Monitoring the Socio-Economic Climate at a Global Scale

Jorge Louçã

Lisbon University Institute

Seminar "Modeling Complex Socio-Economic Systems and Crises", organized by the ETH Zurich CCSS - Center of Competence for Coping with Crises in Socio Economic Systems

8 March 2011

Summary

I - The Observatorium

- main goal
- collecting vs. analyzing data

II - On-going research

- topic detection in on-line newspapers
- dynamics of scientific trends
- political opinion dynamics
- socio-economic “climate” at a global scale

III - Open source library of tools and free data for computational social science

The Observatorium - main goal

"Intelligence is ten million rules."

Douglas Lenat

The Observatorium - main goal

Goal:

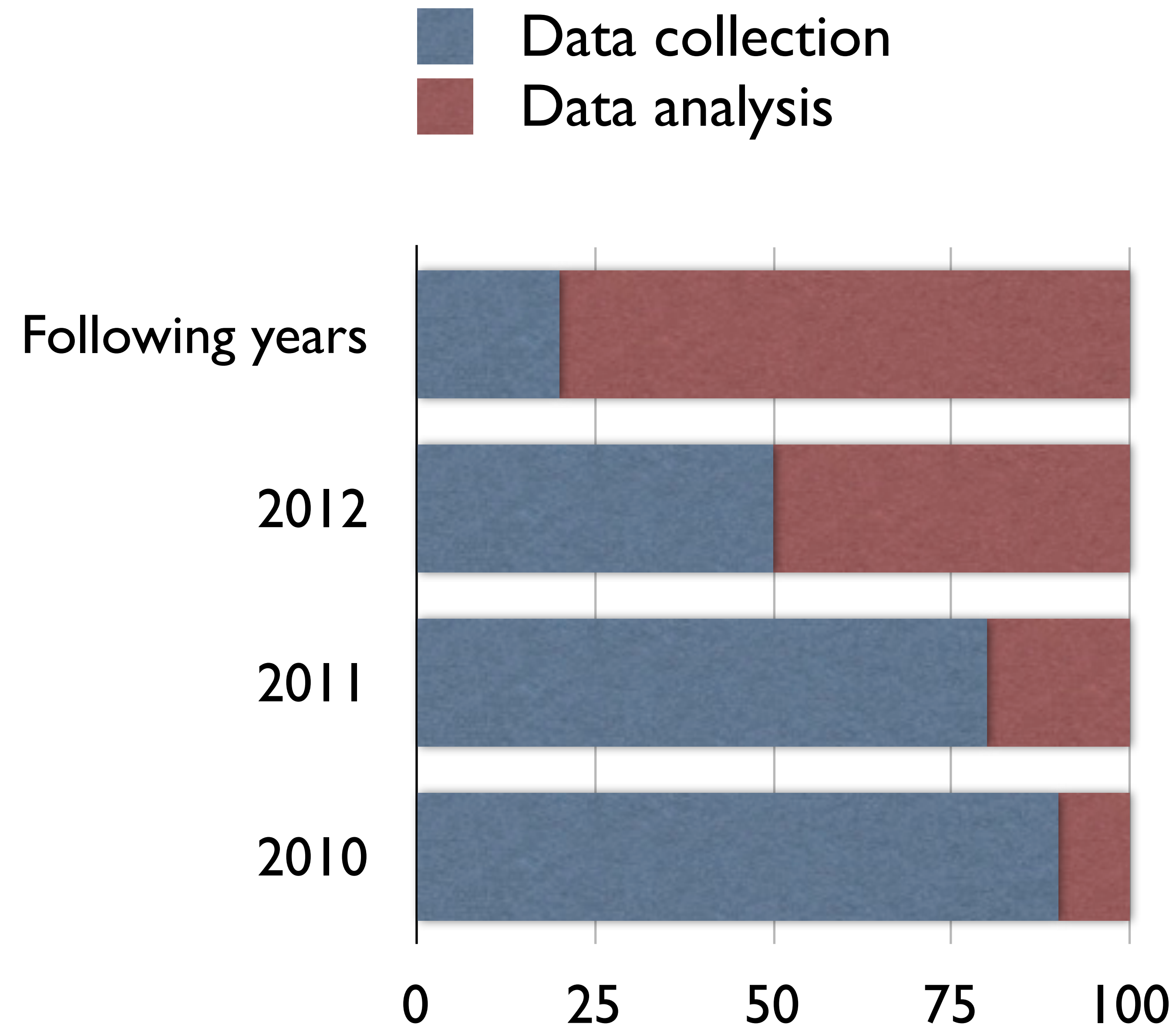
understand more about how and what are people discussing - what are their interests, topics, opinions, arguments, structure of communication

at a global level...

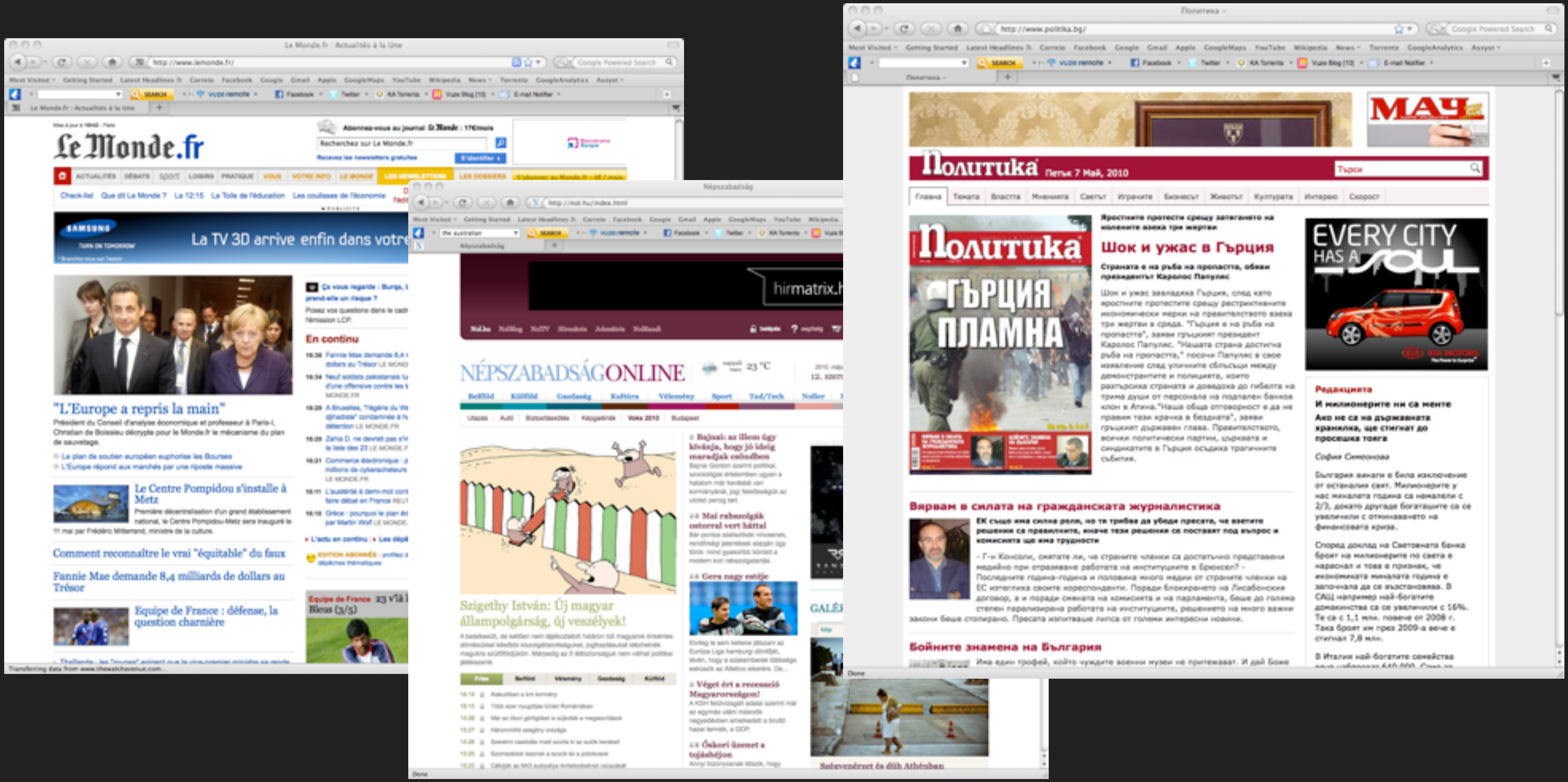
Approach:

data-driven

The Observatorium - collecting vs. analyzing data



On-going research - topic detection in on-line newspapers


Le Monde.fr
La TV 3D arrive enfin dans votre
"L'Europe a repris la main"
En continu
Le Centre Pompidou s'installe à Metz
Comment reconnaître le vrai "équitable" du faux
Fannie Mae demande 8,4 milliards de dollars au Trésor
Equipe de France : défense, la question charnière
NÉPSZABADSÁGONLINE
Szigethy István: Új magyar állampolgárság, új veszélyek!
Политика
Шок и ужас в Гърция
ГЪРЦИЯ ПЛАМНА
Вярвам в силата на гражданската журналистика
Бойните знамена на България
EVERY CITY HAS A SOUL
KIA MOTORS

On Feb 17th: 27,9GB (of text) corresponding to 424.850 news

On-going research - topic detection in on-line newspapers

Frequent approaches: automatic categorisation of text from linguistics, natural language, statistics and information retrieval. Strategies:

- regression models
- Bayesian approaches
- nearest neighbor classification
- neural networks
- hierarchical clustering

(Miao and Qiu, 2010) (Sole et al., 2010)

On-going research - topic detection in on-line newspapers

Frequent approaches: automatic categorisation of text from linguistics, natural language, statistics and information retrieval. Strategies:

- regression models
- Bayesian approaches
- nearest neighbor classification
- neural networks
- hierarchical clustering

(Miao and Qiu, 2010) (Sole et al., 2010)

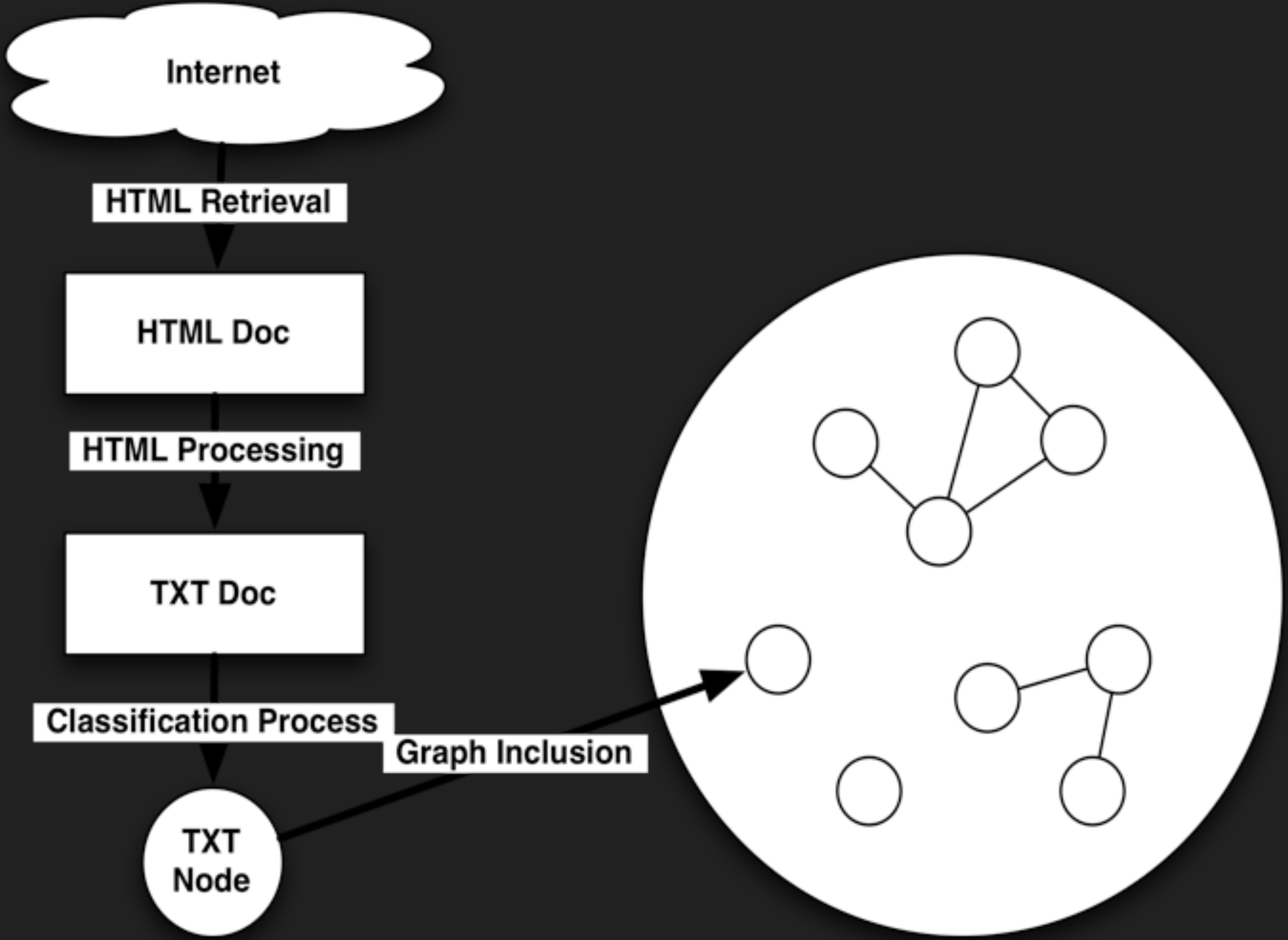
Supervision is needed !

