

From individual behavior to social process: using human traces for social good

Esteban Moro

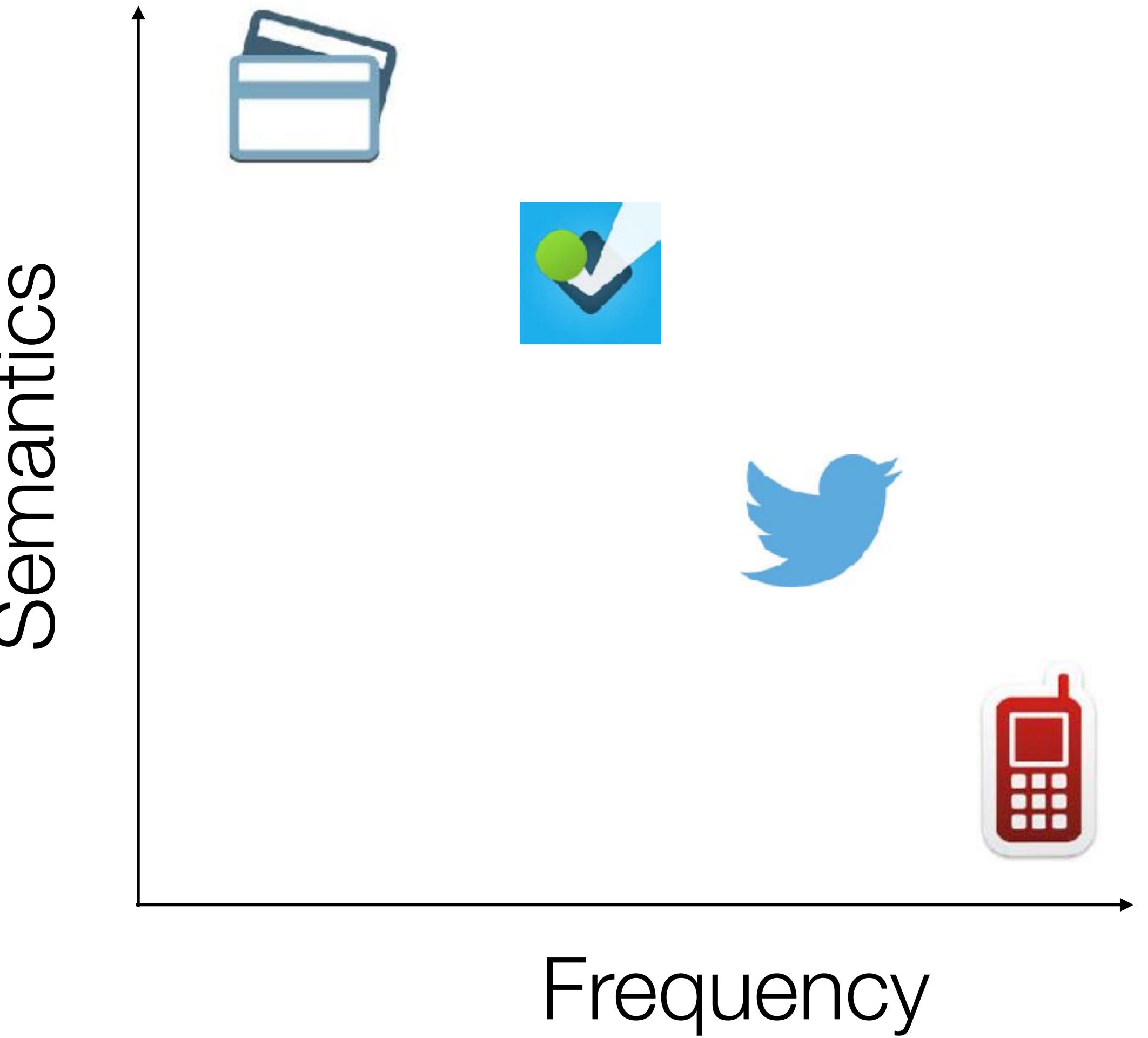


Universidad
Carlos III de Madrid

ICCS 2018

Sources of human traces

- Social networks:
 - Twitter, Facebook, Foursquare, etc.
- Financial data
 - Transfers
 - Credit card transactions
- Mobile phone:
 - CDRs (calls/SMS), network events, etc.
 - Phone sensors
 - Apps
- Satellite data



Why alternative data?

- Better spatial-temporal resolution
- Faster answers
- Cheaper
- Availability

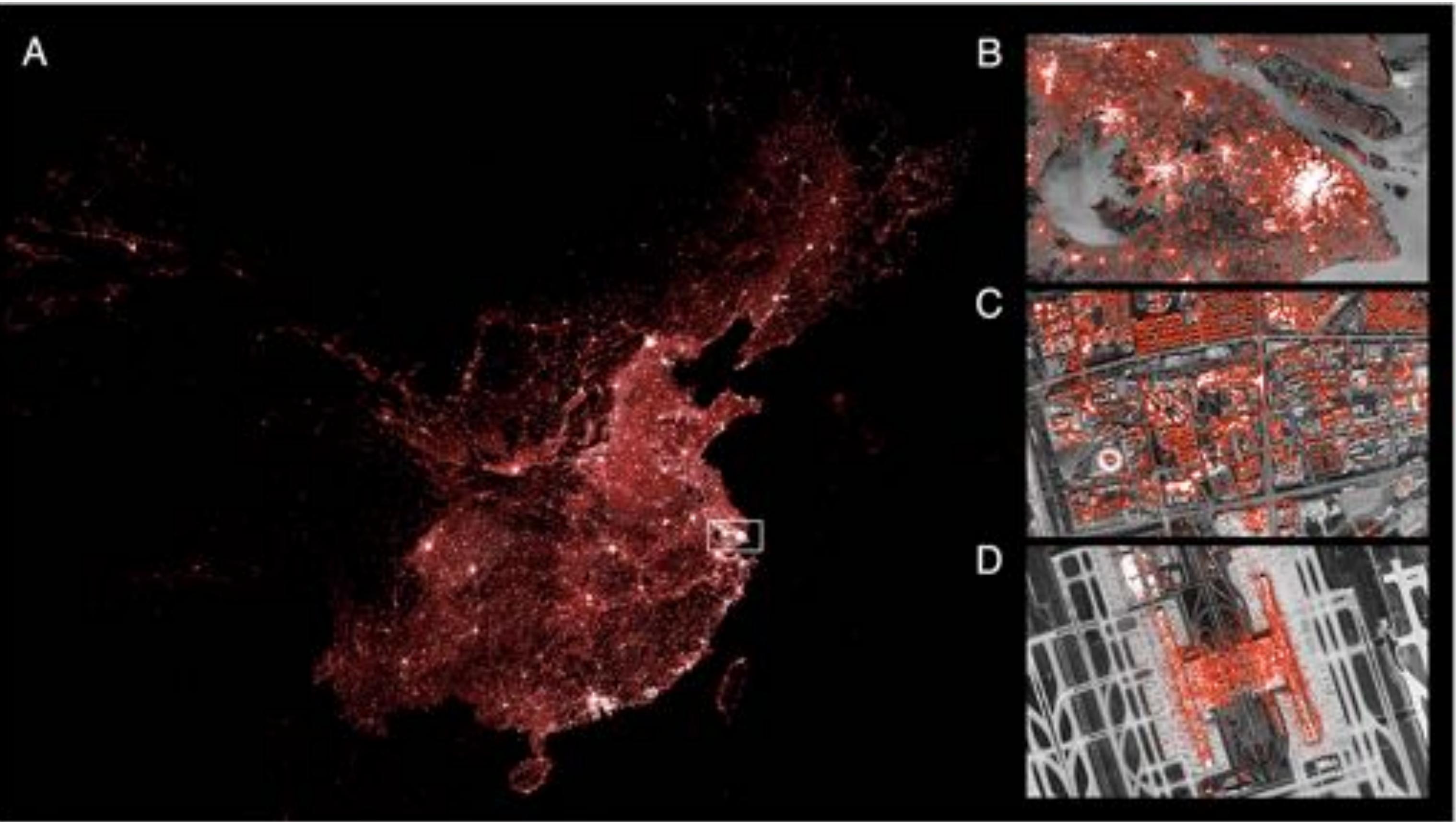
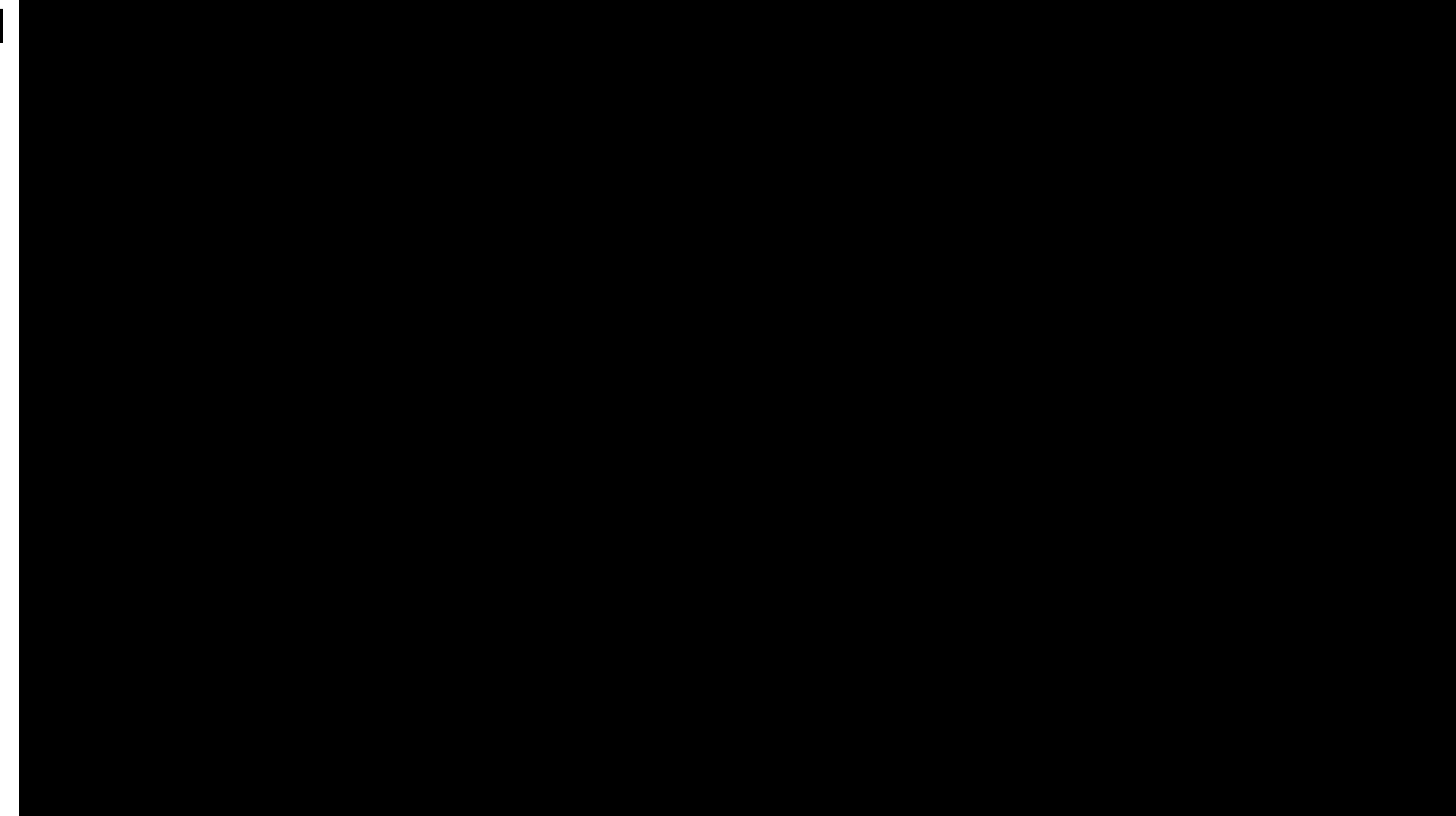


Figure 1 Spatial-temporal big data reflects human activity at different scales. **(A)** At the national level: Data points depict the fact that most of China's population is concentrated in large cities in the east. As the brightness of the spot increases, the aggregation of the data points (population) increases. **(B)** At the regional level: This figure shows urban clusters in the Yangtze River Delta. **(C)** At the zone level: Zhangjiang Hi-tech Park in Shanghai. **(D)** At the building level: Pudong Airport in Shanghai. Maps were created using C and Datamaps (<https://github.com/ericfischer/datamaps>), and the remote sensing images were derived from Baidu Maps.

