

# Modeling Spread of Disease from Social Interactions



UNIVERSITY *of*  
ROCHESTER

Adam Sadilek  
Henry Kautz  
Vincent Silenzio





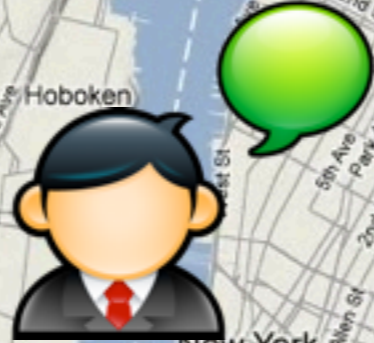
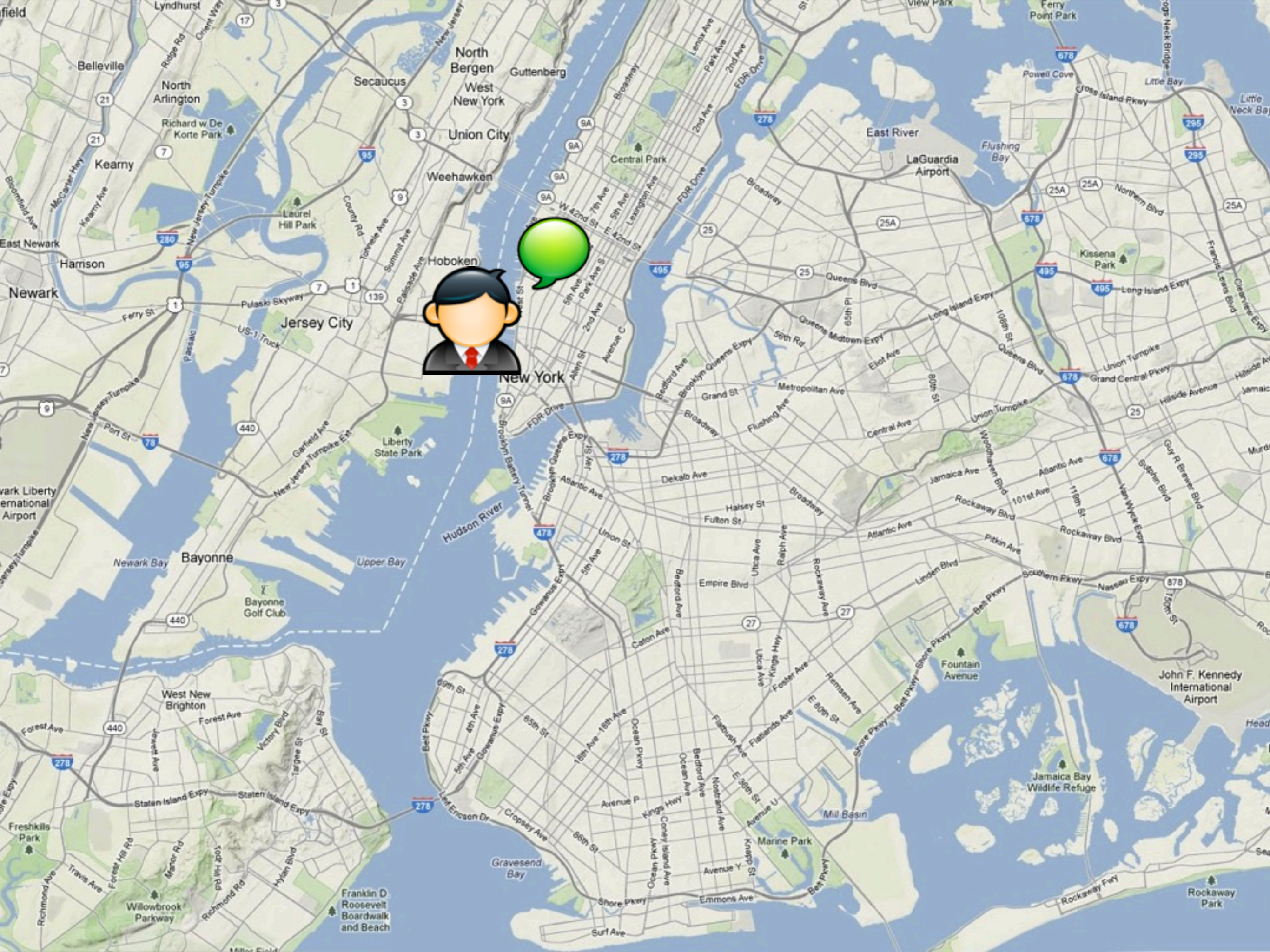












North Bergen  
West New York

Union City

Weehawken

Hoboken

Jersey City

New York

East River

LaGuardia Airport

John F. Kennedy International Airport

Jamaica Bay Wildlife Refuge

Franklin D. Roosevelt Boardwalk and Beach

Newark

Liberty International Airport

Freshkills Park

Willowbrook Parkway

West New Brighton

Bayonne

East Newark

Kearny

Belleville

field

Richmond Ave

Forest Ave

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

Travis Ave

Forest Hill Rd

440

Newark Bay

Passaic

Harrison

Richard w De Korte Park

North Arlington

Lyndhurst

Miller Field

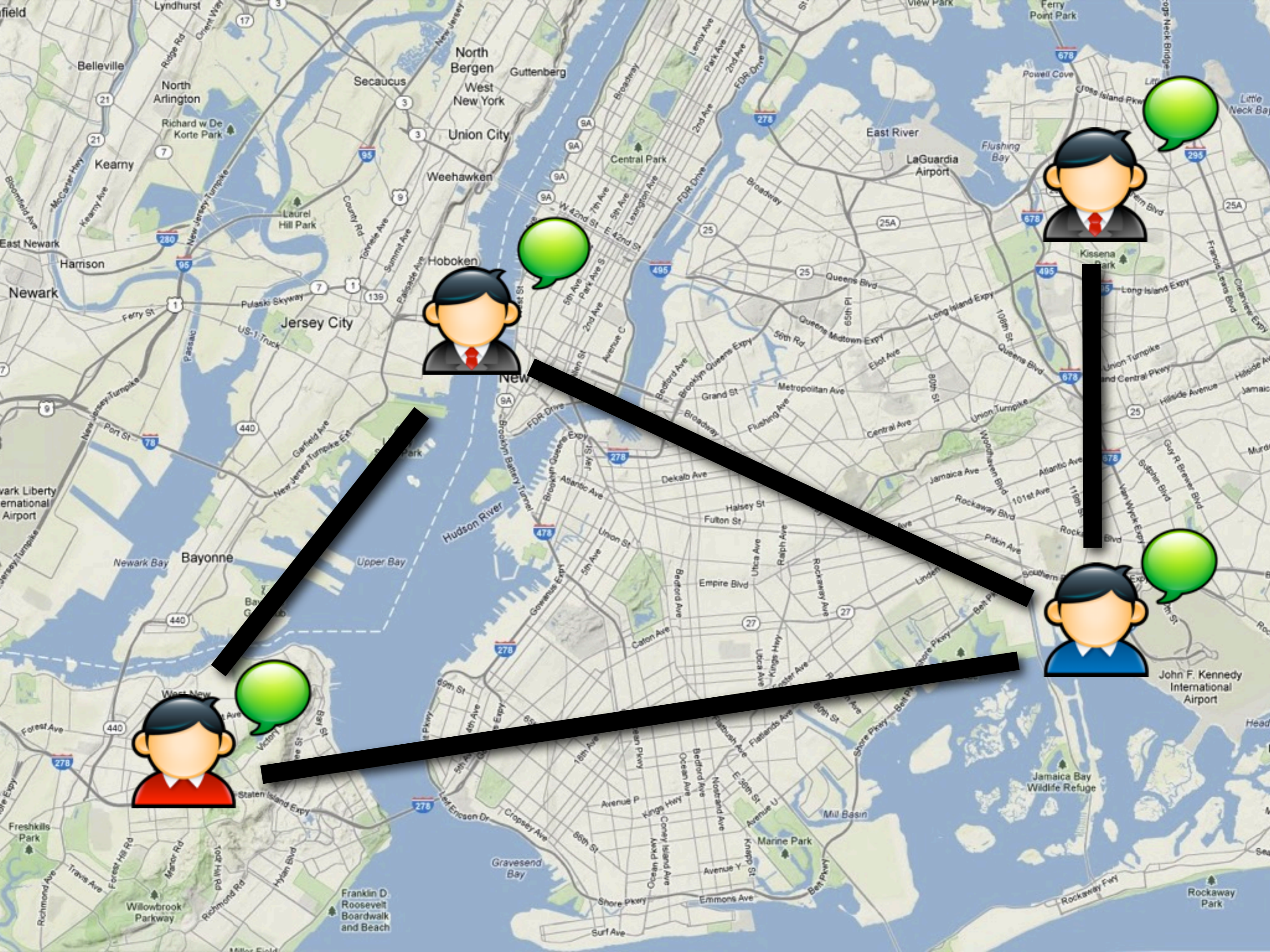
Travis Ave

Forest Hill Rd



















# The Data

- **New York City**

# The Data

- **New York City**
- **16M tweets** over a month



# The Data

- **New York City**
- **16M tweets** over a month
  - **0.6M unique users**

# The Data

- 6K **geo-active** users

# The Data

- 6K **geo-active** users
- 2.5M **tweets** by geo-active users

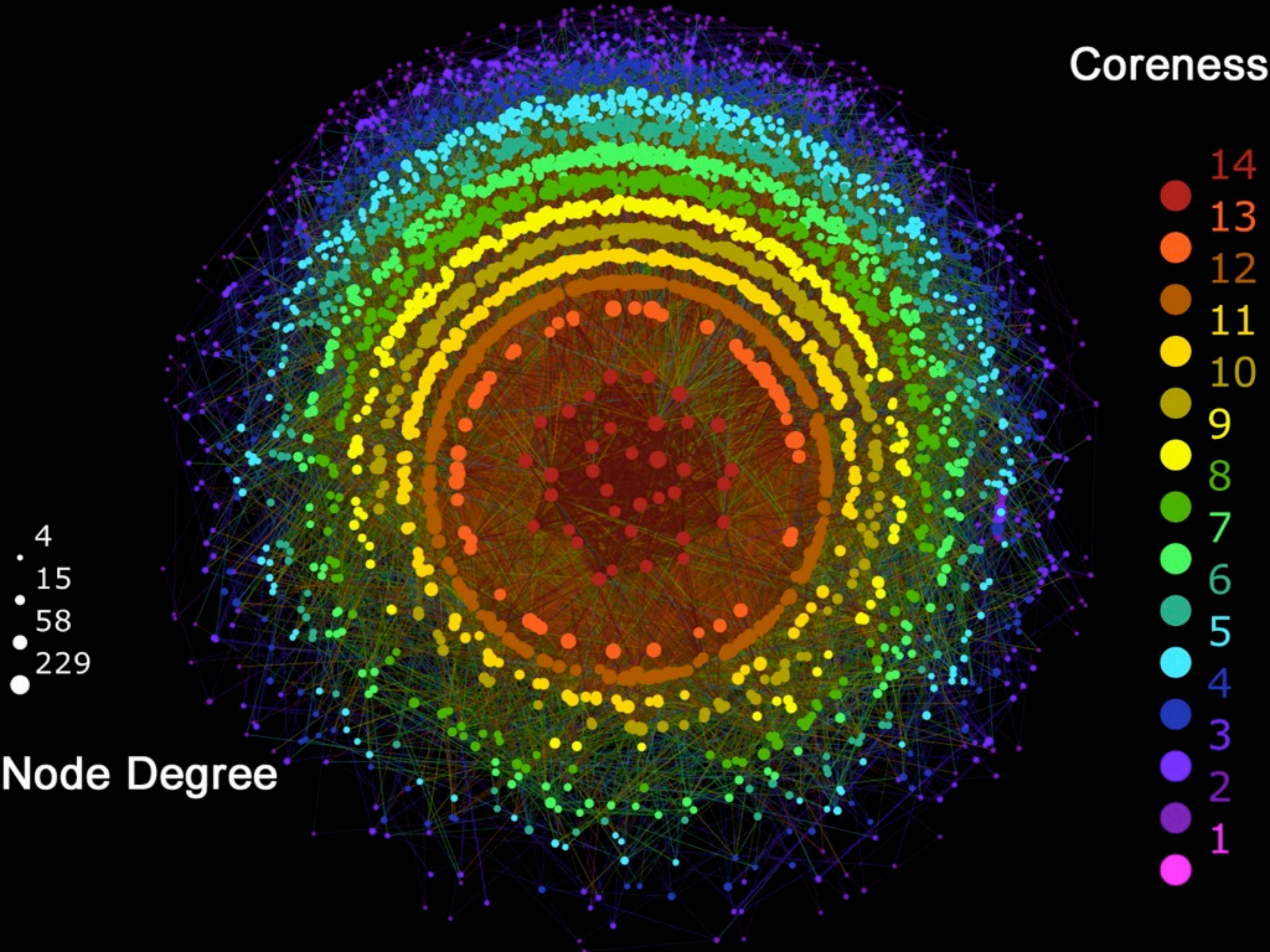
# The Data

- 6K **geo-active** users
- 2.5M **tweets** by geo-active users
  - 103K “**follows**” relationships

