

# **Digital Social Research**

## **an introduction**

**it will introduce you to the ways in which  
important social issues can be researched  
on the basis of digital traces of human activity**

**Daniel Alexandrov  
Irina Busurkina**

## **TWO PARTS OF THE PRESENTATION**

- 1. What digital social research is about? An example**
  
- 2. How the course will work?  
Themes, projects, grades...**

# The example of digital research

**How to turn noisy chat on Twitter into the data on well-being, stress and mortality?**

**You can look it up and read the materials on their site. It's great work done at UPenn**

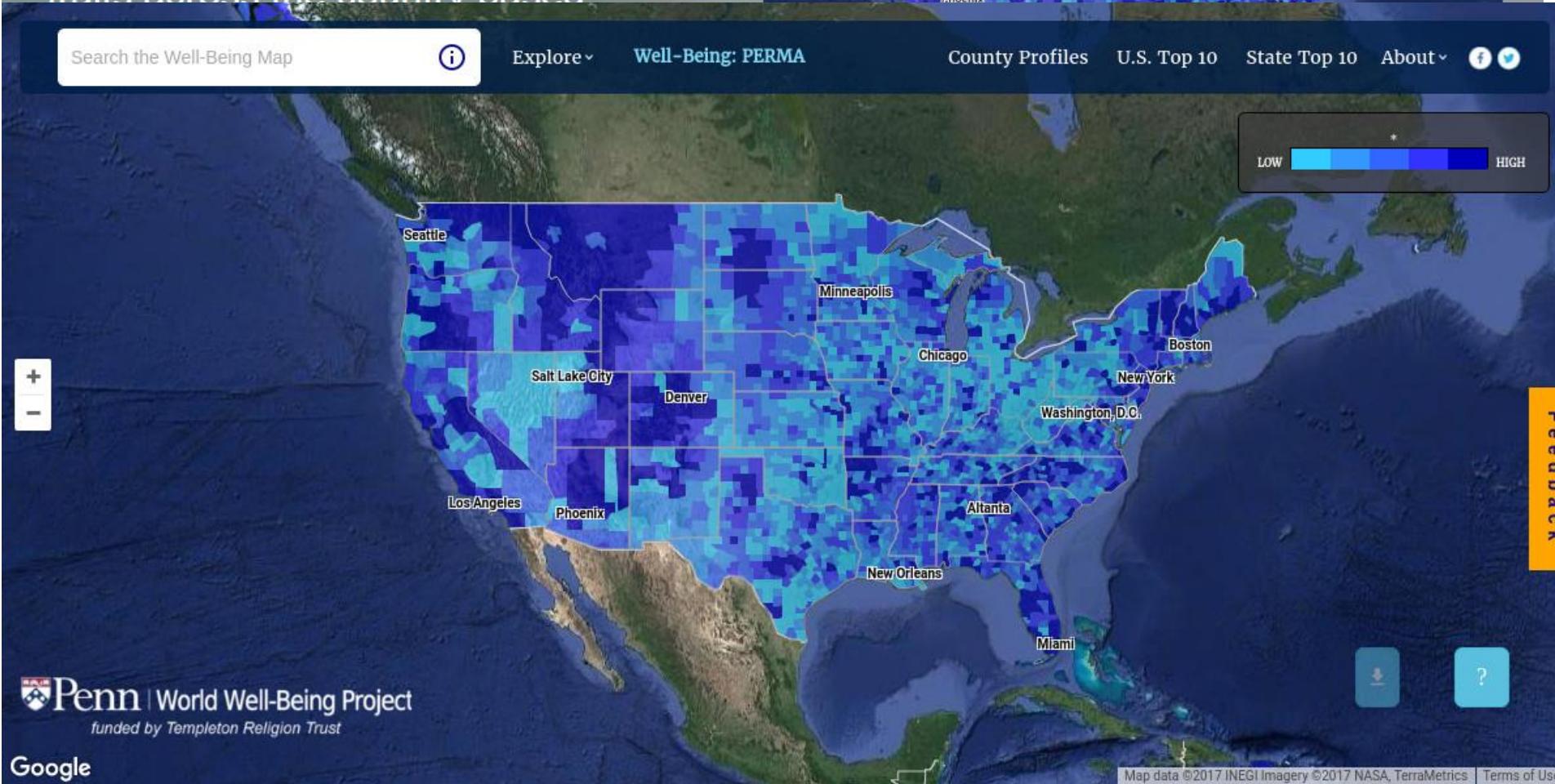
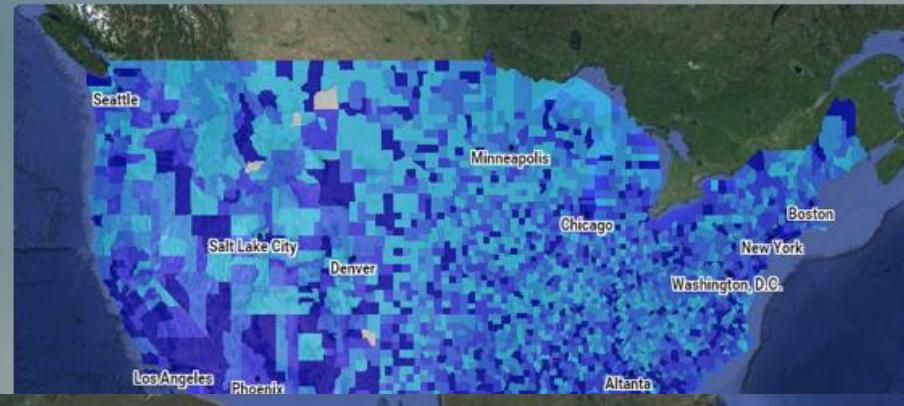
**WELCOME TO THE WORLD WELL-BEING PROJECT**

