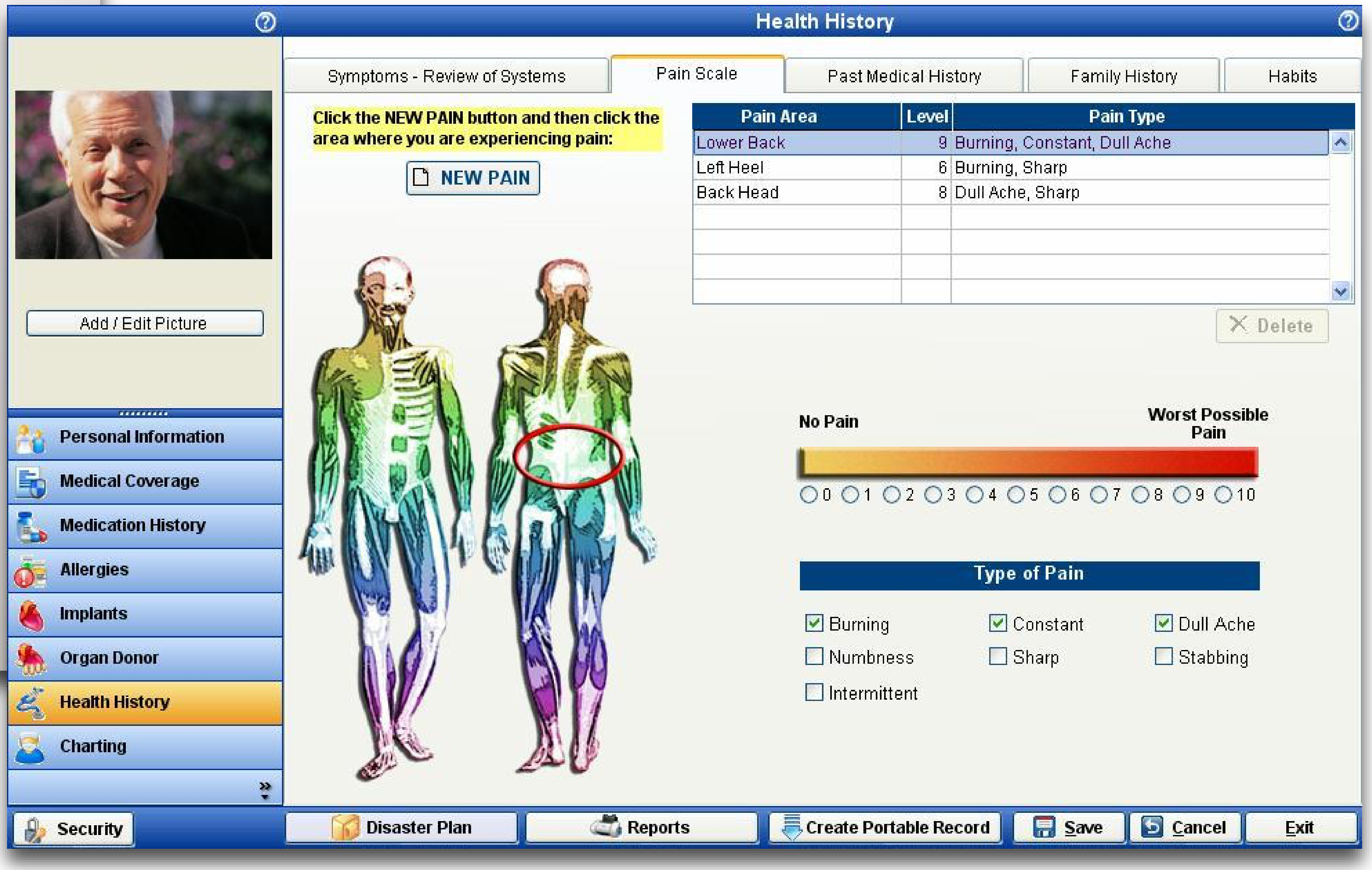
RQ: Can we identify at-risk YBMSM and predict future prevention engagement through social media use patterns

Aim 3

Predictive Analytics in Healthcare

Big Data Source:
Electronic Medical Records (EMRs)

