

Correlation of Forum Activity and real world events based on a social network dataset

Cedric Waldburger*

Department of Electrical Engineering
ETH Zurich
ctwaldburger@student.ust.hk

Christoffer Hirsimaa

Department of Computer Science
KTH Stockholm
cahirsimaa@stu.ust.hk

May 16, 2011

Abstract

In the beginning of web2.0 there were not as many sites as today where people were able to discuss current events. One popular platform to discuss news was aSmallWorld.net, an invitation-only social network that was started even before facebook. We built a Visualizer to show forum activity on aSmallWorld.net between 2004 and 2009 and we tried to find a correlation to real life events by comparing the visualized data of the event's time span with the activity before the event had happened. We then tried to conclude in which cases and which years the network activity corresponded to real life events and whether it was still true during the last few years or not. We concluded that in earlier years there was a correlation between events and the forum activity, but in later years the connection is weaker or even nonexistent.

*thanks to ASmallWorld Holding, New York for providing the data

Contents

1	Goal	3
1.1	About ASmallWorld	3
1.2	Our objective	3
1.3	Facts & Figures	4
2	Approach	4
2.1	Getting the Data	4
2.2	Building the Visualizer	4
2.3	Setting up the Experiment	5
2.4	Environment and Code	5
3	Experiments - News Events	6
3.1	2005	6
3.2	2006	10
3.3	2007	13
3.4	2008	16
3.5	2009	20
4	Conclusion	22

1 Goal

1.1 About ASmallWorld

ASmallWorld.net (ASW) is a unique social media combining a by-invitation-only social network with a content delivery platform targeting the "upper end" of the social media market. Founded in 2004 even before facebook became popular, ASW stayed the market leader in this segment. As An integrated media company, ASW is an ideal match for advertisers seeking to target the world's tastemakers and develop heightened mindshare with this sophisticated and influential group.

Some of the site's features are:

- member profiles and the ability to 'connect' with each other
- a lifestyle and travel magazine called ASmallMagazine
- an eBay-like platform where members trade everything from computers to Houses and Yachts
- a GeoLocator, where member log their trips, ask for sightseeing advice and find selected services for every destination in the Travel Guides
- a Forum where people seek business advice, look for like-minded members all over the globe or discuss current happenings
- some more, less used features

With ASW existing before many of the other web 2.0 sites, it's forum was and is heavily used also to discuss politics and current news. Rumors say, that in 2004 and 2005, ASW members would be faster at posting news to the site's forum than journalists had their news on their respective news websites and ASW functioned as an 'information hub'.

Based on this rumor, we decided to see whether it is possible to find correlations between the forum's activity and the real world happenings.

1.2 Our objective

In order to find out, how correlated ASW's forum activity and real life news are, we are building a tool to visualize forum activity over a certain time span using the Google Maps API.

We'll use the visualized activity to

- have a look at the world wide activity to see if there is a substantial part of it is coming from the particular area of the real life's event
- zoom in on that particular region and compare the time span of the real life's event with the weeks before to see if activity increased

We'll then try to reason and conclude where and when the activity on the forum increased due to real life events.

1.3 Facts & Figures

- **Total number of records:** 254 622
- **First record:** June 7, 2004
- **Last record:** February 2, 2011 (but there's a gap in 2010 and we therefore decided to only look at 2004-2009)
- **database size and type:** 19.5 MB, sqlite3
- **geography:** ASW's user base is concentrated in Europe and the east and west coast of the United States (with a main concentration in New York/Boston and Los Angeles/San Francisco)

2 Approach

2.1 Getting the Data

ASW uses a SQL-based database to store their data. We were granted to access all their data from mid 2004 until late 2009. Using scripting languages, we converted the database dump into a sqlite database that we could afterwards use locally and online without setting up a database server. The database contained one table with each record containing this information:

<code>id</code>	<code>post_id</code>	<code>creation_date</code>	<code>member_id</code>	<code>lng (longitude)</code>	<code>lat (latitude)</code>
:	:	:	:	:	:

2.2 Building the Visualizer

We used the Google Maps API in javascript to put out circles on a map. We basically put out two date pickers so you could define a time period to visualize in a specified speed. By starting the animation it starts to iterate over each date and for each date make an asynchronous AJAX call to retrieve the posts from the server. The iteration is done by `setInterval` method so that the browser don't freeze during visualisation. Posts are iterated over to create new circles on the map or increase the size and color intensity. The size and opacity are calculated according to the following formulas 1 and 2 for each circle.

$$\text{radius} = 20000 * \log(10 * \text{numberOfPosts}) \quad (1)$$

$$\text{opacity} = 0.2 * \log(\text{numberOfPosts}) \quad (2)$$

These formulas were found through trial and error and they resulted in very detailed images. For more detailed implementation we refer to our publicly available code in <https://github.com/Hirschen/comp621u>.

2.3 Setting up the Experiment

In order to test our hypothesis that real life events and the forum activity are correlated, we decided to only look at the years with a full dataset (2005-2008) and for each year to select 3 events (exception: 4 in 2008, only 2 in 2009):

- an event that was present in public newspapers for approximately 1 week
- an event that was present for approximately 2 weeks
- an event that was discussed in the news for a month or longer

We selected the events based on <http://moldova.org> and we referred to the particular article at the beginning of each year. Then we compared the activity on relevant locations before the event and during its public presence. To make the comparison fair the duration of visualizing before and during events is the same.

We chose events from the following geographical locations (some events are affecting more than one location):

- 9 events affecting Europe
- 4 events affecting the US
- 6 events affecting Asia
- 1 events affecting South America

2.4 Environment and Code

The Visualizer built for our project is currently available at <http://cedricwaldburger.ch/comp621u/main.html> and will be available until 2 weeks after the end of the semester.

The code can be pulled from Github on this address <https://github.com/Hirschen/comp621u>.

3 Experiments - News Events

3.1 2005

Top10 Events in 2005 on moldova.org

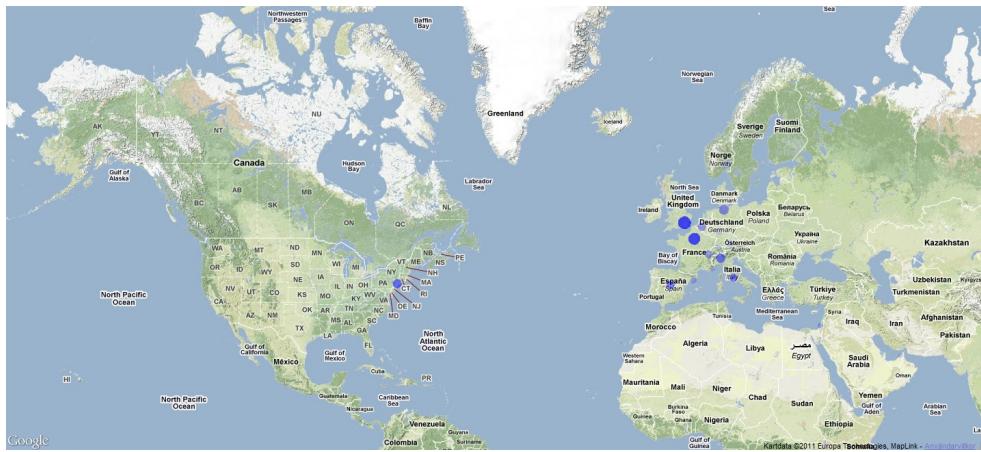
- **Death of Pope John Paul II**

The death of Pope John Paul II marked the end of an era in the life of the Roman-Catholic Church and modern history.

