# Modeling Mass Protest Adoption in Social Network Communities using Geometric Brownian Motion

Fang Jin, Rupinder Paul Khandpur, Nathan Self, Edward Dougherty,

Sheng Guo, Feng Chen, B. Aditya Prakash, Naren Ramakrishnan

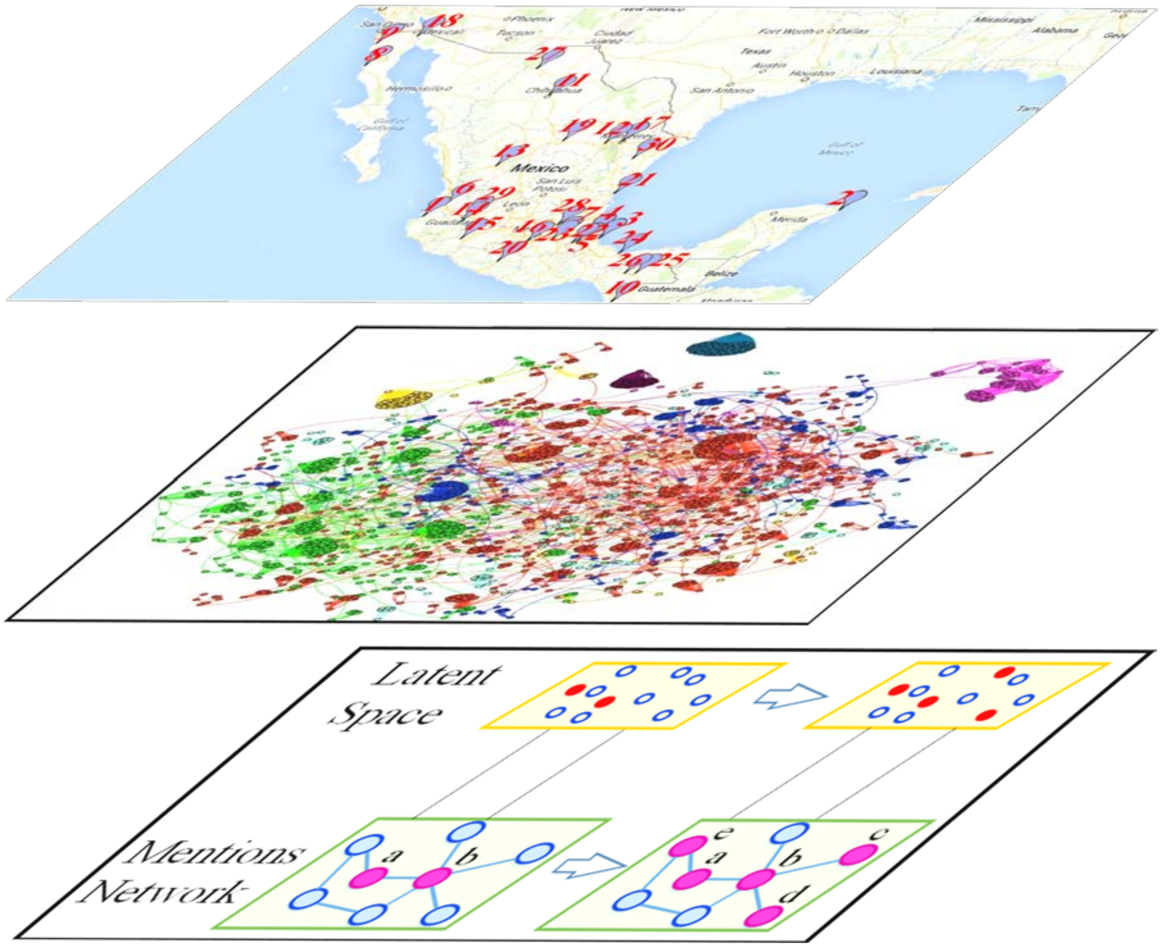
VirginiaTech Linkedin UNIVERSITY AT ALBANY State University of New York

