

# Finding Your Friends and Following Them to Where You Are



UNIVERSITY *of*  
ROCHESTER

Adam Sadilek  
Henry Kautz  
Jeffrey P. Bigham









North Bergen  
West New York  
Union City

Hoboken

New York

Jersey City

North Arlington

Kearny

Newark

Bayonne

West New Brighton

John F. Kennedy International Airport

Jamaica Bay Wildlife Refuge

Franklin D. Roosevelt Boardwalk and Beach

Rockaway Park

Hudson River

East River

LaGuardia Airport

Gravesend Bay

Mil Basin

Marine Park

Fountain Avenue

Richard w De Korte Park

Laurel Hill Park

Central Park

Kissena Park

Garfield Ave

Liberty State Park

FDR Drive

Brooklyn Queens Expy

Queens Midtown Expy

Long Island Expy

Union Turnpike

Grand Central Pkwy

Ferry St

Pulaski Skyway

US-1 Truck

Passaic

W 42nd St

7th Ave

5th Ave

Lexington Ave

FDR Drive

2nd Ave

Port St

New Jersey Turnpike

Garfield Ave

New Jersey Turnpike

Atlantic Ave

Brooklyn Queens Expy

Atlantic Ave

Atlantic Ave

Atlantic Ave

Newark Bay

Upper Bay

Bayonne Golf Club

Hudson River

Atlantic Ave

Atlantic Ave

Atlantic Ave

Atlantic Ave

Atlantic Ave

Forest Ave

Freshkills Park

Willowbrook Parkway

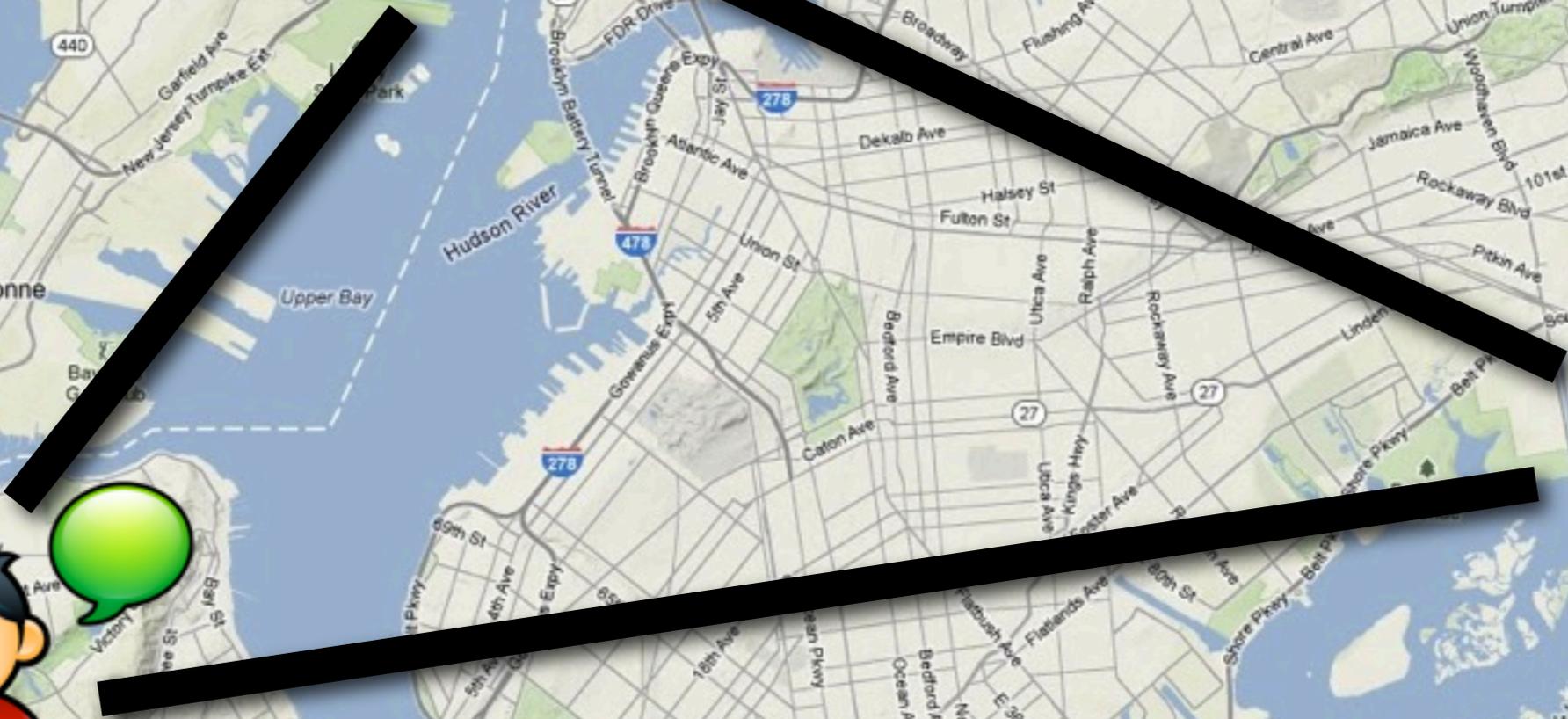
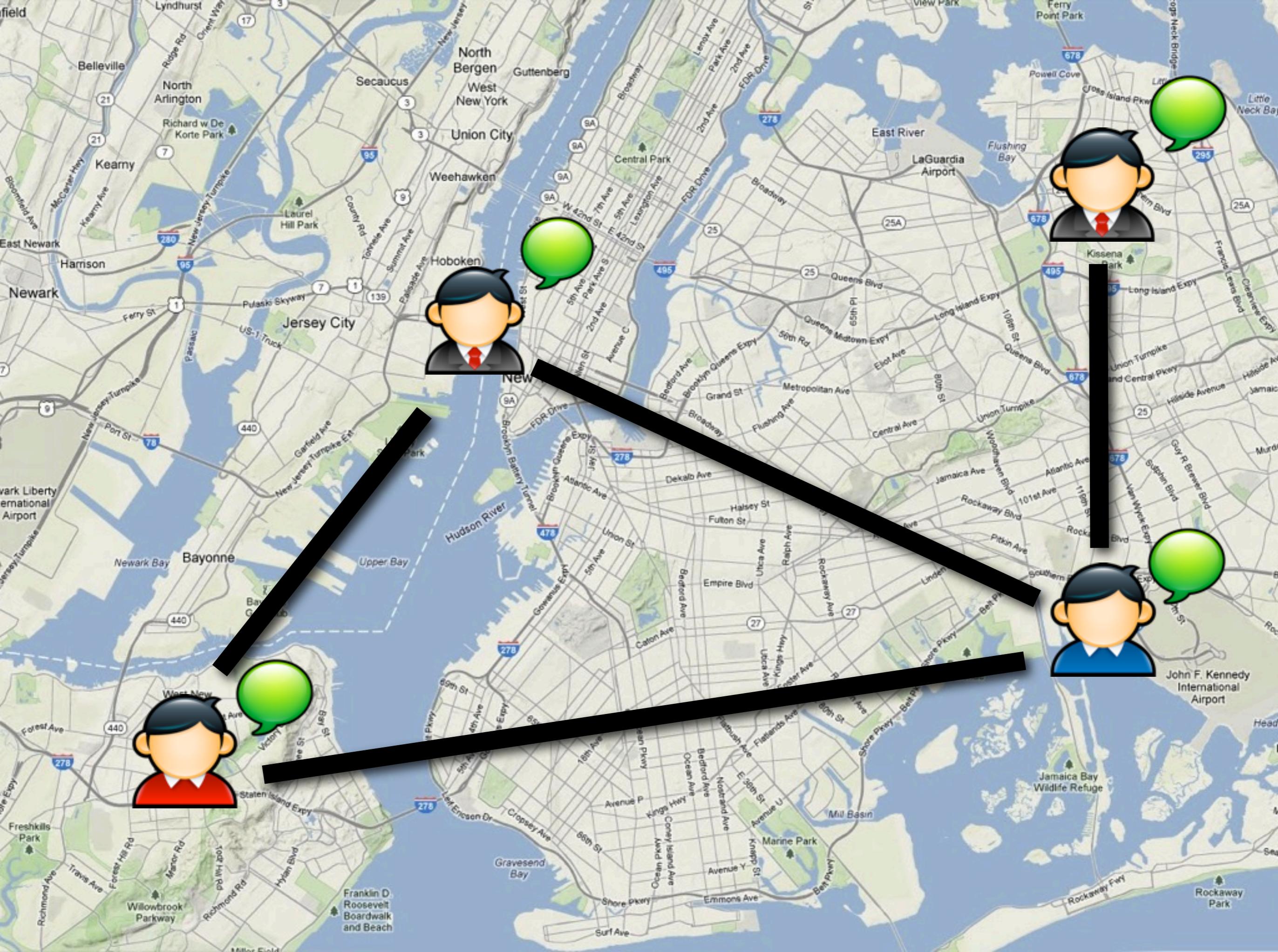
Richmond Rd