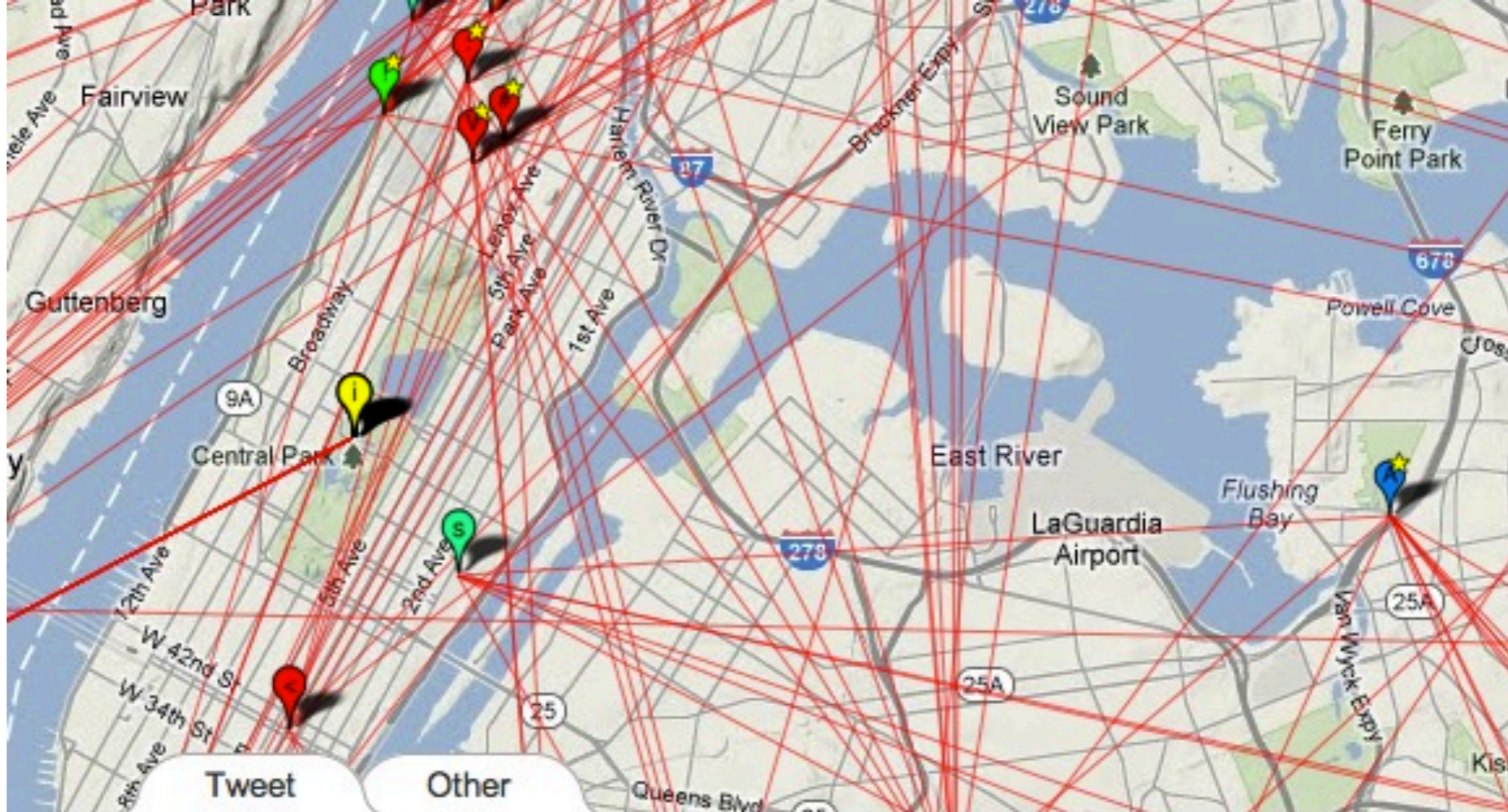
# The Data

- 6K **geo-active** users
- 2.5M **tweets** by geo-active users
  - 103K “**follows**” relationships
  - 32K “**friends**” relationships



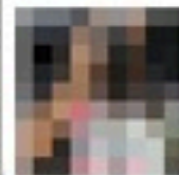






Tweet

Other



(2010-05-18 20:09:58)

“feeling horrible.. this flu had me up till 3am..... Jesus!”



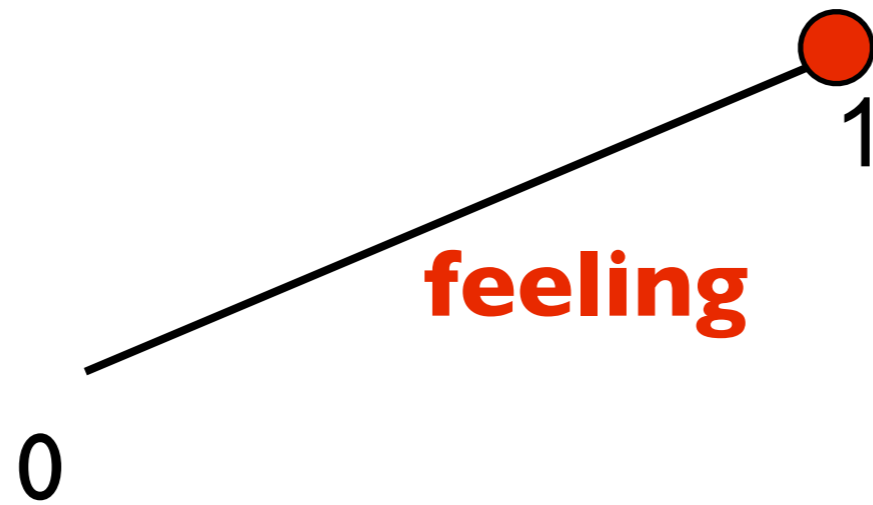


# Spread of Disease

- **Identify** people with **symptoms**
- **Quantify** the impact of:
  - Co-location
  - Social ties
- **Model contagion** with fine granularity

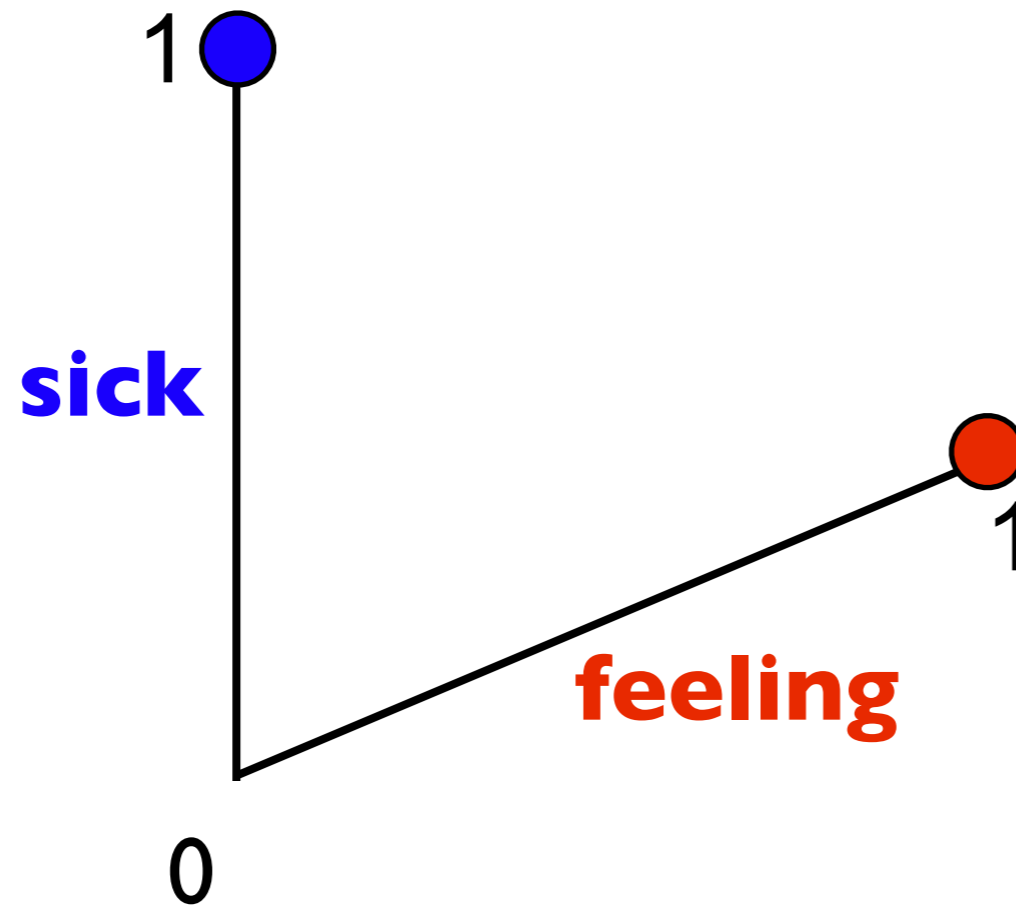
“feeling sick”

“feeling sick”