Pioneering techniques for measuring psychological well-being and physical health based on language in social media



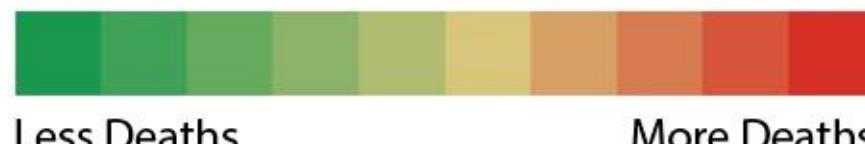
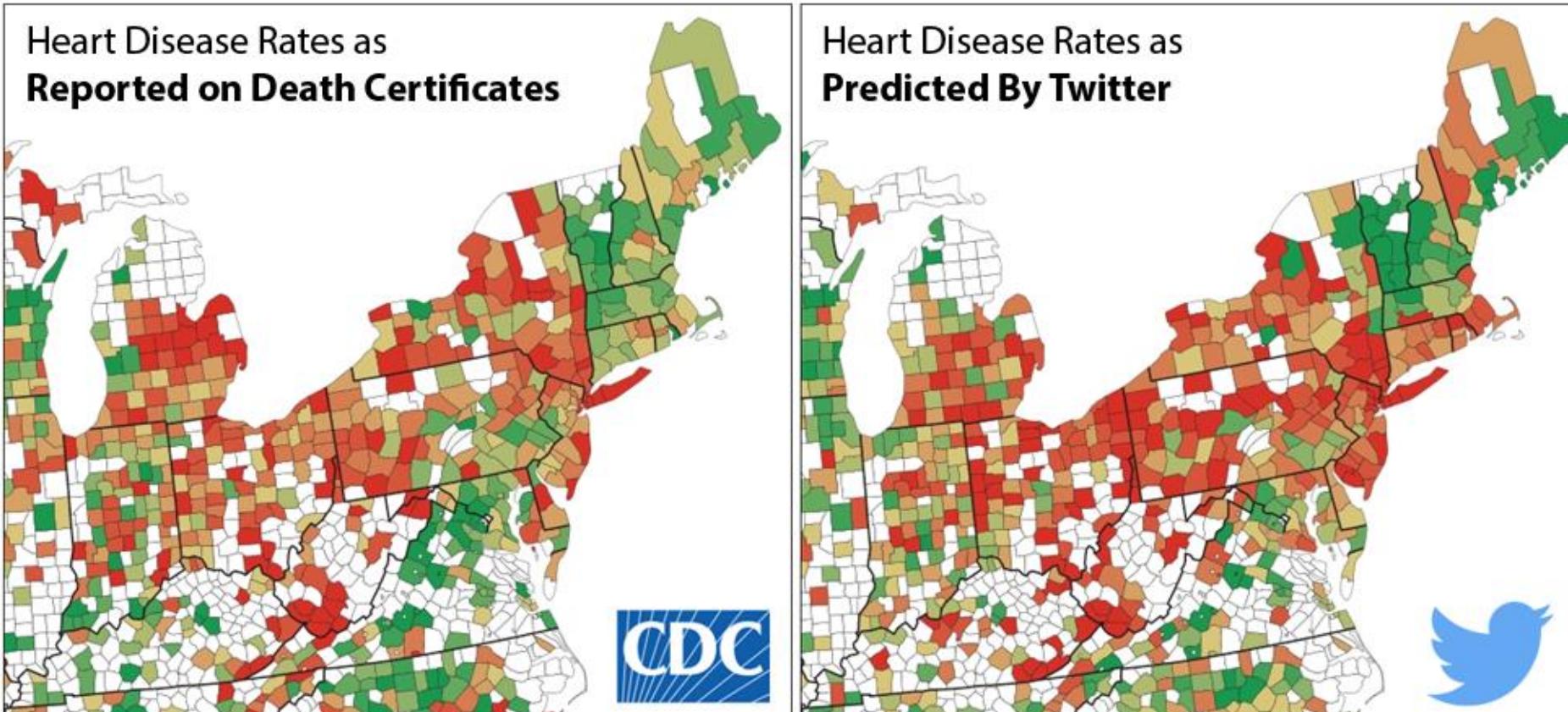
# THE WELL-BEING MAP IS HERE!