On-going research - topic detection in on-line newspapers

A non-semantic methodology for extraction and real-time analysis of documents from the Internet

On-going research - topic detection in on-line newspapers

A non-semantic methodology for extraction and real-time analysis of documents from the Internet



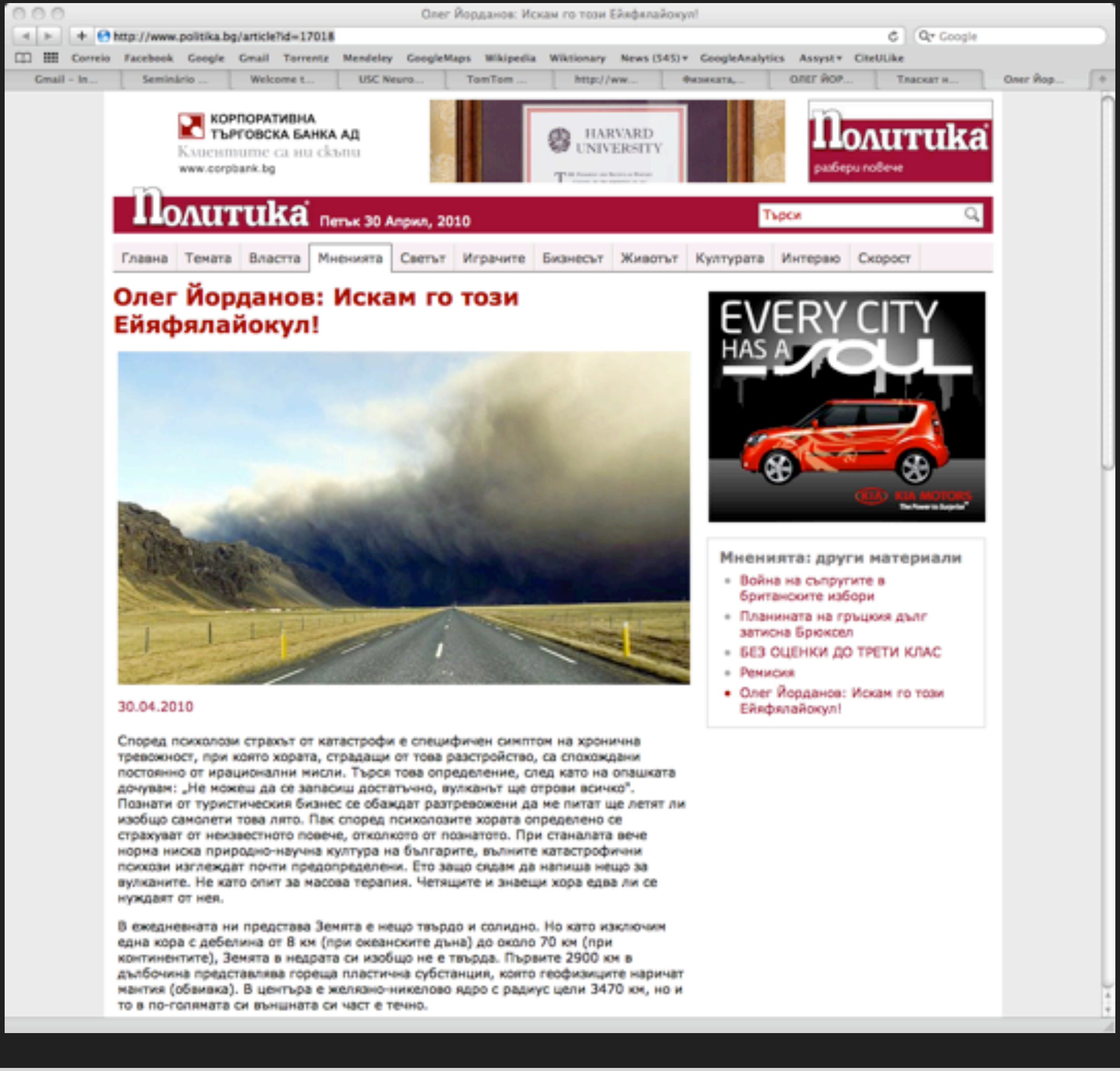
On-going research - topic detection in on-line newspapers

Extraction of the text from each HTML file by employing the Text to Tag ratio (Weninger, 2008)

http://www.politika.bg/article?id=17018

On-going research - topic detection in on-line newspapers

Extraction of the text from each HTML file by employing the Text to Tag ratio (Weninger, 2008)



Олег Йорданов: Искам го този Ейяфялайокул!

30.04.2010

Според психолози страхът от катастрофи е специфичен симптом на хронична тревожност, при която хората, страдащи от това разстройство, са спохождани постоянно от ирационални мисли. Търся това определение, след като на опашката дочувам: „Не можеш да се запасиш достатъчно, вулканът ще отрови всичко". Познати от туристическия бизнес се обаждат разтревожени да ме питат ще летят ли изобщо самолети това лято. Пак според психолозите хората определено се страхуват от неизвестното повече, отколкото от познатото. При станалата вече норма ниска природно-научна култура на българите, вълните катастрофични психози изглеждат почти предопределени. Ето защо сядам да напиша нещо за вулканите. Не като опит за масова терапия. Четящите и знаещи хора едва ли се нуждаят от нея.

В ежедневната ни представа Земята е нещо твърдо и солидно. Но като изключим една кора с дебелина от 8 км (при океанските дъна) до около 70 км (при континентите), Земята в недрата си изобщо не е твърда. Първите 2900 км в дълбочина представлява гореща пластична субстанция, която геофизиците наричат мантия (обвивка). В центъра е желязно-никелово ядро с радиус цели 3470 км, но и то в по-голямата си външната си част е течно.

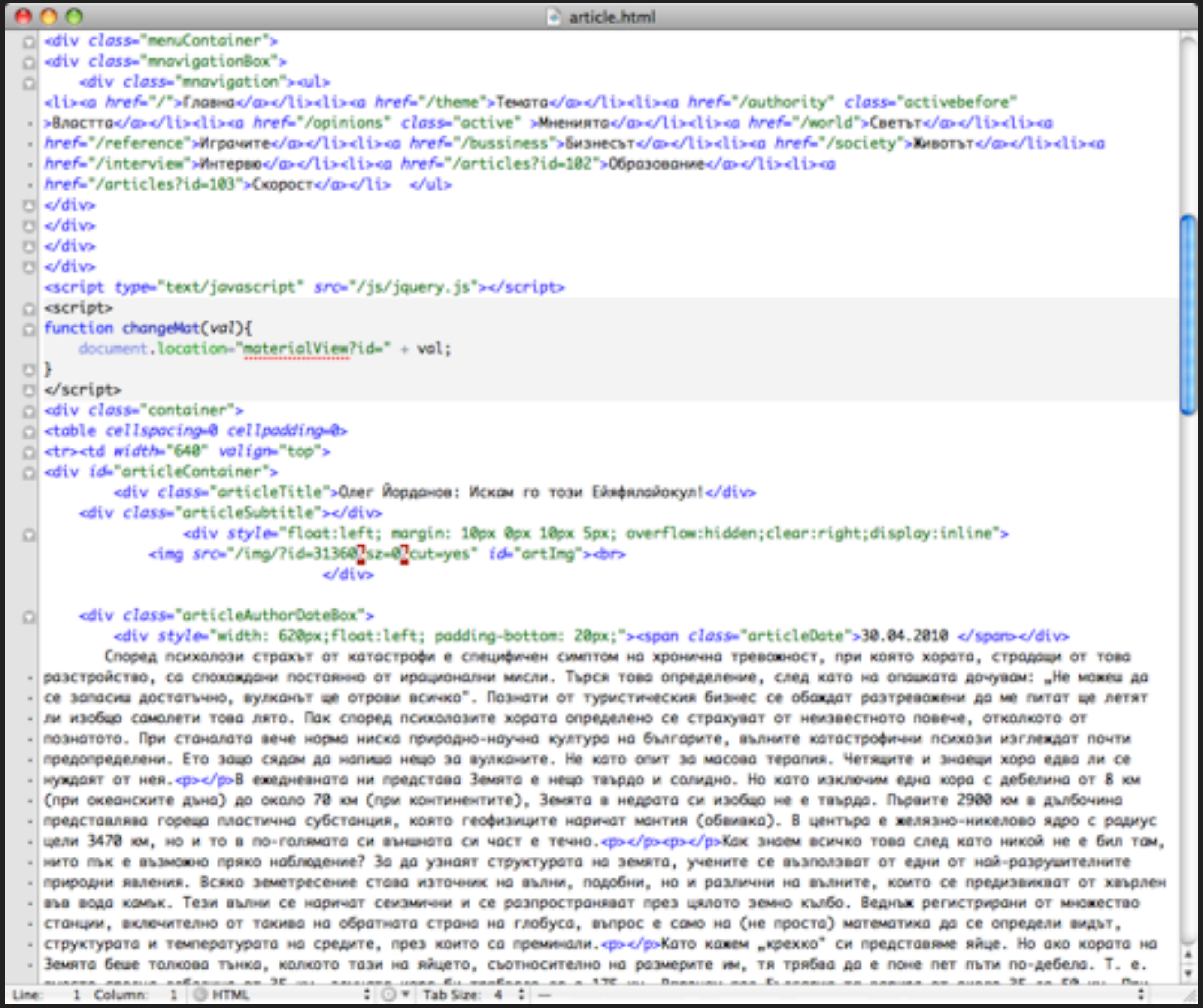
Мненията: други материали

- Война на съпругите в британските избори
- Планината на гръцкия дълг затисна Брюксел
- БЕЗ ОЦЕНКИ ДО ТРЕТИ КЛАС
- Ремисия
- Олег Йорданов: Искам го този Ейяфялайокул!

http://www.politika.bg/article?id=17018

On-going research - topic detection in on-line newspapers

Extraction of the text from each HTML file by employing the Text to Tag ratio (Weninger, 2008)

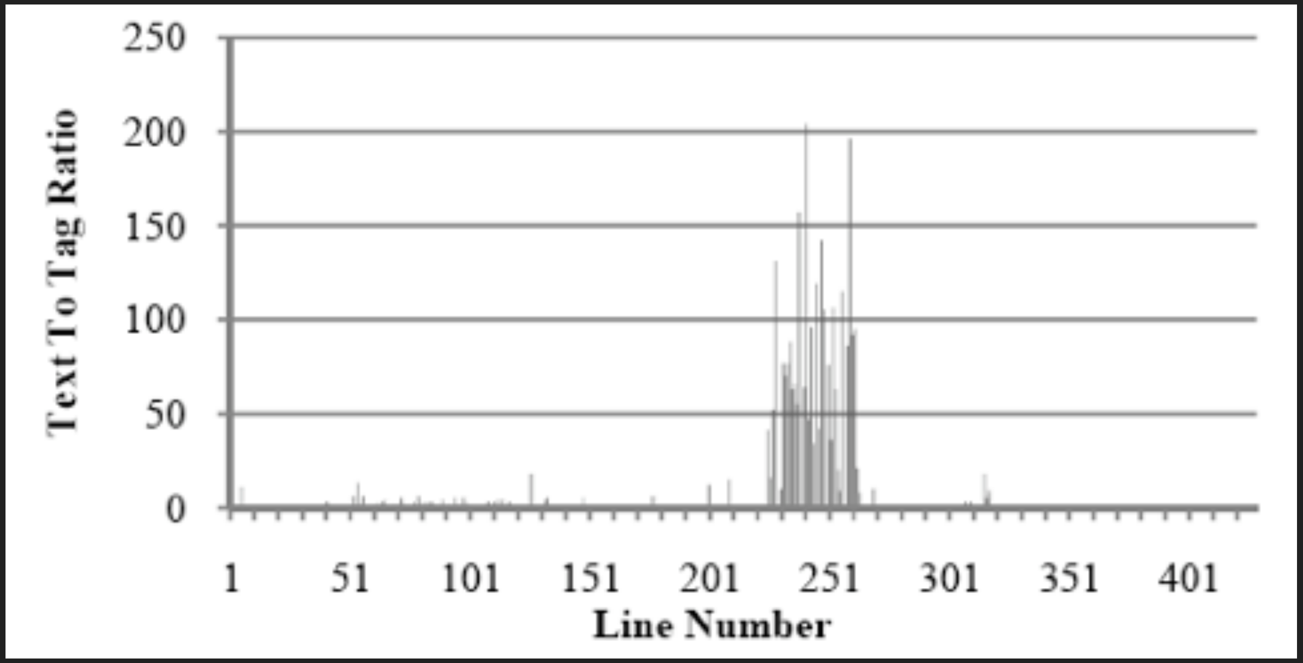


```
<div class="menuContainer">
<div class="mnavigationBox">
    <div class="mnavigation"><ul>
<li><a href="/">Главна</a></li><li><a href="/theme">Темата</a></li><li><a href="/authority" class="activebefore"
>Властта</a></li><li><a href="/opinions" class="active" >Мненията</a></li><li><a href="/world">Светът</a></li><li><a
href="/reference">Играчите</a></li><li><a href="/bussiness">Бизнесът</a></li><li><a href="/society">Животът</a></li><li><a
href="/interview">Интервю</a></li><li><a href="/articles?id=102">Образование</a></li><li><a
href="/articles?id=103">Скорост</a></li>  </ul>
</div>
</div>
</div>
</div>
<script type="text/javascript" src="/js/jquery.js"></script>
<script>
function changeMat(val){
    document.location="materialView?id=" + val;
}
</script>
<div class="container">
<table cellspacing=0 cellpadding=0>
<tr><td width="640" valign="top">
<div id="articleContainer">
        <div class="articleTitle">Олег Йорданов: Искам го този Ейяфялайокул!</div>
    <div class="articleSubtitle"></div>
            <div style="float:left; margin: 10px 0px 10px 5px; overflow:hidden;clear:right;display:inline">
        <img src="/img/?id=31360&sz=0&cut=yes" id="artImg"><br>
                    </div>

    <div class="articleAuthorDateBox">
        <div style="width: 620px;float:left; padding-bottom: 20px;"><span class="articleDate">30.04.2010 </span></div>
       Според психолози страхът от катастрофи е специфичен симптом на хронична тревожност, при която хората, страдащи от това
разстройство, са спохождани постоянно от ирационални мисли. Търся това определение, след като на опашката дочувам: „Не можеш да
се запасиш достатъчно, вулканът ще отрови всичко". Познати от туристическия бизнес се обаждат разтревожени да ме питат ще летят
ли изобщо самолети това лято. Пак според психолозите хората определено се страхуват от неизвестното повече, отколкото от
познатото. При显 станалата вече норма ниска природно-научна култура на българите, вълните катастрофични психози изглеждат почти
предопределени. Ето защо сядам да напиша нещо за вулканите. Не като опит за масова терапия. Четящите и знаещи хора едва ли се
нуждаят от нея.<p></p>В ежедневната ни представа Земята е нещо твърдо и солидно. Но като изключим една кора с дебелина от 8 км
(при океанските дъна) до около 70 км (при континентите), Земята в недрата си изобщо не е твърда. Първите 2900 км в дълбочина
представлява гореща пластична субстанция, която геофизиците наричат мантия (обвивка). В центъра е желязно-никелово ядро с радиус
цели 3470 км, но и то в по-голямата си външната си част е течно.<p></p><p></p>Как знаем всичко това след като никой не е бил там,
нито пък е възможно пряко наблюдение? За да узнаят структурата на земята, учените се възползват от едни от най-разрушителните
природни явления. Всяко земетресение става източник на вълни, подобни, но и различни на вълните, които се предизвикват от хвърлен
във вода камък. Тези вълни се наричат сеизмични и се разпространяват през цялото земно кълбо. Веднъж регистрирани от множество
станции, включително от такива на обратната страна на глобуса, въпрос е само на (не проста) математика да се определи видът,
структурата и температурата на средите, през които са преминали.<p></p>Като кажем „крехко" си представяме яйце. Но ако кората на
Земята беше толкова тънка, колкото тази на яйцето, съотносително на размерите им, тя трябва да е поне пет пъти по-дебела. Т. е.
```