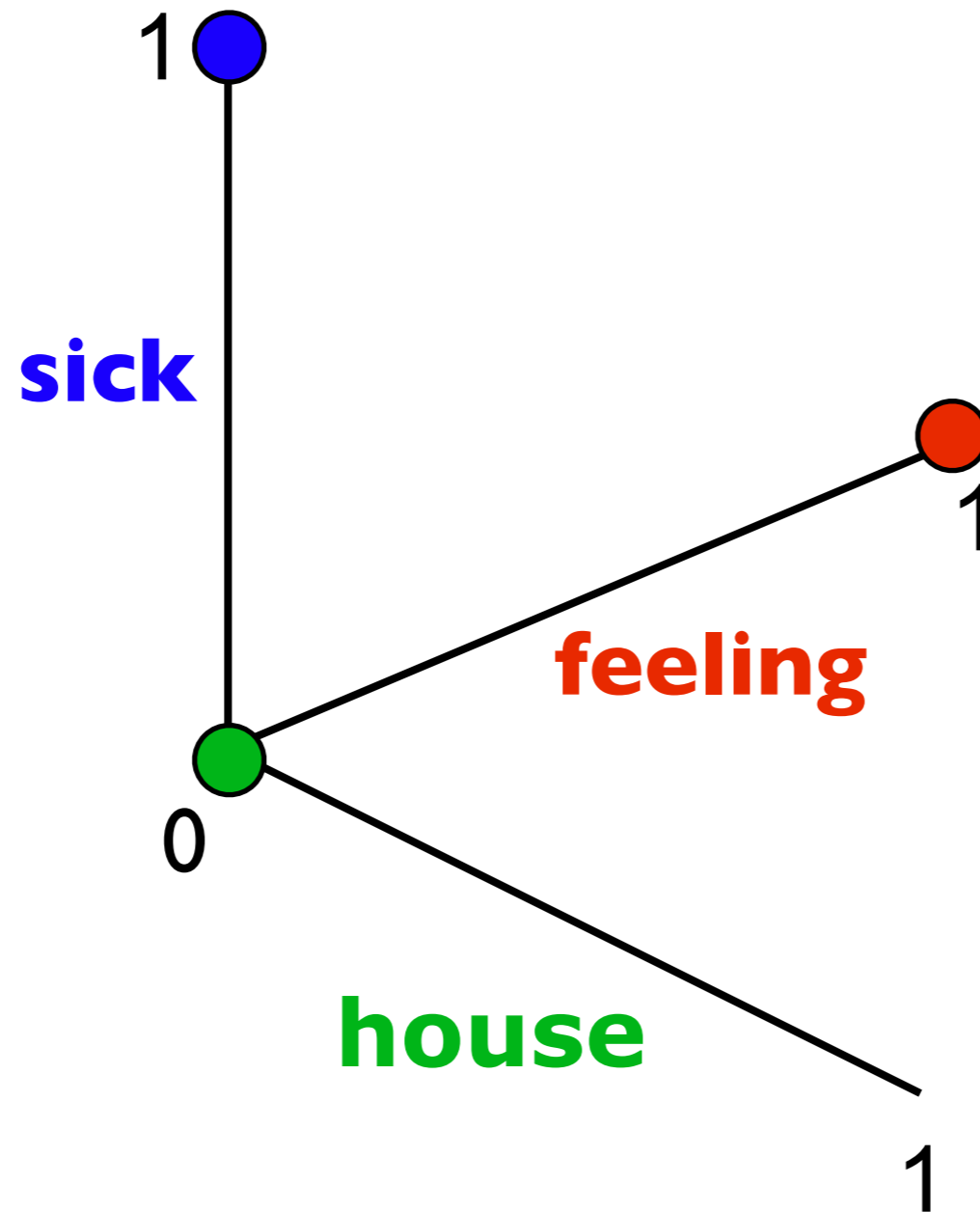




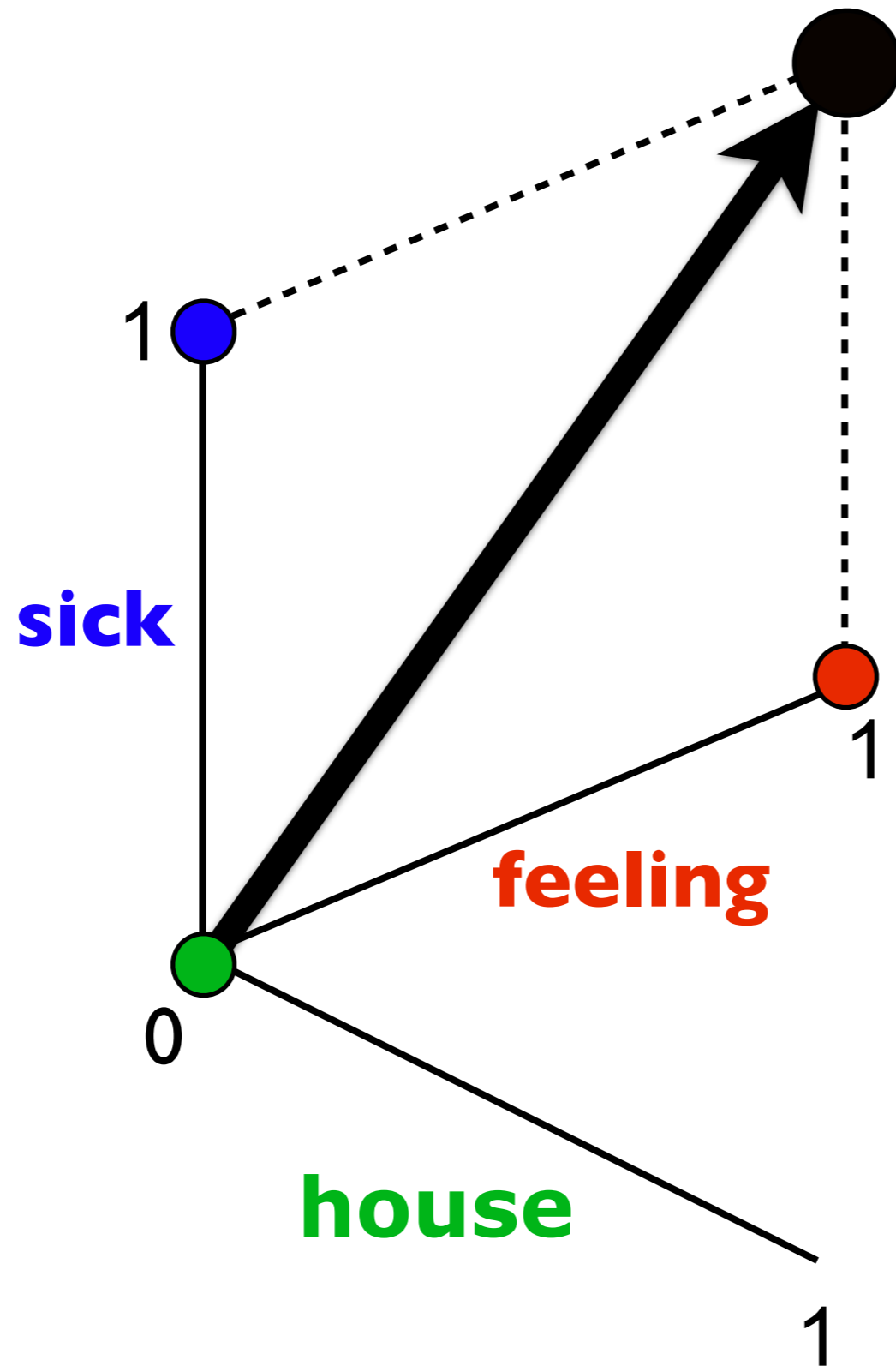
“feeling sick”



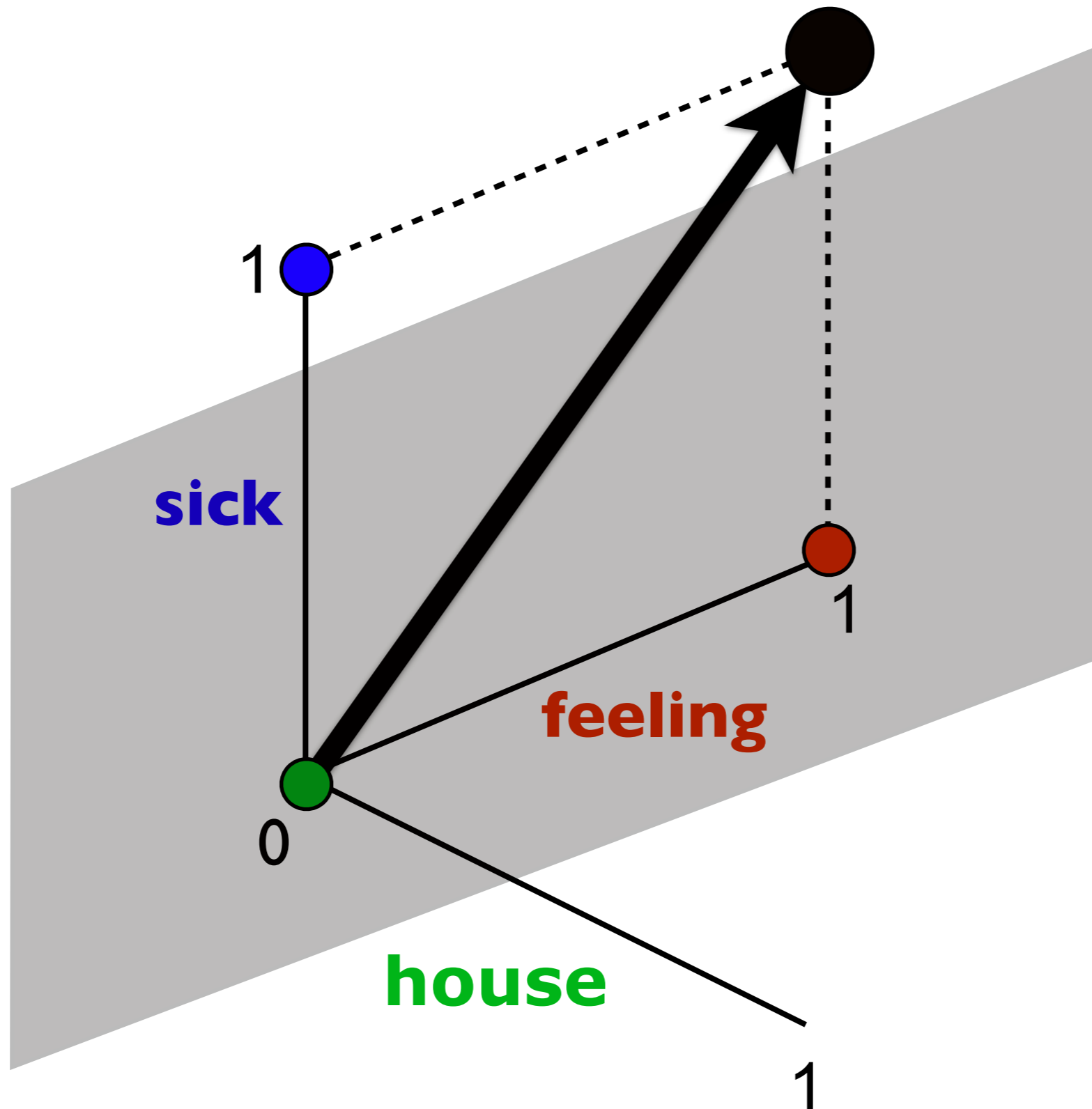
“feeling sick”



“feeling sick”



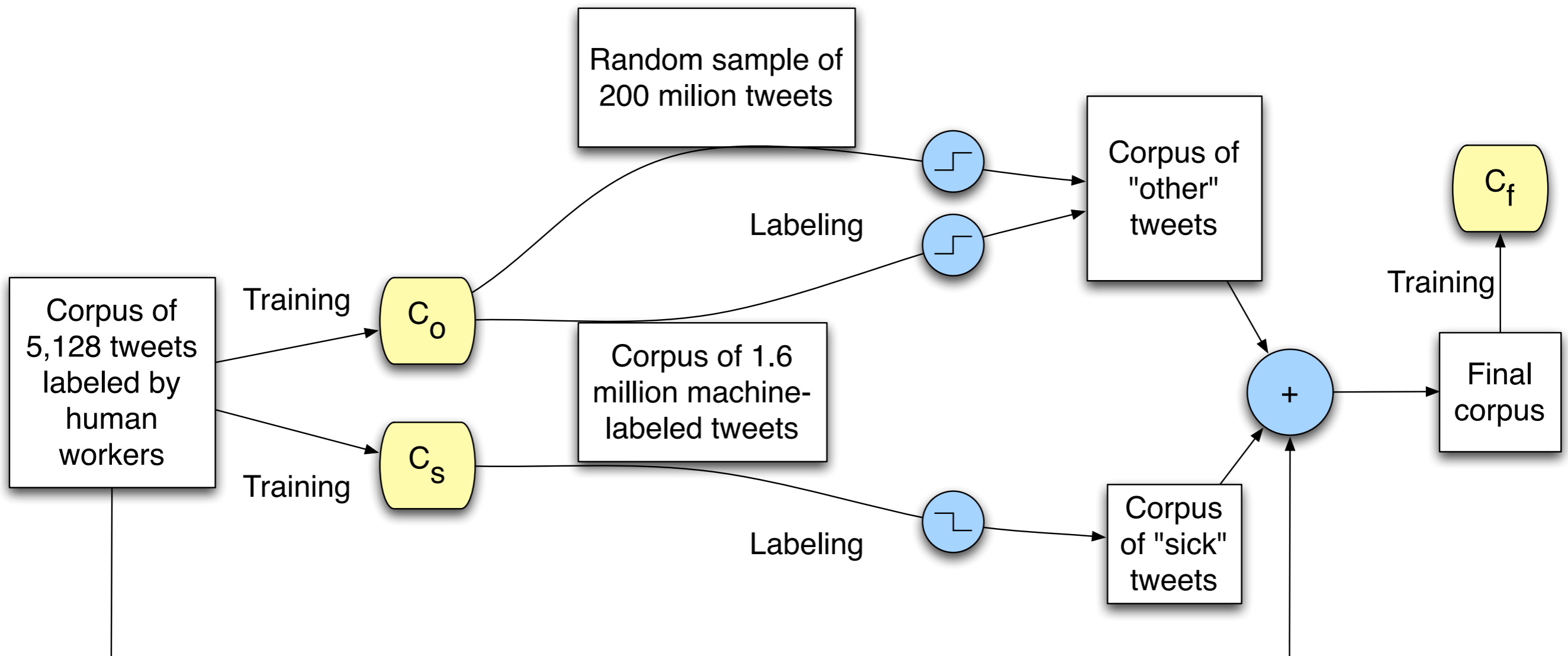
“feeling sick”





# Cascade SVM

# Cascade SVM





# SVM Features

Positive Features		Negative Features	
Feature	Weight	Feature	Weight
sick	0.9579	sick of	-0.4005
headache	0.5249	you	-0.3662
flu	0.5051	lol	-0.3017
fever	0.3879	love	-0.1753
feel	0.3451	i feel your	-0.1416
coughing	0.2917	so sick of	-0.0887
being sick	0.1919	bieber fever	-0.1026
better	0.1988	smoking	-0.0980
being	0.1943	i'm sick of	-0.0894
stomach	0.1703	pressure	-0.0837
and my	0.1687	massage	-0.0726
infection	0.1686	i love	-0.0719
morning	0.1647	pregnant	-0.0639

# SVM Robustness

- Held-out set of 700,000 tweets
- 0.98 precision
- 0.97 recall





**CherishmySTYLE** [1.37667832]

On my way to work! I feel soooooo sick!!!  
I better get it together b4 tonight!!!