Click here to explore our interactive map that displays measurements for different traits across the country based



# PUBLIC MOOD and DEATH FROM HEART DISEASE

## correlation of Twitter data and ground truth data on aggregate level of counties



# PUBLIC MOOD and DEATH FROM HEART DISEASE

## Twitter topics predictive of higher death rate

Twitter Topics Positively Correlated With County-Level AHD Mortality

Hostility,  
Aggression

bullshit  
shits  
bitches  
**fuck**  
**fuckin**  
**damn**  
**fucked**  
fucks  
**fucking**  
**ass**  
**shit**  
shitty  
dude  
pissed

$r = .18$

dick  
motherfucker  
**ass**  
pussy  
**fuckin**  
asshole  
**bitch**  
bitches  
**shit**  
**fuck**  
dumb  
assholes  
cunt

$r = .21$

**fuck**  
shitty  
**bitch**  
idiot  
bitches  
annoying  
bullshit  
**stupid**  
**retarded**  
pissed  
**hate**  
kidding  
**shit**

$r = .27$

Hate,  
Interpersonal  
Tension

jealousy  
mad  
bitches  
envy  
**hate**  
lovers  
**haters**  
famous  
hater  
phase  
hated  
ya'll

$r = .16$

nasty  
pieces  
**games**  
faced  
fake  
**head**  
**bullshit**  
shit  
**drama**  
liars  
sneeze

$r = .17$

grr  
**passion**  
grrr  
pit  
absolutely  
officially  
**hate**  
mondays  
burning  
**hates**  
despise  
mentioned  
**hating**  
**fuckin**

$r = .21$

Boredom,  
Fatigue

sooooooo  
boring  
**text**  
entertain  
insanely  
yawn  
entertainment  
**extremely**  
**bored**  
stiff  
boredom  
incredibly  
entertained  
**bore**

$r = .18$

sore  
bed  
**soooo**  
extremely  
sleep  
**tired**  
sooooo  
freaking  
yawn  
**soooo**  
tire

$r = .18$

**bed**  
bath  
goodnight  
**tired**  
curl  
sleepy  
**sleep**  
laying  
outta  
**ready**  
exhausted  
**crawl**  
**shower**  
layin  
cuddle

$r = .20$

# PUBLIC MOOD and DEATH FROM HEART DISEASE

## Twitter topics predictive of lower death rate

Twitter Topics Negatively Correlated With County-Level AHD Mortality



# **Our course will introduce you to the studies of important social issues based on digital traces of human activities**

