Fake news

Exploring and visualizing cases from euvsdisinfo.eu

Jesper Henrichsen

IT University

December 10, 2017

Problem statement

- How can it be visualized where and when fake news are created, and what countries are victims in terms of being mentioned in the content?

Problem statement

- How can it be visualized where and when fake news are created, and what countries are victims in terms of being mentioned in the content?
- What are the patterns that can be seen if social media data is used as scale of impact of fake news stories?

Resource euvsdisinfo.eu

- Campaign setup by the *European External Action Service East Stratcom Task Force*

Resource euvsdisinfo.eu

- Campaign setup by the *European External Action Service East Stratcom Task Force*
- Composed of 14 full time staff, and a network of 400 journalists, NGO's, Think Tanks etc.

Resource euvsdisinfo.eu

- Campaign setup by the *European External Action Service East Stratcom Task Force*
- Composed of 14 full time staff, and a network of 400 journalists, NGO's, Think Tanks etc.
- Objective to "challenge Russia's ongoing disinformation campaigns"

Resource euvsdisinfo.eu

- Campaign setup by the *European External Action Service East Stratcom Task Force*
- Composed of 14 full time staff, and a network of 400 journalists, NGO's, Think Tanks etc.
- Objective to "challenge Russia's ongoing disinformation campaigns"
- Started in 2015 and ongoing

Resource euvsdisinfo.eu

- Campaign setup by the *European External Action Service East Stratcom Task Force*
- Composed of 14 full time staff, and a network of 400 journalists, NGO's, Think Tanks etc.
- Objective to "challenge Russia's ongoing disinformation campaigns"
- Started in 2015 and ongoing
- About 3400 news flagged as fake news

Example of cases at euvsdisinfo.eu

02.11.2017	Brussels are closing the door they opened with visa freedom for Georgians	Rezonansi	Europe, Georgia
02.11.2017	EU wants to ban information about the of country of origin on the food labels	Vlastenecké Noviny, eOdborar.cz	Italy, EU
02.11.2017	The West supported the terrorists during hostage crisis in Dubrovka in Moscow ("Nord-Ost attack") in 2002 and in the school of Beslan, North Ossetia, in 2004	Vremya pokazhet @Pervyi kanal, 19:15	Russia, The West
02.11.2017	The US destroyed the European values and culture, so now Russia is the only flagship of the European civilization	Vremya pokazhet @Pervyi kanal' TV-channel, 41:08	Europe, US
01.11.2017	The EU teaches journalists how to properly inform about Islam and migrants	ac24.cz	Ireland, Italy, Austria, Slovenia, Hungary, Greece, EU, Germany, Spain
31.10.2017	Finland wants Russia to join the European Centre of Excellence for Countering Hybrid Threats	Sergei Lavrov, Russian Foreign Ministry's website	Russia, Finland
31.10.2017	The West, primarily the United States, is collecting biological material in Russia to create a biological weapon that destroys the Russians.	Mesto vstrechi @NTV TV-channel, 1:12:48	Russia, The West, US
31.10.2017	Czech MP prefers there were 5 million Muslim migrants in Czech Republic rather than 5 million voters of Czech president Miloš Zeman	BezPolitickeKorektnosti	Czech Republic
31.10.2017	Estonia has opened a new military base in the town of Tapa under the pretext of fear of a Russian attack.	cz.Sputniknews	Russia, Baltic states, Estonia

Example of cases at euvsdisinfo.eu

EU wants to ban information about the of country of origin on the food labels

Summary of Disinformation

EU wants to ban information about the of country of origin on food labels. This measure is applied in order to fight against the nationalistic tendencies in member states.

VIEW ORIGINAL PUBLICATION / MEDIA

Reported in:
Issue 86

Date:
02.11.2017

Language:
Czech

Country of Origin:
Italy, EU

Reported by:
Prague Security Studies Institute (PSSI)

Keywords:
European Commission

Disinforming outlet:
Vlastenecké Noviny, eOdborar.cz

Disproof

The original text from Italian website Afffaritalini is about a quarrel between the European Commision and Italy, since Italy started to demand from producers to write on food labels where the product was produced and packed.

According to the European Commission, the current approach is that the country of origin or place of provenance labelling on food is voluntary, unless its absence could mislead consumers.