Why alternative data?

Deville, P, et al. (2014). Dynamic population mapping using mobile phone data.
PNAS 111(45), 15888–15893. <http://doi.org/10.1073/pnas.1408439111>

- Better spatial-temporal resolution
- Faster answers
- Cheaper
- Availability
- Different questions



Why alternative data?

- Complements official data

The screenshot shows the homepage of the CROS website. At the top right, there is a search bar with the placeholder "Search" and links for "U.S. Department of Commerce | BI", "Legal notice", "Cookies", "Contact on Europa", and "Search on Europa". On the left side, there is a vertical sidebar with a decorative pattern and a list of categories: "A", "C", "C", "N", "C", "A", "R", "E", "F", "N", "L", "S", "D". The main content area features the European Commission logo and the CROS logo. The title "CROS" is displayed prominently, followed by the subtitle "Collaboration in Research and Methodology for Official Statistics". Below this, a breadcrumb navigation shows "European Commission > Eurostat > CROS > Big data > Papers on big data in official statistics". The main heading "Papers on big data in official statistics" is centered above a list of papers. A sidebar on the left contains a section titled "Big data" with links to various resources like "Assessing the Quality of Mobile Phone Data as a Source of Statistics", "BDCOMP", "Big Data Initiatives", "ESTP training courses", "Estatnetbigdata", "Ethical Review Big Data and official statistics", "European Big Data Hackathon", and "Events". To the right of the main content, there is a list of recommended papers.

U.S. Department of Commerce | BI
Legal notice | Cookies | Contact on Europa | Search on Europa

CROS

Collaboration in Research and Methodology for Official Statistics

European Commission > Eurostat > CROS > Big data > Papers on big data in official statistics

A TO Z GROUPS EVENTS NEWS HELP Login

Papers on big data in official statistics

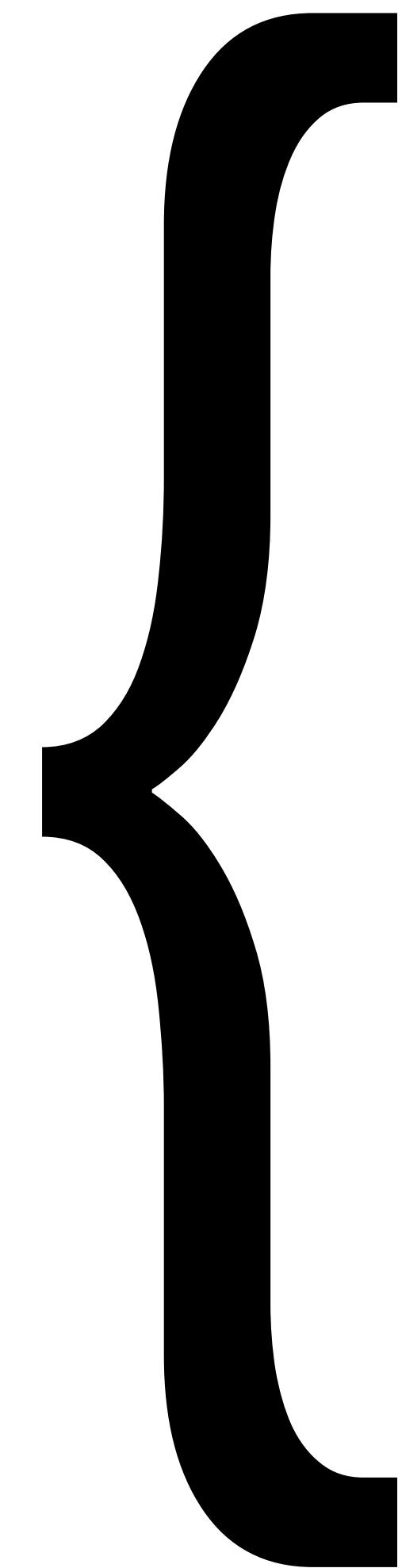
Big data

- Assessing the Quality of Mobile Phone Data as a Source of Statistics
- BDCOMP
- Big Data Initiatives
- ESTP training courses
- Estatnetbigdata
- Ethical Review Big Data and official statistics
- European Big Data Hackathon
- Events

Below, please find some interesting papers on big data with relevance to official statistics:

Beręsewicz, M. (2016) Internet data sources for real estate market statistics

- ESS Big Data Action Plan and Roadmap 1.0
- Business Case Vision Implementing Project Big Data
- Structuring risks and solutions in the use of big data sources for producing official statistics – Analysis based on a risk framework
- Analysis of methodologies for using the Internet for the collection of information society and other statistics
- Estimating Population Density Distribution from Network-based Mobile Phone Data
- Improving prediction of unemployment statistics with Google trends: preliminary experiments
- Methods for treating selectivity in big data sources
- Use of web activity evidence to increase timeliness (IAOS2014)



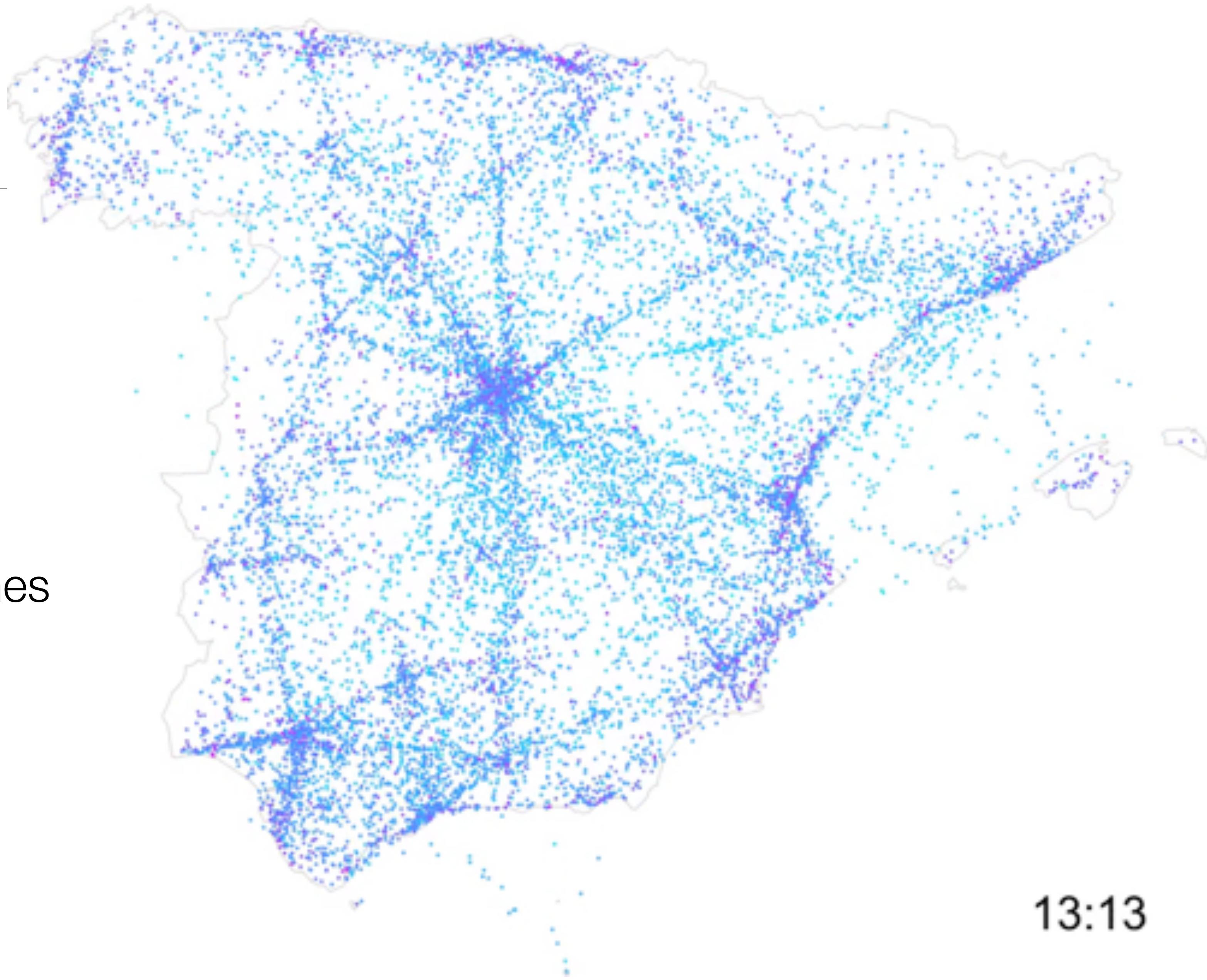
Unemployment

Human Development

Gender Gap

Unemployment

- ~20 Million Geolocalized tweets
- Use mobility to find “efficient economic zones”
- Fingerprints of unemployment:
 - Twitter penetration
 - Social interactions between zones
 - Mobility between zones
 - Daily patterns of activity
 - Tweets’ content in each zone