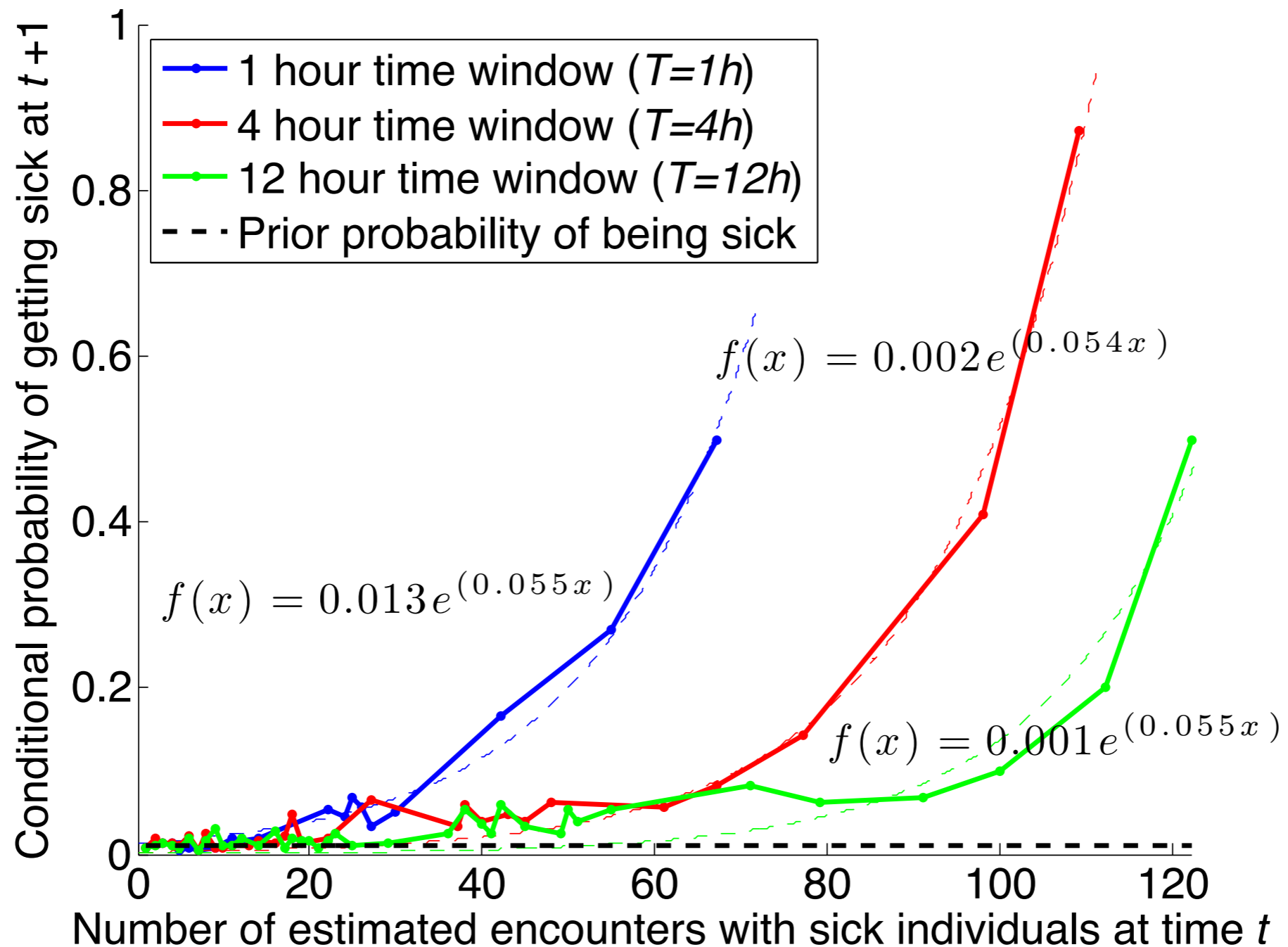


**lafemmefatalexx** [0.74721562]

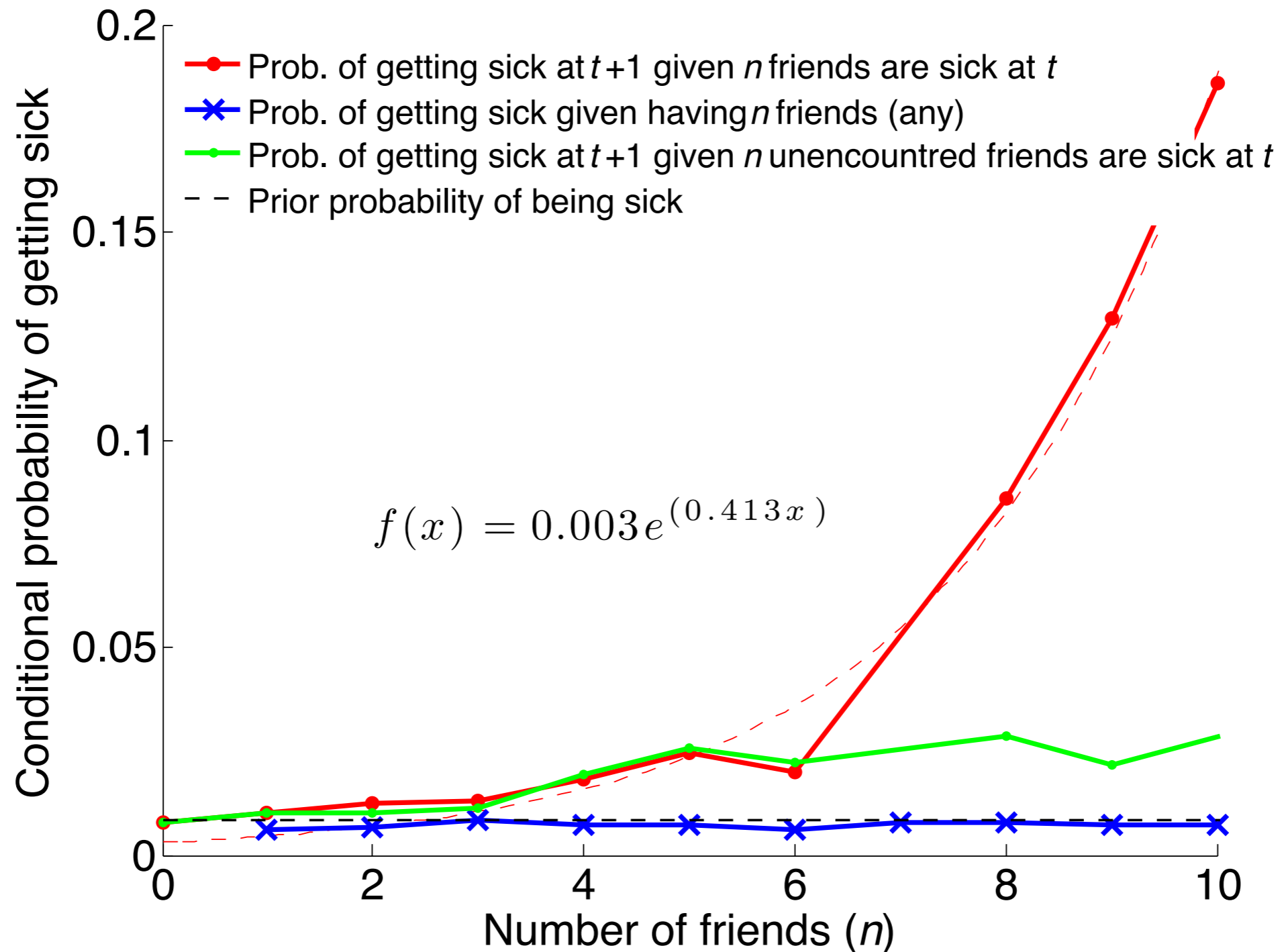
My feet hurt soooooo badddd I need to be carried home  
waaaaahhhhhhhhhhhhhhhhhhhhh never wearing heels again.



# Impact of Co-Location



# Impact of Friendships

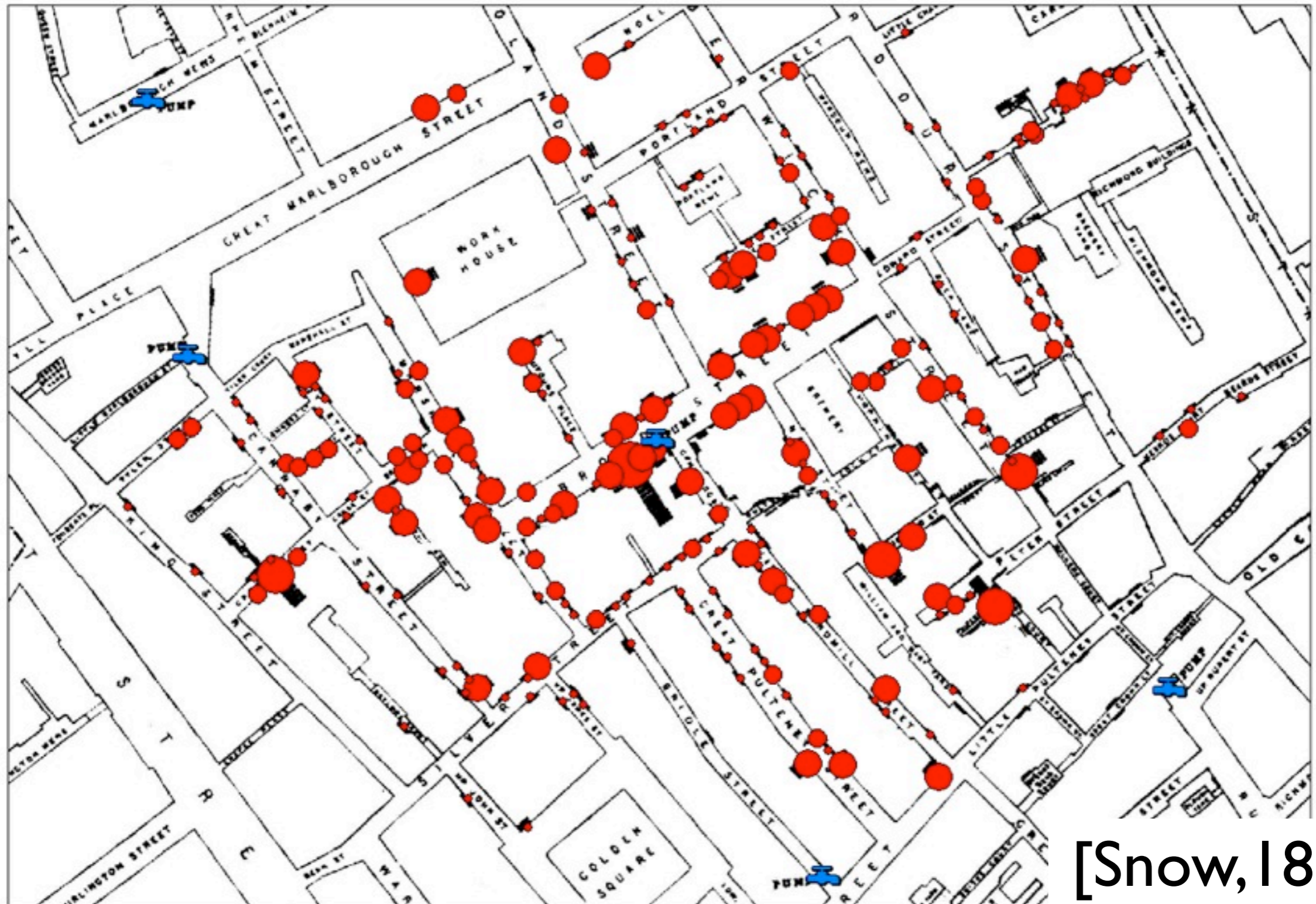


# Twitter Health

- Aggregate accuracy comparable with:
  - Google Flu Trends ( $R = 0.73$ )
  - CDC statistics
- ✦ **we can model fine-grained interactions between specific individuals**



# Health Insights



[Snow, 1855]



