Relation Tracker - tracking the main entities and their relations through time

M. Besher Massri, Inna Novalija, Marko Grobelnik

Artificial Intelligence Laboratory

Jožef Stefan Institute

SiKDD, October 11th,2018

Introduction

- **Motivation**: The need for tools that provide simple and holistic summary about a topic.
- **Goals**:
  - Tracking the changes in main entities [people and organization] within each topic through time.
  - Detecting the characteristics of the relationship between entities.
  - Visualizing information through different graphs to get more insight about the topic.

Artificial Intelligence Laboratory
Jožef Stefan Institute
ailab.ijs.si

Outline

- Data
- Methodology
  - Clustering data
  - Choosing the main entities
  - Detecting the characteristic of a relationship
- Visualizing the results
  - Features and main functionality
  - Displaying relation information
- Demo
- Future work

Artificial Intelligence Laboratory
Jožef Stefan Institute
ailab.ijs.si

Data

- ∼1.8 million events from event registry
- Cover the period Jan 2015 → July 2016
- The following information were collected from each event:
  - Basic info: Title, Uri, Date.
  - Total article within that event (in all languages).
  - List of Wikified concepts extracted from the articles using Wikifier:
    - Entity concepts: person/organization.
    - Non-entity concepts: keywords like computer, toy, robot, etc.

Artificial Intelligence Laboratory
Jožef Stefan Institute
ailab.ijs.si

Methodology

Clustering data

- K-means clustering
  - Each event is represented as a sparse vector of its non-entity concepts.
- =>100 cluster of variety of topics like technology and education
  - Centroid vector carry out the characteristics of that topic, as a set of weighted features.
- Naming clusters:
  - Forming a text based on the centroid vector of each cluster.
  - Using Dmoz classifier to classify that text into list of top categories.

Artificial Intelligence Laboratory
Jožef Stefan Institute
ailab.ijs.si

Methodology

Choosing main entities

At any given time interval *D*, the importance value of each entity is calculated throughout three stages:

1. An initial important value:

$$Imp_{init}(c)_D = \sum_{\substack{e \in E_D \\ e \text{ has concept } c}} S_{c,e} \; * \sum_{\substack{e \in E_D \\ e \text{ has concept } c}} A_e$$

- E_D : The set of events that occur in the time interval *D*.
- A_e : The number of articles that event *e* contains.
- $S_{c,e}$: The TF-IDF score of concept *c* at event *e*.

Artificial Intelligence Laboratory
Jožef Stefan Institute
ailab.ijs.si

Methodology

Choosing main entities (cont.)

2. Negative importance value:
   - introduced to avoid the case of nominating entities only from one big event

$$Imp_{neg}(c)_D = \sum_{c' \in N_{c,D}} \frac{Imp_{init}(c')_D}{|N_{c',D}|}$$

   - $N_{c,D}$: The set of neighbors of concept *c* during the interval of time *D*.

- Final importance value; their difference:

$$Imp_{final}(c)_D = Imp_{init}(c)_D - Imp_{neg}(c)_D$$

Artificial Intelligence Laboratory
Jožef Stefan Institute
ailab.ijs.si

Methodology

Detecting the characteristics of the relationship

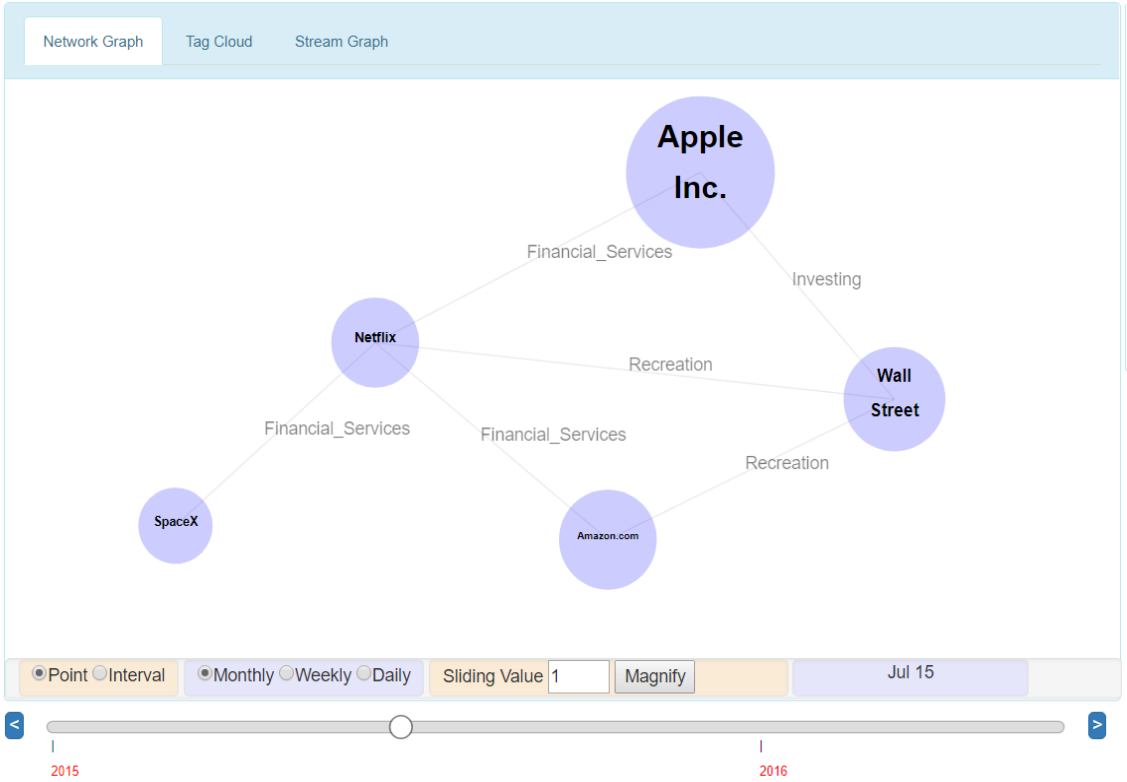
- Modelling the relationship between any two entities through a vector of words where two entities are collocated:
  - Aggregate all the non-entity concepts from the shared events between the two entities.
  - Each concept has a value depends on the number of events it occurs in and its score in those events.
  - Top concepts were chosen as the main characteristics of the relationship.
- Top shared events were also extracted based on the calculated value of each of its non-entity concepts.
- Concepts were summarized by Dmoz classifier to define the main topic of the shared events.

Artificial Intelligence Laboratory
Jožef Stefan Institute
ailab.ijs.si

Visualizing the results

Features and main functionality

- Interactive network graph that features the main entities at the selected interval of time:
  - Nodes: the main entities at each interval of time.
  - Size of nodes represent the importance value of each entity, scaled to a suitable ratio.
  - Color representing the type, whether it's a person[red], or organization [blue].
  - edges: indicating the existence of a shared events between them in the time interval.
  - Thickness of links reflect the number of shared events.
  - Labels are the main topic of the shared events (summary of the characteristics of their relationship).



Top companies in July 2015 and their relations under the business topic.

Artificial Intelligence Laboratory
Jožef Stefan Institute
ailab.ijs.si

Visualizing the results

Features and main functionality (cont.)

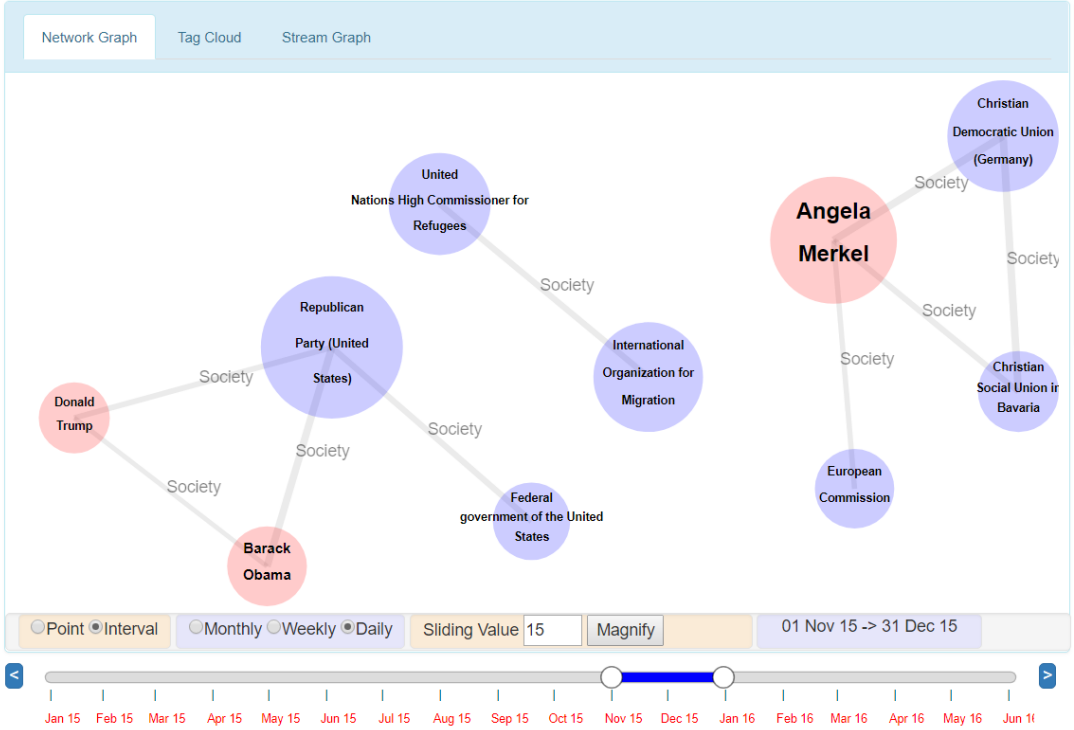
- Slide bar to move through the timeline and update the graph at each step
  - Different scales for moving; monthly/weekly/daily scale.
  - point search/ interval search in time.
  - Magnifying a certain interval to take a closer look at it.

