

# CDC Health Information Innovation Consortium August Forum

**Brian Lee, MPH**

Chief Public Health Informatics Officer  
Office of Public Health Scientific Services

Tuesday, August 1, 2017

10:00-11:00am ET

Chamblee 106, Room 1A + Adobe Connect

Dial-in 1-800-593-0825; Participant code 7695550#



Centers for Disease Control and Prevention  
Office of Public Health Scientific Services

## August Presenters



Matthew Hogben, PhD is the Branch Chief of the Social and Behavioral Research and Evaluation Branch. He has a PhD in Social Psychology from State University of New York. He has been a scientist in the Division of STD Prevention since the late 1990s. His research interests are sexual health, specifically promotion, care-seeking and provision; innovative models of partner services for persons infected with STD or HIV; and models and mechanisms for integrating program activity and program-focused research.

## August Presenters



Karen Kroeger, PhD is a Research Anthropologist in the Social and Behavioral Research and Evaluation Branch of the Division of STD Prevention. She conducted ethnographic fieldwork in Indonesia among female sex workers and received her PhD in 2000 from Washington University in St. Louis. She began her career at CDC in 2001 as a postdoctoral fellow in STD prevention, and from 2004-2009, she was a Behavioral Scientist CDC's Global AIDS program. She has conducted qualitative research, rapid ethnographic assessment, and evaluation among vulnerable populations of female sex workers, men who have sex with men, and persons who use drugs in the US and abroad.

## August Presenters



Elsworth (Els) Campbell is a health scientist in the Laboratory Branch of the Division of HIV/AIDS Prevention. Els holds bachelor's and master's degrees in biology from UC San Diego and is currently pursuing a PhD in biology at Penn State University. Els began working at CDC as a PhD student and Oak Ridge Institute for Science Education (ORISE) fellow in 2013 and recently transitioned to a full-time Associate Service Fellowship.

# INNOVATIVE SURVEILLANCE AND ASSESSMENT TECHNIQUES USING SOCIAL MEDIA TO INFORM STD AND HIV PREVENTION ACTION

Matthew Hogben, Branch Chief  
Karen Kroeger, Team Lead  
Social and Behavioral Research and Evaluation Branch  
Division of STD Prevention  
CDC Health Information Innovation Consortium Forum  
August 1, 2017

- ▶ Pre-exposure prophylaxis (PrEP) reduces HIV transmission, yet only 2-5% of Americans who may benefit from it are using it
- ▶ Men who have sex with men (MSM) using PrEP show high rates of sexually transmitted infections (STI)
- ▶ Divergence of traditional STD prevention from HIV prevention raises new questions:
  - ▶ Is STI an indicator for PrEP? (as an STD program intervention)
  - ▶ What is the impact of PrEP on STI prevention?

## BACKGROUND

- ▶ To evaluate social media site REDDIT as a novel data source for monitoring behaviors and attitudes relevant to PrEP
- ▶ To explore networks of attitudes and beliefs surrounding STD prevention in the context of PrEP

## PURPOSE OF THE PROJECT

NCHHSTP-Epidemico Kickoff Deck 10-17-2016.pdf - Adobe Acrobat Reader DC

File Edit View Window Help

Home Tools Document

150%

Supplemental Information

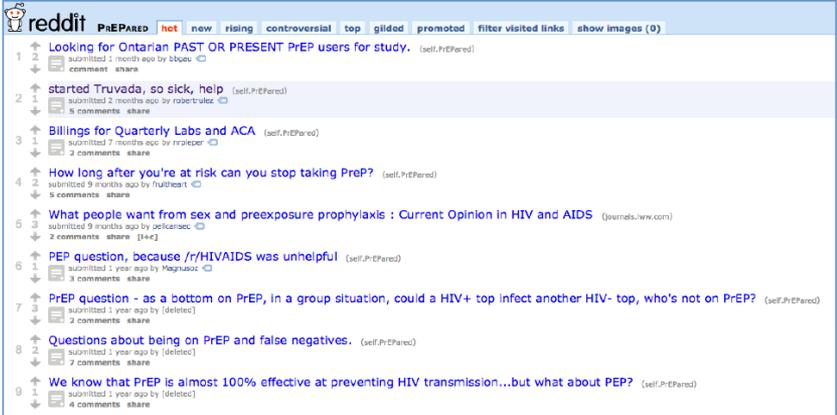
8

# ANATOMY OF REDDIT

## DISSECTING REDDIT

Within a subreddit, users see a list of submissions. These can be statements, questions, images or links to other websites.

In the PrEPared subreddit, we frequently find people asking about symptoms, experiences, sources and perceptions related to pre-exposure prophylaxis.



The screenshot shows the r/PrEPared subreddit interface. The header includes the subreddit name 'PrEPared' and navigation tabs like 'hot', 'new', 'rising', 'controversial', 'top', 'gilded', and 'promoted'. Below the header is a list of 9 submissions, each with a title, submission time, and user information. The submissions are:

- Looking for Ontario PAST OR PRESENT PrEP users for study. (self.PrEPared)
- started Truvada, so sick, help (self.PrEPared)
- Billings for Quarterly Labs and ACA (self.PrEPared)
- How long after you're at risk can you stop taking PrEP? (self.PrEPared)
- What people want from sex and preexposure prophylaxis : Current Opinion in HIV and AIDS (journals.lww.com)
- PEP question, because /r/HIVAIDS was unhelpful (self.PrEPared)
- PrEP question - as a bottom on PrEP, in a group situation, could a HIV+ top infect another HIV- top, who's not on PrEP? (self.PrEPared)
- Questions about being on PrEP and false negatives. (self.PrEPared)
- We know that PrEP is almost 100% effective at preventing HIV transmission...but what about PEP? (self.PrEPared)

Store and share files in the Document Cloud  
Learn More

Search the web and Windows

11:35 AM 7/20/2017

REDDIT IS AN OPEN FORUM PLATFORM THAT ALLOWS USERS TO SUBMIT AND COMMENT ON LINKS, IMAGES, AND STATEMENTS FROM ACROSS THE INTERNET.

NCHHSTP-Epidemico Kickoff Deck 10-17-2016.pdf - Adobe Acrobat Reader DC

File Edit View Window Help

Home Tools Document

Supplemental Information 9

# ANATOMY OF REDDIT

## DISSECTING REDDIT

Within a submission, users can comment in a threaded, hierarchical branching format.

**Subreddit** — GET +TESTED STD comments show Images (2)

**Submission Title** — Is this herpes? (self.STD) **Submission Author**  
submitted 5 days ago \* (last edited 5 days ago) by throwawaaayyy5739

**Submission Text** — <http://imgur.com/qH7rim4> I had unprotected oral sex and protected piv one week ago. I noticed that my urethra was red and it made me nervous. It's not a blister or lesion and it doesn't hurt to touch or urinate, I just decided to look inside and saw that it was red. (I tested negative for Gonnoreha and chlyamydia.)

8 comments source share hide all child comments

all 8 comments subscribe

sorted by: best

navigate by: submitter moderator friend me admin highlighted tagged gilded IAMA images videos popular new

**Comment Text** — [-] Maggost 1 point 5 days ago **Comment Author**  
Herpes in that location is pretty rare, looks like an UTI.

permlink source embed save-RES hide child comments

**Comment Text** — [-] throwawaaayyy5739 1 point 5 days ago **Comment Author**  
Hey, thank you for replying :, put me at ease

permlink source save-RES parent

Note: We use the term 'post' to refer to submissions and comments

Comment

Fill & Sign

Store and share files in the Document Cloud

Learn More

Search the web and Windows

12:30 PM 7/20/2017

# DISSECTING REDDIT POSTS

- ▶ Data collection and processing
  - ▶ Refine research questions, identify concepts of interest
  - ▶ Define sampling frame (subreddits -e.g. askgaybros, PrEPared)
  - ▶ Develop search and processing taxonomies
    - ▶ 20 search terms (product names, STI names, concepts from STI prevention)
    - ▶ Up to 40 concepts (e.g. stigma, side effects, adherence)
- ▶ Data analysis and output
  - ▶ Process queries (one time .csv data file with meta-tagged data)
  - ▶ Data explorer dashboard to conduct analysis/output

## STEPS IN THE PROCESS

- ▶ What were the main topics of discussion regarding PrEP in 2015?
- ▶ How much uptake/use of PrEP was there vs. information seeking?
- ▶ What are main barriers to PrEP uptake?
- ▶ How are MSM talking with sex partners about PrEP?
- ▶ When discussing PrEP how often are STDs other than HIV discussed, which STDs are mentioned, and in what context?

REFINED RESEARCH QUESTIONS,

- ▶ Temporal visualization of mentions of PrEP and STIs
- ▶ Visualization of distribution of mentions across digital space
- ▶ Visualization of relationships between topics
- ▶ Provide a view of the “size or volume” of conversations about PrEP (but not the “shape or meaning” of the conversation)

DATA EXPLORER (DASHBOARD)

A series of white, parallel diagonal lines of varying lengths and positions, located in the bottom right corner of the slide, creating a modern, abstract graphic element.



10/26/21 Evidence-Mining Deck 10.10.2020.pdf - Adobe Acrobat Reader DC

Home Tools Document

3

# REVIEW SEARCH TERMS

GATHERING DATA

Term	Search Term
PfEP	PfEP, exposure prophylaxis
Truvada	truvada, tenofovir, emtricitabine
Condoms	Condoms, condom
Syphilis	Syphilis
Gonorrhoea	Gonorrhoea
Chlamydia	Chlamydia
Herpes	Herpes, HSV
HPV	HPV
HIV	HIV, AIDS
STD	STD, STDs, STIs, STI, sexually transmitted
Testing (STD/HIV)	test, tested, testing
Positive Serostatus	Positive, poz, undetectable viral load
Negative serostatus	negative, clean

Comment  
Fill & Sign

Share and share this in the Document Cloud  
Learn More

10/26/21 Evidence-Mining Deck 10.10.2020.pdf - Adobe Acrobat Reader DC

Home Tools Document

4

# REVIEW CONCEPT TERMS

TAGGING CONCEPTS

Term	Synonyms
Hookups	hookups, hook ups, hookup, hookups, hooked up
Partner	partnr, partners, boyfriend, husband,
Bottoms	bottom, bottoms, bottoming
Tops	top, tops, topping
Oral Sex	oral, blow, blew, blowjob, went down, going down
Group Sex	threesome, threesomes, foursome, foursomes, orgies...
Bareback	bareback, raw, bare
Open Relationship	open, poly, polyamorous,
Communication/disclosure	disclose, disclosure, tell your, tell my, tell their...
Sex and Drugs	meth, crystal, tina, chem sex, PNP, party and play, poppers,
Mobile App Use	gridr, scruff, manhunt, jackd, tinder
Adherence/timing	adhere, adherence, every day, on time, missed my, missed a...
Patient-provider communication	doctor, doctors, nurse, nurses, clinic, clinics, RN, GP,
Stigma	slut, whore, truvada whore
Cost	copay, copays, coupon, insurance, confidentiality, cost, price...
Side effects	fatigue, kidney, GI problems, bloating, bloated, weird dreams...

