Uncovering News-Twitter Reciprocity via Interaction Patterns

Yue Ning[1] Sathappan Muthiah[1]
Ravi Tandon[2] Naren Ramakrishnan[1]

[1]Discovery Analytics Center, Department of Computer Science, Virginia Tech

[2]Now with Department of Electrical and Computer Engineering, The University of Arizona

Outline

Problem Introduction

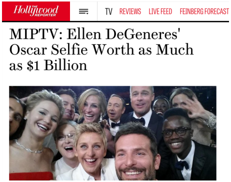
Social Media

News Media

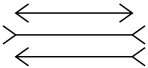
Problem Introduction

Social Media

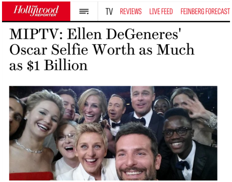
News Media



Problem Introduction



Social Media **News Media**



Motivation

- News -> Twitter
- Twitter -> News Media
- Explosion of information to comment/feed upon
- Cause for variations in such interdependencies
  - Temporal popularity of a "topic"
  - Geo-location (Africa vs Asia)

Motivation

- News -> Twitter
- Twitter -> News Media
- Explosion of information to comment/feed upon
- Cause for variations in such interdependencies
    - Temporal popularity of a "topic"
    - Geo-location (Africa vs Asia)

Motivation

- News -> Twitter
- Twitter -> News Media
- Explosion of information to comment/feed upon
- Cause for variations in such interdependencies
  - Temporal popularity of a "topic"
  - Geo-location (Africa vs Asia)

Motivation

- News -> Twitter
- Twitter -> News Media
- Explosion of information to comment/feed upon
- Cause for variations in such interdependencies
  - Temporal popularity of a "topic"
  - Geo-location (Africa vs Asia)

Motivation

- News -> Twitter
- Twitter -> News Media
- Explosion of information to comment/feed upon
- Cause for variations in such interdependencies
    - Temporal popularity of a "topic"
    - Geo-location (Africa vs Asia)

Motivation

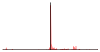
- News -> Twitter
- Twitter -> News Media
- Explosion of information to comment/feed upon
- Cause for variations in such interdependencies
  - Temporal popularity of a "topic"
  - Geo-location (Africa vs Asia)

One Example

D_1

1/27/2013: Fire at Kiss night-club: security guards tried to stop nightclub output

One Example

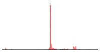


N

D_1

1/27/2013: Fire at Kiss night-club: security guards tried to stop nightclub output

One Example



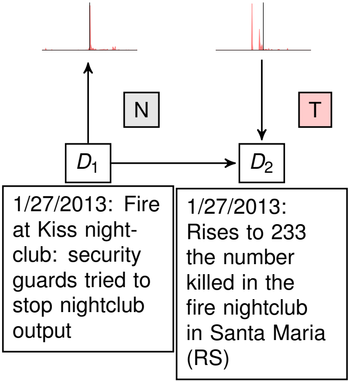
N

D_1 → D_2

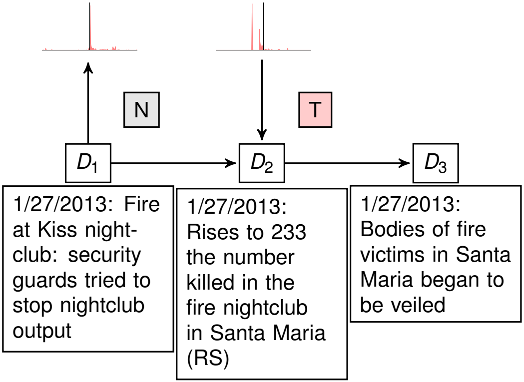
1/27/2013: Fire at Kiss nightclub: security guards tried to stop nightclub output

1/27/2013: Rises to 233 the number killed in the fire nightclub in Santa Maria (RS)

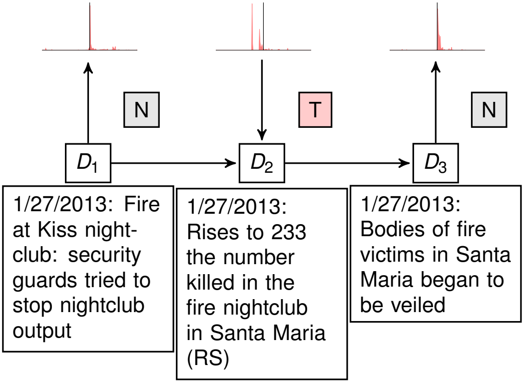
One Example


N
T
D_1
D_2
1/27/2013: Fire at Kiss night-club: security guards tried to stop nightclub output
1/27/2013: Rises to 233 the number killed in the fire nightclub in Santa Maria (RS)