# #Yosoy132 Protest

# #Yosoy132 Protest

# #Yosoy132 Protest

# Motivation




Protest examples

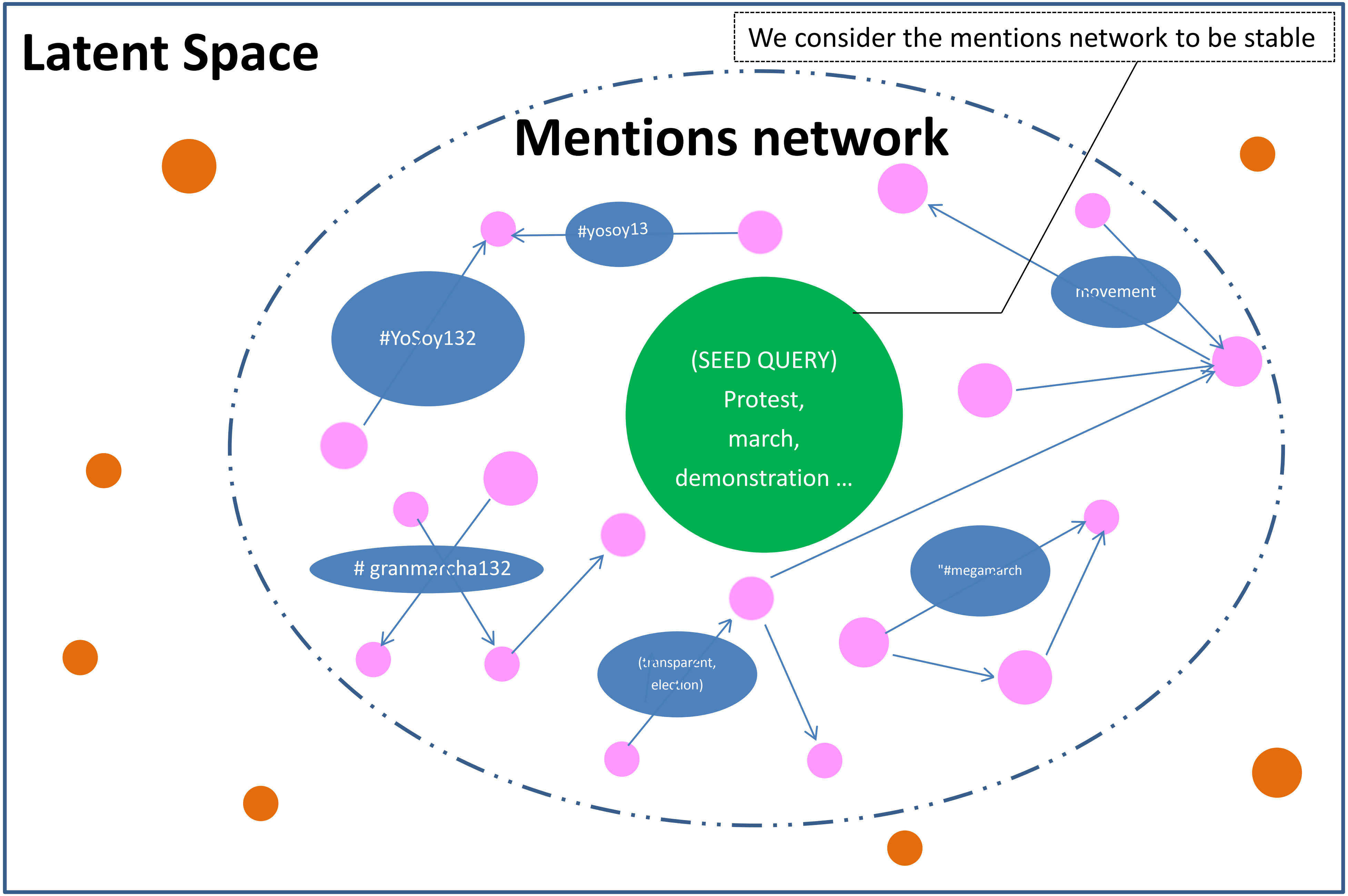
Mass movement in Twitter

Propagation model

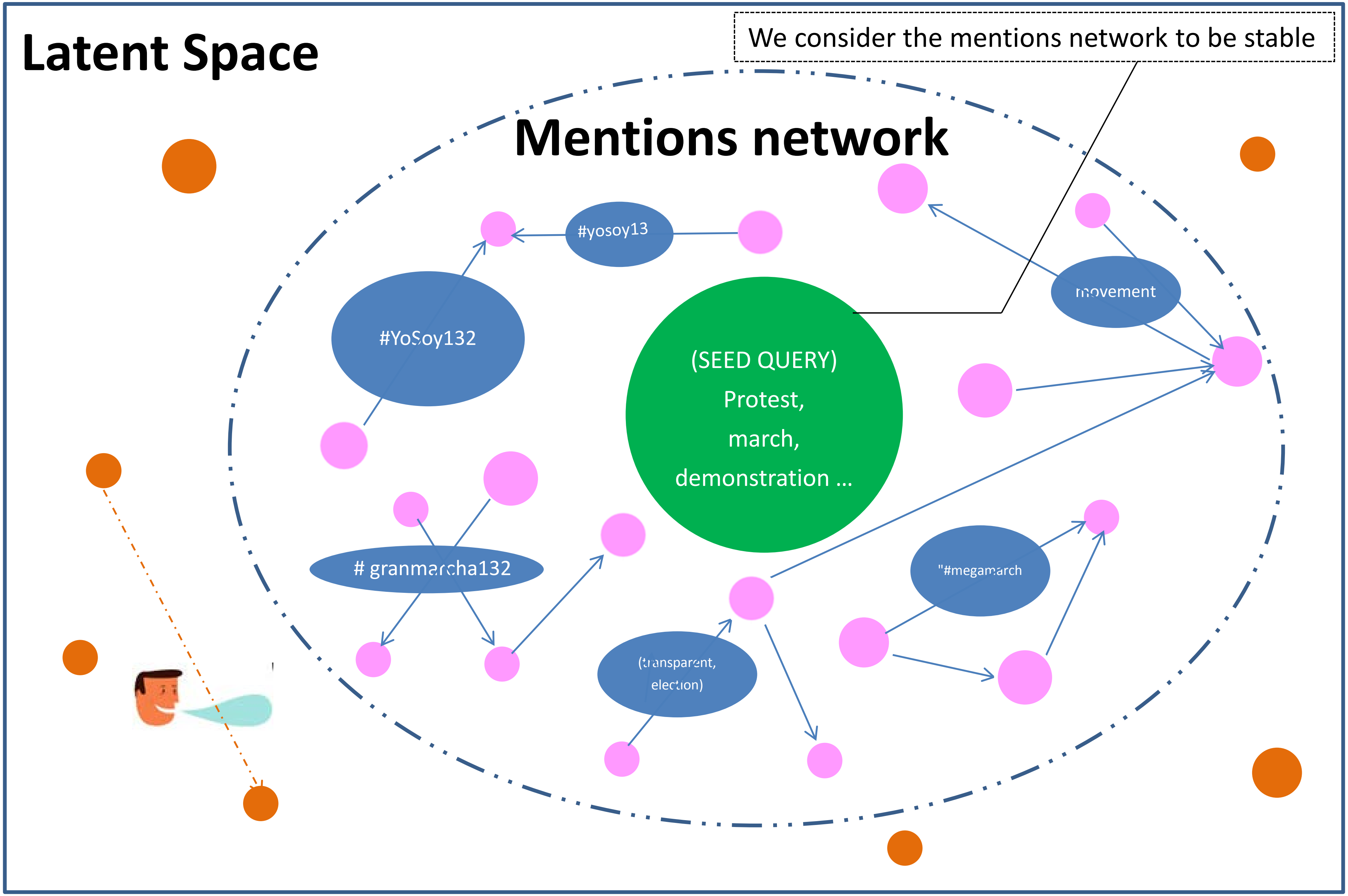
**Goal:**

- Model the growth of protest participants within a social network.
- Understand the underlying social network and structural dynamics.