Comment  
Fill & Sign

Share and share this in the Document Cloud  
Learn More

# SEARCH TERMS AND CONCEPTS

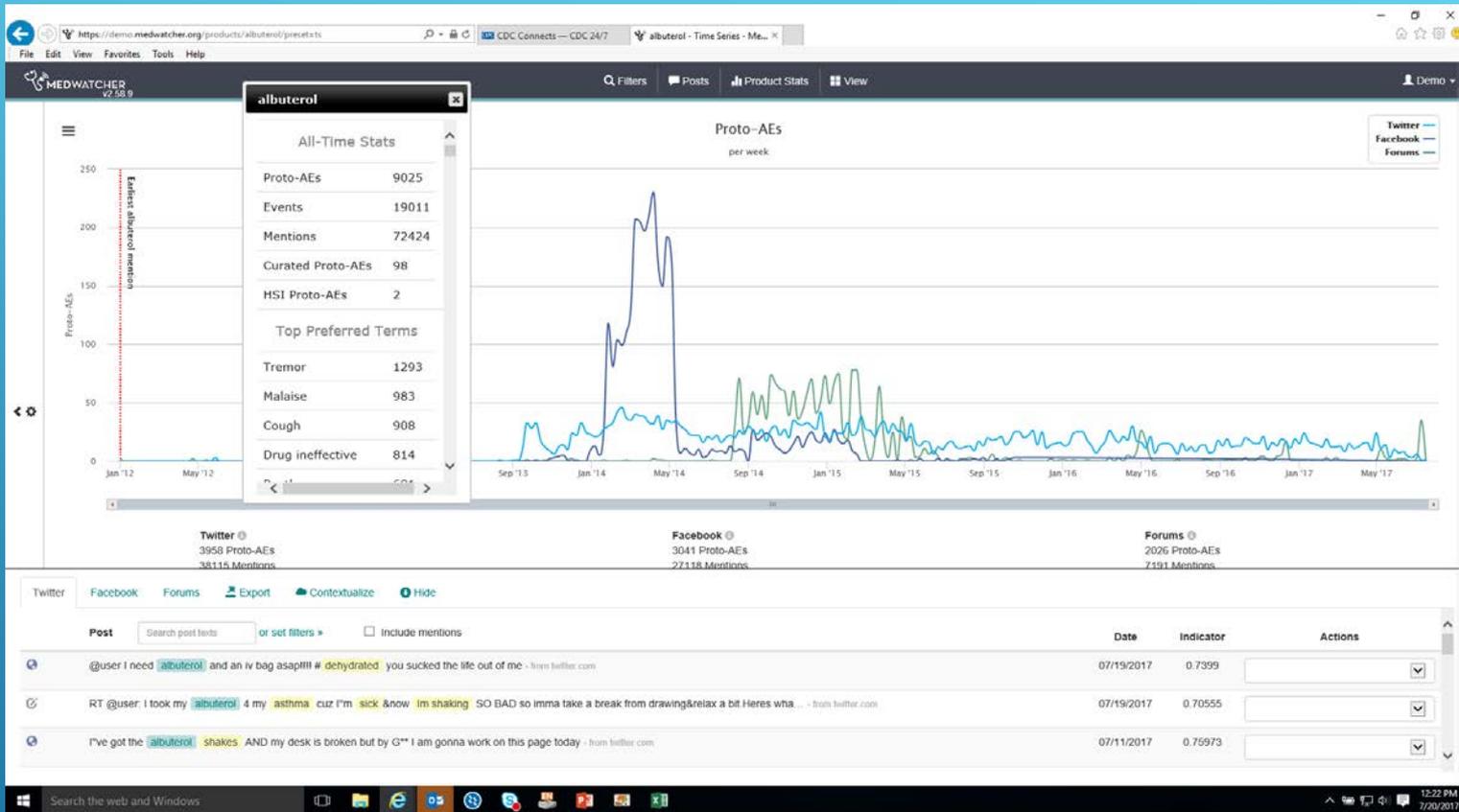
NCHHSTP\_Reddit\_Data\_(BAH\_Proprietary) - Excel

id	author	subreddit	title	post_text	created_utc	Datetime	type	terms	groups	STI_menti	Product_n	STI_or_Product_mention
3			Delivered to CDC-NCHHSTP under contract # 200-2016-M-91892, BAH Proprietary.									
4	3fd6io	Baconbits4e	askgaybros	Question About Anal	1.44E+12	7/31/2011 2:19	submission	[STD, Partner, Positive Ser	(Serostatus, STI, Sexu	1	0	1
6	3fed4l	Virginbottom	askgaybros	First gay experience	1.44E+12	7/31/2011 9:22	submission	[Condoms, Oral Sex, Botton	(Product, Sexual Beha	0	1	1
7	3ffk0a	[deleted]	PrEPared	Is it immoral to advocate PrEP wh	1.44E+12	7/31/2011 18:54	submission	[Patient-provider communi	(Product, Product Inf	0	1	1
8	3fg2w4	asaawwww	askgaybros	what is it with guys trying to fuck	1.44E+12	7/31/2011 21:30	submission	[Condoms]	(Product)	0	1	1
9	3fg4n8	cirrus_sc	askgaybros	How do you decide when (if ever	1.44E+12	7/31/2011 21:46	submission	[STD, Condoms, Testing [ST	(Product, STI, Service,	1	1	1
10	3fg1tm	GeorgeGnyc02	askgaybros	No condom, but PrEP? A couple	1.44E+12	8/1/2011 0:03	submission	[Condoms, HIV, Tops, Grou	(Product, STI, Sexual I	1	1	1
11	3fg09e	jsuce6e	gaybros	In Some Ways, AIDS in America	1.44E+12	8/1/2011 0:45	submission	[HIV]	(STI)	1	0	1
12	3fgyzf	Letha_Arajh	askgaybros	People are saying I'm overreactin	1.44E+12	8/1/2011 2:31	submission	[Condoms, Positive Serosta	(Serostatus, Product,	0	1	1
13	3fhg9b	scaredthrowaway	askgaybros	I'm scared	1.44E+12	8/1/2011 5:44	submission	[Negative serostatus, Testi	(Serostatus, STI, Servi	1	0	1
14	3fjn7f	Obversa	askgaybros	[23F] Question about Blood Dona	1.44E+12	8/1/2011 21:01	submission	[STD, Condoms, Oral Sex, T	(Product, STI, Service,	1	1	1
15	3fkvyp	fwioudwif	gaybros	Best STD Dating Sites for STD	1.44E+12	8/2/2011 3:38	submission	[STD]	(STI)	1	0	1
16	3fkvft	gaidatingexpert	askgaybros	Should we use protection? Hello	1.44E+12	8/2/2011 4:02	submission	[Condoms, Oral Sex]	(Product, Sexual Beha	0	1	1
17	3fkypj	theszak	gay	The strategy. BEFORE sex get	1.44E+12	8/2/2011 4:03	submission	[STD, Testing (STD/HIV)]	(STI, Service)	1	0	1
18	3f8i2i	Oz_33	gaybros	"Here's What We Actually Know	1.44E+12	8/2/2011 5:45	submission	[HIV]	(STI)	1	0	1
19	3f8laf	nod3s	TopsAndBottoms	Ex ruined my hole?	1.44E+12	8/2/2011 6:13	submission	[Partner, Condoms, Botton	(Product, Sexual Beha	0	1	1
20	3f8fv38	britnit	gay	Aging and HIV issues	1.44E+12	8/2/2011 11:03	submission	[STD, Patient-provider com	(Sexual Behaviors, ST	1	0	1
21	3f8penp	kokuinjmatto	gaybros	Any bros that live in Amsterdam?	1.44E+12	8/3/2011 3:36	submission	[Condoms, PrEP]	(Product)	0	1	1
22	3f8ykv	robertulez	askgaybros	Went bareback for the first time.	1.44E+12	8/3/2011 7:06	submission	[STD, Bareback, Condoms, T	(Product, STI, Service,	1	1	1
23	3f8qpx	justraging	askgaybros	just an angry and scared rant	1.44E+12	8/3/2011 17:29	submission	[Condoms, HIV, Patient-prc	(Serostatus, Product,	1	1	1
24	3f8t23	advocate4avocado	PrEPared	Taking PrEP into my own hands	1.44E+12	8/3/2011 22:57	submission	[Patient-provider communi	(Product, Product Inf	1	1	1
25	3f8ucep	bihere	askgaybros	23M here, not able to top my boy	1.44E+12	8/4/2011 5:33	submission	[Partner, Adherence/timing	(Product, Product Inf	0	1	1
26	3f8vkhw	blissed_out_cossac	askgaybros	What's the mechanics of catching	1.44E+12	8/4/2011 14:03	submission	[STD, Oral Sex]	(STI, Sexual Behavior	1	0	1
27	3f8v9bf	cjrutherford	askgaybros	Am I crazy? Am I being too hard	1.44E+12	8/4/2011 17:02	submission	[Condoms, Communication	(Product, Sexual Beha	0	1	1
28	3f8wbw	red_lambda	gay	I contracted Herpes	1.44E+12	8/4/2011 17:20	submission	[Herpes]	(STI)	1	0	1
29	3f8qdqo	Fcb12	askgaybros	Starting on PrEP?	1.44E+12	8/4/2011 18:56	submission	[Cost, PrEP]	(Product, Product Inf	0	1	1
30	3f8xc3g	throwaway8491919	askgaybros	"Relationship" help?	1.44E+12	8/4/2011 21:25	submission	[Partner, Condoms, Oral Se	(Product, Sexual Beha	0	1	1
31	3f8ywg6	dandelion-floatie	PrEPared	PrEP Interaction with other medic	1.44E+12	8/5/2011 5:08	submission	[PrEP]	(Product)	0	1	1
32	3g14jc	uboaaaaaaaaaaaaaa	askgaybros	Celibacy and not letting anyone	1.44E+12	8/5/2011 18:00	submission	[STD, Testing (STD/HIV)]	(STI, Service)	1	0	1
33	3g306r	analbeadsjim	gay	check out this gay condom compa	1.44E+12	8/6/2011 2:30	submission	[Condoms]	(Product)	0	1	1
34	3g30tp	trifurcate	gay	"9/11 is the justification of AIDS	1.44E+12	8/6/2011 2:36	submission	[HIV]	(STI)	1	0	1
35	3g3e1f	prinsengracht	askgaybros	Basic Questions	1.44E+12	8/6/2011 4:39	submission	[STD, Mobile App Use, Top	(STI, Sexual Behavior	0	1	1
36	3g3qr6	Mootank	askgaybros	Advice.. hes still in a relationship	1.44E+12	8/6/2011 6:58	submission	[Partner, Adherence/timing	(Serostatus, STI, Prodi	1	0	1
37	3g3rez	[deleted]	askgaybros	I was diagnosed as HIV+ yesterda	1.44E+12	8/6/2011 18:50	submission	[HIV]	(STI)	1	0	1
38	3g668l	boynthetwild	askgaybros	Do you think this subreddit shou	1.44E+12	8/6/2011 20:36	submission	[HIV]	(STI)	1	0	1
39	3g7dme	apollo151	askgaybros	"The Normal Heart"	1.44E+12	8/7/2011 2:53	submission	[HIV]	(STI)	1	0	1
40	3g9t10	helpful_person	askgaybros	Lube suggestions?	1.44E+12	8/7/2011 15:26	submission	[Partner, Condoms]	(Product, Sexual Beha	0	1	1
41	3gaxdb	gaythrowaway65	gay	Had my first gay experience, wor	1.44E+12	8/8/2011 1:27	submission	[STD, Condoms, Oral Sex]	(Product, STI, Sexual I	1	1	1

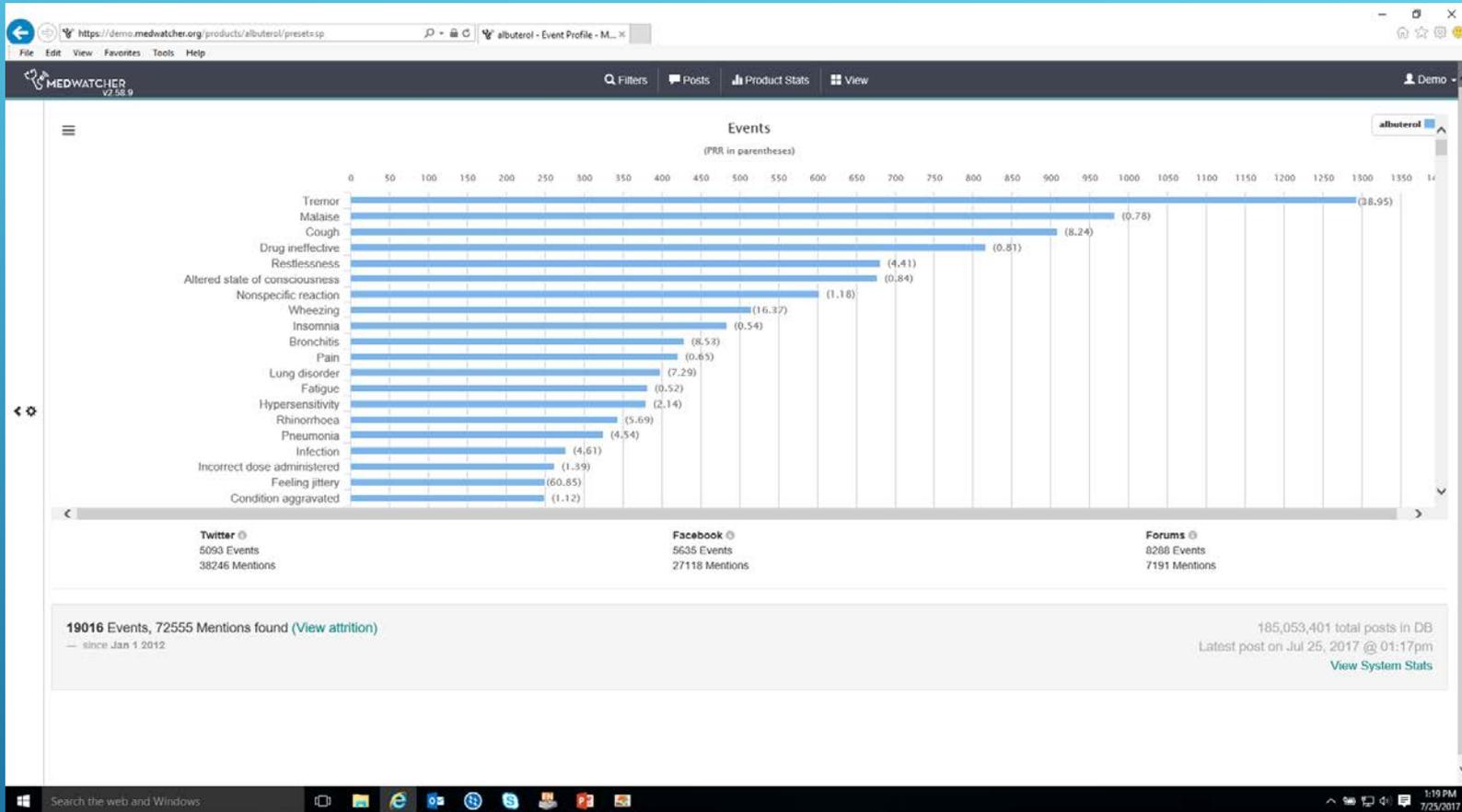
8363 POSTS FROM 6 SUBREDDITS  
 (2411 MENTIONS OF "PREP")  
 JAN 1-DEC 31, 2015