Artificial Intelligence Laboratory
Jožef Stefan Institute
ailab.ijs.si

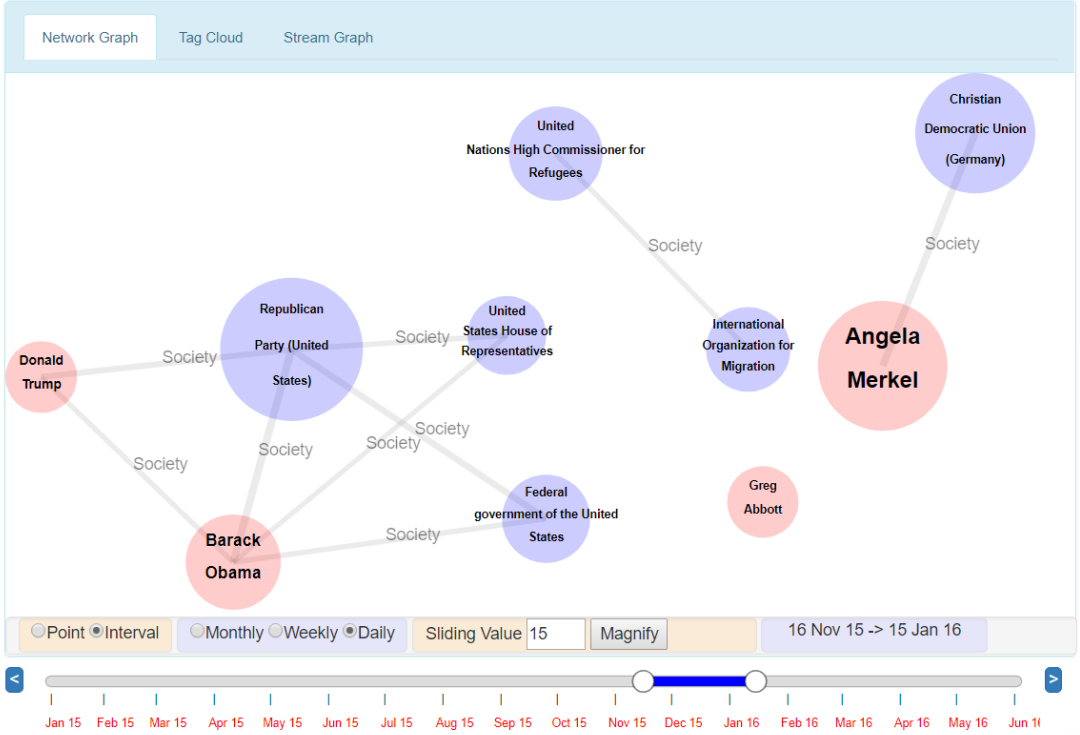
Visualizing the results

Features and main functionality (cont.)

Top entities for the last two months of 2015 under the refugee topic.