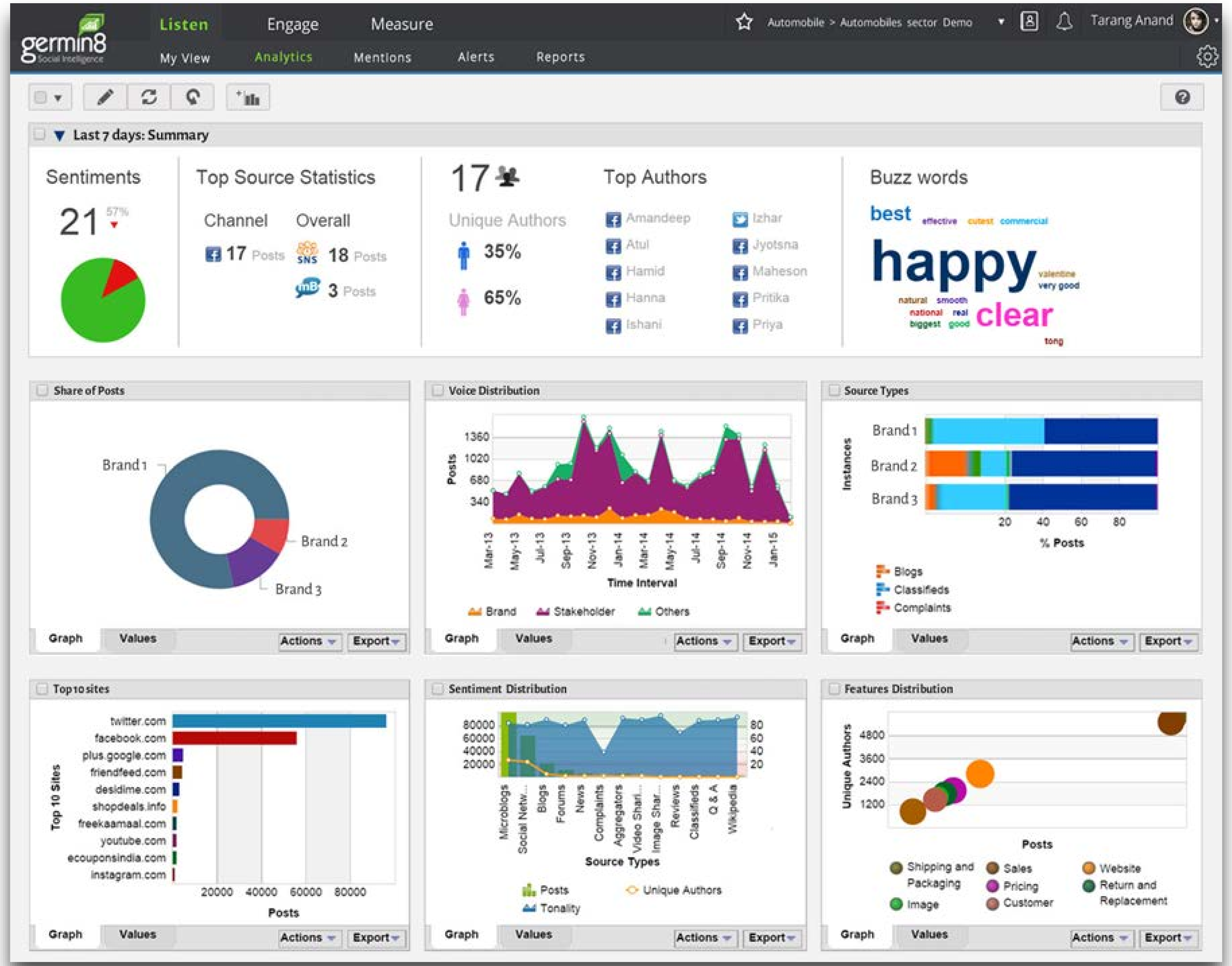




Aim 3

Predictive Analytics in Healthcare

Big Data Source: Social Media?