The Regulation introduces mandatory origin labelling for fresh meat from sheep, goat, poultry and pigs. As of 1 April 2015, with some exemptions, the Member State or third country where the animal was reared and slaughtered will appear on the label of such meats.

For foods bearing origin indications, the country of origin or place of provenance of the main ingredients must also be listed if those ingredients originate from a different

Example of cases at euvsdisinfo.eu

První českojazyčné noviny | Založeno v roce 1789 | Vydavatel: Česká expedice, z. s. | ISSN 2533-4328 | Web je monitorován fízly a cenzory teroristického centra CTHH

VLASTENECKÉNOVINY.CZ

NEOTŘELÉ INFORMACE BEZ MASKY PRO LIDI S VLASTNÍM POSTOJEM

EVROPA, ZAHRANIČNÍ, ZPRÁVY

Výplod bruselského mozku: Označování země původu potravin musí být zakázáno. Je to nacionalistické

Autor článku: -byi- • 02. 11. 2017 v 07:45

To se mi líbí 1 tis.

ŘÍM | EU požaduje zákaz označení země původu na výrobcích. Prý je to nacionalistické, informuje server Českého portálu eOdborar.cz s odkazem na italský affaritaliani.it.

Dokud budu prezidentem, žádné předčasné volby nebudou

Zloděj Bakala za pomocí ožralé pizdy Marksové vede pomlouvačnou štvanici proti SPD

POLITICKÁ ARÉNA

Podpoří Babiš a Okamura Zemana v prezidentských volbách? A co za to?

Další sloupky →

Páteční glosy 10. 11. 2017

Další glosy →

Scraping

For every case in the list of cases

```
1  def get_all_cases(url=None, start_pos=0, end_pos=4000):
2      if url is None:
3          url = 'https://euvsdisinfo.eu/disinformation-cases/?
               offset=%s'
4      cases = []
5      for page in range(start_pos, end_pos, 10):
6          time.sleep(1)
7          print("Get page: %s" % (url % page))
8          page_html = req.get(url % page)
9          soup = bs(page_html.text, 'lxml')
10         case_links = soup.find_all("div", class_='disinfo-db
               -cell cell-title')[1:]
11         for link in [ x.find('a').attrs['href'] for x in
               case_links ]:
12             cases.append(link)
13
14     print("Number of cases found: %s" % len(cases))
15     return cases
```

Scraping

Scrape the article info

```
1       fields = {}
2
3       fields['title'] = soup.find("h2").text.strip()
4       main_info = soup.find("div", class_="report-main-info")
5       fields['summary'] = main_info.find("div", class_="report
            -summary-text").text.strip()
6       fields['disproof'] = main_info.find("div", class_="
            report-disproof-text").text.strip()
7       fields['source'] = main_info.find("div", class_="report-
            disinfo-link").find("a").attrs['href']
```

Scraping

Wrapped in commandline interface

```
$ ./news_scraper.py -h
    usage: news_scraper.py [-h] [-o o] [-p p] csv

    A scraper for the fake news available at
    euvsdisinfo.eu.

    positional arguments:
      csv                 The csv file to write output to

    optional arguments:
      -h, --help          Show this help message and exit.
      -o o, --offset o    The offset page to start scraping
                          news.
      -p p, --pages p     The number of pages to fetch,
                          each page includes at most
                          10 fake news cases.
```

Scraping

Including description from meta tags

```
1  metatag_list = [
2      ("name", "description", "content"),
3      ("property" "og:title", "content"),
4      ("property", "og:description", "content"),
5      ("name", "twitter:title", "content"),
6      ("name", "twitter:description", "content"),
7      ("name", "language", "content"),
8      ("name", "keywords", "content"),
9      ("name", "subject", "content"),
10     ("name", "topic", "content"),
11     ("name", "summary", "content"),
12     ("name", "subtitle", "content"),
13     ("itemprop", "name", "content"),
14     ("itemprop", "description", "content"),
15 ]
```

Location tagging using Stanford Named Entity Recognition

- Pretrained models for arabic, chinese, english, french, german, and spanish. (No russian).

Location tagging using Stanford Named Entity Recognition

- Pretrained models for arabic, chinese, english, french, german, and spanish. (No russian).
- Attempts to recognize locations and people in text.