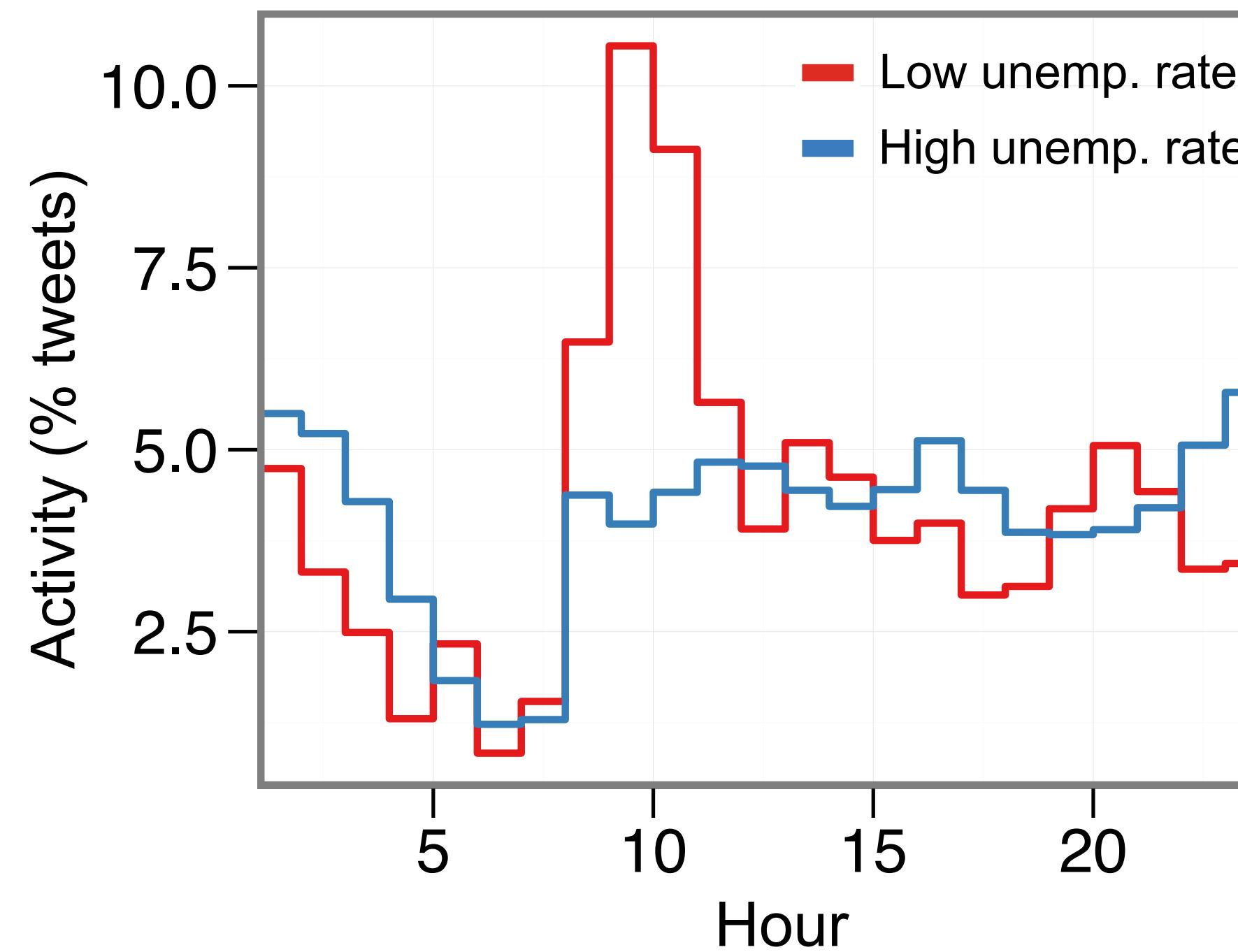


13:13

Llorente, A., Garcia-Herranz, M., Cebrian, M., & Moro, E. (2015). Social media fingerprints of unemployment. PLoS ONE, 10(5), e0128692. <http://doi.org/10.1371/journal.pone.0128692>

Unemployment

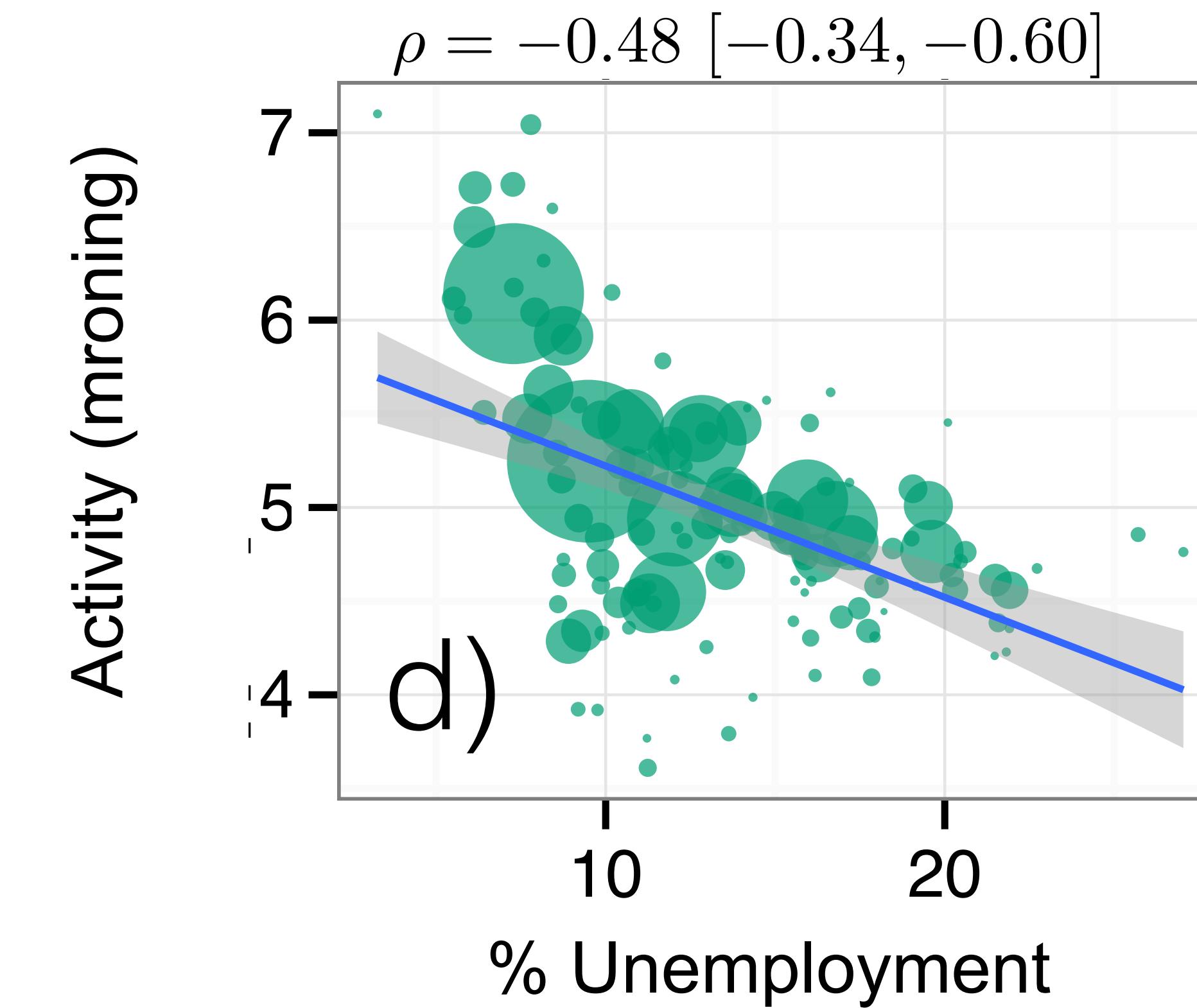
- Daily patterns of activity:
 - Early morning activity is correlated with unemployment



Just arrived to work, mondays are too hard...