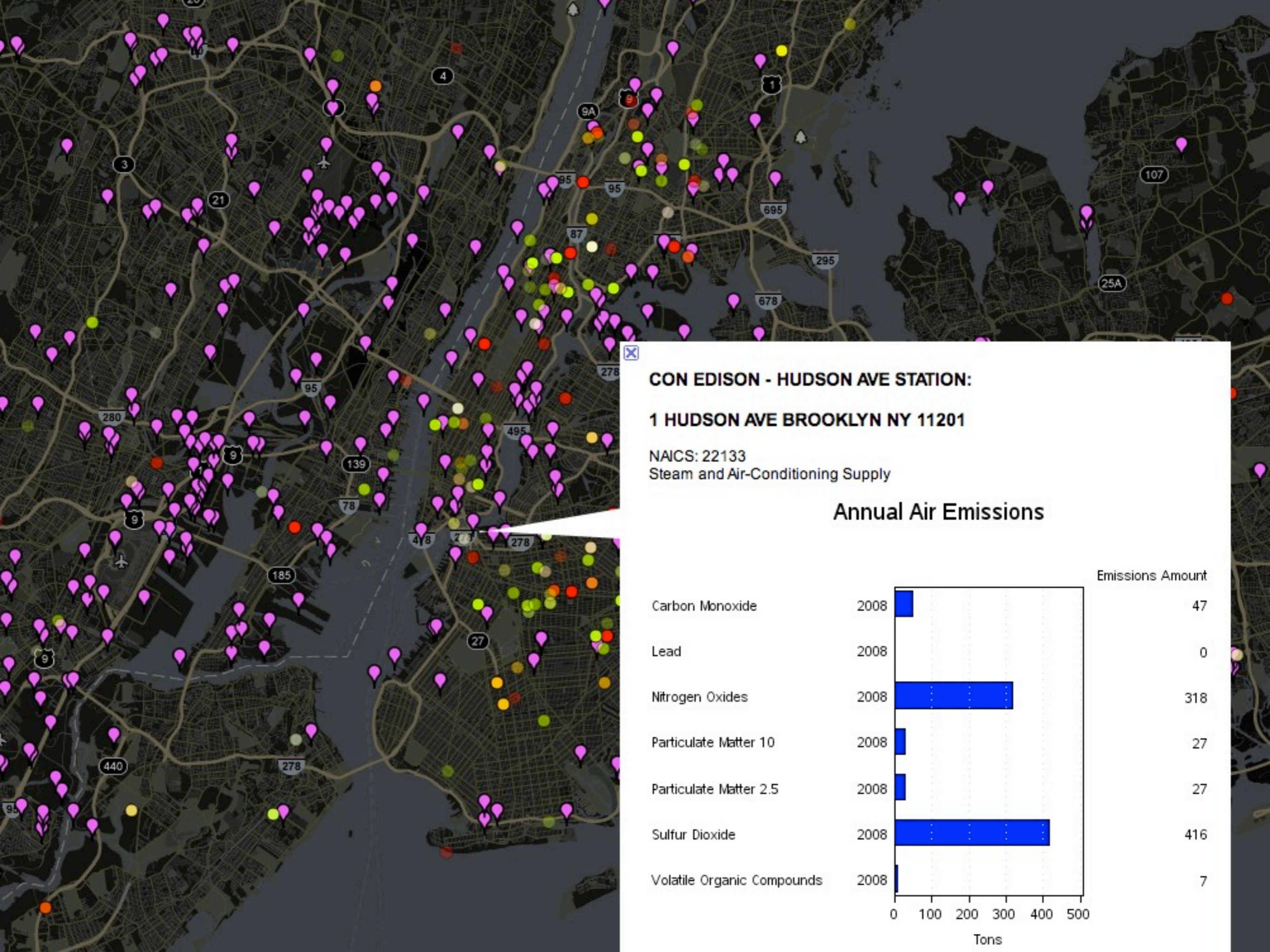










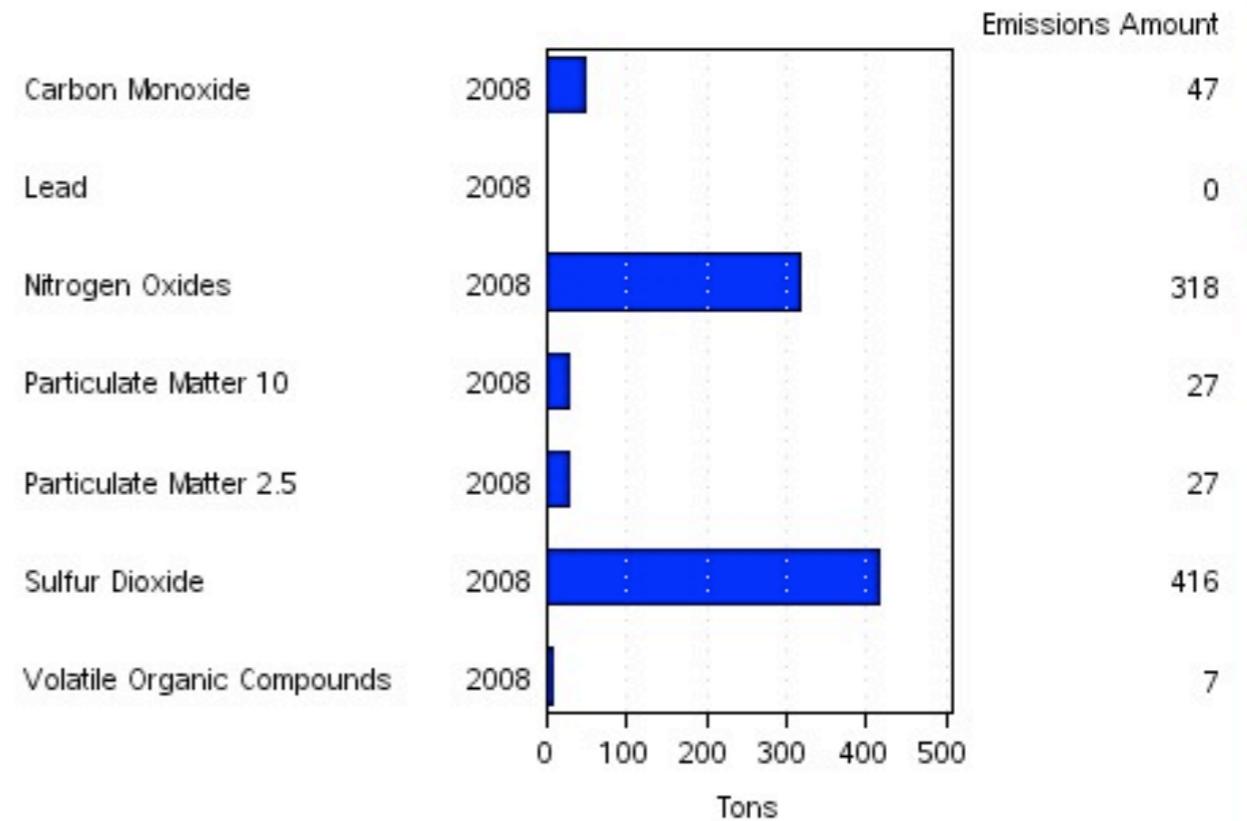


**CON EDISON - HUDSON AVE STATION:**

**1 HUDSON AVE BROOKLYN NY 11201**

NAICS: 22133  
Steam and Air-Conditioning Supply

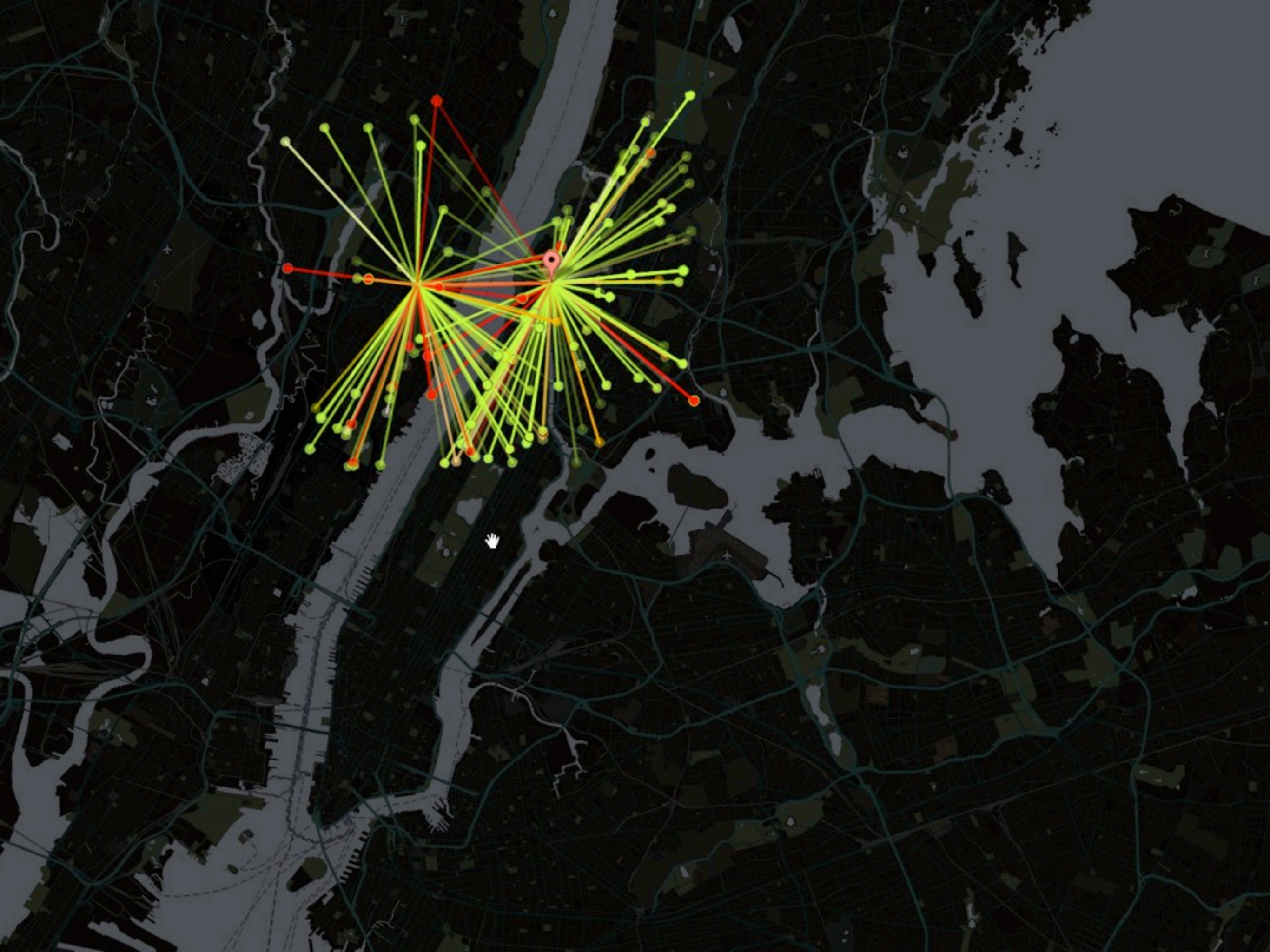
**Annual Air Emissions**



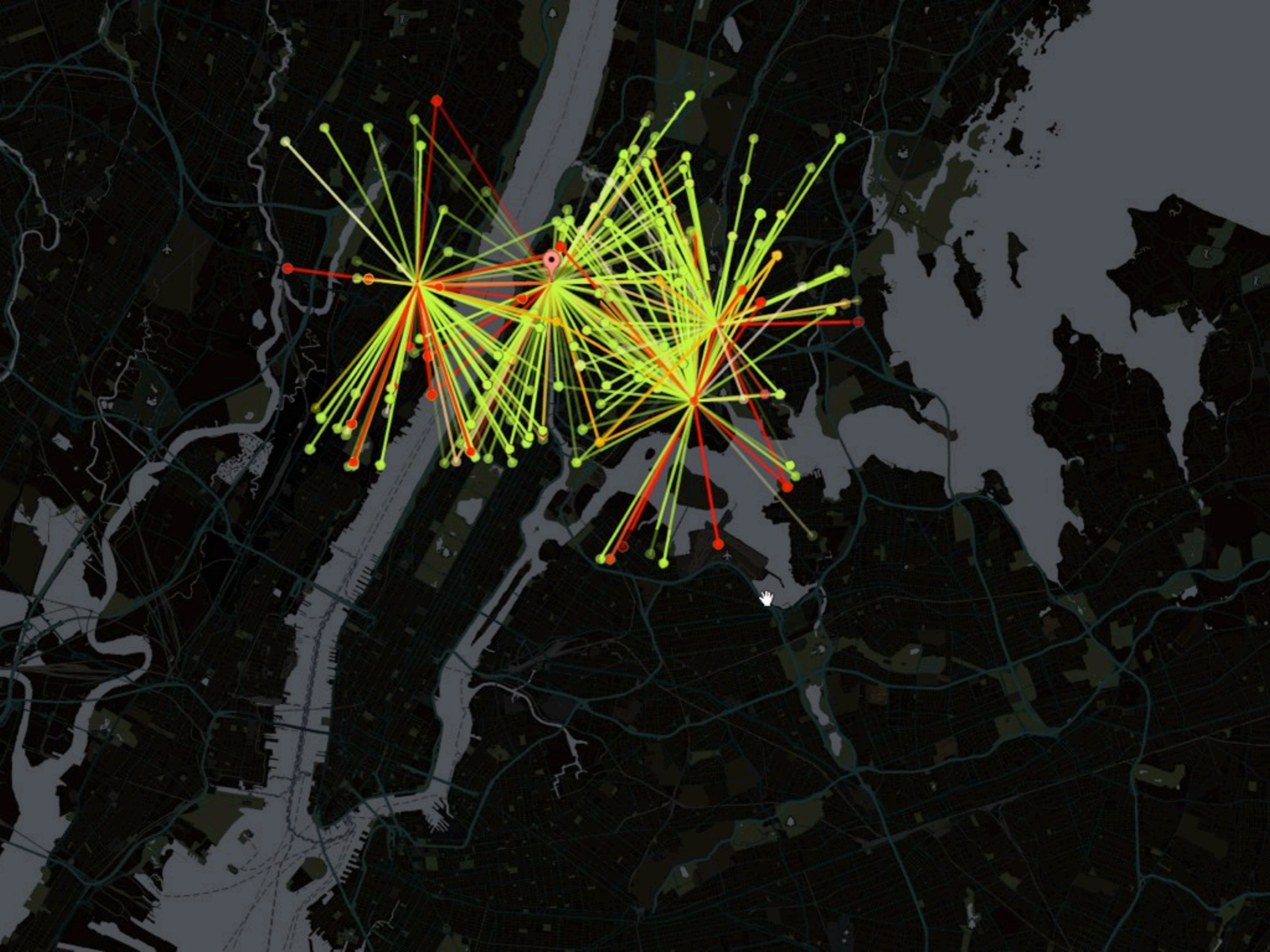
















# Summary



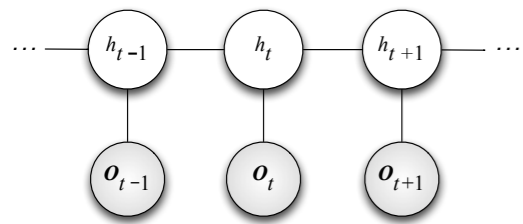




# Summary



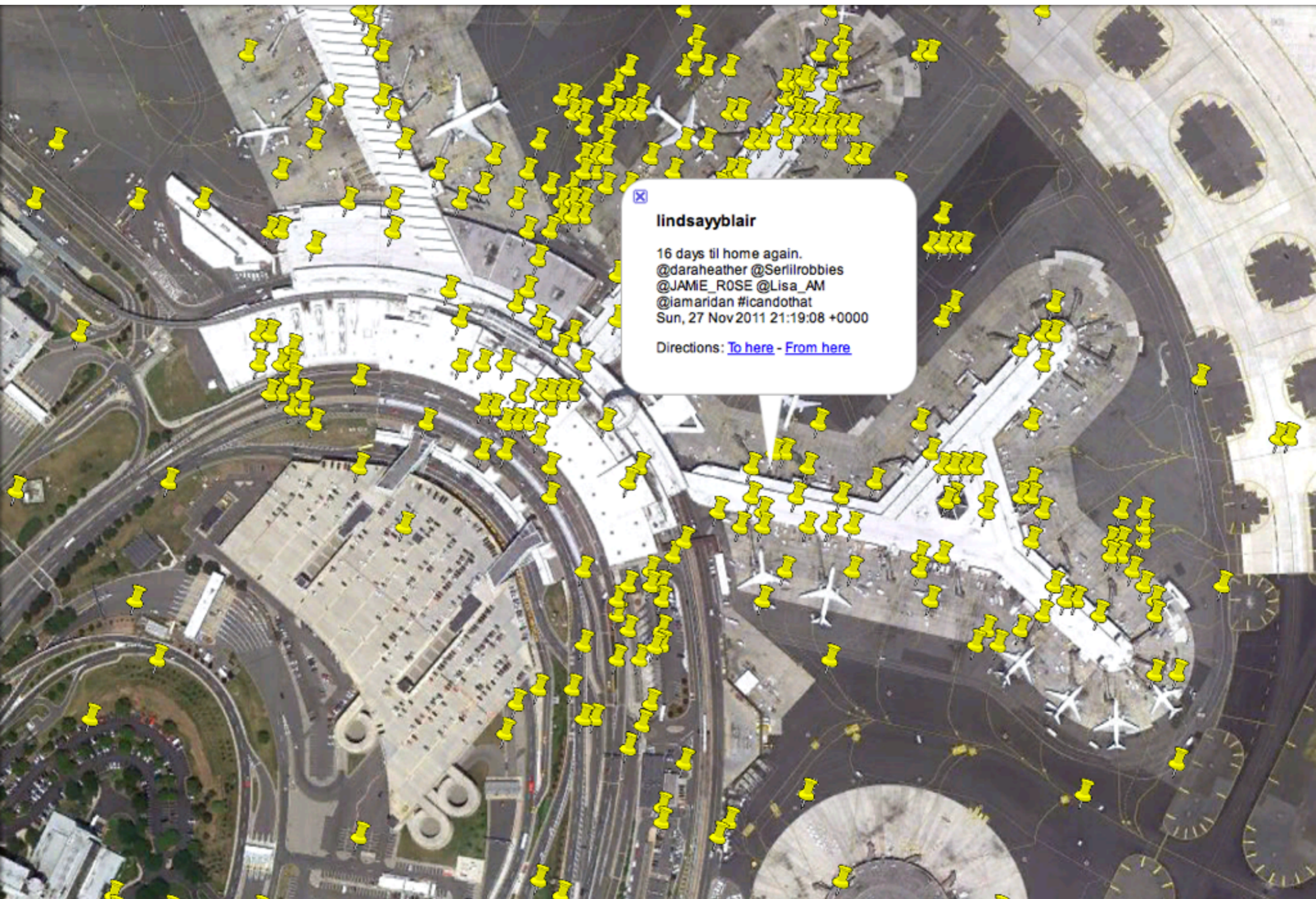