Location tagging using Stanford Named Entity Recognition

Concatenating all available text for an article and feed to NER tagger

```
1  for line in csvreader:
2      content = ""
3      content += " " + line[header['keywords']]
4      content += " " + line[header['title']]
5      content += " " + line[header['summary']]
6      content += " " + line[header['metatags_content']]
7      tagged = tagCell(content)
8      locations    = toString(filterTags(tagged, ["LOCATION"]))
9      miscs        = toString(filterTags(tagged, ["MISC"]))
10     people       = toString(filterTags(tagged, ["PERSON"]))
```

Location tagging using Stanford Named Entity Recognition

title	**language**	**locations**
Crimea decided its own fate	Russian	Crimea,Russia,China,India
Ukraine is a neo-Nazi state	Russian	Ukraine,West
NATO is encircling Russia.	Czech	Russia
Ukraine is governed by Nazis.	Russian	Ukraine
Sweden wants to leave the EU.	Spain	Sweden
Georgia is a US colony.	Georgian	Georgia,US
Ukraine has a Nazi identity.	Russian	Ukraine
Ukraine is a part of Russia.	Russian	Ukraine,Russia
Savchenko is a US spy.	Russian	US
Nazis control Ukraine.	Russian	Ukraine
NATO kills Serbian children.	English	PravdaReport,West,Russia,Having, Syria,russia,Serbia,Yugoslavia
Ukraine is governed by nazis.	Russian	Ukraine
All Ukraine is Russia.	Russian	Ukraine,Russia

Location tagging using Stanford Named Entity Recognition

language	% of articles
Russian	55
Czech	16
English	9
Slovak	2
Georgian	2

location	% of articles
Ukraine	27
Russia	23
USA	21
Europe	8
Syria	5

Visualizing

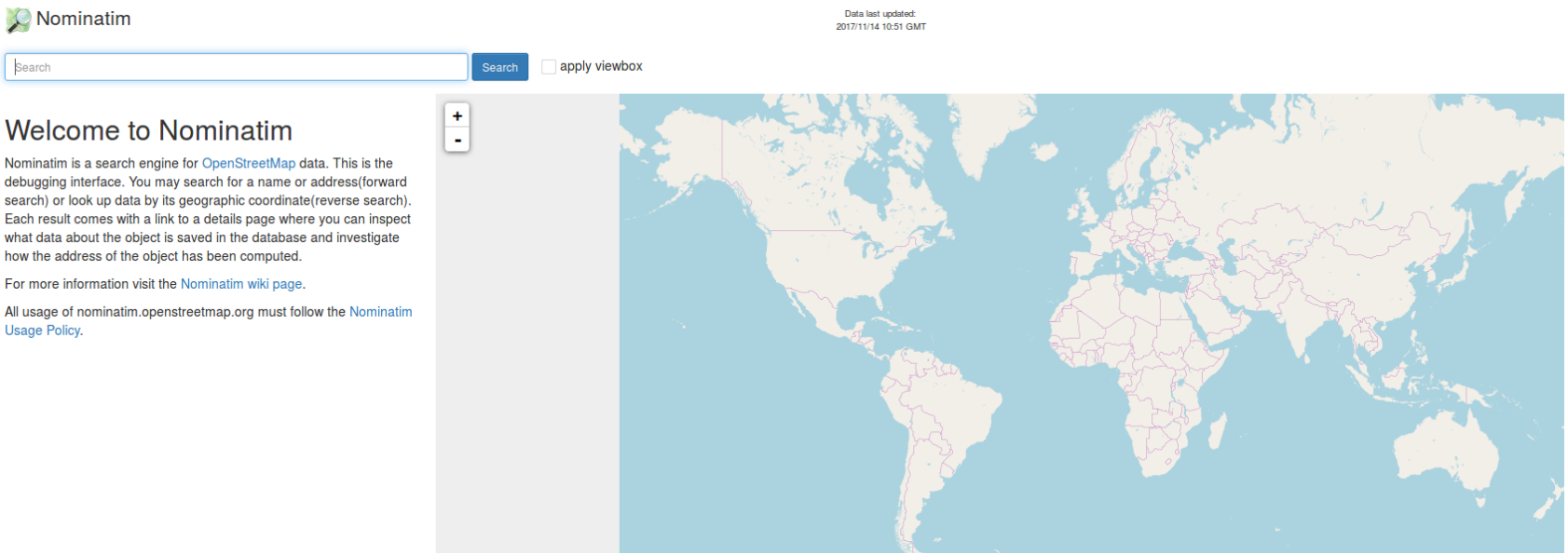
Translating location names to coordinates

Nominatim

Searchengine for searching OpenStreetMaps data by name.