time span: April 2 - April 9

geographical area: Vatican

total number of threads: 347



interpretation: correlated

Compared to the week before, we saw an increased activity in many European countries including the UK, Denmark, the South of France, Spain, Switzerland and Italy. Even though the Death of Pope John

Paul II is definitely an event with worldwide attention, it is reasonable that it was most discussed around central Europe from the Pope was originally from (Poland). We expected to see more activity in Italy, close to the Vatican, however we can definitely say that there's a strong correlation between the Pope's death and activity on ASW in central Europe.

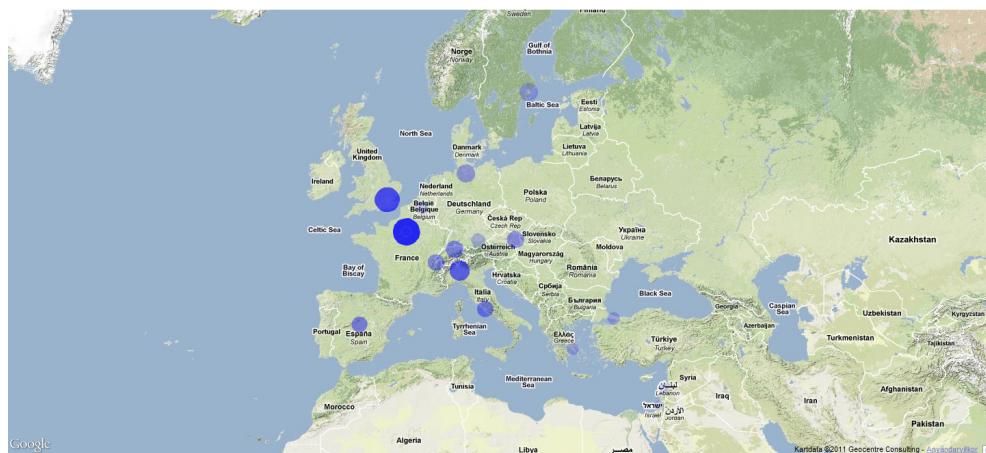
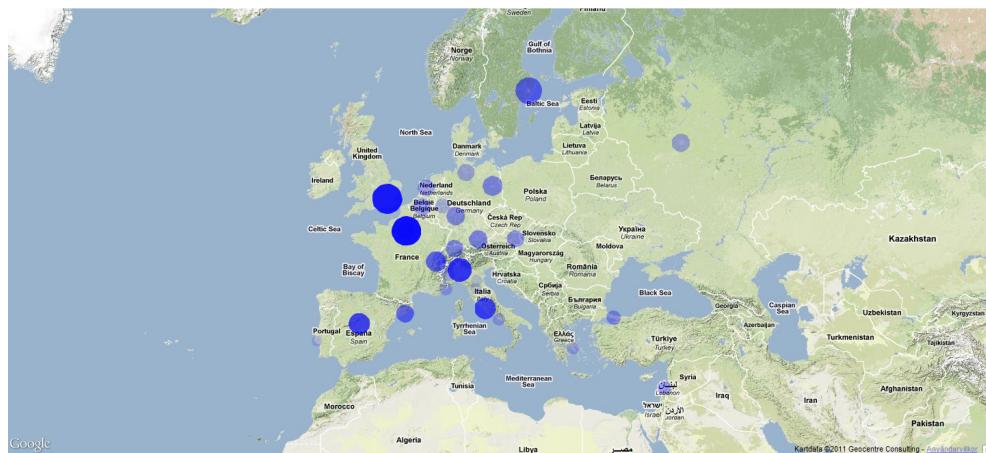
- **Terrorist act in London**

A terrorist act in London overshadowed the G8 summit in Gleneagles, Scotland, and spoiled Britons' joy over London's election as the host city for the 2012 Olympics. Four explosive devices went off in the underground and a bus, leaving 50 people dead and about 700 injured.

time span: July 7 - July 21

geographical area: London, UK

total number of threads: 498



interpretation: correlated

On the first day of the event there's a massive amount of activity in Paris and London, which proofs that the event and the network activity are definitely correlated. During the week after the event there was more activity in central europe and also cities that usually show not much activity were opening forum threads in large numbers - this includes cities in the Netherlands, Belgium, all over Germany and even in Russia.

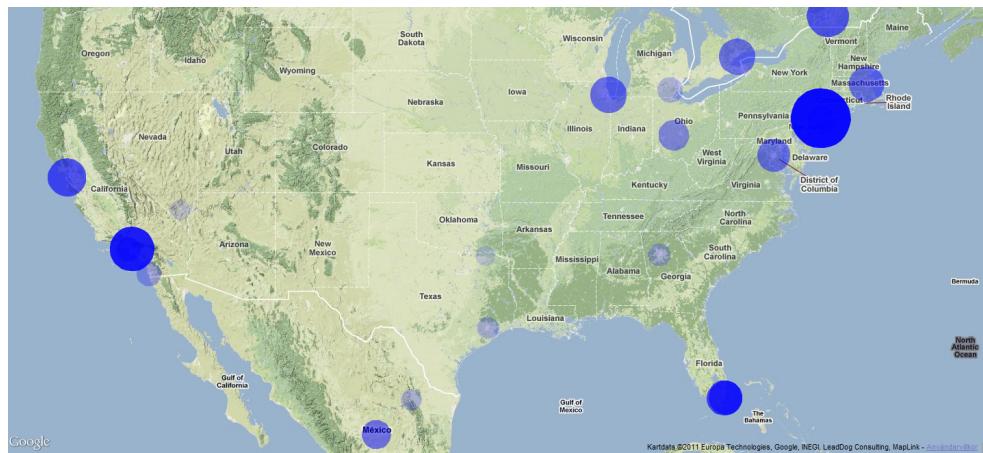
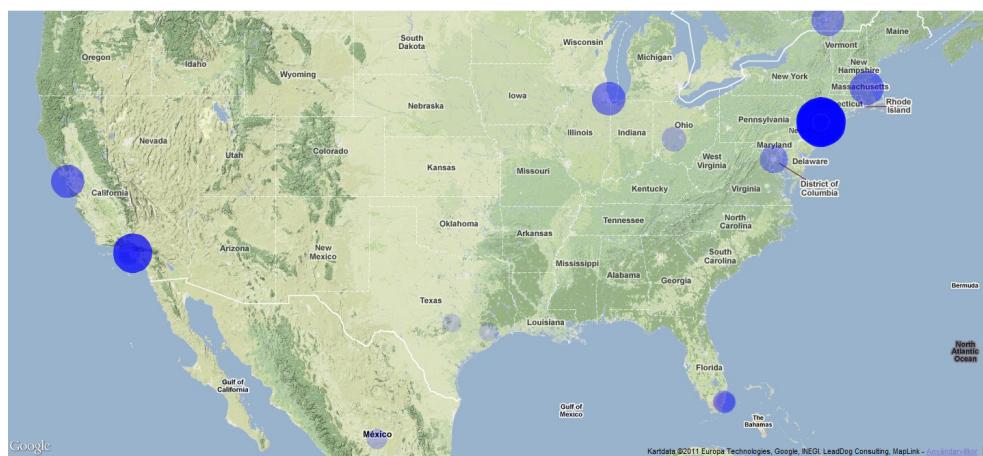
- **Hurricane Katrina**