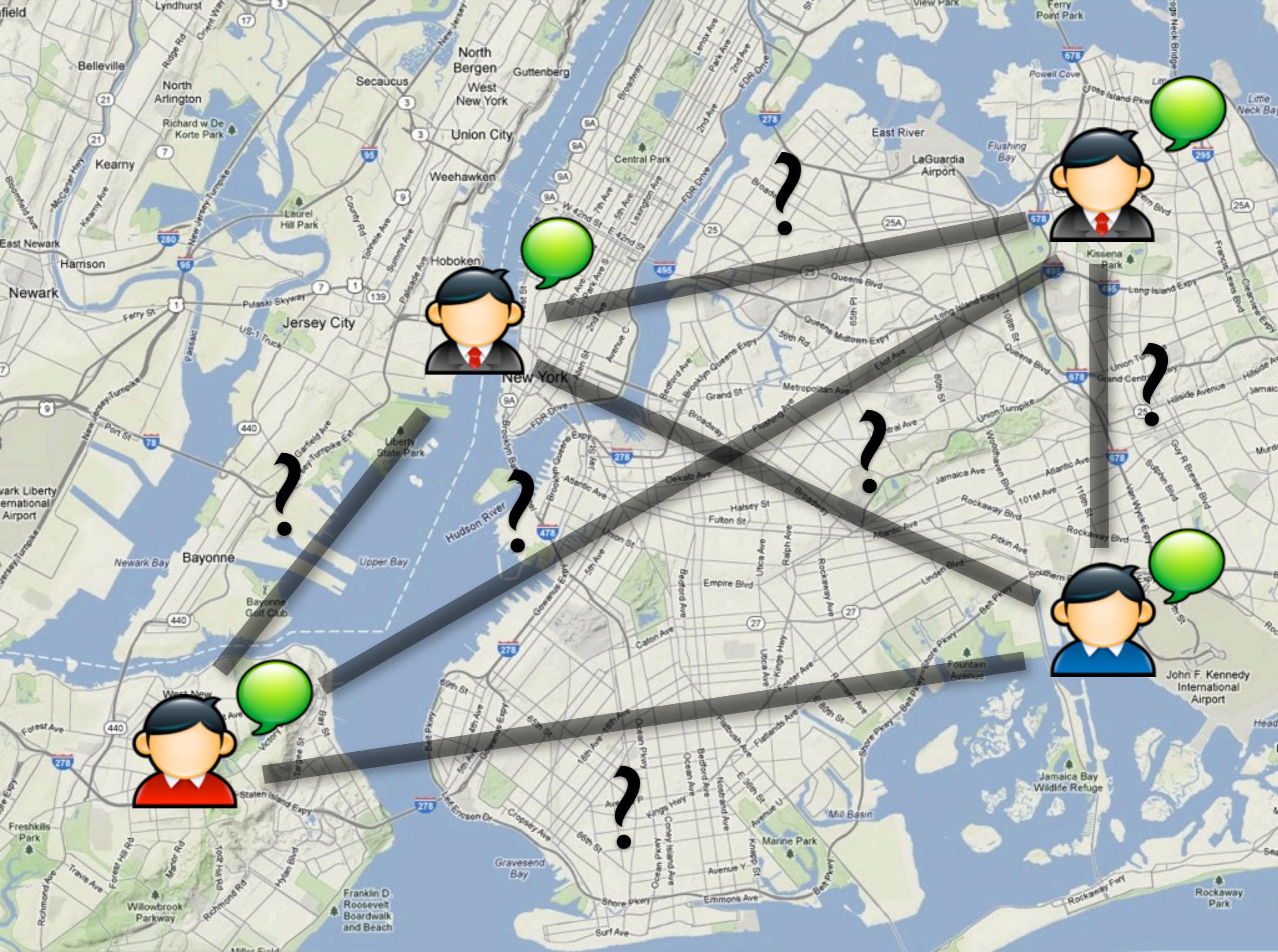
Travis Ave

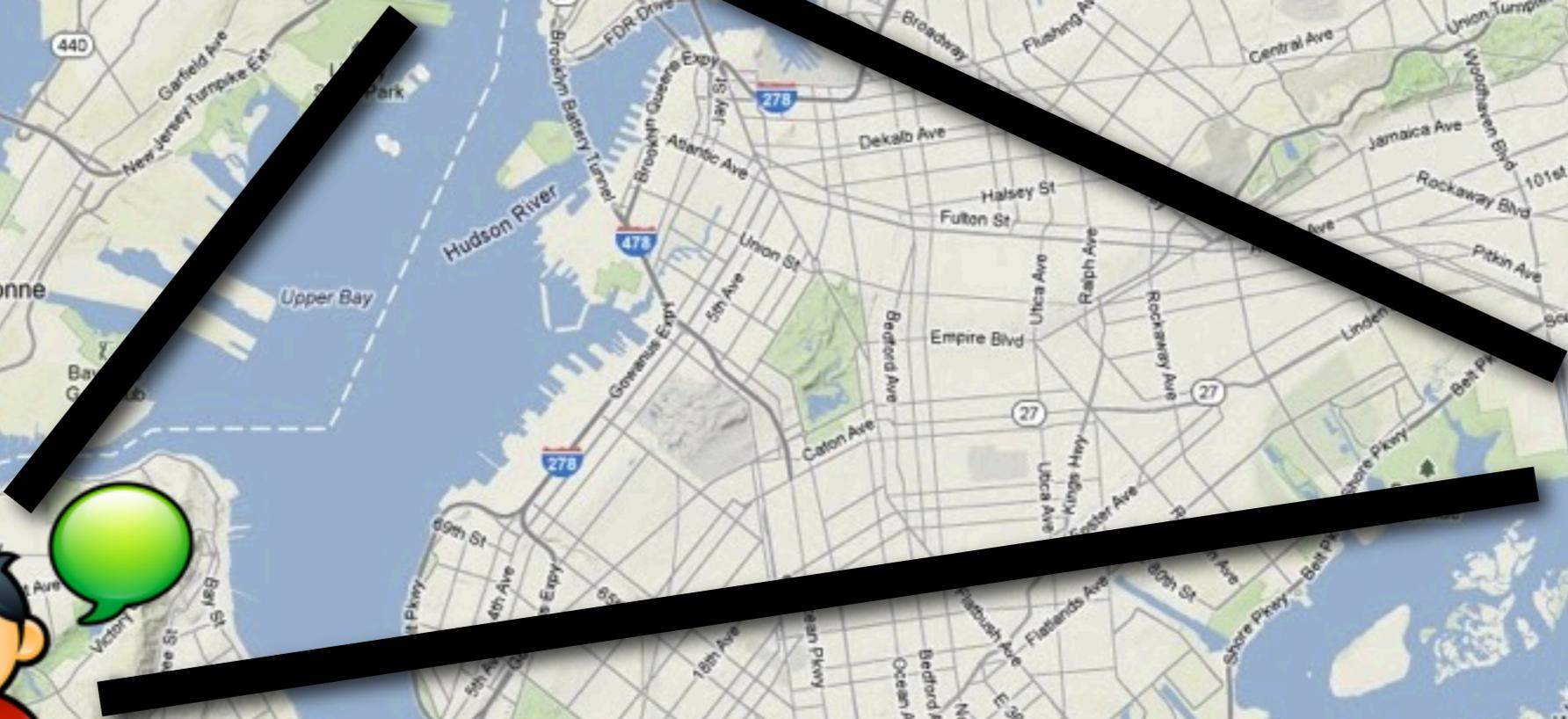
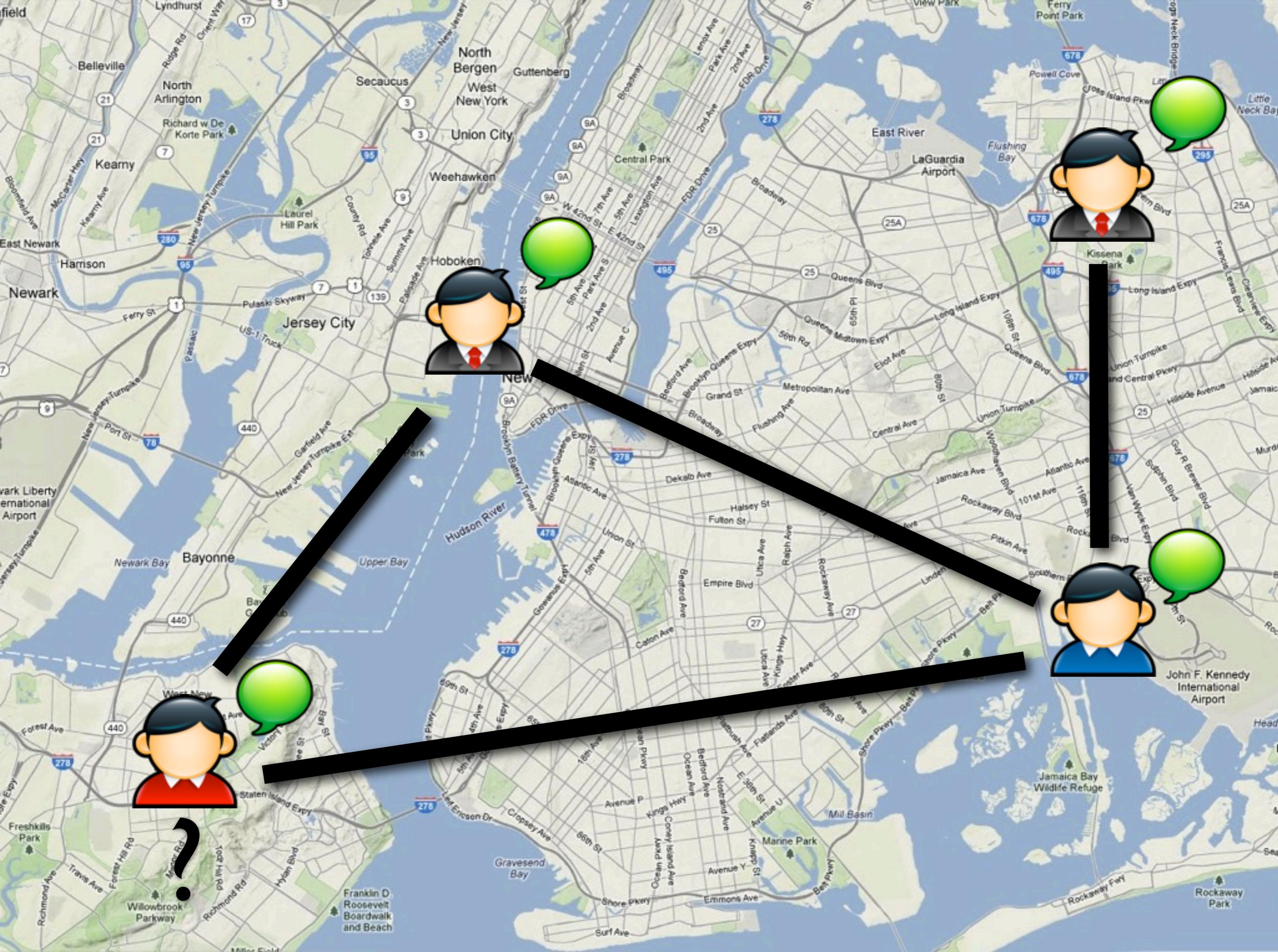
Manor Rd

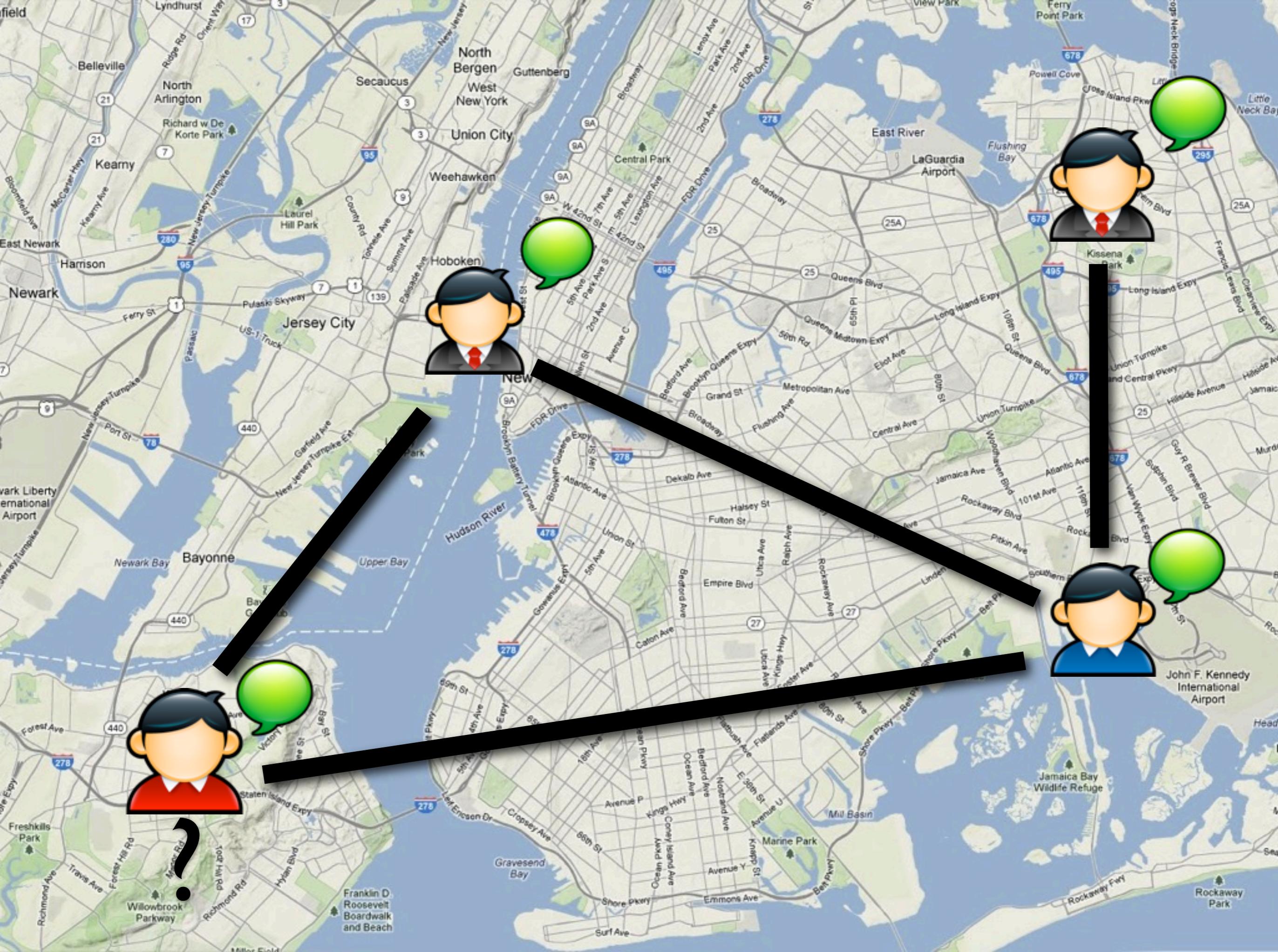
Richmond Ave













© 2012 Adam Sadilek

There was a man dancing on the train. He kept swinging his bookbag. Now I have a bruise on my leg. - \_ -



There was a man dancing on the train. He kept swinging his bookbag. Now I have a bruise on my leg. - \_ -

Train to work





# The Data

- **New York City & Los Angeles**

# The Data

- **New York City & Los Angeles**
- **26M tweets** over a month

# The Data

- **New York City & Los Angeles**
- **26M tweets** over a month
  - **1.2M unique users**

# The Data

- **New York City & Los Angeles**
- **26M tweets** over a month
  - **1.2M unique users**
  - **7.6M geo-tagged tweets**