Visualizing

Translating location names to coordinates



Visualizing

Weigh stories by their social media spread

Facebook graph API

```
1       print(" " * indent + "}")
2
3   def get_shares(url):
4       enc_url = encode(url)
5       access_token = get_access_token()
6       fields = "og_object{engagement}"
```

Visualizing

Weigh stories by their social media spread

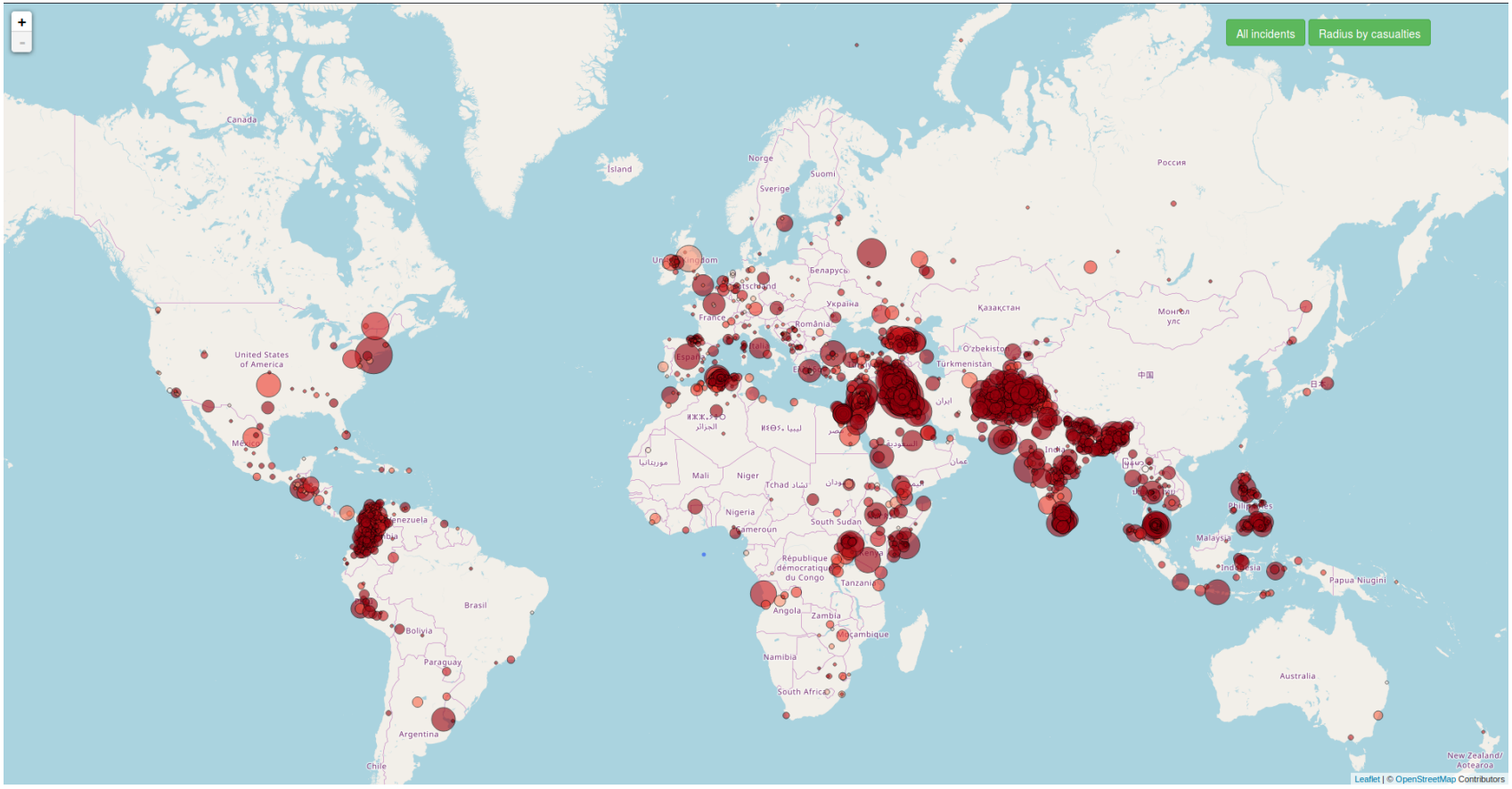