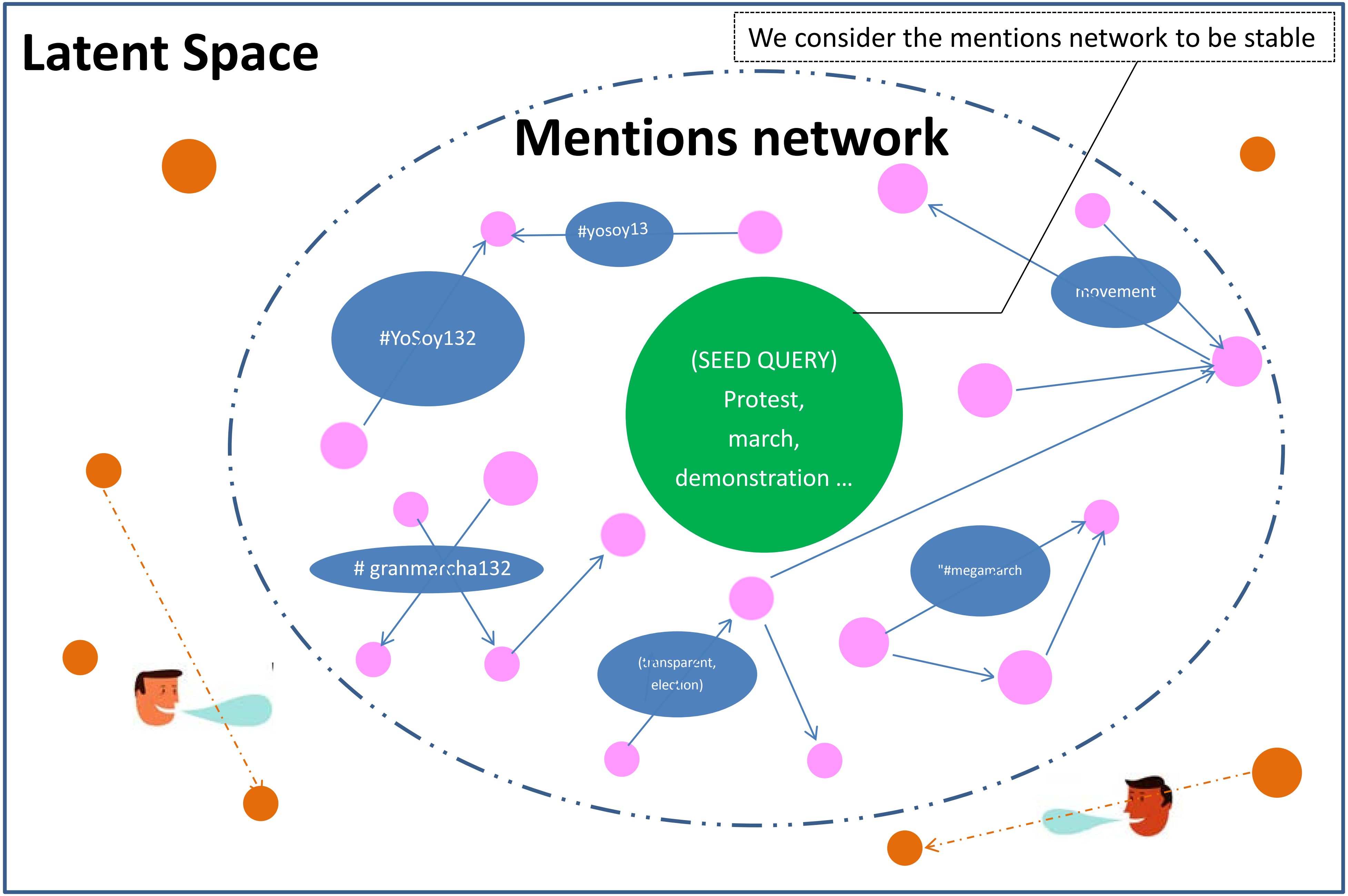
# Approach: Bispace model



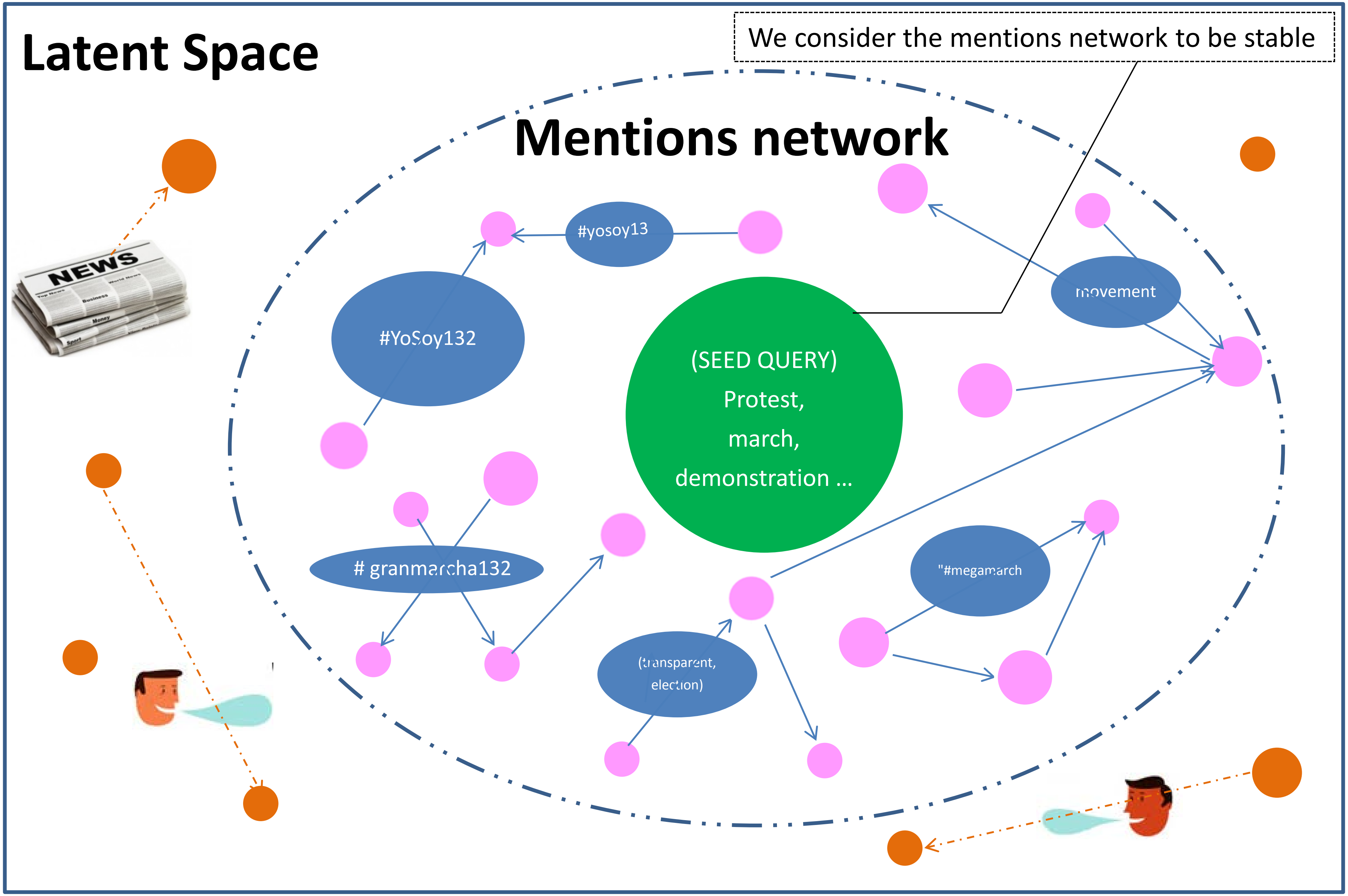
# Approach: Bispace model



# Approach: Bispace model

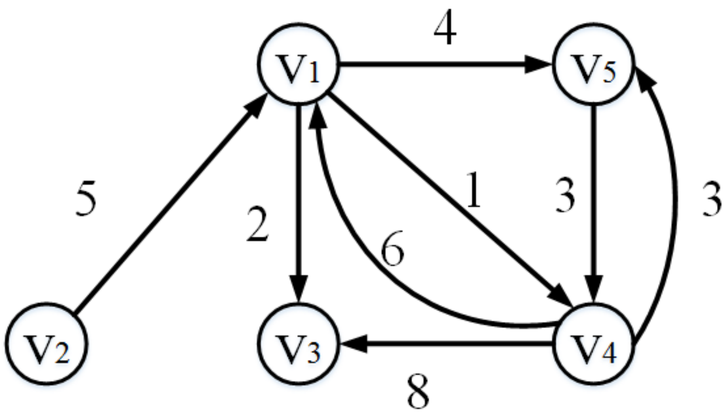


# Approach: Bispace model



# Geometric Brownian motion (GBM)

## Brownian distance



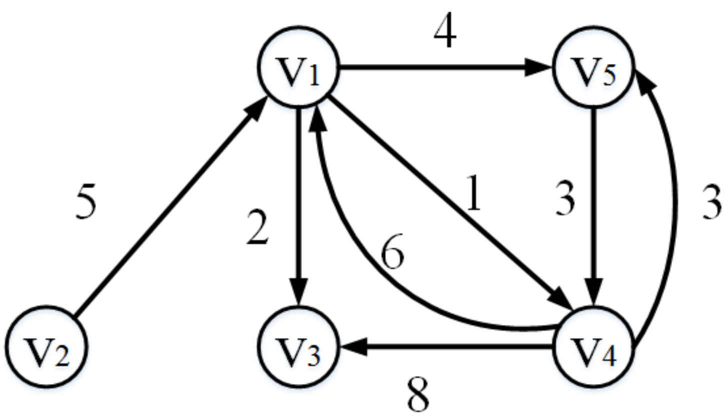

$$d_{ij} = \frac{1}{(\omega_{ij}+1)(\omega_{ji}+1)^{\gamma}(\eta_{ij}+1)^{\gamma}}, \gamma \geq 1$$

$$eg: d_{14} = \frac{1}{(1+1)(6+1)(2+1)}, \gamma = 1$$

- We define the Brownian distance, which depends on the mention frequency and number of common direct neighbors
- Brownian distance is a measurement of the closeness and diffusion possibility of the nodes