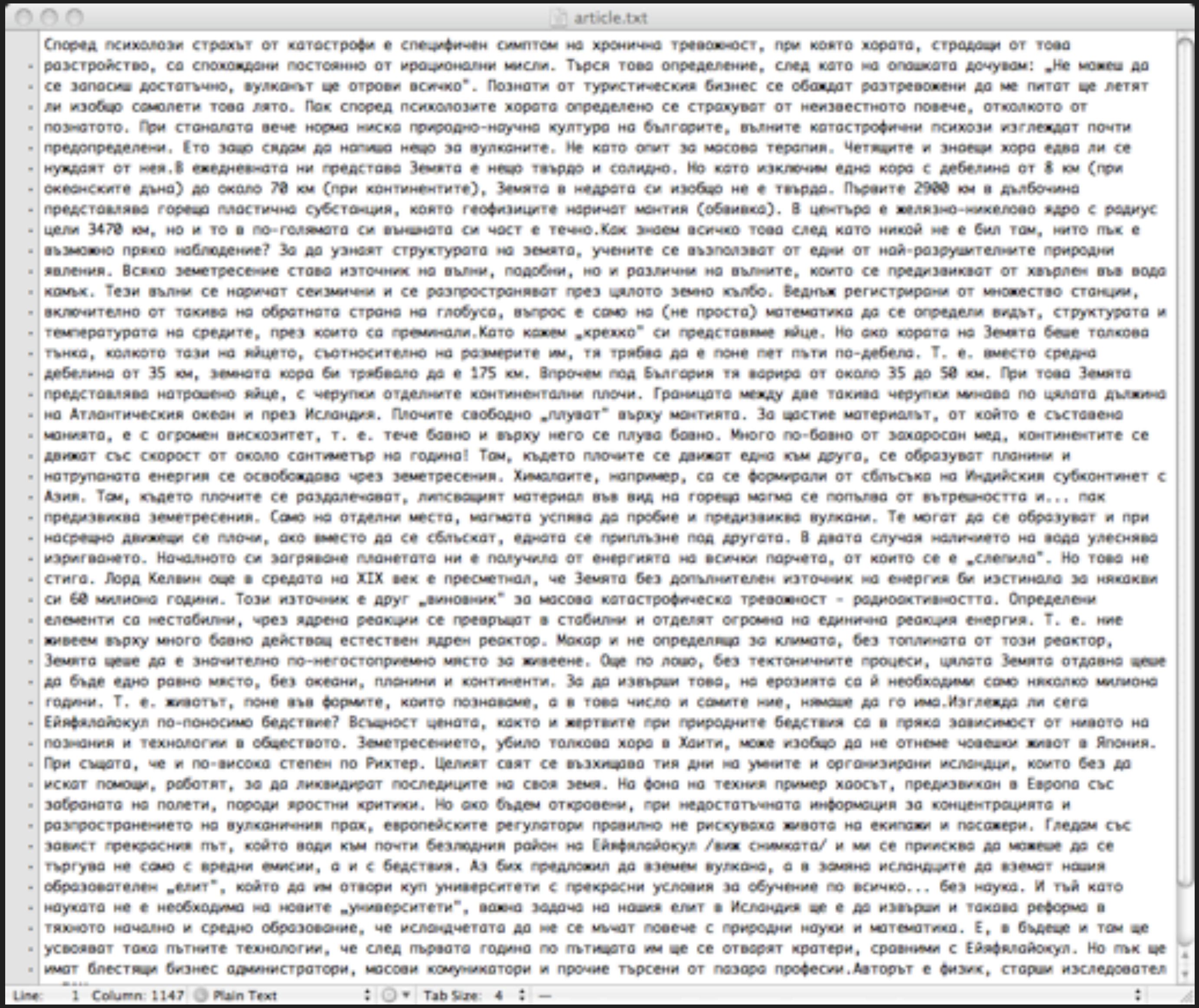
On-going research - topic detection in on-line newspapers

Extraction of the text from each HTML file by employing the Text to Tag ratio (Weninger, 2008)



On-going research - topic detection in on-line newspapers

Extraction of the text from each HTML file by employing the Text to Tag ratio (Weninger, 2008)



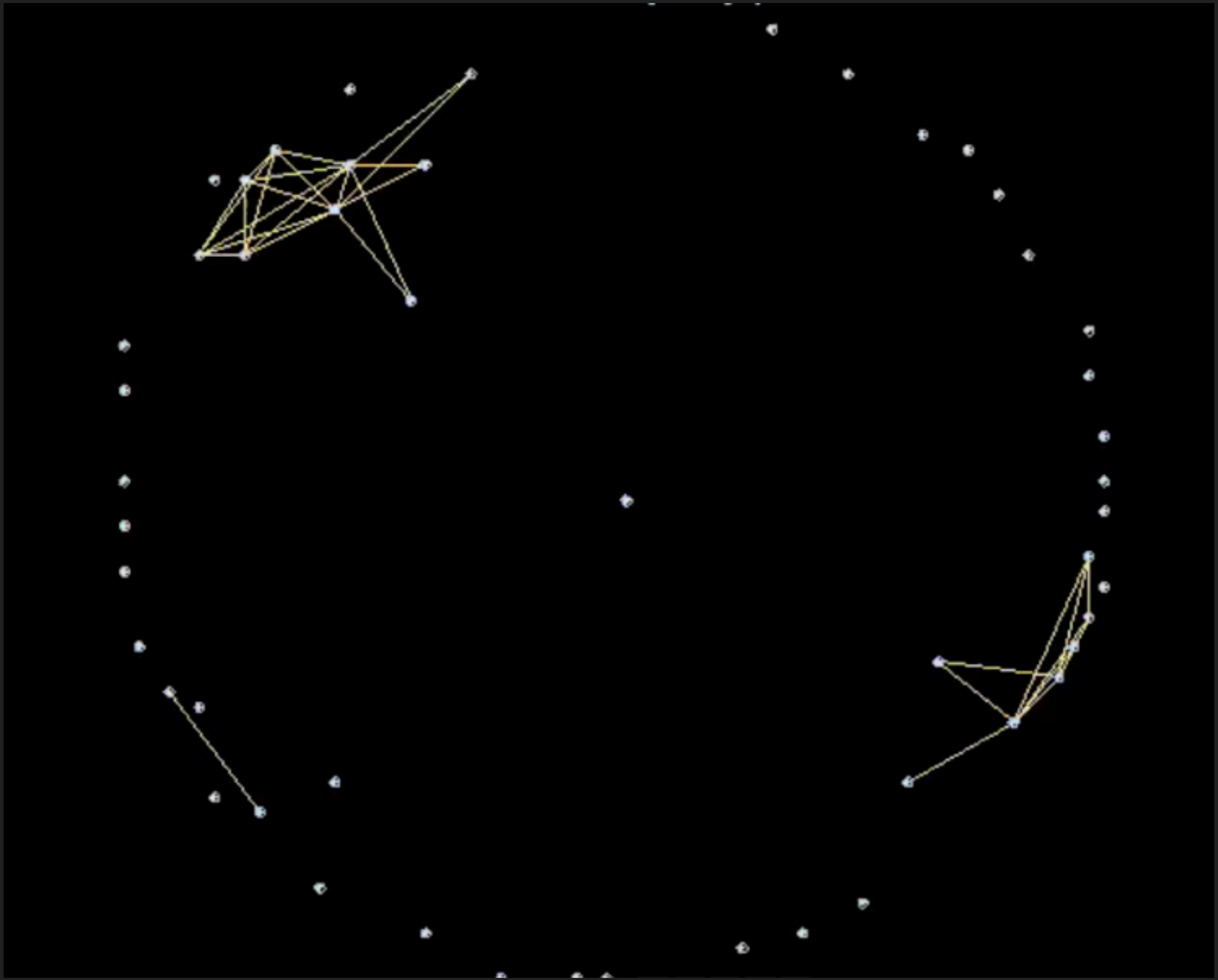
Според психолози страхът от катастрофи е специфичен симптом на хронична тревожност, при която хората, страдащи от това разстройство, са спохождани постоянно от ирационални мисли. Търся това определение, след като на опашката дочувам: „Не можеш да се запасиш достатъчно, вулканът ще отрови всичко". Познати от туристическия бизнес се обаждат разтревожени да ме питат ще летят ли изобщо самолети това лято. Пак според психолозите хората определено се страхуват от неизвестното повече, отколкото от познатото. При станалата вече норма ниска природно-научна култура на българите, вълните катастрофични психози изглеждат почти предопределени. Ето защо сядам да напиша нещо за вулканите. Не като опит за масова терапия. Четящите и знаещи хора едва ли се нуждаят от нея.В ежедневната ни представа Земята е нещо твърдо и солидно. Но като изключим една кора с дебелина от 8 км (при океанските дъна) до около 70 км (при континентите), Земята в недрата си изобщо не е твърда. Първите 2900 км в дълбочина представлява гореща пластична субстанция, която геофизиците наричат мантия (обвивка). В центъра е желязно-никелово ядро с радиус цели 3470 км, но и то в по-голямата си външната си част е течно.Как знаем всичко това след като никой не е бил там, нито пък е възможно пряко наблюдение? За да узнаят структурата на земята, учените се възползват от едни от най-разрушителните природни явления. Всяко земетресение става източник на вълни, подобни, но и различни на вълните, които се предизвикват от хвърлен във вода камък. Тези вълни се наричат сеизмични и се разпространяват през цялото земно кълбо. Веднъж регистрирани от множество станции, включително от такива на обратната страна на глобуса, въпрос е само на (не проста) математика да се определи видът, структурата и температурата на средите, през които са преминали.Като кажем „крехко" си представяме яйце. Но ако кората на Земята беше толкова тънка, колкото тази на яйцето, съотносително на размерите им, тя трябва да е поне пет пъти по-дебела. Т. е. вместо средна дебелина от 35 км, земната кора би трябвало да е 175 км. Впрочем под България тя варира от около 35 до 50 км. При това Земята представлява натрошено яйце, с черупки отделните континентални плочи. Границата между две такива черупки минава по цялата дължина на Атлантическия океан и през Исландия. Плочите свободно „плуват" върху мантията. За щастие материалът, от който е съставена мантията, е с огромен вискозитет, т. е. тече бавно и върху него се плува бавно. Много по-бавно от захаросан мед, континентите се движат със скорост от около сантиметър на година! Там, където плочите се движат една към друга, се образуват планини и натрупаната енергия се освобождава чрез земетресения. Хималаите, например, са се формирали от сблъсъка на Индийския субконтинет с Азия. Там, където плочите се раздалечават, липсващият материал във вид на гореща магма се попълва от вътрешността и... пак предизвиква земетресения. Само на отделни места, магмата успява да пробие и предизвиква вулкани. Те могат да се образуват и при насрещно движещи се плочи, ако вместо да се сблъскат, едната се приплъзне под другата. В двата случая наличието на вода улеснява изригването. Началното си загряване планетата ни е получила от енергията на всички парчета, от които се е „слепила". Но това не стига. Лорд Келвин още в средата на XIX век е пресметнал, че Земята без допълнителен източник на енергия би изстинала за някакви си 60 милиона години. Този източник е друг „виновник" за масова катастрофическа тревожност - радиоактивността. Определени елементи са нестабилни, чрез ядрена реакции се превръщат в стабилни и отделят огромна на единична реакция енергия. Т. е. ние живеем върху много бавно действащ естествен ядрен реактор. Макар и не определящ за климата, без топлината от този реактор, Земята щеше да е значително по-негостоприемно място за живеене. Още по лошо, без тектоничните процеси, цялата Земята отдавна щеше да бъде едно равно място, без океани, планини и континенти. За да извърши това, на ерозията са й необходими само няколко милиона години. Т. е. животът, поне във формите, които познаваме, а в това число и самите ние, нямаше да го има.Изглежда ли сега Ейяфялайокул по-поносимо бедствие? Всъщност цената, както и жертвите при природните бедствия са в пряка зависимост от нивото на познания и технологии в обществото. Земетресението, убило толкова хора в Хаити, може изобщо да не отнеме човешки живот в Япония. При същата, че и по-висока степен по Рихтер. Целият свят се възхищава тия дни на умните и организирани исландци, които без да искат помощи, работят, за да ликвидират последиците на своя земя. На фона на техния пример хаосът, предизвикан в Европа със забраната на полети, породи яростни критики. Но ако бъдем откровени, при недостатъчната информация за концентрацията и разпространението на вулканичния прах, европейските регулатори правилно не рискуваха живота на екипажи и пасажери. Гледам със завист прекрасния път, който води към почти безлюдния район на Ейяфялайокул /виж снимката/ и ми се приисква да можеше да се търгува не само с вредни емисии, а и с бедствия. Аз бих предложил да вземем вулкана, а в замяна исландците да вземат нашия образователен „елит", който да им отвори куп университети с прекрасни условия за обучение по всичко... без наука. И тъй като науката не е необходима на новите „университети", важна задача на нашия елит в Исландия ще е да извърши и такава реформа в тяхното начално и средно образование, че исландчетата да не се мъчат повече с природни науки и математика. Е, в бъдеще и там ще усвояват така пътните технологии, че след първата година по пътищата им ще се отворят кратери, сравними с Ейяфялайокул. Но пък ще имат блестящи бизнес администратори, масови комуникатори и прочие търсени от пазара професии.Авторът е физик, старши изследовател

On-going research - topic detection in on-line newspapers

Graph composition:

- Each news has a set of words, represented by a node
- Each node is added to a graph G
- A node is linked to other nodes according to a distance
- The Jaccard similarity coefficient is used to calculate distances

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$



Problem: real-time monitoring, amount of data

On-going research - topic detection in on-line newspapers

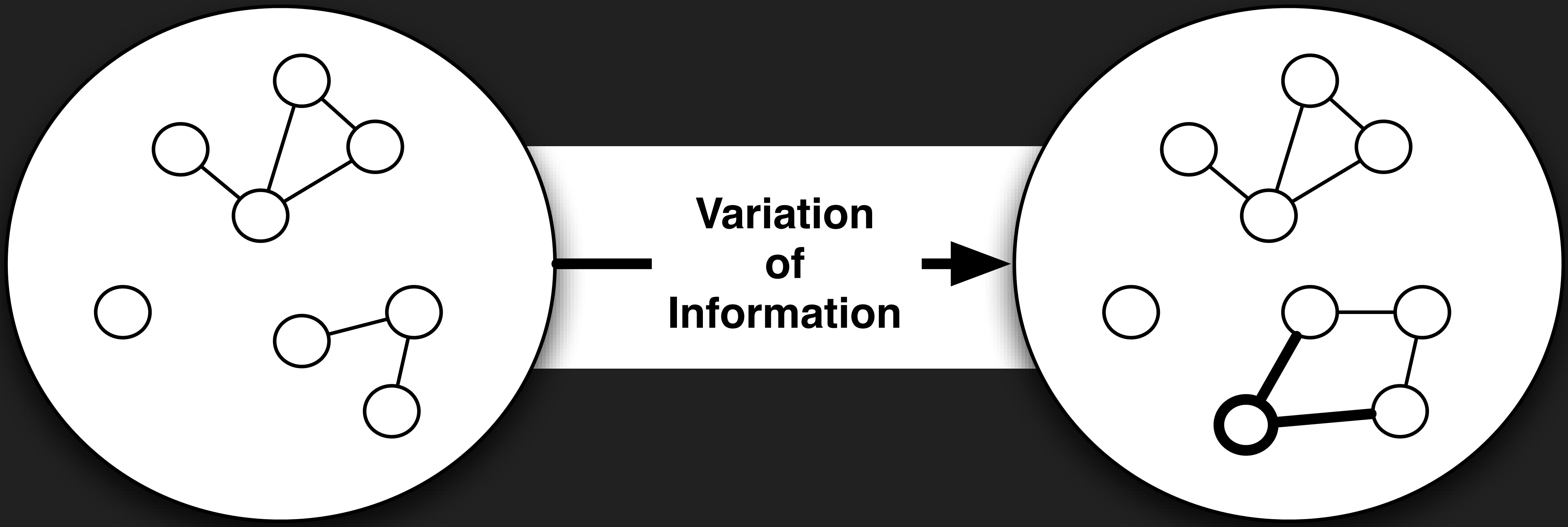
The life expectancy of a node to graph G is given by the *time to live* of this node. Each time we added a node to graph G, the TTL of the nodes that have established connections to it is incremented, and for others this value is decremented.