# The Data

- 11K **geo-active** users

# The Data

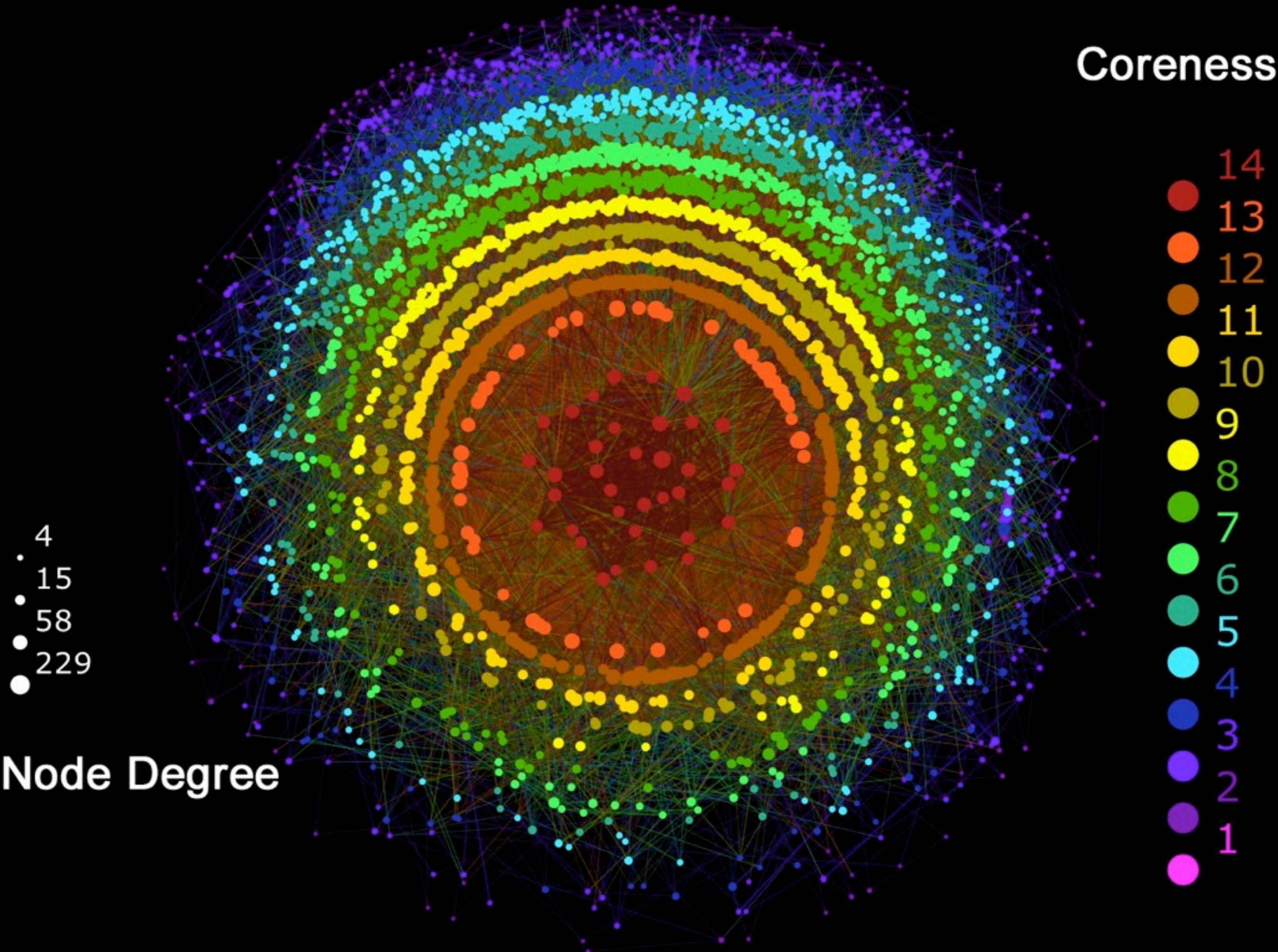
- 11K **geo-active** users
- 4M **tweets** by geo-active users

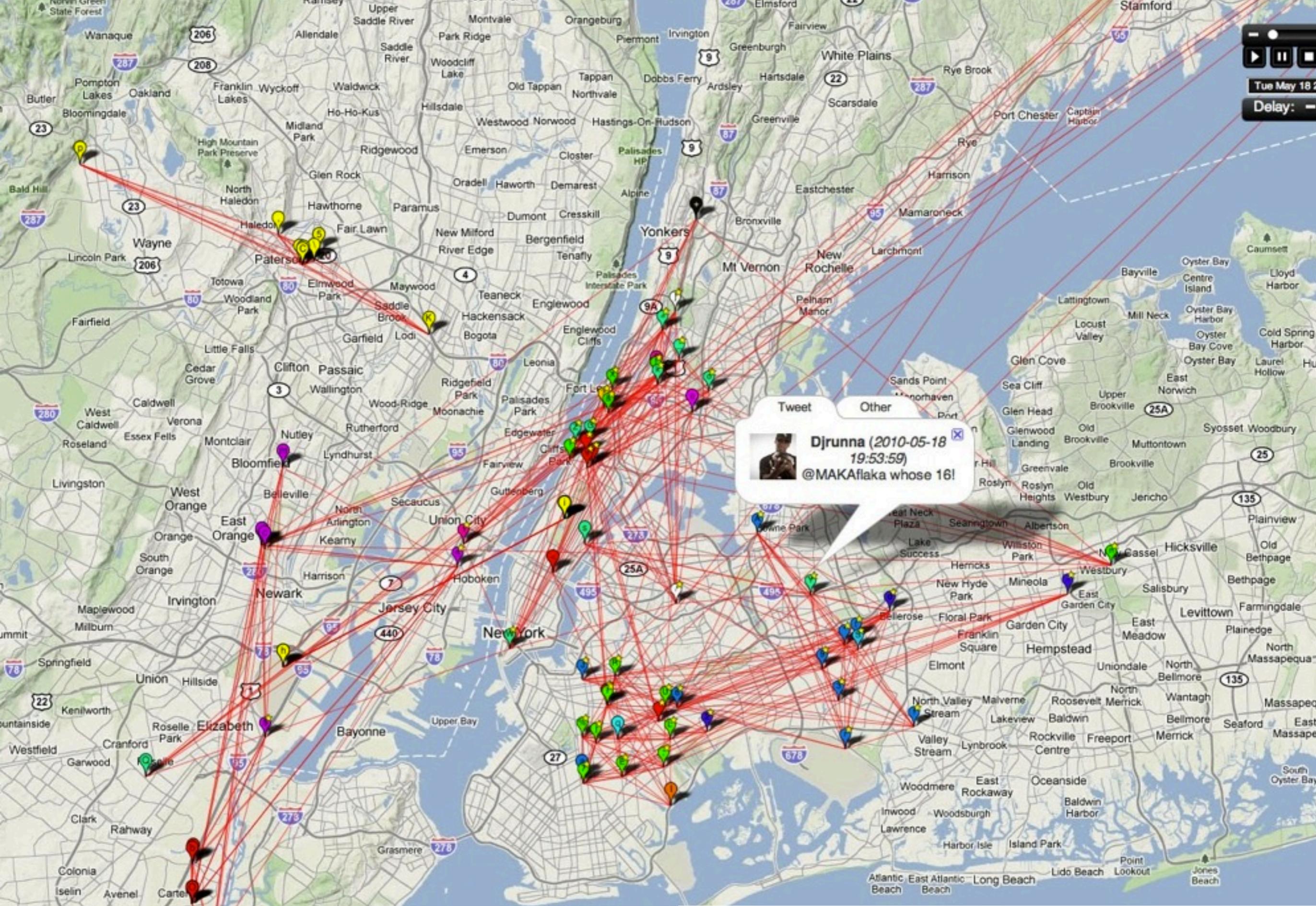
# The Data

- 11K **geo-active** users
- 4M **tweets** by geo-active users
  - 123K “**follows**” relationships

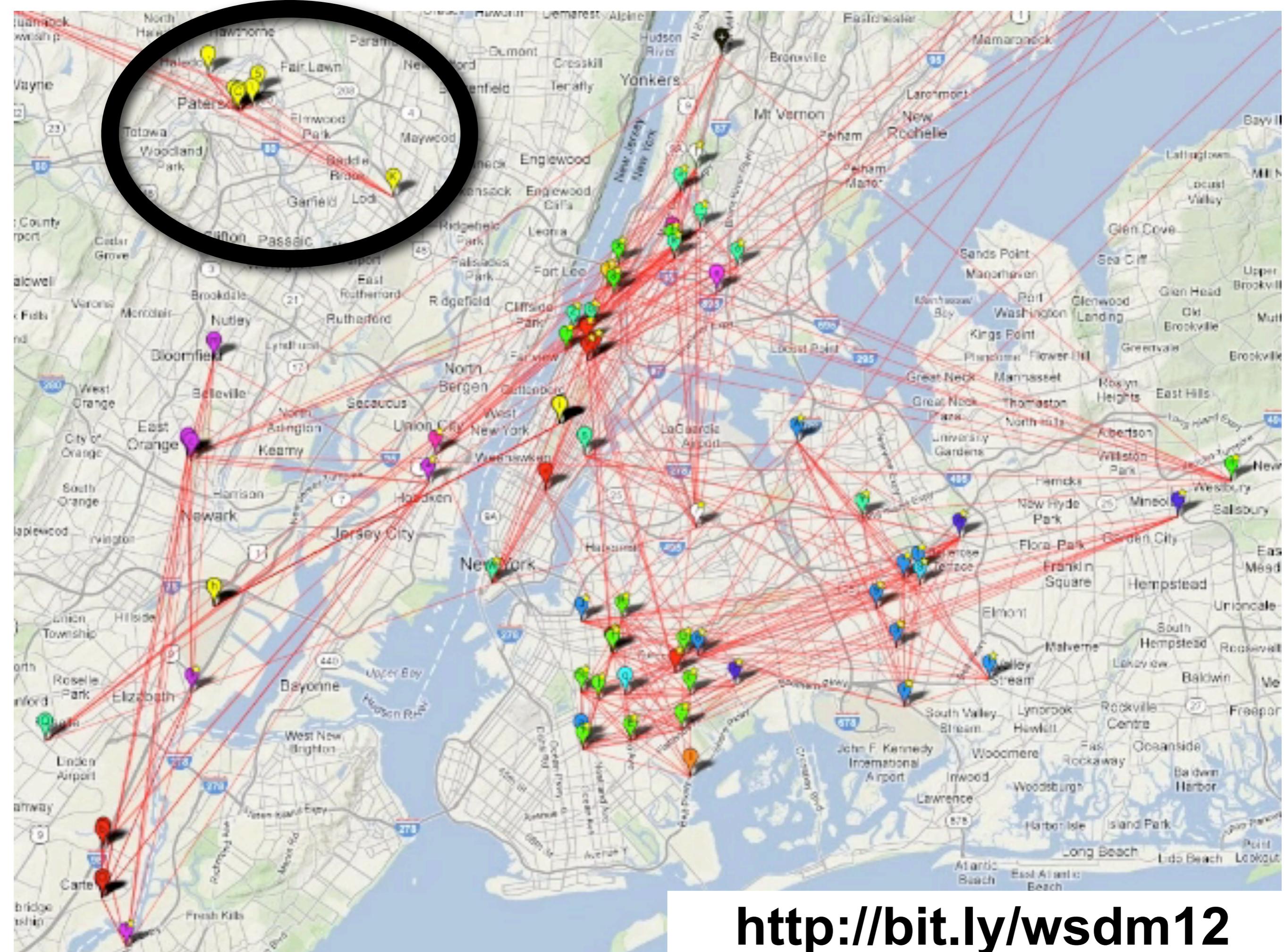
# The Data

- 11K **geo-active** users
- 4M **tweets** by geo-active users
  - 123K “**follows**” relationships
  - 52K “**friends**” relationships