# Geometric Brownian motion (GBM)

## Brownian distance

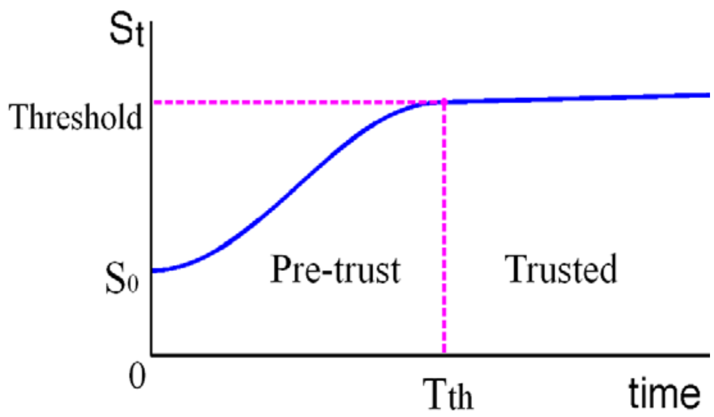



$$d_{ij} = \frac{1}{(\omega_{ij}+1)(\omega_{ji}+1)^{\gamma}(\eta_{ij}+1)^{\gamma}}, \gamma \geq 1$$

$$eg: d_{14} = \frac{1}{(1+1)(6+1)(2+1)}, \gamma = 1$$

- We define the Brownian distance, which depends on the mention frequency and number of common direct neighbors
- Brownian distance is a measurement of the closeness and diffusion possibility of the nodes

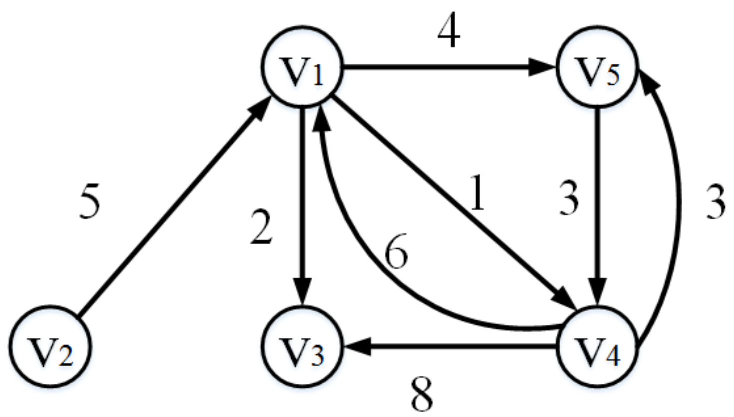
## Trust function




- We define $v^j$ trust function with $v^i$ at time t as $S_t^{ij}$
- We modeled the trust function $S_t$ as a GBM process: $S_t^{ij} \sim GBM$

# Geometric Brownian motion (GBM)

## Brownian distance

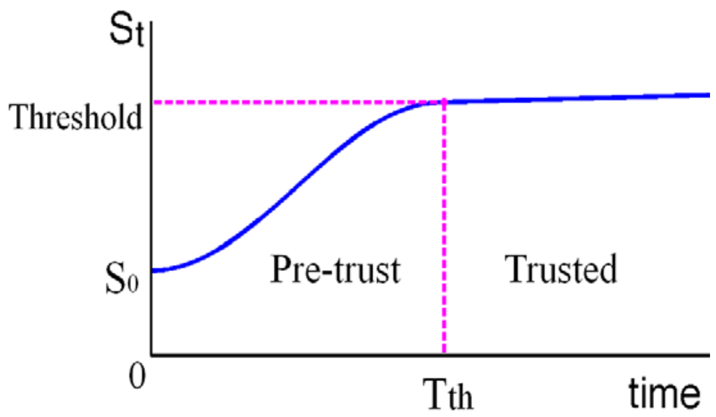



$$d_{ij} = \frac{1}{(\omega_{ij}+1)(\omega_{ji}+1)^{\gamma}(\eta_{ij}+1)^{\gamma}}, \gamma \geq 1$$

$$eg: d_{14} = \frac{1}{(1+1)(6+1)(2+1)}, \gamma = 1$$

- We define the Brownian distance, which depends on the mention frequency and number of common direct neighbors
- Brownian distance is a measurement of the closeness and diffusion possibility of the nodes

## Trust function




- We define $v^j$ trust function with $v^i$ at time t as $S_t^{ij}$
- We modeled the trust function $S_t$ as a GBM process: $S_t^{ij} \sim$ GBM

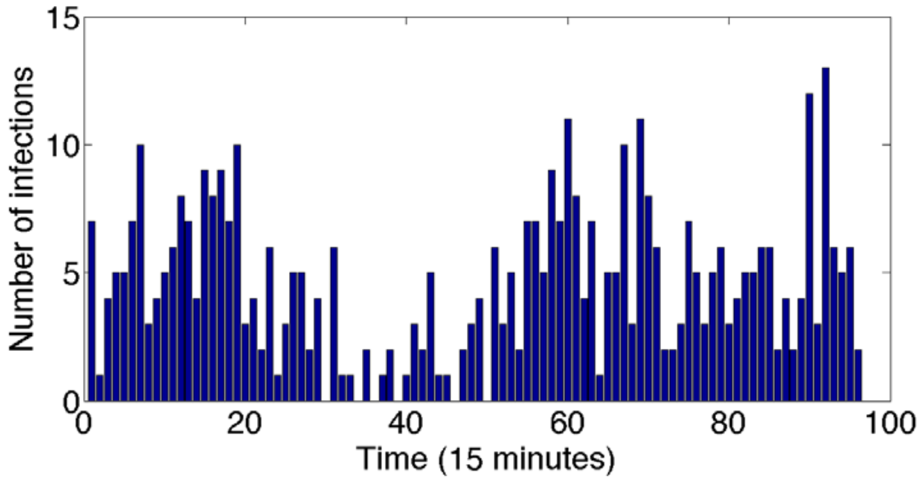
## Infection condition:

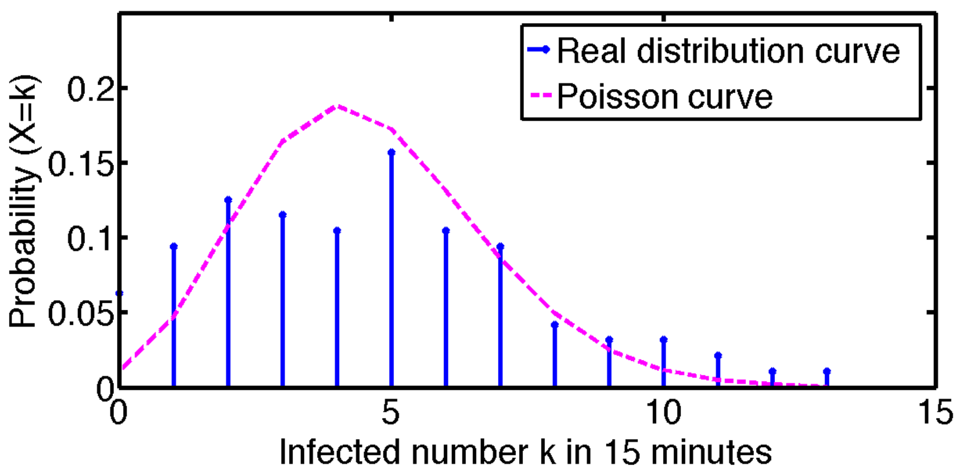
$\ln(S_t^{ij}) >= d_{ij}$

# Latent space: Poisson model

- We assume the probability of the number of newly infected users, X, in a given time interval satisfies the Poisson distribution:

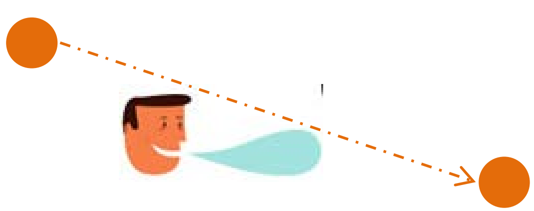
$$\Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

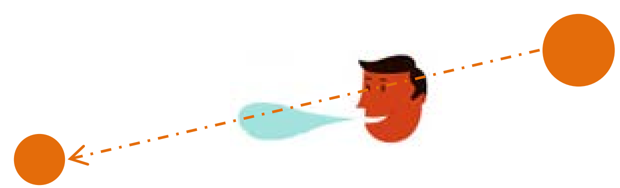


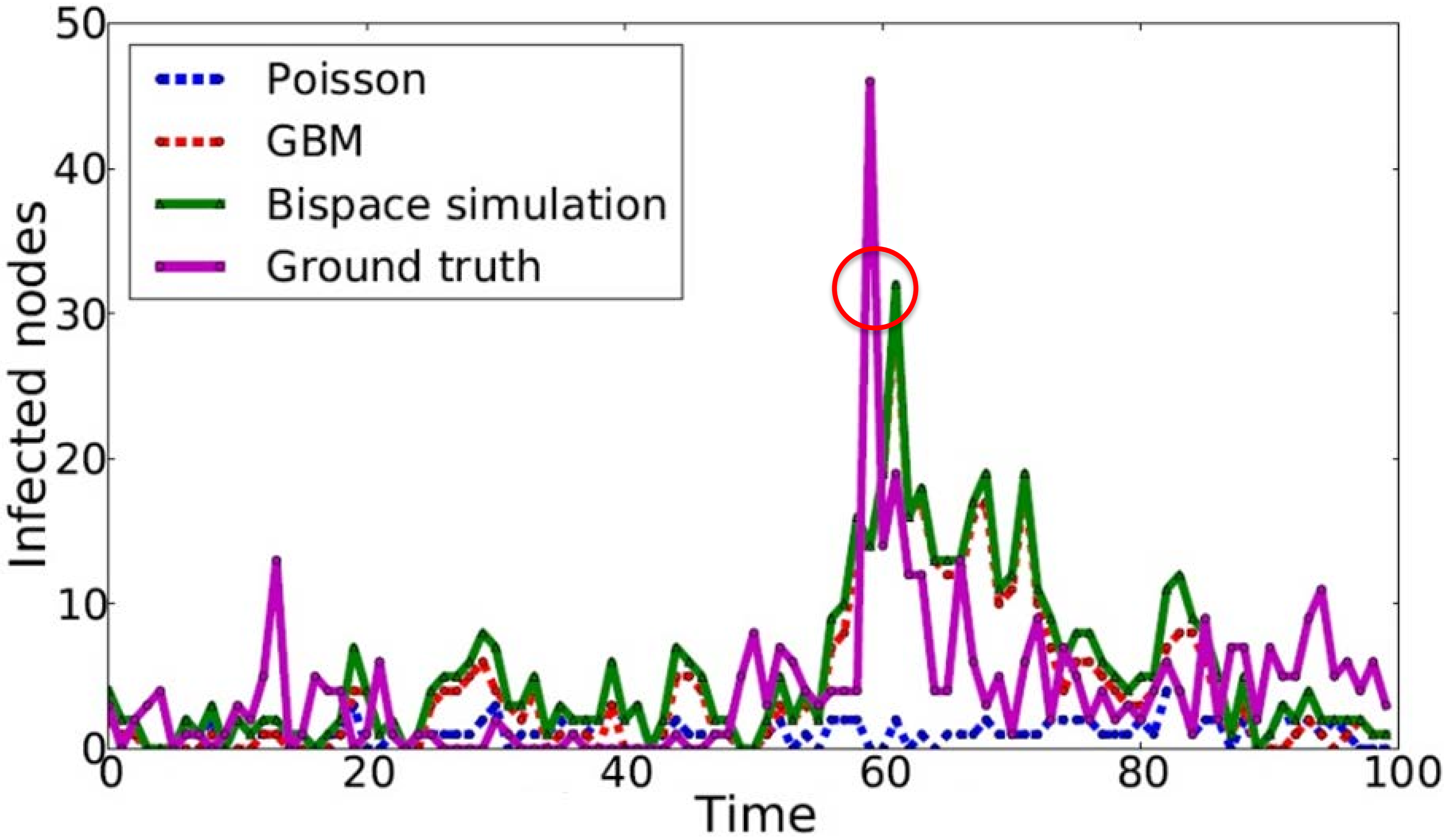

Infected nodes in latent space

Poisson distribution fit (λ = 4.18)