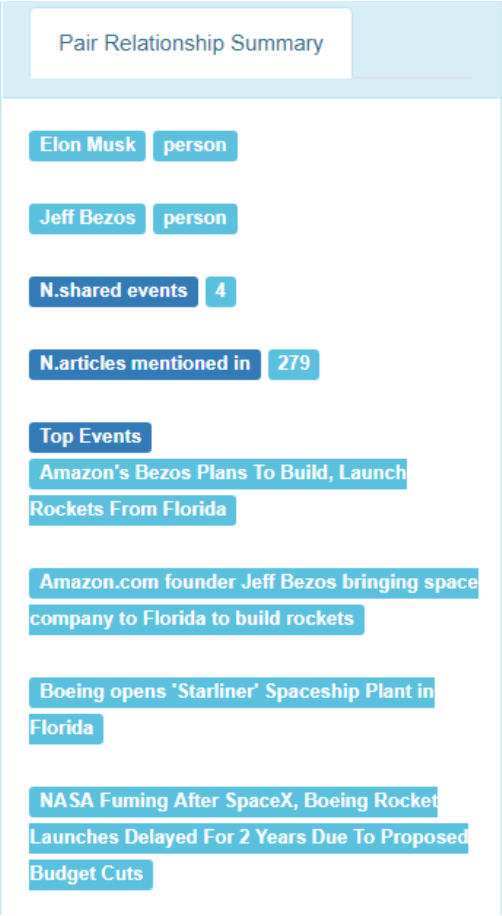
The changes in top entities under the same topic after moving the interval for 15 days.

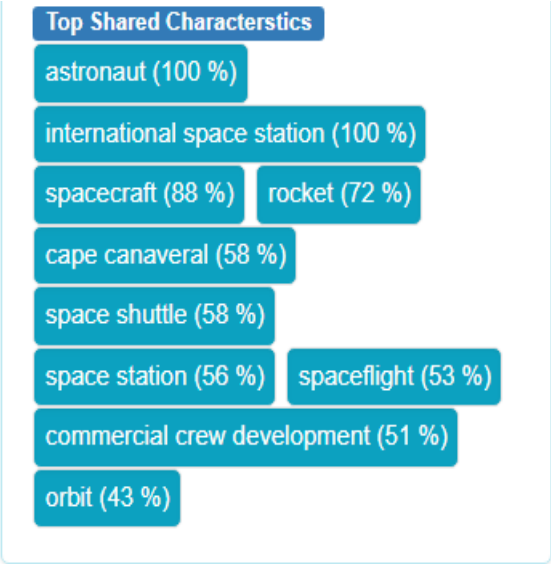


Visualizing the results

Displaying relation information

- **Summary of the relationship whenever the user selects a pair of entities:**
  - Number of shared events and articles.
  - List of top events.
  - Top shared characteristics of the relationship.





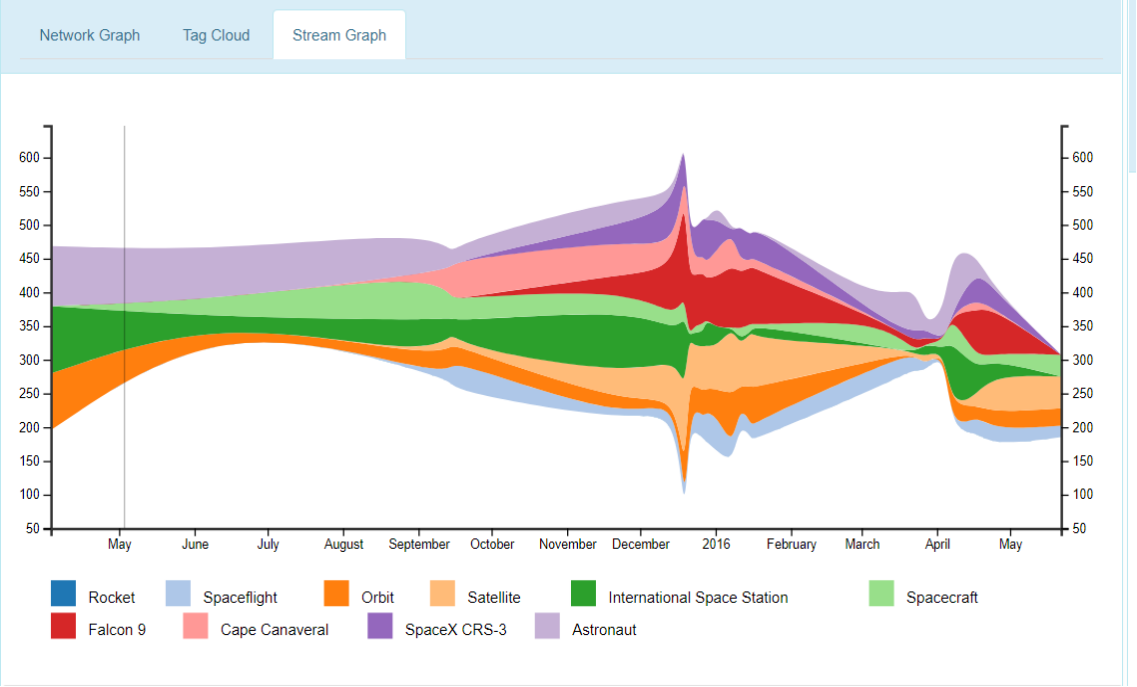
Relationship summary about Jeff Bezos and Elon Musk between January and September 2015 under the Space topic.