One Example

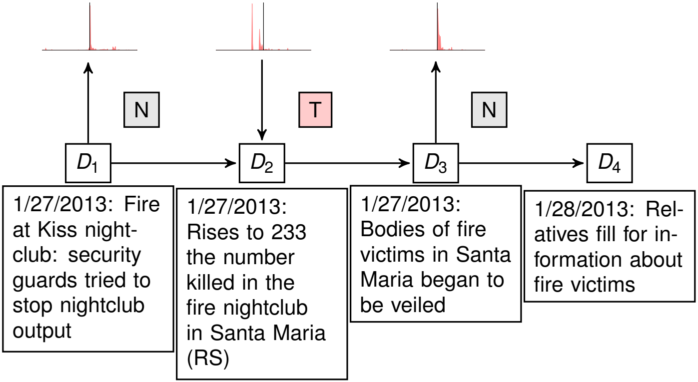

N
T
D_1
D_2
D_3
1/27/2013: Fire at Kiss night-club: security guards tried to stop nightclub output
1/27/2013: Rises to 233 the number killed in the fire nightclub in Santa Maria (RS)
1/27/2013: Bodies of fire victims in Santa Maria began to be veiled

One Example

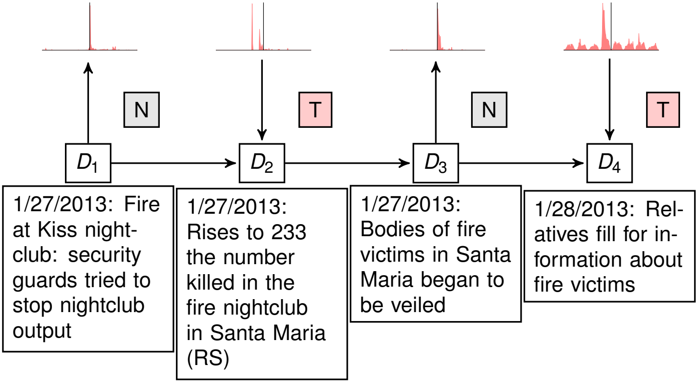


N
T
N
D_1
D_2
D_3
1/27/2013: Fire at Kiss night-club: security guards tried to stop nightclub output
1/27/2013: Rises to 233 the number killed in the fire nightclub in Santa Maria (RS)
1/27/2013: Bodies of fire victims in Santa Maria began to be veiled

One Example

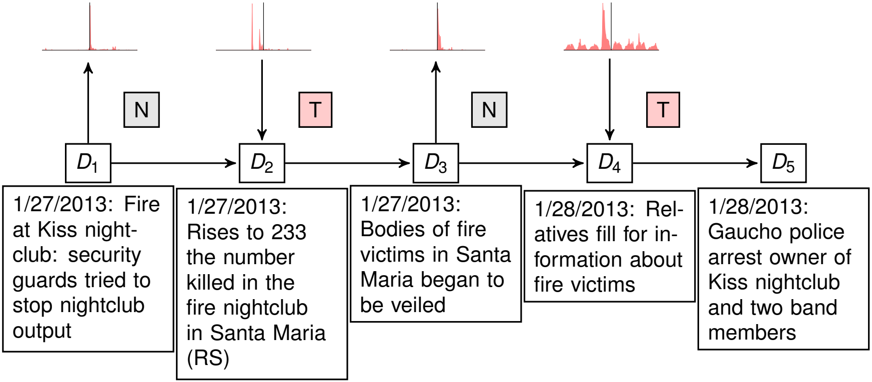

N
T
N
D_1
D_2
D_3
D_4
1/27/2013: Fire at Kiss night-club: security guards tried to stop nightclub output
1/27/2013: Rises to 233 the number killed in the fire nightclub in Santa Maria (RS)
1/27/2013: Bodies of fire victims in Santa Maria began to be veiled
1/28/2013: Rel-atives fill for in-formation about fire victims