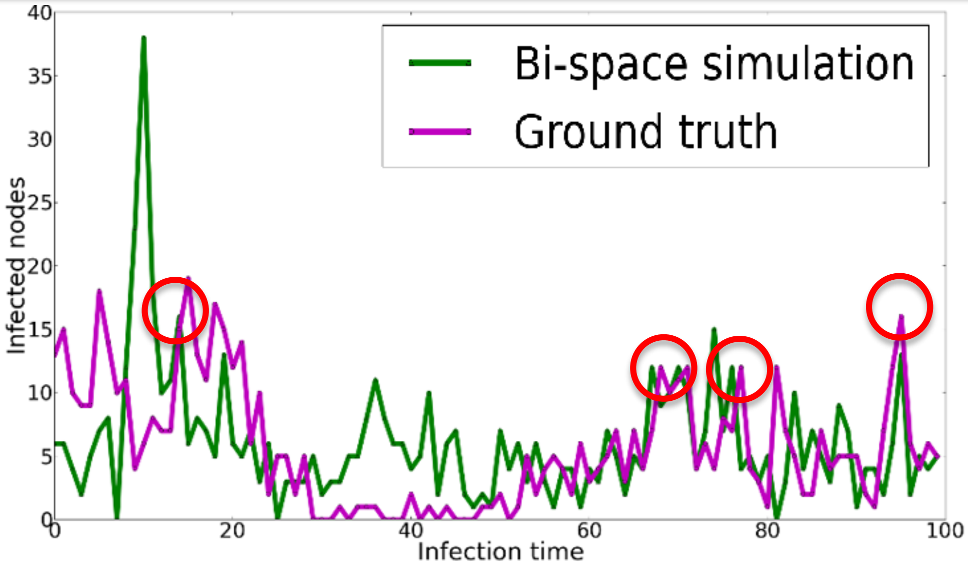



# Simulation Result

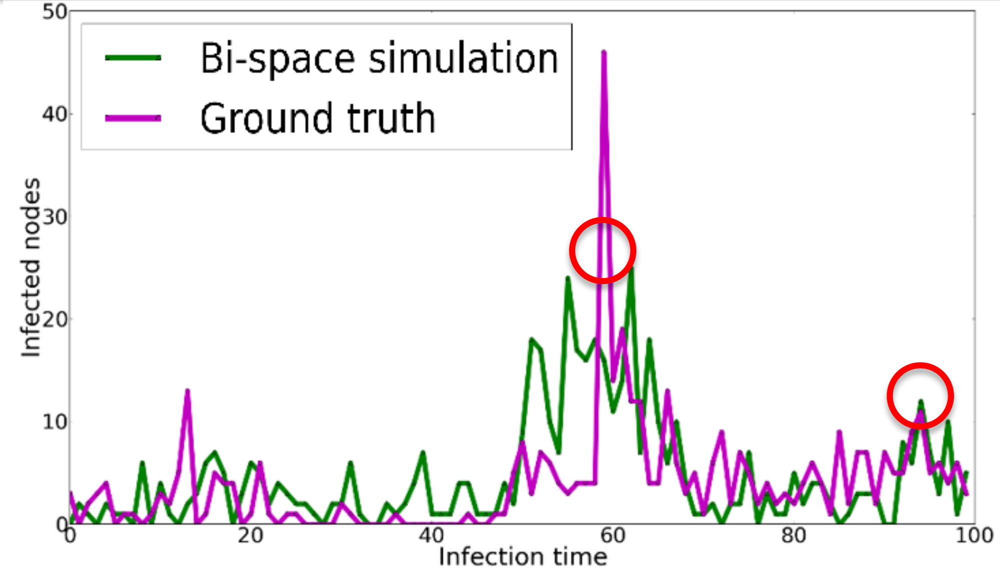



Bispace simulation value = Poisson value + GBM value

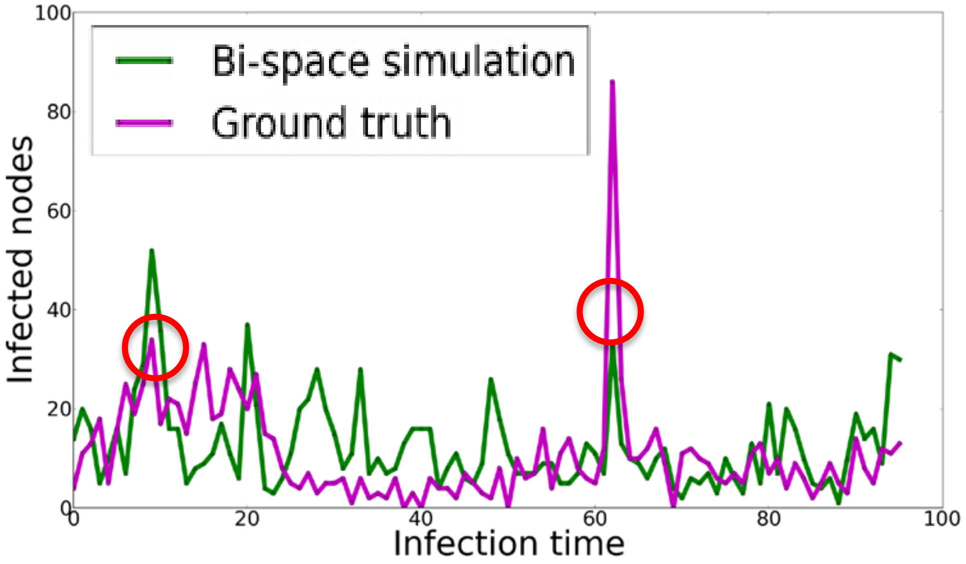
# Simulation results for more cases



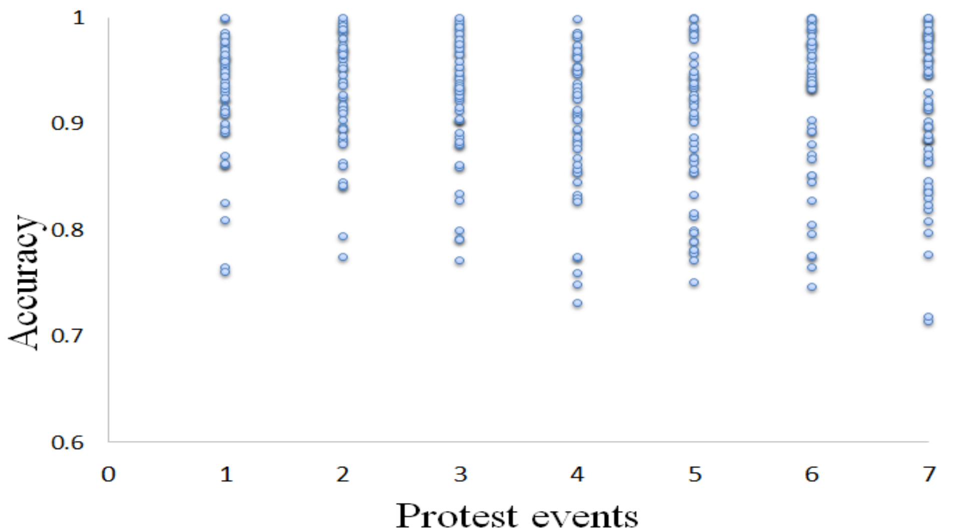

**Mexican Yosoy132 protest**




Colombian anti-government protest

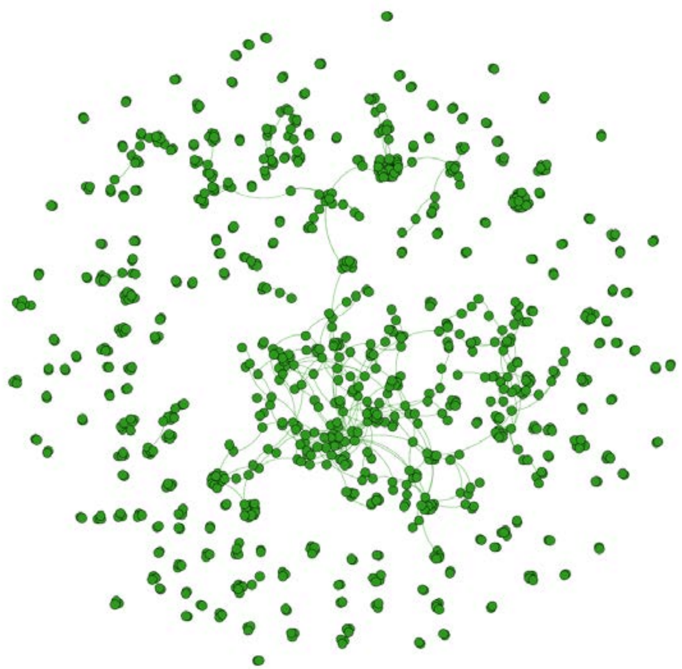



Mexican teachers' protest

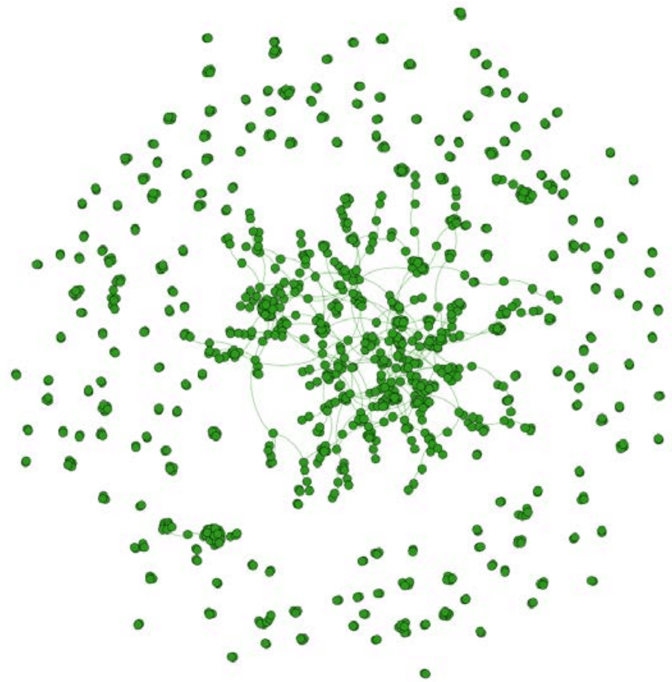



Accuracy for 7 protest events

# Subgraph structure evaluation



Simulated infection subgraph



True infection subgraph

| | Average degree | Diameter | Graph density | Connected components | Average clustering coefficient | Average path length |
|---|---|---|---|---|---|---|
| **Simulation** | 1.791 | 11 | 0.002 | 183 | 0.083 | 4.786 |
| **Ground truth** | 1.726 | 18 | 0.002 | 204 | 0.008 | 6.261 |

GBM simulation results for teacher protest events on Sep 2, 2013

# Conclusion

- The bispace model is effective at capturing the total number of infected nodes. We run 500 times experiments, the average accuracy is 0.90.
- The GBM model can simulate the protest burst, even though not at the precise moment.
