



# **Patterns in Data**

**Nello Cristianini  
University of Bristol**

**“Information Beyond Shannon”  
Venice, December 29<sup>th</sup>, 2008**



# Information *beyond* Shannon

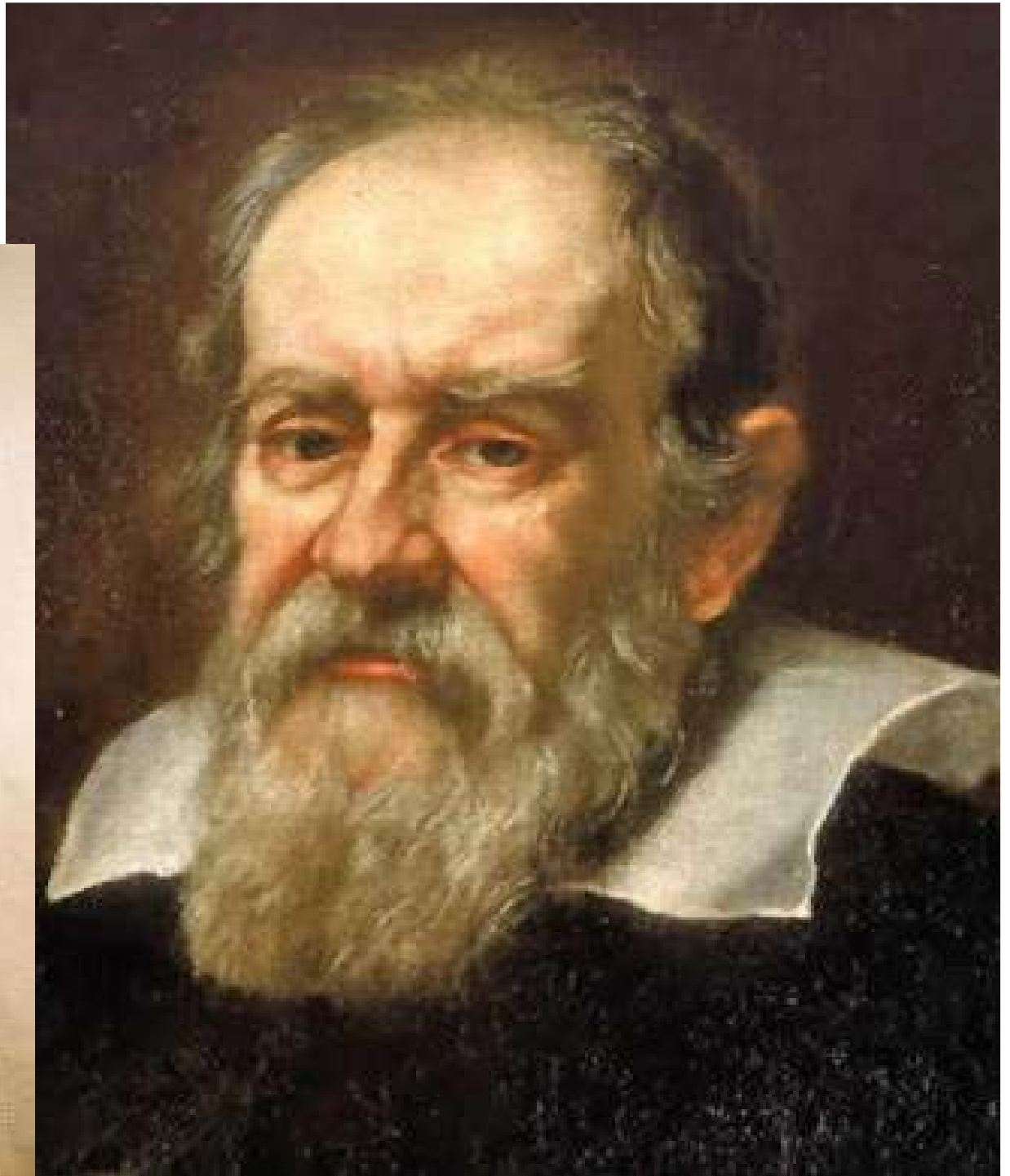
.... *beyond* ....

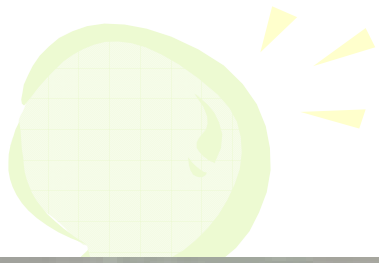


# *beyond*

- Something that I care about, and is about information, and about the future...

# Venice 1609

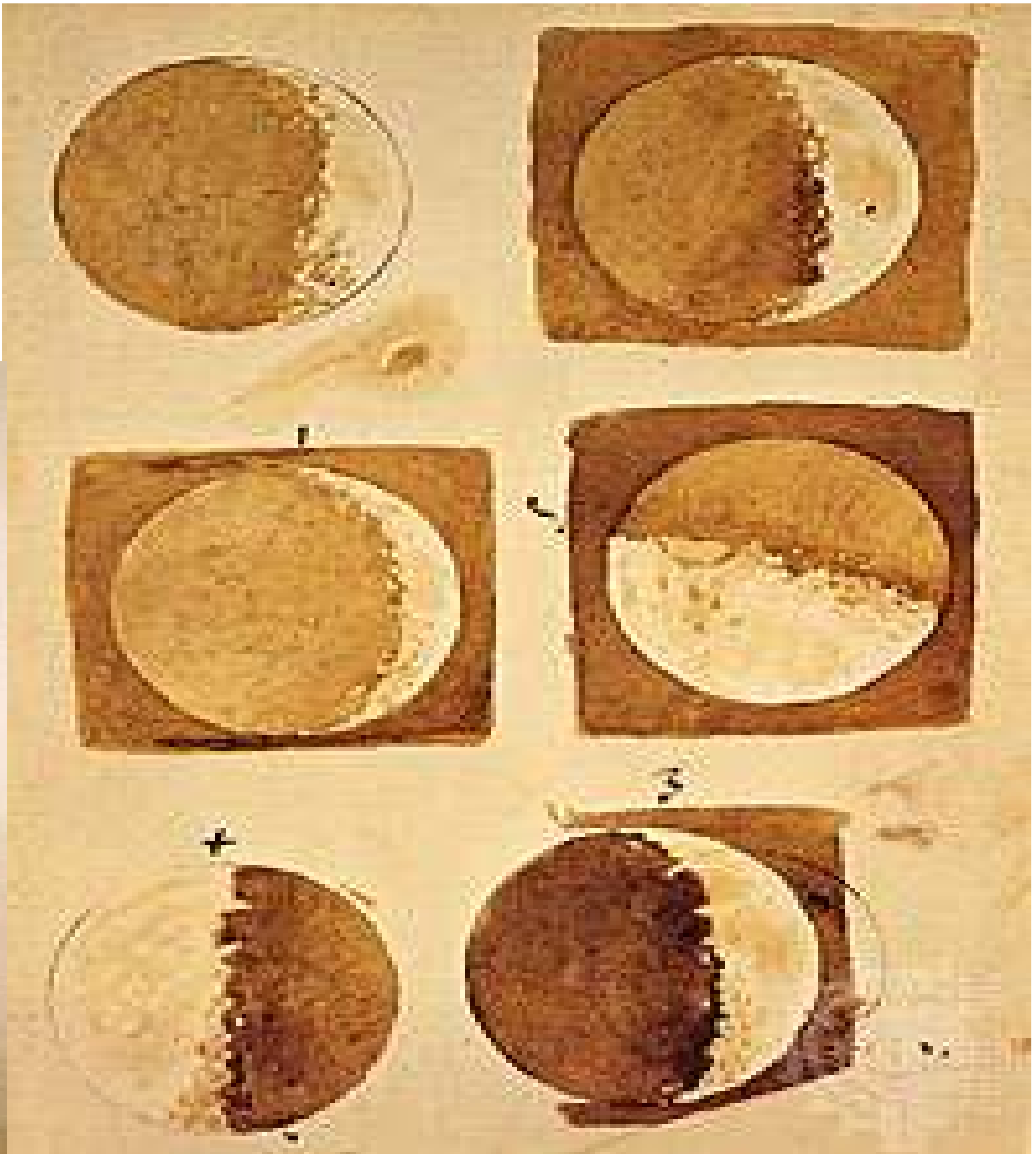




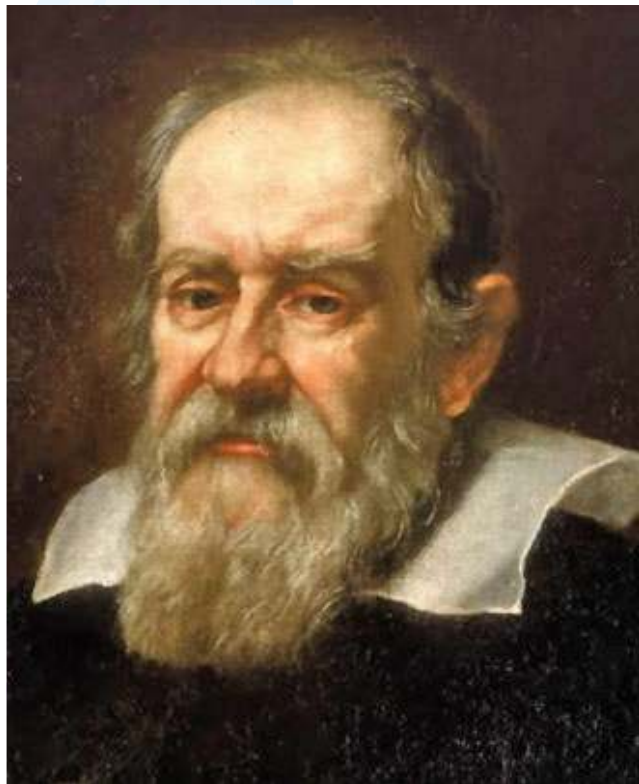
# Venice, 1609



# Discovery



# Publication



## S I D E R E V S N V N C I V S

MAGNA, LONGEQVE ADMIRABILIA  
Spectacula pandens, suspiciendaque proponens  
vnicuique, præfertim verò

*PHILOSOPHIS, atq; ASTRONOMIS, que à*

## G A L I L E O G A L I L E O P A T R I T I O F L O R E N T I N O

Patauini Gymnasij Publico Mathematico

### P E R S P I C I L L I

*Nuper à se reperti beneficio sunt obseruata in LVNÆ FACIE, FIXIS IN-  
NUMERIS, LACTEO CIRCVLO, STELLIS NEBVLOSIS,*

*Apprime verò in*

### Q V A T V O R P L A N E T I S

*Circa IOVIS Stellam disparibus interuallis, atque periodis, celesti-  
tate mirabili circumuolutis; quos, nemini in hanc vsque  
diem cognitos, nouissimè Author depra-  
hendit primus; atque*

## M E D I C E A S I D E R A

N V N C V P A N D O S D E C R E V I T .



VENETIIS, Apud Thomam Baglionum. M D C X.

*Superiorum Permissu, & Priuilegio.*






# Scientific Method

- How do we know what we know?  
(in science)


- 
- One possible way:  
experimental science

- Experiment → Hypothesis → Experiment →  
Hypothesis → Experiment → Hypothesis →  
Experiment → Hypothesis → Experiment →  
Hypothesis → Experiment → Hypothesis
- 




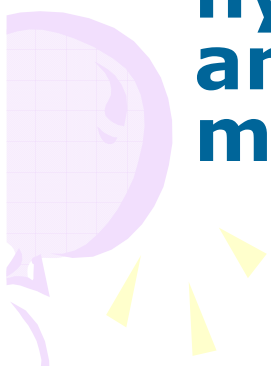
A lightbulb icon with a green grid pattern and yellow rays emanating from it, positioned at the top left of the slide.

# Scientific Method

- Formalising the process of experiment design
  - Formalising the representation of information and knowledge
  - Formalising the process of inductive inference
  - Rigorous (but not perfect) approach to knowledge generation
- 
- A lightbulb icon with a purple grid pattern and yellow rays emanating from it, positioned at the bottom left of the slide.



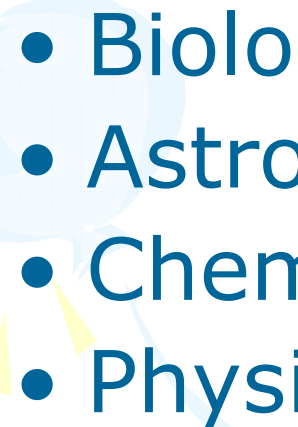
# Scientific Method

- Systematic gathering of information
  - Systematic organisation of knowledge
  - Systematic sharing of knowledge
  - Systematic scrutiny of hypotheses...
- 
- **Data acquisition via carefully designed experiments, and hypothesis generation via careful analysis are at the core of scientific method**
- 
- 

A lightbulb icon with a green grid pattern, yellow rays emanating from the top, and a green base with yellow rays. It is positioned to the left of the title.

# Scientific Method

- Things are changing fast...

- 
- A lightbulb icon with a blue grid pattern, yellow rays emanating from the top, and a blue base with yellow rays. It is positioned to the left of the list.
- Biology
  - Astronomy
  - Chemistry
  - Physics

A lightbulb icon with a purple grid pattern, yellow rays emanating from the top, and a purple base with yellow rays. It is positioned to the left of the text.

- The process is changing...