One Example

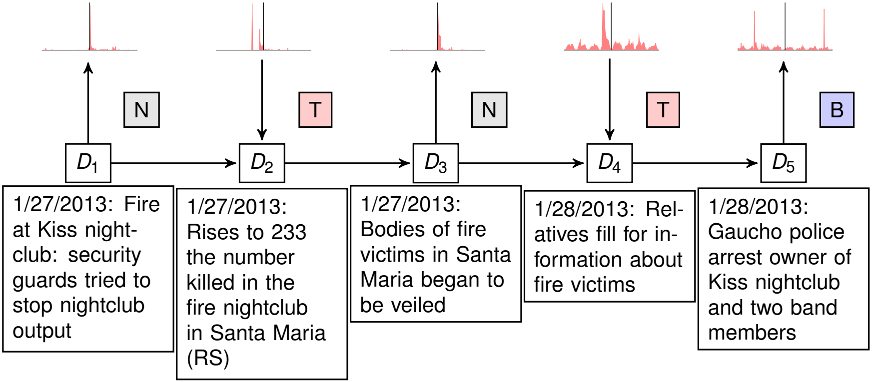


N
T
N
T
D_1
D_2
D_3
D_4
1/27/2013: Fire at Kiss night-club: security guards tried to stop nightclub output
1/27/2013: Rises to 233 the number killed in the fire nightclub in Santa Maria (RS)
1/27/2013: Bodies of fire victims in Santa Maria began to be veiled
1/28/2013: Rel-atives fill for in-formation about fire victims

One Example


N
T
N
T
D_1
D_2
D_3
D_4
D_5
1/27/2013: Fire at Kiss nightclub: security guards tried to stop nightclub output
1/27/2013: Rises to 233 the number killed in the fire nightclub in Santa Maria (RS)
1/27/2013: Bodies of fire victims in Santa Maria began to be veiled
1/28/2013: Relatives fill for information about fire victims
1/28/2013: Gaucho police arrest owner of Kiss nightclub and two band members

One Example


N
T
N
T
B
D_1
D_2
D_3
D_4
D_5
1/27/2013: Fire at Kiss nightclub: security guards tried to stop nightclub output
1/27/2013: Rises to 233 the number killed in the fire nightclub in Santa Maria (RS)
1/27/2013: Bodies of fire victims in Santa Maria began to be veiled
1/28/2013: Relatives fill for information about fire victims
1/28/2013: Gaucho police arrest owner of Kiss nightclub and two band members

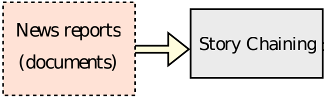
Goals

1. Understanding **the type of information flow** between news and Twitter.
2. **Chaining** similar news articles together.
3. Identifying **major interaction patterns**
   - **Cluster** story chains and understanding their differences
   - **Identify main topics** of interest within such clusters.

System Framework

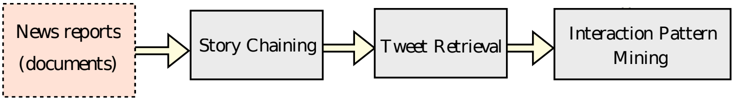
News reports
(documents)