On-going research - topic detection in on-line newspapers

The variation of information evaluates the distance between different clusters. Variation in information measures the amount of information lost and gained when changing from one version of the graph to a new one.

VI = entropy of the 1st graph + entropy of the 2nd – mutual information of both

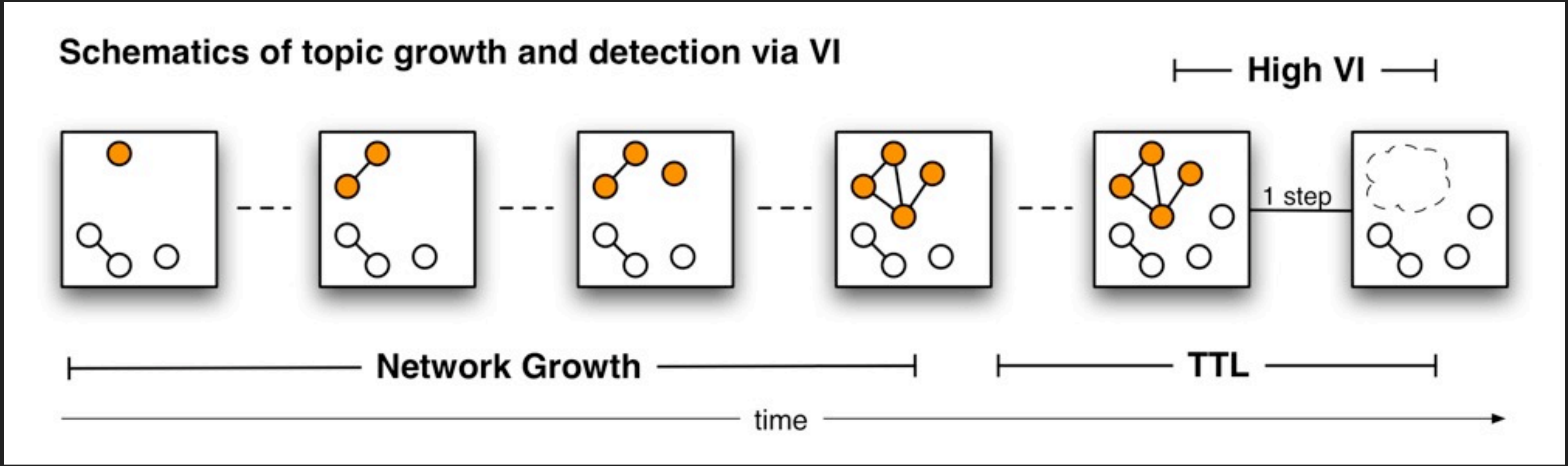
(Meilã, 2007)