▶ <https://demo.medwatcher.org/>

DEMONSTRATION



# DEMO EXAMPLE: MEDWATCHER



# DEMO EXAMPLE: MEDWATCHER

- ▶ Use completed dashboard to explore the volume and content of conversations about PrEP and STIs
  - ▶ Attitudes, perceptions
  - ▶ Barriers, facilitators
  - ▶ Context of PrEP use (e.g. behaviors, sex partners)
- ▶ Evaluate potential utility of subreddits and search terms for future surveillance use
- ▶ Evaluate potential for network analysis (topical content) and subsequent qualitative analysis

## NEXT STEPS

- ▶ DSTDP
  - ▶ Kyle Bernstein
  - ▶ Sevgi Aral
- ▶ Epidemico
  - ▶ Michael Gilbert
  - ▶ Chi Bahk
- ▶ CDC Health Information Innovation Consortium

## ACKNOWLEDGEMENTS

# Building Capacity for Automated Phylodynamic Analysis

A look at the 2015 HIV Outbreak in rural Indiana

Elis Campbell, MS

Computational Biologist

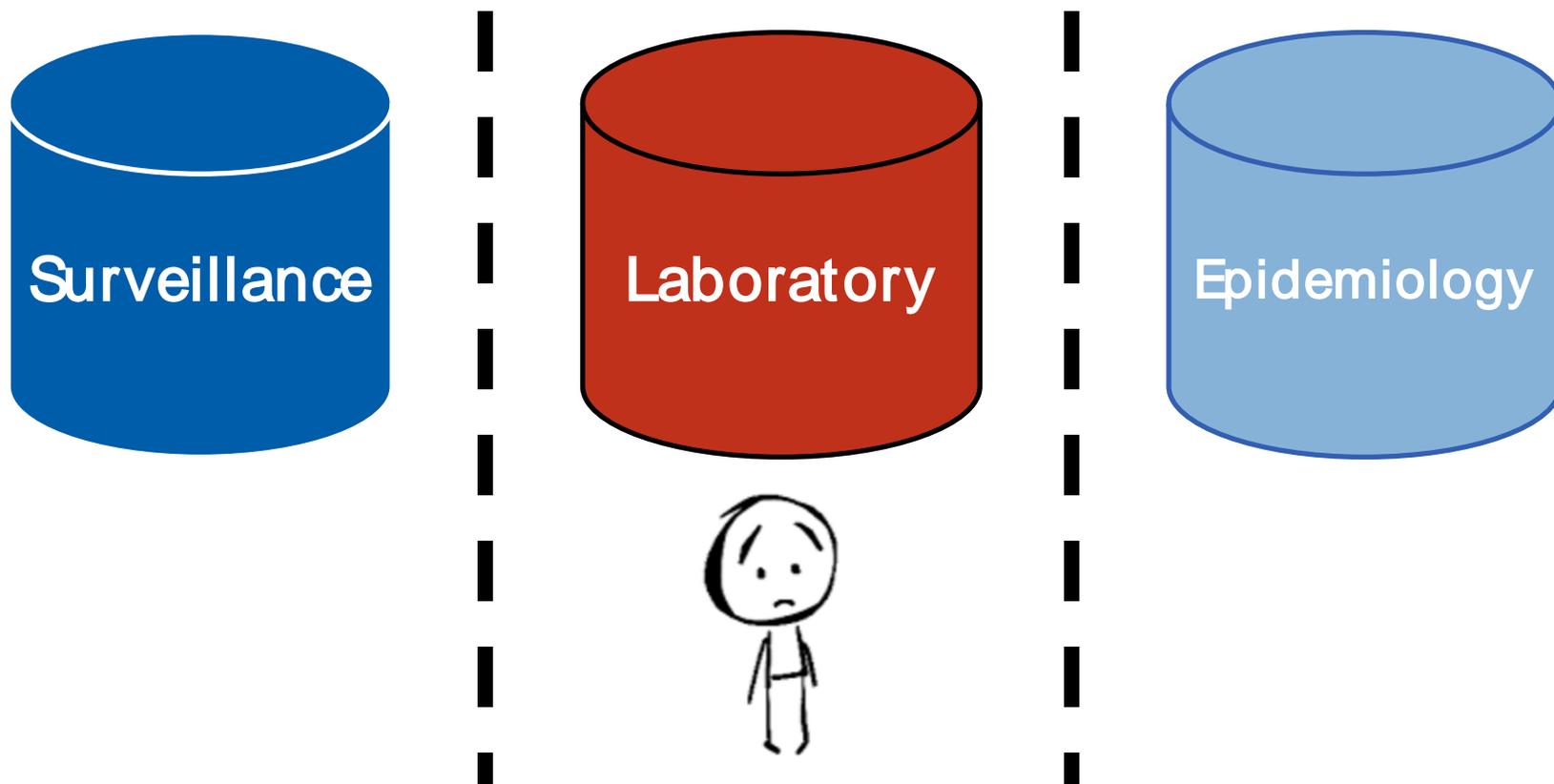
Division of HIV/AIDS Prevention

National Center for HIV/AIDS, Viral Hepatitis, STDs, and Tuberculosis Prevention

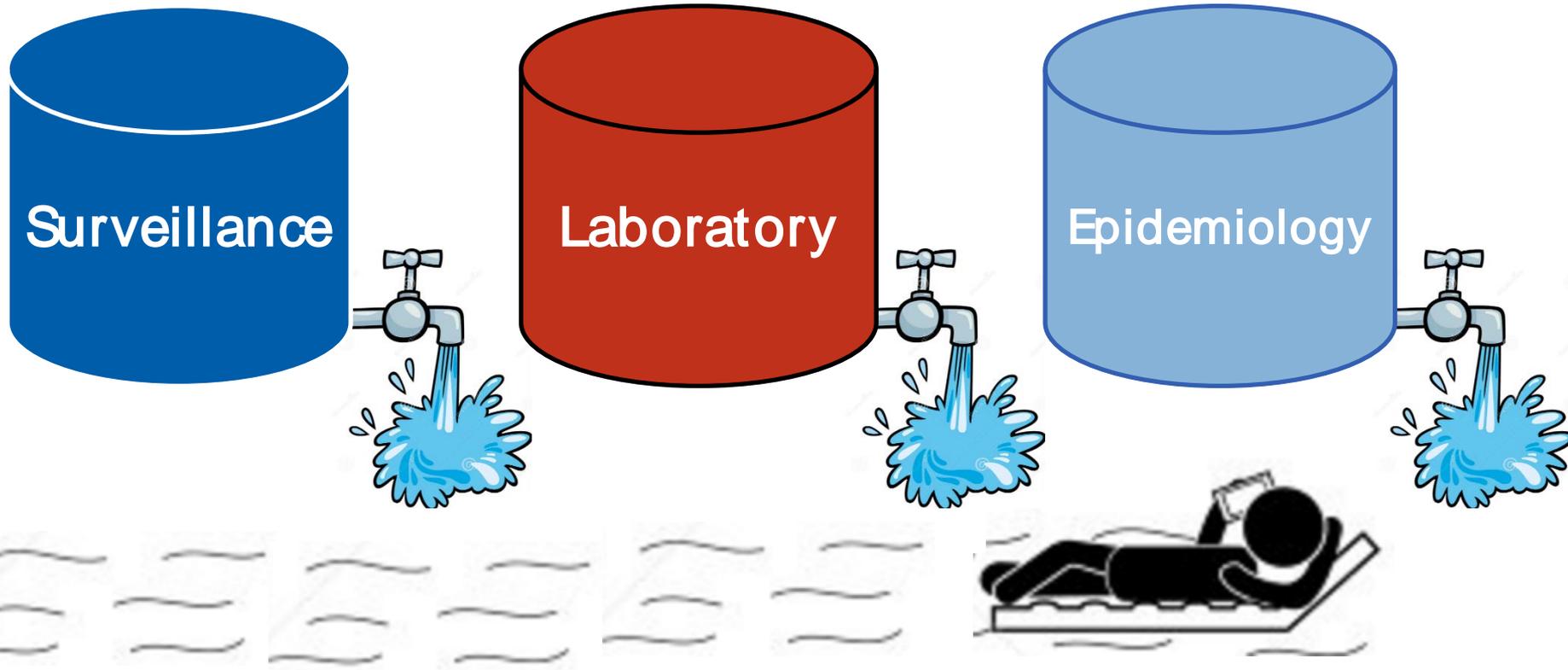
CDC Health Information Innovation Consortium