Hurricane Katrina in the United States showed that the world's leading super power was unprepared to deal with the aftermath of the natural disaster. The hurricane destroyed the cradle of jazz, New Orleans.

time span: August 23 - September 23

geographical area: USA

total number of threads: 4108



interpretation: correlated

The month after Hurricane Katrina, New York showed a lot more activity than usual (the bulb's size is logarithmic to the number of posts) and all over the US new circles popped up, also in areas where there's almost no activity otherwise. This is true for the northern and southern part and also for the east and West Coast. There's no activity in New Orleans which is probably due to a very small user base in that area.

3.2 2006

Top10 Events in 2006 on moldova.org

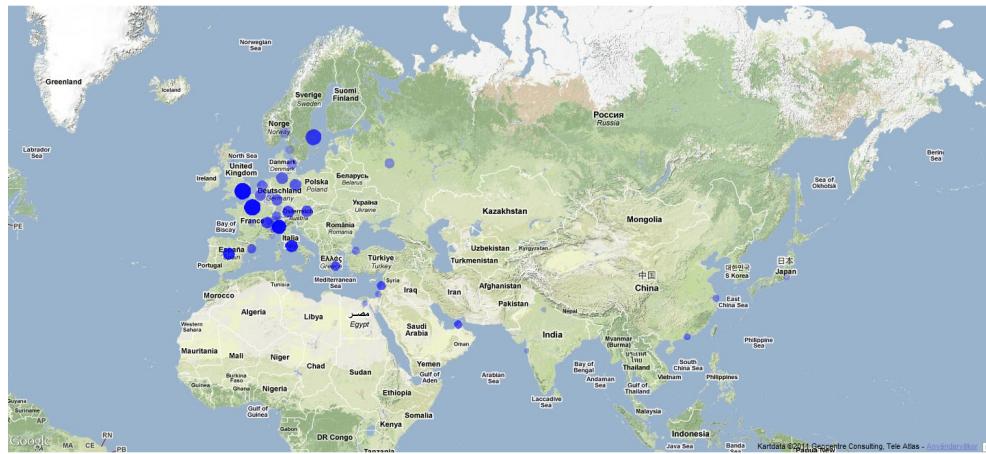
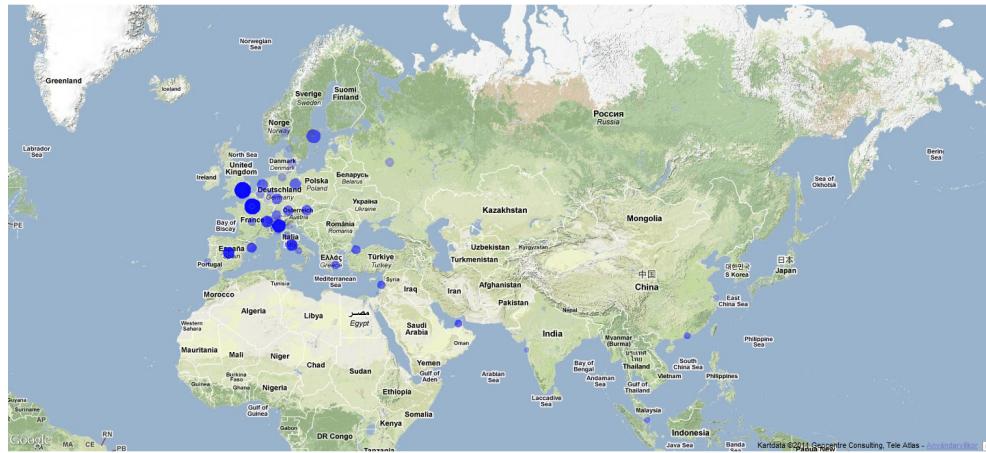
- **Bird Flu**

People watched nervously as the infamous disease spread from Asia to Europe. But flu season died down before disaster struck. The human being was in a big danger, as a current H5N1 strain is a fast-mutating, highly pathogenic avian influenza virus (HPAI) found in multiple bird species.

time span: January 1 - January 14

geographical area: Asia

total number of threads: 25930



interpretation: not correlated

There's more activity in Greece, Portugal, Scandinavia, Egypt, Syria, Singapur and Shanghai. As for Europe this could be due to minor local

activities in those countries and as for Asia there's not a big increase to normal activity and the network activity therefore was probably not correlated to the Bird Flu.

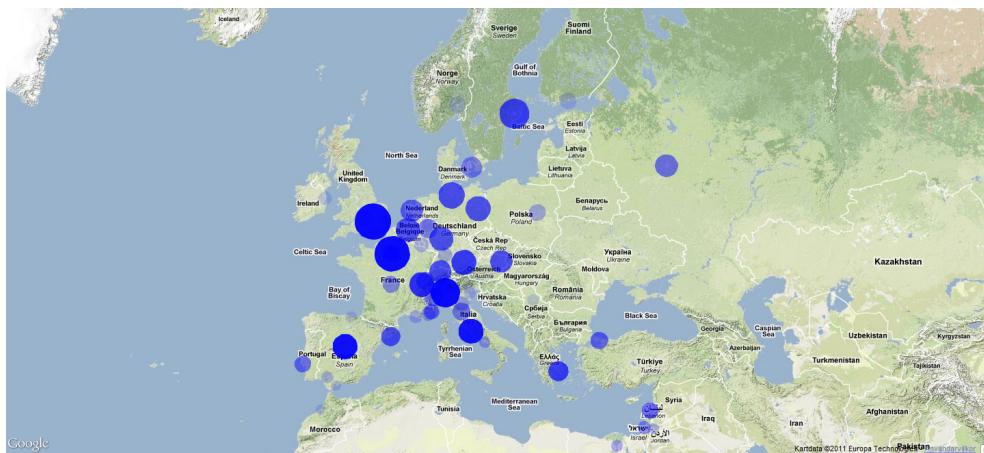
- **Soccer World Cup**

Italy wins the World Cup 5-3 vs. France.

time span: June 9 - July 9

geographical area: Central Europe (Cup held in Germany)

total number of threads: 4326



interpretation: correlated

Even the month before the world cup, there's a lot more activity in europe than usual. During the world cup almost every ASW member city was active in the forum. The existing circles got a bit bigger during this month and there's new activity in the north of Spain, South

of France, all over Italy, Ireland, Scandinavia and many others. This month showed so much more activity in smaller and also bigger european cities compared to other months that it is very likely to be related to the World Cup.