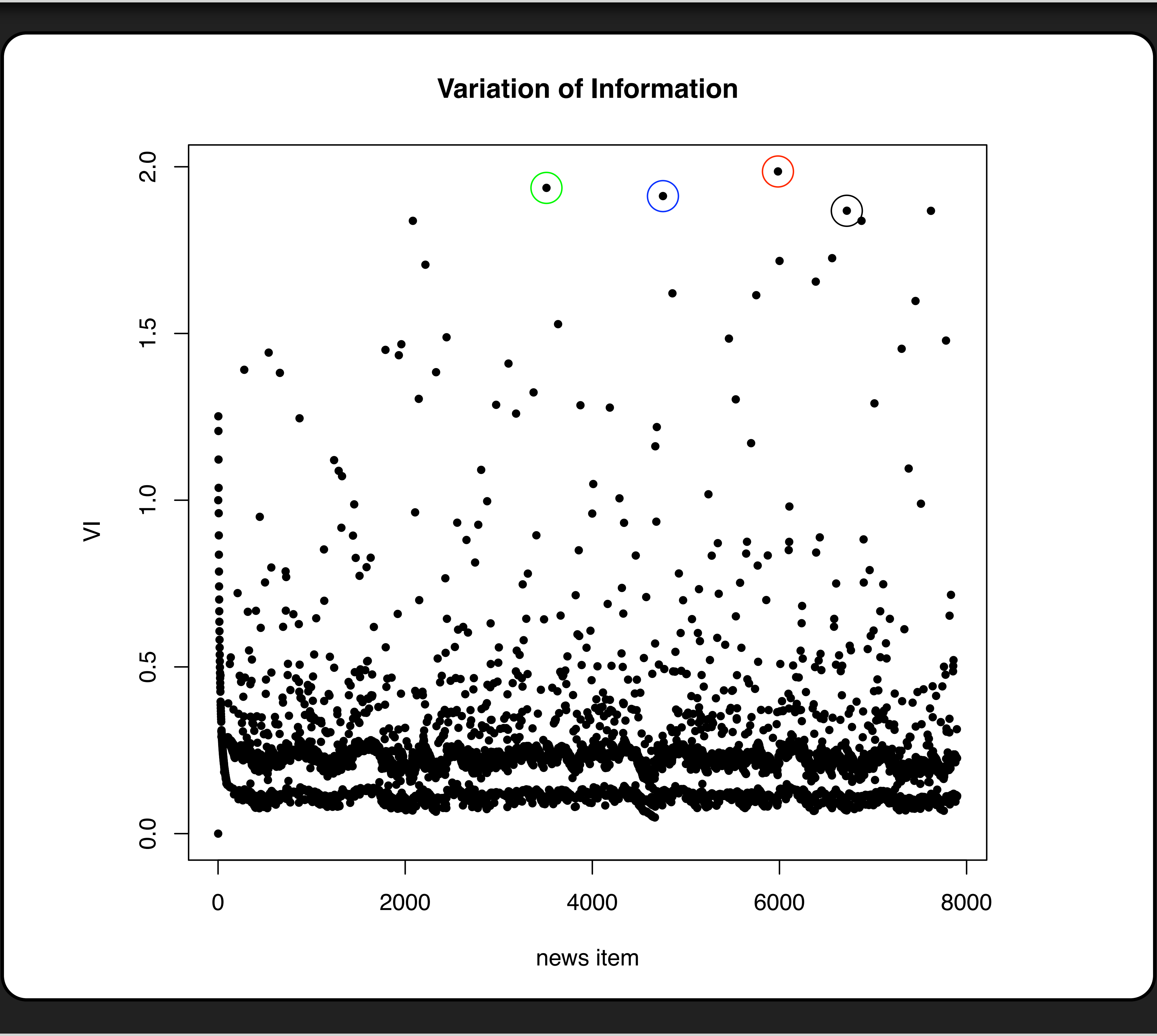
Variation of Information

On-going research - topic detection in on-line newspapers

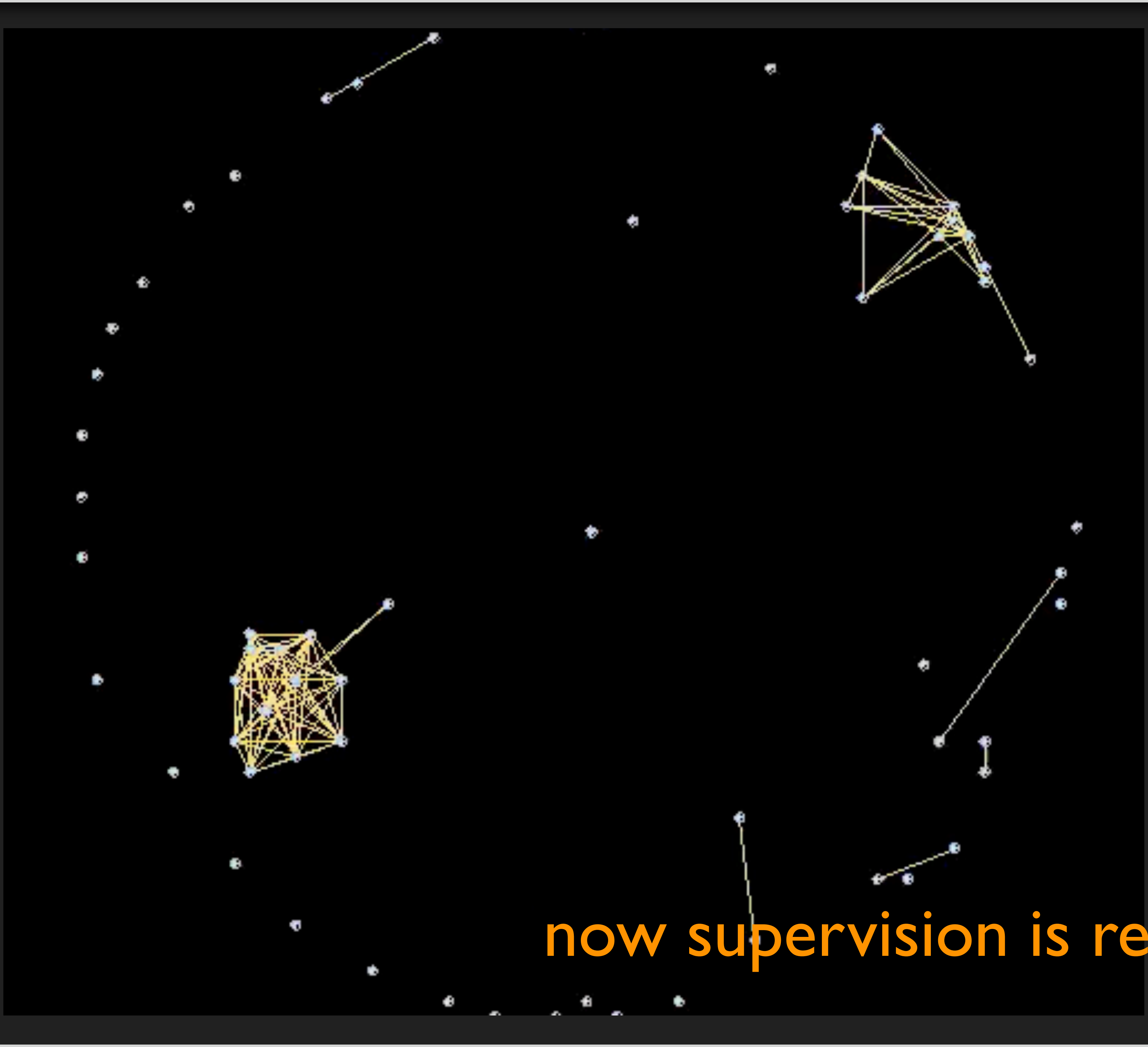
TTL + VI



On-going research - topic detection in on-line newspapers

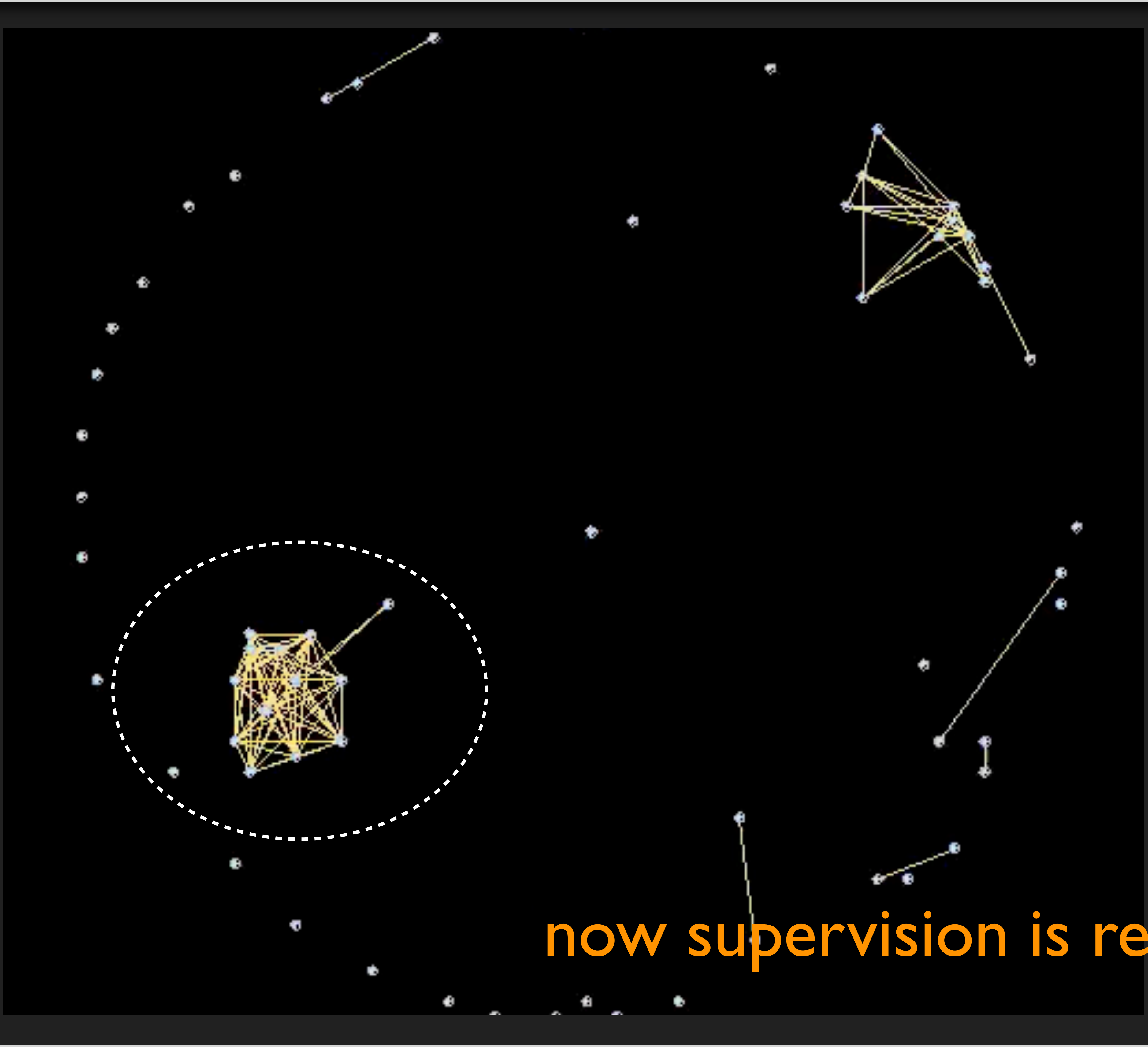


On-going research - topic detection in on-line newspapers



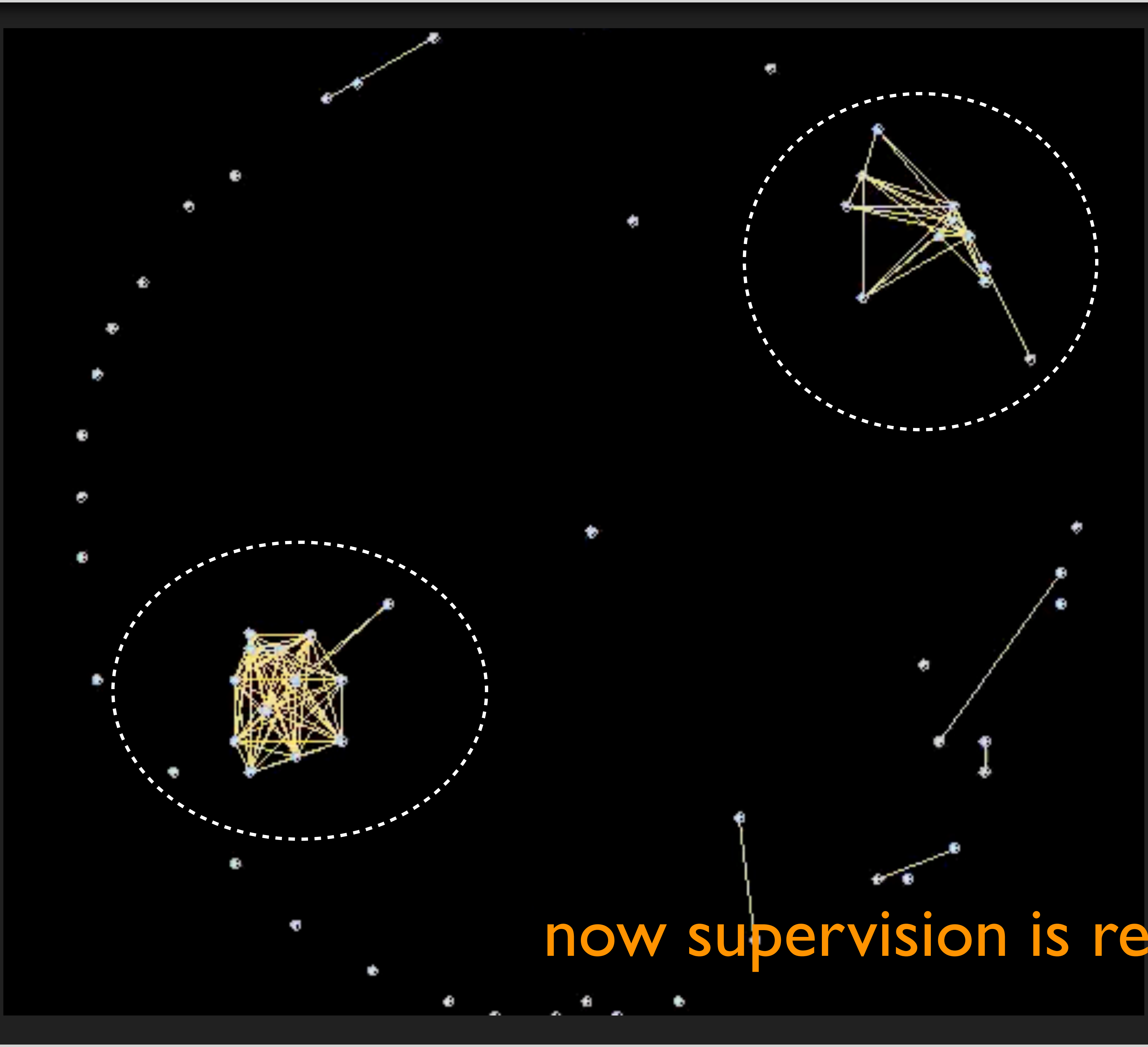
now supervision is required...

On-going research - topic detection in on-line newspapers



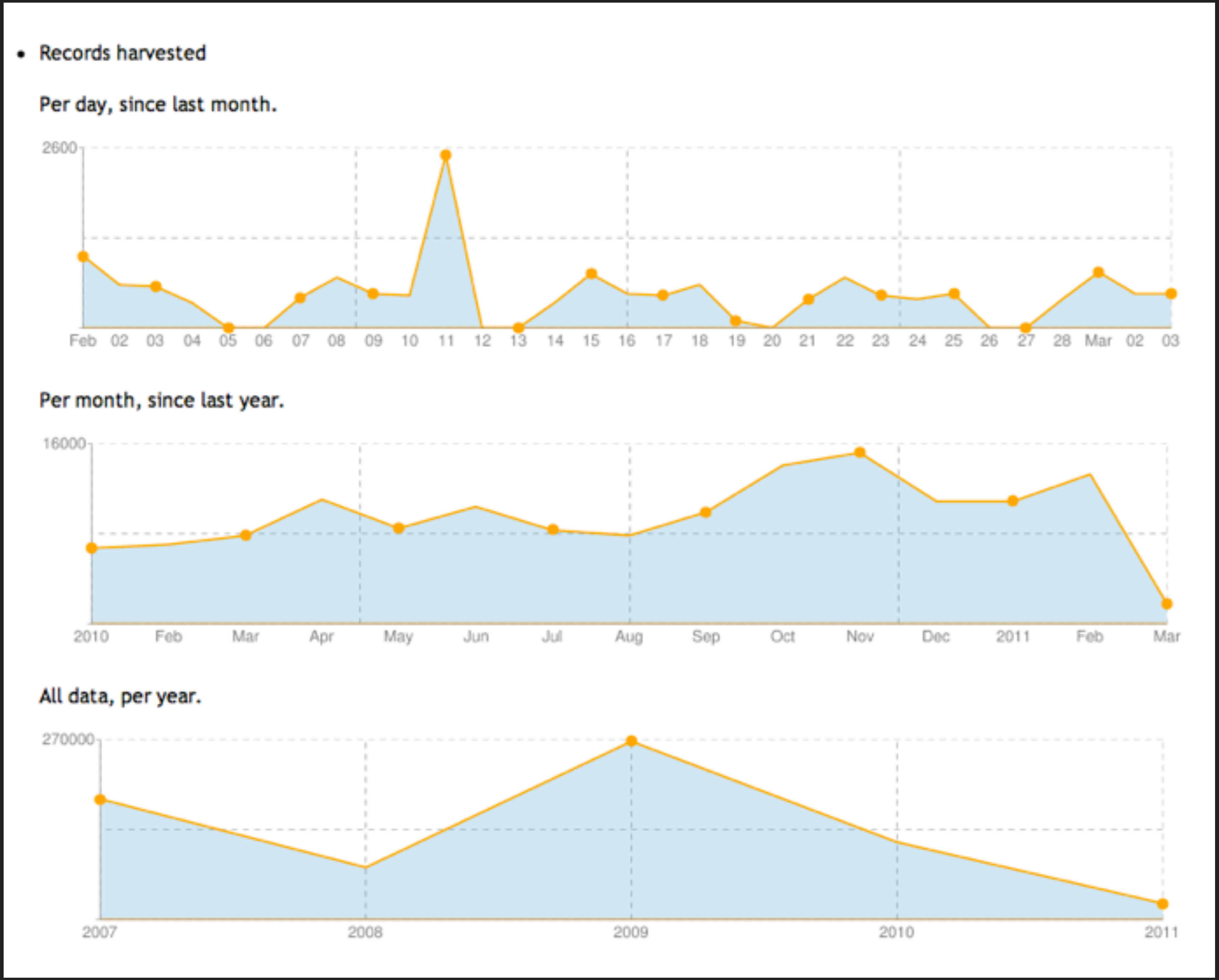
now supervision is required...

On-going research - topic detection in on-line newspapers



now supervision is required...

On-going research - dynamics of scientific trends



- Records harvested

Per day, since last month.

Per month, since last year.

All data, per year.

http://work.theobservatorium.eu/science/

On-going research - political opinion dynamics

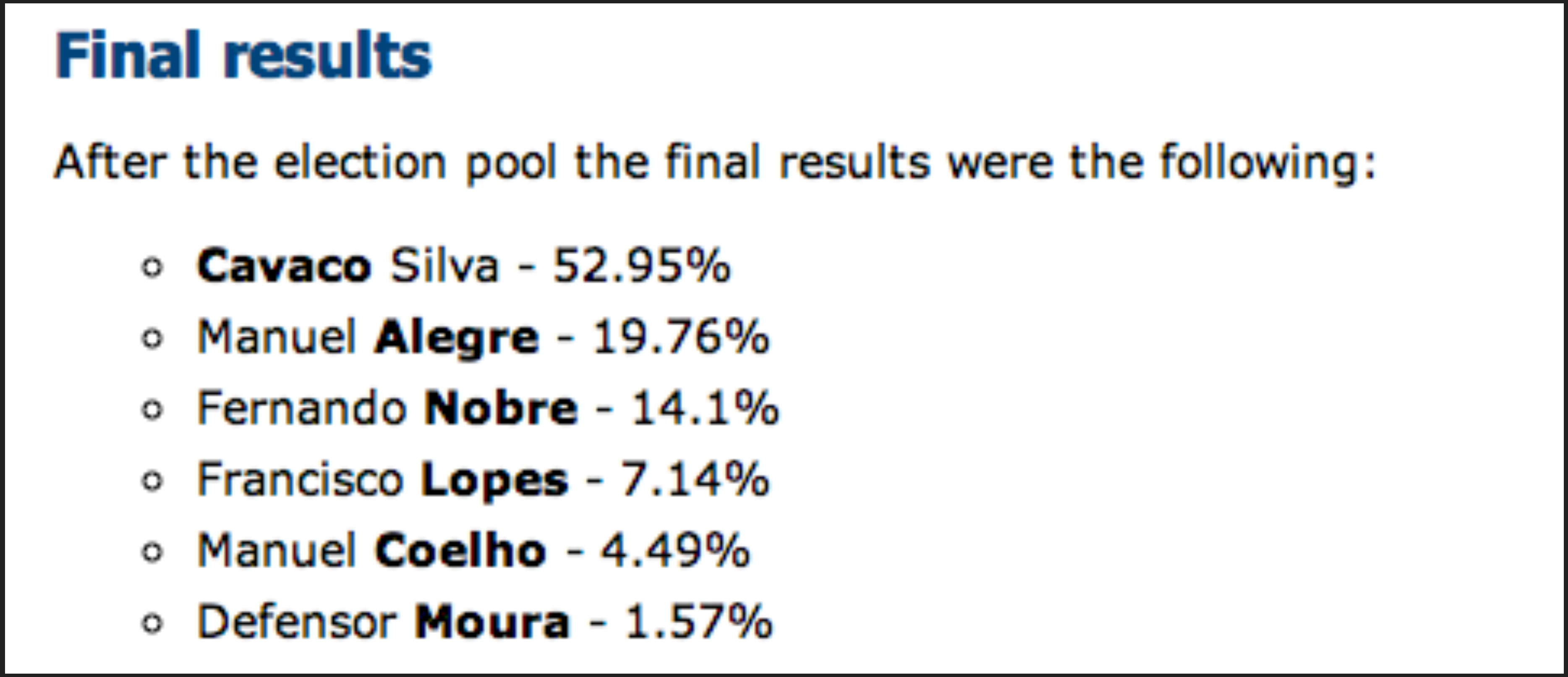
A case study of presidential elections in January 2011

Data collected from the 1st of November 2010 to the 22nd January 2011

Collection of:

- News from a representative set of portuguese media - TV stations; radio stations; newspapers and news agencies.
- Twits
- Facebook

On-going research - political opinion dynamics



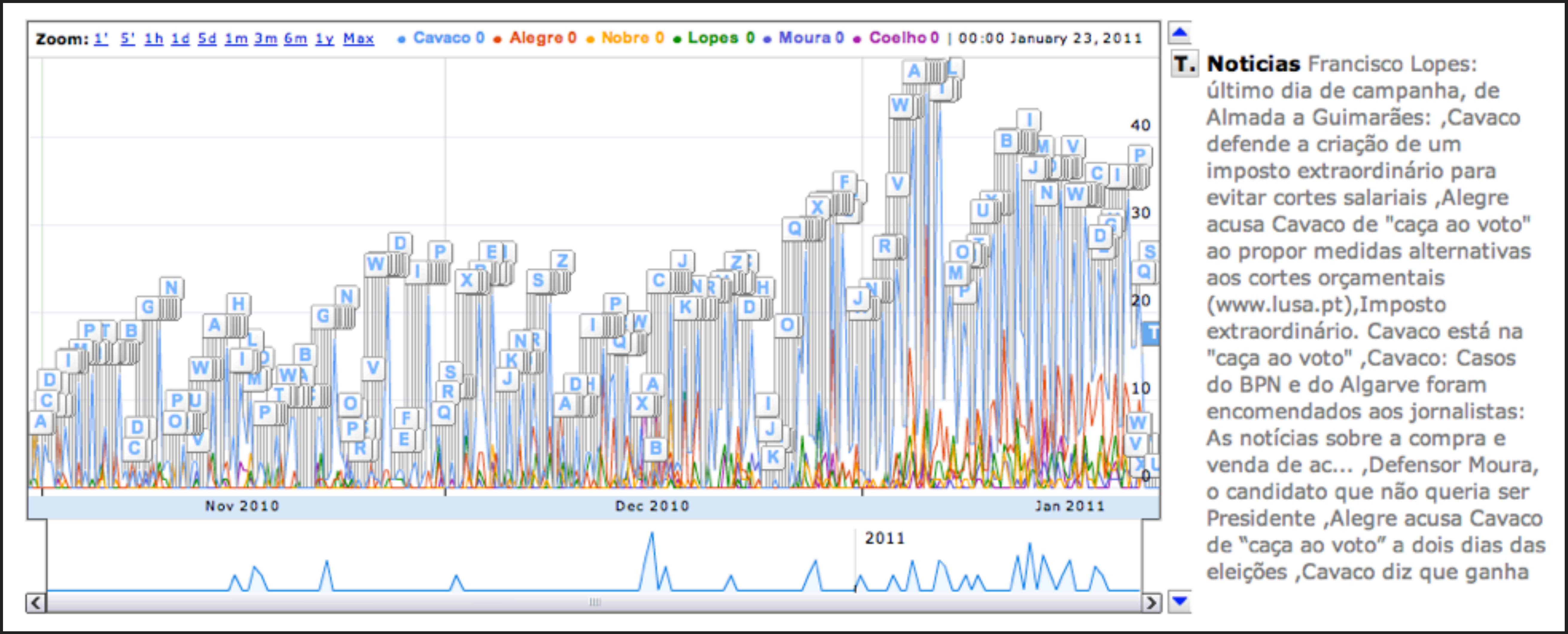
Final results

After the election pool the final results were the following:

- **Cavaco** Silva - 52.95%
- Manuel **Alegre** - 19.76%
- Fernando **Nobre** - 14.1%
- Francisco **Lopes** - 7.14%
- Manuel **Coelho** - 4.49%
- Defensor **Moura** - 1.57%

On-going research - political opinion dynamics

News

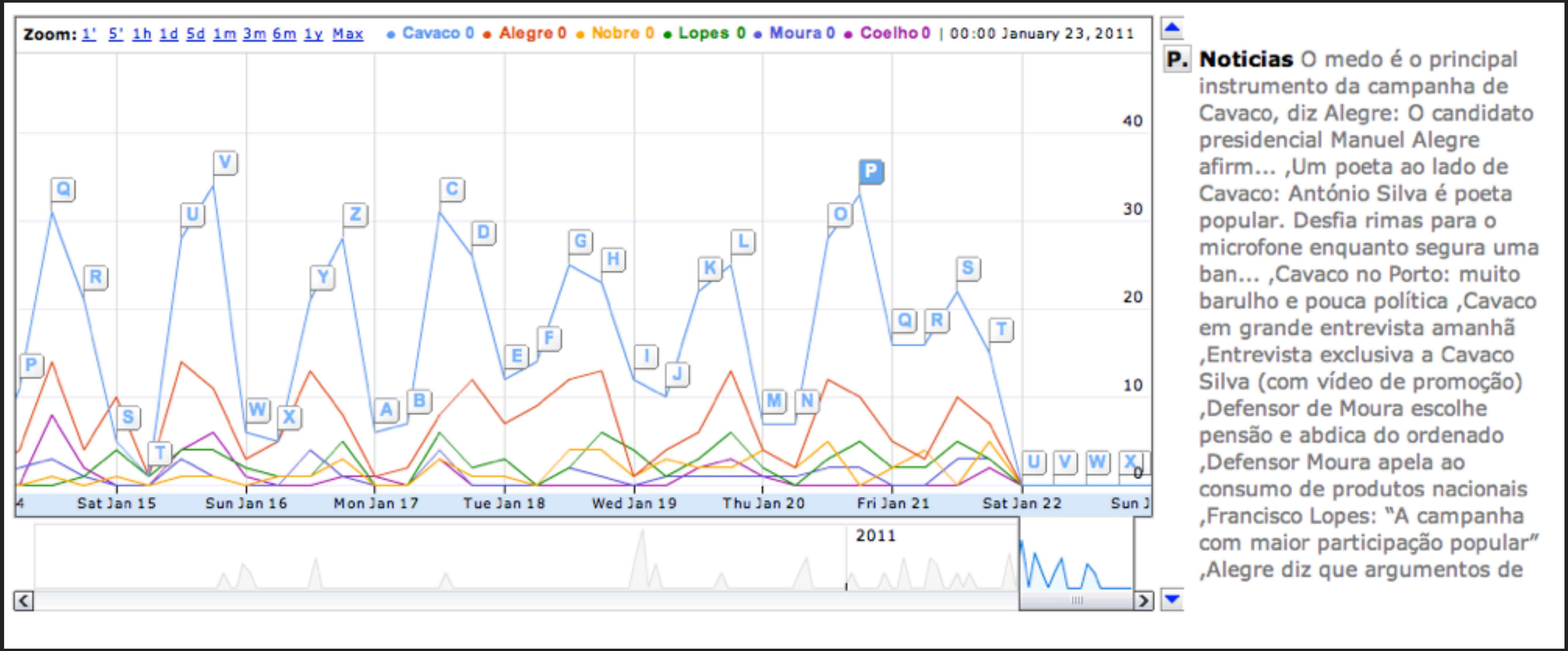


Zoom: 1' 5' 1h 1d 5d 1m 3m 6m 1y Max
● Cavaco 0 ● Alegre 0 ● Nobre 0 ● Lopes 0 ● Moura 0 ● Coelho 0 | 00:00 January 23, 2011
40
30
20
10
0
Nov 2010
Dec 2010
Jan 2011
2011

T. Noticias Francisco Lopes: último dia de campanha, de Almada a Guimarães: ,Cavaco defende a criação de um imposto extraordinário para evitar cortes salariais ,Alegre acusa Cavaco de "caça ao voto" ao propor medidas alternativas aos cortes orçamentais (www.lusa.pt),Imposto extraordinário. Cavaco está na "caça ao voto" ,Cavaco: Casos do BPN e do Algarve foram encomendados aos jornalistas: As notícias sobre a compra e venda de ac... ,Defensor Moura, o candidato que não queria ser Presidente ,Alegre acusa Cavaco de "caça ao voto" a dois dias das eleições ,Cavaco diz que ganha

On-going research - political opinion dynamics

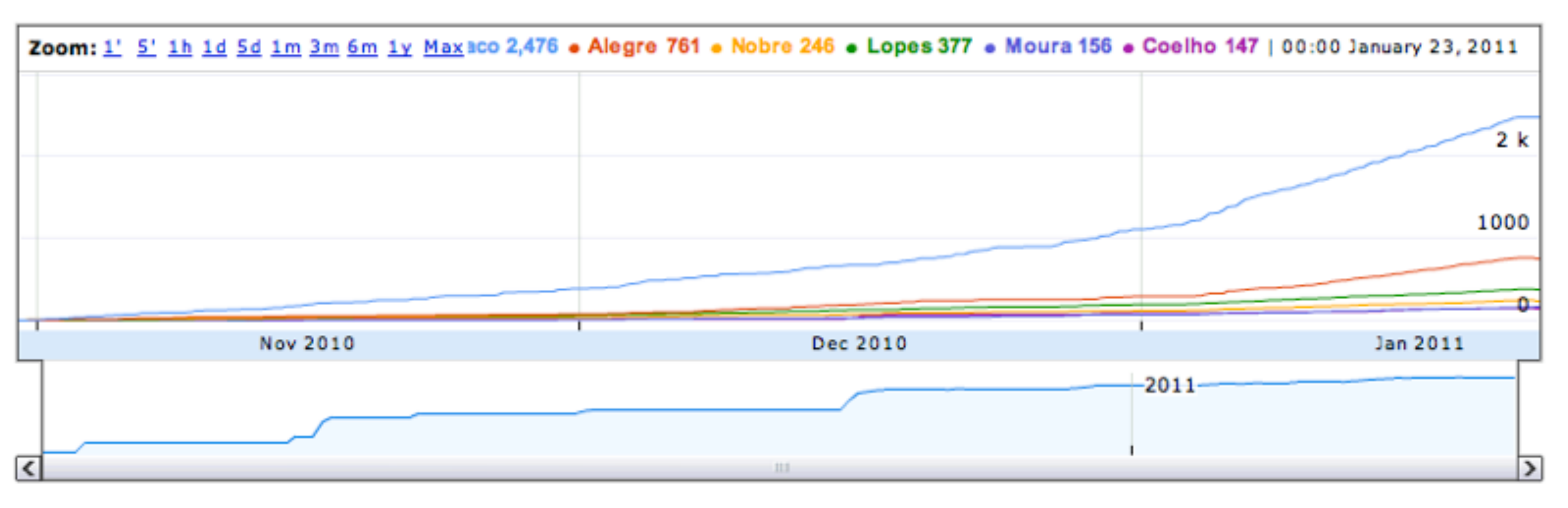
News


Zoom: 1' 5' 1h 1d 5d 1m 3m 6m 1y Max
● Cavaco 0 ● Alegre 0 ● Nobre 0 ● Lopes 0 ● Moura 0 ● Coelho 0 | 00:00 January 23, 2011
40
30
20
10
0
P Q R S T U V W X Y Z A B C D E F G H I J K L M N O P Q R S T U V W X
4
Sat Jan 15
Sun Jan 16
Mon Jan 17
Tue Jan 18
Wed Jan 19
Thu Jan 20
Fri Jan 21
Sat Jan 22
Sun J
2011
P.

Noticias O medo é o principal instrumento da campanha de Cavaco, diz Alegre: O candidato presidencial Manuel Alegre afirm... ,Um poeta ao lado de Cavaco: António Silva é poeta popular. Desfia rimas para o microfone enquanto segura uma ban... ,Cavaco no Porto: muito barulho e pouca política ,Cavaco em grande entrevista amanhã ,Entrevista exclusiva a Cavaco Silva (com vídeo de promoção) ,Defensor de Moura escolhe pensão e abdica do ordenado ,Defensor Moura apela ao consumo de produtos nacionais ,Francisco Lopes: "A campanha com maior participação popular" ,Alegre diz que argumentos de

On-going research - political opinion dynamics

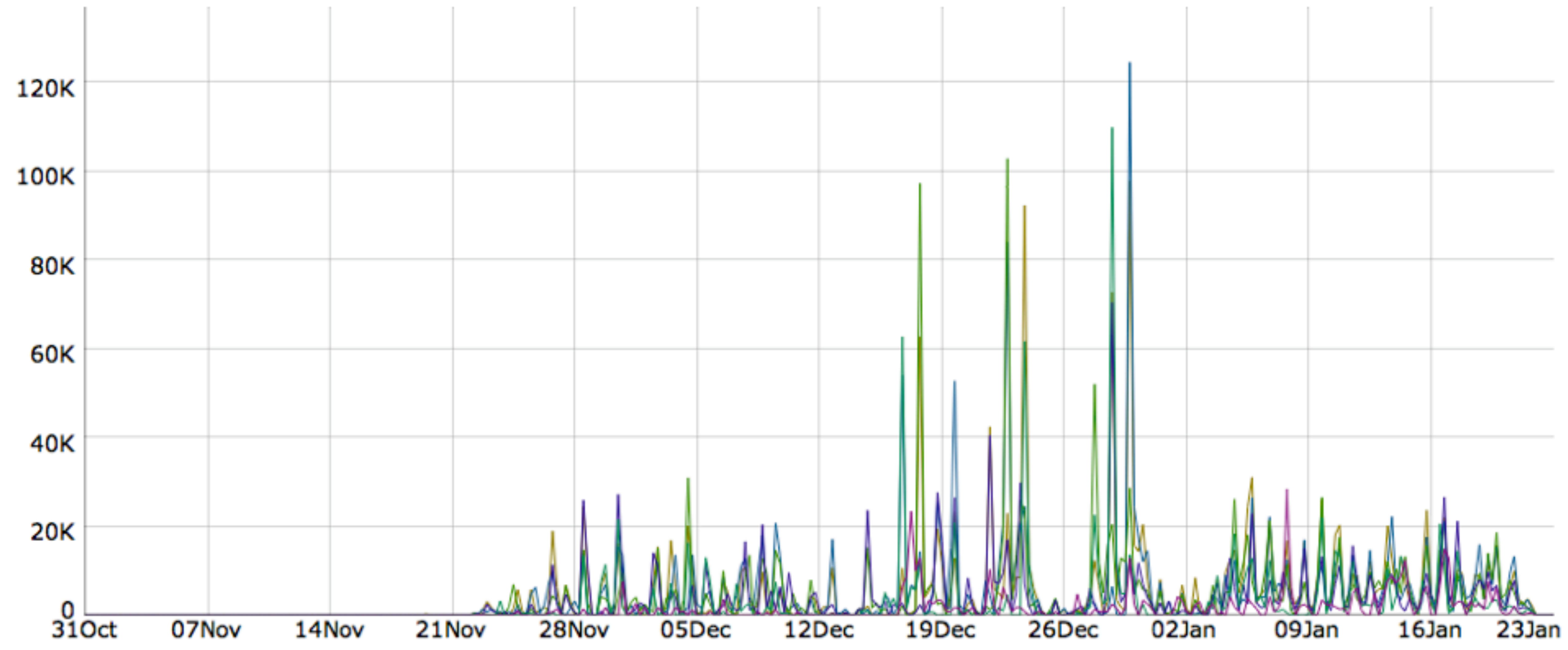
News



On-going research - political opinion dynamics

News

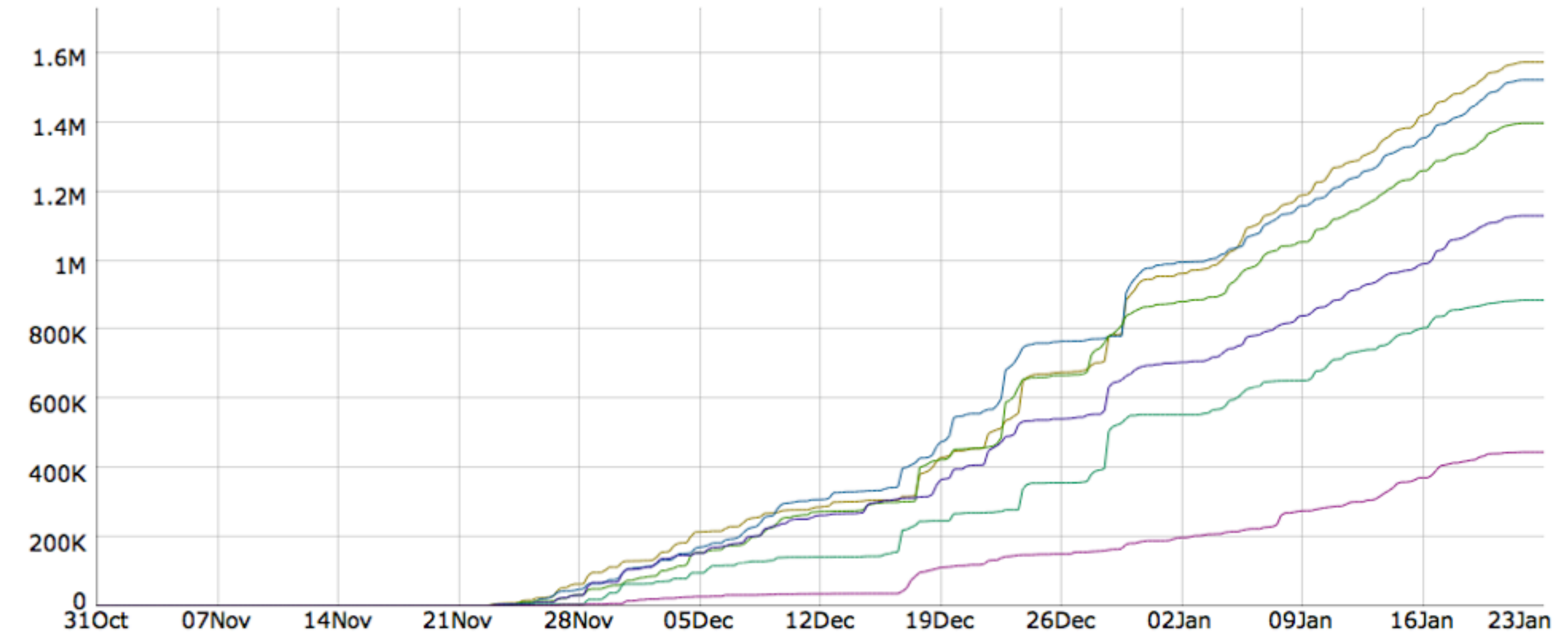
Figure 1. Daily number of tweets per candidate posted on Twitter.