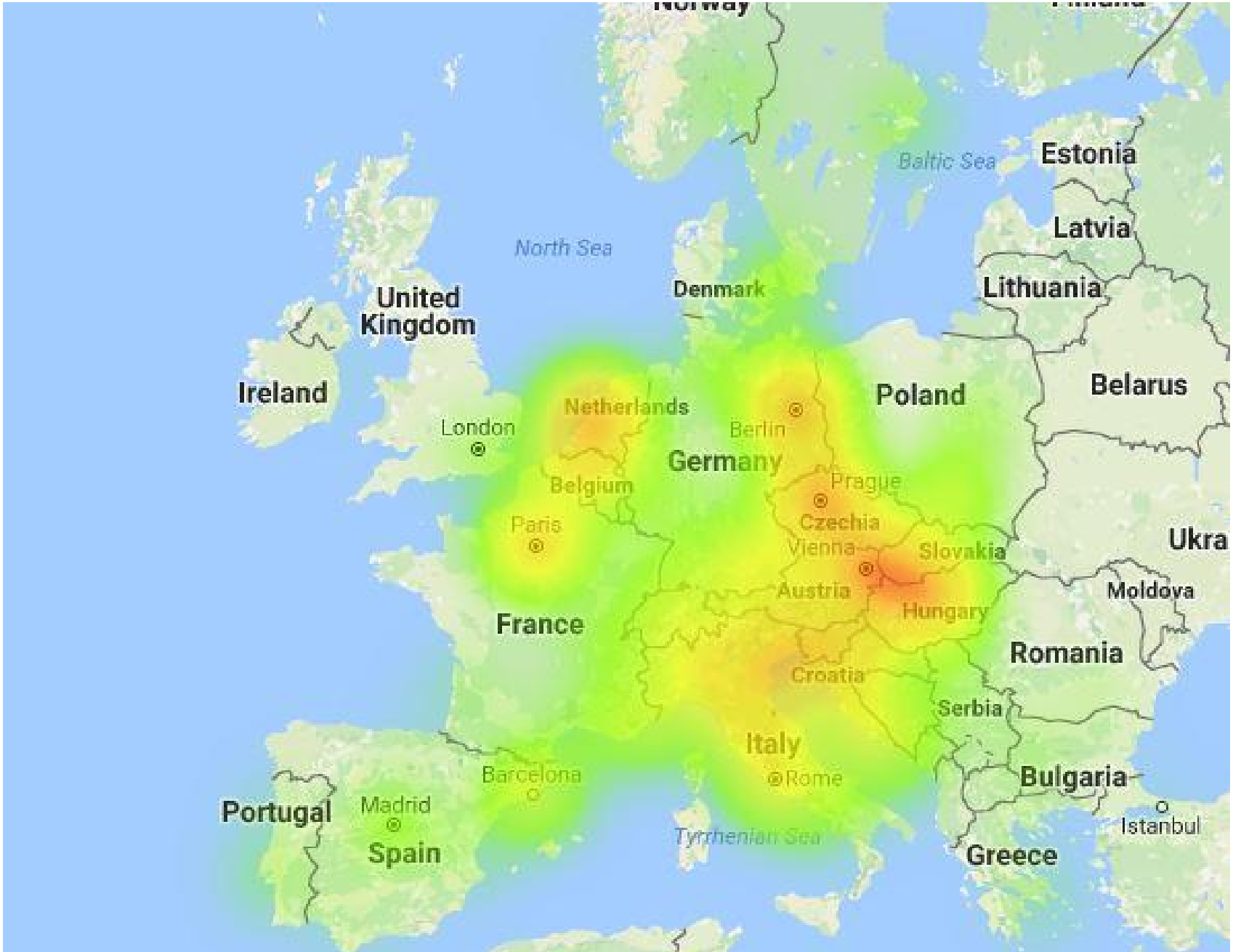
Facebook graph API

```
In [3]: run requests_fb.py

In [4]: get_shares("http://www.vlasteneckenoviny.cz/?p=190576")
{
  id : http://www.vlasteneckenoviny.cz/?p=190576
  og_object : {
    engagement : {
      social_sentence : 1K people like this.
      count : 1029
    }
    id : 1537911602953150
  }
}
```

Visualizing

Visualizing



Visualizing