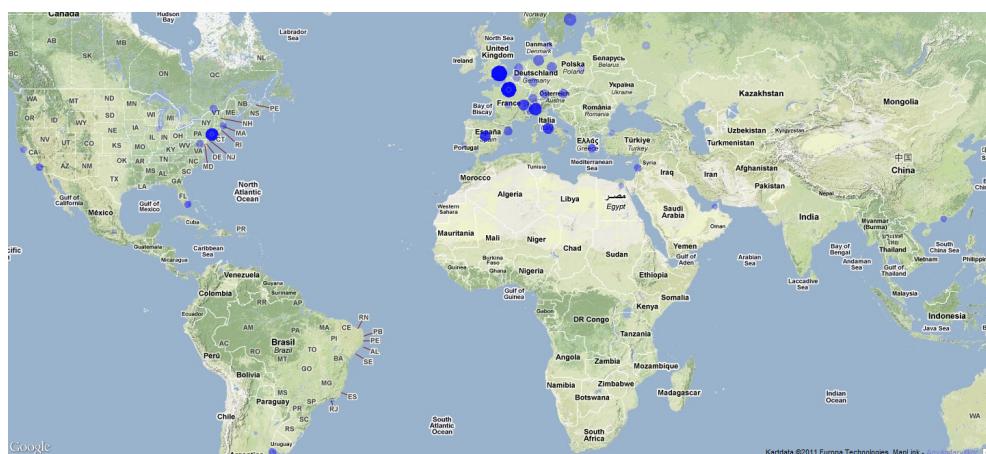
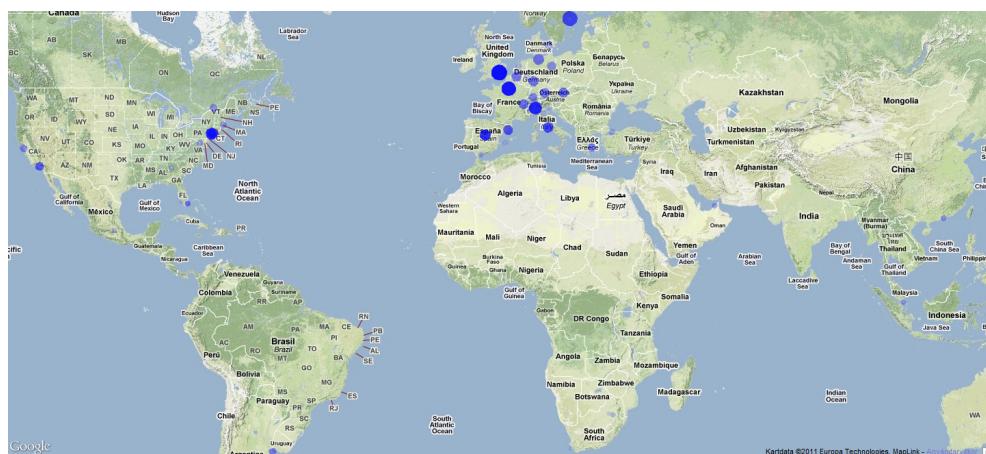
- **Saddam Hussein sentenced to death**

Saddam Hussein sentenced to death by hanging by an Iraqi court. Earlier in this year he was charged with genocide by an Iraqi court for a campaign against Iraq's Kurdish population in 1988.

time span: November 5 - November 12

geographical area: Iraq

total number of threads: 1212



interpretation: not correlated

There are no posts at all from Iraq (maybe due to a limited number of users in this area) and also no drastic changes worldwide. In this case,

forum activity is probably not related to Saddam Hussein's sentence to death.

3.3 2007

Top10 Events in 2007 on moldova.org

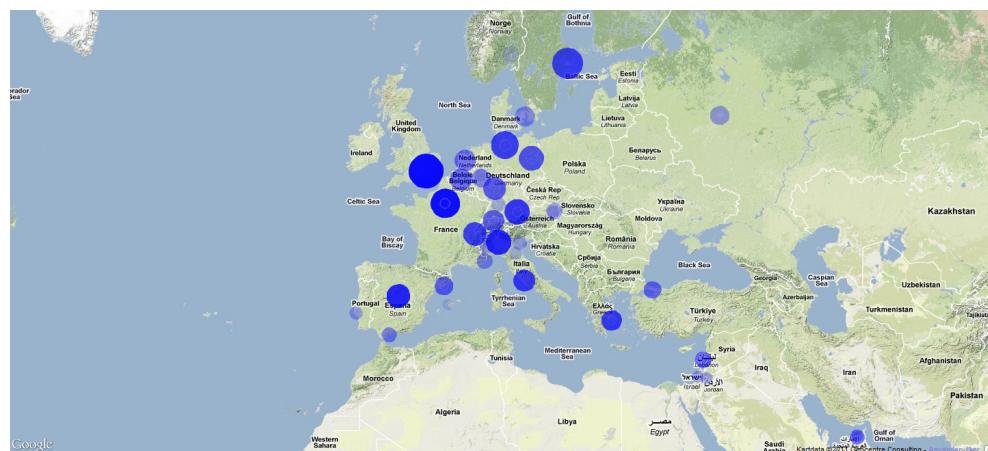
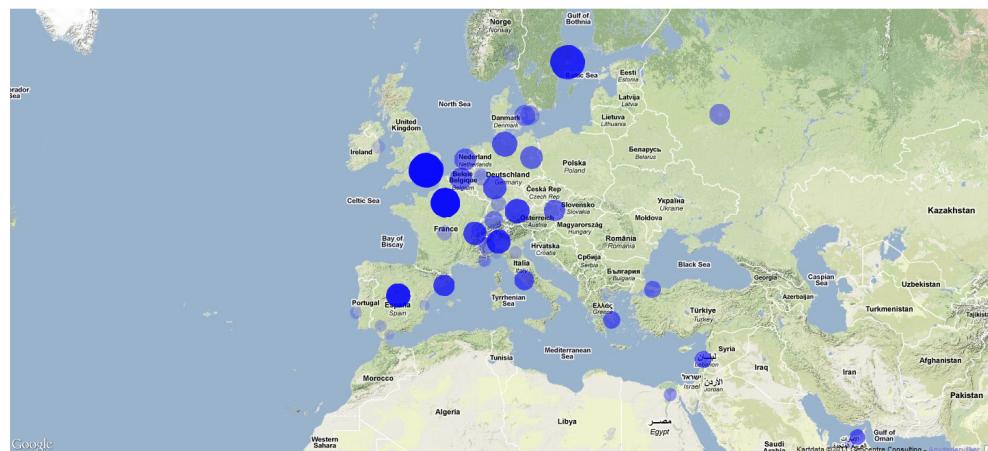
- **Gordon Brown becomes Prime Minister**

Gordon Brown replaces Tony Blair as the prime minister of Great Britain (June 27). "Let the work of change begin," said Brown in the day he was elected.

time span: June 27 - July 4

geographical area: UK

total number of threads: 2098



interpretation: not correlated

No difference between the week before and the week of election can be

seen. This is probably due to the fact that there was a lot of press coverage and discussion already the week before his election. Therefore the forum activity is probably not related to the actual event of his election, but may be related to his popularity during that time period (which we found hard to measure and therefore conclude that the event of his election and the forum activity are not correlated in this case).