- In the bispace propagation model, the GBM process is more influential than the Poisson process.
- The GBM model performs well for these network structure metrics: average degree, diameter, graph density, connected components and average path length
- Capturing the average clustering coefficient needs improvement.

# Thanks!

Fang Jin: jfang8@cs.vt.edu

# Data description

**GSR**
- Gold standard report (GSR) contains: location, date, type of protest, url

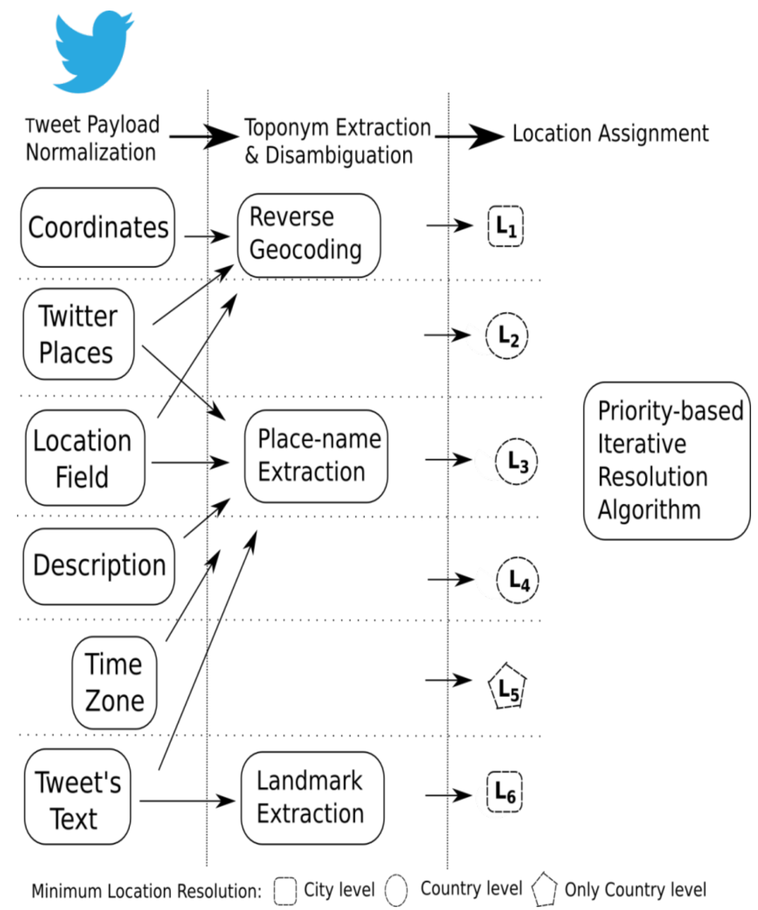
**hashtags**
- For prominent protests, GSR has hashtags

**Filtering**
- Filter tweets by hashtags like “#LaMarchaYoSoy132, #ReformaEducativa”

**Geocoding**
- Geocoding the filtered tweets by content and properties of its users

**Infected nodes**
- We extract and geolocate 20227830 users, assuming they are infected users

**Initial mentions network**
- Given one day and one country, get the mentions network




Geocoding process

# GBM propagation

$$ln(S_t^{ij}) \sim \mathcal{N}((\mu - \frac{\sigma^2}{2})t, \sigma^2 t)$$

At time t, $ln(S_t^{ij}) \geq d_{ij}$

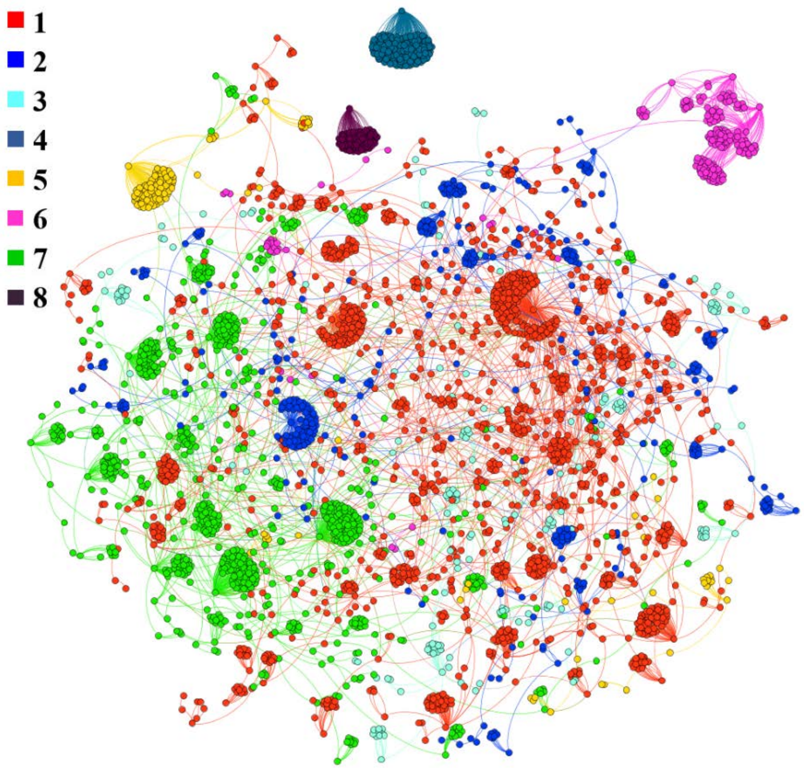
At time $t+\delta t$, $ln(S_{t+\delta t}^{ij}) \sim \mathcal{N}((\mu - \frac{\sigma^2}{2})(t+\delta t), \sigma^2(t+\delta t))$