August 1<sup>st</sup>, 2017

## Program Perspective – Data Silos

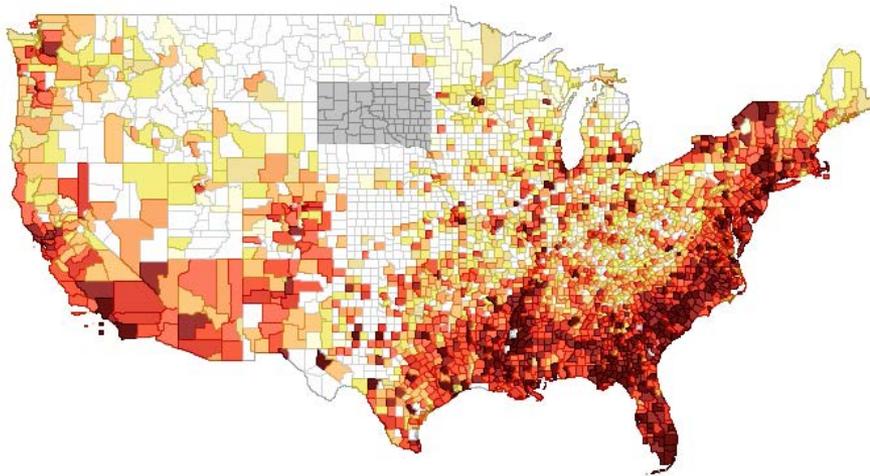


## Program Perspective – Data Lake

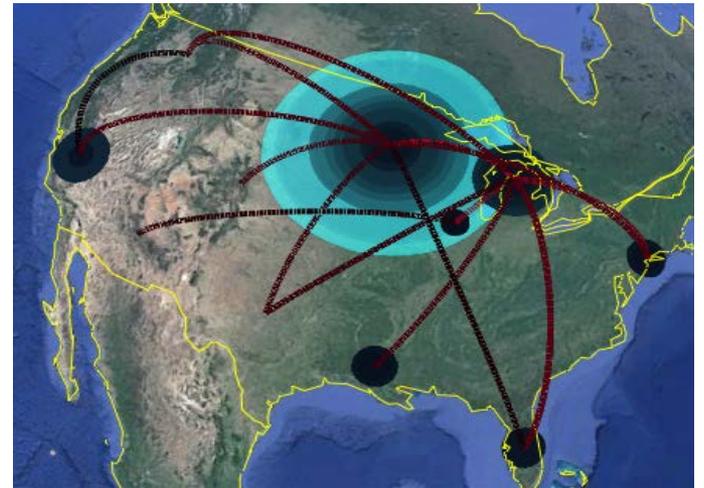


# Connecting the Dots

Incidence Surveillance



Molecular Surveillance



# HIV-TRACE – Quantifying difference between sequences

Independently align each sequence to a reference

Master Reference Sequence

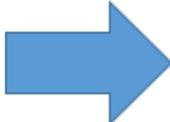


HXB2

Sample Sequence



HIV Sequences in FASTA file



Calculate distance between all possible pairs

Sequence A



Sequence B



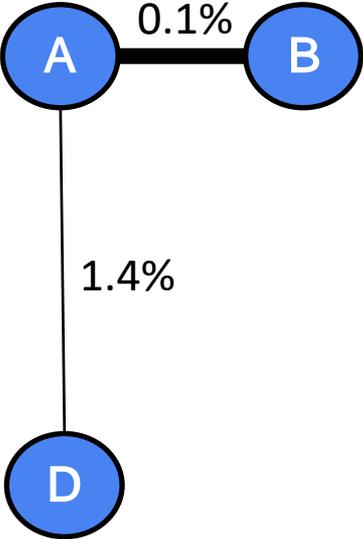
Threshold Filtration



ID1	ID2	Distance
A	B	0.001
A	D	0.014

# HIV-TRACE Results – Simple Example

ID1	ID2	Distance
A	B	0.001
A	D	0.014



## Why build transmission networks?

### Results directly comparable to traditional contact tracing data

- Can readily compare genetic distance and contact tracing

### Fine-grain resolution of transmission dynamics

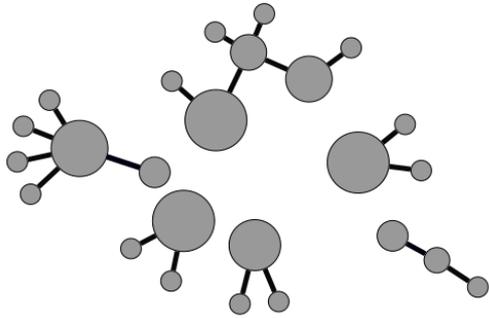
- Inform outbreak-specific intervention efforts
- Inform future prevention strategies

### Identification of *unreported* direct or indirect transmission links

- Aid for partner services and outbreak investigations

# Complementary Perspectives

Epidemiological



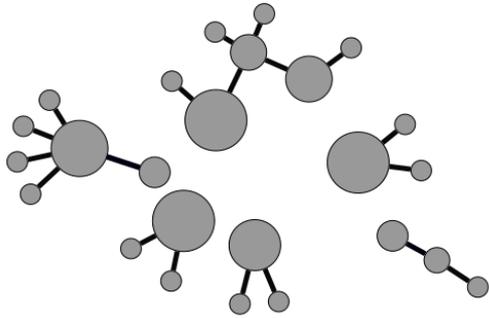
High-Risk  
Sexual  
Network

# Complementary Perspectives

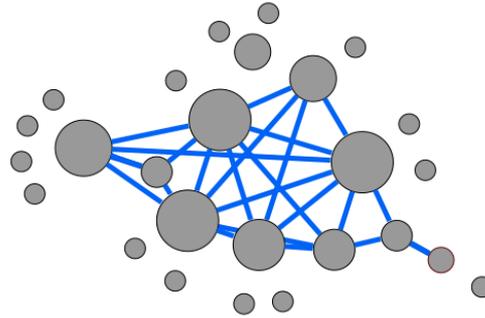
Epidemiological



Laboratory



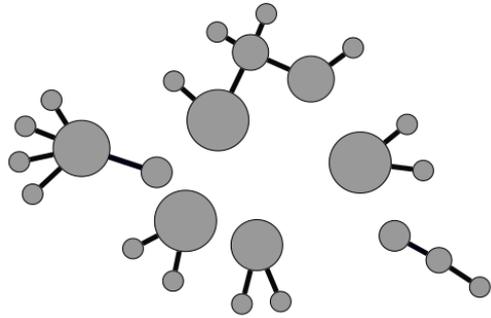
High-Risk  
Sexual  
Network



HIV Genetic Distance  
Network

# Complementary Perspectives

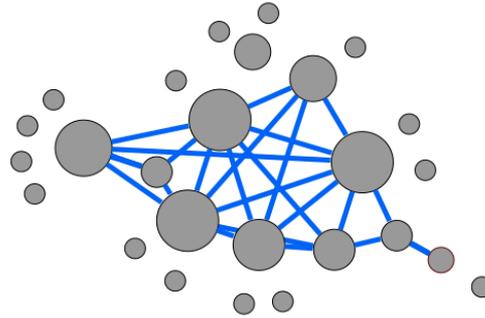
Epidemiological



High-Risk  
Sexual  
Network



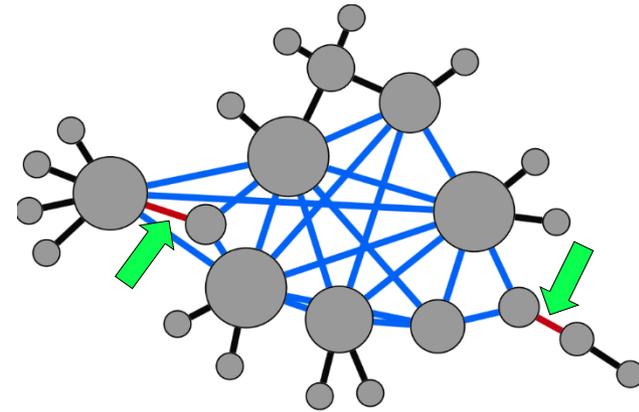
Laboratory



HIV Genetic Distance  
Network



Integrated



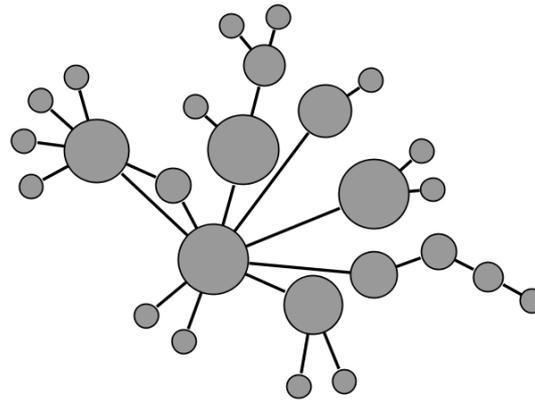
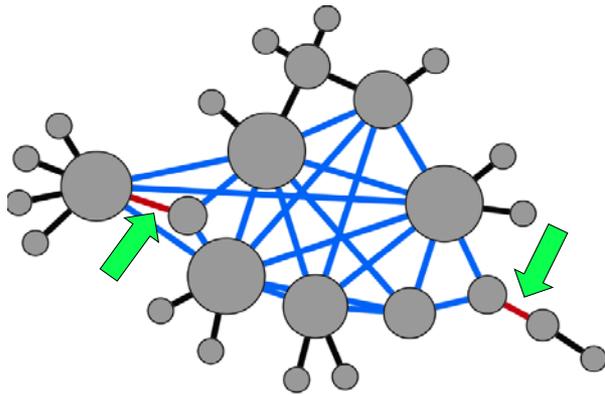
Both

# Analytics Enrichment

Integration



Informatics

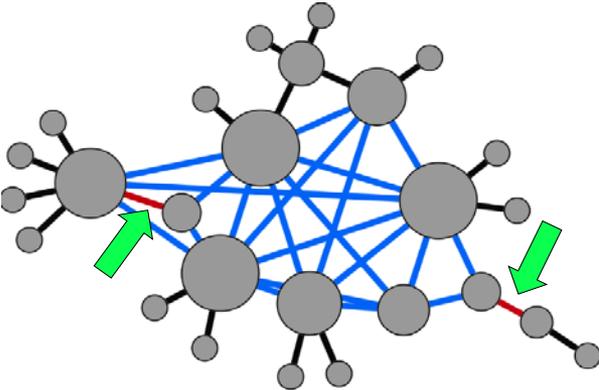


Both Contact & Genetic

Minimum Spanning Tree  
(MST)

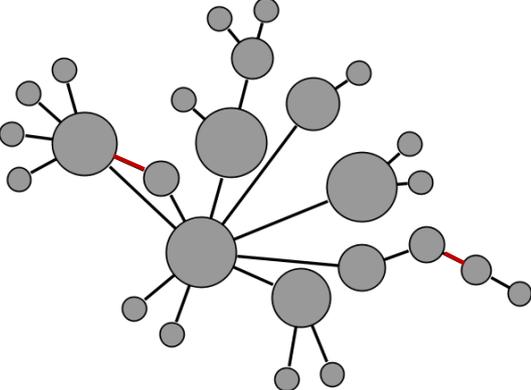
# Decision Support

Integration



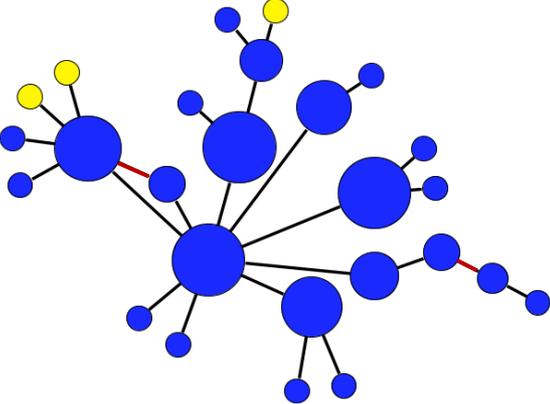
Both Contact & Genetic

Informatics



Minimum Spanning Tree (MST)