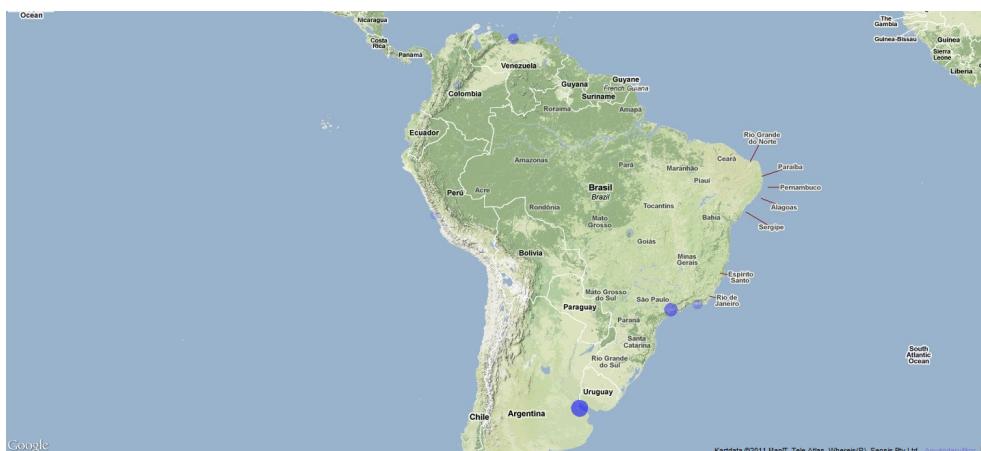
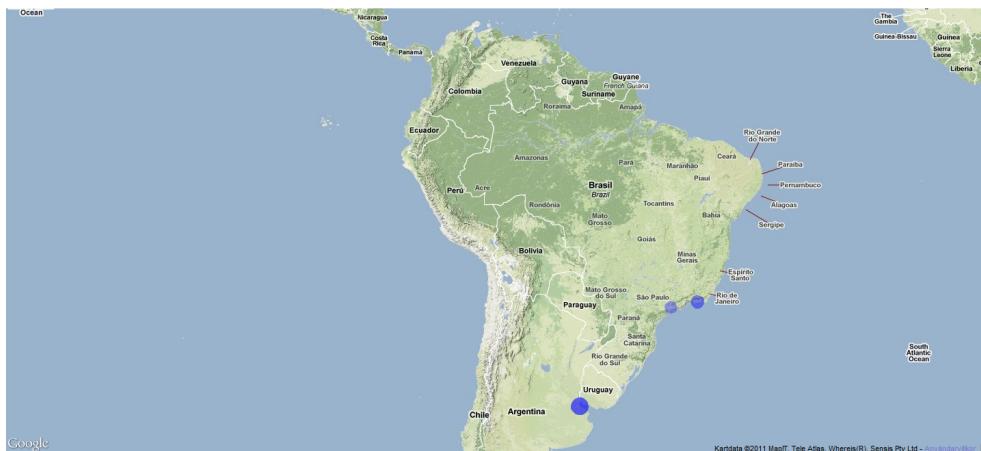
- **Chile Earthquake**

8.0 Earthquake in Peru kills over 500 people, injures over 1,500; 7.7 Earthquake in northern Chile.

time span: August 15 - August 29

geographical area: South America

total number of threads: 3033



interpretation: correlated

There's activity in south american cities that have never shown activity

before - including cities in Venezuela, Columbia, Brasil, Peru and Chile. This is most probably correlated to the earthquakes in Chile.

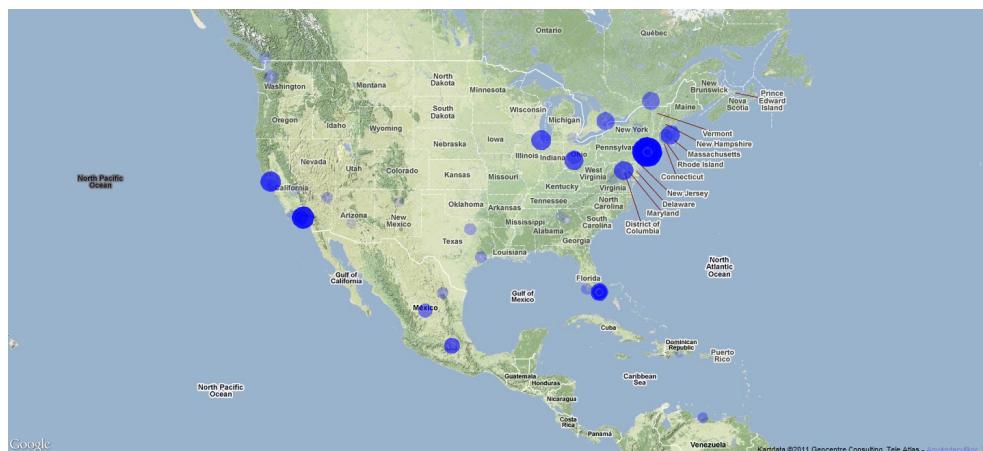
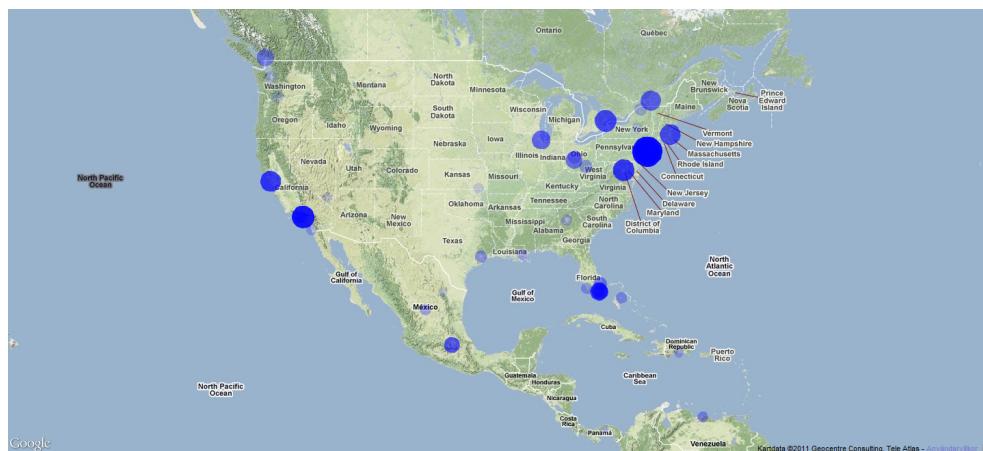
- **Writers Guild goes on strike**

The Writers Guild of America goes on a strike that lasts until February 12, 2008. More than 12,000 writers joined the strike. The strike's goal was to rectify what was perceived as a historical injustice to America: the greatly diminished monetary compensation the writers got in comparison with the profits of the larger studios. The guilds were on strike 100 days.

time span: November - December

geographical area: Los Angeles, USA

total number of threads: 10234



interpretation: not correlated

There's not visual change of the activity during the strike. The Writers

Guild's strike was probably not a often discussed topic on ASW and the activity is therefore not correlated.