- **Quantify** impact of physical **encounters** and **social ties** on public **health**
  - Real world observed via Twitter
  - Fine granularity
  - Real-time
  - No active user participation
  - High accuracy



# Current Work



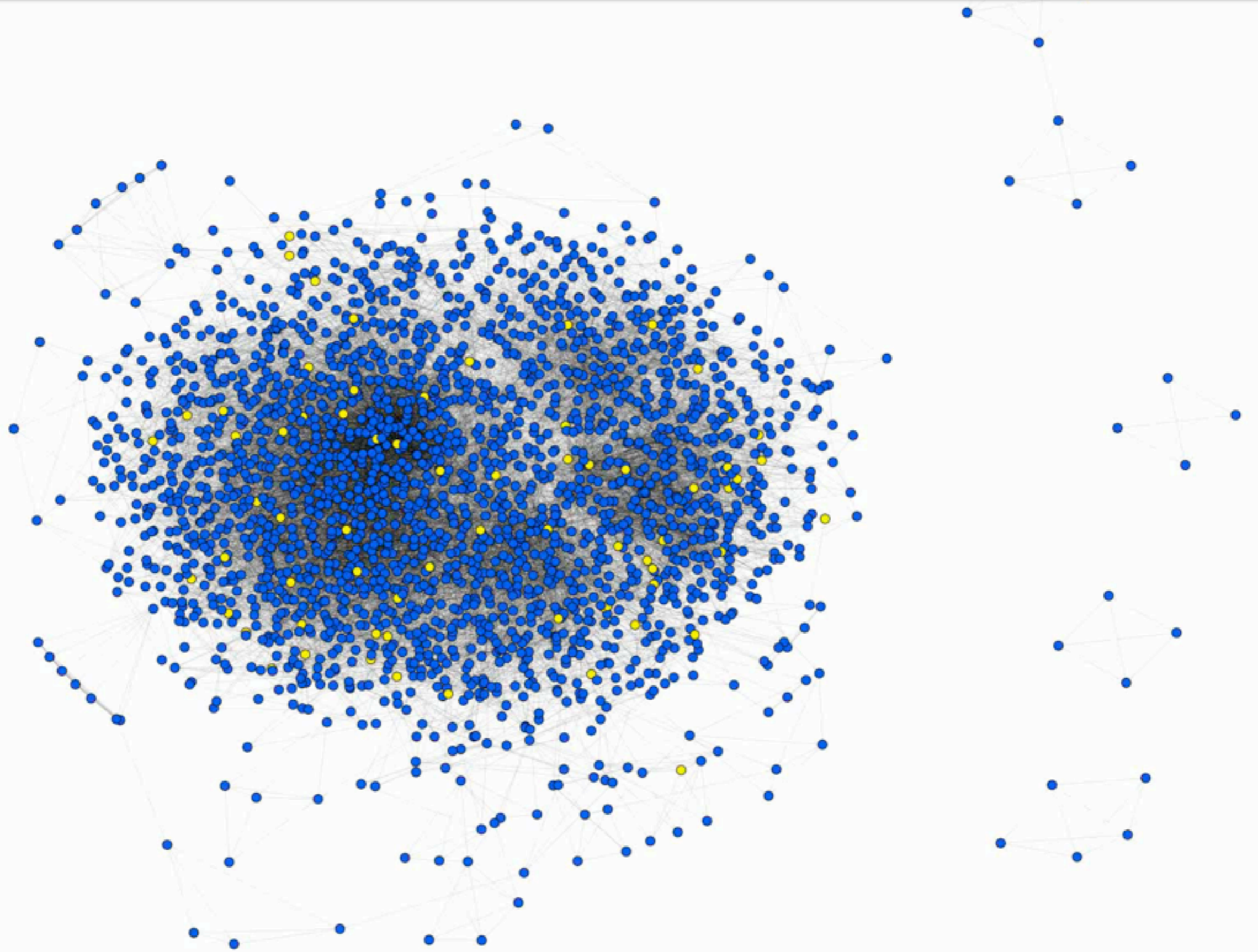
# Disease **Hubs** and **Vectors**



✕  
**lindsayblair**  
16 days til home again.  
@daraheather @Serilrobbies  
@JAMIE\_ROSE @Lisa\_AM  
@iamaridan #icandothat  
Sun, 27 Nov 2011 21:19:08 +0000  
Directions: [To here](#) - [From here](#)