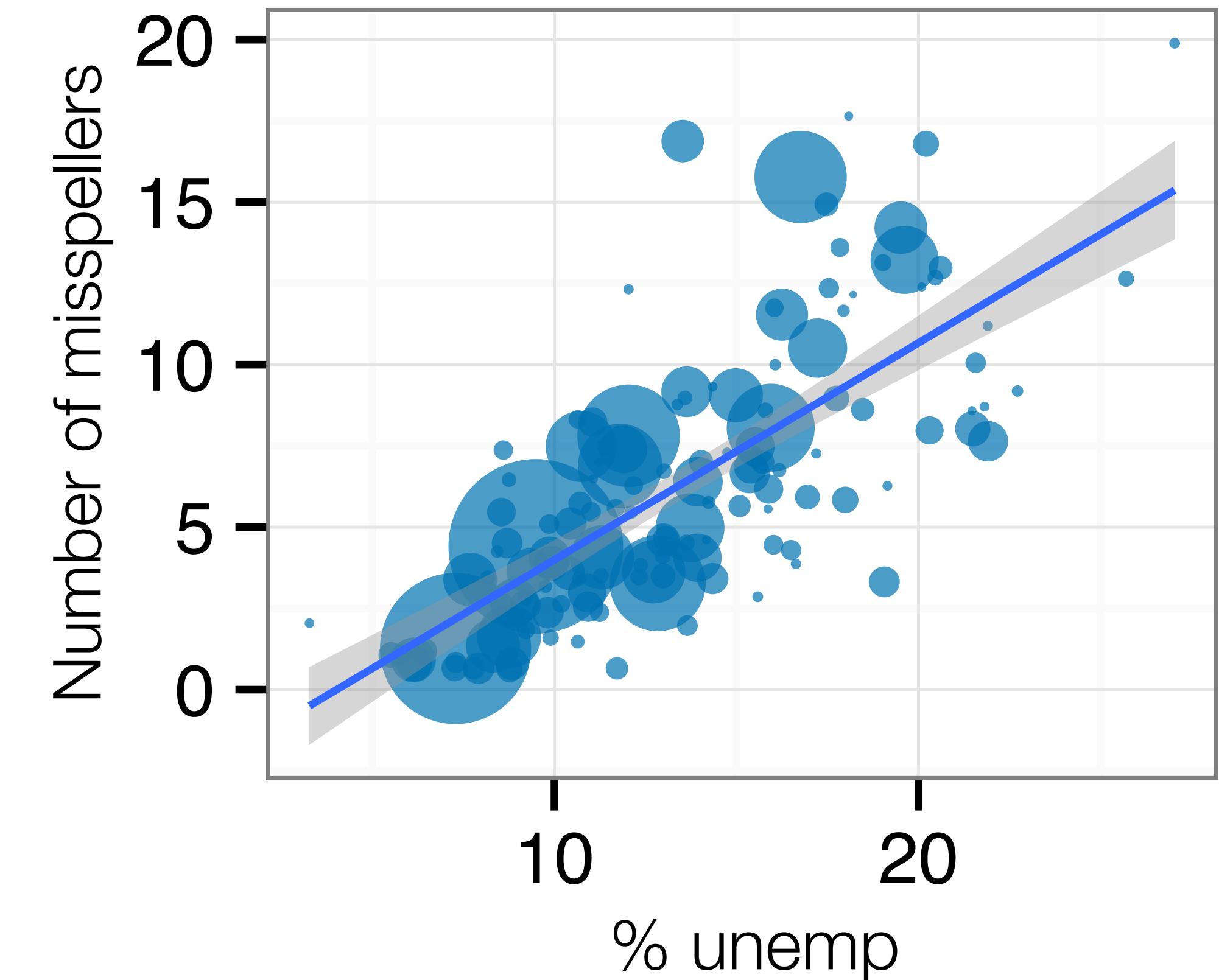
10:43 - 2 de jun. de 2014



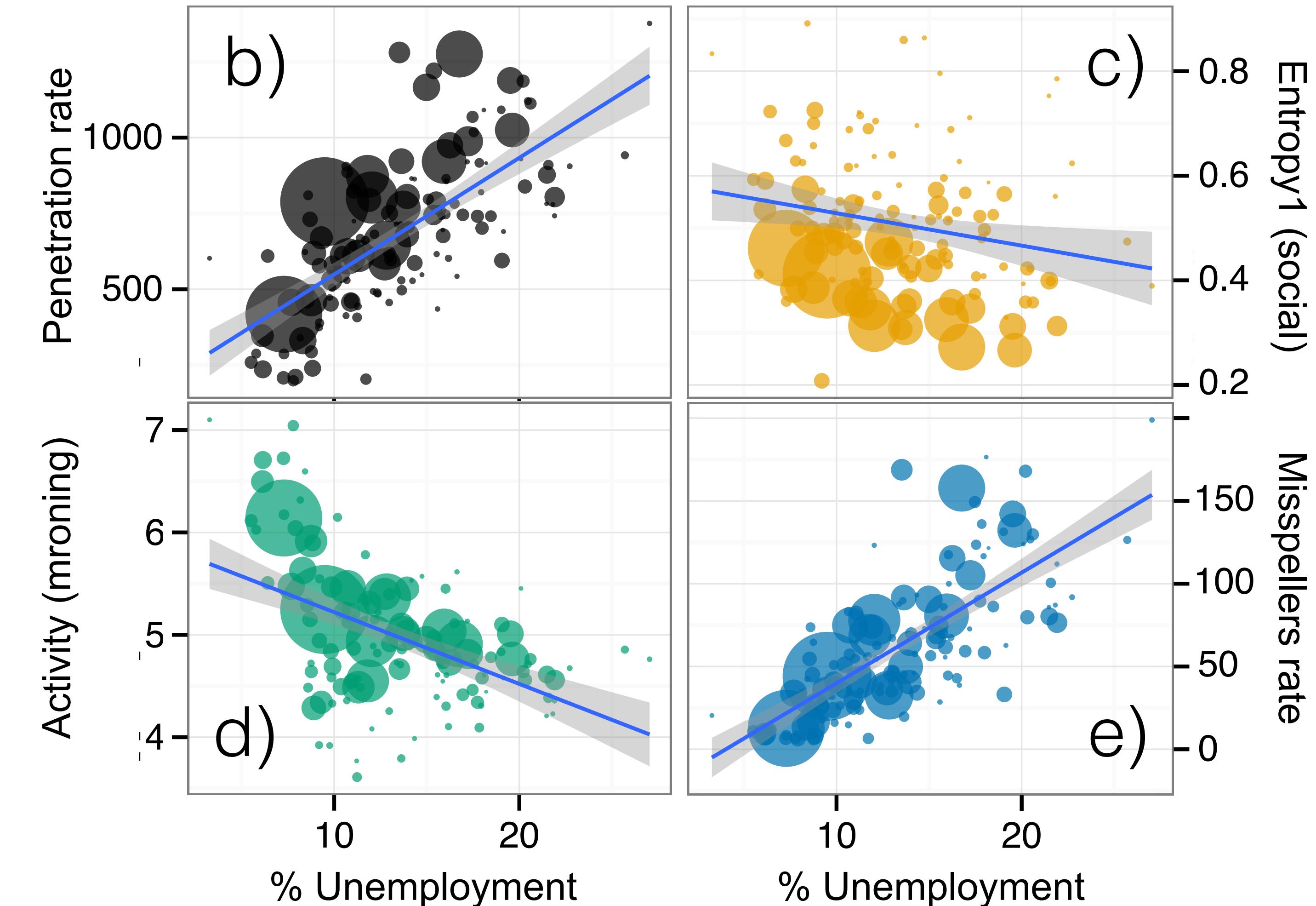
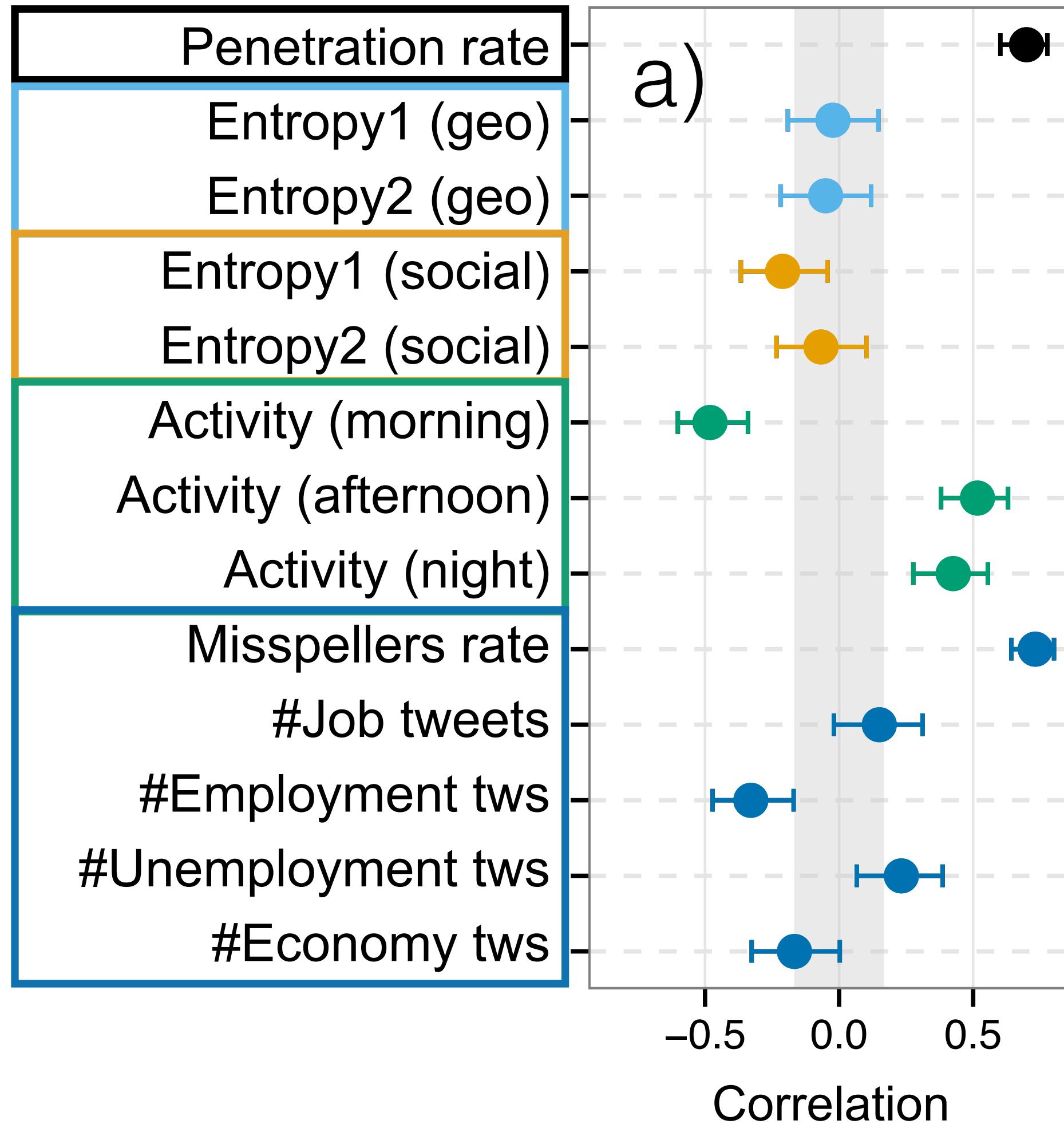
Unemployment

- Tweet's content:
 - Misspelling is highly correlated with unemployment

Tweet	Correct spelling
Alguien se viene con migo aver la vida de PI??	Alguien se viene conmigo a ver la vida de PI??
La quiero mucho y la hecho de menos	La quiero mucho y la echo de menos
Yo llendo a trabajar con este tiempo	Yo yendo a trabajar con este tiempo

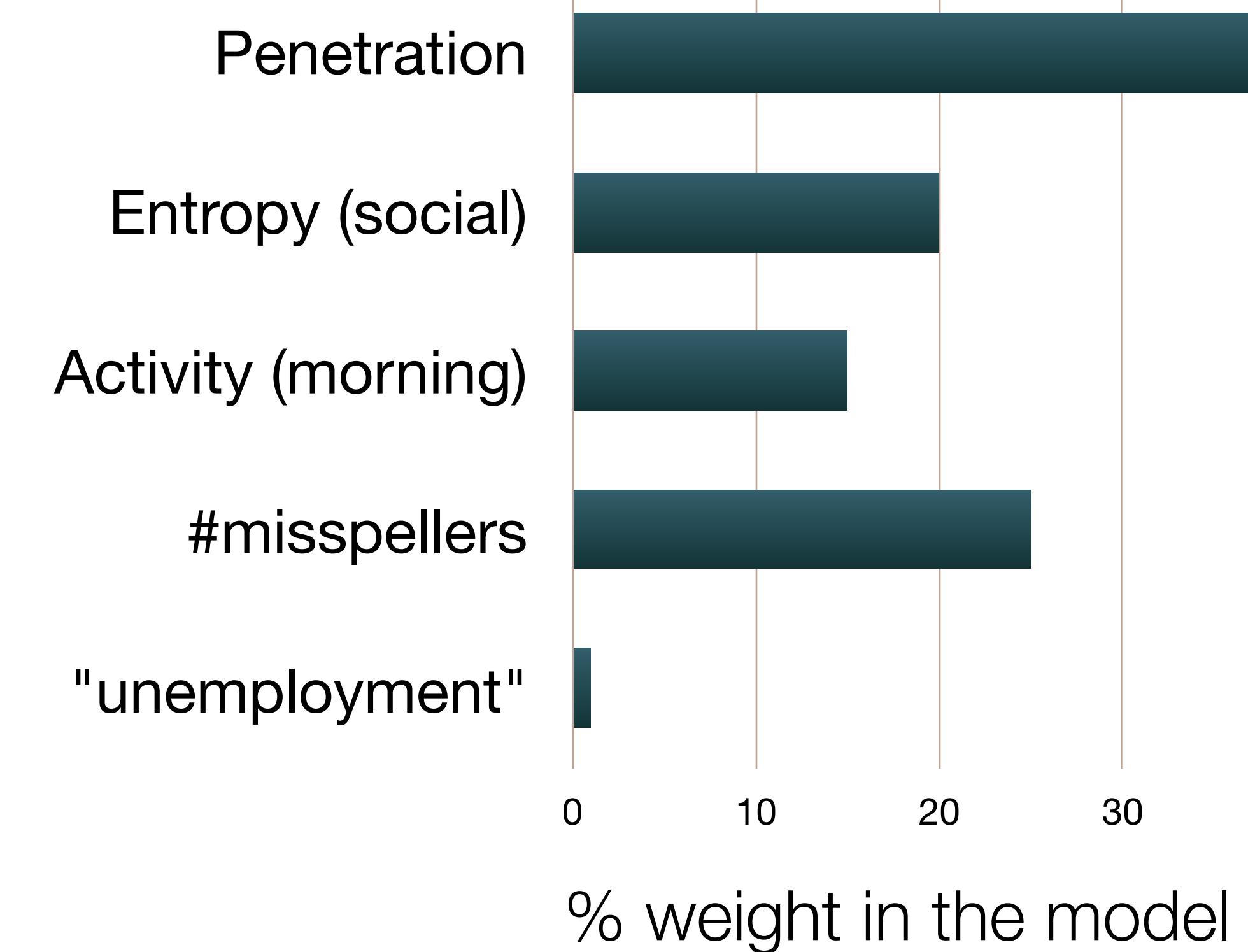
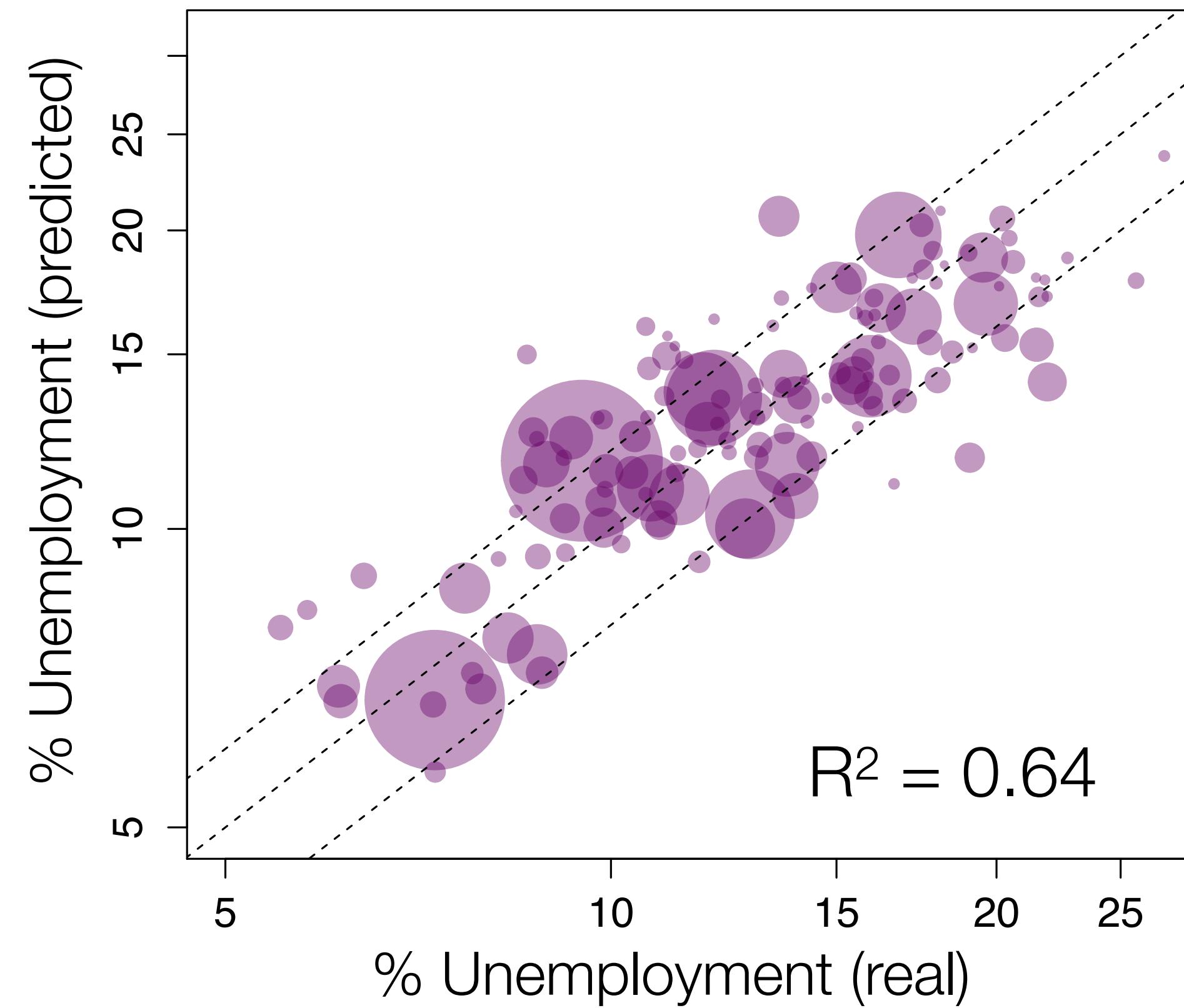


