

### (1) Data Collection



Twitter API

Identify seeds and neighborhoods

Timelines

### (2) Community Detection

Network construction

k-core

Community detection

### (3) Topic Detection

Data cleaning

Generate docs

LDA

### (4) Cluster Similar Messages

Clean tweets and word tokenizing

n-gram analysis

Manual labeling

### (5) Spam Detection

Compute probabilistic table and paths of interest

Classification