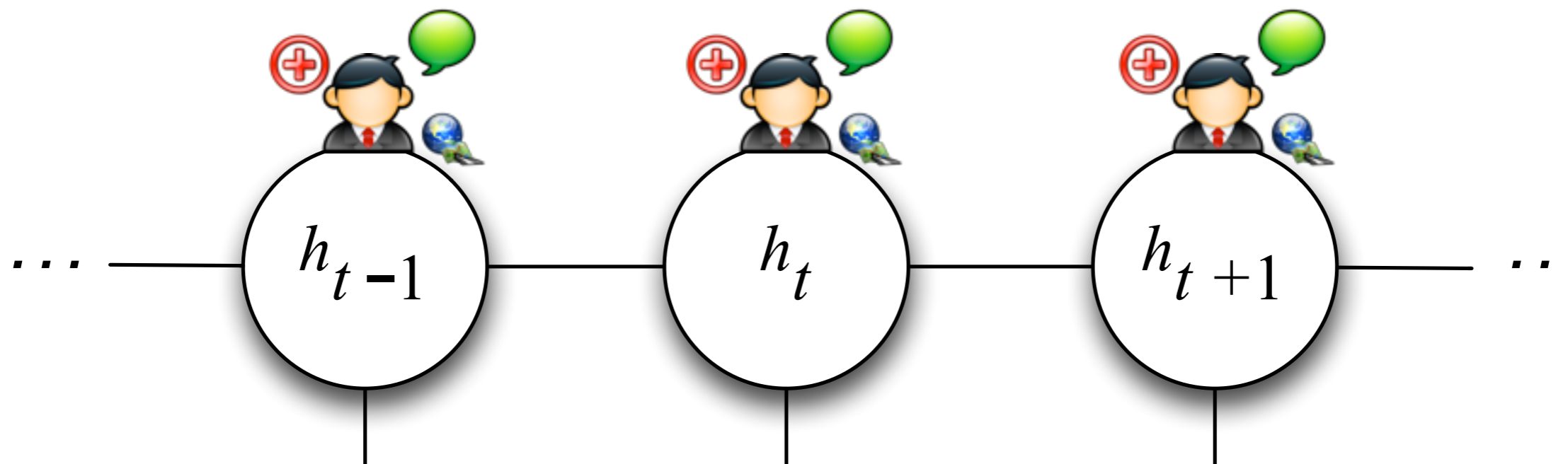


# Refining **Epidemiological Models**

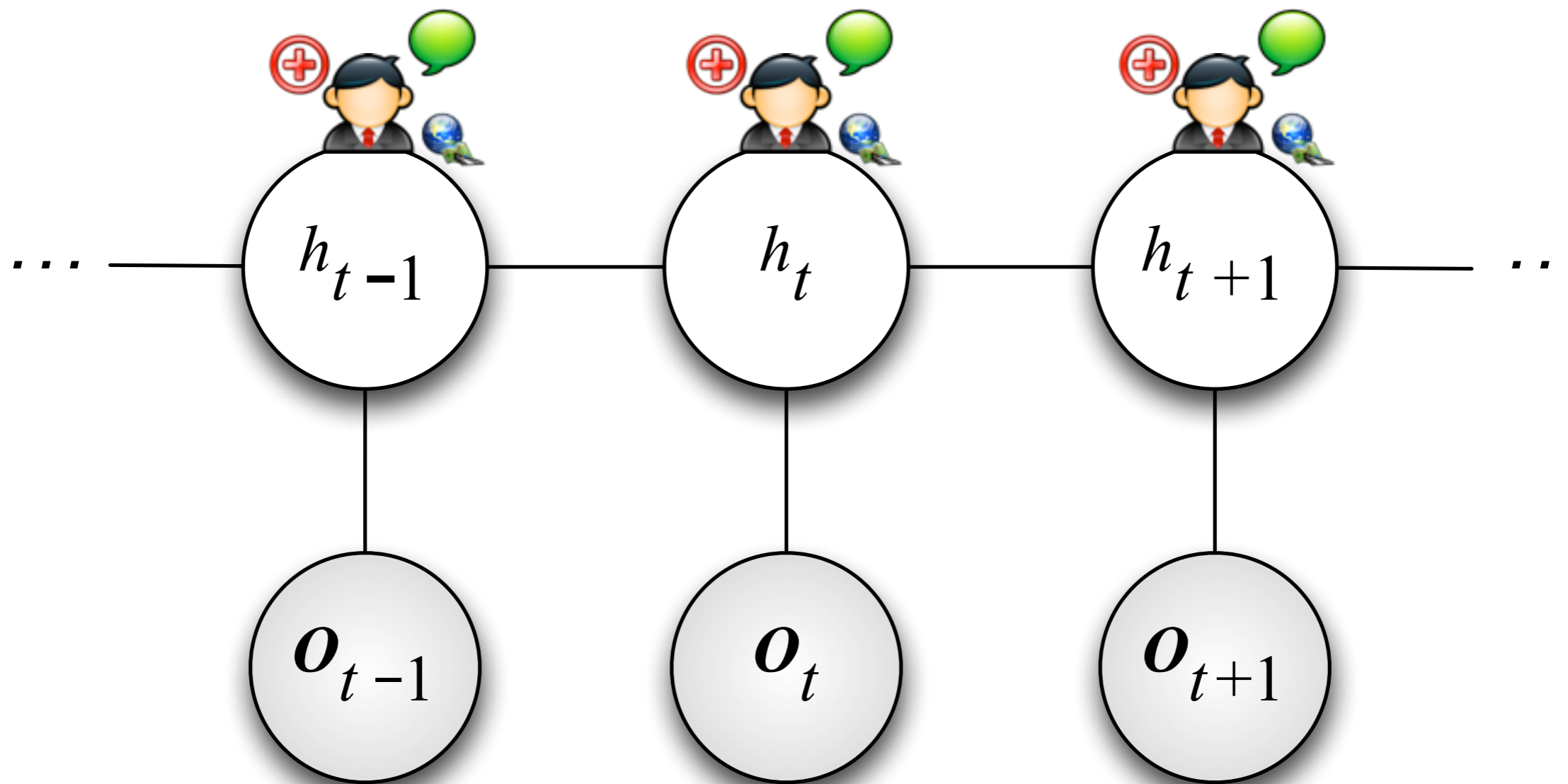




# Health Prediction

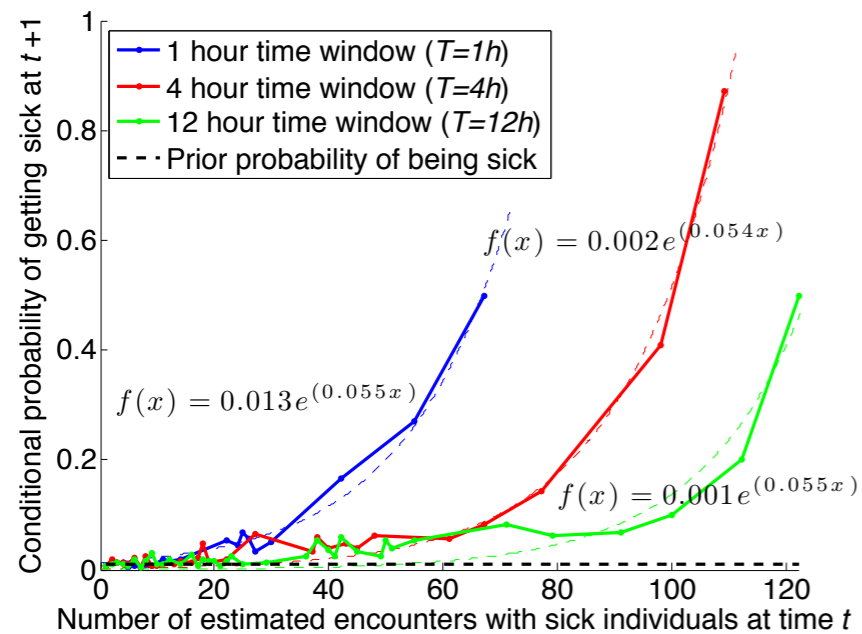
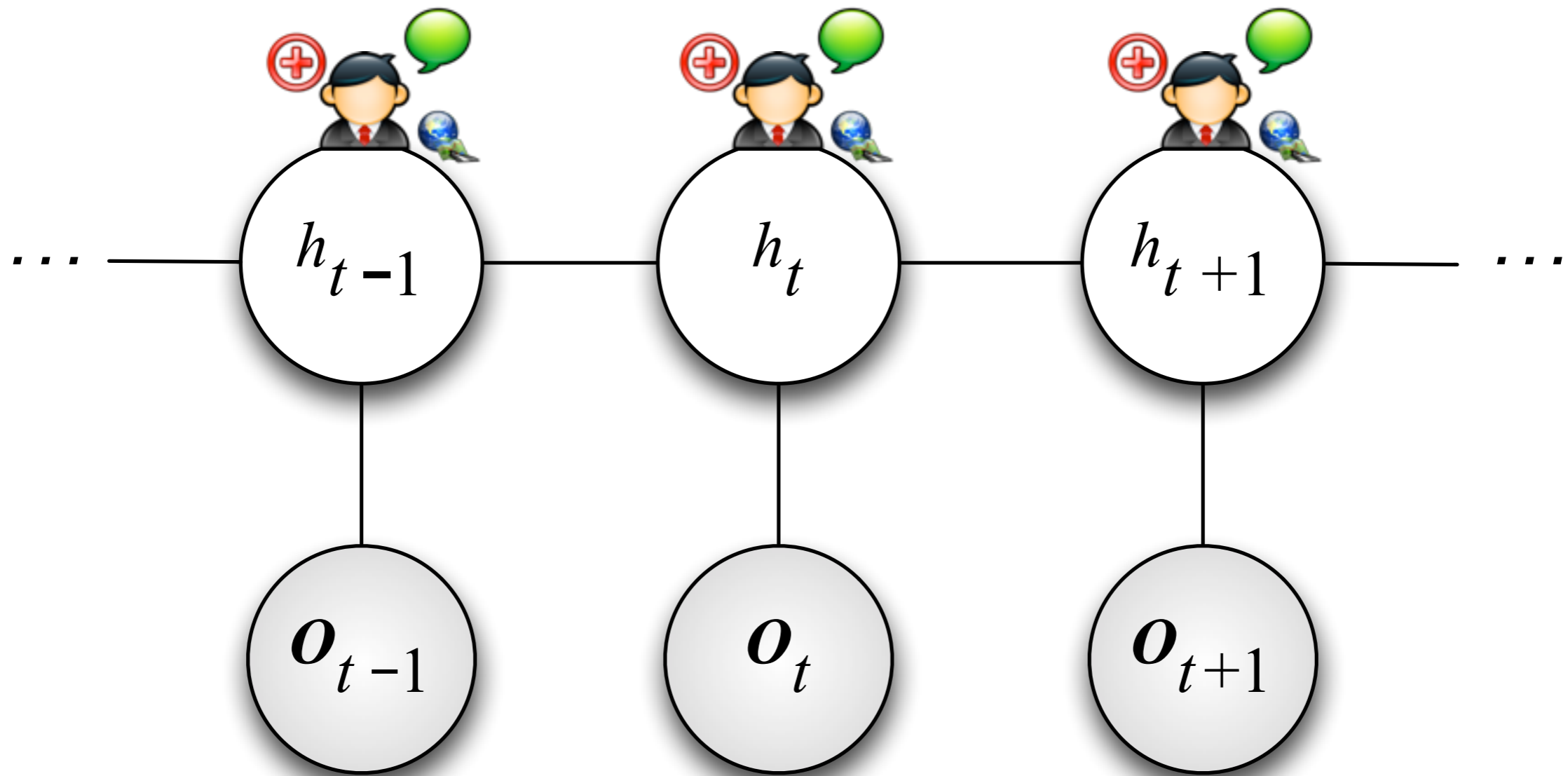


# Health Prediction

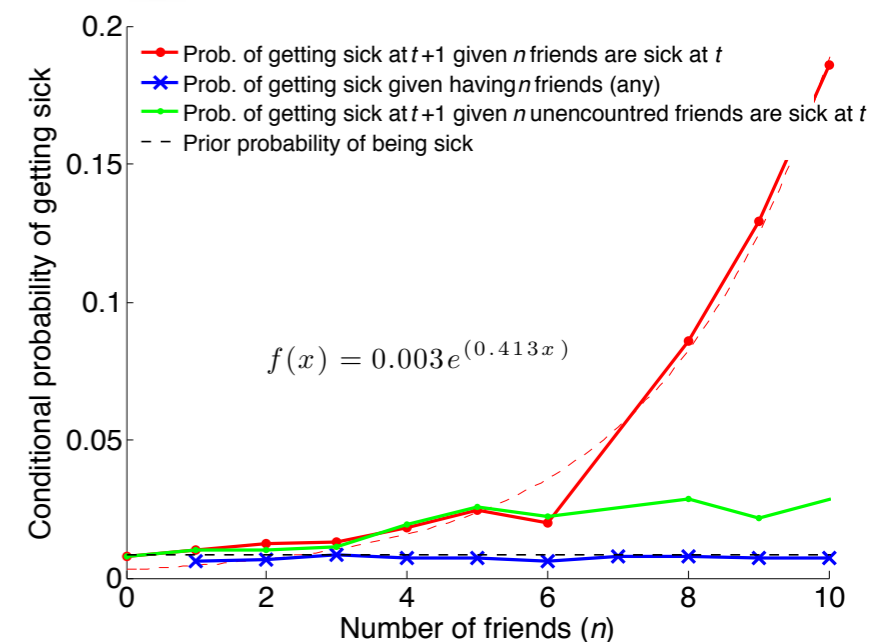
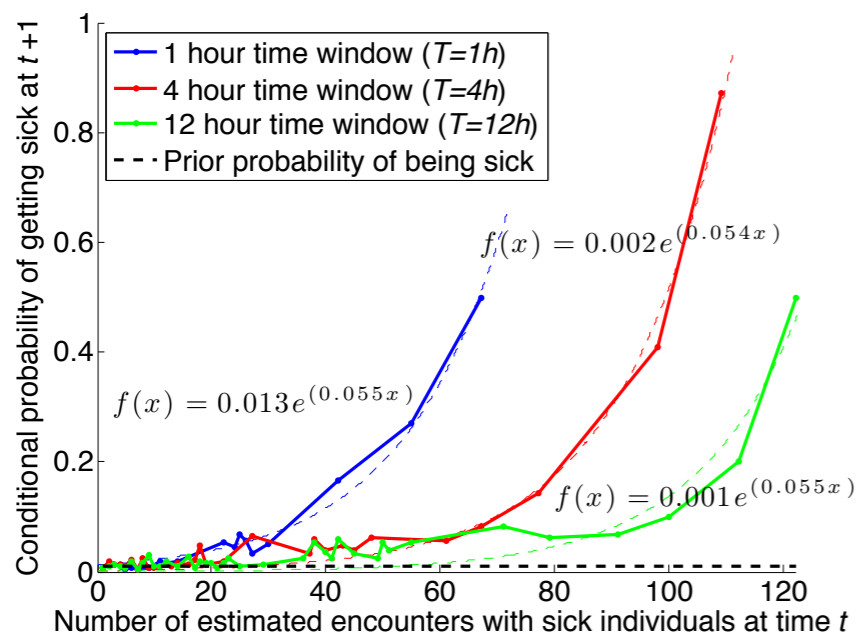
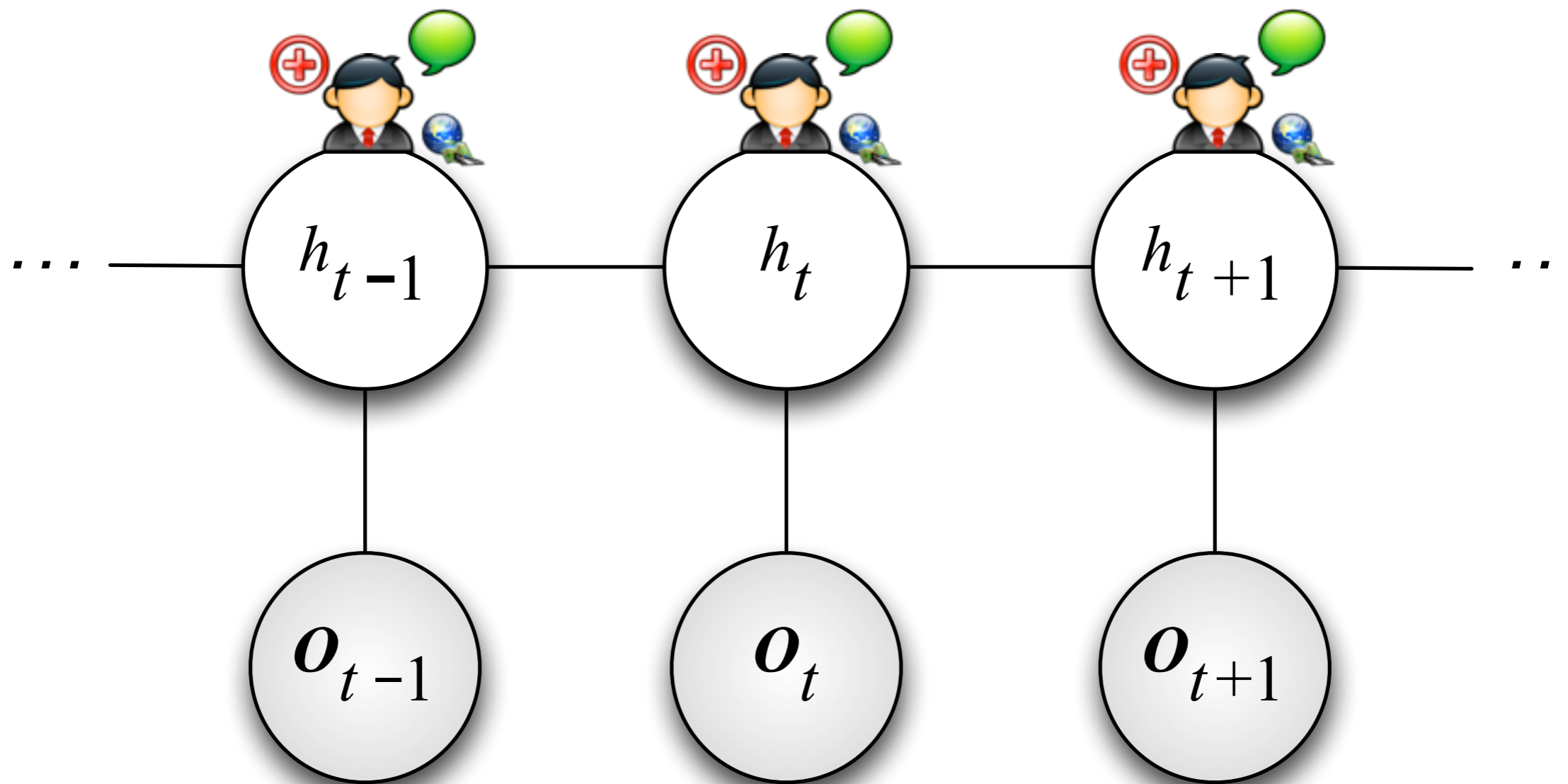