3.4 2008

Top10 Events in 2008 on moldova.org

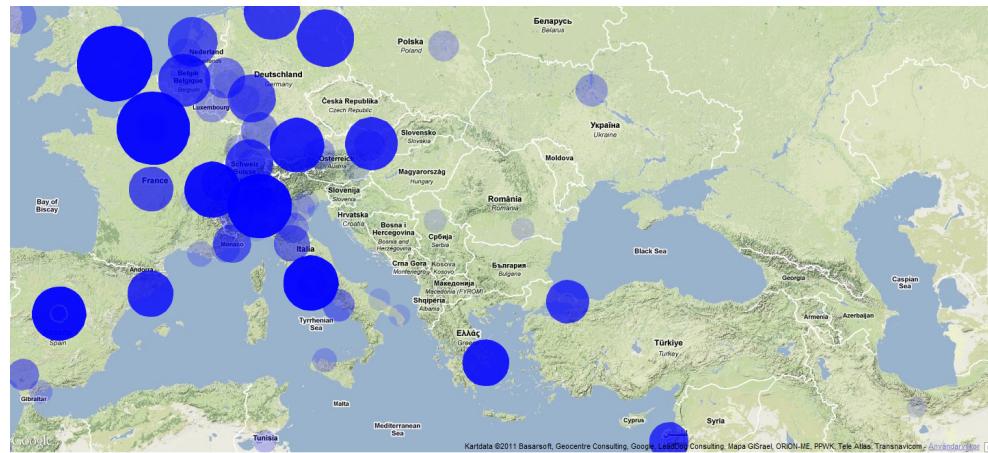
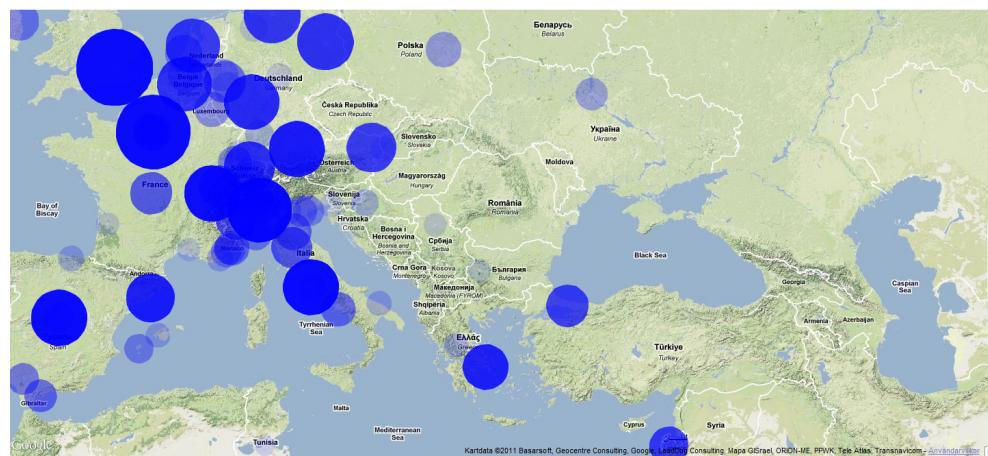
- **Kosovo declares independency**

Kosovo declares independence from Serbia, announced by prime minister Hashim Thaci. Serbian prime minister Vojislav Kostunica says he would never recognize the "false state." International reaction is mixed, with the United States, France, Germany, and Britain indicating that they plan to recognize Kosovo as the world's 195th country.

time span: February 17 - March 17

geographical area: Kosovo

total number of threads: 5250



interpretation: correlated

We found some activity in a city in Kosovo and increased activity in Kosovo's surroundings which makes us think that the forum activity and the happenings in Kosovo are related in this case.

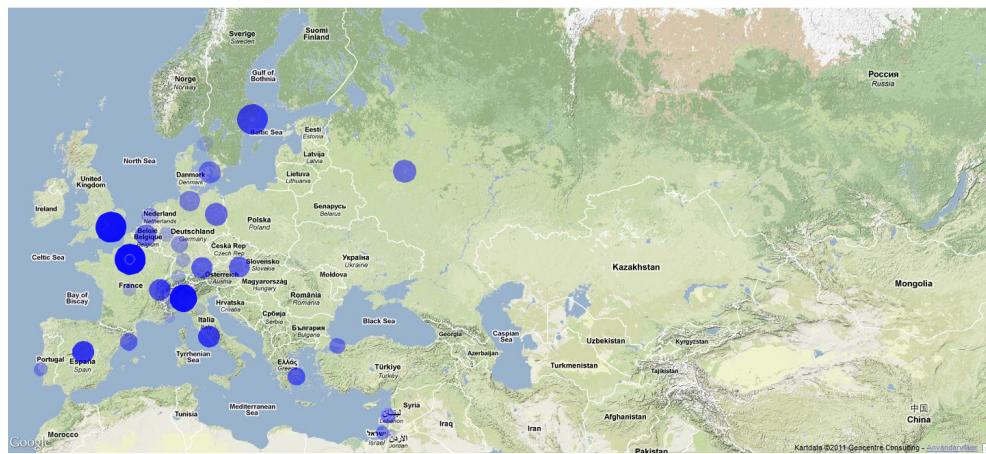
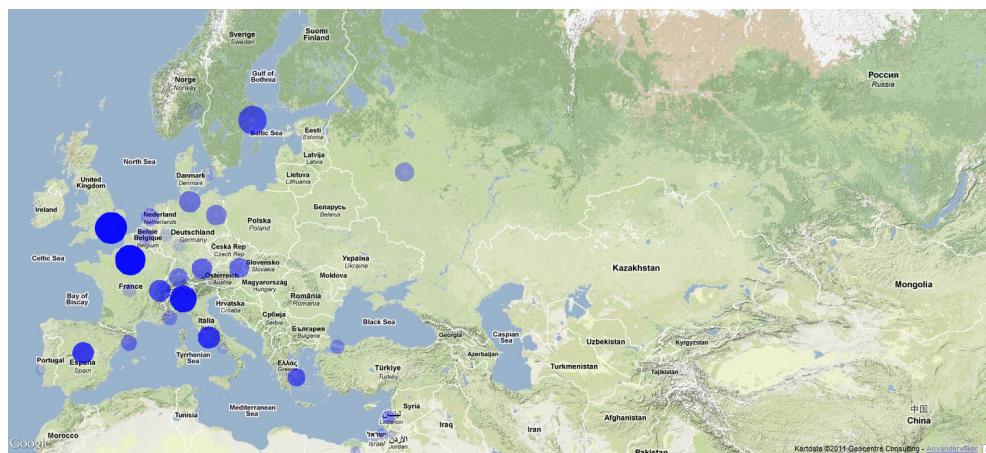
- **Medvedev becomes Russia's President**