Aim 3

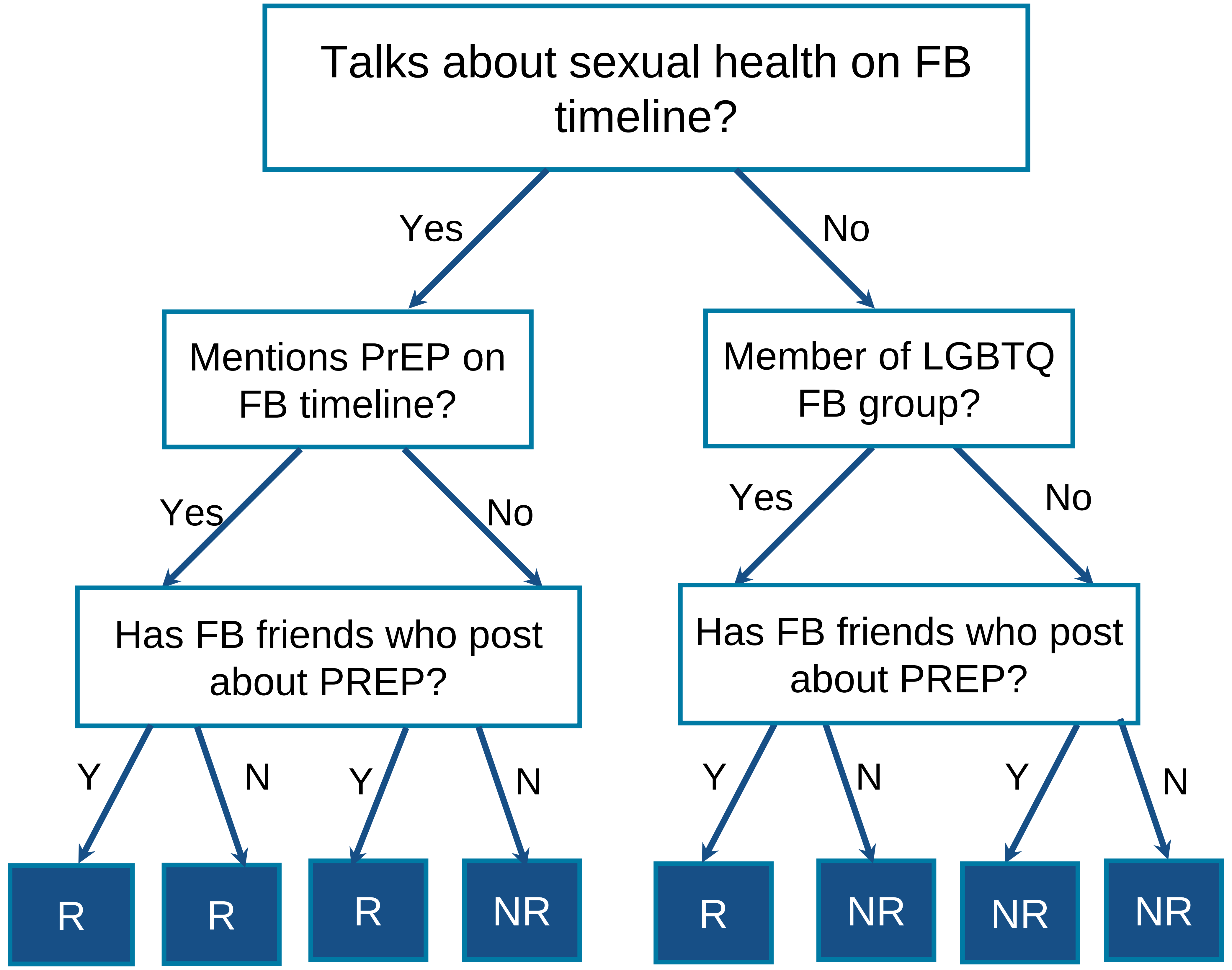
Building a predictive model of HIV prevention/care engagement

Model Details

- Approach: Machine learning techniques (e.g., decision trees, random forests)
- Features: Features of Facebook communication, group affiliations and friendships + EMR data, sex behaviors, substance use, structural vulnerabilities, sexual identity, demographics
- Output: Retention scores/classifications (e.g., probabilities of being retained in PrEP care); and indications of which features are the most accurate predictors

Oversimplified Decision Tree Model

(predicting retention in PrEP care as a function of FB features)



R = retained; NR = not retained

Practical Implications

- Aim 1: ERGMs model effects of configurations that can help us identify pockets (or structural signatures) of social and behavioral norms that may need to be intervened with (or leveraged) to facilitate behavior change in a targeted population.

- Aim 2: Identifies communication patterns that could be flagged and used as catalysts for a near-real time intervention

- Aim 3: Identifies more people at risk for future risks (e.g., retention failure), which will enable more tailored and proactive intervention

Outstanding questions and future work

Future Research

Information diffusion in a competitive information environment

- The problem: Influencing the sexual health attitudes and behaviors of YBMSM is difficult when their online social environments are saturated with all kinds of competing information.