Unemployment



Unemployment

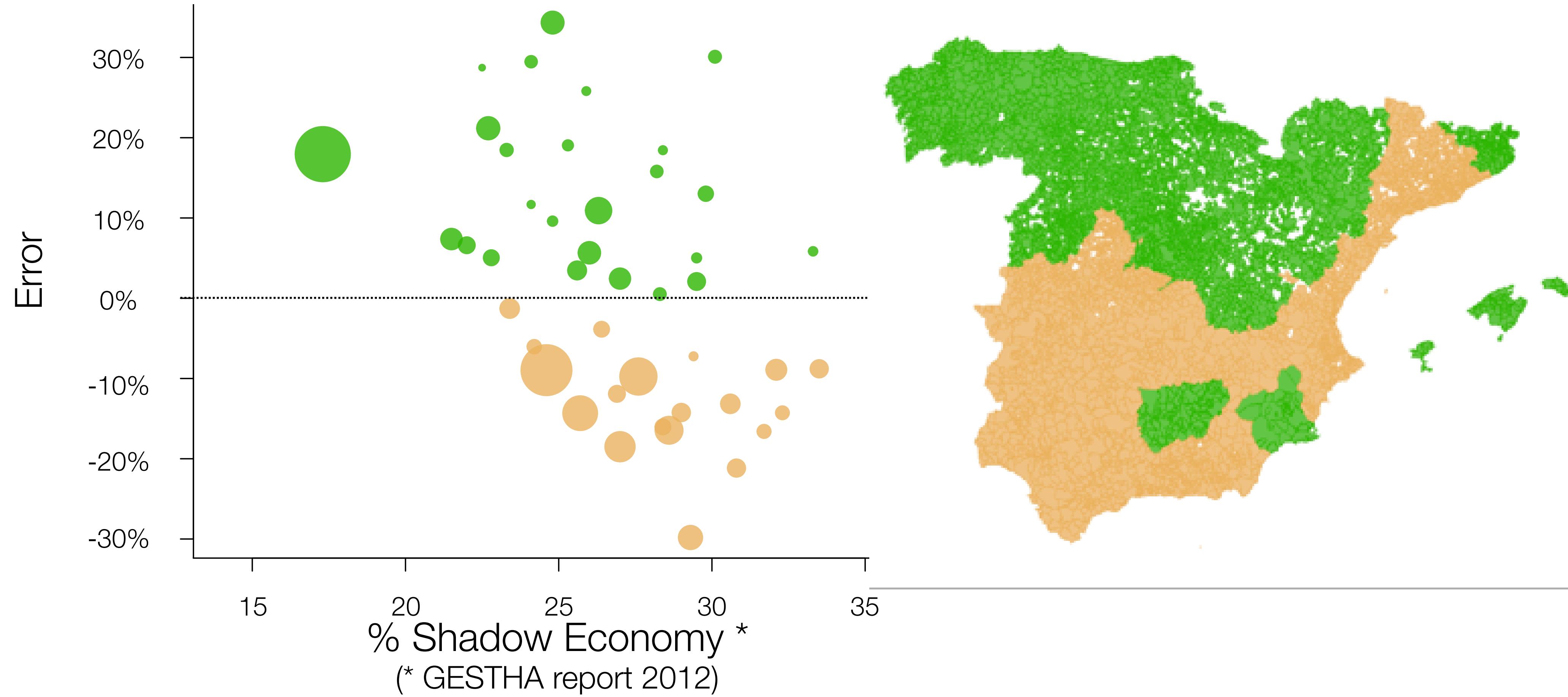
- Explanatory power of the social media fingerprints



Unemployment

- Are we really wrong?

Model Error = Model[variables] - Official unemployment



Use twitter to estimate the Human Development Index (HDI)