Clinical



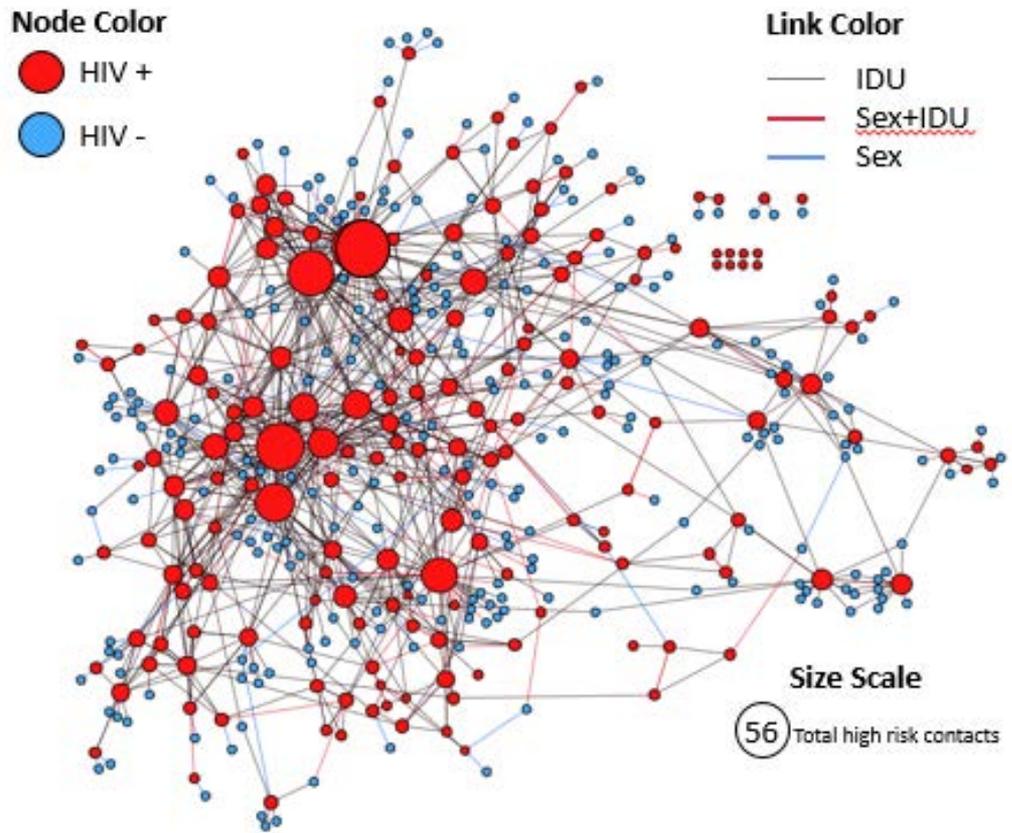
Recency of HIV Infection

- Established
- Acute

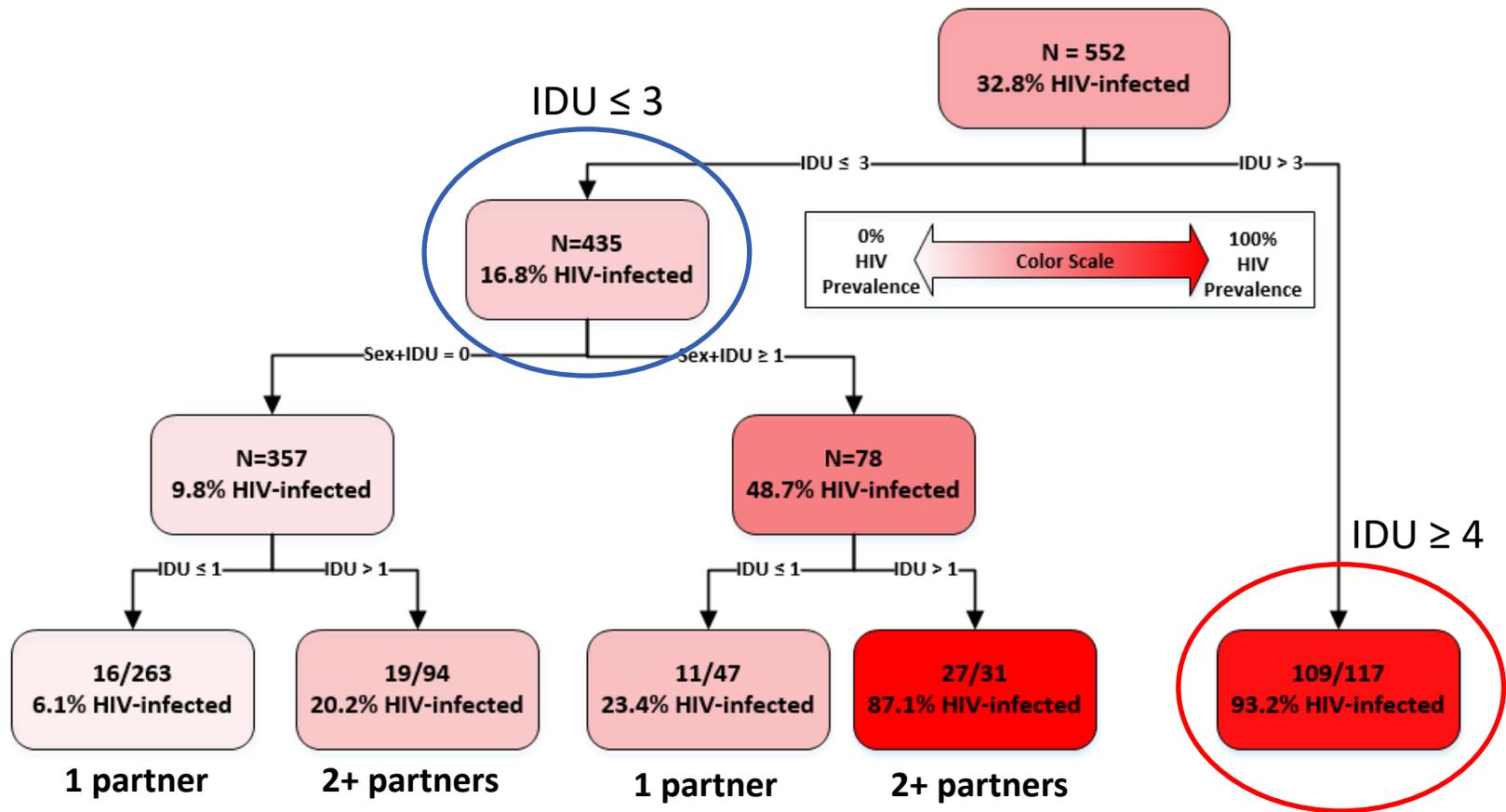
# Outbreak Response – HIV Outbreak among PWID in rural Indiana

## ■ Outbreak Details

- >220 HIV+ diagnoses
- >90% inject drugs
  - 4-15x/day
  - 1-6 partners/injection
- >90% HCV co-infection
- High prevalence of sex exposure

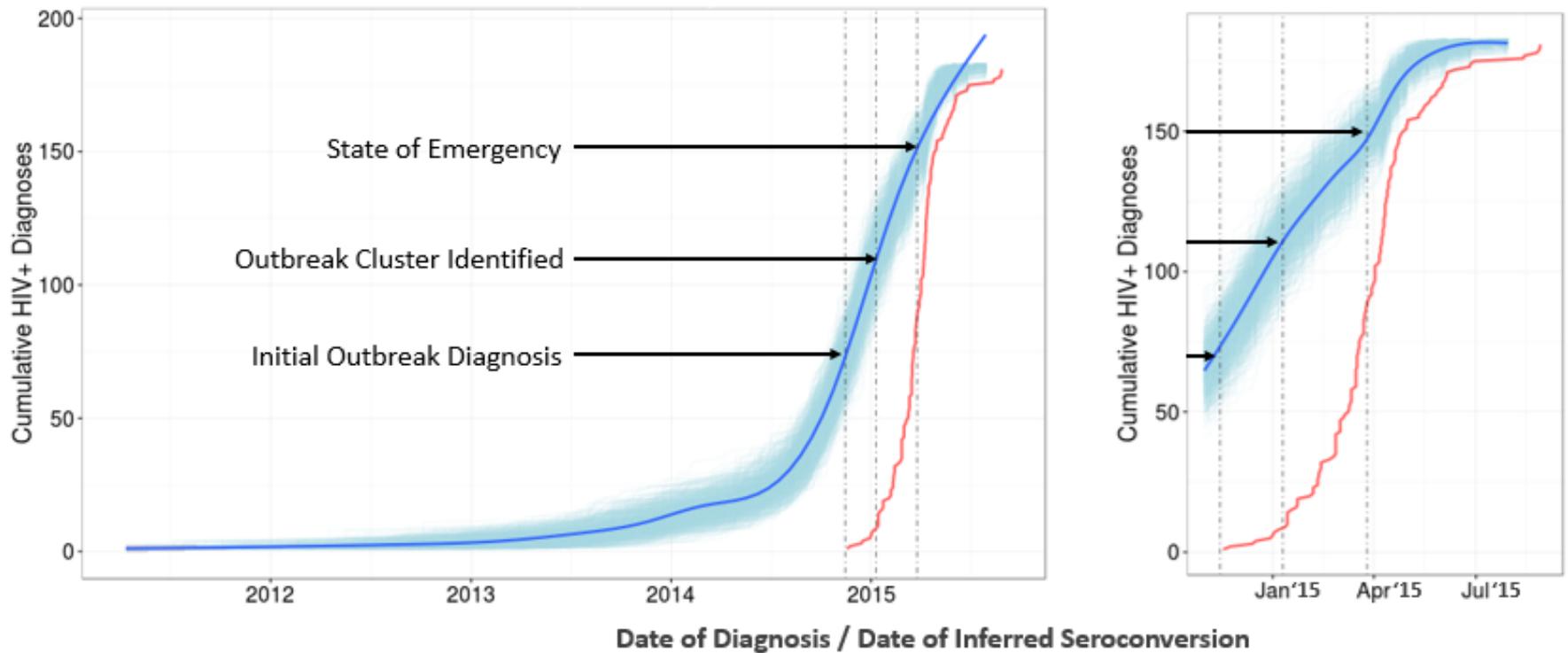


# Decision Support - Machine Learning - HIV Outbreak in Rural Indiana, 2015



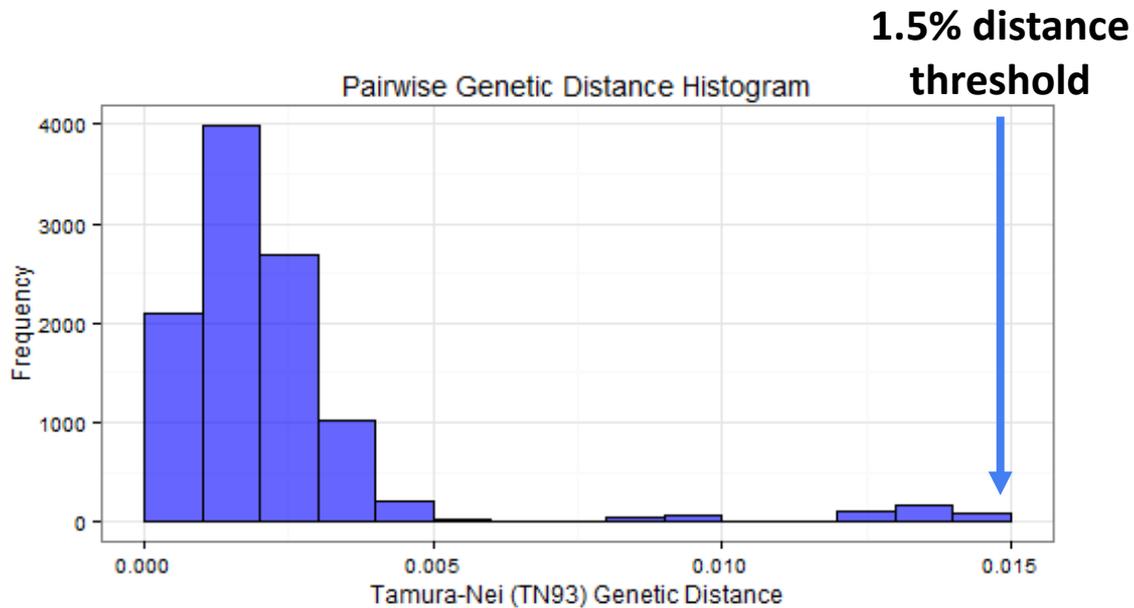
# Outbreak Simulations

## - HIV Outbreak in Rural Indiana, 2015

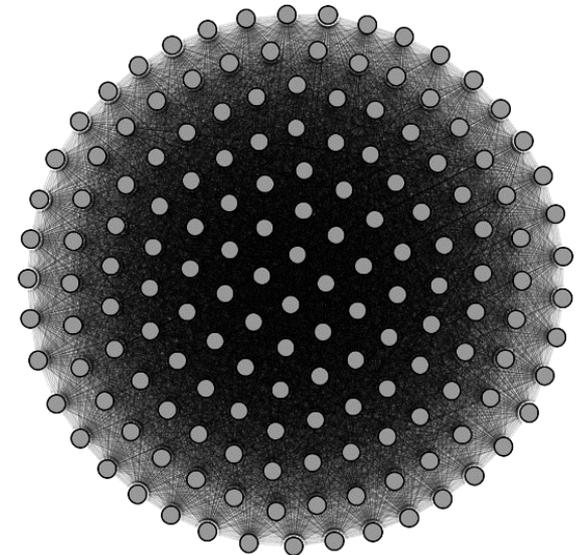


# Outbreak Response – HIV Outbreak among PWID in rural Indiana

- Most HIV sequences are highly similar, representing rapid and recent transmission