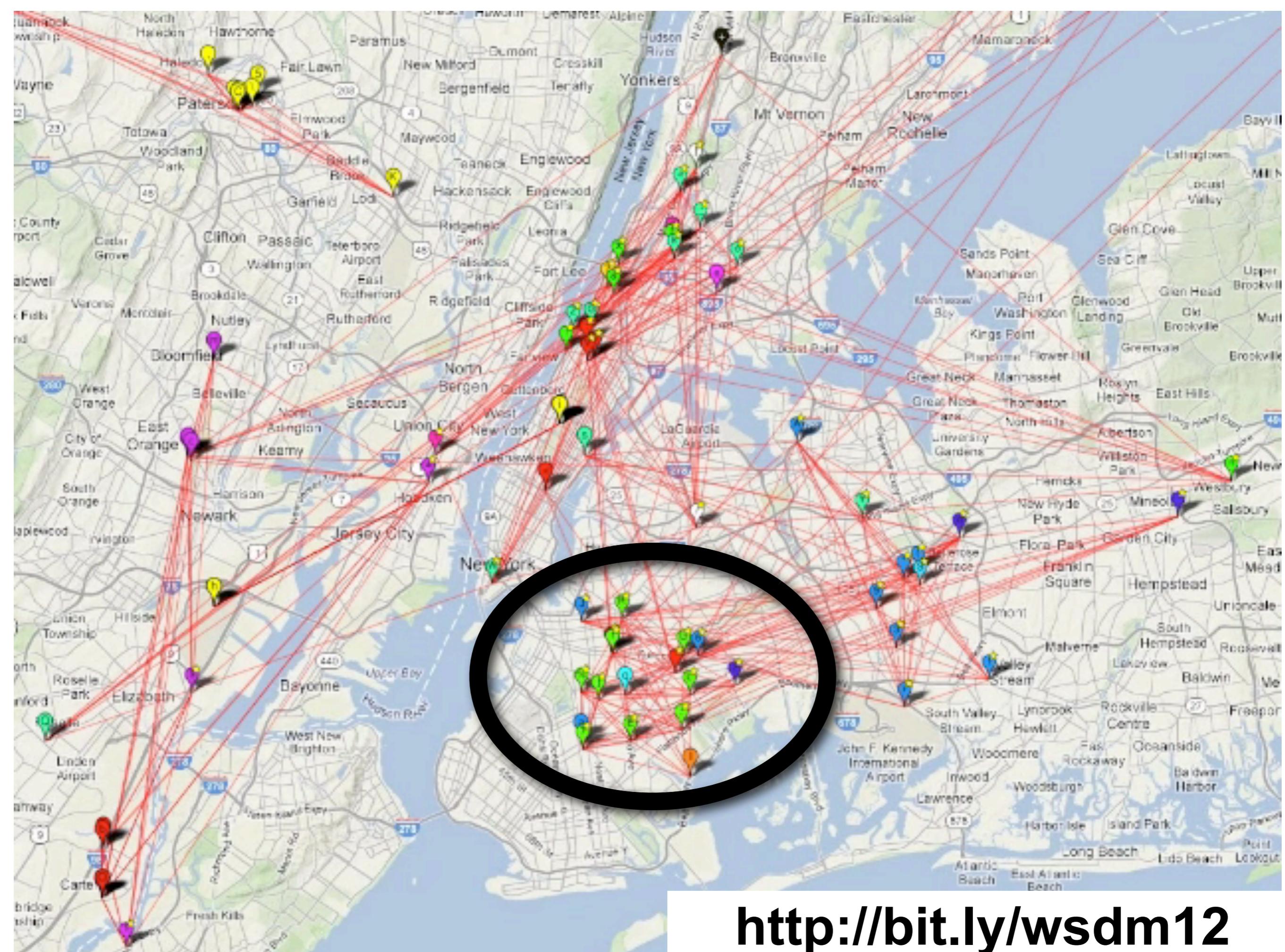




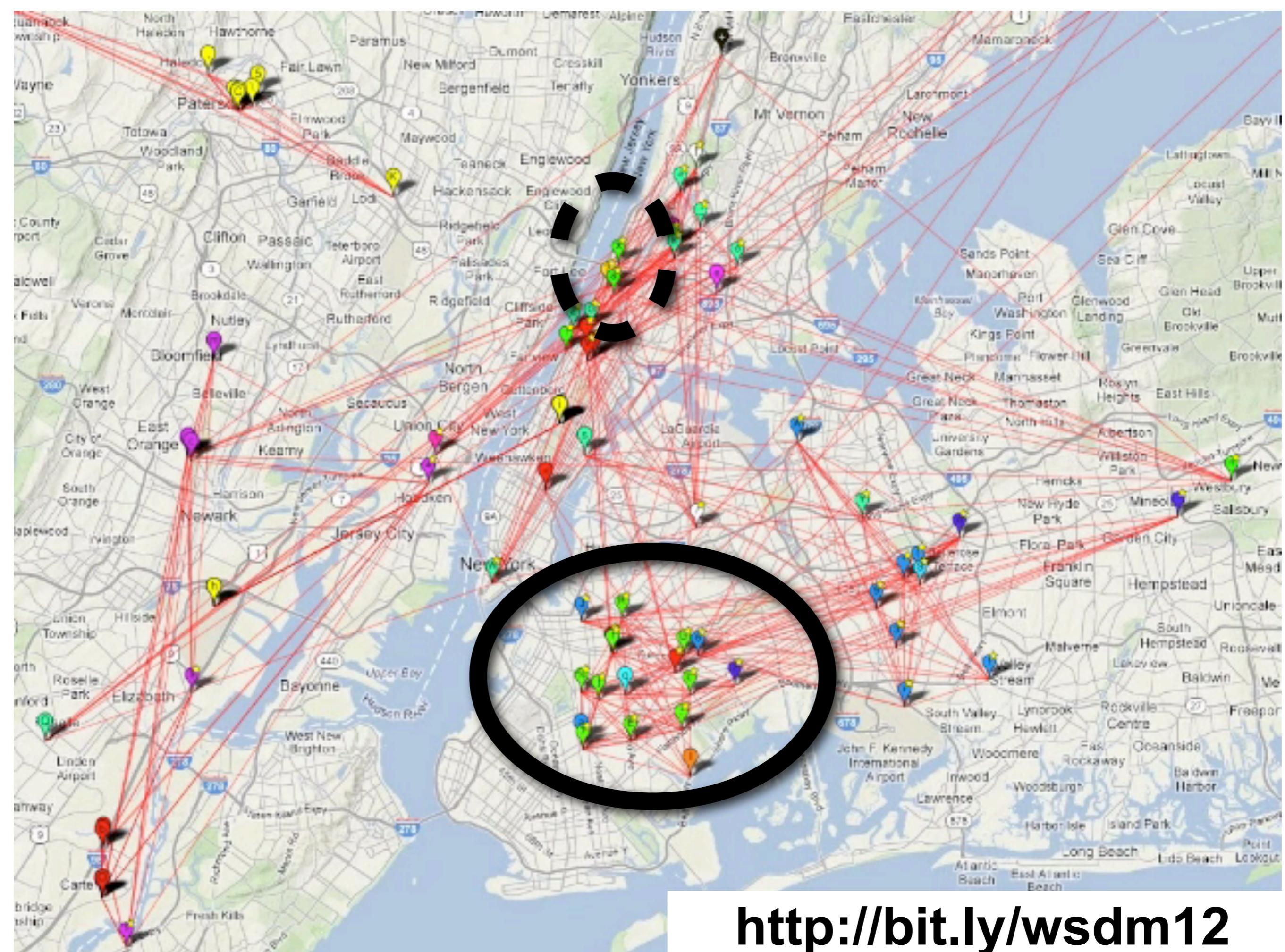
<http://bit.ly/wsdm12>



<http://bit.ly/wsdm12>



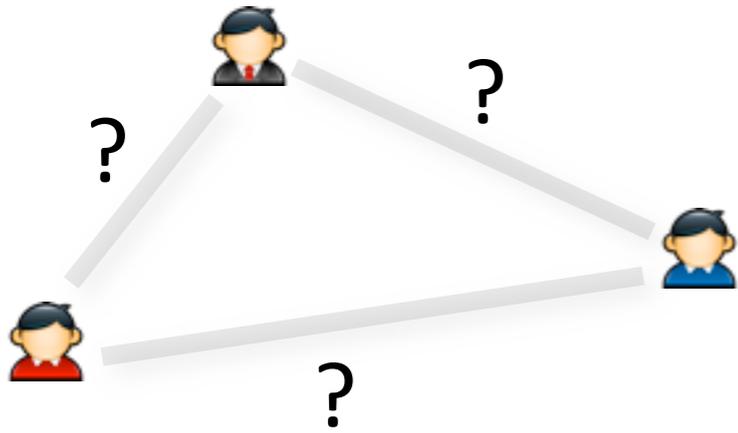
<http://bit.ly/wsdm12>



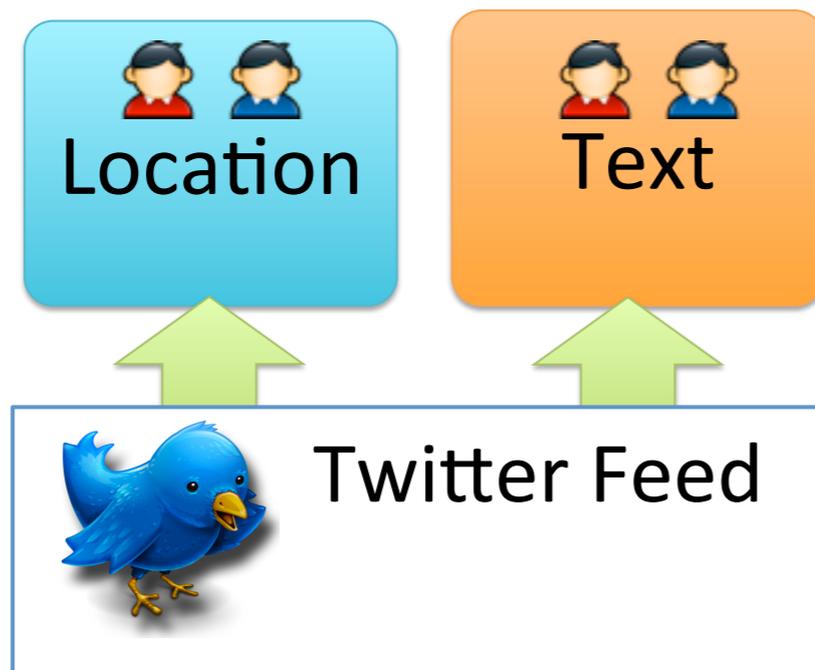
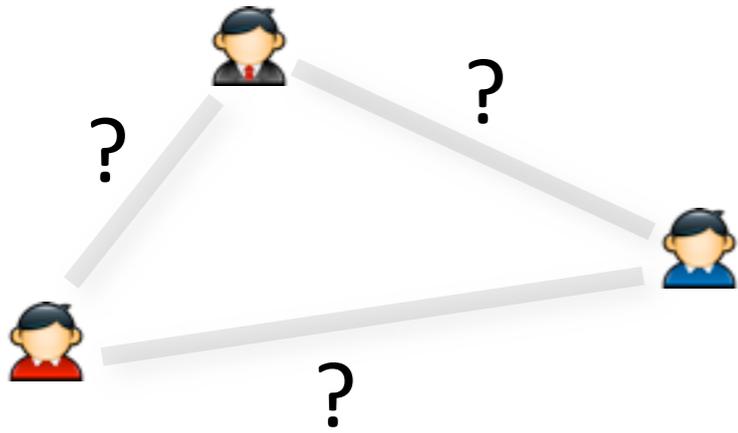
<http://bit.ly/wsdm12>

# Predicting Social Ties

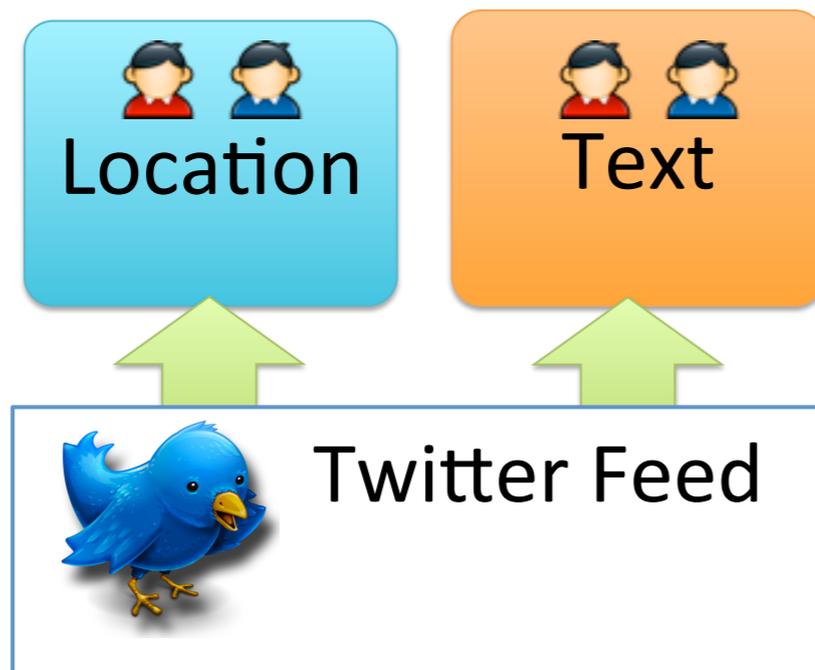
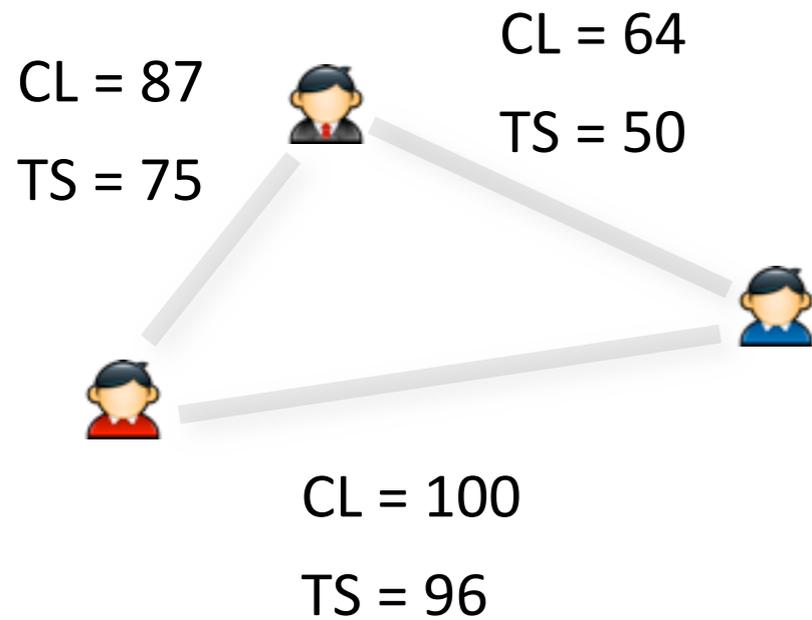
# Predicting Social Ties



