Information Diffusion and External Influence in Networks

Seth A. Myers
Stanford University
samyers@stanford.edu

Chenguang Zhu
Stanford University
cgzhu@cs.stanford.edu

Jure Leskovec
Stanford University
jure@cs.stanford.edu

Introduction

- We model the effect of an unobservable external influence on the diffusion of information across a social network.
- Traditional models only consider pieces of information, or *contagions*, passing from user to user within the network.
- Here, we quantify other sources of information such as mass media that also expose users to contagions.
- We fit our model to URLs diffusing across the Twitter social network.

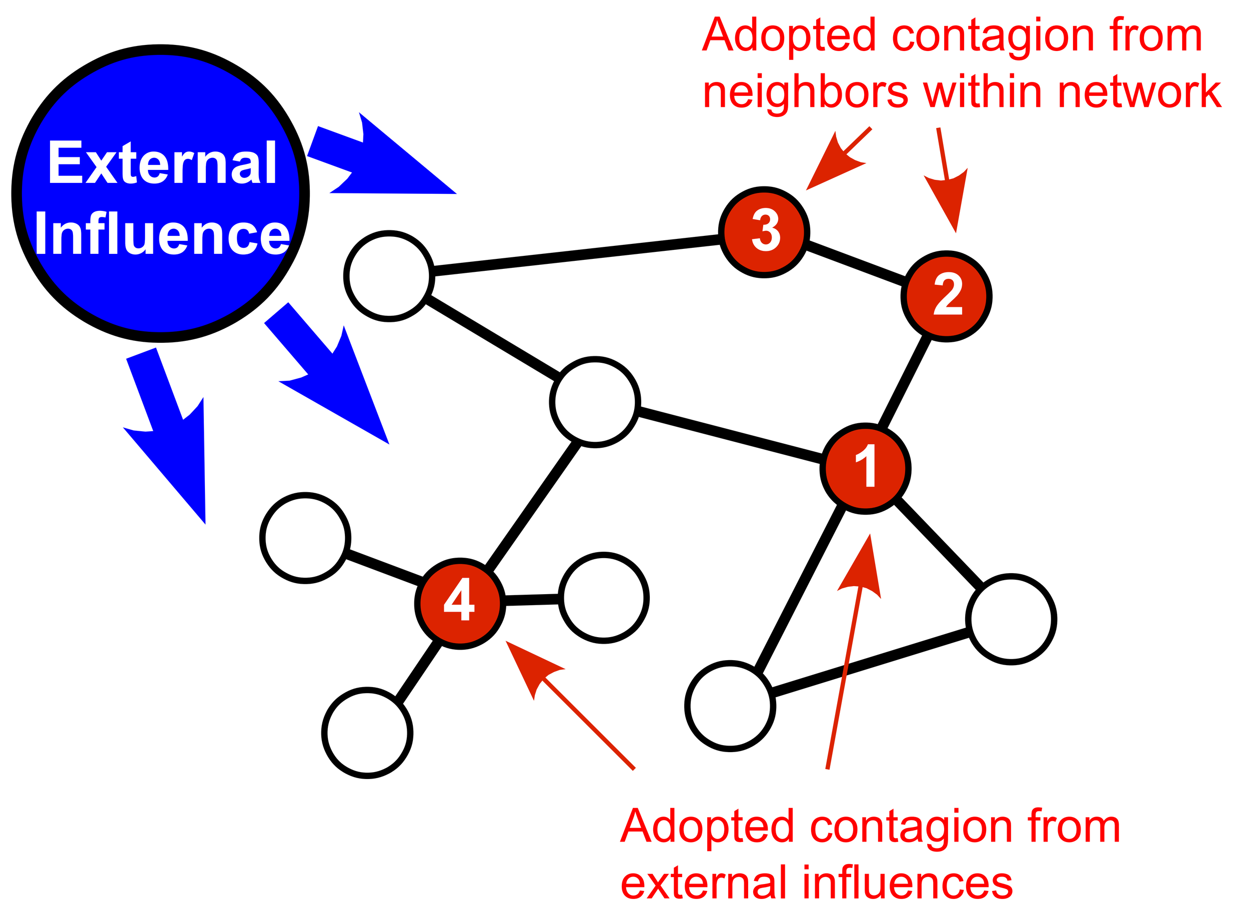

Figure: An example of a single information contagion diffusing through the social network. External influence explains users adopting the contagion before their neighbors

Modeling Network Diffusion

- Each exposure changes a user's perception of the contagion.
- More exposures validate rumors → more likely to adopt
- More exposures make news stale → less likely to adopt
- The exposure curve [Romero '11] gives probability of adopting the contagion given the number of exposures:

$\eta(x)$ = Pr[Adoption of contagion after *x* exposures]

$$=\frac{\rho_1}{\rho_2}\cdot x\cdot\exp\left(1-\frac{x}{\rho_2}\right)$$

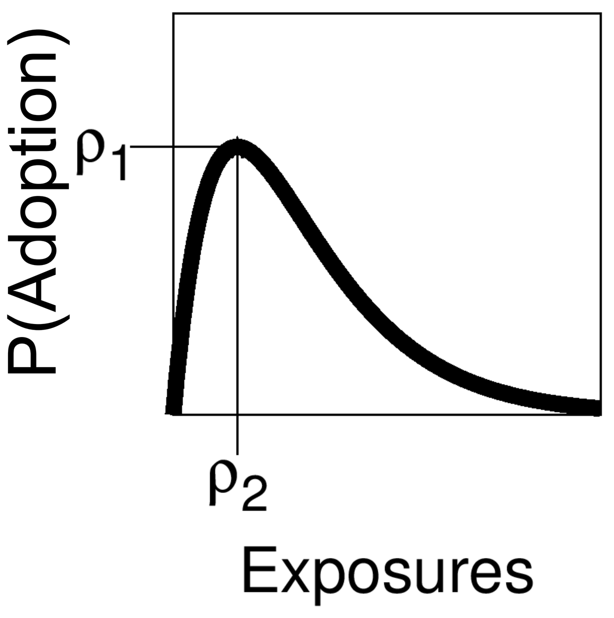

Figure: We fit the exposure curve parameters ρ_1 and ρ_2 for each contagion independently.

Modeling External Influence

- We assume a single external source influencing all users
- The external source's influence is quantified over time, and this is called the event profile:

$\lambda_{ext}(t)$ = Pr[Any user receiving external exposure at time *t*]

Then $\lambda_{ext}(t)$ is fit to observed data for *each* time *t*.

The Complete Model

- For each network neighbor adopting the contagion, the user receives an exposure.
- The external source randomly generates exposures for each user according to the event profile.
- With each exposure, the exposure curve is sampled to decide whether or not the user adopts the contagion.

Fitting the Model

- Given the network and adoption times of each user, the model infers the event profile and exposure curve.
- We iterate between fitting the exposure curve parameters and event profile
- The event profile is fit by matching the expected number of non-adopting nodes to actual number at each time:

$$S(t_k)\approx\sum_{i=1}^{N}\exp\left(-\int_0^{\Lambda_{ext}(t_k)+\Lambda_{net}^{(i)}(t_k)}\eta(y)\,dy\right)$$

where $\Lambda_{ext}(t_k)=\int_0^{t_k}\lambda_{ext}(s)\,ds$

and $\Lambda_{net}^{(i)}(t_k)$ is the number of neighbors of user *i* that have adopted the contagion.

- The exposure curve parameters are fit using maximum likelihood of each user's adoption time.

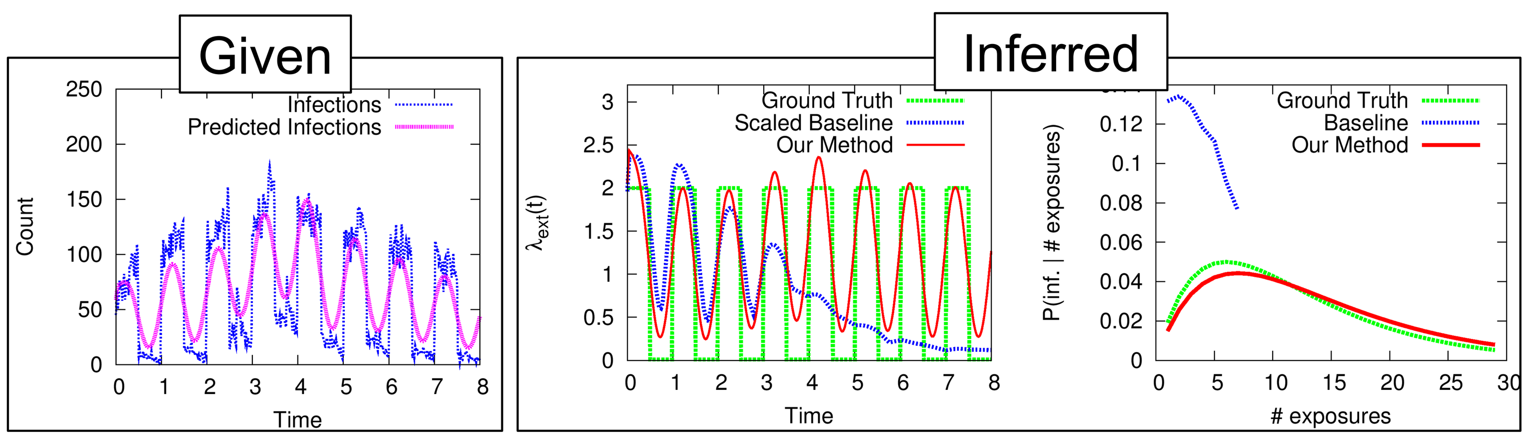

Figure: Inferred model for synthetically generated data.