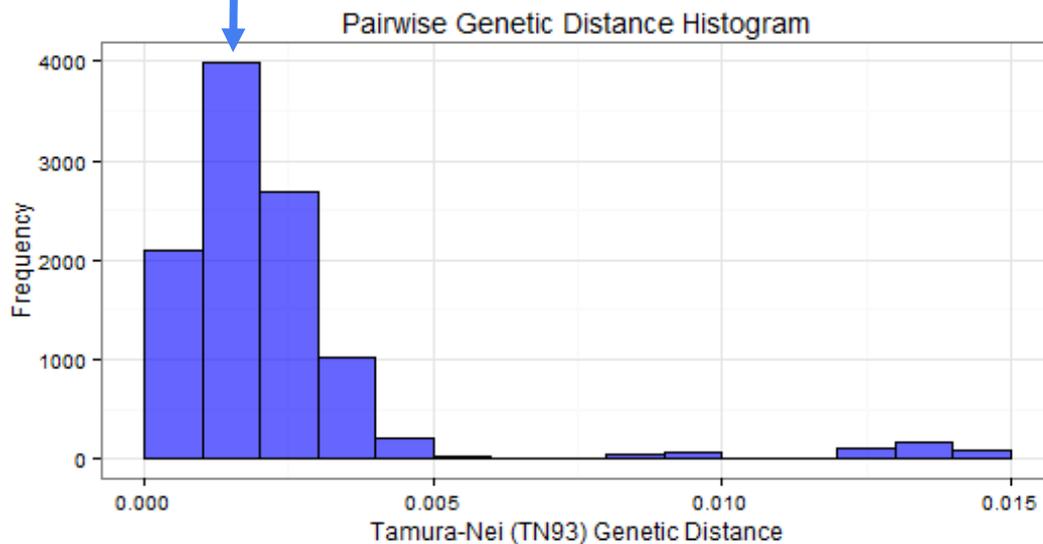
Genetic Distance Network  
polymerase region



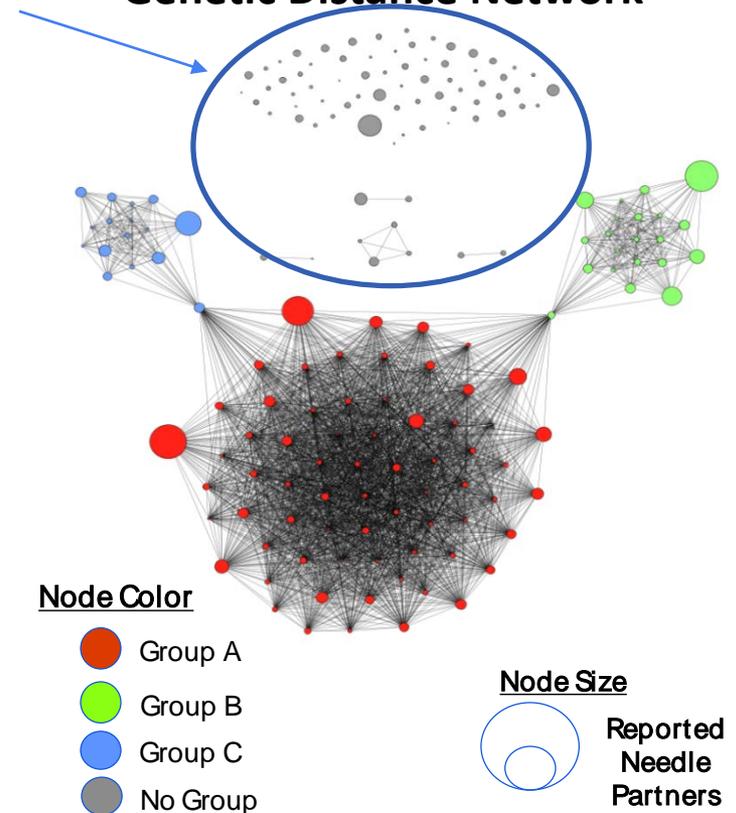
# Outbreak Response – HIV Outbreak among PWID in rural Indiana

- Pruning links with a lower threshold reveals community structure at cost of historical resolution
- Subgroups may be evidence of biological, social, temporal, and/or geographic factors

## 0.1% distance threshold

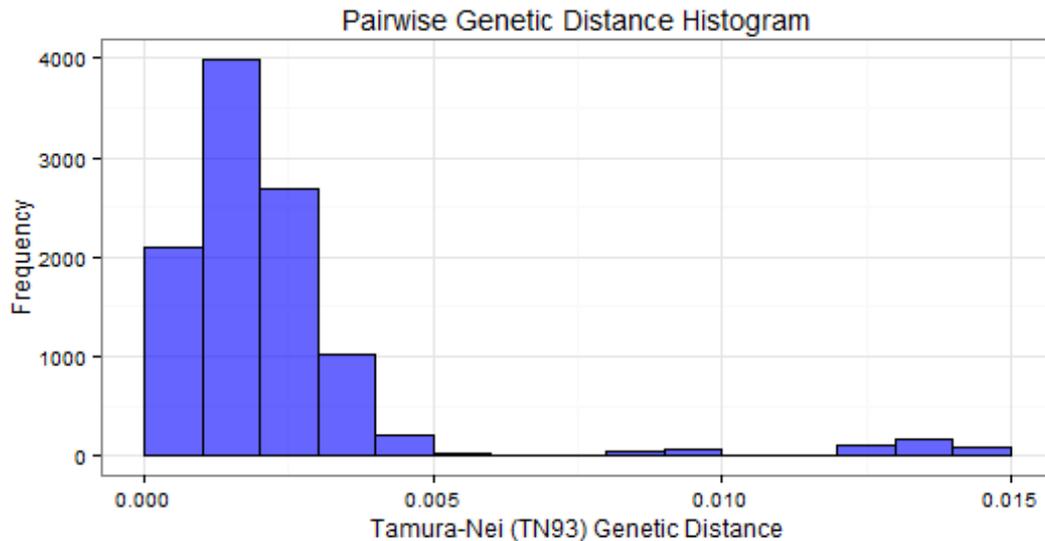


## Genetic Distance Network

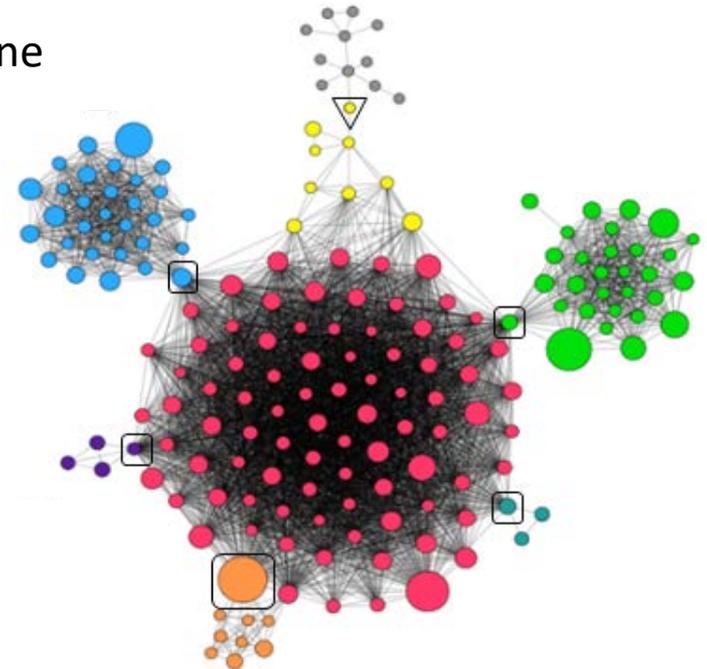


# Outbreak Response – HIV Outbreak among PWID in rural Indiana

- Minimum spanning trees (MSTs) select the most parsimonious links from the distribution of distances
- There can be many equally true MSTs, so we combine many unique MSTs to illustrate uncertainty

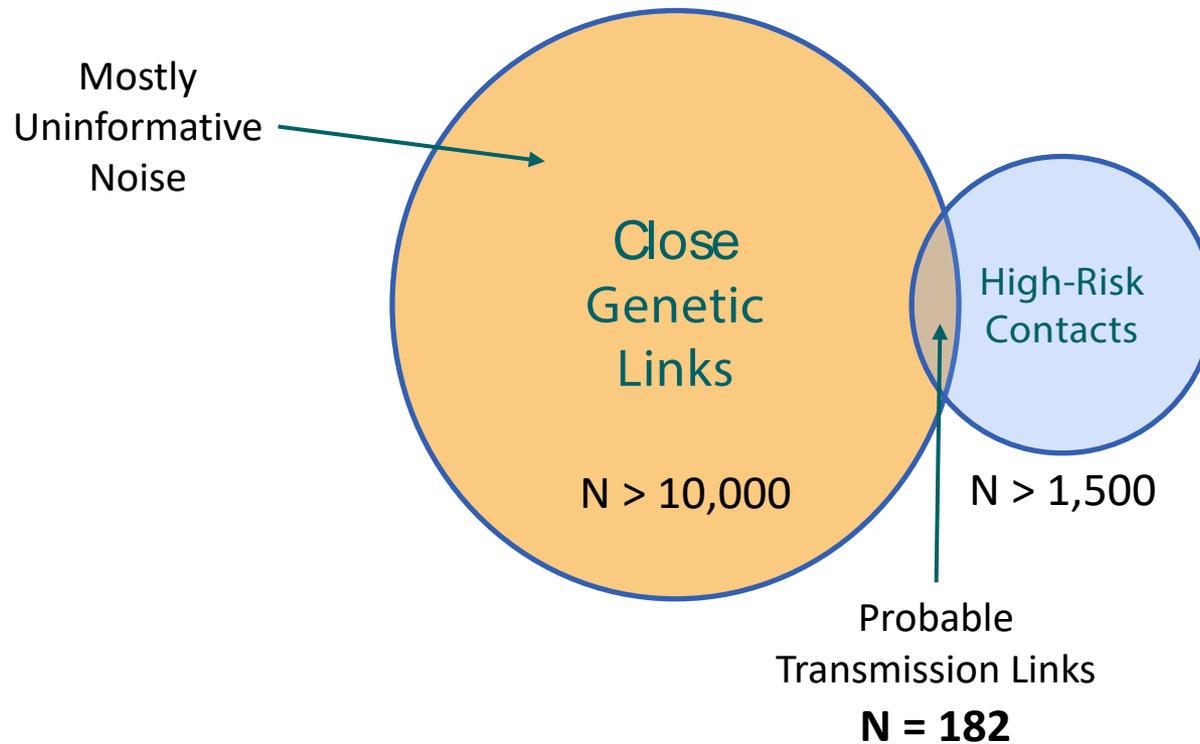


Composite of  
100 unique MSTs



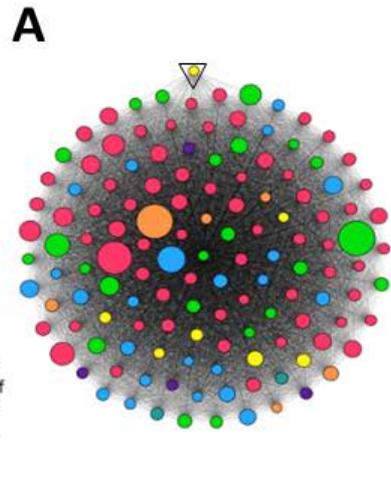
# Complementary Perspectives

## - HIV Outbreak in Rural Indiana, 2015

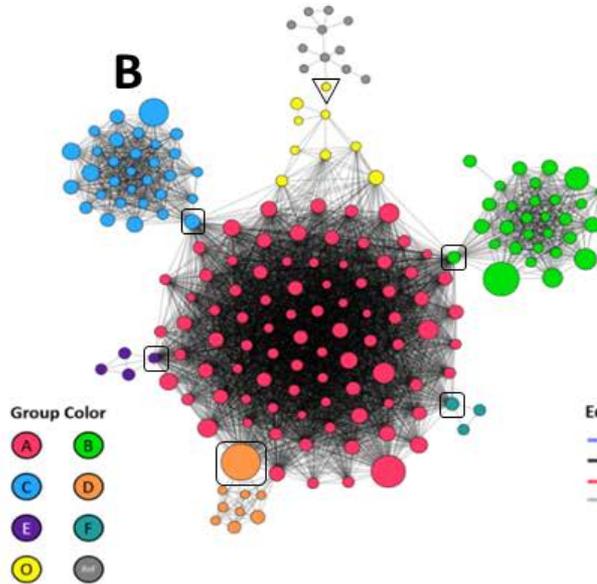


# Outbreak Response – HIV Outbreak among PWID in rural Indiana

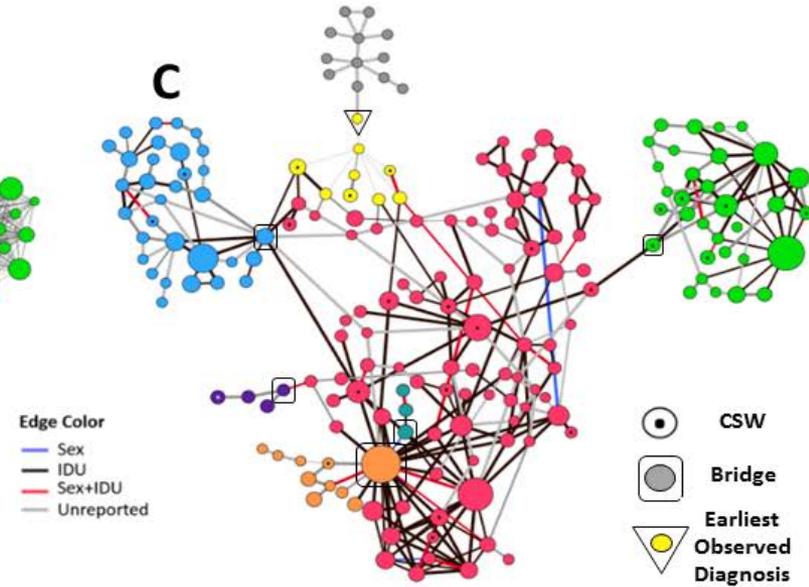
**Genetic Distance Network**  
1.5% distance threshold