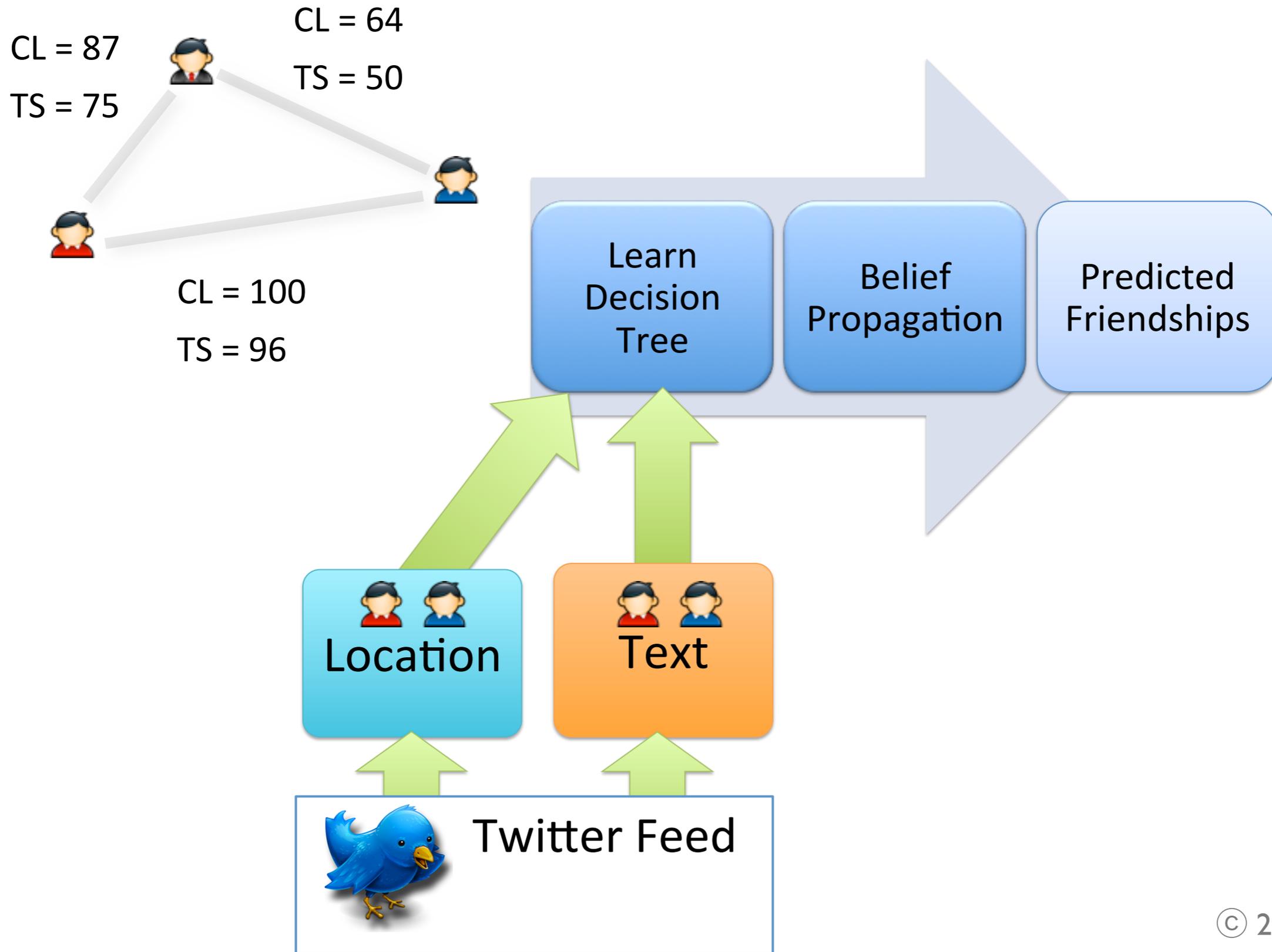
# Predicting Social Ties



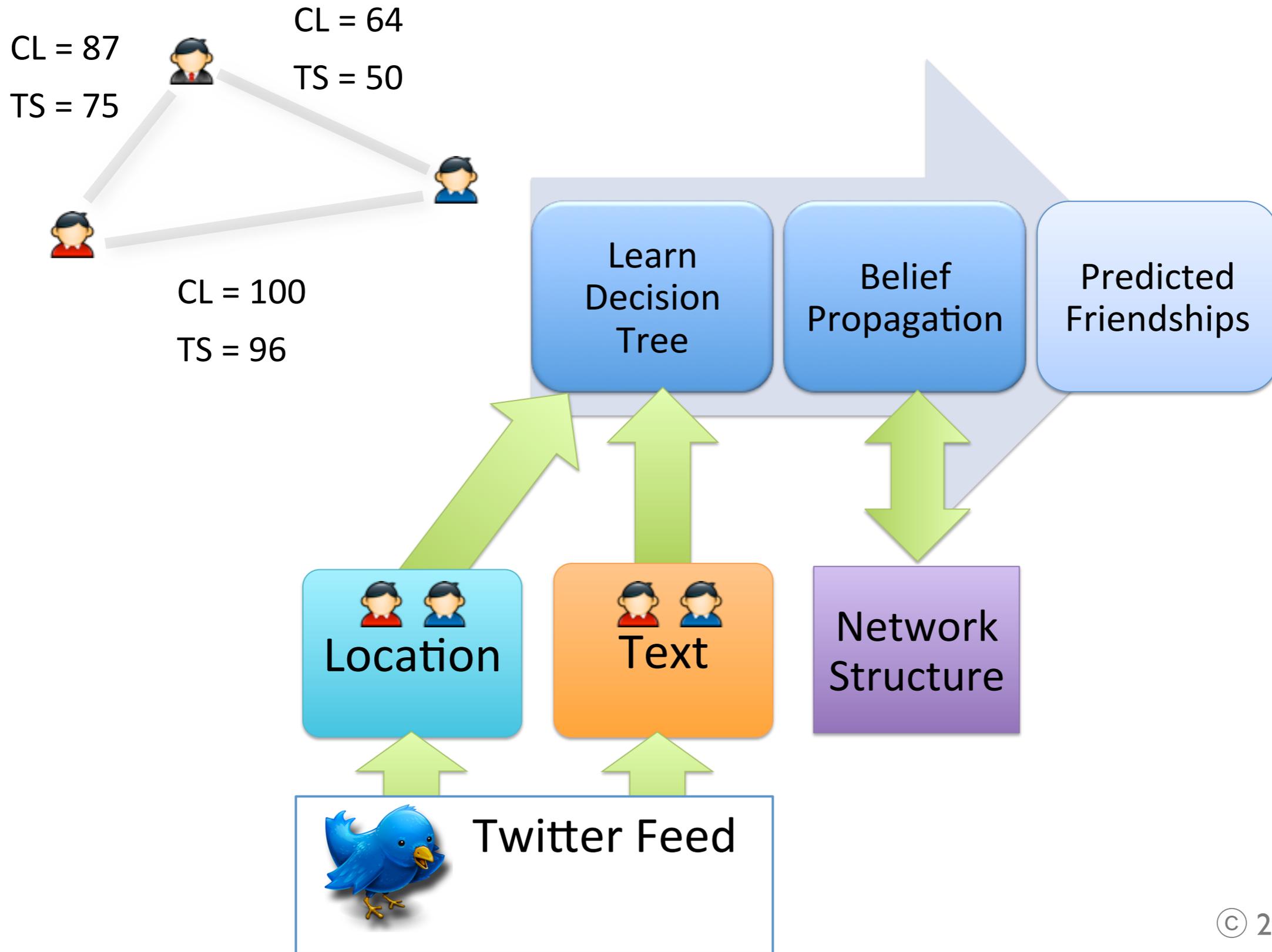
# Predicting Social Ties



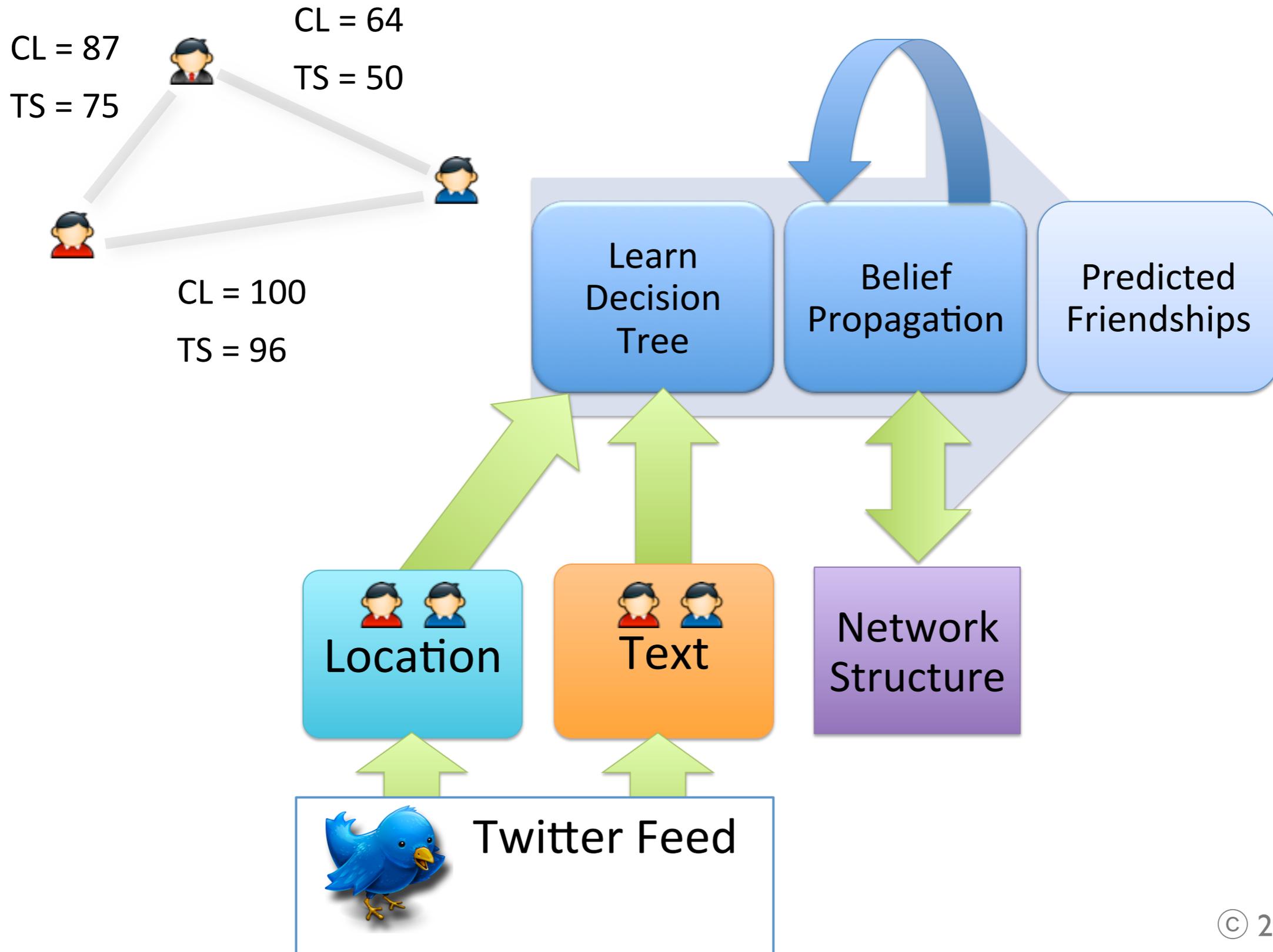
# Predicting Social Ties



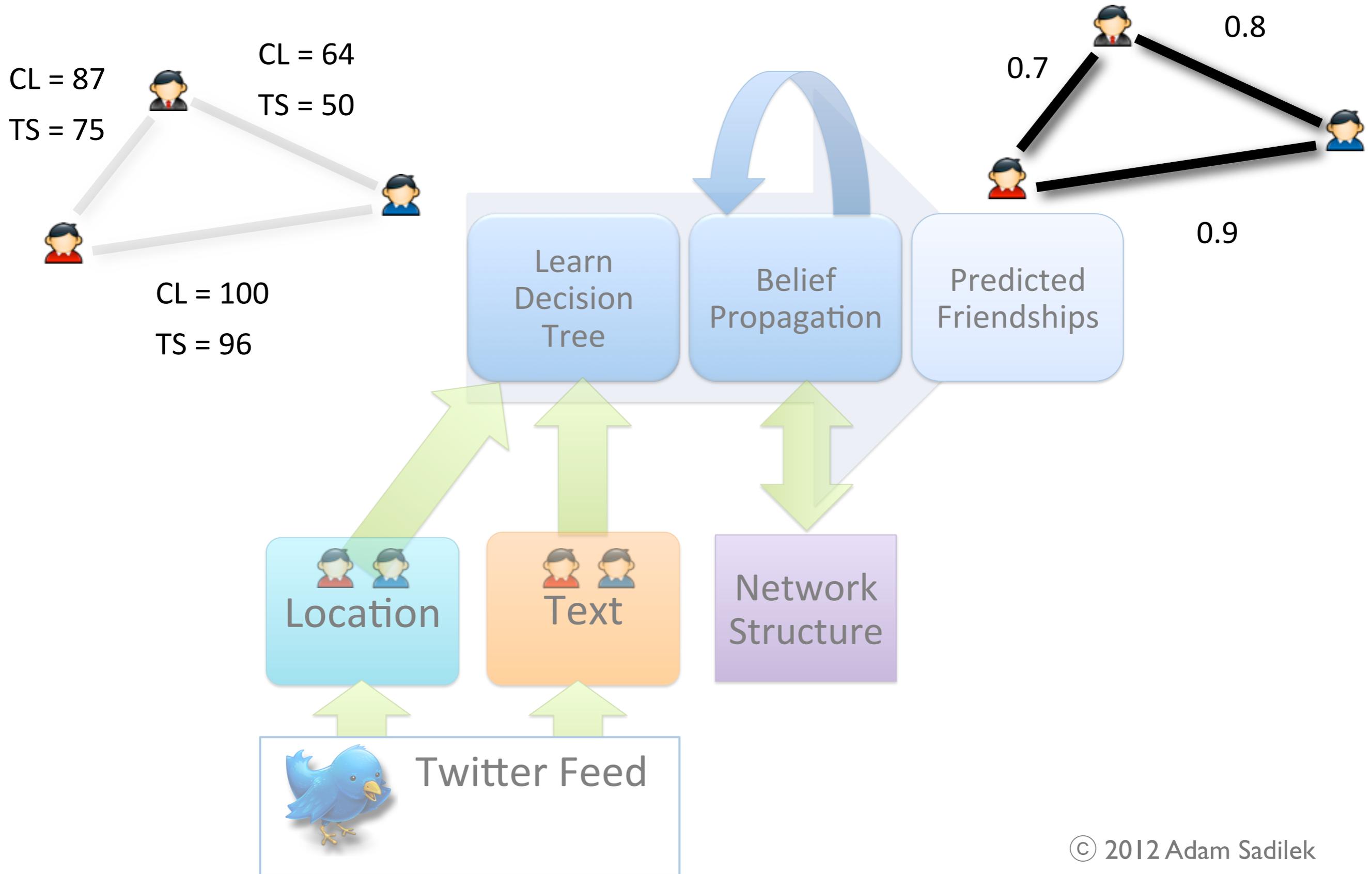
# Predicting Social Ties



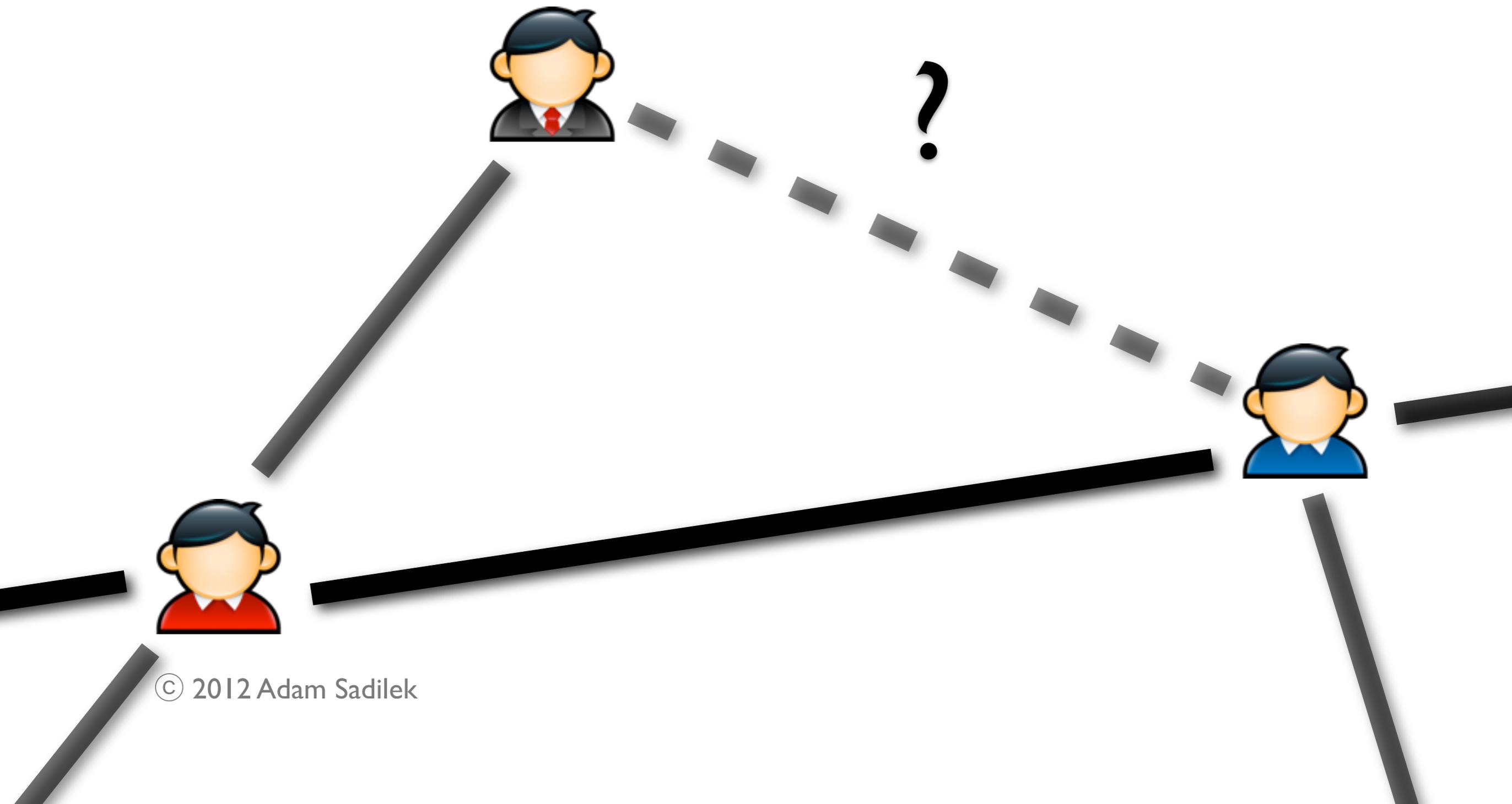
# Predicting Social Ties



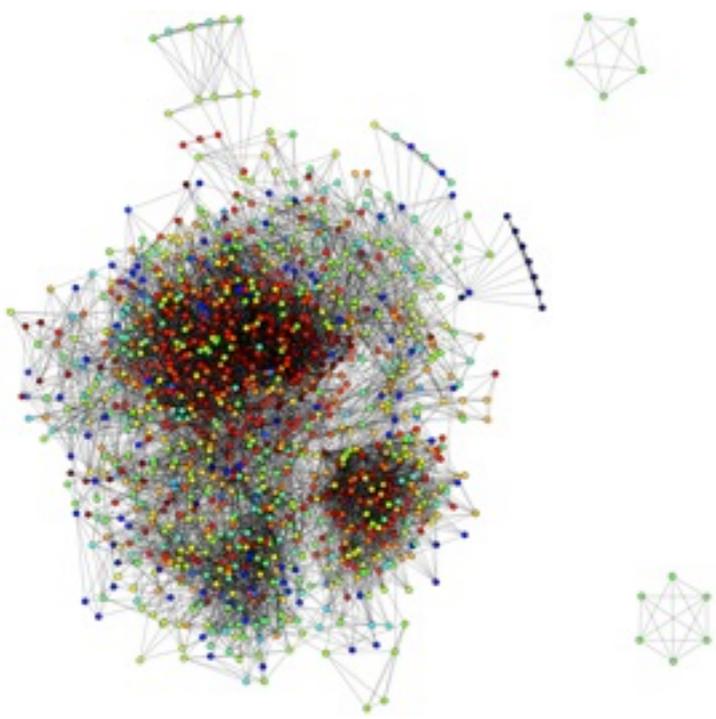
# Predicting Social Ties



# Leveraging Structure

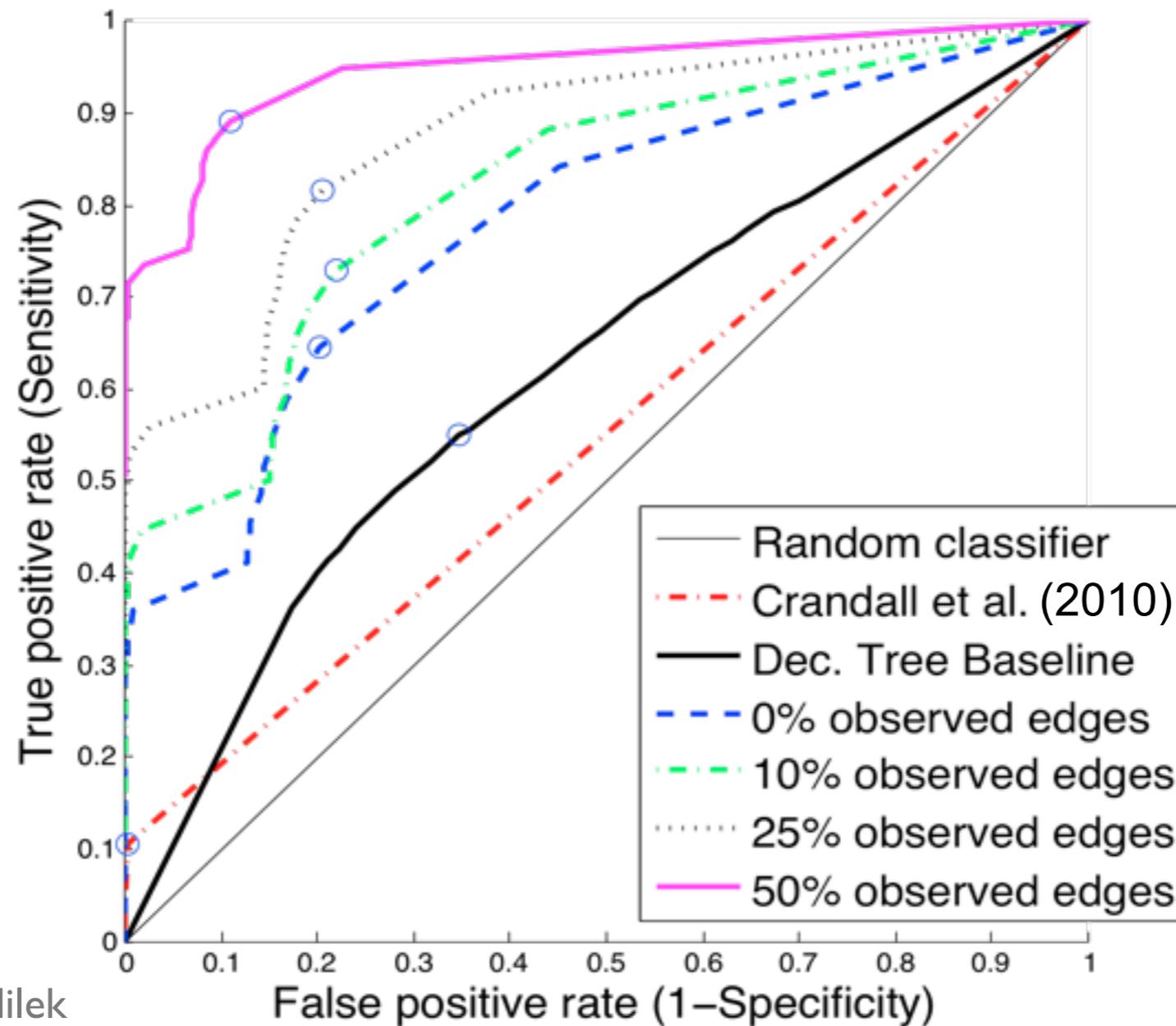