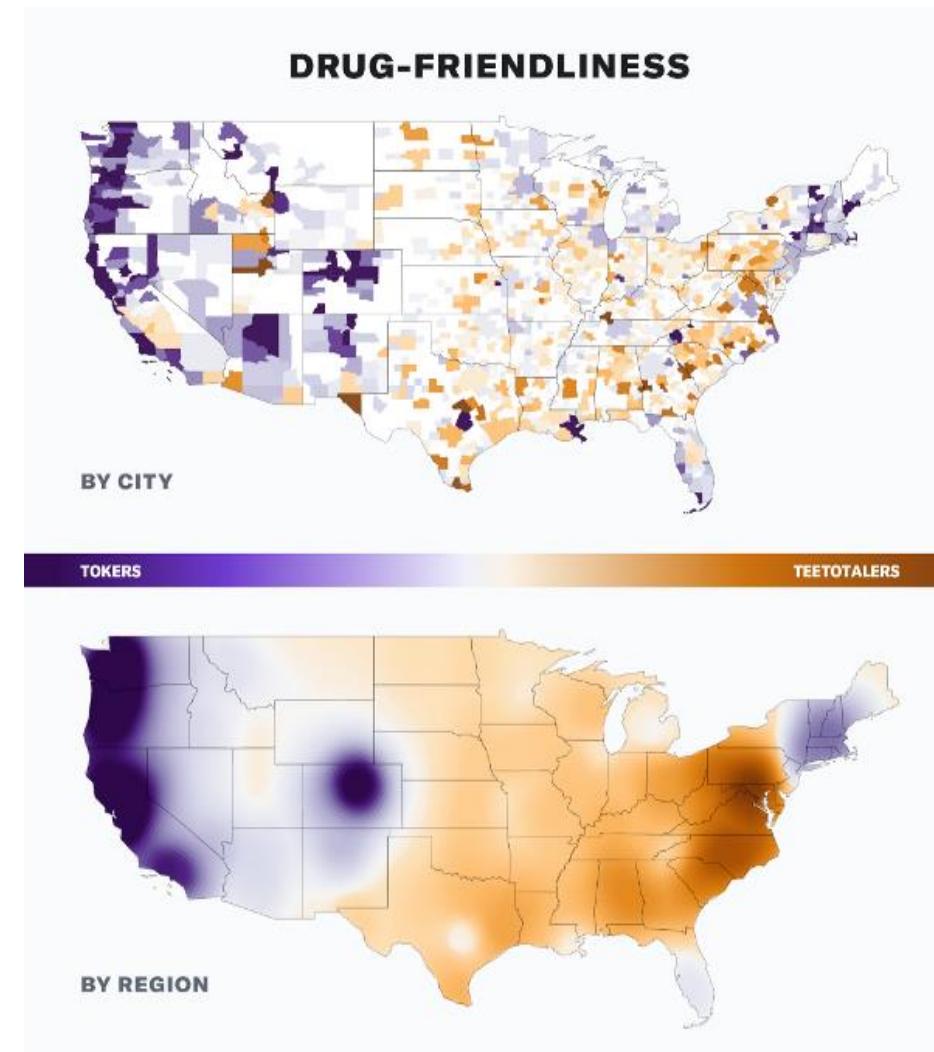
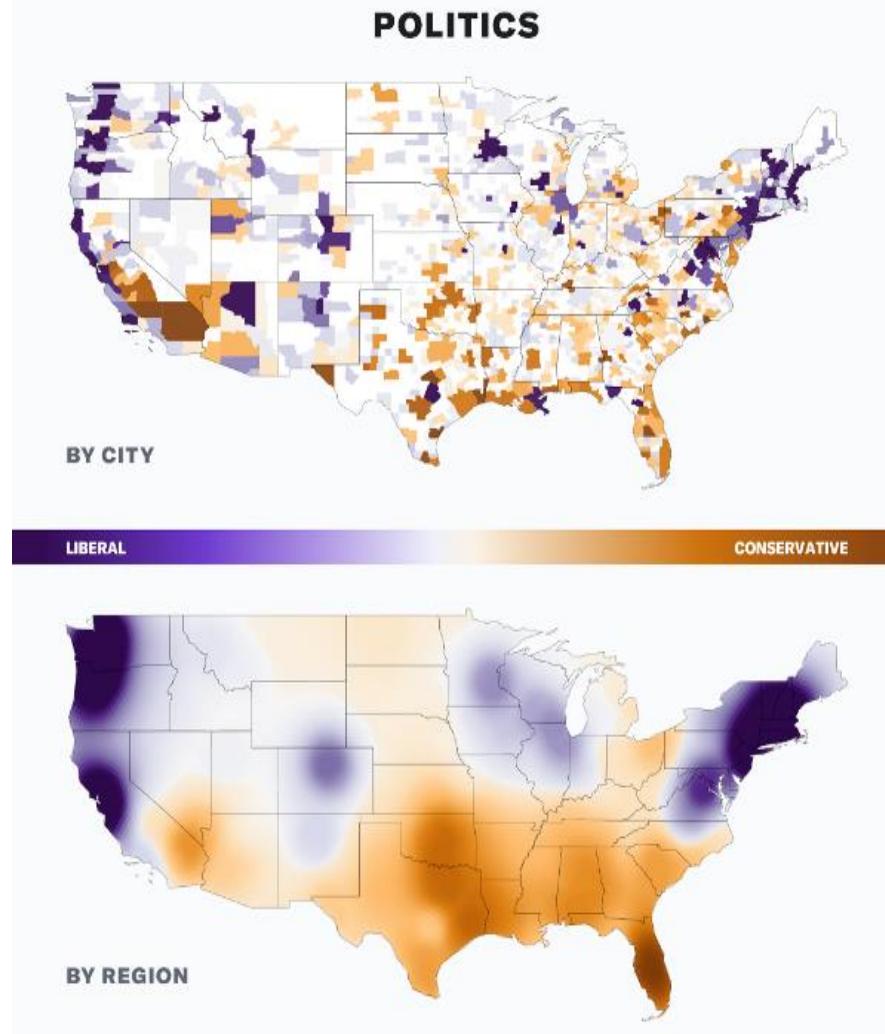
- **it will be hands-on project work and seminar discussions**
- **focus on issues rather than data techniques**
- **no prerequisites**
- **no need to know R or statistics**