BRAZIL



COSTA RICA



MEXICO



NIGERIA



POLAND



COLOMBIA



INDONESIA



NEPAL



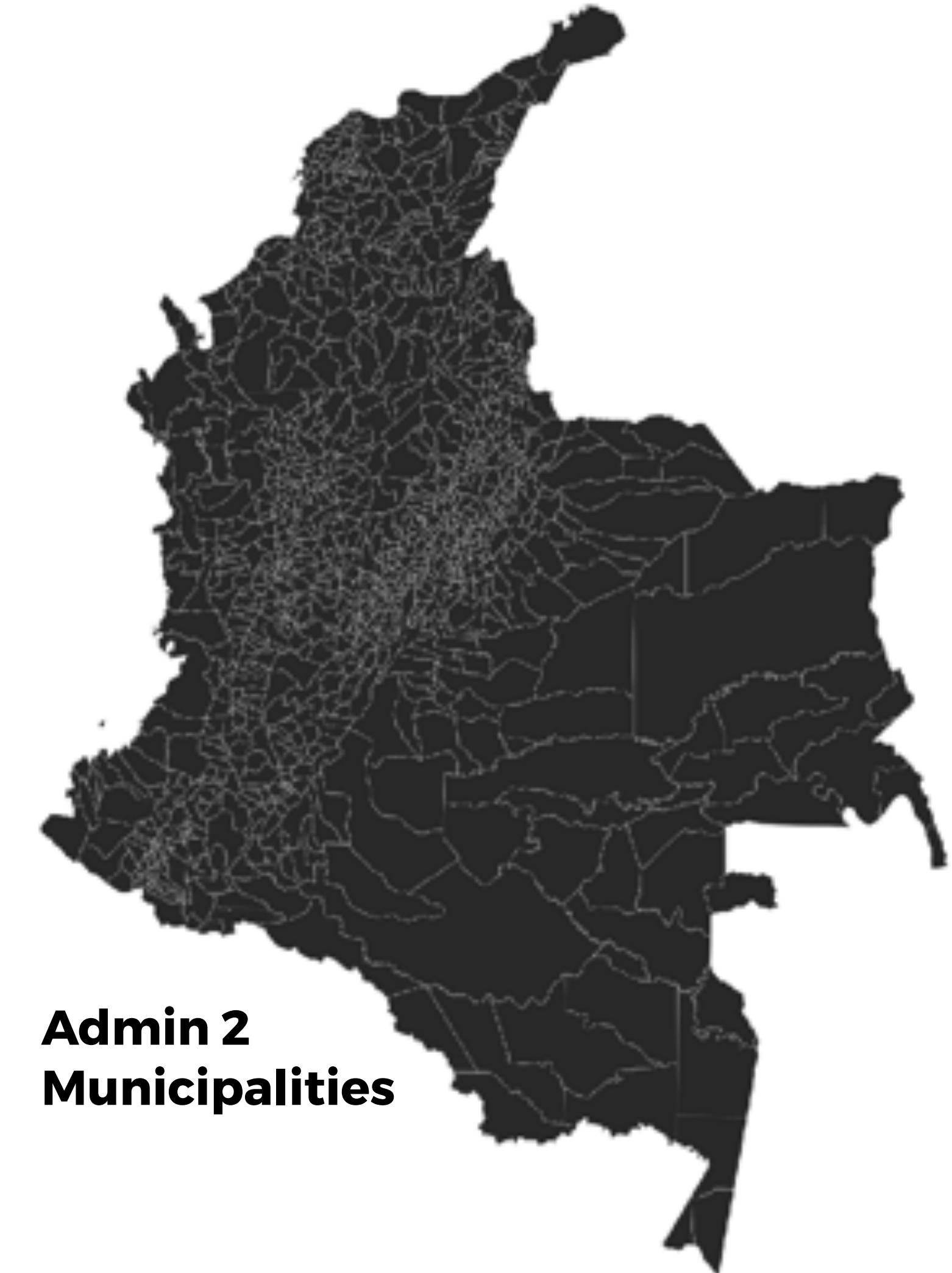
PAKISTAN



Use **twitter** to estimate HDI on sub-national levels

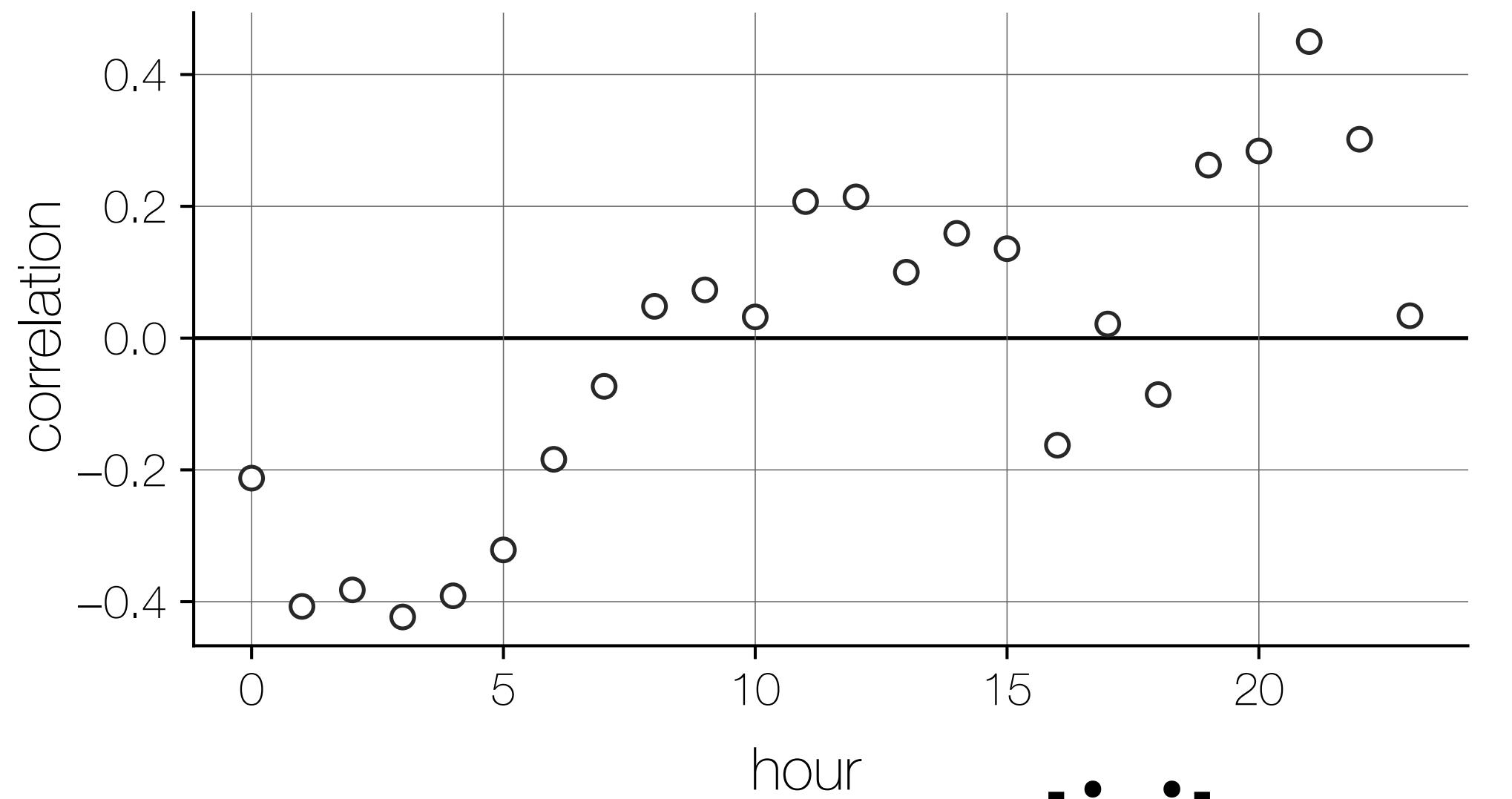


Admin 0
Country



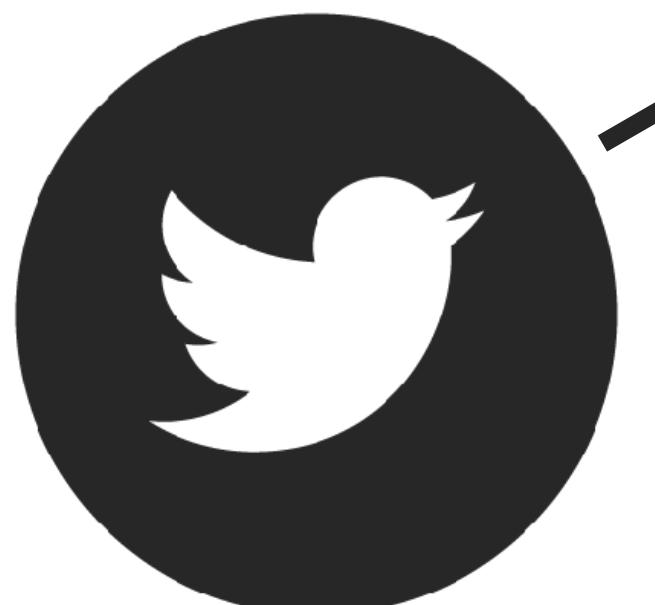
Admin 2
Municipalities

features



activity

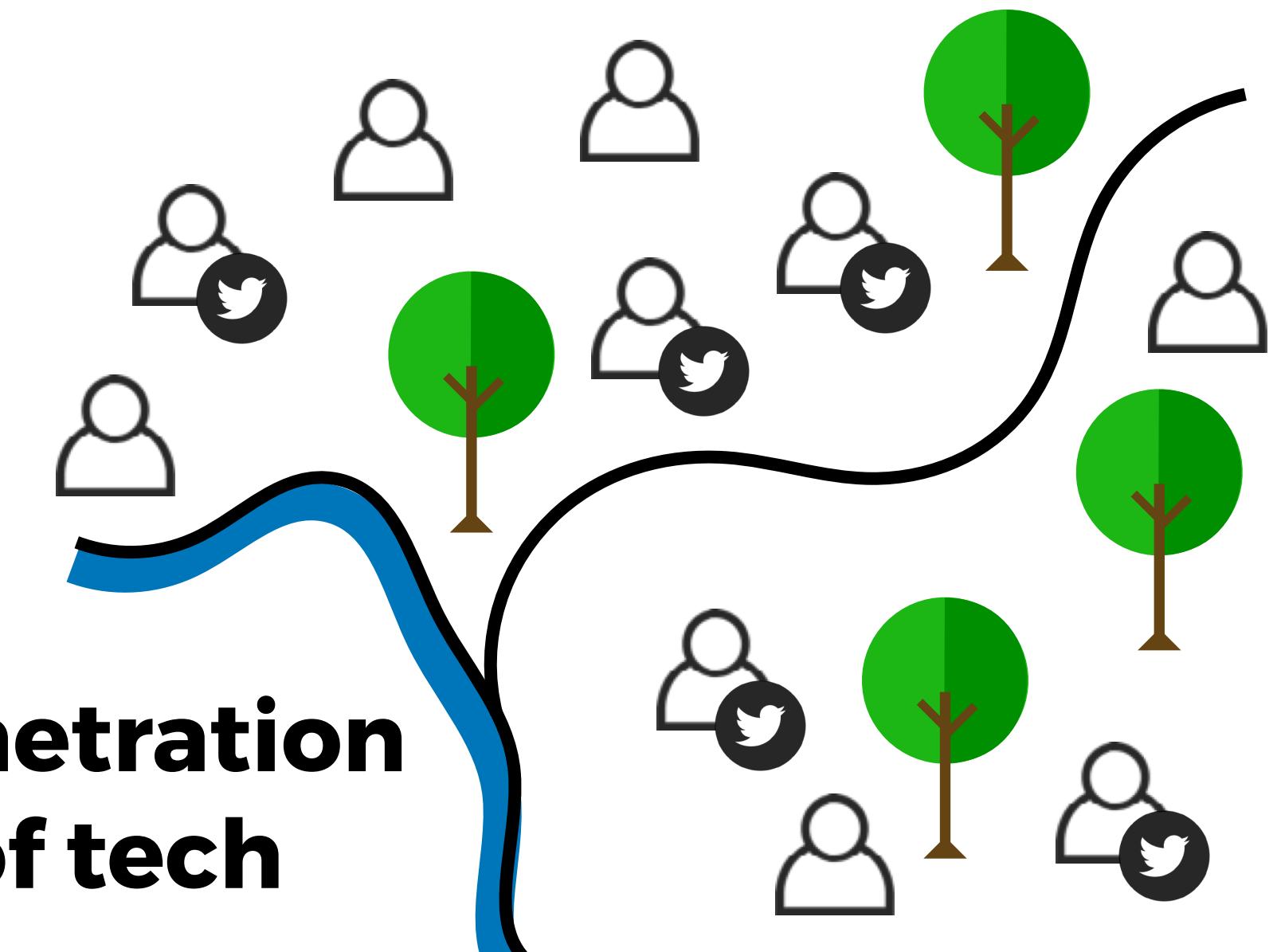
- no. tweets
- morning
- daily
- night



mobility

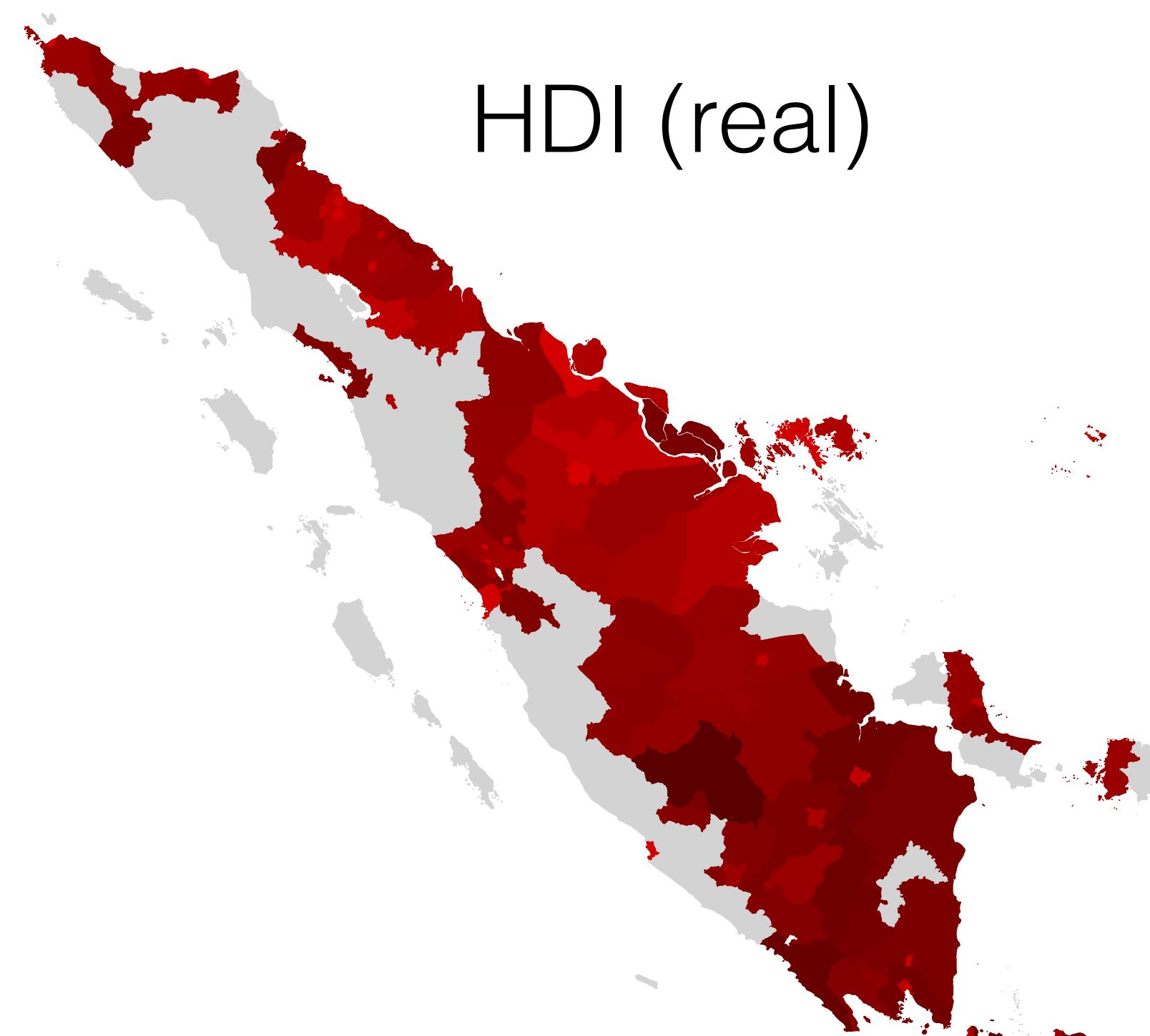


penetration of tech



Human development

- HDI = Human Development index (sub-national)