System Framework

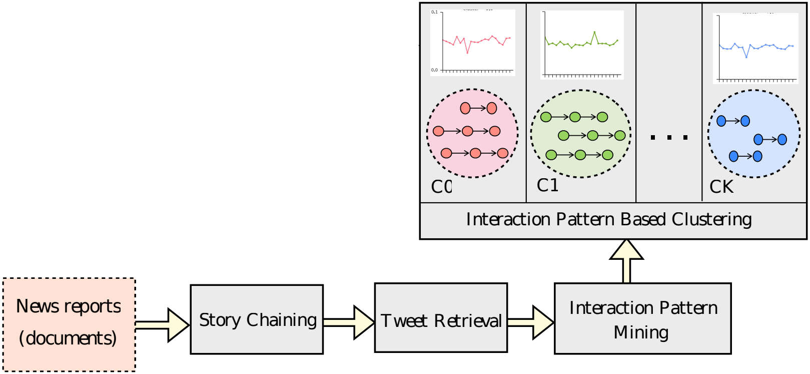


System Framework

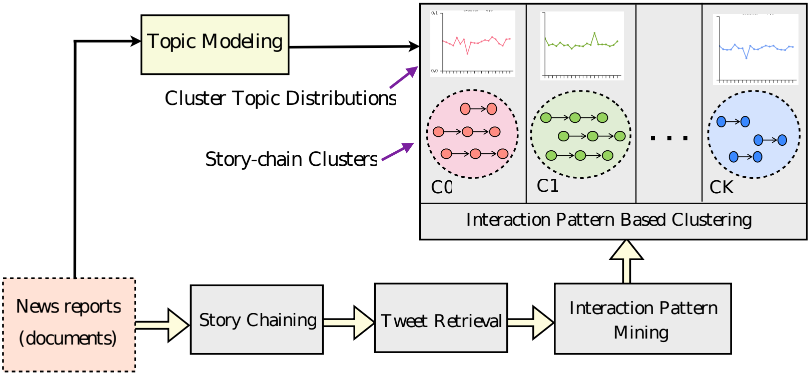
System Framework



System Framework



System Framework



Story Chaining Algorithm

[1] **Goal:** identifying all documents related to a news story and to keep track of the news story as new documents arrive.
Method: To assess if two documents are referring to the same underlying context, we calculate their similarity scores with respect to three features:

- - textual features, denoted by $T(D_i)$
- - spatial features, denoted by $L(D_i)$, e.g. city, state, country
- - actors, denoted by $A(D_i)$, e.g. Hillary Clinton.

---

[1] J. Schlachter, A. Ruvinskya, L. Asencios Reynoso, S. Muthiah, and N. Ramakrishnan, “Leveraging topic models to develop metrics for evaluating the quality of narrative threads extracted from news stories”, in *Proc. of the 6th International Conference on Applied Human Factors and Ergonomics*, AHFE, Elsevier, 2015.

Story Chaining Algorithm (Cont.)

The total weighted similarity measure between two documents, D_i and D_j , is then defined as follows:

$$\text{sim}(D_i, D_j) \triangleq \alpha \underbrace{f(\mathcal{T}(D_i), \mathcal{T}(D_j))}_{\text{textul features}} + \beta \underbrace{f(\mathcal{L}(D_i), \mathcal{L}(D_j))}_{\text{spatial features}} + \eta \underbrace{f(\mathcal{A}(D_i), \mathcal{A}(D_j))}_{\text{actor features}}$$

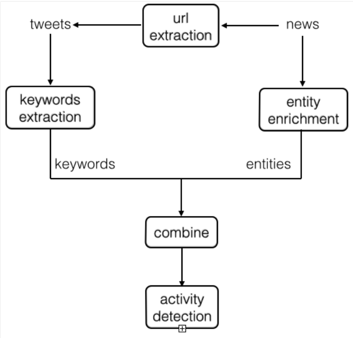
The coherence between a chain C_j and document D_i is defined as

$$\text{coh}(D_i, C_j) = \theta g(\mathcal{L}(D_i), \mathcal{L}(C_j)) + \phi g(\mathcal{A}(D_i), \mathcal{A}(C_j))$$

where g is any similarity measure and the coefficients θ, ϕ are chosen such that $\theta + \phi = 1$.

Twitter Profile for News

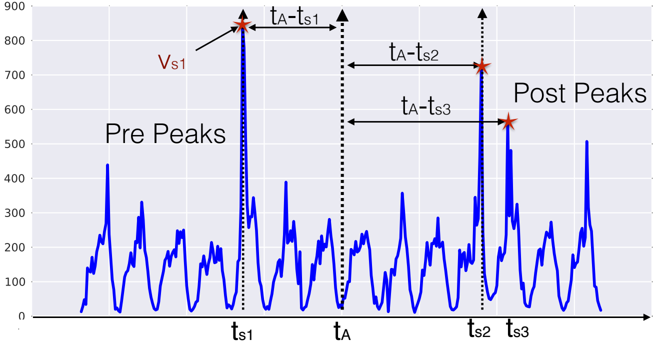
1. Collect tweets based on URL.
2. Extract entity keywords from news.
3. Filter keywords together.
4. Download hourly count metrics.



Interaction Patterns

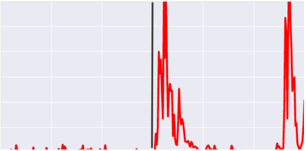
- Peak detection [2]
- Incoming influence ($\mathcal{W}^{\text{pre}}$) and outgoing influence ($\mathcal{W}^{\text{post}}$):

$$\mathcal{W}^{\text{pre}} = \sum_{s \in S_{\text{pre}}} \frac{v_s}{t_A - t_s}, \quad \mathcal{W}^{\text{post}} = \sum_{s \in S_{\text{post}}} \frac{v_s}{t_s - t_A} \tag{1}$$

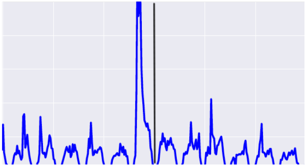


[2] M. Duarte, "Notes on scientific computing for biomechanics and motor control", 2015.