On-going research - political opinion dynamics

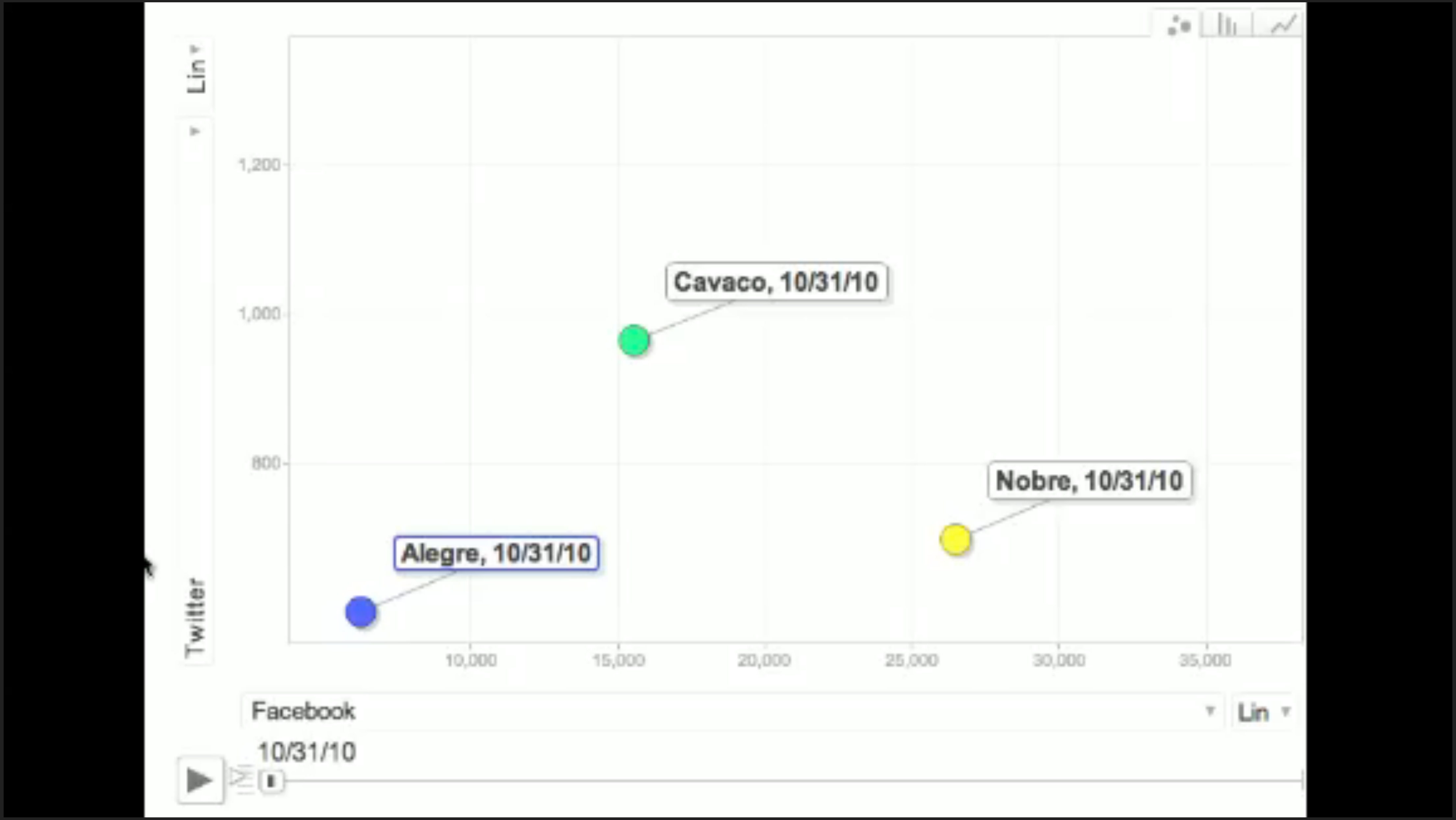
News

Figure 2. Cumulative summing of the previous chart.

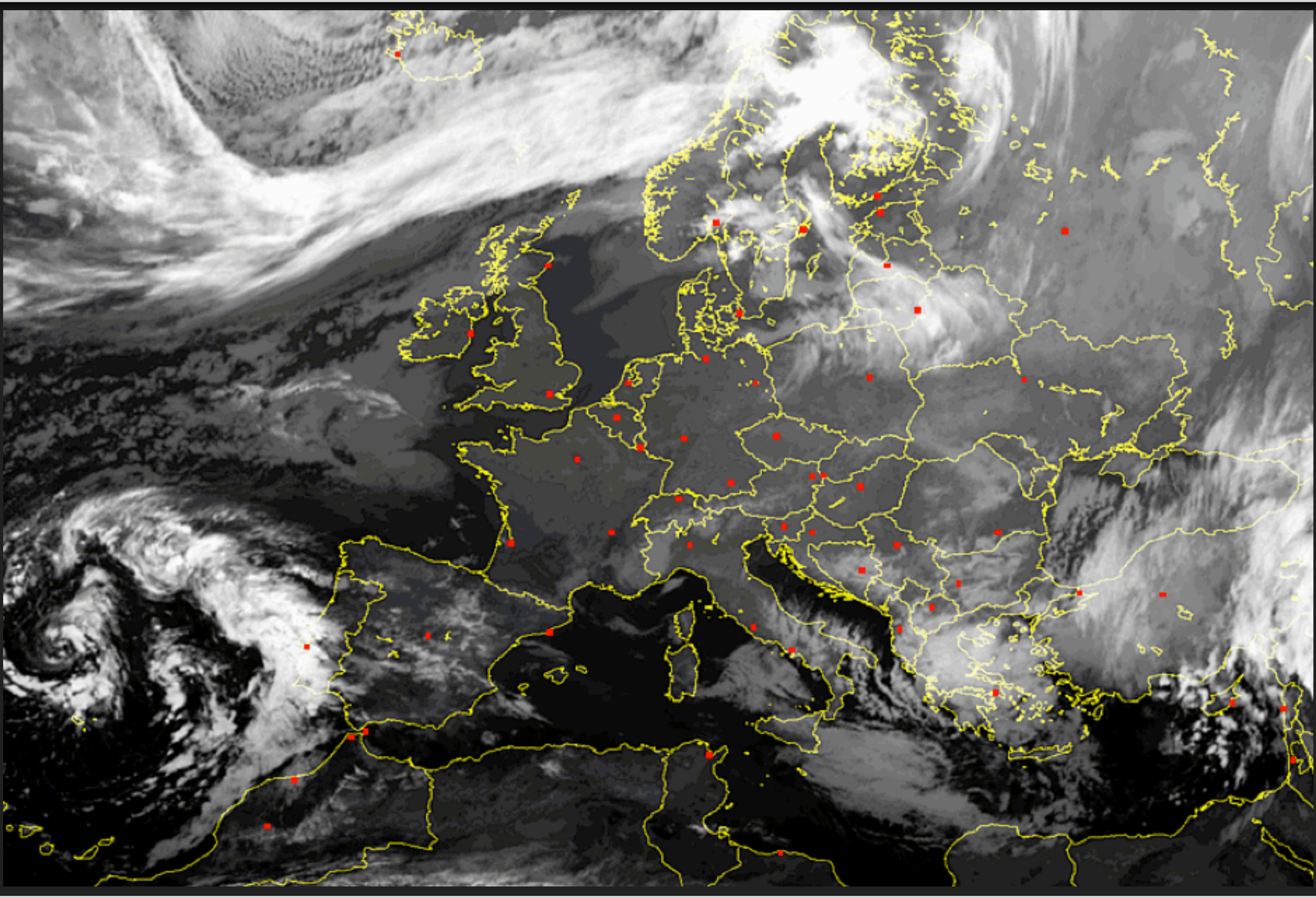


On-going research - political opinion dynamics

Social networks



On-going research - socio-economic “climate” at a global scale



On-going research - socio-economic “climate” at a global scale

Goal:

monitoring financial and economic data

+

monitoring opinion dynamics concerning economy, markets, etc

How to decide what data to collect ?