# SEQUENCING

- Tera bytes



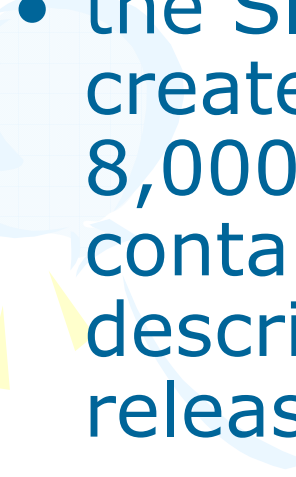
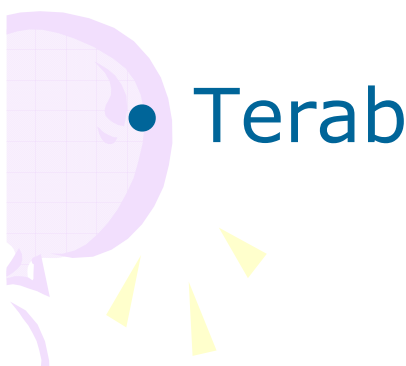
# LHC

- 15 petabytes per year







# Astronomy

- Automated Sky Surveys
    - the Sloan Digital Sky Survey (SDSS) created a 5-wavelength catalogue over 8,000 square degrees of the sky, containing about 200 million objects, described by hundreds of features (data released incrementally to the public).
  - Terabytes...
- 
- 



# Chemistry

- Combinatorial Chemistry

- 
- Hundreds of thousands of compounds can be generated and tested, either by using robotics, or – increasingly - by computer simulations, in what is essentially a survey of entire regions of chemical space, hunting for compounds with a given set of properties.
- 





# To Summarize...

- First we gather GAZILLIONS of data
- Then we let our computers sort it out, and suggest hypotheses...

## A Google DATACENTER



# Many Google DATACENTERS



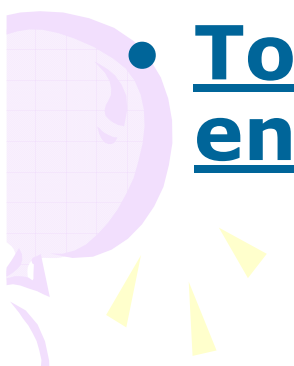


# A Newer Kind of ...?

- High quality experimental information is produced on massive scale, by automated processes
- And the data is analysed by pattern discovery software, which can generate hypotheses in the form of statistical relations discovered in the data





# Pattern Analysis

- Various disciplines have contributed to this
  - Statistics, data mining, machine learning, information theory, ...
  - We can MANAGE and ANALYSE massive amounts of data, of any kind...
  - Text, DNA, graphs, images, times series, ...
  - **To what extent can we automate the entire process of scientific discovery?**
- 





# Pattern Analysis

- The set of principles behind most successful algorithms in PA is actually quite small
  - Efficient search and optimisation in vast pattern spaces
  - Effective control of multiple testing issues, by use of sharp statistical tools
  - Principles that unify all, from suffix trees to itemset mining, from graphical models to kernel methods
- 
- 



# Science

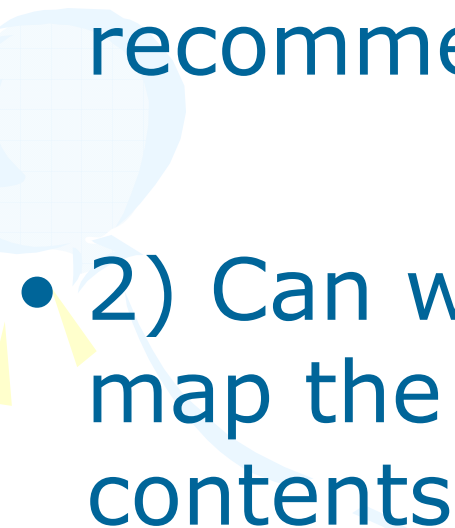
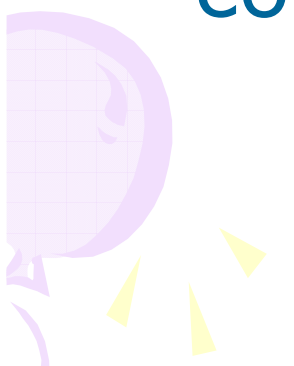
- How can this vast toolbox, and this conceptual framework, help us to acquire more scientific knowledge?

- 
- I will show you an example from the most improbable of domains... the SOCIAL sciences...
- 





# EXAMPLES FROM BRISTOL

- 1) Can we make an agent that reads all news in all languages, than makes recommendations?
  - 2) Can we use modern AI tools to map the mediasphere, based on its contents?
- 
- 



# Media System

- How do we know what we know?  
(in society)

# NewsPatterns - Saturday 21 June 2008

Monitoring 454 News Outlets and 6 Languages.  
 Read more...



- Read (18,79)
- Barack Obama (12,290)
- Jamie Lynn Spears (18)
- Other (61,89)

- People Today**
- Read (73)
  - Barack Obama (45)
  - John McCain (23)
  - Bush (20)
  - Britney Spears (19)
  - Jamie Lynn Spears (18)**
  - Robert Mugabe (14)
  - Amy Winehouse (11)
  - Candice Rice (10)
  - Tim Russert (8)

- Organisations Today**



## Jamie Lynn Spears (18)

**Jamie Lynn gives birth**  
 20 Jun 2008 04:16:27 GMT

It's a girl, y'all! Jamie Lynn Spears gave birth to Maddie Briann this morning at Mississippi Southwest Regional Medical Center in McComb, Miss. It's the first child for the 17-year old Zoey 101 star and her fiance, 19-year old  
 Read more...

**Jamie Lynn Spears Has a Baby Girl**  
 20 Jun 2008 04:16:14 GMT

There is a new baby in the Spears clan! Jamie Lynn Spears and her fiancé Casey Aldridge welcomed a baby girl Thursday morning. PEOPLE confirms exclusively that the baby was named Maddie Briann. "Just the family was there," says a source about the birth around 9:30 a.m. at a hospital in Mississippi, near her Louisiana hometown. "Everyone  
 Read more...

**Report: Spears Gives Birth To Girl, Maddie**  
 20 Jun 2008 04:15:22 GMT

People magazine reports that Britney Spears' younger sister, Jamie Lynn Spears, has given birth.  
 Read more...

Search Today's News


**Locations Today**

- U.S. (157)
- China (58)
- New York (27)
- Japan (27)
- Washington (26)
- London (24)
- United States (21)
- Iraq (20)
- India (20)
- Britain (26)
- More >

**Sequence**

- jung sun sui ky (3)
- ile de france (4)
- nicolas sarkozy (4)
- the treaty of lissabon (3)
- the heads of state (4)
- summit in brussels (3)
- of the european union (3)
- secretary of state for sports lemand laporte (2)
- the second round of G2
- around paris (3)
- More >

**Sequence from Translated Spanish**



NewsPatterns - Mozilla Firefox  
 http://newspatterns.enn.bris.ac.uk/

# NewsPatterns - Saturday 21 June 2008

Monitoring 454 News Outlets and 6 Languages.  
 Read more...

Map | Timeline | **Social Network** | Bar Chart | Text Only

**Jamie Lynn Spears Activity**



**Jamie Lynn Spears Share**



- Read (18,79)
- Barack Obama (12,29)
- Jamie Lynn Spears (7,9)
- Other (61,36)

**People Today**

- Read (73)
- Barack Obama (45)
- John McCain (23)
- Bush (20)
- Britney Spears (19)
- Jamie Lynn Spears (18)
- Robert Muzsa (14)
- Ami Winehouse (11)
- Candolezza Rice (10)
- Tim Russert (9)

**Organisations Today**



## Jamie Lynn Spears (18)

### Jamie Lynn gives birth

20 Jun 2008 04:18:57 GMT

It's a girl, y'all! Jamie Lynn Spears gave birth to Maddie Briann this morning at Mississippi Southwest Regional Medical Center in McComb, Miss. It's the first child for the 17-year old Zoey 101 star and her fiance, 19-year old.  
 Read more...

### Jamie Lynn Spears Has a Baby Girl

20 Jun 2008 04:18:14 GMT

There is a new baby in the Spears clan! Jamie Lynn Spears and her fiance Casey Aldridge welcomed a baby girl Thursday morning. PEOPLE confirms exclusively that the baby was named Maddie Briann.  
 "Just the family was there," said a source about the birth around 9:30 a.m. at a hospital in Mississippi, near her Louisiana hometown. "Everybody

Search Today's News

**Locations Today**

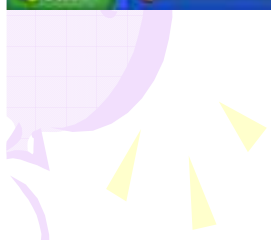
- U.S. (157)
- China (58)
- New York (27)
- Japan (37)
- Washington (26)
- London (24)
- United States (21)
- Iraq (20)
- India (20)
- Britain (26)

**Sequence**

- jung jin soo lee (3)
- ile de france (4)
- nicolas Sarkozy (6)
- the treaty of Lisbon (3)
- the heads of state (4)
- summit in brussels (2)
- of the european union (3)
- secretary of state for sports Bernard laporte (2)
- the second round of G5
- around paris (3)

Sequence from Translated Spanish

Transferring data from newspatterns.enn.bris.ac.uk...



# NewsPatterns - Saturday 21 June 2008

Monitoring 454 News Outlets and 6 Languages.  
Read more...

Map Timeline Social Network Bar Chart Text Only



## Jamie Lynn Spears (18)

### Jamie Lynn gives birth

20 Jun 2008 04:18:27 GMT

It's a girl, y'all! Jamie Lynn Spears gave birth to Maddie Briann this morning at Mississippi Southwest Regional Medical Center in McComb, Miss. It's the first child for the 17-year old Zoey 101 star and her fiance, 19-year old  
[Read more...](#)

### Jamie Lynn Spears Has a Baby Girl

20 Jun 2008 04:16:14 GMT

There is a new baby in the Spears clan! Jamie Lynn Spears and her fiancé Casey Aldridge welcomed a baby girl Thursday morning. PEOPLE confirms exclusively that the baby was named Maddie Briann. "Just the family was there," says a source about the birth around 9:30 a.m. at a hospital in Mississippi, near her Louisiana hometown. "Everyone  
[Read more...](#)

### Report: Spears Gives Birth To Girl, Maddie

20 Jun 2008 04:15:22 GMT

People magazine reports that Britney Spears' younger sister, Jamie Lynn Spears, has given birth.  
[Read more...](#)

Search Today's News

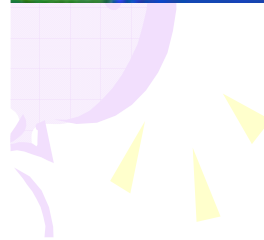
Locations Today

- U.S. (157)
- China (58)
- New York (27)
- Japan (37)
- Washington (26)
- London (24)
- United States (21)
- Iraq (20)
- India (20)
- Britain (26)
- More >

Sequence

- Jung can see fly (3)
- ile de france (4)
- nicolas Sarkozy (6)
- the treaty of lissabon (3)
- the heads of state (4)
- summit in brussels (3)
- of the european union (3)
- secretary of state for sports Bernard laporte (2)
- the second round of G8
- around paris (3)
- More >

Sequence from Translated Spanish





# Election Watch'08

Election Watch automatically summarises news about the 2008 US Presidential race. [More](#)



## Read Activity



## Read Share



- Read (31.2%)
- Barack Obama (19.2%)
- John McCain (9.8%)
- Other (39.7%)



Heat Map Map Timeline Bar Chart Word Cloud Text Only



## Read (73)

### China to raise fuel prices

20 Jun 2008 04:13:53 GMT

Read full story for latest details.  
[Read more...](#)

### Thailand braces for anti-government protests

20 Jun 2008 04:13:53 GMT

Read full story for latest details.  
[Read more...](#)

### French economy forecast to halt

20 Jun 2008 04:13:53 GMT

Read full story for latest details.  
[Read more...](#)

## Search

## Politicians

- Read (73)
- Barack Obama (45)
- John McCain (23)
- Bush (20)
- Robert Mugabe (14)
- Condoleezza Rice (10)
- Tim Russert (9)
- Gordon Brown (8)
- Ran (7)
- Michelle Obama (7)
- [More >](#)

Celebrity Watch - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Getting Started Latest Headlines INI laptop login

# Celebrity Watch

Celebrity Watch summarises the World's hottest celebrity gossip news! [Read more...](#)

David Beckham Activity

David Beckham Share

■ Read (25.1%)  
■ Barack Obama (15.5%)  
■ David Beckham (7.9%)  
■ Other (51.5%)

PASCAL

Transferring data from celebritwatch.ens.bris.ac.uk...

Map Timeline **Social Network** Bar Chart Text Only

David Beckham (8)

### San Francisco crowd goes nuts as David Beckham takes it all off

20 Jun 2008 04:13:55 GMT

Don't miss this video of a San Francisco crowd shrieking and squealing with delight as a gigantic, almost-nude shot of David Beckham was revealed inch-by-inch at Macy's in Union Square yesterday. The huge billboard is an ad for Emporio Armani underwear, but the promoters kept it draped until showtime and then dropped the curtain slowly as the crowd went wild. It was better than a strip joint and you didn't have to stick any money in his drawers. AddThis Button BEGIN addthis\_pub = 'mizizb'; AddThis Button END

[Read more...](#)

### David Beckham's hot manhood

20 Jun 2009 04:11:04 GMT

Search

Who's Hot

- < Less
- Read (73)
- Barack Obama (45)
- John McCain (23)
- Bush (20)
- Britney Spears (19)
- Jamie Lynn Spears (18)
- Amy Winehouse (11)
- Condoleezza Rice (10)
- Jamie Lynn (9)
- David Beckham (8)
- Gordon Brown (8)
- Charlie Sheen (8)
- Ran (7)
- Sienna Miller (7)
- Denise Richards (7)
- Laws (6)
- Steve Carell (6)
- Hugo Chavez (6)



# Found In Translation

FOUND  
In Translation

Found in Translation monitors  
news outlets in languages.  
[More...](#)

Yr Hw  
Fun Ce  
Yr Hw  
Fun Centre

## EU Activity



## EU Share



- ★ EU (51.9%)
- ★ Belgium (2.0%)
- ★ Cyprus (0.5%)
- ★ Other (45.6%)

## Weather

- ★ EU (286)
- ★ Belgium (11)
- ★ Cyprus (3)
- ★ Czech Republic (35)
- ★ Denmark (7)
- ★ Estonia (3)
- ★ France (50)
- ★ Germany (8)
- ★ Greece (18)
- ★ Hungary (1)
- ★ Italy (50)
- ★ Lithuania (2)
- ★ Netherlands (12)
- ★ Portugal (17)
- ★ Romania (2)
- ★ Slovenia (11)
- ★ Spain (8)
- ★ Sweden (8)
- ★ United Kingdom (19)

Weather EU News Local News Text Only



## EU (286)

### EU aid to Bulgaria longer suspended

The European Union has frozen aid to EU Member Bulgaria because of fraud and corruption. The aid will total EUR 560 million and was intended for agriculture and infrastructure.