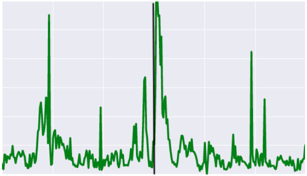
Interaction States



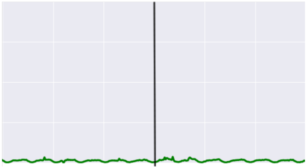
N



T



B



E

Interaction States (Cont.)

$$\text{State}(D_i) = \begin{cases} N, & \text{if } \mathcal{W}^{\text{pre}} < \rho,\ \mathcal{W}^{\text{post}} \geq (1+\lambda)\mathcal{W}^{\text{pre}} \\ E, & \text{if } \mathcal{W}^{\text{pre}} < \rho,\ \mathcal{W}^{\text{post}} < (1+\lambda)\mathcal{W}^{\text{pre}} \\ T, & \text{if } \mathcal{W}^{\text{pre}} \geq \rho,\ \mathcal{W}^{\text{post}} < (1+\lambda)\mathcal{W}^{\text{pre}} \\ B, & \text{if } \mathcal{W}^{\text{pre}} \geq \rho,\ \mathcal{W}^{\text{post}} \geq (1+\lambda)\mathcal{W}^{\text{pre}} \end{cases}$$

Interaction States (Cont.)

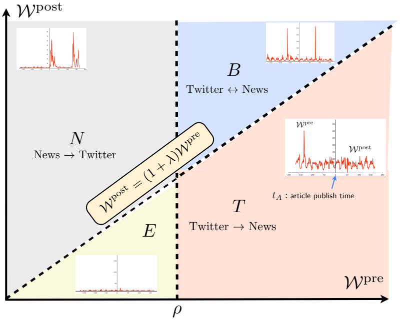
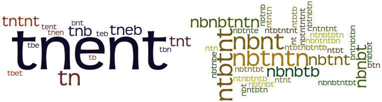


Figure: Geometric Interpretation of States

Clustering on Encoded Chains

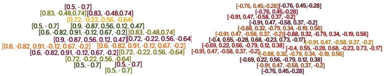
Clustering via qualitative encoding (e.g. “NNTNTBBE”)

- Levenshtein distance
- Jaro-Winkler distance
- Ratcliff-Obershelp pattern recognition



Clustering via quantitative encoding (e.g. “0.5, -0.9, 0.88,0.3,-0.4”)

- Multi-dimensional Dynamic Time Warping(DTW)



Interpretation from a Different Dimension

- Are sports events always related with Bi-directional Interactions?
- Do Twitter users focus more on sports and entertainment?

- Latent Dirichlet Allocation (LDA) for hidden topic analysis on clusters. The topic distributions for one cluster is defined by:

$$\mathbf{C}_{j,k} = \frac{\sum_{d_{ij} \in c_j} n_{d_{ij}} \theta(d_{ij}, k)}{\sum_{d_{ij}} n_{d_{ij}}}, \tag{2}$$

where

- $n_{d_{ij}}$ refers to the frequency of d_i in cluster C_j.
- $\theta(d_{ij}, k)$ refers to the topic proportions for this document.
- k is the topic index.

Interpretation from a Different Dimension

- Are sports events always related with Bi-directional Interactions?
- Do Twitter users focus more on sports and entertainment?

- Latent Dirichlet Allocation (LDA) for hidden topic analysis on clusters. The topic distributions for one cluster is defined by:

$$\mathbf{C}_{j,k} = \frac{\sum_{d_{ij} \in c_j} n_{d_{ij}} \theta(d_{ij}, k)}{\sum_{d_{ij}} n_{d_{ij}}}, \tag{2}$$

where

- $n_{d_{ij}}$ refers to the frequency of d_i in cluster C_j.
- $\theta(d_{ij}, k)$ refers to the topic proportions for this document.
- k is the topic index.

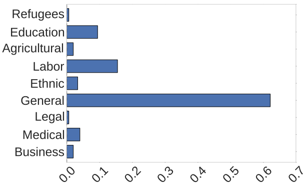
Dataset

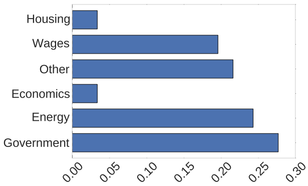
Real data from Brazil during the period from Nov. 2012 to Sep. 2013:

- Protest related articles: GSR
- Other articles: NON-GSR

Table: Statistical properties of GSR and Non-GSR chains.