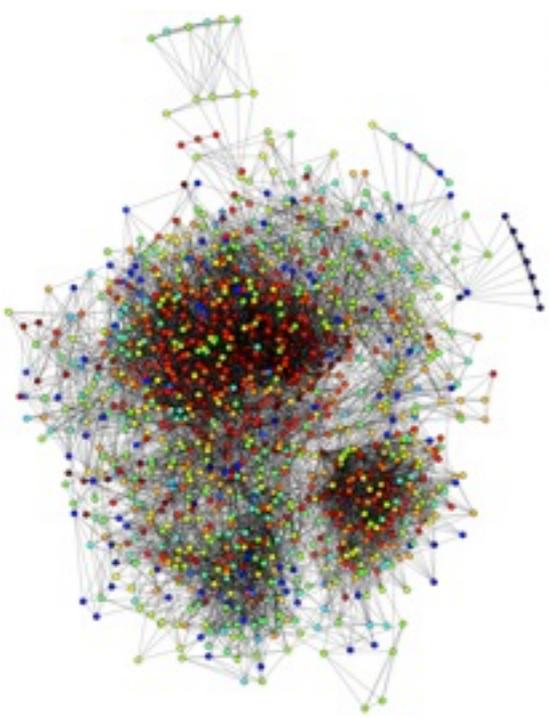


# Evaluation



- Cross-validation
  - New York City → Los Angeles
  - Los Angeles → New York City

# Predicting Social Ties





# Predicting Location



# Predicting Location

- Friends ~ noisy “sensors” of one’s location



# Predicting Location

- Find significant locations



# Predicting Location

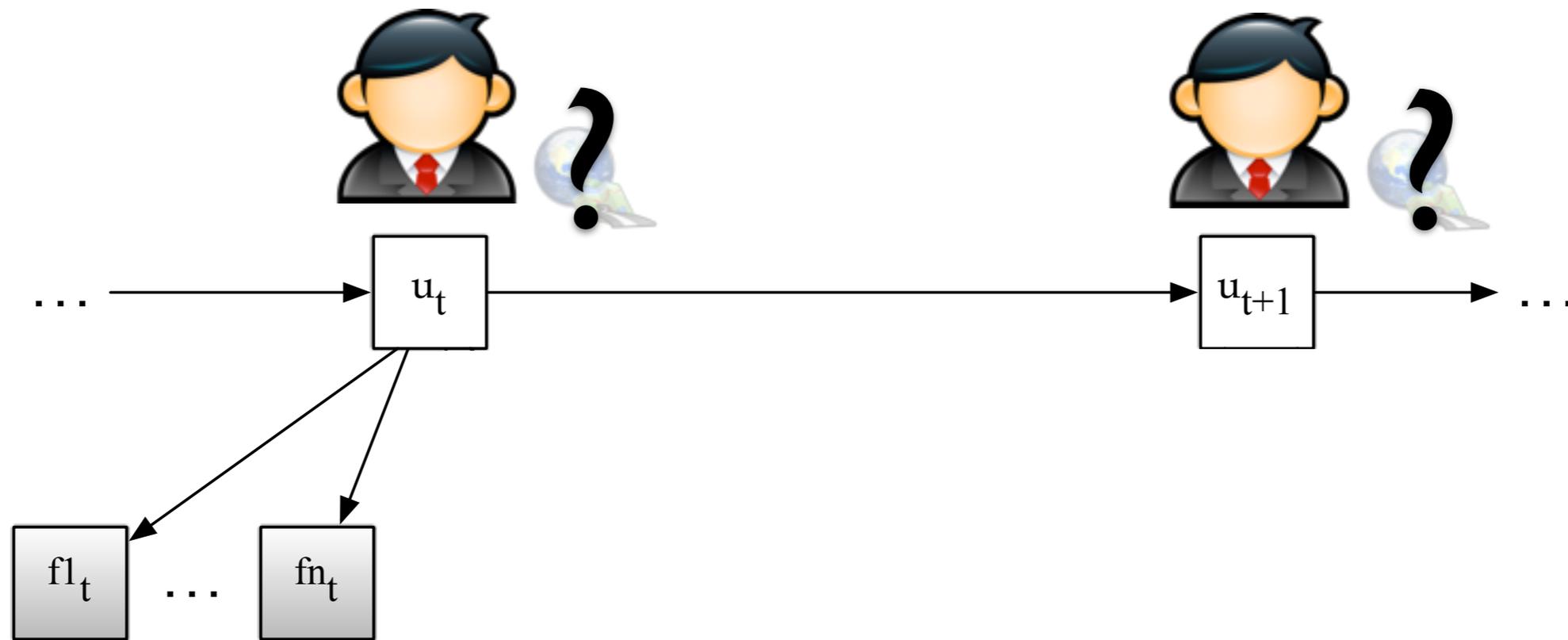
- Find significant locations
  - 90K with 100m granularity