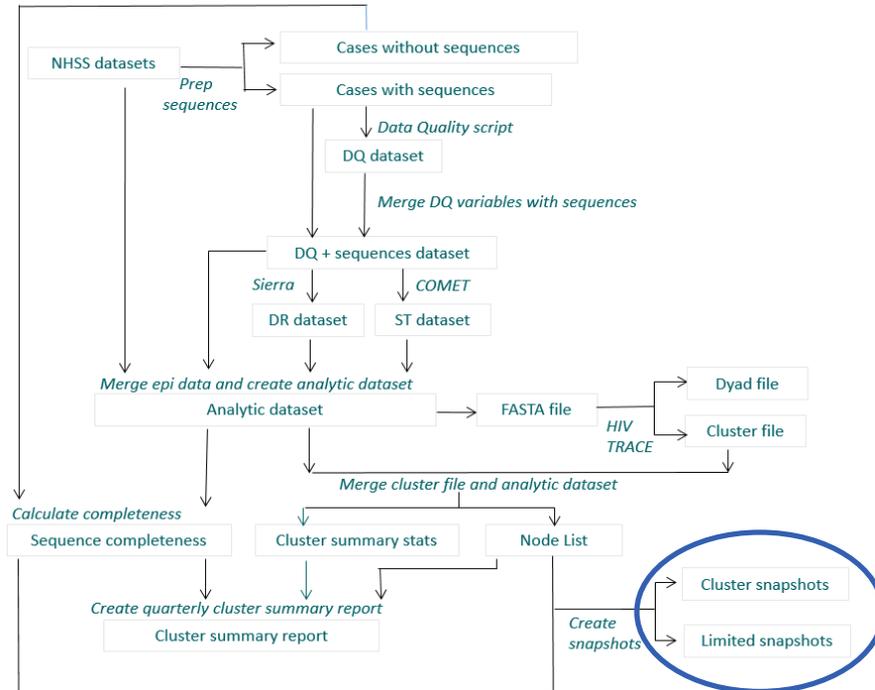
**Minimum Spanning Trees**  
Mean Distance <0.1%



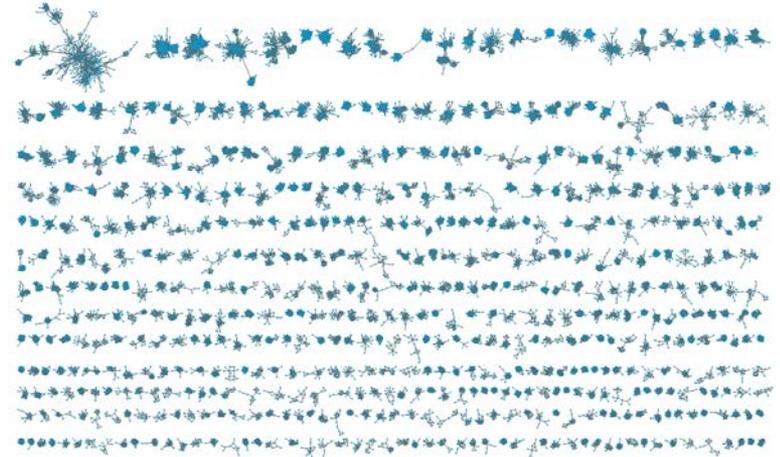
**Inferred Transmission Network**



# Integration is Complex and Resource Intensive



## HIV Surveillance Data Integration Workflow



PRIMARY JURISDICTION: OHIO      DIAGNOSIS DATES INCLUDED: 12/31/2012–12/31/2015

Table 1: Overall information about analysis and data completeness.

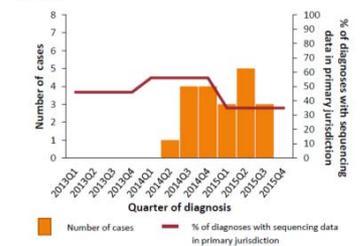
Dataset date:	December 2015	Date of Analysis:	April 2016	Threshold for cluster inclusion:	0.5%
% of diagnoses with HIV sequencing data available in primary jurisdiction, by year:					
2013:	46%	2014:	56%	2015:	35%
Complete lab reporting of CD4+ and VL for primary jurisdiction during 2013–2015? Yes					

Figure 1. Maps of cases by residence at diagnosis.



ted data

Figure 2. Epidemiologic curve of cases associated with cluster and percent of diagnoses with sequencing data by quarter, 2013–2015 (N = 20).



# Data Prep

## Trifacta Wrangler

- **Summary statistics**

- Interactive visualization<sup>1</sup>

- **Data cleaning**

- Guided by machine learning<sup>2</sup>
- Live preview of changes<sup>3</sup>

- **Optimized for collaboration**

- Shareable scripts<sup>4</sup>

- **Personal & Enterprise versions available**

<https://www.trifacta.com/start-wrangling/>

The screenshot displays the Trifacta Wrangler interface. At the top, it shows the data source 'amind5' with 75 columns and 3,282 rows. The main view is a data grid with columns: '#', 'id', 'ARC', 'uppercase\_Tribe\_name', 'Collection\_period', 'Collection\_period!', 'ARC', and 'Tribe\_Abbrev'. A summary statistics panel is visible on the left, showing histograms for '#', 'Collection\_period', and 'Collection\_period!'. A 'SUGGESTIONS' panel at the bottom shows three suggestions: 'Extract on: {digit}{4}', 'Replace on: {(start){digit}{4}(end)}', and 'Countpattern on: {digit}{4}'. Each suggestion includes a live preview of the data changes.

4

```
SCRIPT
splitrows col: column1 on: '\n' quote: '\"'
split col: column1 on: ',' Limit: 19 quote: '\"'
replace col: * on: '\"' with: '' global: true
header
settype col: CD_Ethnicity type: 'string'
settype col: ID_Profile type: 'string'
```

# Exploratory Analysis Alpine Chorus

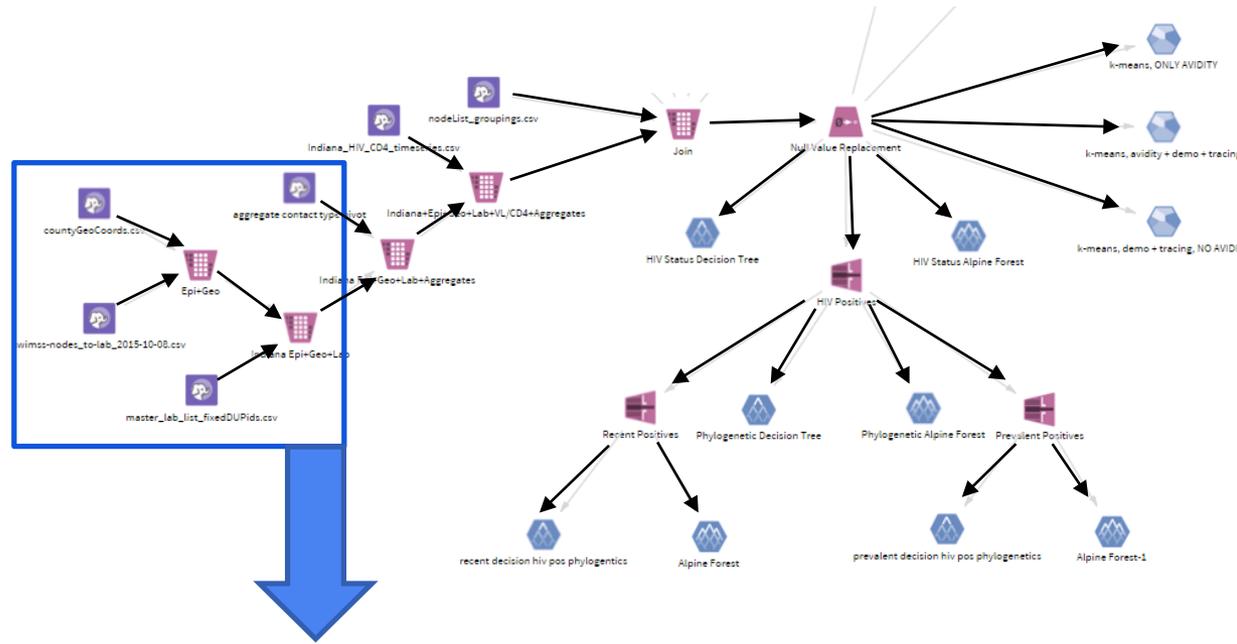
Powerful collaborative and analytics tools via web forms

Intuitive workflow editor with drag and drop “operators”

Built-in analytics (decision trees, linear regression, PCA, k-means, and dozens more)

Custom code turned into drag & drop operators (R, Python, Java)

Open source and Enterprise



COLUMNS FOR MATCHING ROWS FROM EACH TABLE

[add condition](#)

epi person list

master inventory list

CD\_Gender [chararray]

=

gender [chararray]

[delete](#)

age [int]

=

age [int]

[delete](#)

swimmsID [chararray]

=

id\_profile [chararray]

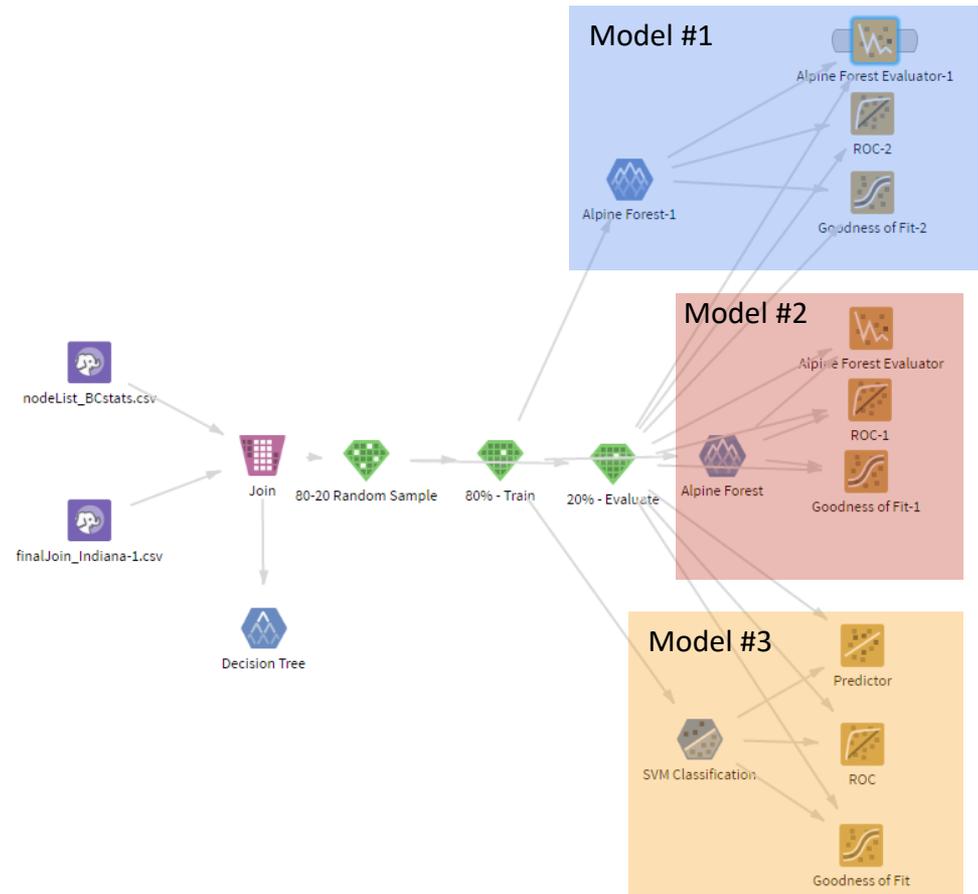
[delete](#)

[github.com/alpinedatalabs](https://github.com/alpinedatalabs)

# Automated Solution

Enables reproducible machine learning analytics

...without the programmer!