Category	% of Twitter starts	Avg-Time-Lag(hour)
GSR Chains	40%	10.95
Non-GSR Chains	73%	5.26





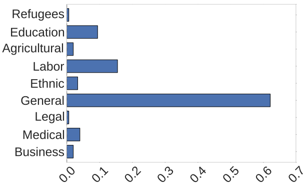
Dataset (Cont.)

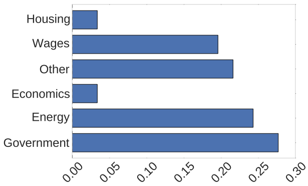
Real data from Brazil during the period from Nov. 2012 to Sep. 2013:

- Protest related articles: GSR
- Other articles: NON-GSR

Table: Statistical properties of GSR and Non-GSR chains.

Category	% of Twitter starts	Avg-Time-Lag(hour)
GSR Chains	40%	10.95
Non-GSR Chains	73%	5.26





GSR Dataset

Category	% News starts	% of Twitter starts
Housing related protests	100%	0%
Agriculture	100%	0%
Medical	74%	26%
Other (religious & cultural)	60%	40%
General Population	30%	70%
Govt. Policies	23%	77%

Table: % of Twitter, News starts for GSR story-chains

Cluster Results

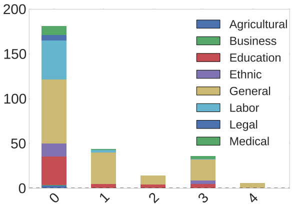


Figure: Population Distribution of Clusters (K-Medoids)

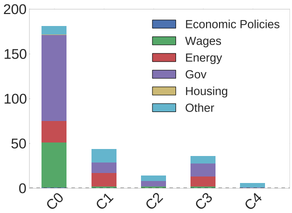


Figure: Event Type Distribution of Clusters (K-Medoids)

Topics in Clusters

ID	Frequent Sub-patterns	Top Topics
C0	“NBNBTNTN”, “NTNTN”	Local Events
C1	“NT”, “NTNT”	Local Events
C2	“TNT”	Local Events, Ads, Technology
C3	“T”, “TB”	Others, Protest, Sports
C4	“TNENT”, “TEB”	Protest, Government, Entertainment

Table: Top topics for clusters

Topic Distributions for Clusters

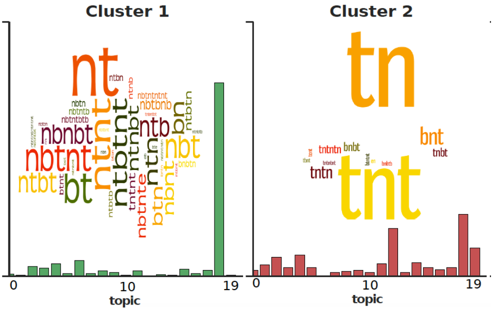


Figure: Topic distributions of 2 interaction pattern clusters. The X-axis labels refer to topic numbers

Main Influencer

We define the *influence weight* of a story chain as the average of the difference of pre- and post- influence weights:

$$\frac{\sum_i (\mathcal{W}_i^{\text{pre}} - \mathcal{W}_i^{\text{post}})}{n}$$

where the summation is over n, the number of articles in a chain.

Main Influencer (Cont.)

Table: Story Chains with Interaction Patterns and Main Influencer

ID	IP	IW	MI	Story Summary
SC1	TT	0.514	Twitter	"Marco Feliciano protest at church door"
SC2	TN	0.48	Twitter	"25%Teachers are on strike."
SC3	NNNNBNTBN	-0.422	News	"Fire in Kiss Nightclub in Santa Maria "
SC4	NBNNTN	-0.405	News	"Governor decree official mourning"
SC5	TTTNN	5.0e-05	Both	"Nadal back to Brazil"
SC6	NNTNTTNTN	-1.7e-04	Both	"Nissan sells more than 100 thousand"

Conclusion

- A new framework for discovering the direction of information flow over time across **news and Twitter**.
- Uncover the **interaction patterns** over stories and test our proposed method on real data.
- Cluster on encoded story chains and discover topics

Observation 1

Twitter as a social network platform serves as a fast way to draw attention from public for many social events such as sports news.

Observation 2

News media is quicker to report events regarding political, economical and business issues.

Thank you!
Q&A