Russia in 2008 chooses another president - Dmitri A. Medvedev, a former aide to Russian president Vladimir Putin, wins the presidential election in a landslide. Putin will remain in a position of power, serving as Medvedev's prime minister.

time span: March 2 - March 9

geographical area: Russia

total number of threads: 1353



interpretation: not correlated

Although there's a certain change of activity in the area around Moscow,

we don't think that this change is big enough to ensure that the activity increase was related to Medvedev's election and is not just noise. We therefore think that the forum activity and this event are not correlated.

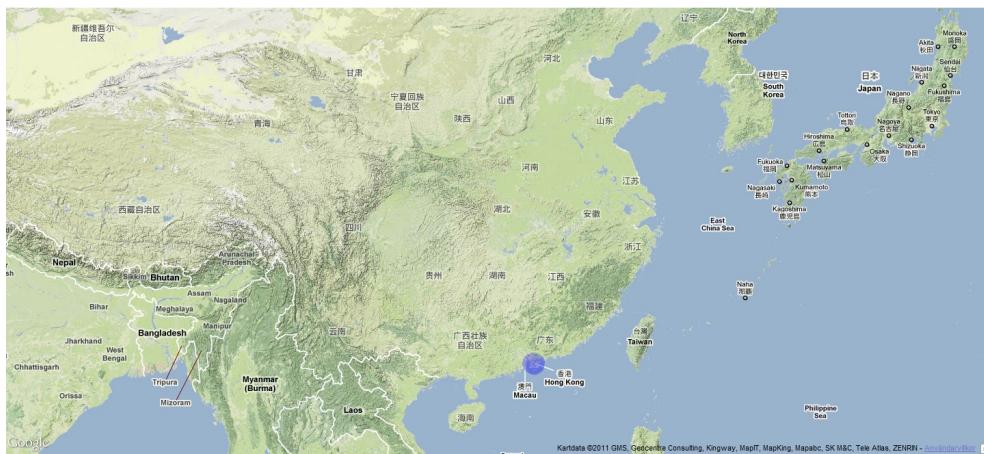
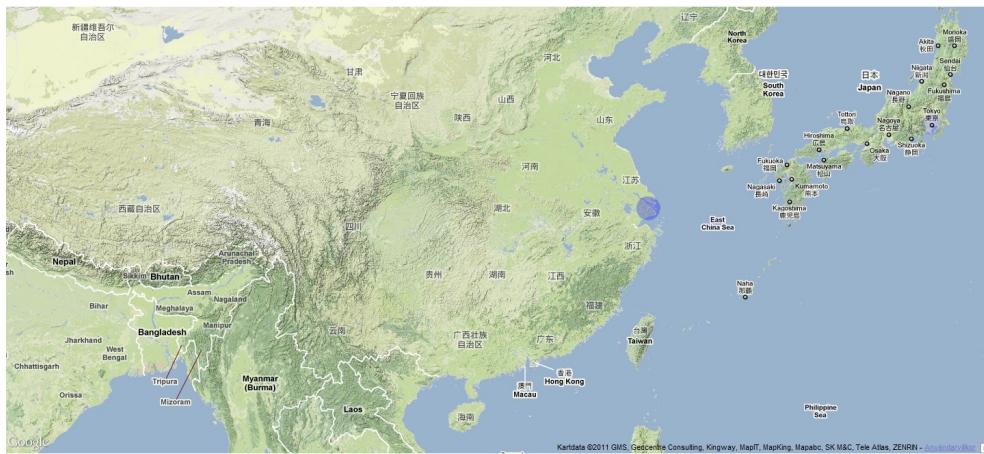
- **China's Earthquake**

On May 12, 2008, a magnitude 7.9 earthquake struck the Sichuan Province of China, killing more than 69,000 people, making it China's worst disaster in more than 30 years.

time span: May 12 - May 26

geographical area: China

total number of threads: 1660



interpretation: not correlated

There's almost no activity in Asia during this period and therefore the forum activity was not related to the real world events.

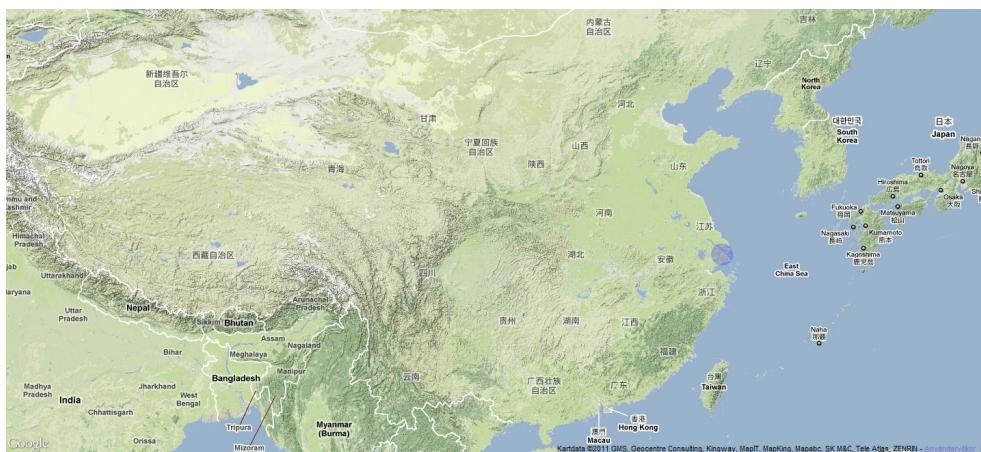
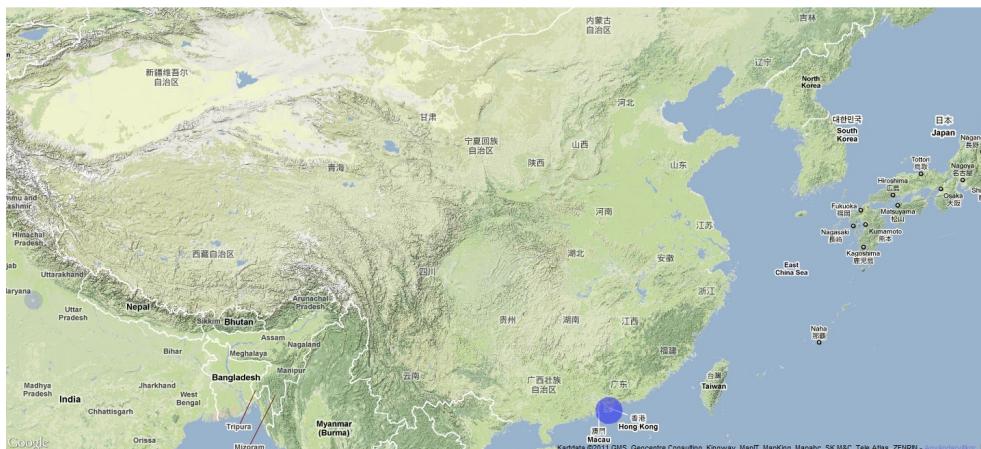
- **Summer Olympic Games**