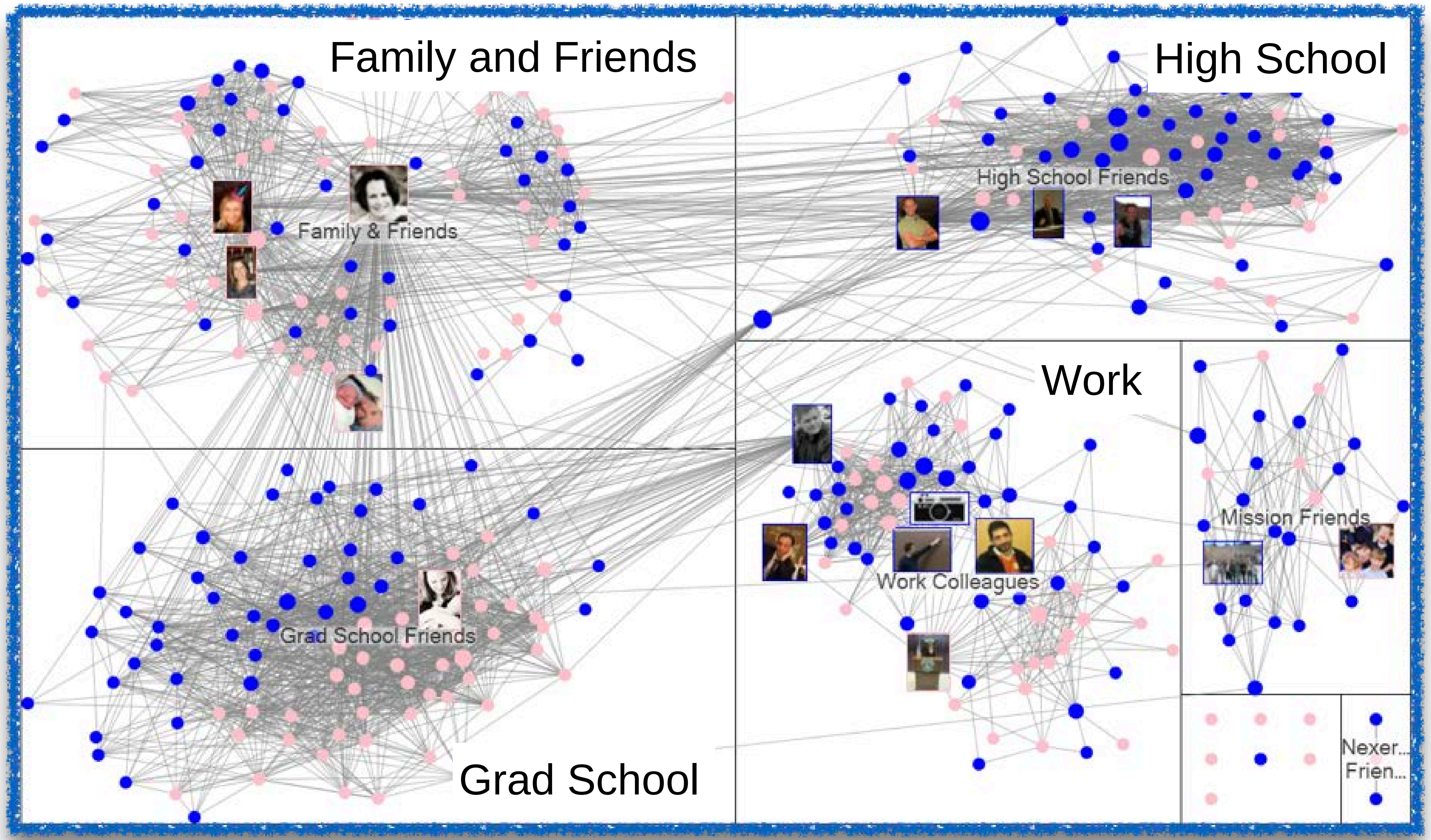
- The project: Can we model what will happen to the diffusion process when information competes with or complements one another and use those insights to design more effective diffusion interventions in organic online networks?

Future Research

The relationship between online and offline network dynamics

- The problem: Eliciting offline social networks over time is incredibly resource-intensive. But turnover in those networks have been linked to critical HIV risk and prevention outcomes.

- The project: Can we predict turnover in offline networks (e.g., the introduction of a new sex partner) from dynamic features of online social networks.



Acknowledgements:

National Institute of Child Health and Human Development (K99 funder)

John Schneider, University of Chicago (K99 mentor, postdoc advisor)

Kayo Fujimoto, University of Texas Health Science (K99 co-mentor)

Rayid Ghani, University of Chicago (K99 co-mentor)

Ian Holloway, UCLA (K99 consultant)

Tom Valente, Kayla de Haye, and the Center for Applied Network Analysis (CANA), USC

* Special thanks to: USC Center for Artificial Intelligence in Society (CAIS)

lyoung1@medicine.bsd.uchicago.edu