On-going research - socio-economic “climate” at a global scale

GDP – Gross Domestic Product

“economy strenght”; data from the International Monetary Fund, relative to 2008

DMC – Domestic Market Capitalization

“market size”; data from the World Federation of Exchanges, relative to 2008

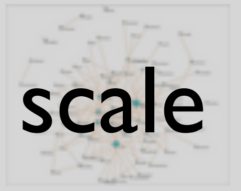
AI - Financial Market Activity

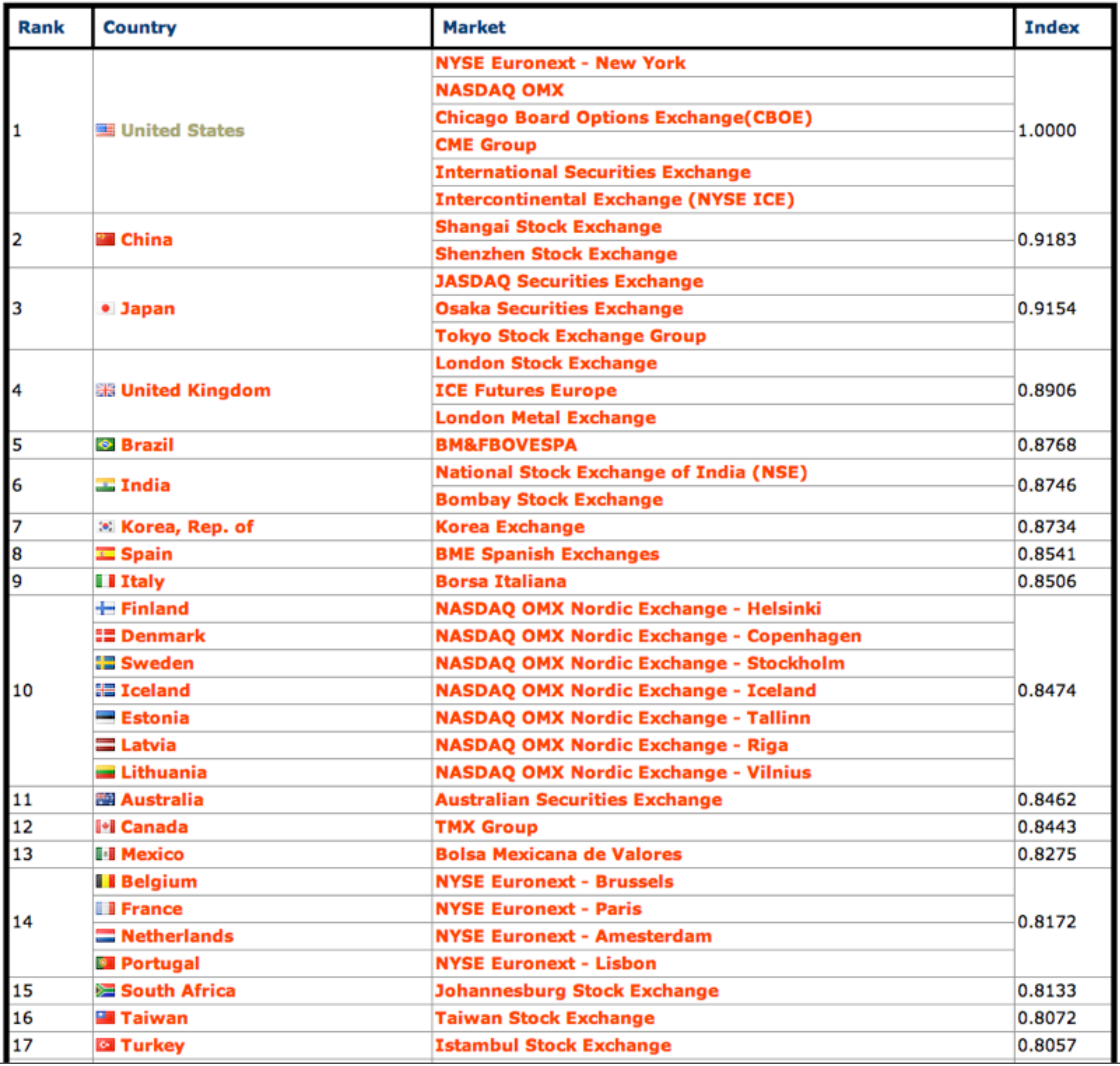
stocks, bonds and derivatives traded

$$TI = \frac{1}{4}AI + \frac{2}{4}DMCI + \frac{1}{4}GDPI$$

On

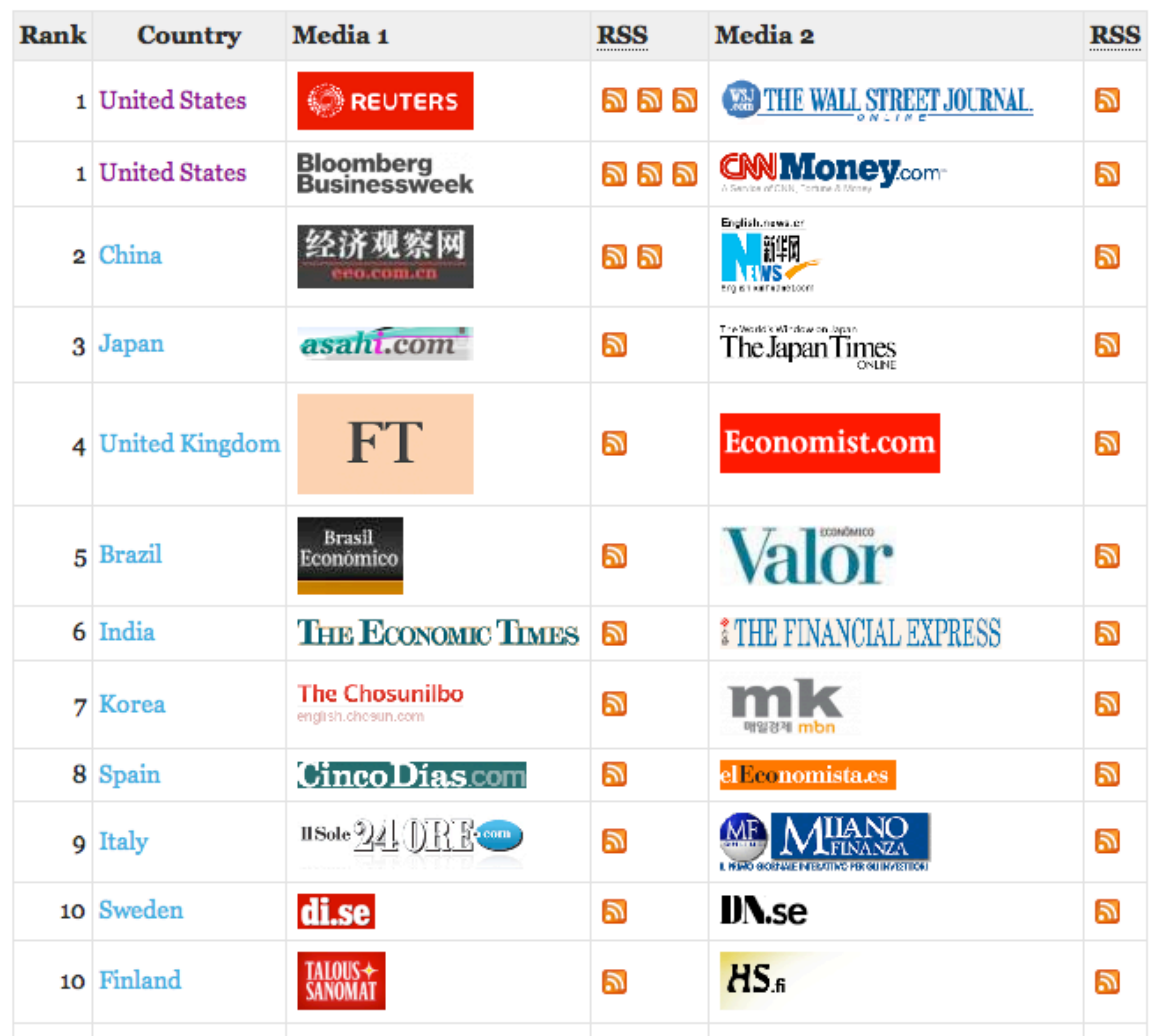
l scale





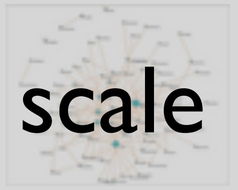
Rank	Country	Market	Index
1	United States	NYSE Euronext - New York NASDAQ OMX Chicago Board Options Exchange(CBOE) CME Group International Securities Exchange Intercontinental Exchange (NYSE ICE)	1.0000
2	China	Shangai Stock Exchange Shenzhen Stock Exchange	0.9183
3	Japan	JASDAQ Securities Exchange Osaka Securities Exchange Tokyo Stock Exchange Group	0.9154
4	United Kingdom	London Stock Exchange ICE Futures Europe London Metal Exchange	0.8906
5	Brazil	BM&FBOVESPA	0.8768
6	India	National Stock Exchange of India (NSE) Bombay Stock Exchange	0.8746
7	Korea, Rep. of	Korea Exchange	0.8734
8	Spain	BME Spanish Exchanges	0.8541
9	Italy	Borsa Italiana	0.8506
10	Finland	NASDAQ OMX Nordic Exchange - Helsinki	0.8474
	Denmark	NASDAQ OMX Nordic Exchange - Copenhagen	
	Sweden	NASDAQ OMX Nordic Exchange - Stockholm	
	Iceland	NASDAQ OMX Nordic Exchange - Iceland	
	Estonia	NASDAQ OMX Nordic Exchange - Tallinn	
	Latvia	NASDAQ OMX Nordic Exchange - Riga	
	Lithuania	NASDAQ OMX Nordic Exchange - Vilnius	
11	Australia	Australian Securities Exchange	0.8462
12	Canada	TMX Group	0.8443
13	Mexico	Bolsa Mexicana de Valores	0.8275
14	Belgium	NYSE Euronext - Brussels	0.8172
	France	NYSE Euronext - Paris	
	Netherlands	NYSE Euronext - Amesterdam	
	Portugal	NYSE Euronext - Lisbon	
15	South Africa	Johannesburg Stock Exchange	0.8133
16	Taiwan	Taiwan Stock Exchange	0.8072
17	Turkey	Istambul Stock Exchange	0.8057

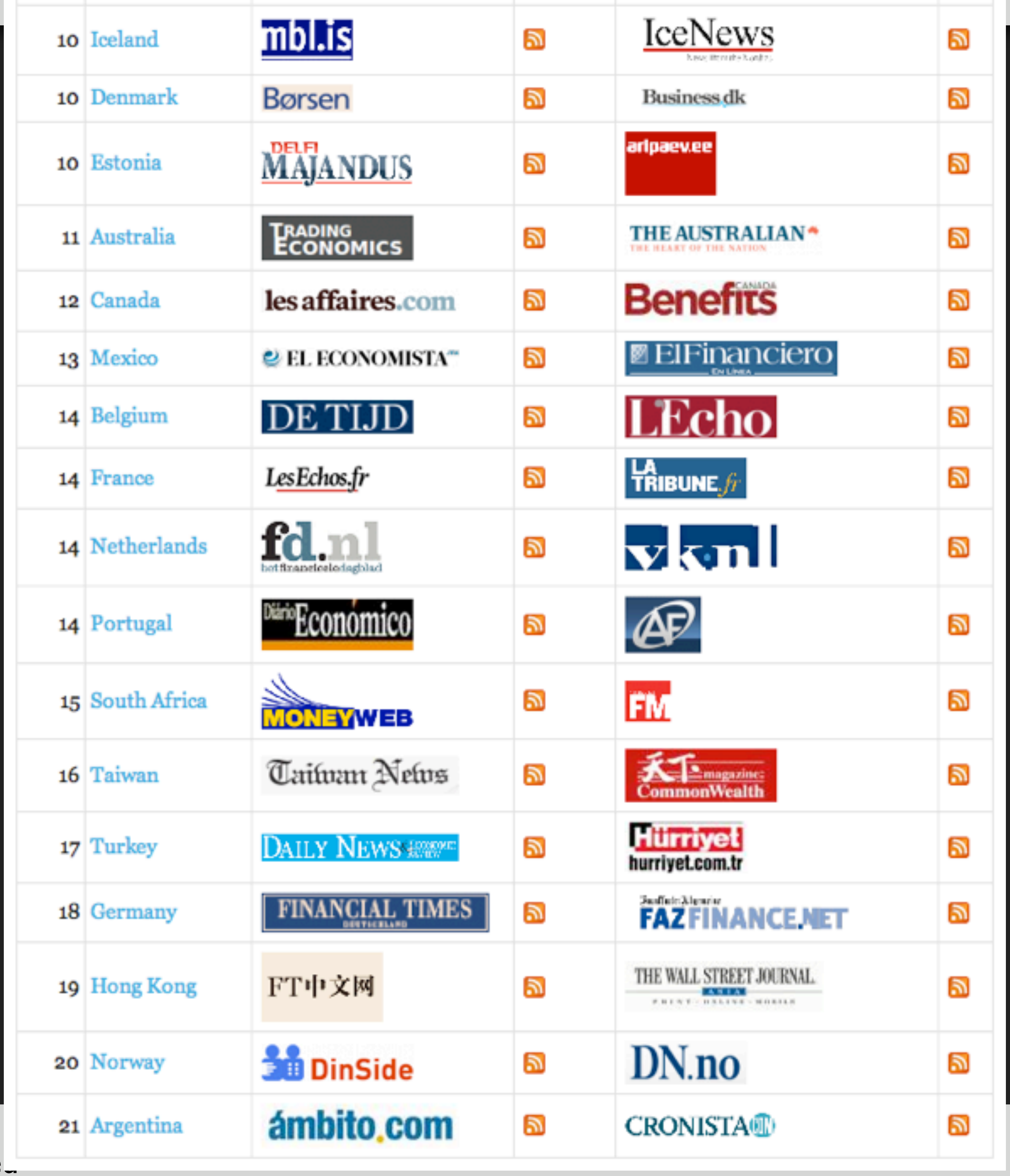
On-going research - socio-economic “climate” at a global scale



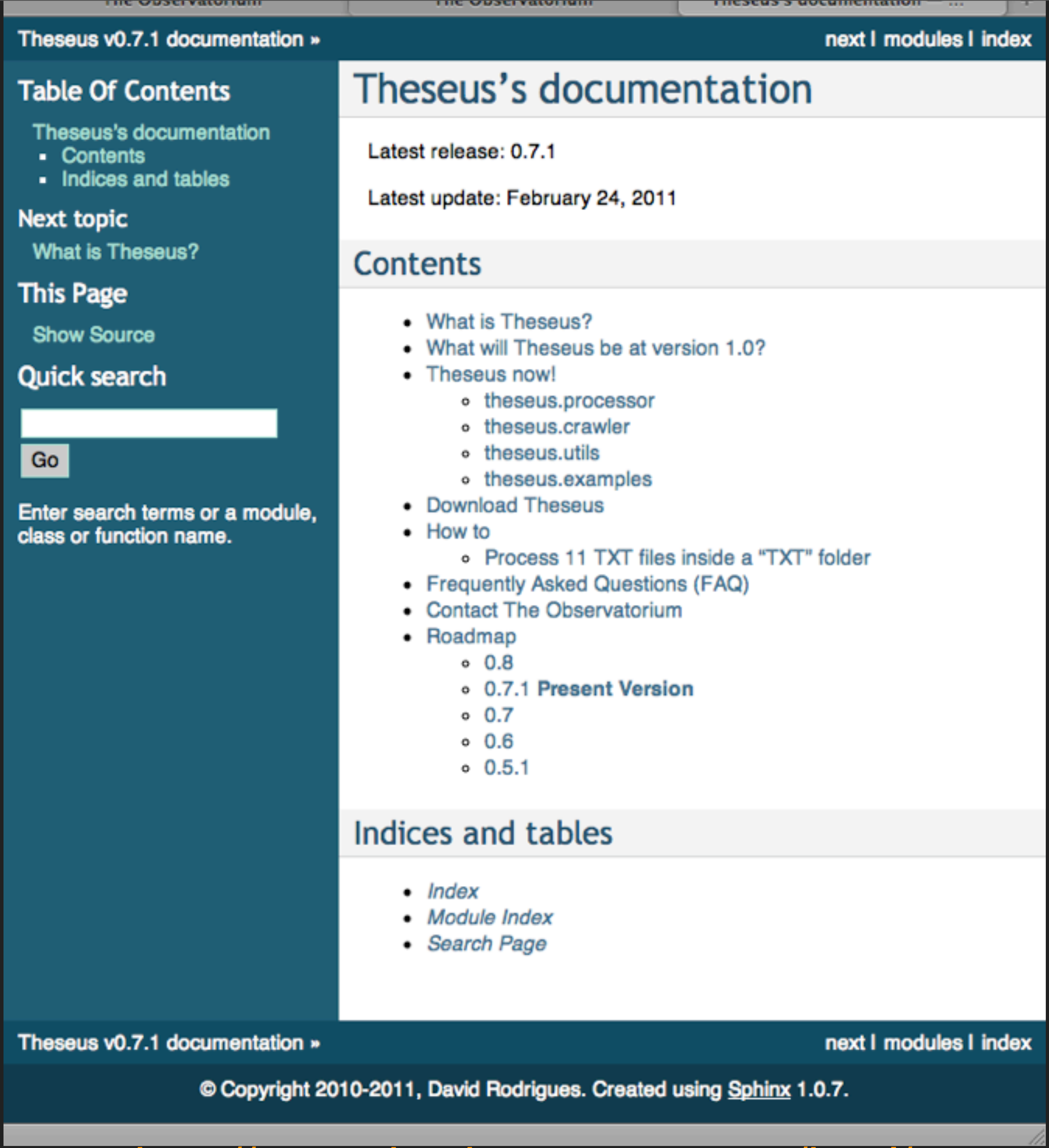
Rank	Country	Media 1	RSS	Media 2	RSS
1	United States	REUTERS	◙ ◙ ◙	THE WALL STREET JOURNAL ONLINE	◙
1	United States	Bloomberg Businessweek	◙ ◙ ◙	CNNMoney.com	◙
2	China	经济观察网 eeo.com.cn	◙ ◙	English.news.cn News	◙
3	Japan	asahi.com	◙	The Japan Times ONLINE	◙
4	United Kingdom	FT	◙	Economist.com	◙
5	Brazil	Brasil Econômico	◙	Valor Econômico	◙
6	India	The Economic Times	◙	The Financial Express	◙
7	Korea	The Chosunilbo english.chosun.com	◙	mk mbn	◙
8	Spain	CincoDias.com	◙	elEconomista.es	◙
9	Italy	Il Sole 24 ORE.com	◙	MF Milano Finanza	◙
10	Sweden	di.se	◙	DN.se	◙
10	Finland	Talous Sanomat	◙	HS.fi	◙

On-going research - socio-economic “climate” at a global scale





10	Iceland	mbl.is		IceNews	
10	Denmark	Børsen		Business.dk	
10	Estonia	DELFI MAJANDUS		arIpaev.ee	
11	Australia	TRADING ECONOMICS		THE AUSTRALIAN THE HEART OF THE NATION	
12	Canada	les affaires.com		CANADA Benefits	
13	Mexico	EL ECONOMISTA		ElFinanciero EN LÍNEA	
14	Belgium	DE TIJD		L'Echo	
14	France	LesEchos.fr		LA TRIBUNE.fr	
14	Netherlands	fd.nl het financieele dagblad		vk.nl	
14	Portugal	Diário Económico			
15	South Africa	MONEYWEB		FM	
16	Taiwan	Taiwan News		天下 magazine CommonWealth	
17	Turkey	DAILY NEWS & ECONOMIC REVIEW		Hürriyet hurriyet.com.tr	
18	Germany	FINANCIAL TIMES DEUTSCHLAND		Frankfurter Allgemeine FAZ FINANCE.NET	
19	Hong Kong	FT中文网		THE WALL STREET JOURNAL ASIA PRINT · ONLINE · MOBILE	
20	Norway	DinSide		DN.no	
21	Argentina	ámbito.com		CRONISTA.com	



Theseus v0.7.1 documentation » next | modules | index

Table Of Contents

- Theseus's documentation
  - Contents
  - Indices and tables

Next topic

What is Theseus?

This Page

Show Source

Quick search

Enter search terms or a module, class or function name.

Theseus's documentation

Latest release: 0.7.1

Latest update: February 24, 2011

Contents

Indices and tables

- *Index*
- *Module Index*
- *Search Page*

Theseus v0.7.1 documentation » next | modules | index

http://www.theobservatorium.eu/html/

The team

António Fonseca

David Rodrigues

José António Silva

Manuel Tânger

Pedro Loureiro

Jorge Louçã

software is available under open source licence

data is available under request

fet11

The European
Future Technologies
Conference and Exhibition

4-6 May 2011
Budapest, Hungary

Science

beyond fiction

mmission
n Society and Media

Thank you !