**In order to have real hands-on group work we will focus on four particular research questions to answer using digital data.**

**First, we use secondary data to analyze and discuss two questions:**

- **Which cultural regions in the US we can distinguish? We will focus on the US because it has a lot of data published already**
- **Which macro-regions in the world we can distinguish based on the similarities of nations as revealed by digital data?**

**You will study these maps (the data is from online dating site OKCupid) and similar sources in order to build your own vision of political and cultural division of the United States.**



You will read several articles like the one below and compare the results in order to build your own vision of political and cultural division of world into macro-regions

Samoilenko et al. *EPJ Data Science* (2016) 5:9  
DOI 10.1140/epjds/s13688-016-0070-8



REGULAR ARTICLE

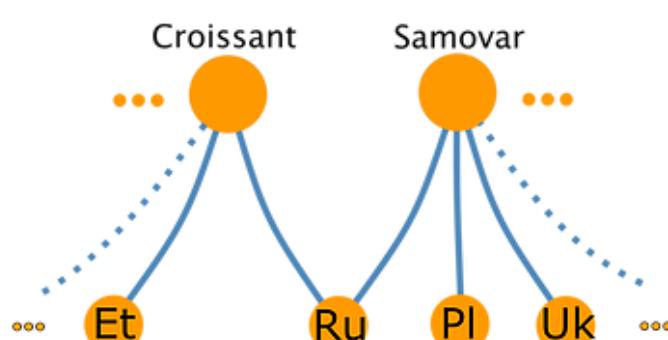
EPJ Data Science  
a SpringerOpen Journal