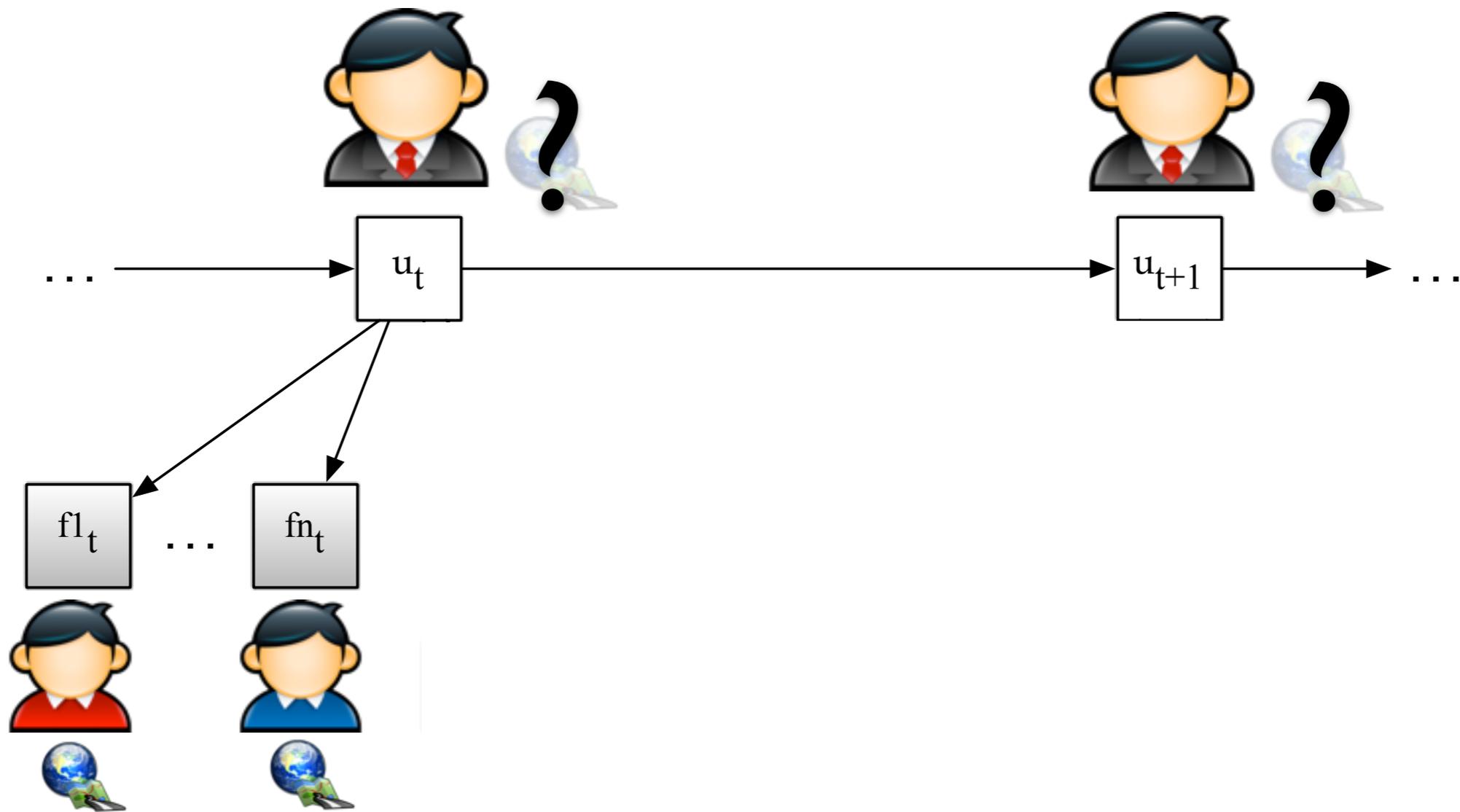


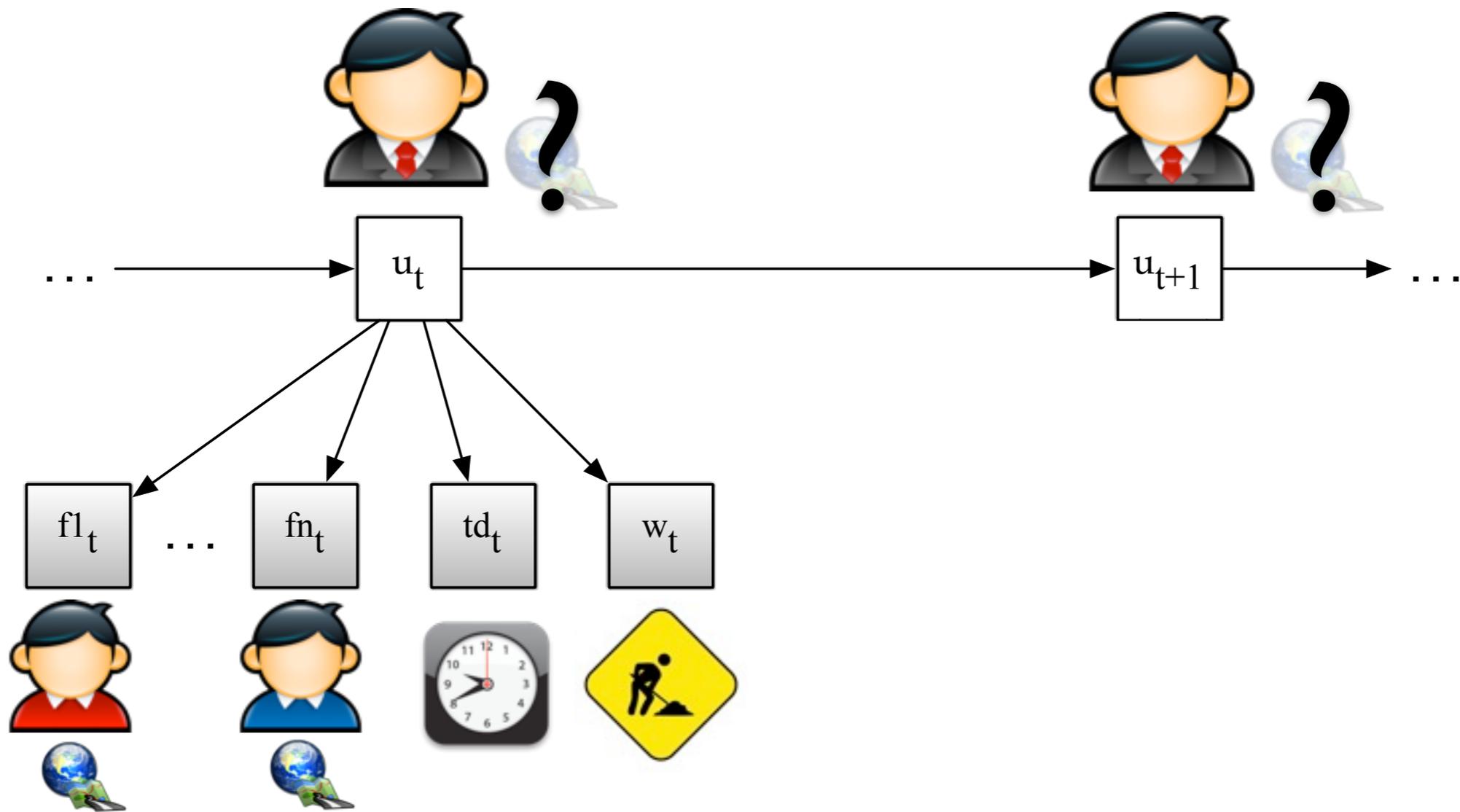
# Predicting Location

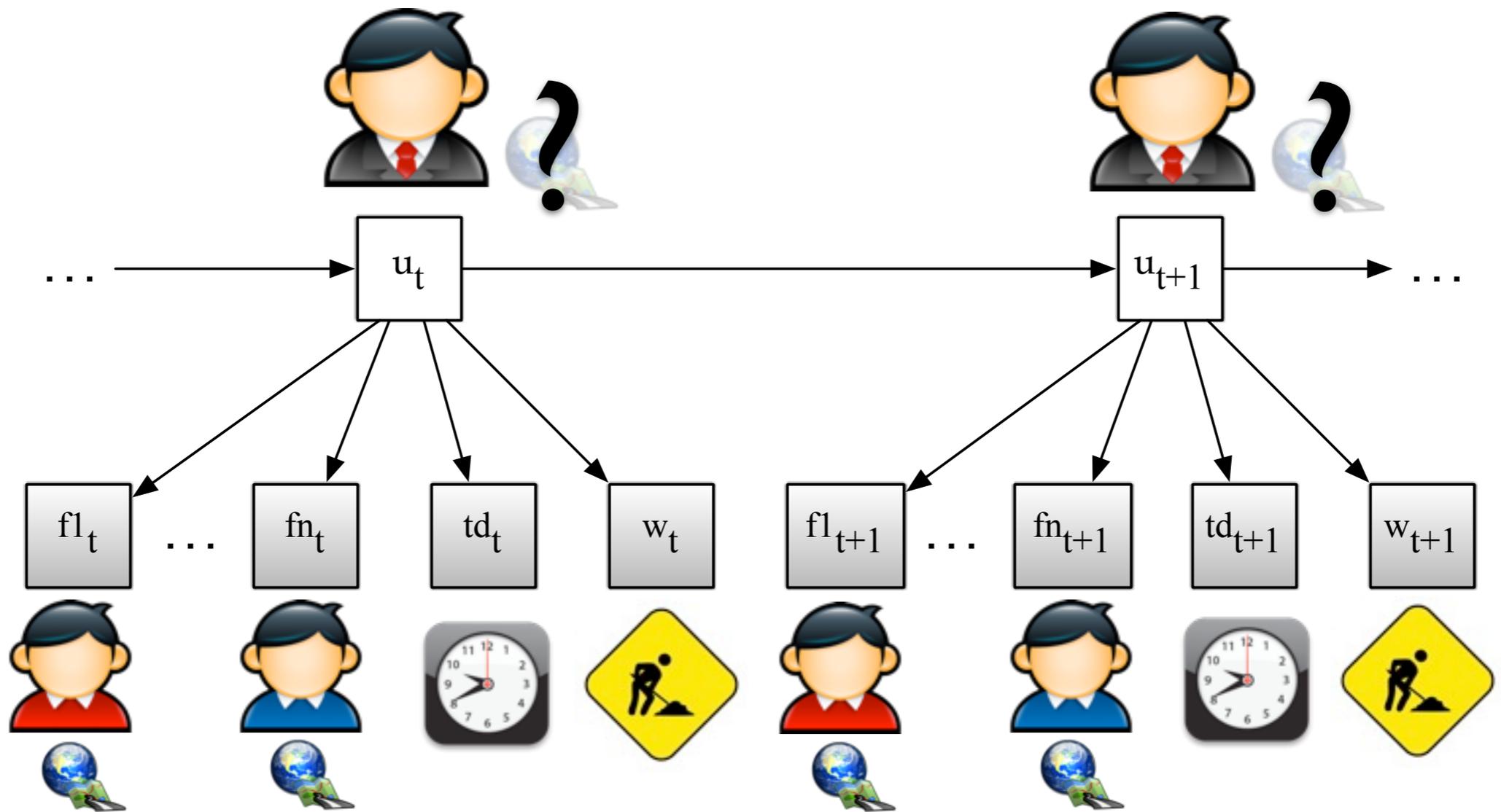
- Find significant locations
  - 90K with 100m granularity
  - 26K visited  $>4$  times













# Evaluation

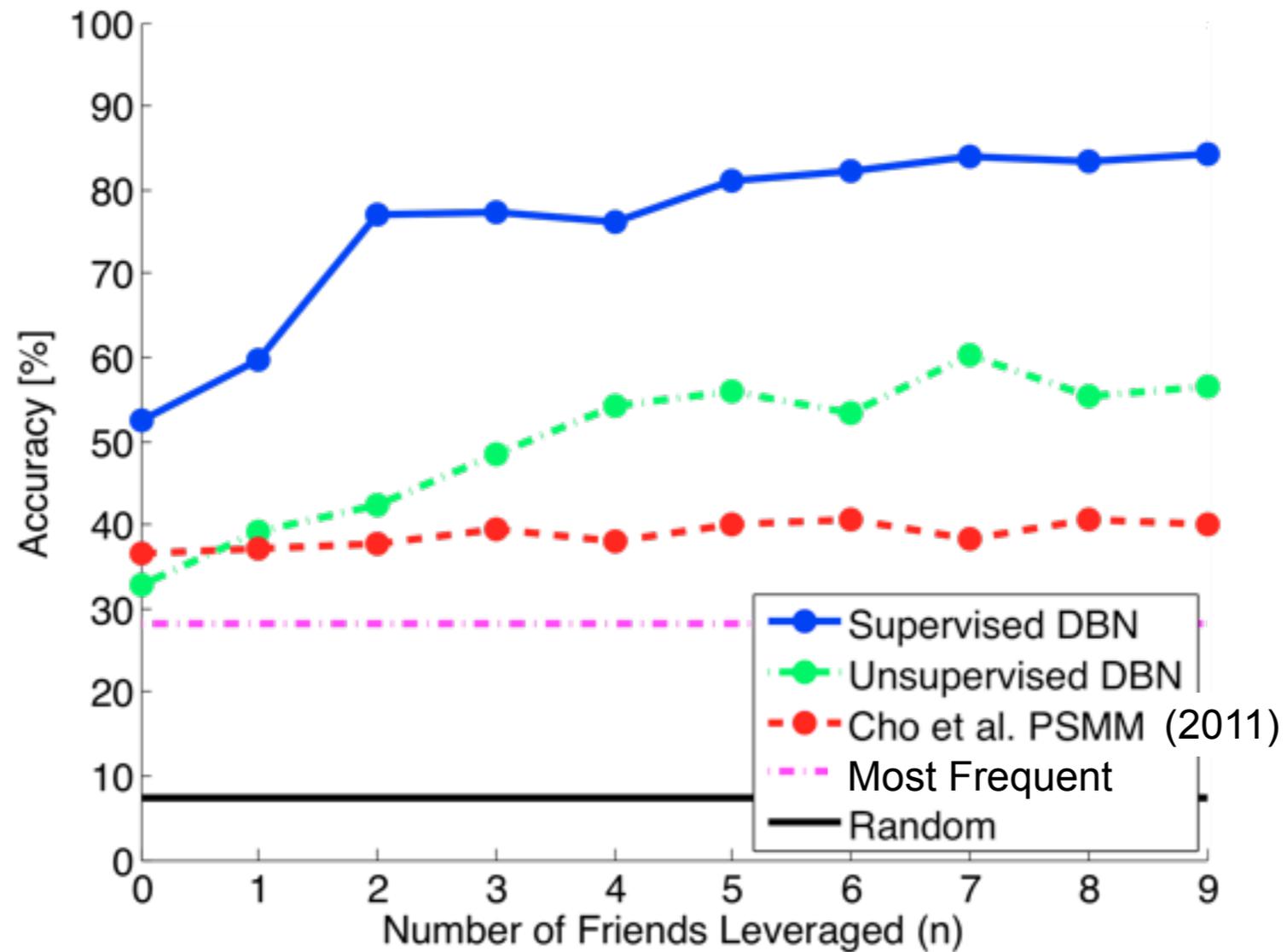


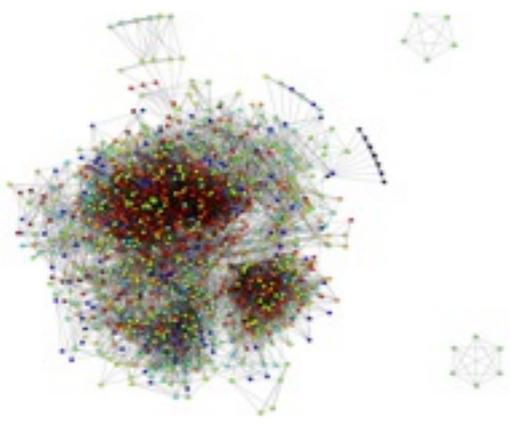
# Evaluation

- Cross-validation over all users
- Train on first 3 weeks, test on the 4<sup>th</sup>



# Predicting Location

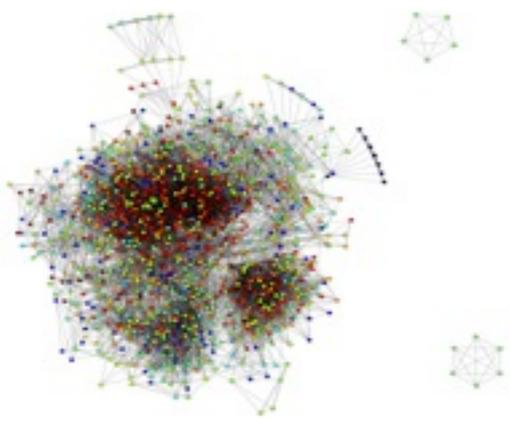




# Summary



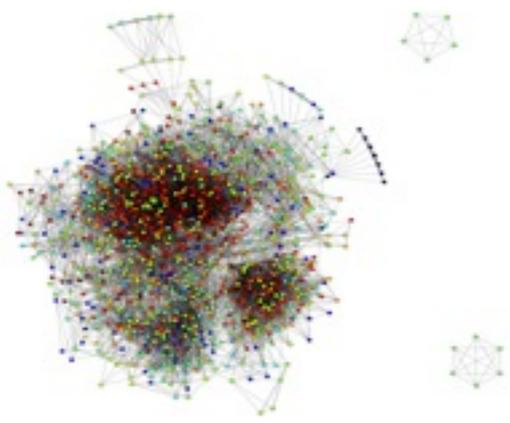
- Predicting **social ties** and **location**



# Summary



- Predicting **social ties** and **location**
  - Large scale



# Summary



- Predicting **social ties** and **location**
  - Large scale
- Evaluation on Twitter data from NYC and LA

# Future Work

- **Unify** our two approaches
- **Fine-grained** modeling
  - General
  - Fashion, gossip, epidemics

# Finding Your Friends and Following Them to Where You Are



UNIVERSITY *of*  
ROCHESTER

Adam Sadilek  
Henry Kautz  
Jeffrey P. Bigham

$$y_{1:t}^* = \operatorname{argmax}_{y_{1:t}} \log (\operatorname{Pr}(y_{1:t}|x_{1:t}))$$

$$\theta^* = \operatorname{argmax}_{\theta} \log (\operatorname{Pr}(x_{1:t}, y_{1:t}|\theta))$$

$$\theta^* = \operatorname{argmax}_{\theta} \log \sum_{y_{1:t}} \operatorname{Pr}(x_{1:t}, y_{1:t}|\theta)$$