# Network Visualization

## Centrifuge - Enterprise

Network visualization, exploration, and animation

### Dynamic model generation

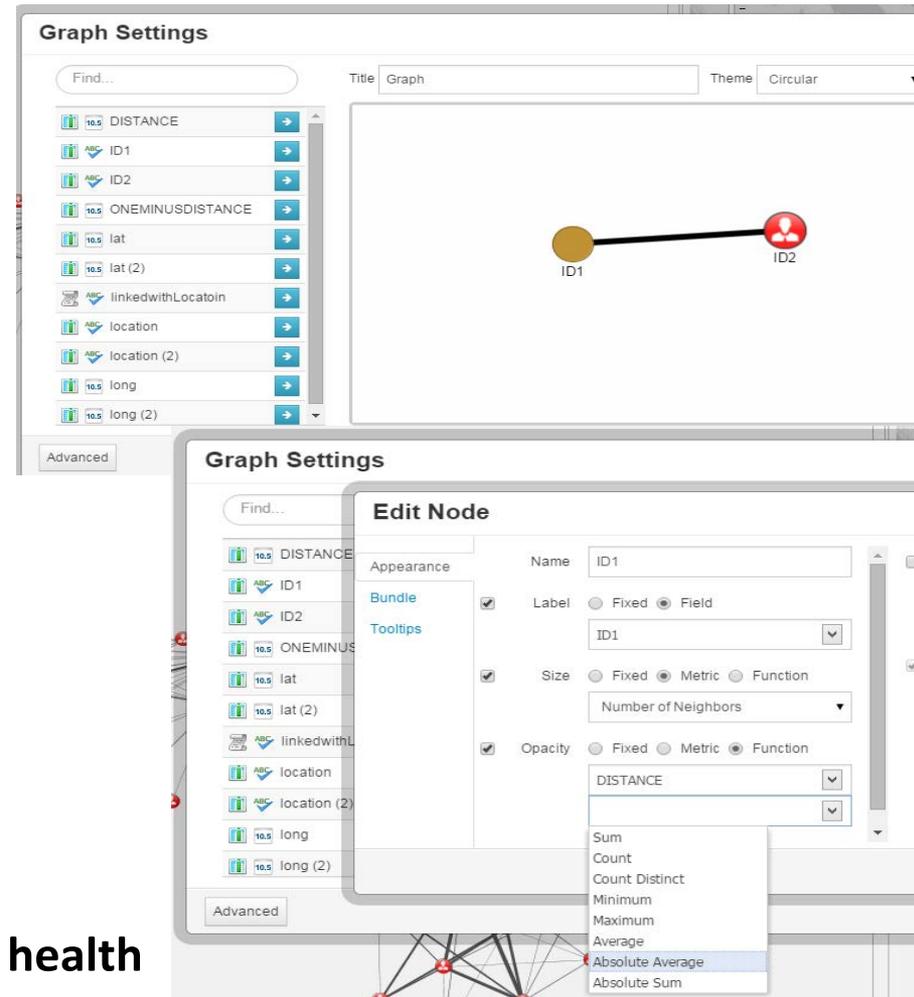
- Person  $\leftrightarrow$  Person
- County  $\leftrightarrow$  County

### Variables mapped via dropdown

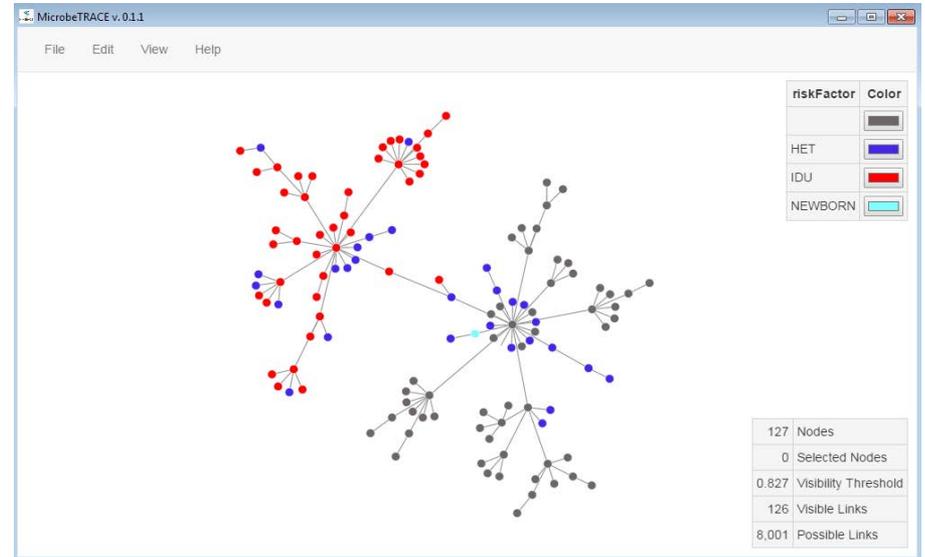
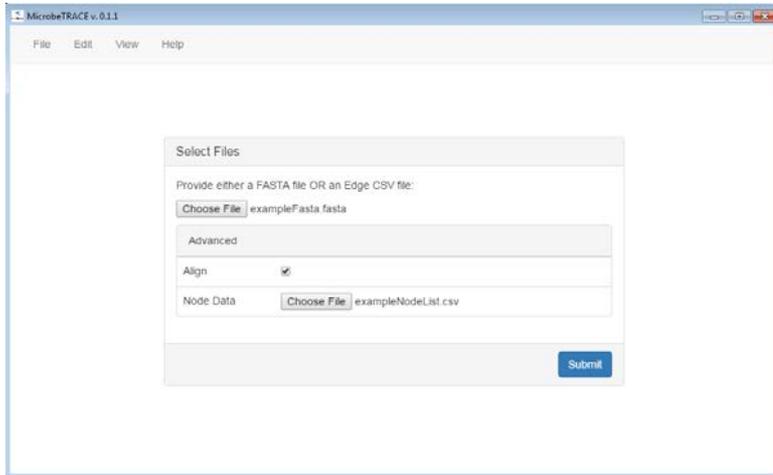
- Shape = Risk Factor
- Color = Infection Status
- Size = # of Reported Partners
- Link width = genetic distance

### Pattern Recognition

Low-cost 'lite' version on the way for public health



# MicrobeTRACE



Sequences + Epidemiologic Data



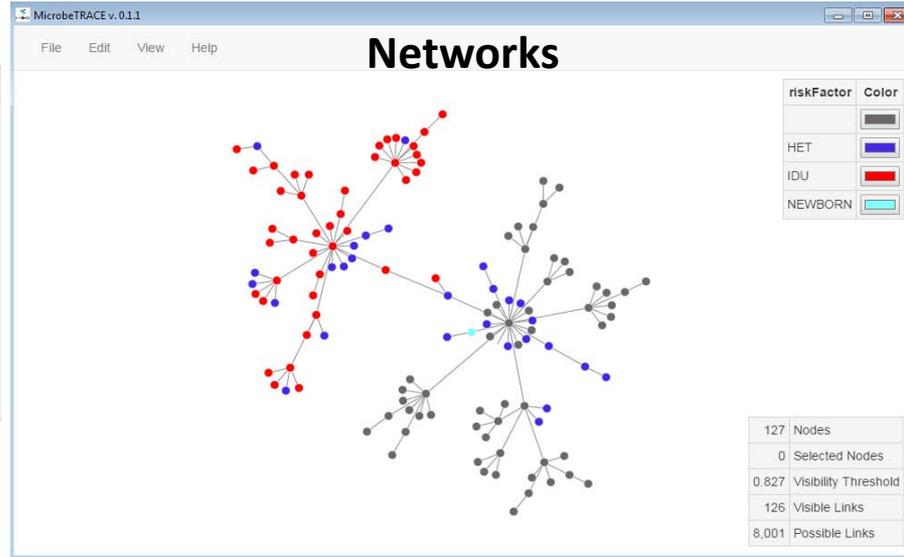
Interactive Exploration of  
Transmission Networks

[github.com/cdcgov/microbetrace](https://github.com/cdcgov/microbetrace)

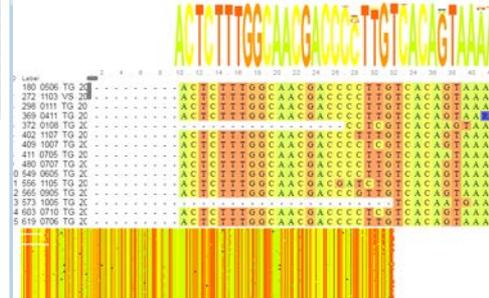
# MicrobeTRACE



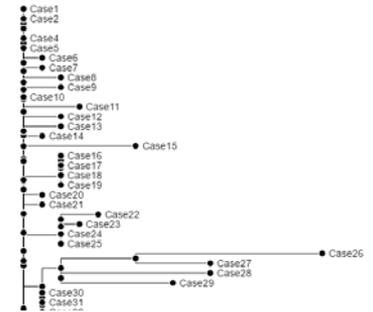
Distance Histogram



## Alignment Viewer



## Tree Viewer



Maps

[github.com/cdcgov/microbetrace](https://github.com/cdcgov/microbetrace)

# Potential for Enhancement

## **Business**

Partner with vendors for licensing innovation

- Public health “ROI” doesn’t translate to increased budget

## **Admin**

Network administration and security must catch up and keep pace

## **Technology**

Dynamic redaction of PII

- Submit necessary data for automated analysis
- Rapidly integrate sensitive data locally

# On the Shoulders of Giants

## Indiana

- Indiana State Department of Health Scott County Health Department
- Clark County Health Department
- Foundations Family Medicine
- Indiana University School of Medicine, Division of Infectious Diseases
- University of Louisville School of Medicine, Division of Infectious Diseases
- NCHHSTP, Division of HIV/AIDS Prevention NCHHSTP, Division of Viral Hepatitis NCEZID, Division of Scientific Resources, Specimen Management Branch and Biotech Core Facility
- Epidemic Intelligence Service

## Surveillance

- HIV Incidence and Case Surveillance Branch (HICSB)
  - Molecular HIV Surveillance
- State, Territory, Local, and Tribal Public Health Departments

## STOP Study

- NYC, NC, and SF public health
- NCHHSTP, Division of HIV/AIDS Prevention (DHAP)

## Commercial Laboratories

- Quest
- LabCorp

## Software and Support

- SciComp, ITSO and OCISO at CDC
- Leidos Inc.
  - Centrifuge Systems
  - Trifacta
  - Alpine Data Labs
- UCSD Viral Evolution Group (VEG)
- Stanford Visualization Group
- Seattle Interactive Data Lab
- Stanford HIV Drug Resistance DB Team
- Los Alamos National Laboratories
- Gephi Consortium

## Items of interest

- Data Visualization Scientific Advisory Group [\[link\]](#)
- CDC GitHub organization - <https://github.com/CDCgov/>
- CodeForAtlanta – 8/13 – [\[link\]](#)
- CodeDay – Nov 11/12 – Registrations up [\[link\]](#)
- CDC Tracking Network Enviro Health App Challenge <http://www.envirohealthchallenge.com/> - winners Sep 1
- Next Forum – November 7, 10am-11am ET- Share ideas you would like to have covered
- Please take 90 seconds to share your feedback, <https://epiinfowebsurvey.cdc.gov/Home/aa515f48-55f7-4a14-9b82-1a5bd2f96124>. We'd love to hear from you on improving the quarterly forums.



# Thank You & Questions

**For more information please contact Centers for Disease Control and Prevention**

1600 Clifton Road NE, Atlanta, GA 30333

Telephone: 1-800-CDC-INFO (232-4636)/TTY: 1-888-232-6348

Visit: [www.cdc.gov](http://www.cdc.gov) | Contact CDC at: 1-800-CDC-INFO or [www.cdc.gov/info](http://www.cdc.gov/info)

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.



Centers for Disease Control and Prevention  
Office of Public Health Scientific Services