[Read more...](#)

### Greenland votes on autonomy

In Greenland on Tuesday, the polls are open for a referendum on extending the autonomy of the capital island to Denmark.

[Read more...](#)

### Yellow River is heavily polluted

A third of the Yellow River in China is so polluted by industrial waste water not only is inappropriate as drinking water, but also for the industry or agriculture.

[Read more...](#)

### A woman washes up baby by toilets

A British woman to the judge said Monday that her new-born baby toilets have inadvertently by the year. The woman said she thought she had diarrhoea.

[Read more...](#)

## Search

## Local-News

- ★ EU (295)
- ★ Belgium (11)
- ★ Czech Republic (31)
- ★ Denmark (5)
- ★ Estonia (3)
- ★ France (50)
- ★ Germany (5)
- ★ Greece (21)
- ★ Hungary (1)
- ★ Italy (50)
- ★ Lithuania (2)
- ★ Netherlands (12)
- ★ Portugal (11)
- ★ Romania (2)
- ★ Slovenia (11)
- ★ Spain (8)
- ★ Sweden (23)
- ★ United Kingdom (19)

## EU-News

- ★ EU (298)
- ★ Belgium (11)
- ★ Czech Republic (31)
- ★ Denmark (5)
- ★ Estonia (3)
- ★ France (50)
- ★ Germany (8)
- ★ Greece (21)
- ★ Hungary (1)
- ★ Italy (50)
- ★ Lithuania (2)
- ★ Netherlands (12)
- ★ Portugal (11)

SMART

# Found In Translation

FOUND  
In Translation

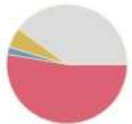
Found in Translation monitors  
news outlets in languages.  
[More...](#)

Yr Hw  
Fun Ce.  
Yr Hw  
Fun Centre

## Belgium Activity



## Belgium Share



- ★ EU (52.7%)
- ★ Belgium (2.0%)
- ★ Czech Republic (5.5%)
- ★ Other (39.8%)

## Weather

- ★ EU (286)
- ★ Belgium (11)
- ★ Cyprus (3)
- ★ Czech Republic (35)
- ★ Denmark (7)
- ★ Estonia (3)
- ★ France (50)
- ★ Germany (8)
- ★ Greece (18)
- ★ Hungary (1)
- ★ Italy (50)
- ★ Lithuania (2)
- ★ Netherlands (12)
- ★ Portugal (17)
- ★ Romania (2)
- ★ Slovenia (11)
- ★ Spain (8)
- ★ Sweden (8)
- ★ United Kingdom (19)

Weather | EU News | Local News | Text Only



## Belgium (11)

### EU aid to Bulgaria longer suspended

The European Union has frozen aid to EU Member Bulgaria because of fraud and corruption. The aid will total EUR 560 million and was intended for agriculture and infrastructure.

[Read more...](#)

### Greenland votes on autonomy

In Greenland on Tuesday, the polls are open for a referendum on extending the autonomy of the capital island to Denmark.

[Read more...](#)

### Yellow River is heavily polluted

A third of the Yellow River in China is so polluted by industrial waste water not only is inappropriate as drinking water, but also for the industry or agriculture.

[Read more...](#)

### A woman washes up baby by toilets

A British woman to the judge said Monday that her new-born baby toilets have inadvertently by the year. The woman said she thought she had diarrhoea.

[Read more...](#)

## Search

## Local-News



- ★ EU (295)
- ★ Belgium (11)
- ★ Czech Republic (31)
- ★ Denmark (5)
- ★ Estonia (3)
- ★ France (50)
- ★ Germany (5)
- ★ Greece (21)
- ★ Hungary (1)
- ★ Italy (50)
- ★ Lithuania (2)
- ★ Netherlands (12)
- ★ Portugal (11)
- ★ Romania (2)
- ★ Slovenia (11)
- ★ Spain (8)
- ★ Sweden (23)
- ★ United Kingdom (19)

## EU-News

- ★ EU (298)
- ★ Belgium (11)
- ★ Czech Republic (31)
- ★ Denmark (5)
- ★ Estonia (3)
- ★ France (50)
- ★ Germany (8)
- ★ Greece (21)
- ★ Hungary (1)
- ★ Italy (50)
- ★ Lithuania (2)
- ★ Netherlands (12)




# Mapping the Mediasphere

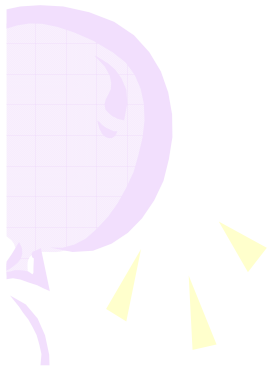
- We are interested in describing the mediasphere as a single, unique, system
  - We are interested in the flow of information through the system
  - And in any other patterns, both dynamic and static, that can be detected in its contents
- 
- 



# Tracking memes

- We track memes as they travel from outlet to outlet

- 
- We use them to reconstruct the invisible influence structure of the system





# Mediasphere

- The mediasphere is the integrated system of global news media, including tens of thousands of broadcast, press, online outlets, in dozens of languages.
- How big is it?

# The Size of the Mediasphere

9138 outlets appeared at some point in Google News

(does this include foreign outlets?) –

they have 22 languages – but different set of languages from us

If we sum together the sizes declared by google for each language

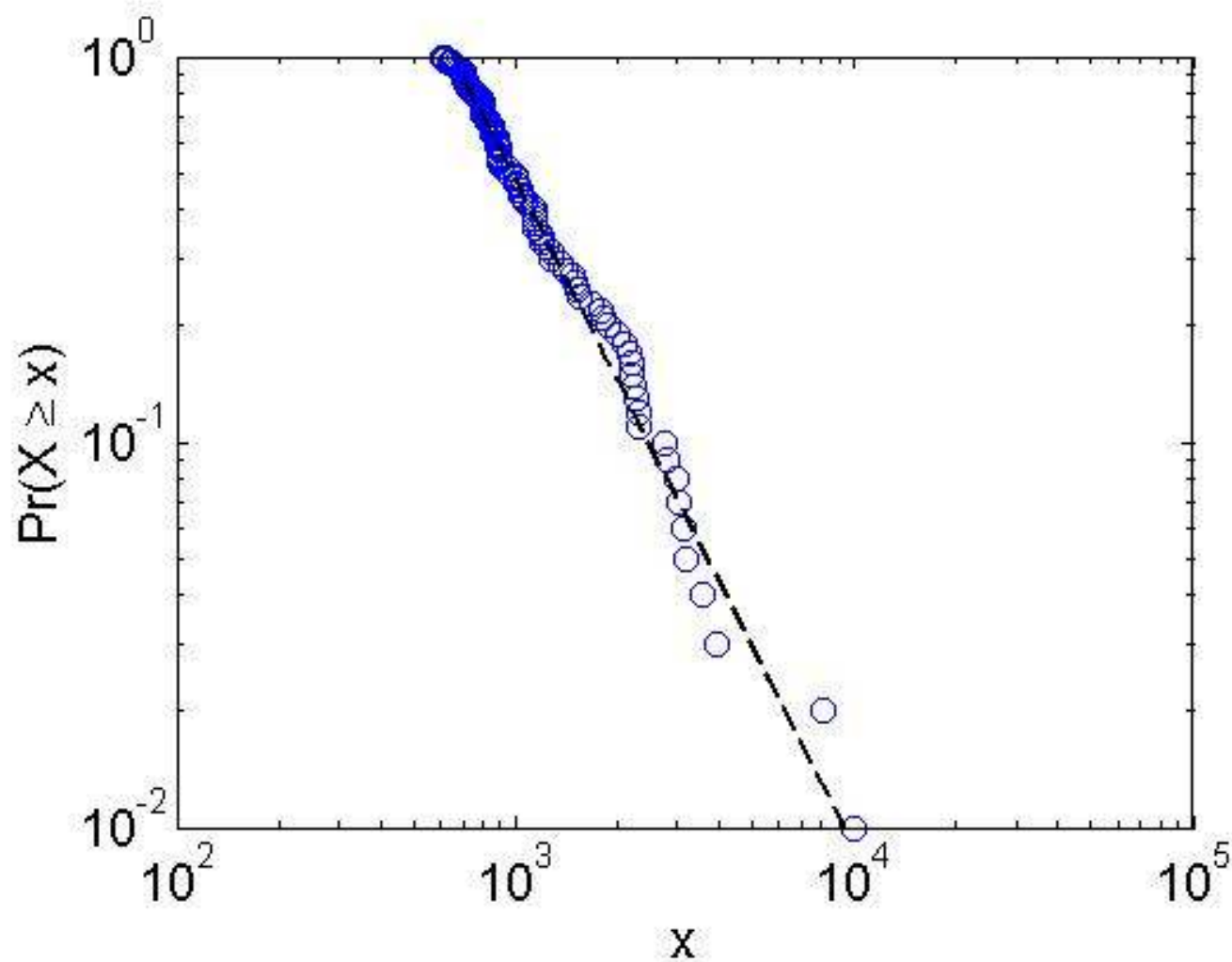
(they are all approx, we get a similar figure – so we know they over-state a bit)

English	4500
China	1000
Korea	1000
Spanish	700
German	700
Japan	610
French	500
Belgie	400
Russia	400
cesca rep	400

Nederland	400
Norge	400
Taiwan	350
Polska	300
Italian	250
Turkie	200
Portugal	200
Arabic	150
Sverige	100
Israel	100
Greece	100
IndiaB	20
SUM	12780