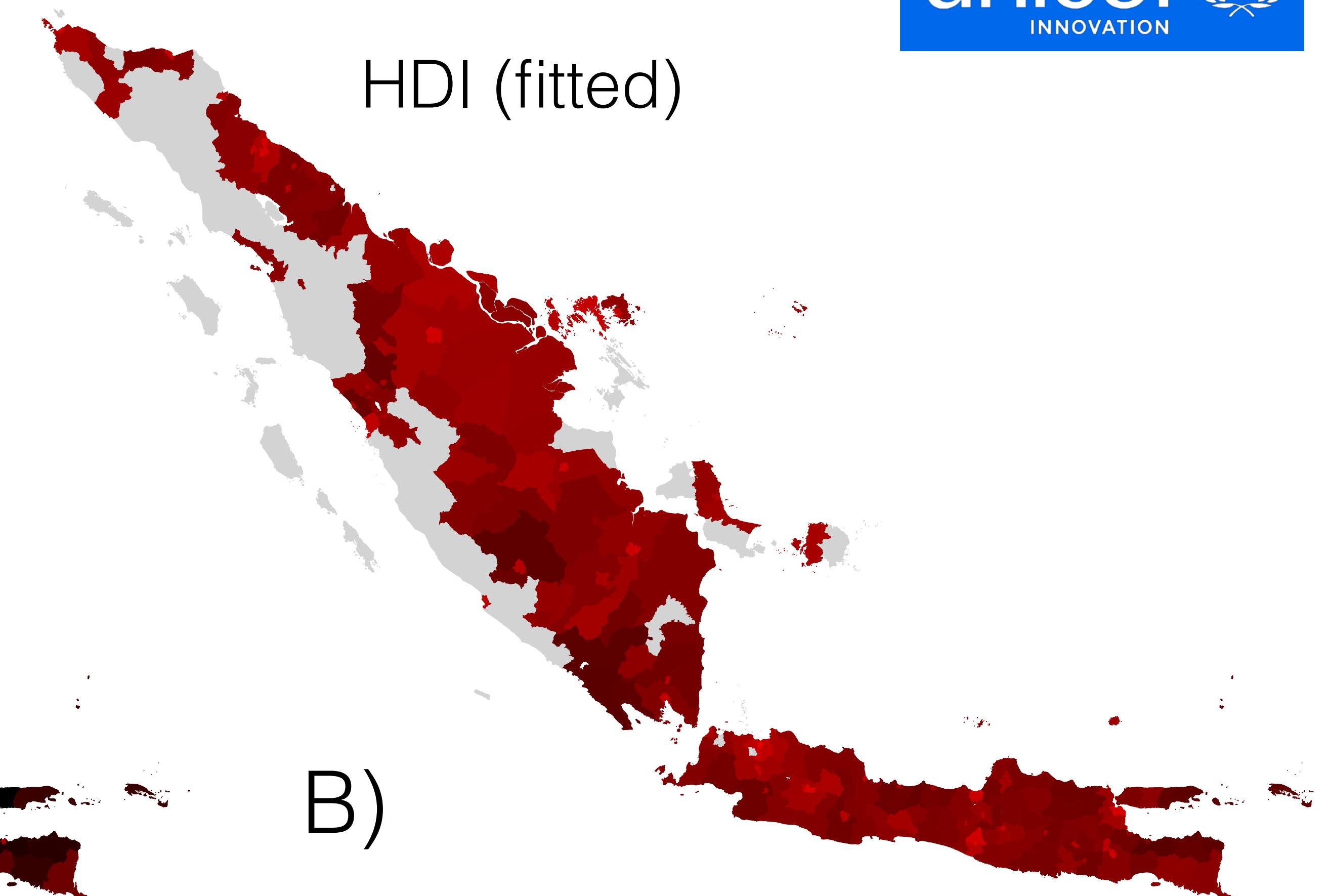
A)



@estebanmoro



Universidad
Carlos III de Madrid

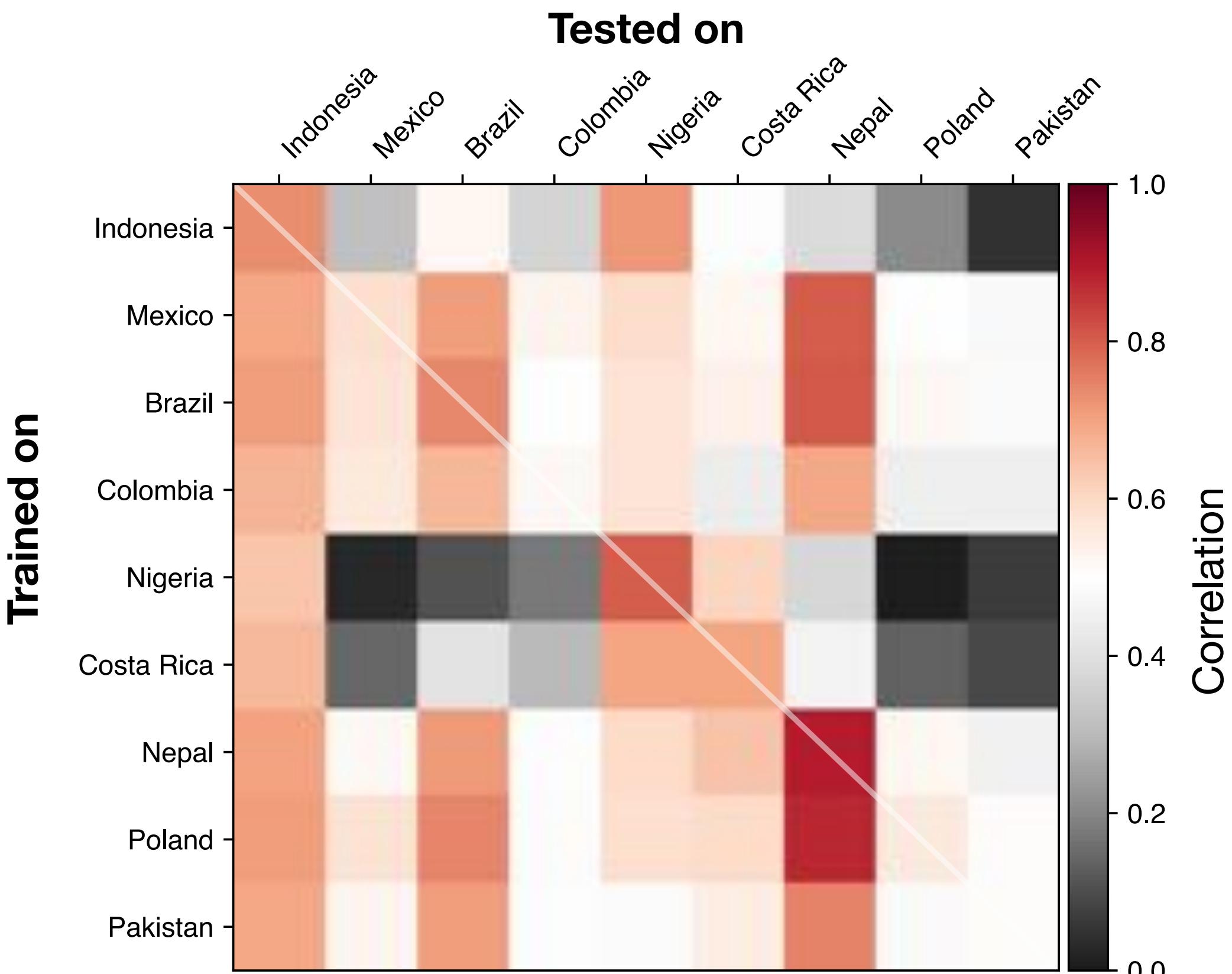
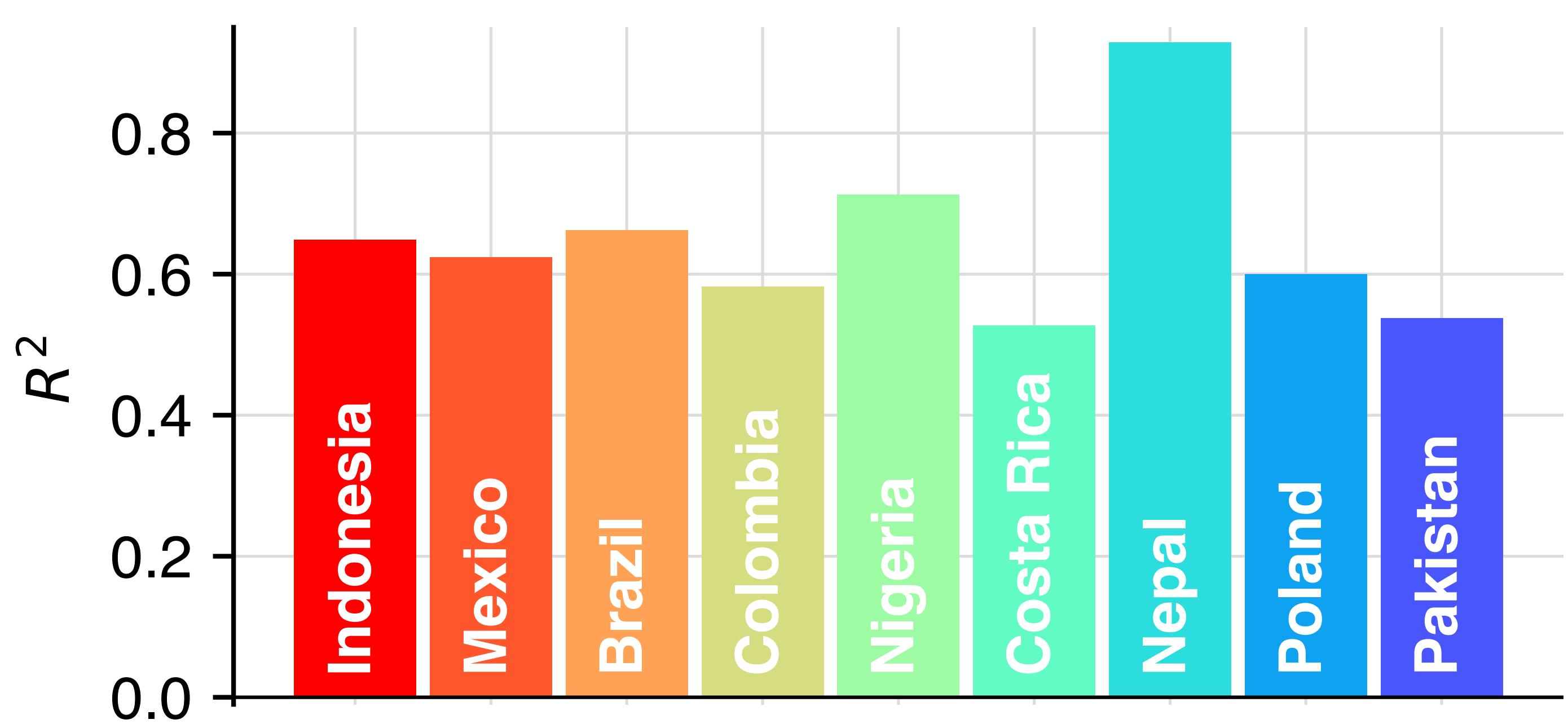


B)