Artificial Intelligence Laboratory
Jožef Stefan Institute
ailab.ijs.si

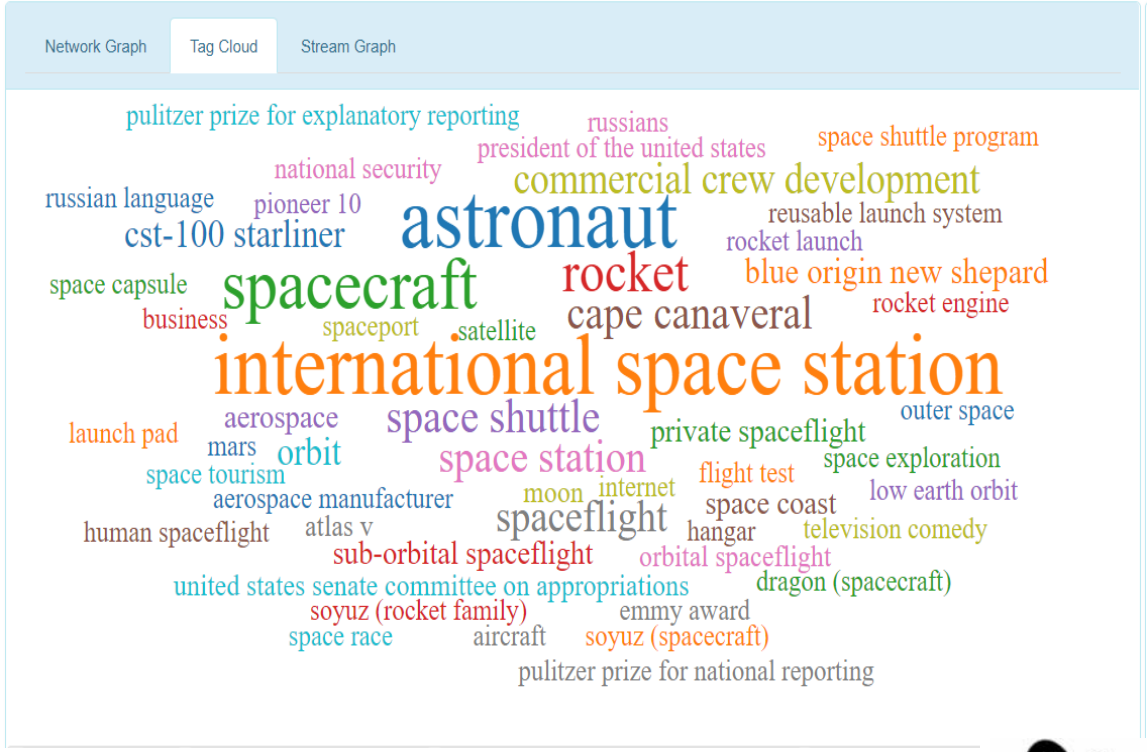
Visualizing the results

Displaying relation information(cont.)