# Circulation of top 100 global newspapers









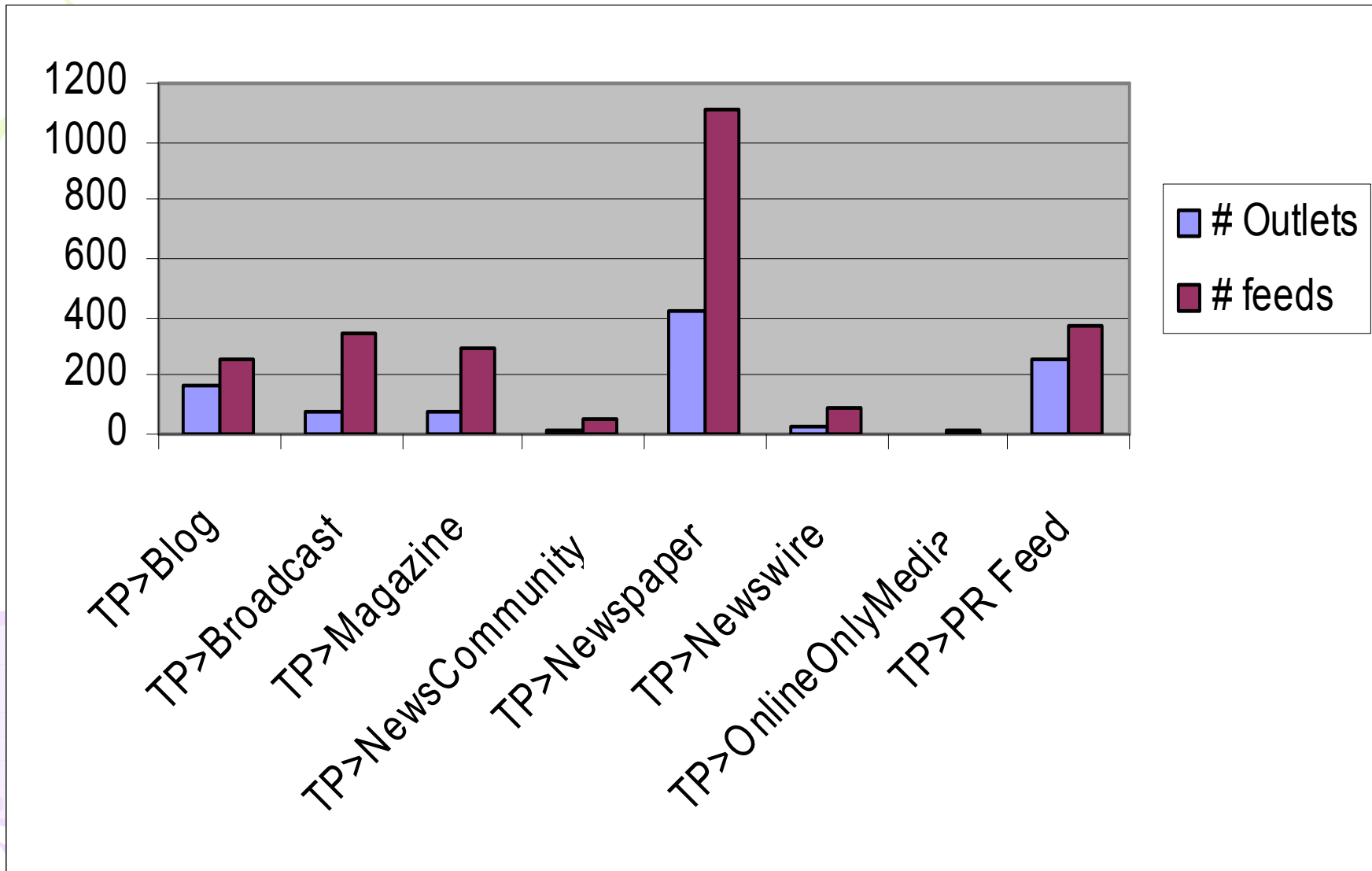
# Our Data

- For period: april to nov 2008  
We have 1007 outlets

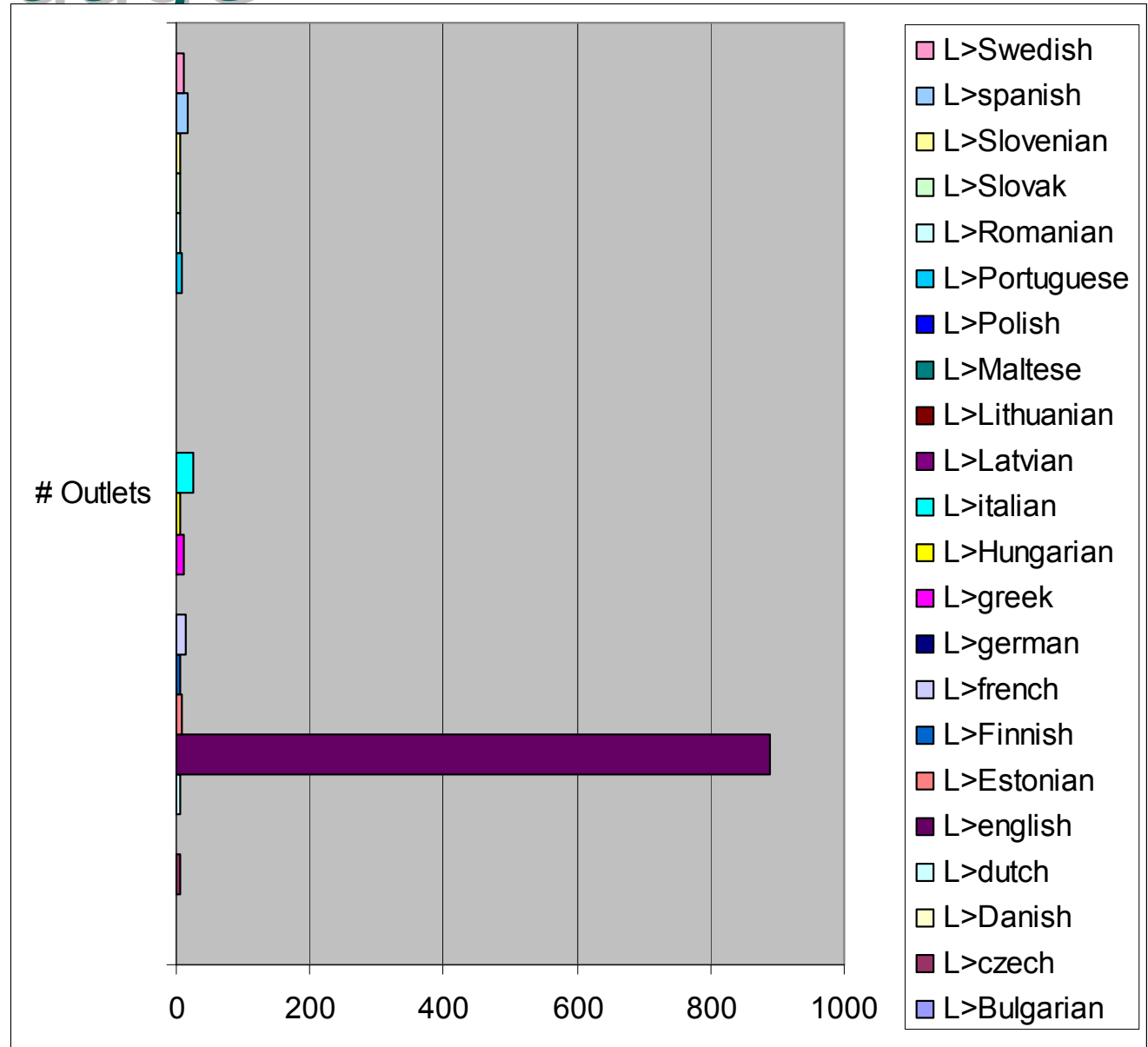
- 
- for 507 we have complete coverage (all top stories for the entire period)
  - For 723 we have 80% coverage (all top stories for at least 80% of the period).
- 

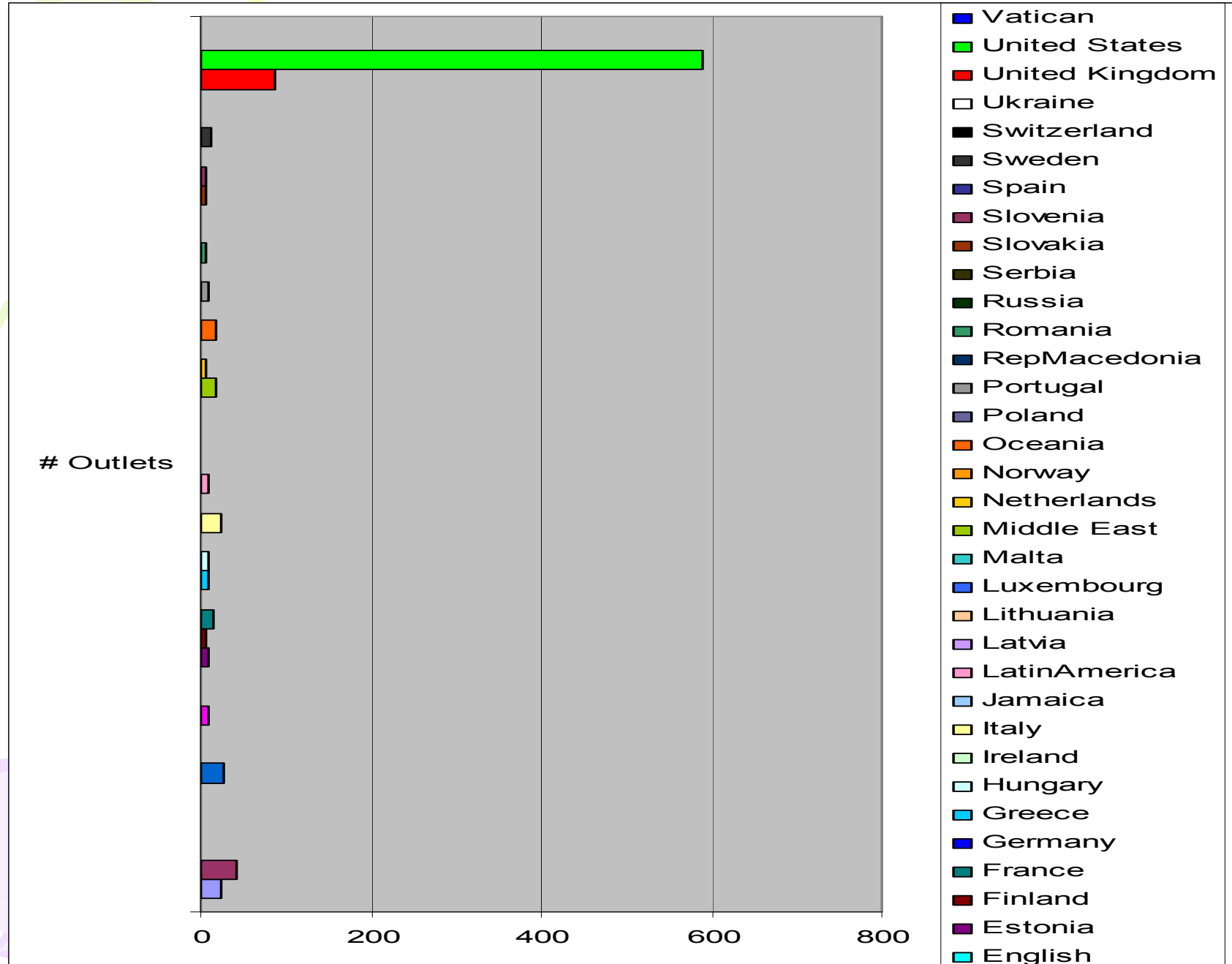


# Our Data – By Media Type



# By Language




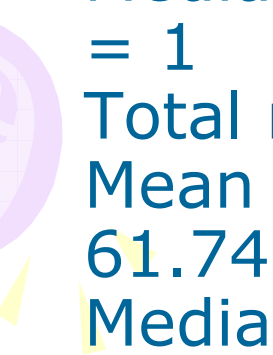


- Vatican
- United States
- United Kingdom
- Ukraine
- Switzerland
- Sweden
- Spain
- Slovenia
- Slovakia
- Serbia
- Russia
- Romania
- RepMacedonia
- Portugal
- Poland
- Oceania
- Norway
- Netherlands
- Middle East
- Malta
- Luxembourg
- Lithuania
- Latvia
- LatinAmerica
- Jamaica
- Italy
- Ireland
- Hungary
- Greece
- Germany
- France
- Finland
- Estonia
- English

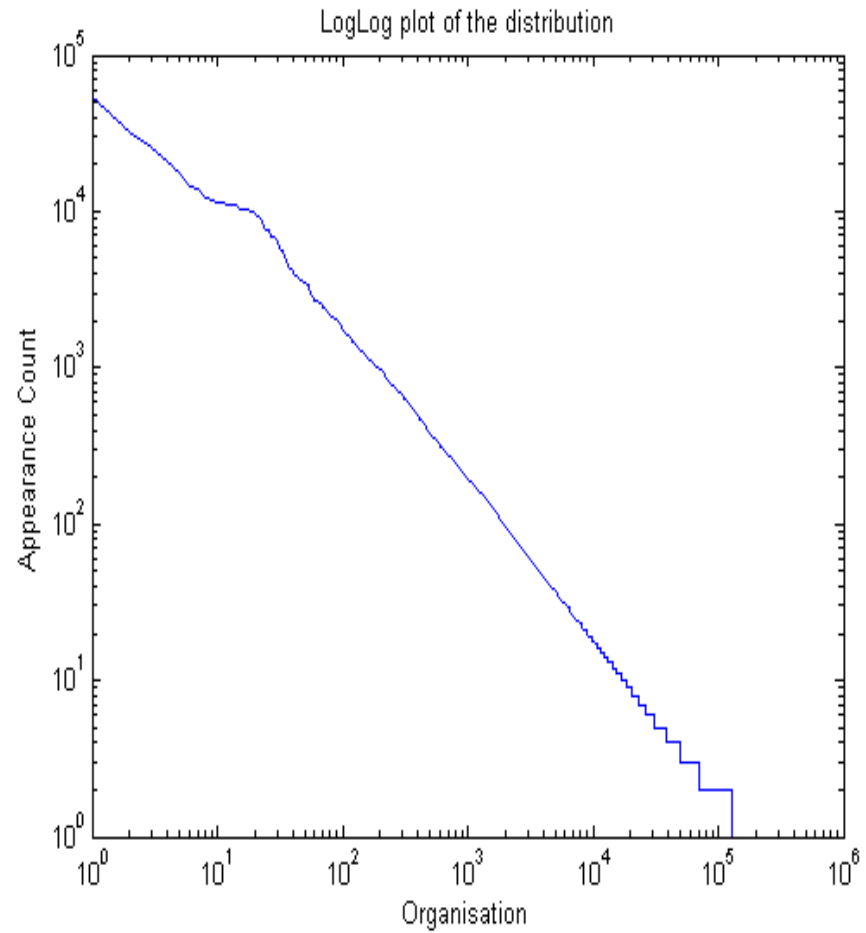


# The Contents of the Mediasphere

- NAMED ENTITIES:

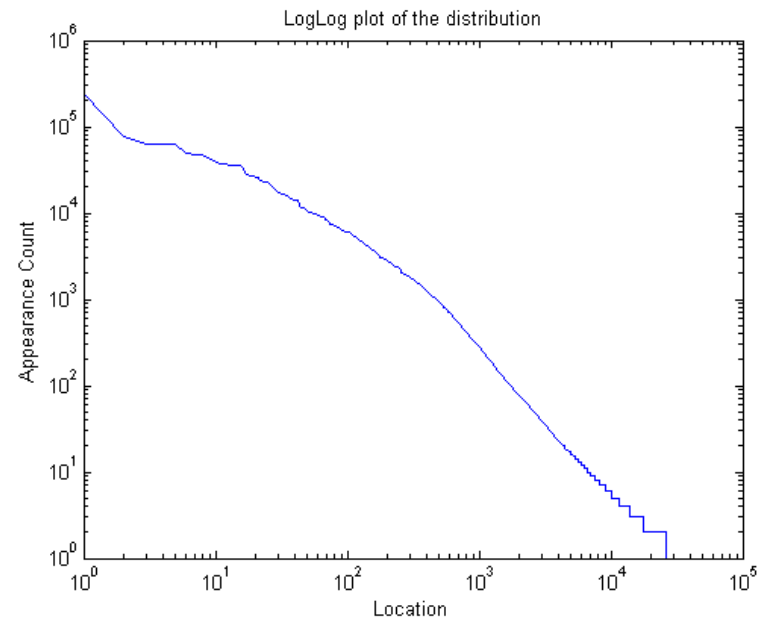
- 
- Mean number of appearances per person = 6.0995
  - Median number of appearances per person = 1
  - Total number of people = 447221
  - Mean number of appearances per organisation = 5.3305
  - Median number of appearances per organisation = 1
  - Total number of organisations = 420637
  - Mean number of appearances per location = 61.7462
  - Median number of appearances per location = 1
- 