Summer Olympic Games 2008 that took place in Beijing, China. It was the XXIX Olympiad. The international multi-sport event officially opened on August 6 and ended on August 24, featuring 10,500 athletes who struggled for the Olympic Gold medal in 302 events in 28 sports. The Summer Olympic Games 2008 was really worth remembering, showing remarkable performances of the athletes: 43 new world records and 132 new Olympic records. Besides, representative of 87 countries won at least one medal at these Olympics, an unprecedented achievement.

time span: August 8 - 24

geographical area: Beijing, China

total number of threads: 1124



interpretation: not correlated

The only activity seen in China moved from Hong Kong to Shanghai

during this period. Not correlated to what happens in Beijing at this time.

3.5 2009

Top10 Events in 2009 on moldova.org

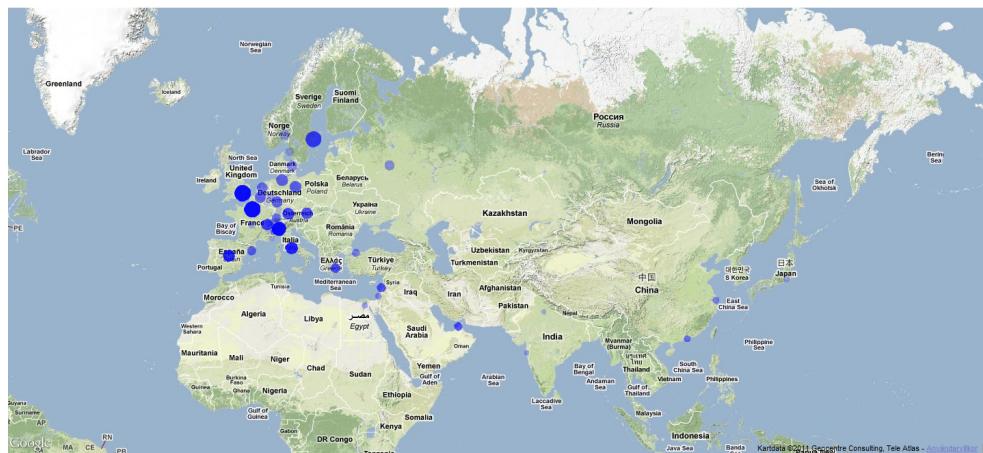
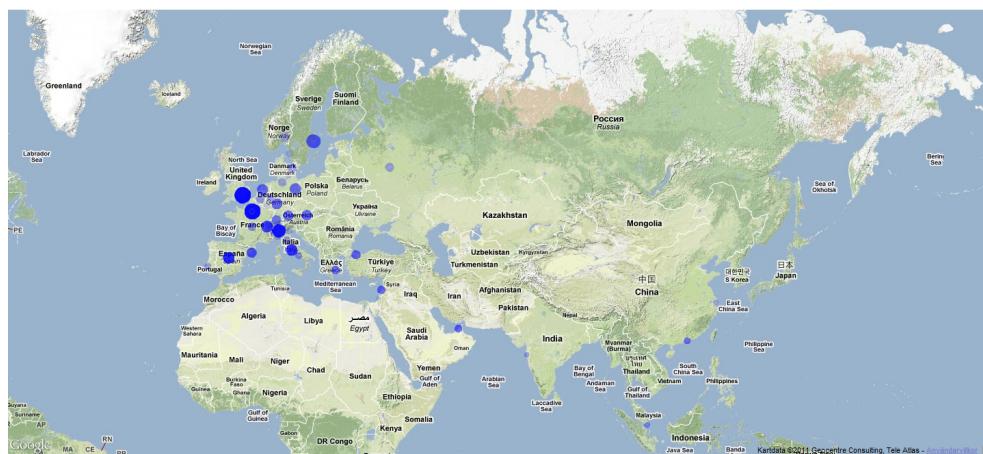
- **Swine Flu**

The "swine flu", the H1N1 influenza strain is the first condition deemed a global pandemic since the Hong Kong flu of 1967 to 1968. According to the latest WHO statistics, the virus has killed more than 18,000 people since it appeared in April 2009 and were 622,482 infected (approximate).

time span: November 1 2008 - February 31 2009

geographical area: global

total number of threads:



interpretation: probably not correlated

We could not conclude whether the event and the forum activity were correlated as the time span is probably too big to make any distinction

- **Michael Jackson's death**

Michael Jackson the most famous US singer dies after he suffered cardiac arrest at his home in the Holmby Hills neighborhood in Los Angeles, California. Later the Los Angeles County Coroner declared Jackson's death a homicide caused by the combination of drugs in his body. Conrad Murray is the doctor that had to take care of Michael's health before his death, and in 2011 he is still under charge of involuntary manslaughter. Internet traffic reaches unprecedented levels after entertainer Michael Jackson's death triggers an outpouring of worldwide grief.

time span: June 25 - July 2

geographical area: USA

total number of threads: 38724

interpretation: maybe correlated

There's so much data for this period, that the browser can not execute all the data. We could not determine whether this actually real data and related to MJ's death or if there's something wrong with the database at this point.

4 Conclusion

Out of the 15 events we were looking at, we rated

- 6 as 'correlated' or 'probably correlated'
- 8 as 'not correlated' or 'probably not correlated'
- 1 as 'maybe correlated'

which shows that most of the events would not correlate with forum activity.

However, during our research we discovered that there's probably not enough users in eastern regions so that we should also look at the outcome as if we only looked at european- and us-affecting events. Out of the 9 events in the US and central Europe, we rated

- 5 as 'correlated' or 'probably correlated'
- 3 as 'not correlated' or 'probably not correlated'
- 1 as 'maybe correlated'

We also saw that correlation decreases with time - events in 2004 are more likely to be correlated to real world events than events in 2008 and 2009. This is probably due to a decreasing popularity of the network as well as improved news sites that allow people to retrieve information and discuss outside the network. Also, during

We therefore conclude that there was a strong correlation between real world events and forum activity on ASW in the first few years but now, that it has become very easy to blog and discuss ideas everywhere on the web, it has become less popular to discuss real world events on ASW.