**GBM parameter estimation**

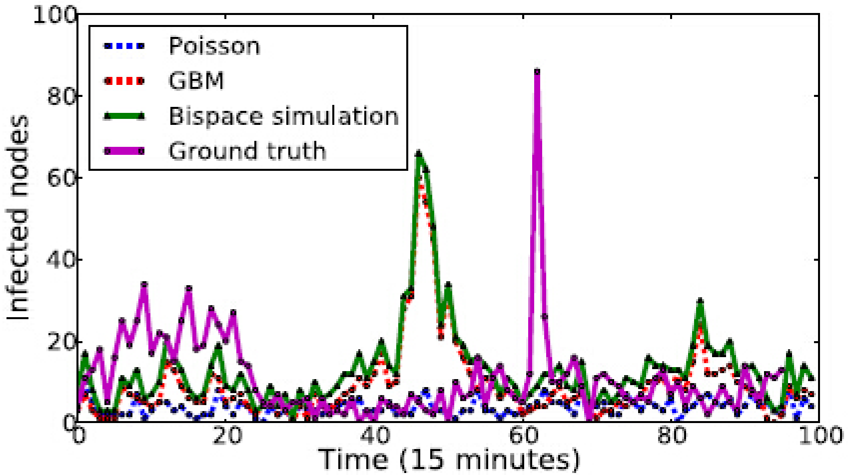
$$\mathcal{L}(\theta, \sigma^2 \mid v_1, \ldots, v_n) = \prod_{j=1}^{n} \frac{1}{\sigma\sqrt{2\pi\tau_j}} exp(-\frac{(x_j - (\mu - \frac{\sigma^2}{2})\tau_j)^2}{2\sigma^2\tau_j})$$

# GBM with community

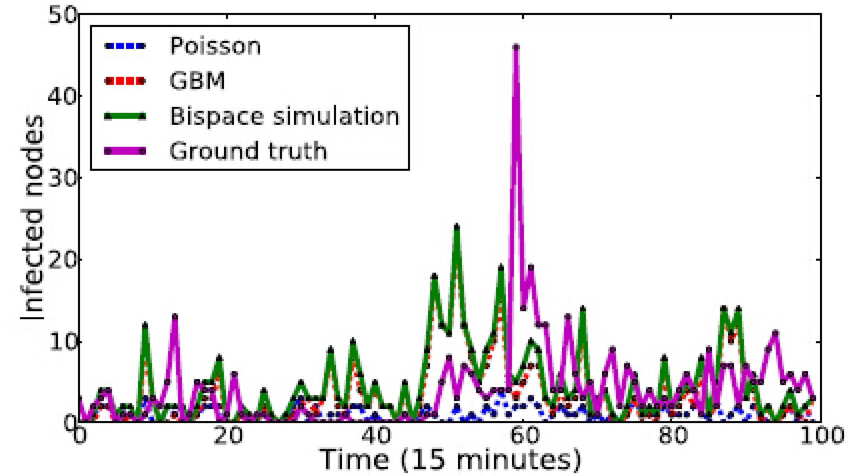
- <u>Assumptions</u>:
  - Each community has its own parameter
  - If $v_i$ and $v_j$ are in the same community $c_i$, the propagation will follow $c_i$ parameters.
  - Propagation from one community to another happens as per the source community's model parameters.



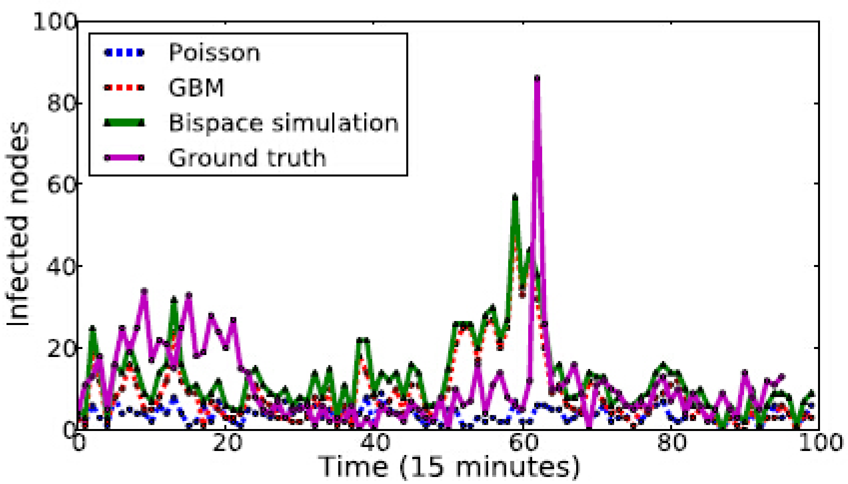
# Results



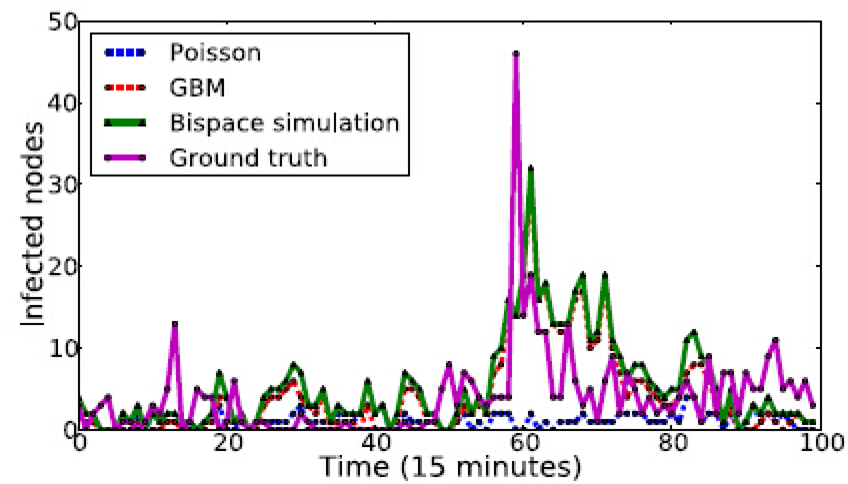

(a) Simulation without community




(a) Simulation without community




(b) Simulation with community




(b) Simulation with community

Mexico teachers' protest

Colombia anti-government protest

# Model set up

- Select one protest event from GSR, find the protest hashtags, keywords, location, beginning day, ending day
- Geocoding tweets
- Initial mentions network: specify the date to be the beginning day, and country to be the protest country, collect all the mentions network of that day and that country, as the initial mentions network.
- Infected users: filter tweets by hashtags/keywords within the protest country.
- Latent space users: Twitter users out of mentions network.