# Distribution of Entities

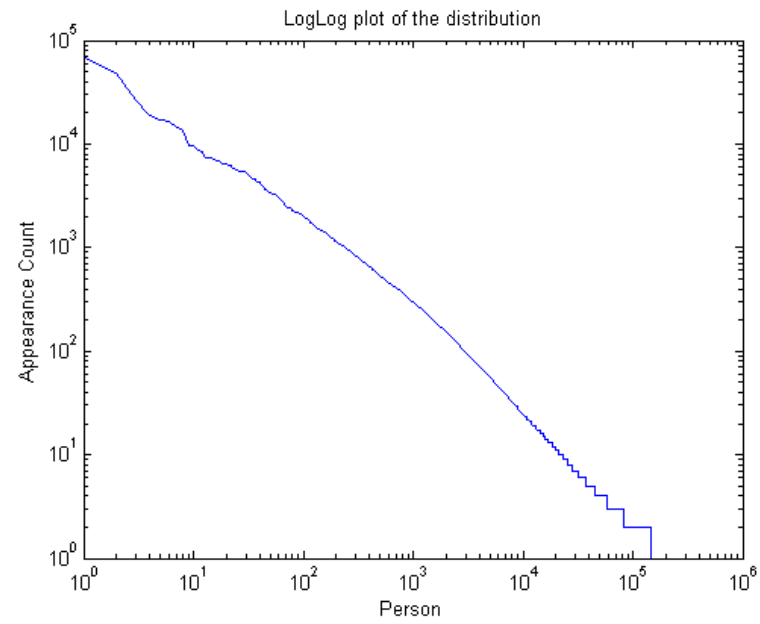
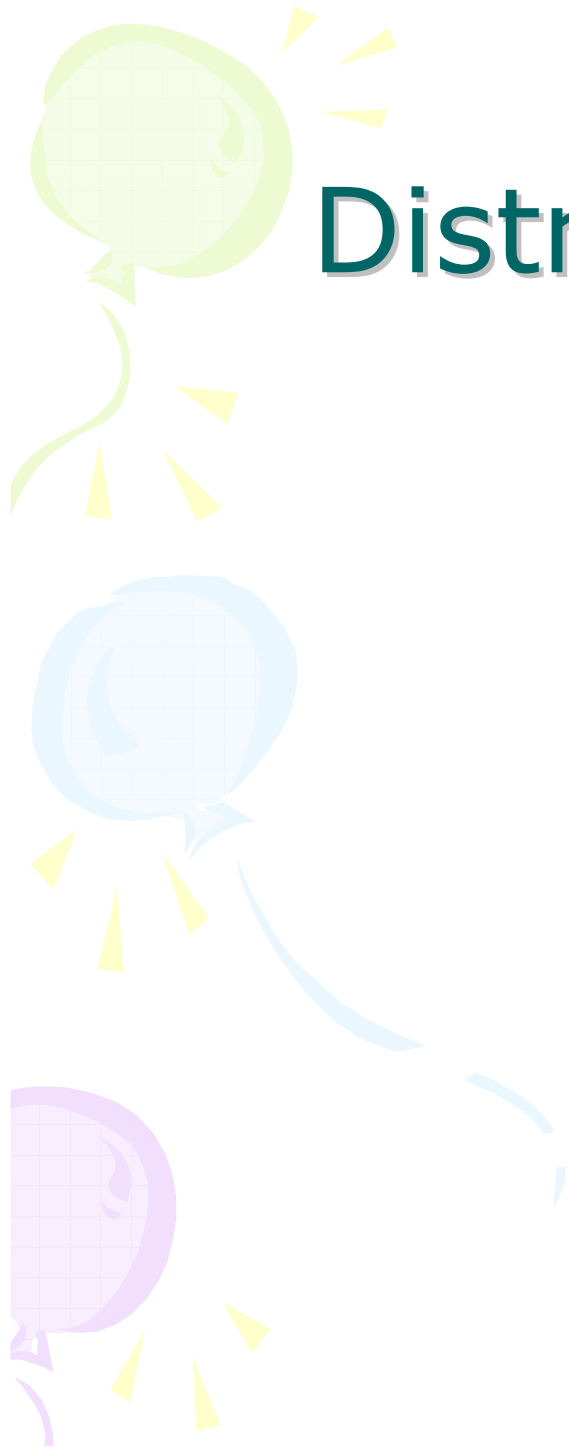




# Distribution of Entities

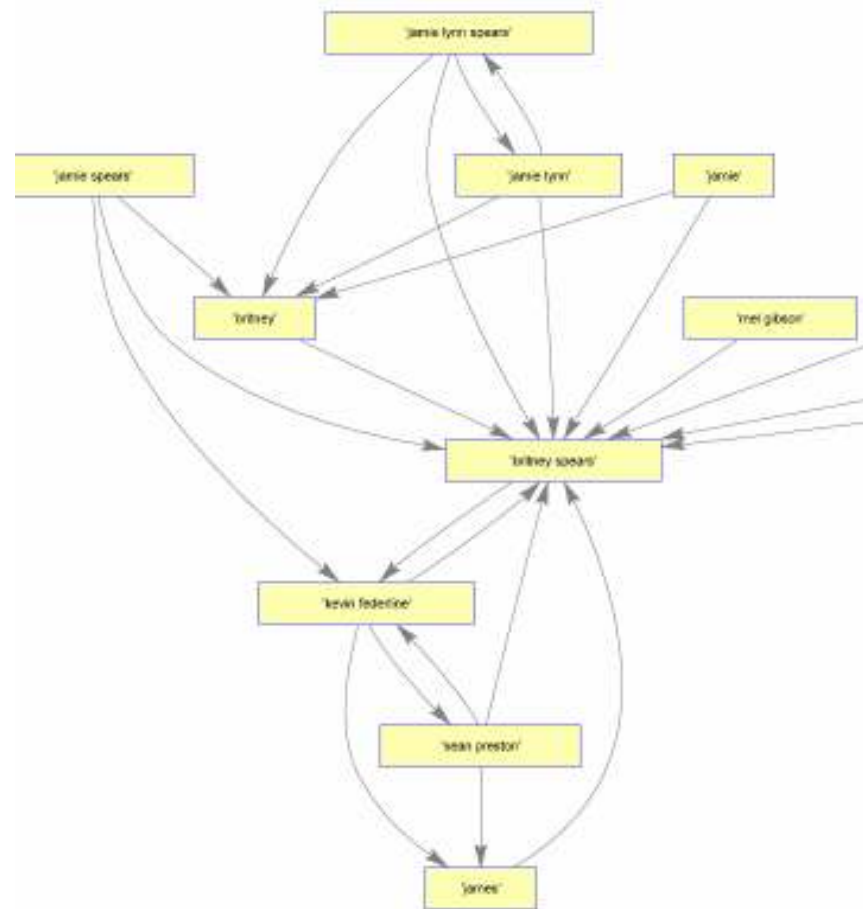


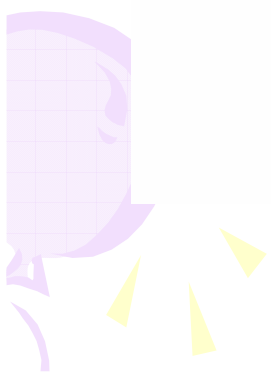
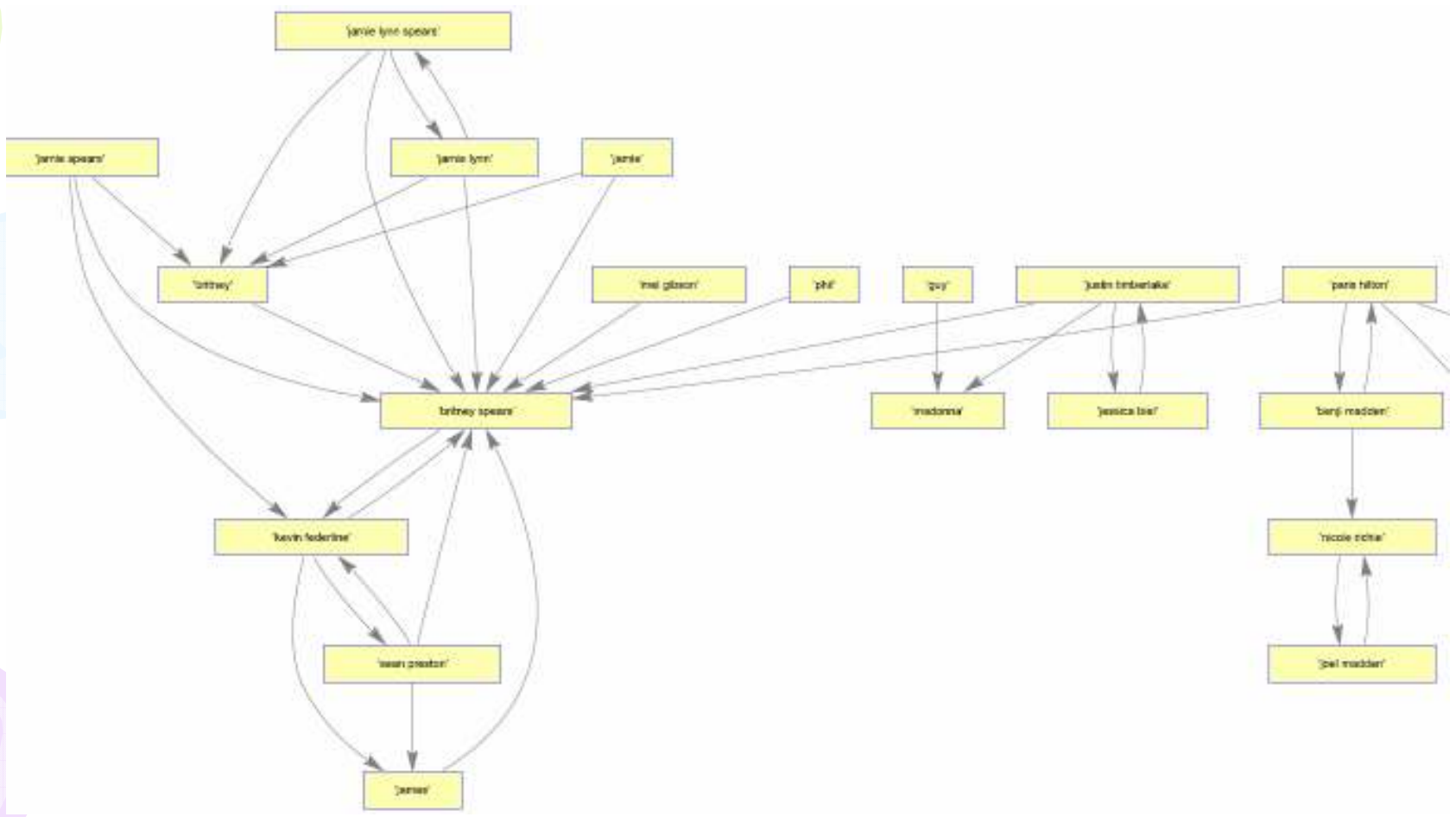
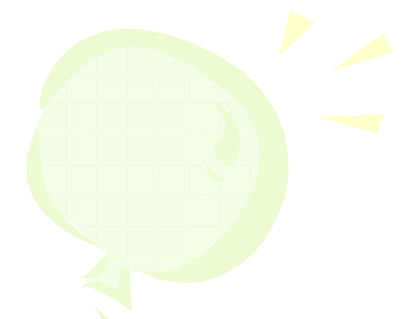
# Distribution of Entities



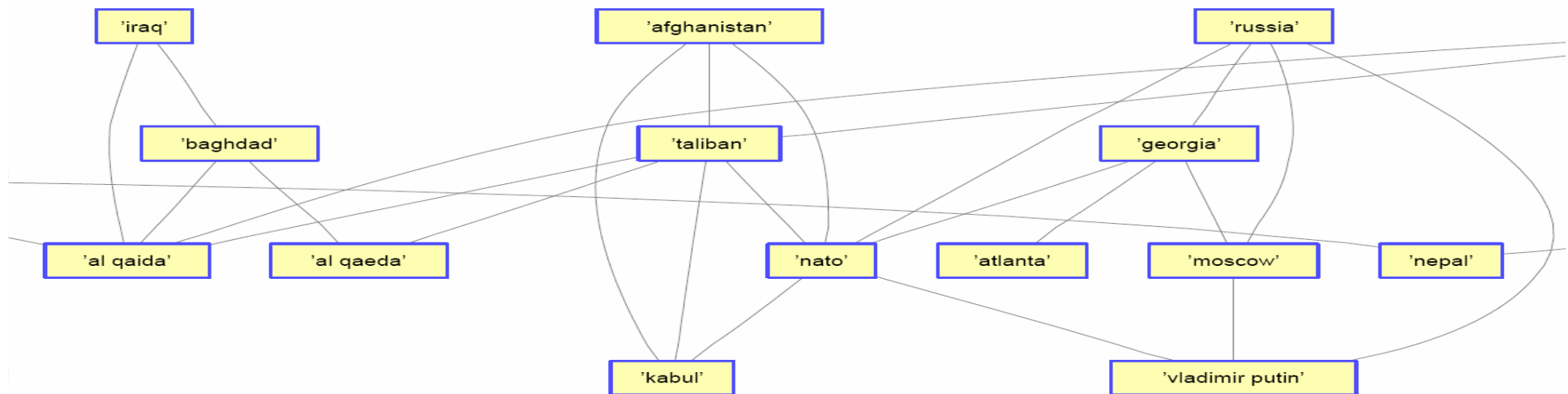
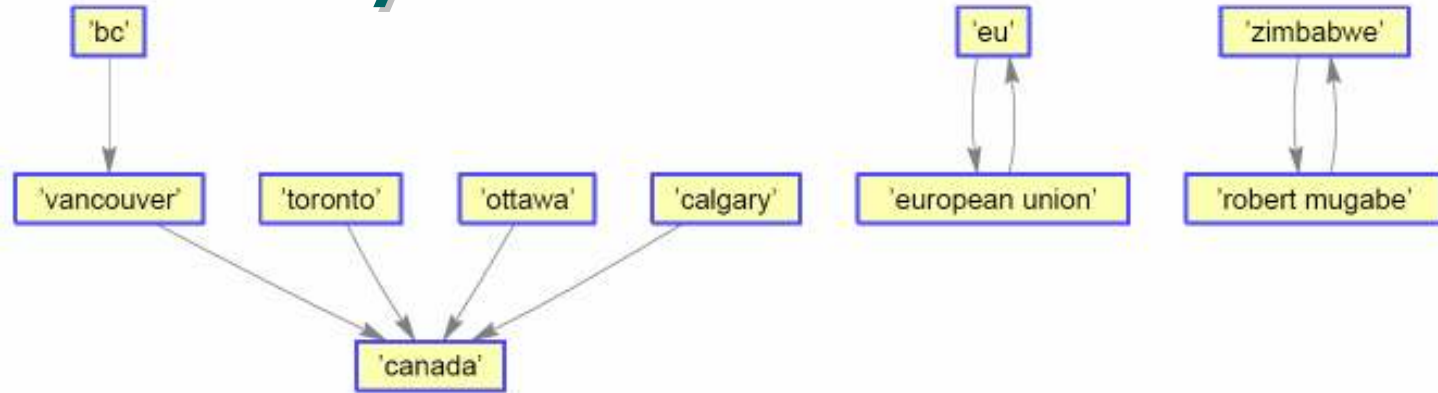
# Co-Mention Network