Open Access

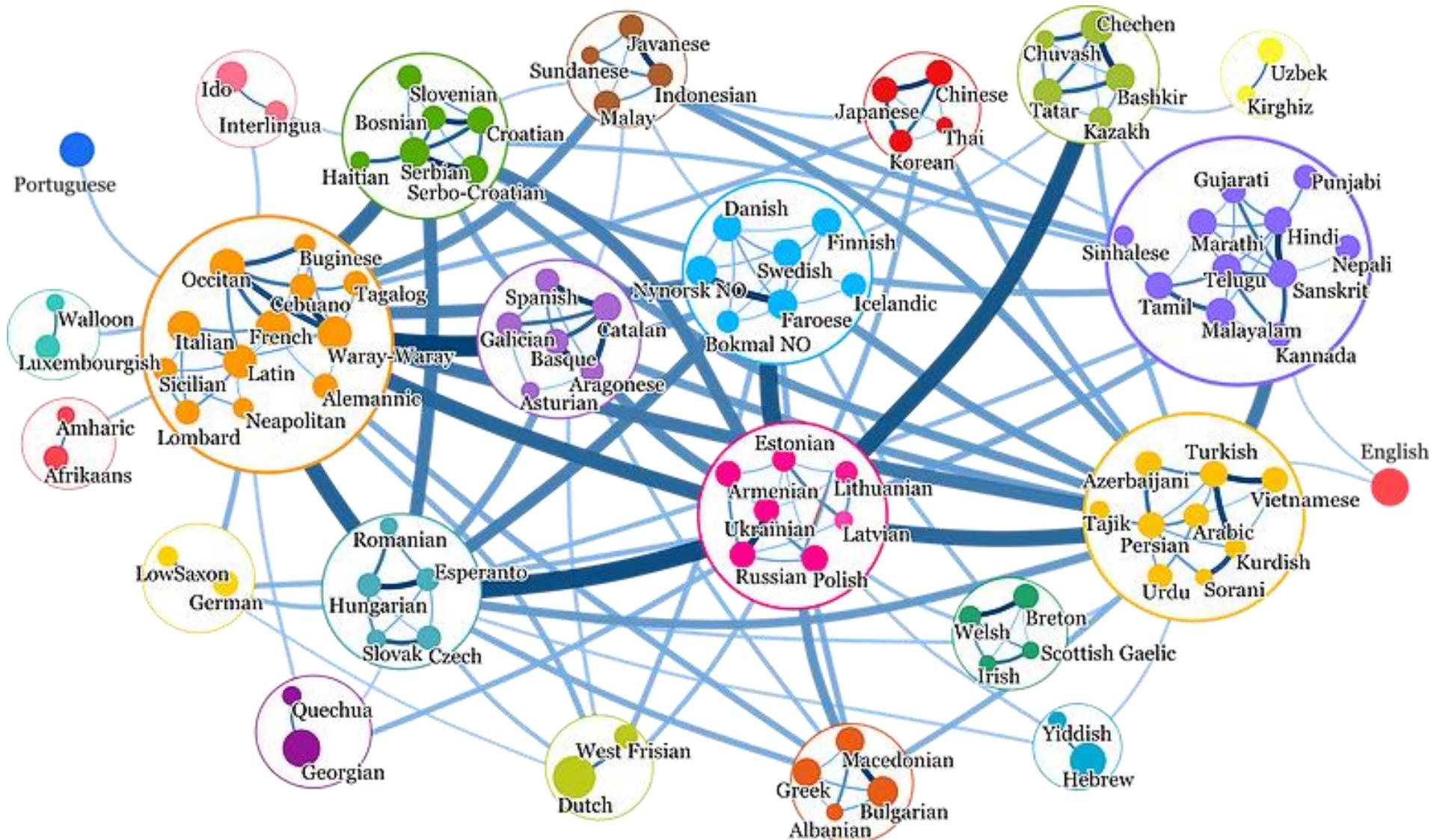


## Linguistic neighbourhoods: explaining cultural borders on Wikipedia through multilingual co-editing activity

Anna Samoilenko<sup>1,2\*</sup>, Fariba Karimi<sup>1</sup>, Daniel Edler<sup>3</sup>, Jérôme Kunegis<sup>2</sup> and Markus Strohmaier<sup>1,2</sup>



# THE CLUSTERS OF NATIONS IN COEDITING IN WIKIPEDIA



# MACRO-REGIONS AS SEEN BY COEDITING IN WIKIPEDIA

Mapping bilateral information interests using the activity of Wikipedia editors

Fariba Karimi,<sup>1,\*</sup> Ludvig Bohlin,<sup>2</sup> Anna Samolenko,<sup>1</sup> Martin Rosvall,<sup>2</sup> and Andrea Lancichinetti<sup>2</sup>

<sup>1</sup>Leibniz Institute for the Social Sciences, Cologne, Germany

<sup>2</sup>Integrated Science Lab, Department of Physics, Umeå University, SE-901 87 Umeå, Sweden