Real World Data - Twitter

- Across *all* tweets on Twitter in Jan. 2011 we observed:
  - 18,000 high-volume URLs tweeted a total of 2.67M times by 1.08M users
  - 103M edges connecting users
- Each URL is a different contagion; we fit the model to each one independently.

Results

- The inferred event profiles capture real world events:

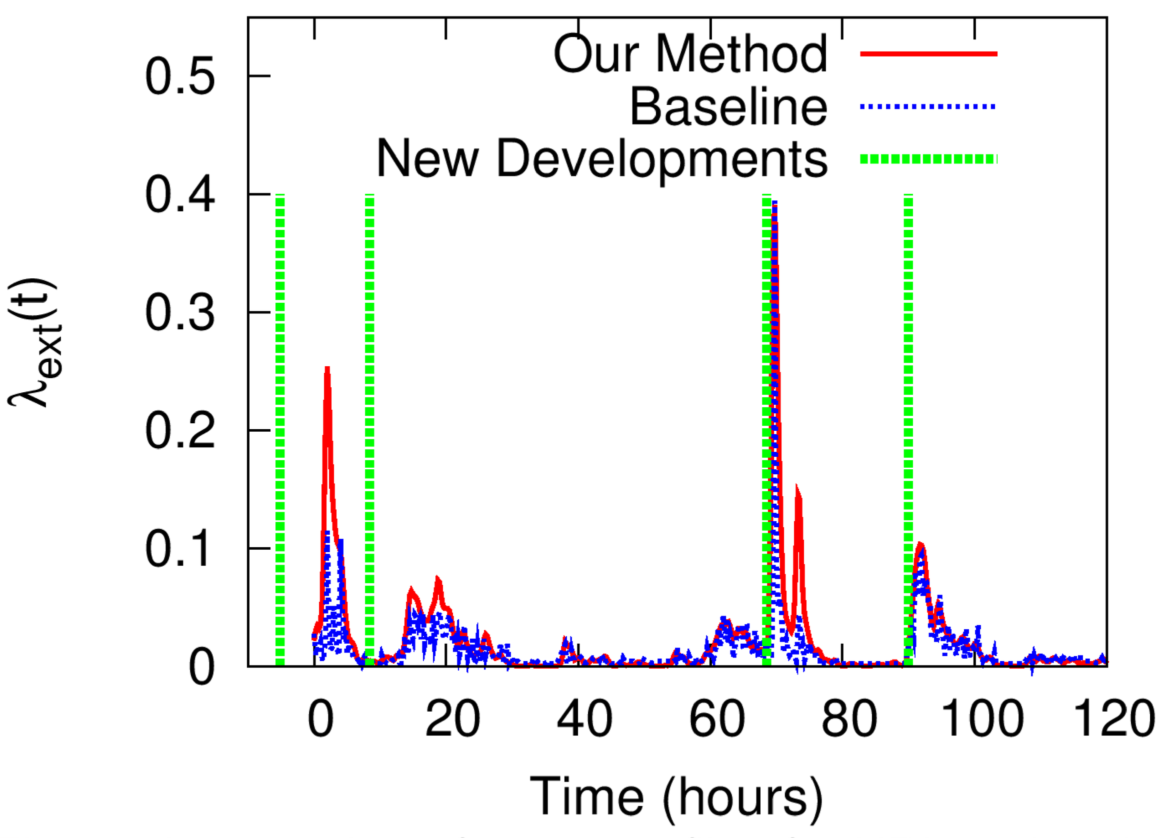

Figure: Aggregation of event profiles for URLs related to the shooting of Rep. Gabby Giffords. Green lines indicate major developments in the story, and they correspond to spikes in the event profile.

- Inferred event profiles match Google keyword search volumes 30% better than the baseline.
- Comparing news categories:
  - World news URLs' exposure curves drop off sooner (they become stale quicker).
  - Art news URLs' exposure curves drop off slower.
  - Political news is the most externally driven with 47% of all exposures from external source.
  - Entertainment news URLs are the least externally driven at only 18%.
- Across all URLs, 29% of all exposures came from the external source.