- The co-mention network can be used to link significantly associated entities, and hence be used to measure their social CENTRALITY





# Entity Networks



# Detecting Changes

## Election Watch'08

Election Watch monitors 595 outlets for news concerning the 2008 US Presidential election. [More](#)



### John McCain Activity

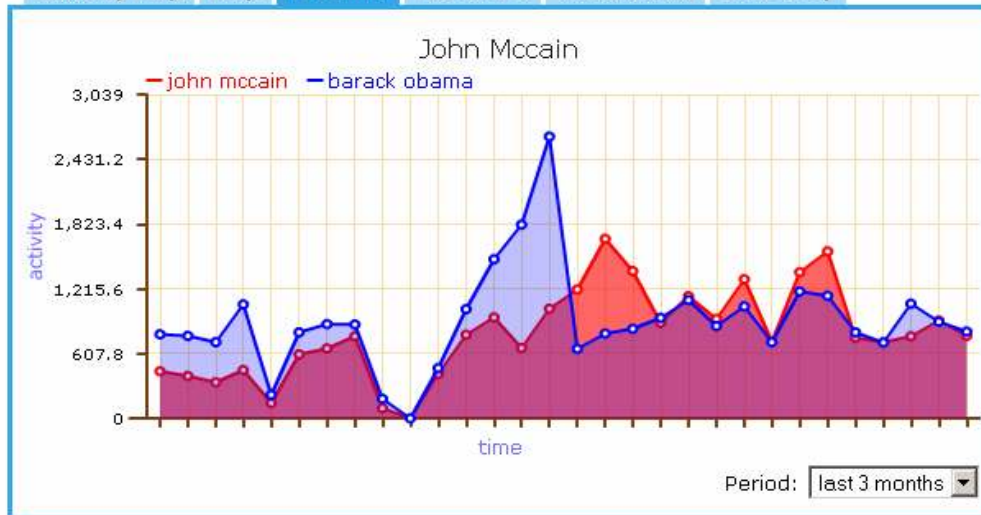


### John McCain Share



- ★ [John McCain \(27.3%\)](#)
- ★ [Barack Obama \(25.5%\)](#)
- ★ [Sarah Palin \(9.8%\)](#)
- ★ [Other \(37.3%\)](#)

Activity Map Map Timeline Bar Chart Word Cloud Text Only



### Search

### Politicians

- ★ [John McCain \(153\)](#)
- ★ [Barack Obama \(143\)](#)
- ★ [Sarah Palin \(55\)](#)
- ★ [Gordon Brown \(28\)](#)
- ★ [Bill \(23\)](#)
- ★ [Joe Biden \(23\)](#)
- ★ [David Letterman \(21\)](#)
- ★ [Clinton \(21\)](#)
- ★ [Brown \(20\)](#)
- ★ [Al Maliki \(20\)](#)



## John McCain (153)

### The politics of hate



13 Oct 2008 04:25:26 GMT

The ugly desperation of the McCain campaign rampages further out of control: Media Bubble, Oct. 8 -- John McCain's bid for the Oval Office suffered another stunning blow yesterday when



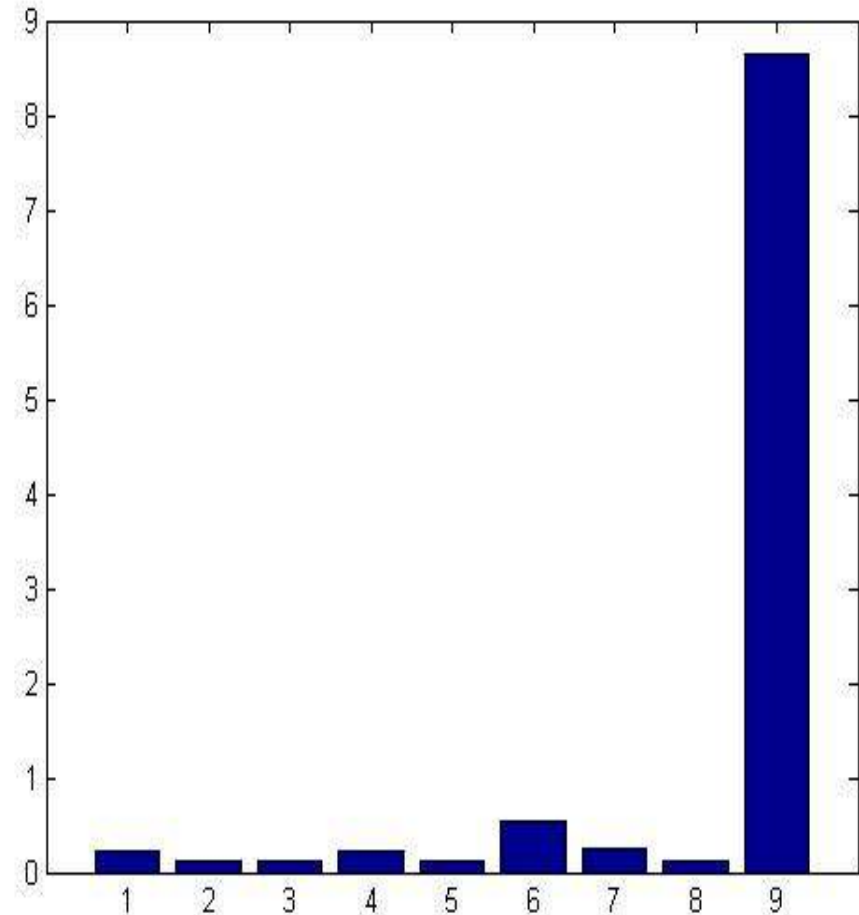


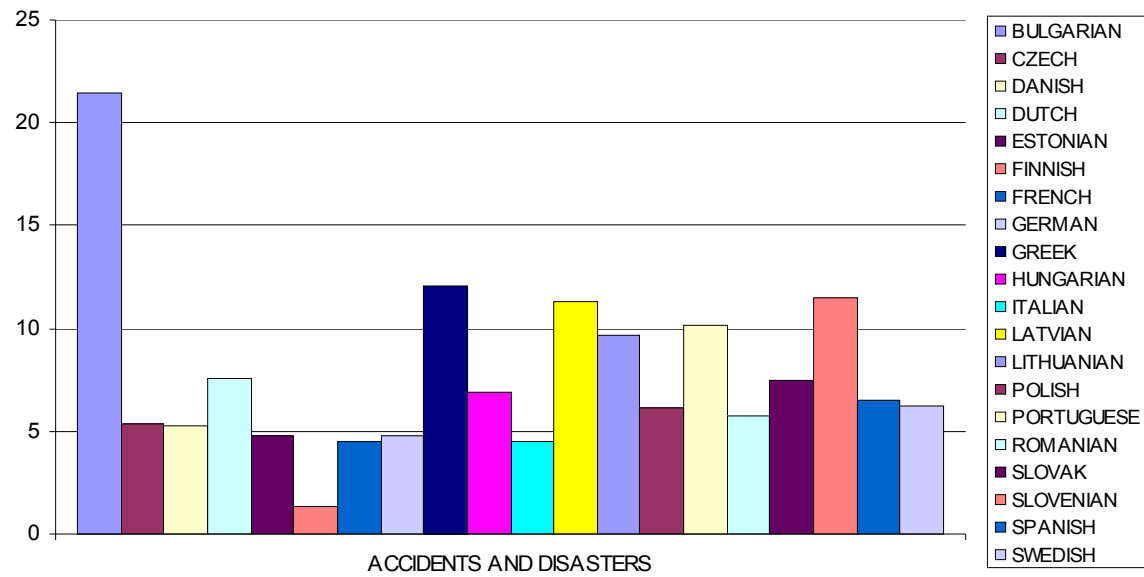
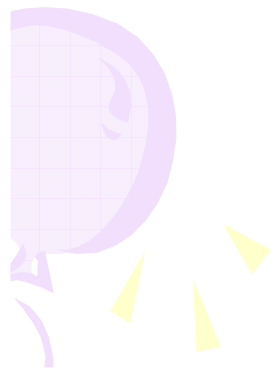
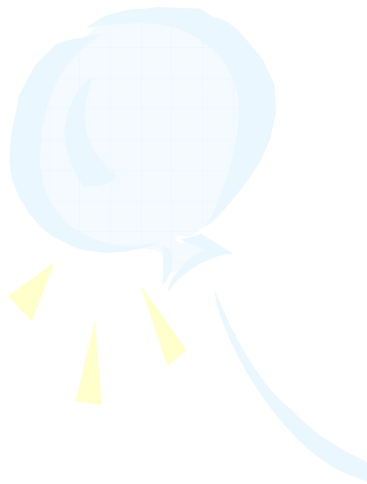
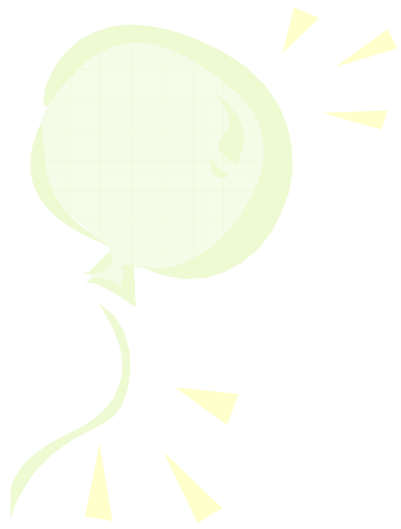
# Pope-ularity

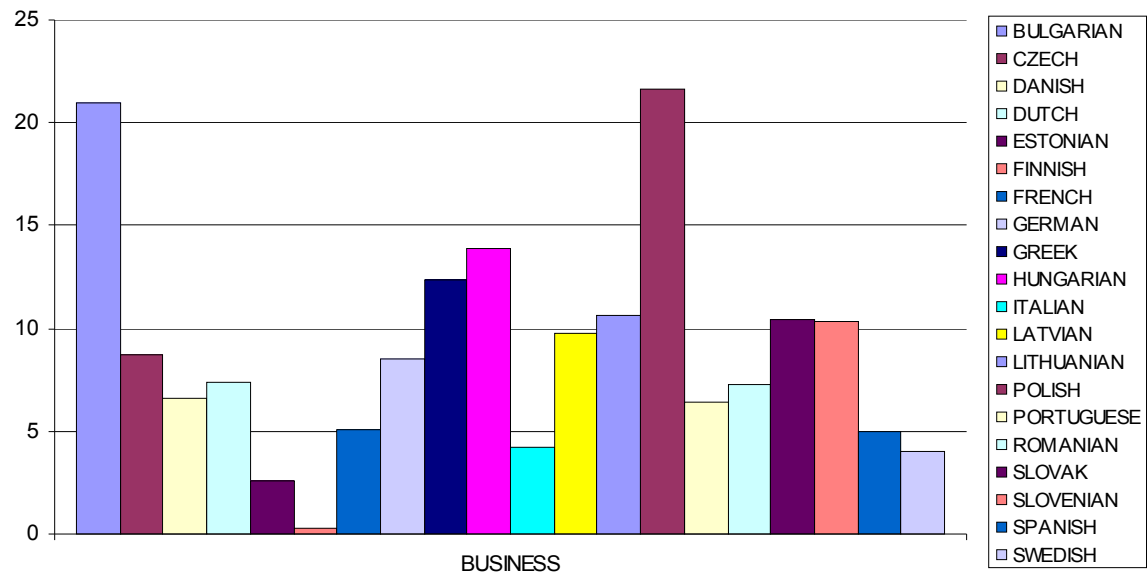
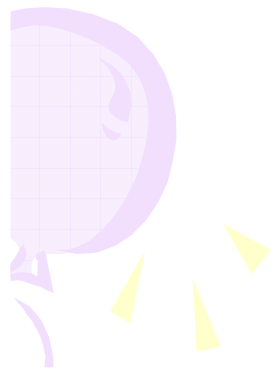
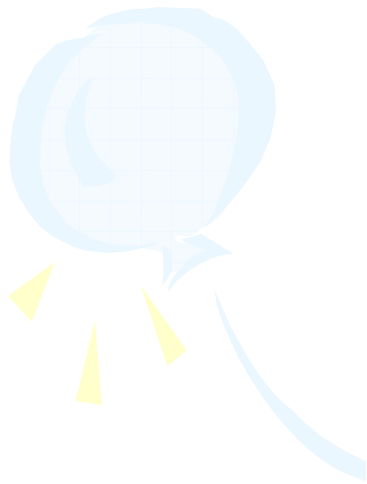
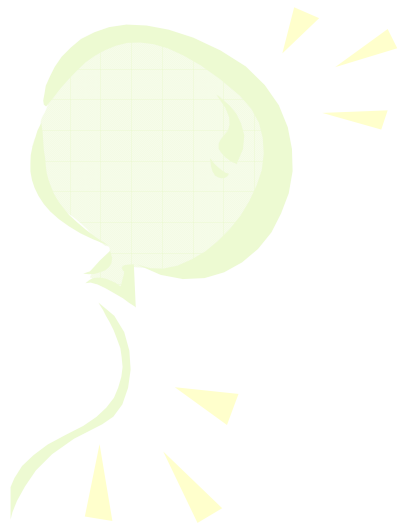
- In the US outlets, in the period april to november 2008, we measured the fraction of newsitems mentioning the Pope (in Content? Just in title-description?)
  - The fraction is
    - 0.102% for english speaking outlets
    - 0.377% for spanish speaking outlets
- 
- 