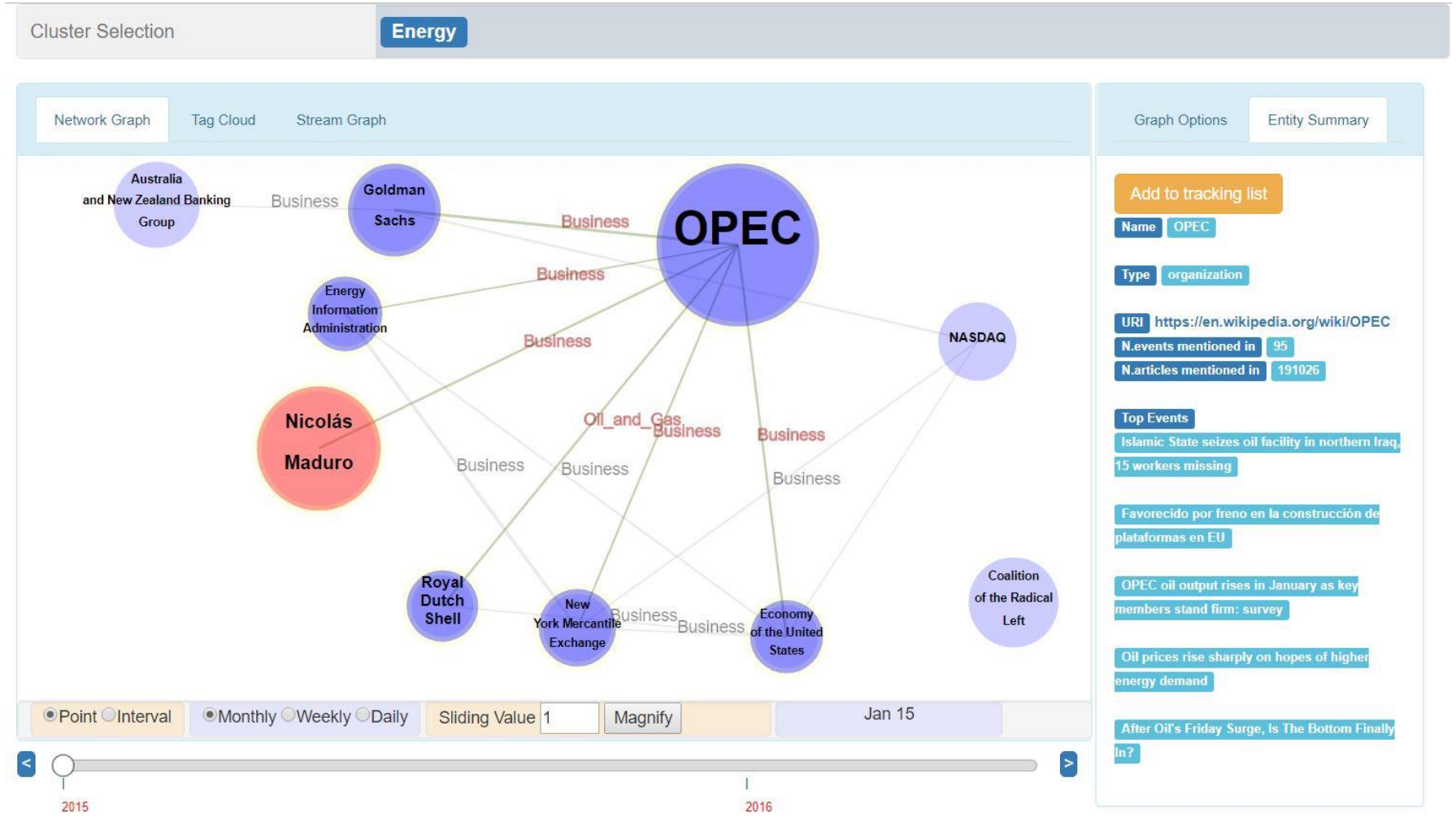
- **Stream graph for detecting how the importance of those top features with respect to the relationship is changing through time.**



- **Tag cloud for visualizing the set of all characteristics that affect the relationship to give a big picture about it.**



Demo


Cluster Selection
Energy
Network Graph
Tag Cloud
Stream Graph
Graph Options
Entity Summary
Australia and New Zealand Banking Group
Business
Goldman Sachs
Business
OPEC
Business
Energy Information Administration
Business
NASDAQ
Nicolás Maduro
Oil_and_Gas
Business
Business
Business
Business
Business
Royal Dutch Shell
New York Mercantile Exchange
Business
Business
Economy of the United States
Coalition of the Radical Left
Point
Interval
Monthly
Weekly
Daily
Sliding Value
1
Magnify
Jan 15
2015
2016
Add to tracking list
Name
OPEC
Type
organization
URI
https://en.wikipedia.org/wiki/OPEC
N.events mentioned in
95
N.articles mentioned in
191026
Top Events
Islamic State seizes oil facility in northern Iraq, 15 workers missing
Favorecido por freno en la construcción de plataformas en EU
OPEC oil output rises in January as key members stand firm: survey
Oil prices rise sharply on hopes of higher energy demand
After Oil's Friday Surge, Is The Bottom Finally In?

Future work

- Improve the methodology for relation extraction and observing relations in time.
- Provide a generic search for topics by concepts or keywords.

Acknowledgements

This work was supported by the euBusinessGraph (ICT-732003- IA) project.