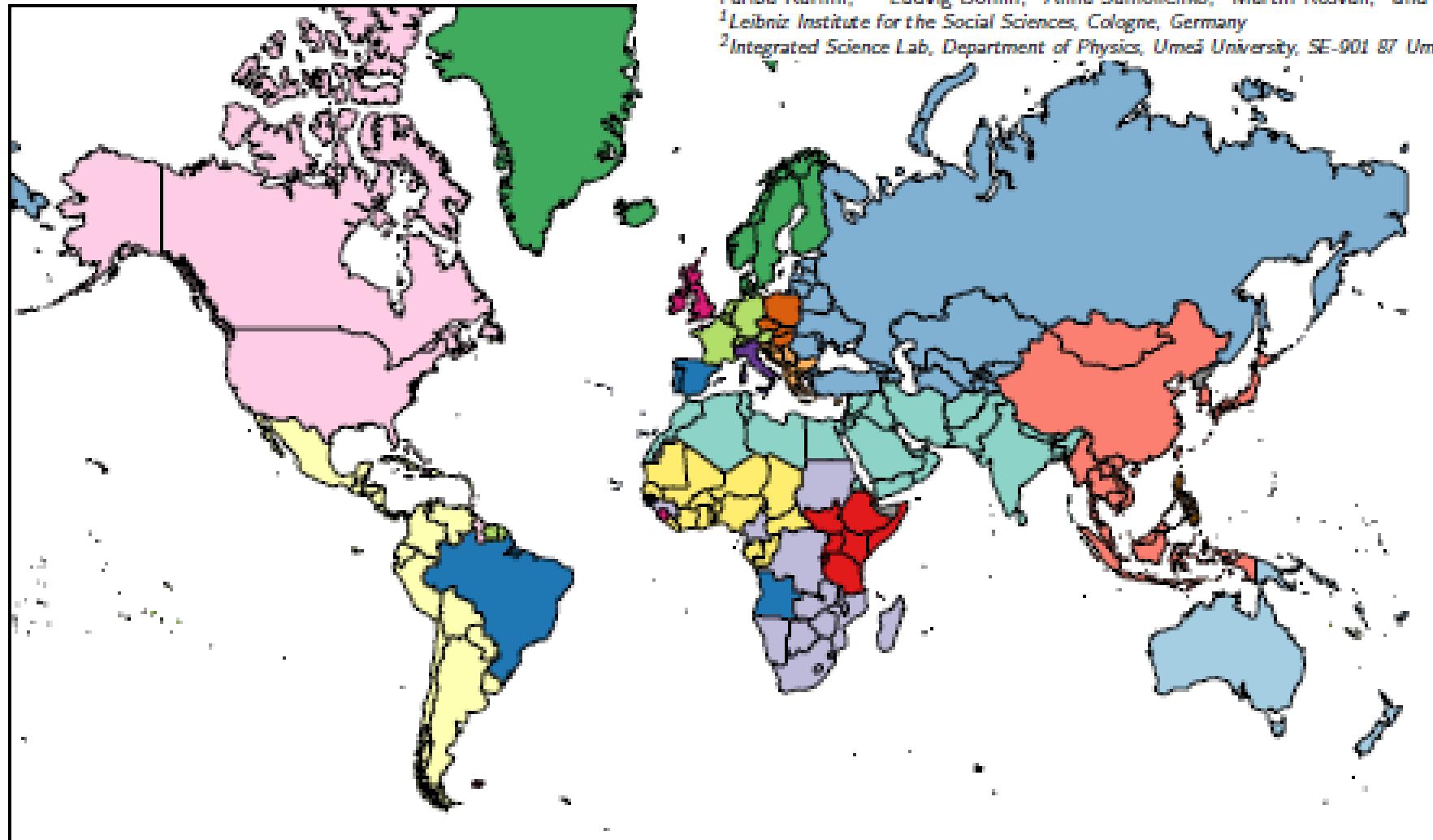


Figure 2 | World map of Information Interests. Countries that belong to the same cluster have the same color. Countries colored with gray do not belong to any cluster. The map suggests that the division of countries can be explained by a combination of cultural and geopolitical features.

**Second, you will collect your own data to study two questions:**

- How one can use **Яндекс Поиск** and **Google Search** in studying social and economic problems / questions. We will focus on Russia and its regions.
- How people perceive privacy and security on social network sites, and how they present themselves on their VK pages.

## **Tasks:**

- **You will have to do your homework personally and submit it in a written form**
- **You will work in groups to discuss group results and then present them at a seminar**
- **In collecting original data you will have to do your share of collecting, but the analysis will be done in groups**

## **Grading:**

- **You will get points for your individual homework submitted in a written form**
- **You will get points for group work**
- **You will get points for presentations and discussion at seminars**
- **You will get points for writing a final essay**

## **Advantages:**

- **You will learn to use various digital sources to study important problems, and it will be useful for any future job dealing with digital data**
- **You will learn important things about the world (and Russia) and human behavior**
- **Our efforts in group work with original data – if you like it - may lead to a student group project**
- **You might think of a topic for your senior paper (курсовая) and your final thesis (BKP)**