# Pope-ularity

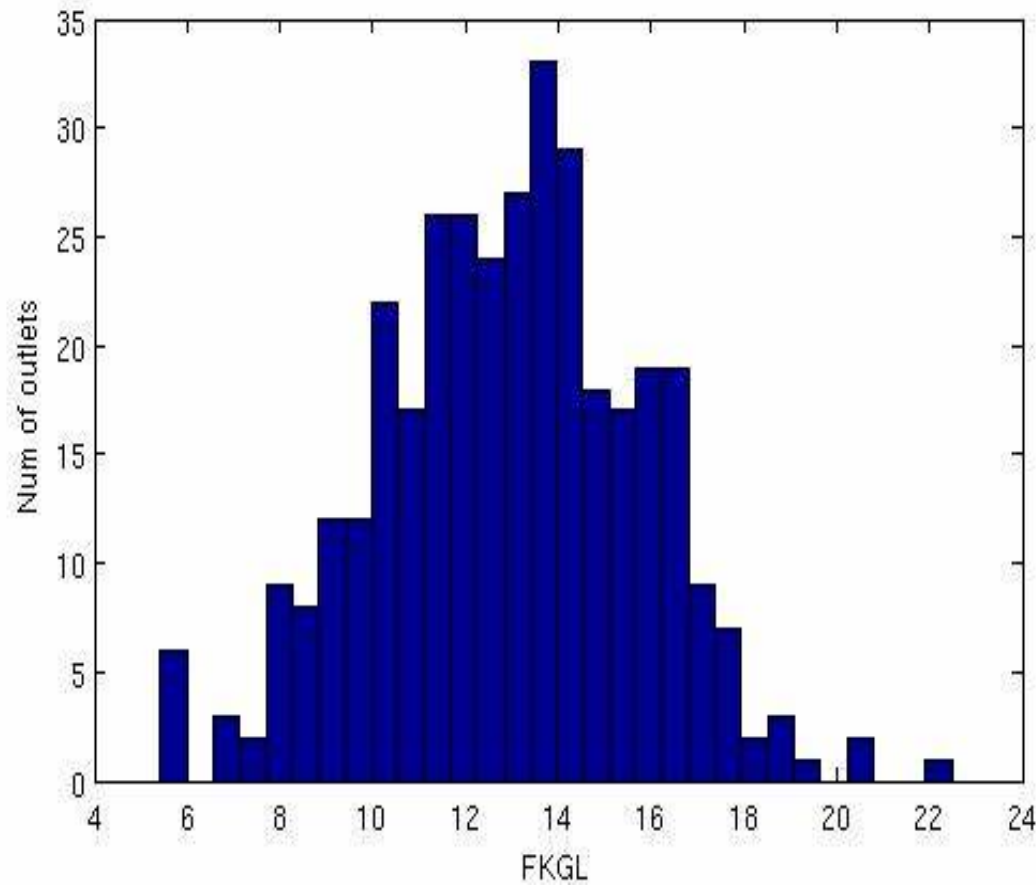
1. 'AFRICA'
2. 'UNITED KINGDOM'
3. 'ASIA'
4. 'CANADA'
5. 'UNITED STATES'
6. 'ITALY'
7. 'FRANCE'
8. 'GERMANY'
9. 'VATICAN'







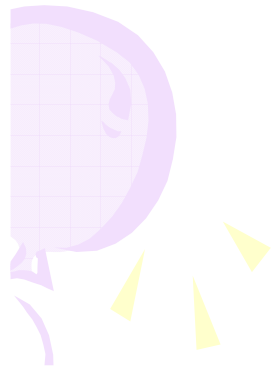
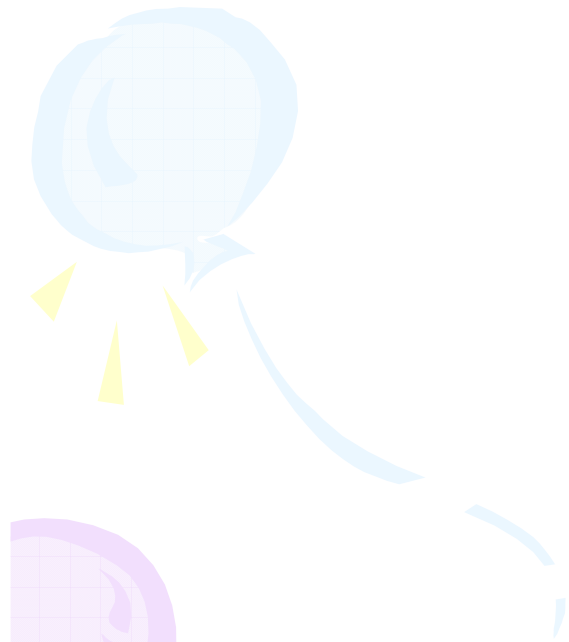
# Readability Index