# Health Prediction



# Health Prediction



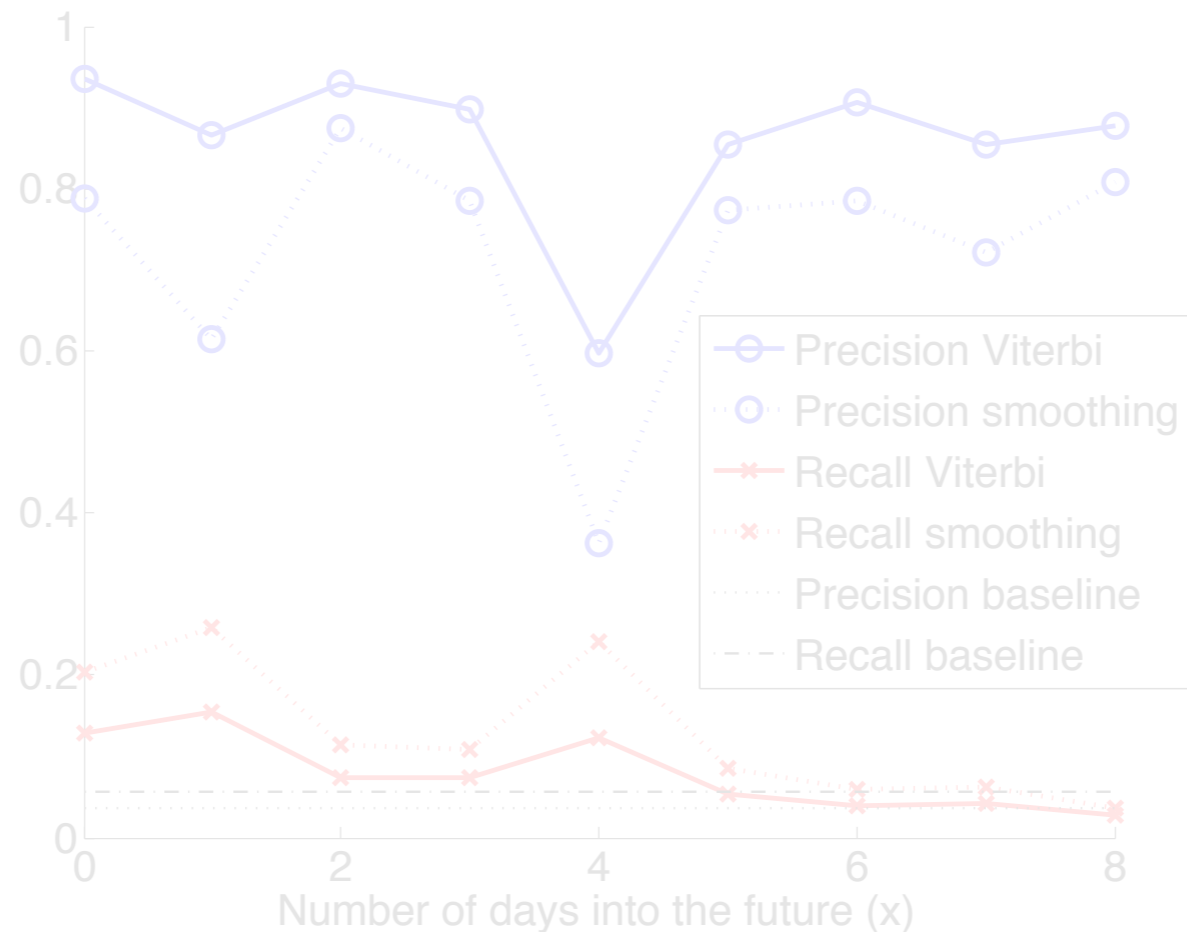
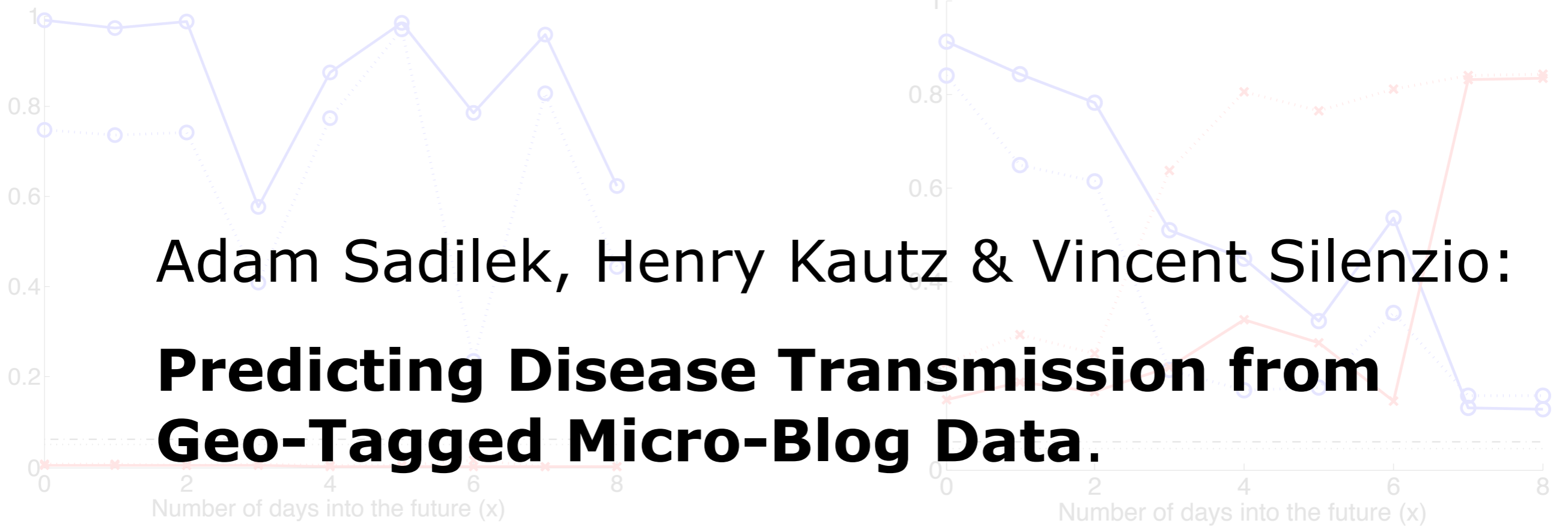


Friends Only

Co-Location Only

Adam Sadilek, Henry Kautz & Vincent Silenzio:  
**Predicting Disease Transmission from  
Geo-Tagged Micro-Blog Data.**

AAAI 2012 in Toronto



# Demo

[bit.ly/](https://bit.ly/twitterHealth)**twitterHealth**



# Modeling Spread of Disease from Social Interactions



[bit.ly/twitterHealth](http://bit.ly/twitterHealth)

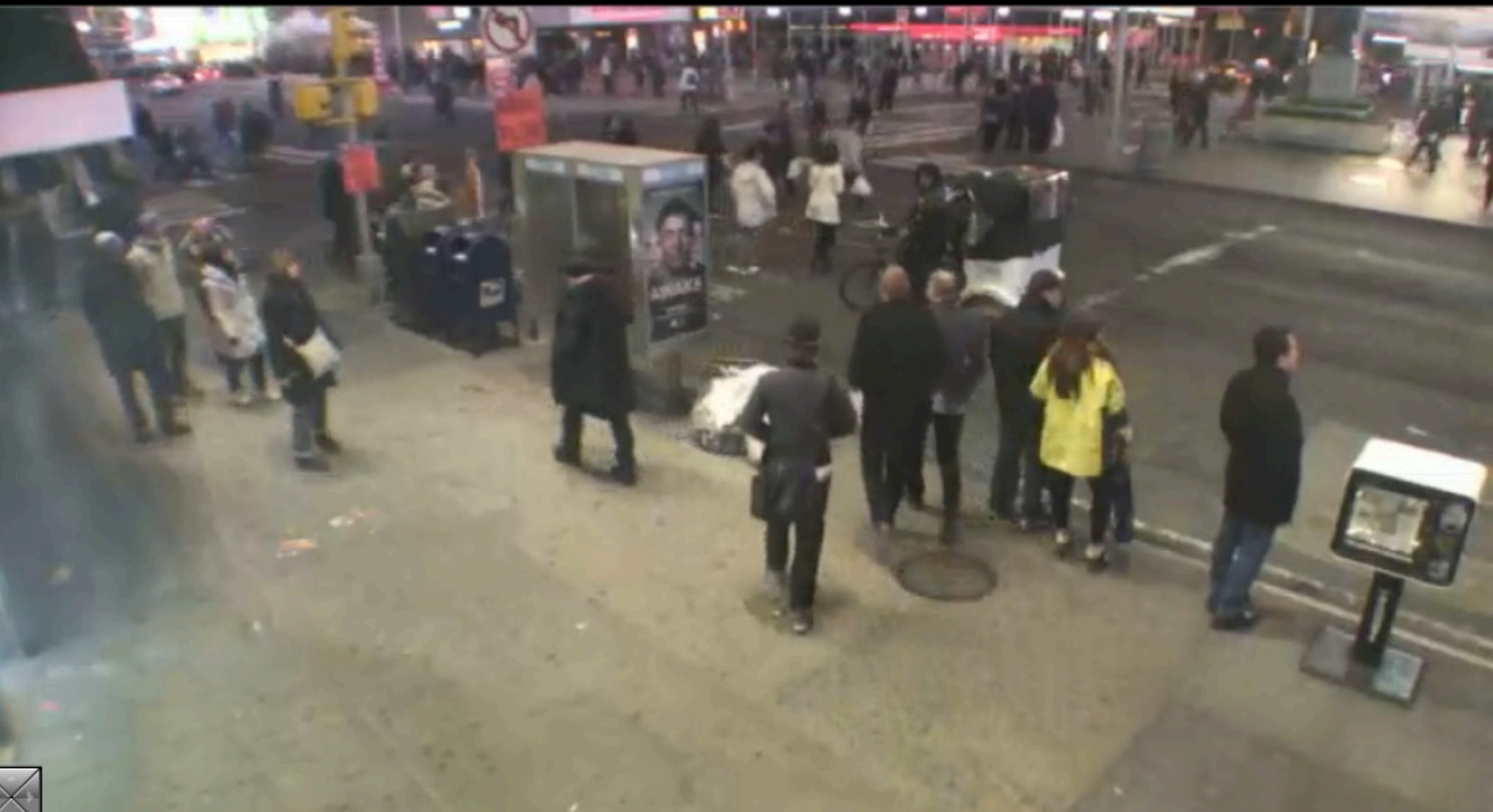
Adam Sadilek  
Henry Kautz  
Vincent Silenzio





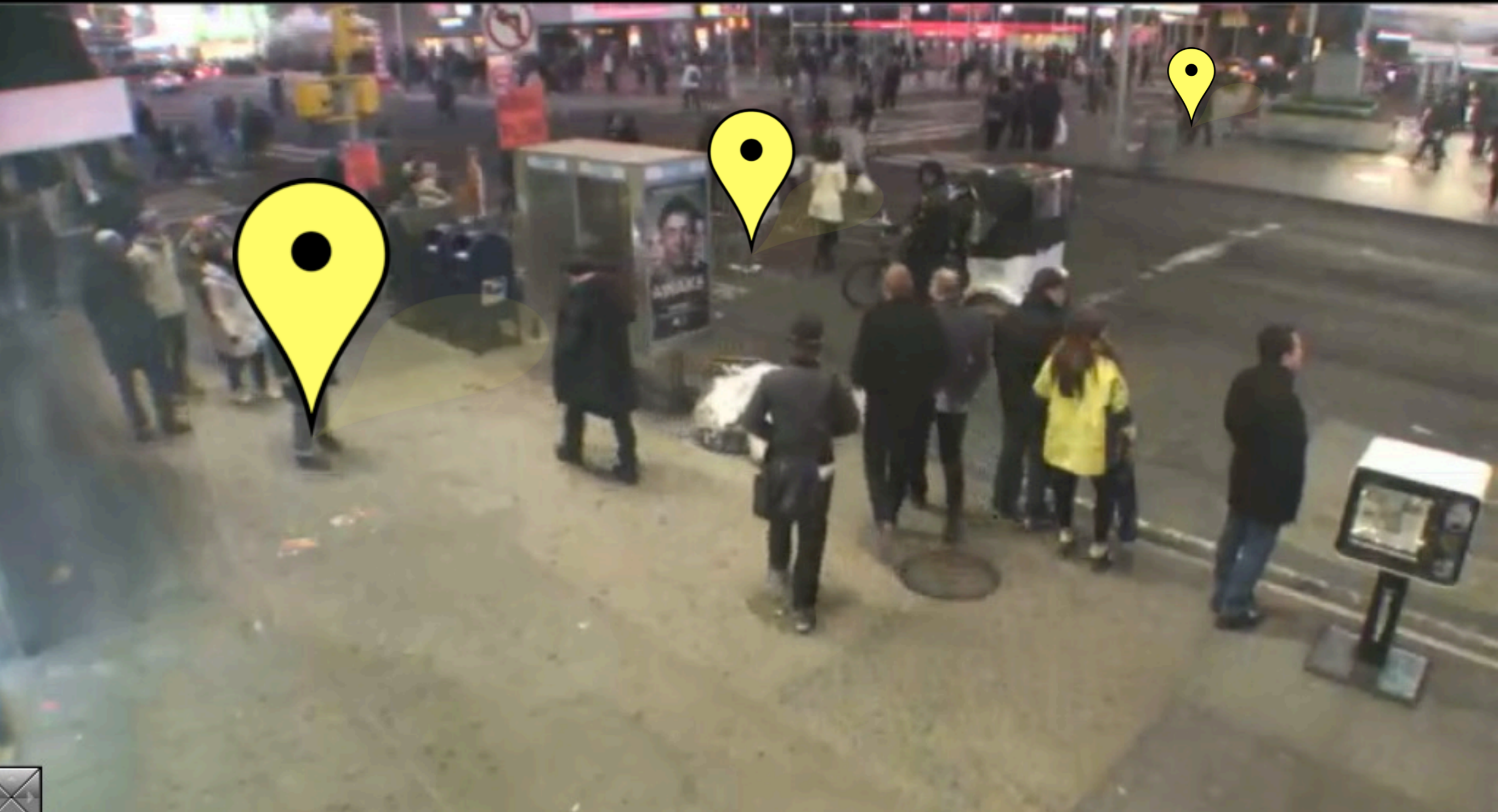


# Modeling “Hidden” Population





# Modeling “Hidden” Population





# Modeling “Hidden” Population



**Terrible  
headache!**