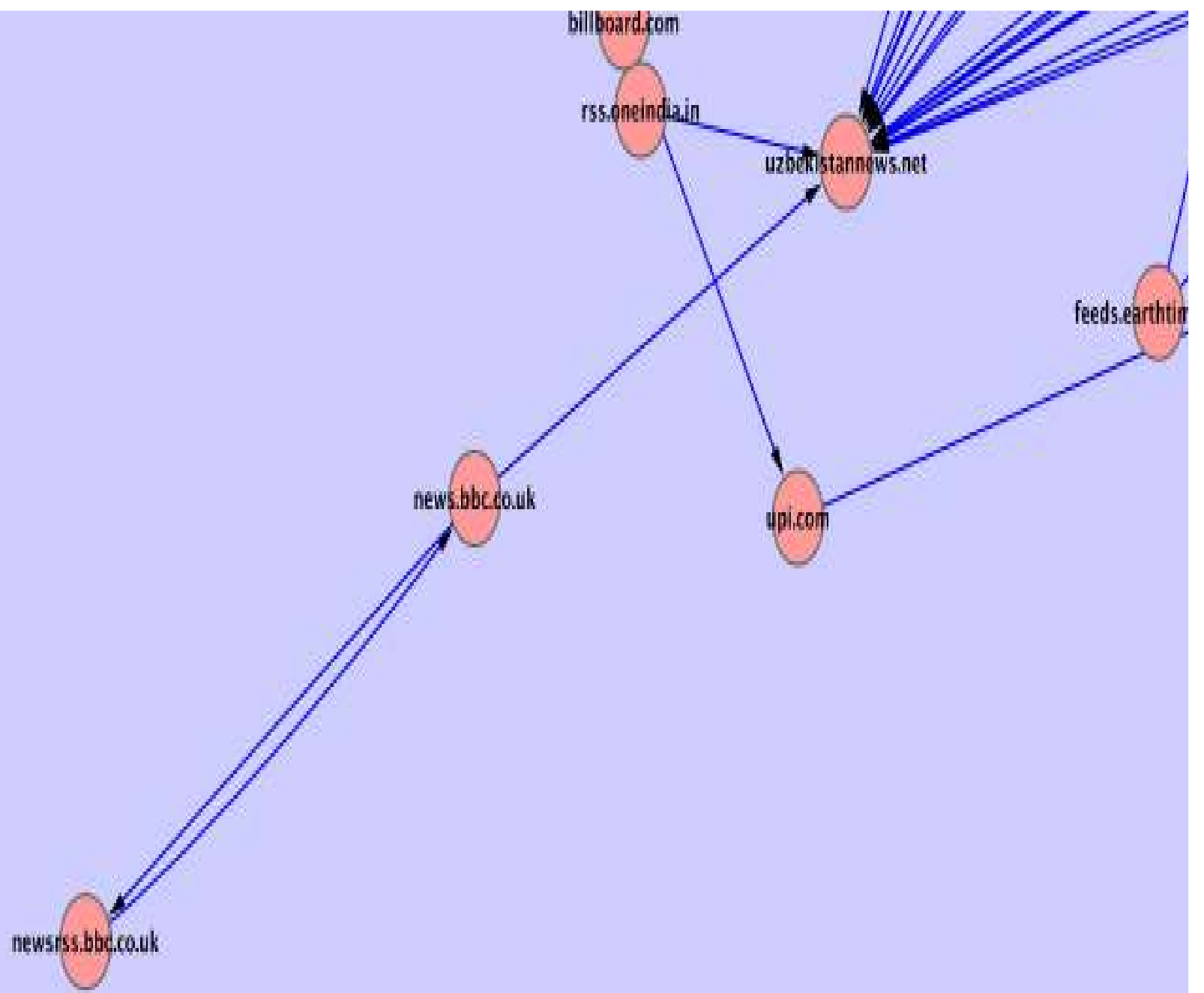
COM	11.97	2.49
EDU	14.47	2.36

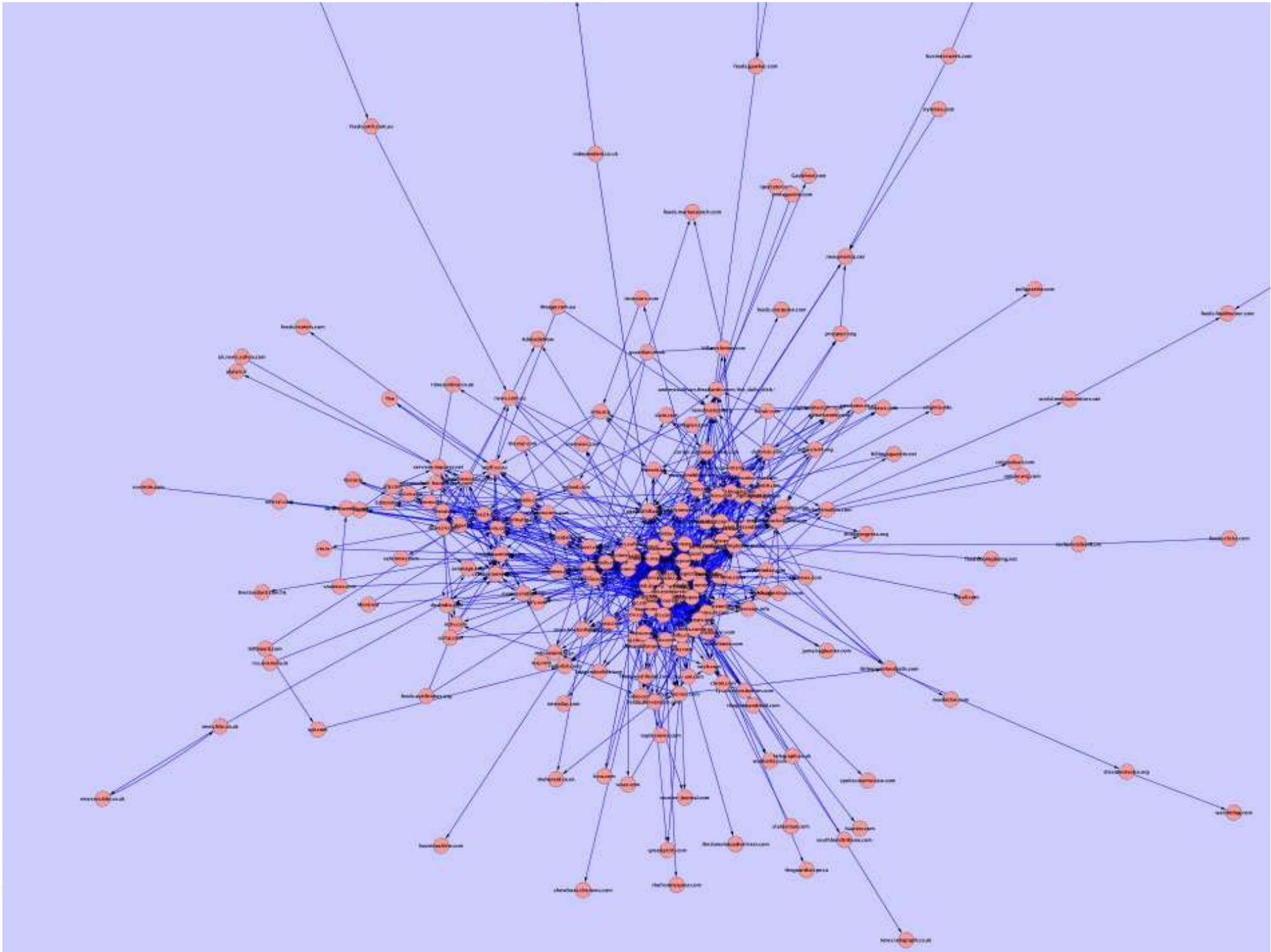


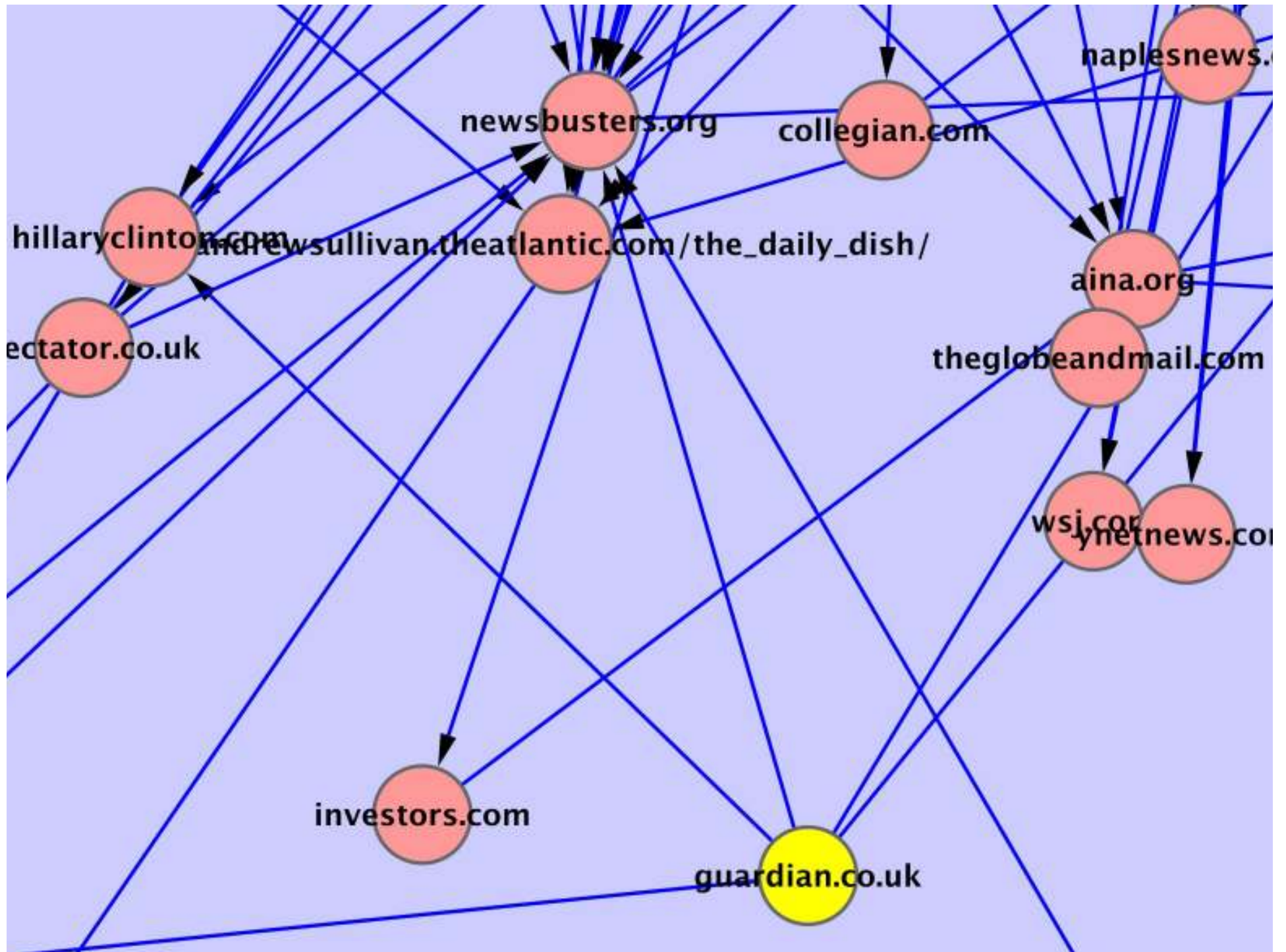
# Text re-Use Networks

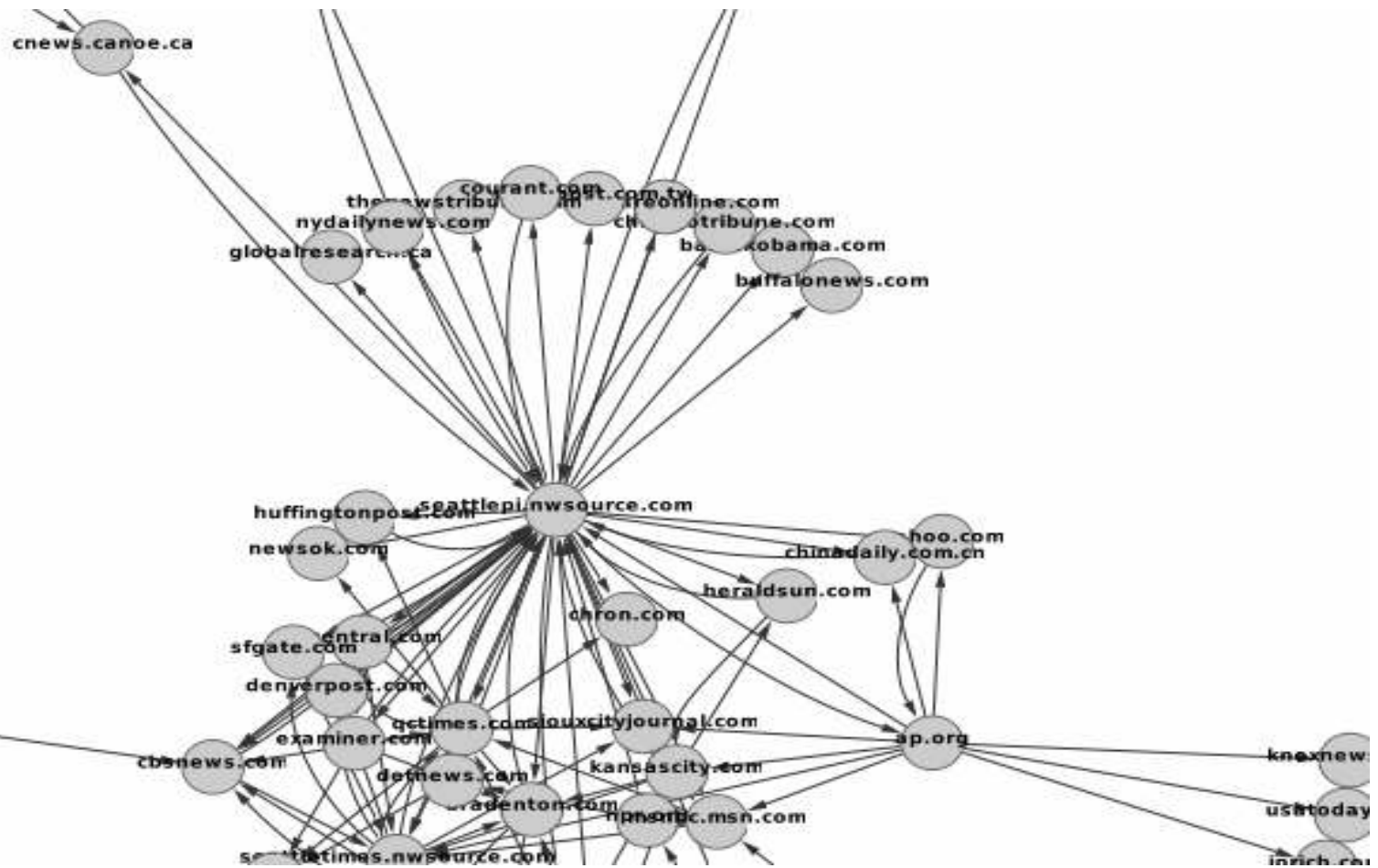
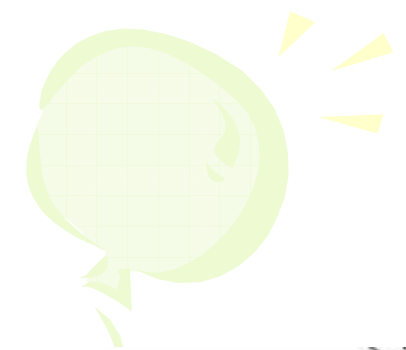




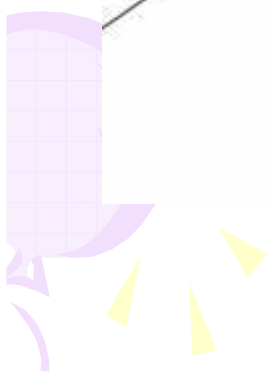
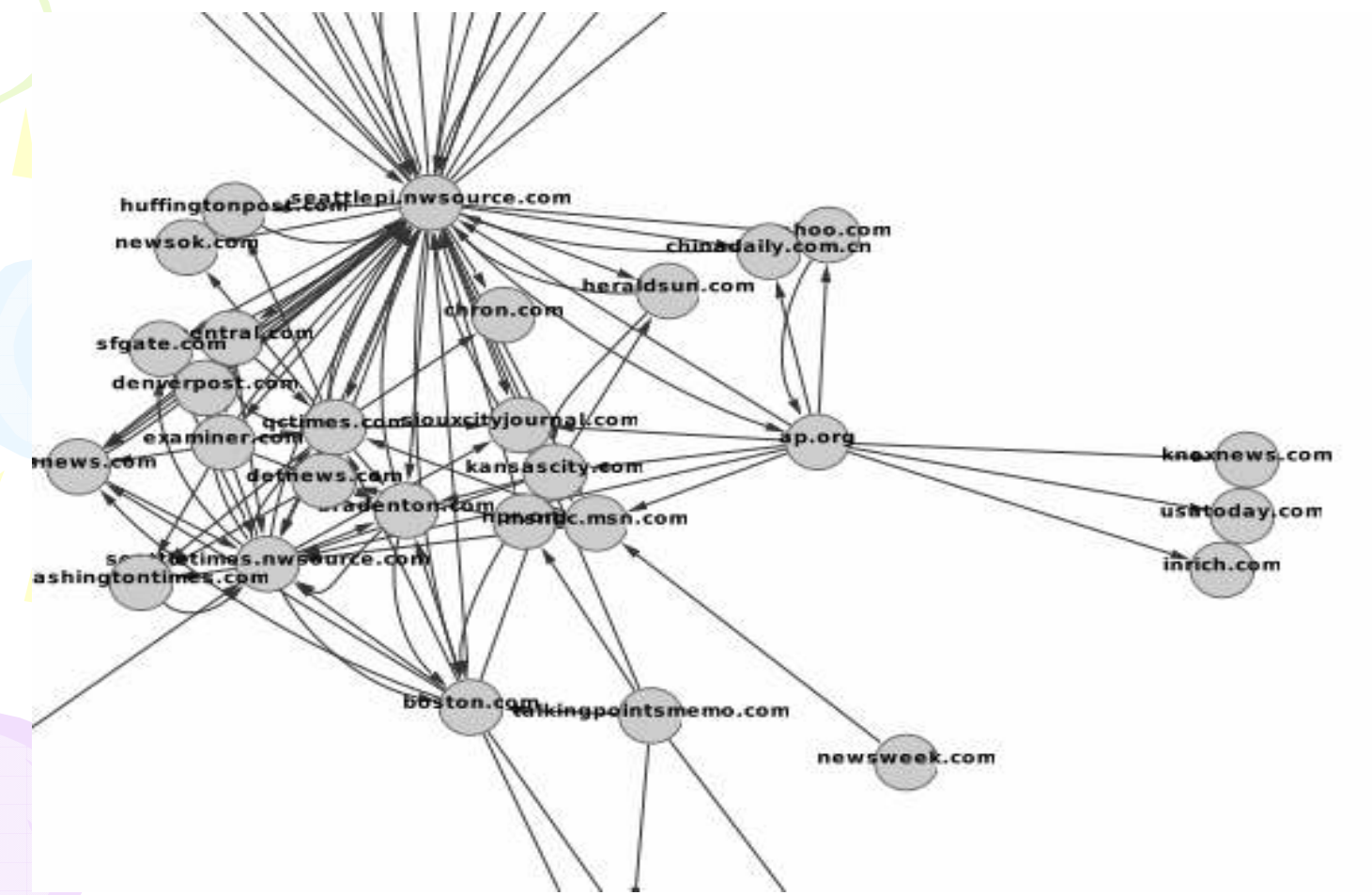
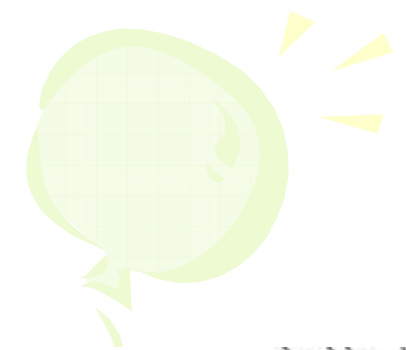






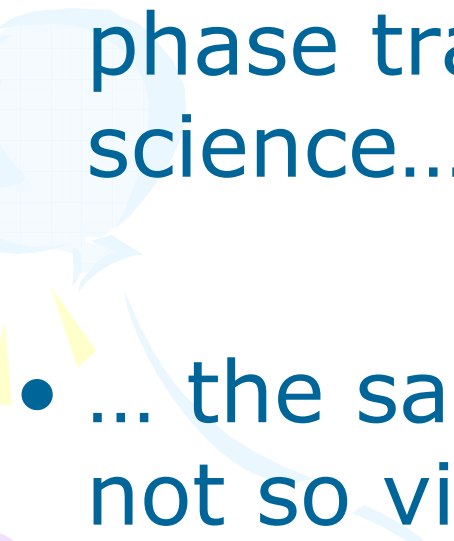
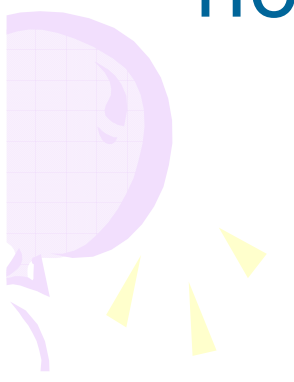








# However...

- As the gathering and analysis of vast amounts of information is enabling a phase transition in the way we do science...
  - ... the same force is behind another – not so visible – phase transition.
- 
- 



# Privacy in the data-driven age...

Surveillance works  
exactly in the same way


...







# Conclusion

- Science is changing (again)
  - Social sciences may benefit enormously from AI technology
  - Much of science can be done automatically
  - Warning: same is happening to other aspects of life
  - Data revolution is moving faster than our philosophical and legal frameworks...
- 
- 