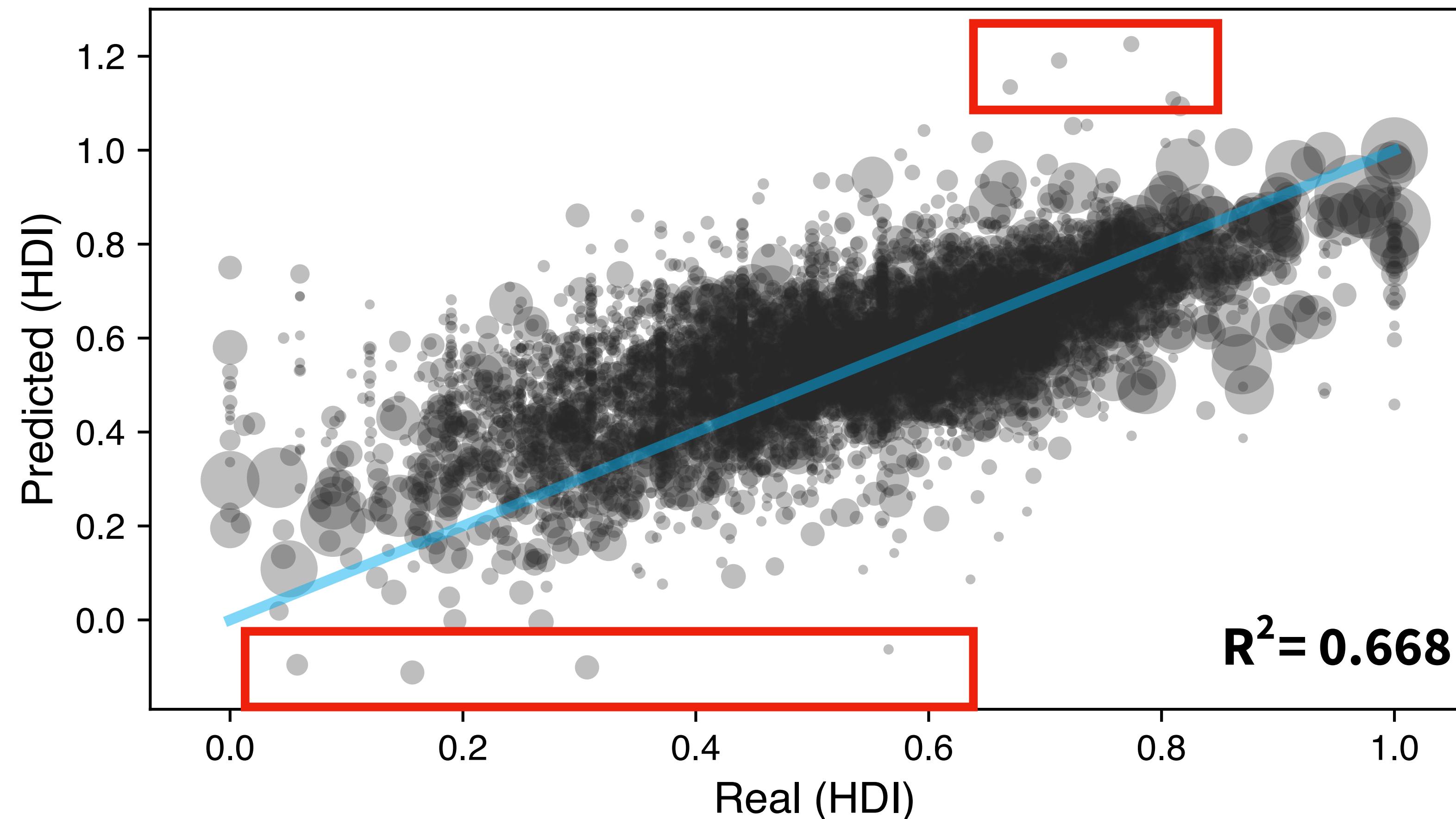
Human development

- Can be re-use these models?



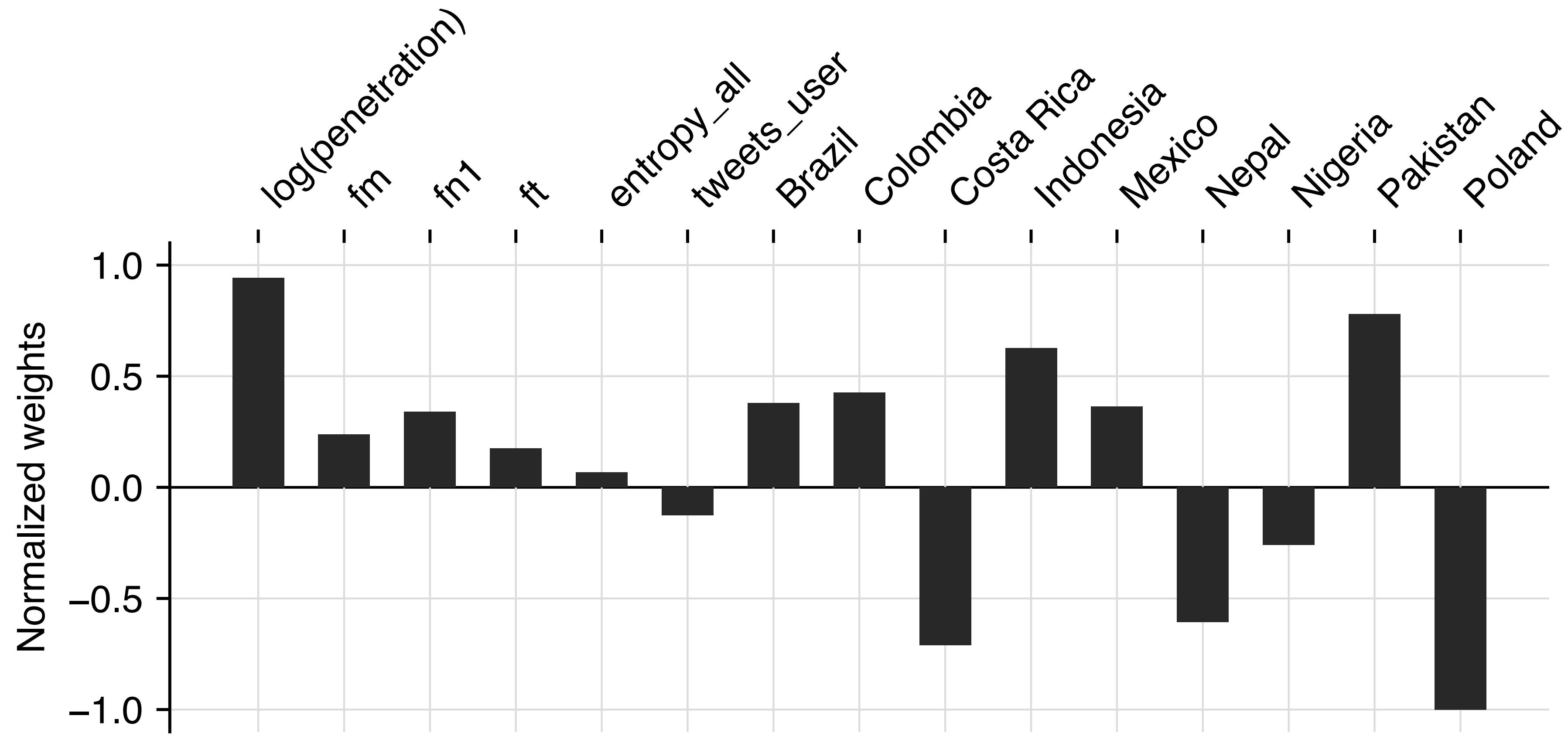
Über model

Building one model for all countries



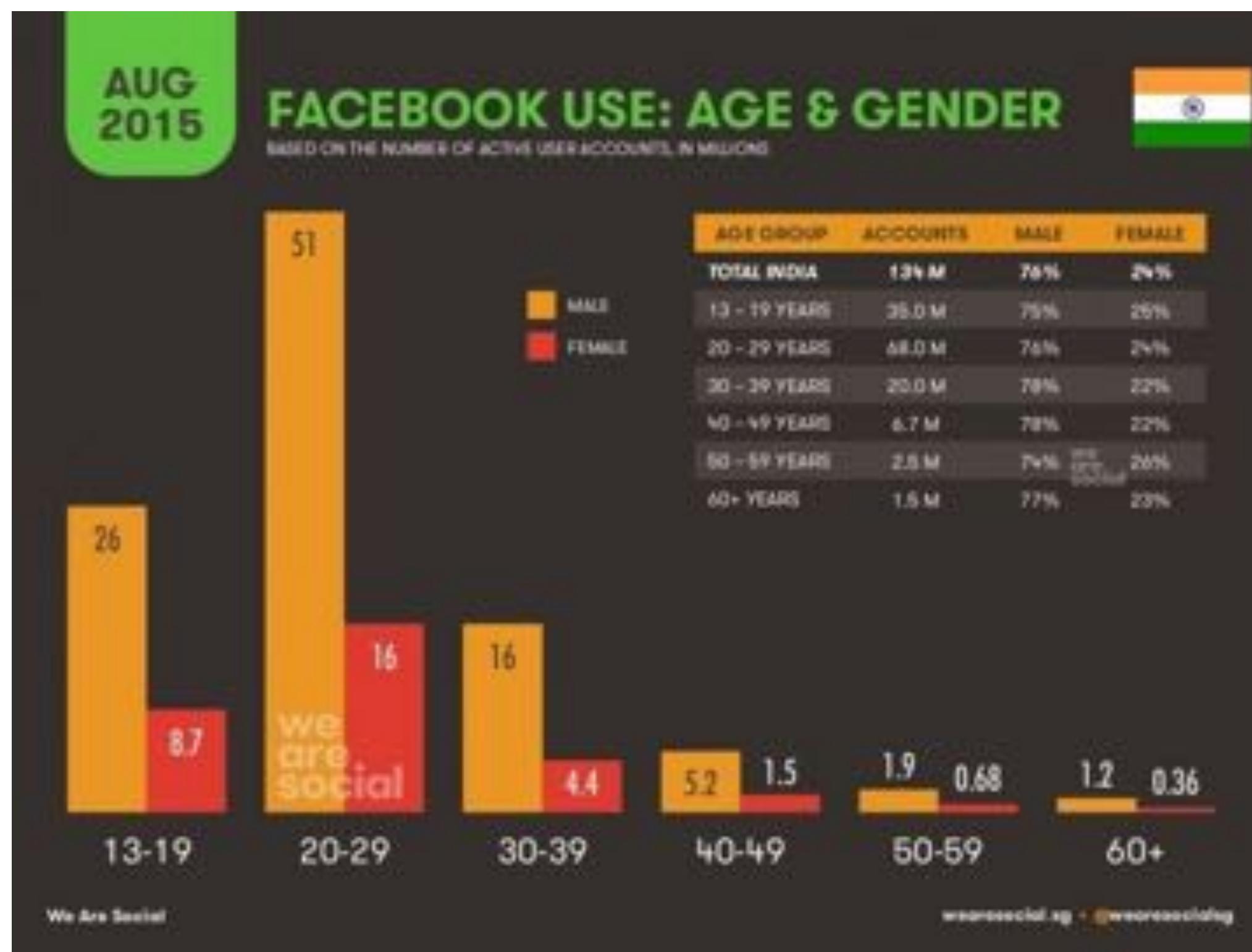
Über model

Importance of features



Facebook gender divide

- Can we understand **gender digital divide** in the whole world?



The Digital Divide Is Still a Thing
Percentage of individuals using the Internet in 2015



Facebook gender divide

- Use the Marketing Application Programming Interface to collect:
 - Number of Facebook users by age and gender in each 217 countries
 - Calculate the “Facebook gender divide” metric

$$FGD_c = \log \left(\frac{R_{Male,c}}{R_{Female,c}} \right),$$

- Find the main explanatory variables for that divide

The screenshot shows the '2. Targeting' section of the Facebook Marketing API. It includes fields for Location (United States, Everywhere selected), Age (24-60), Sex (Male checked, Female checked), Keywords (Marketing), Education (All selected), Workplaces (input field), Relationship (Single, In a Relationship, Engaged, Married), Interested In (Men, Women), and Languages (input field). A summary at the bottom indicates an approximate reach of 48,500 users matching the specified criteria.

Targeting
By default, Facebook targets all users 18 and older in the default location. You can change any targeting specification you wish.

Location
Facebook Ads uses IP address and a user's profile information to determine a user's location.

Age

Sex

Keywords
Keywords are based on information users put in their Facebook profiles, such as Activities, Favorite Books, TV Shows, Movies, etc.

Education

Workplaces
Enter a company, organization or other workplace.

Relationship

Interested In

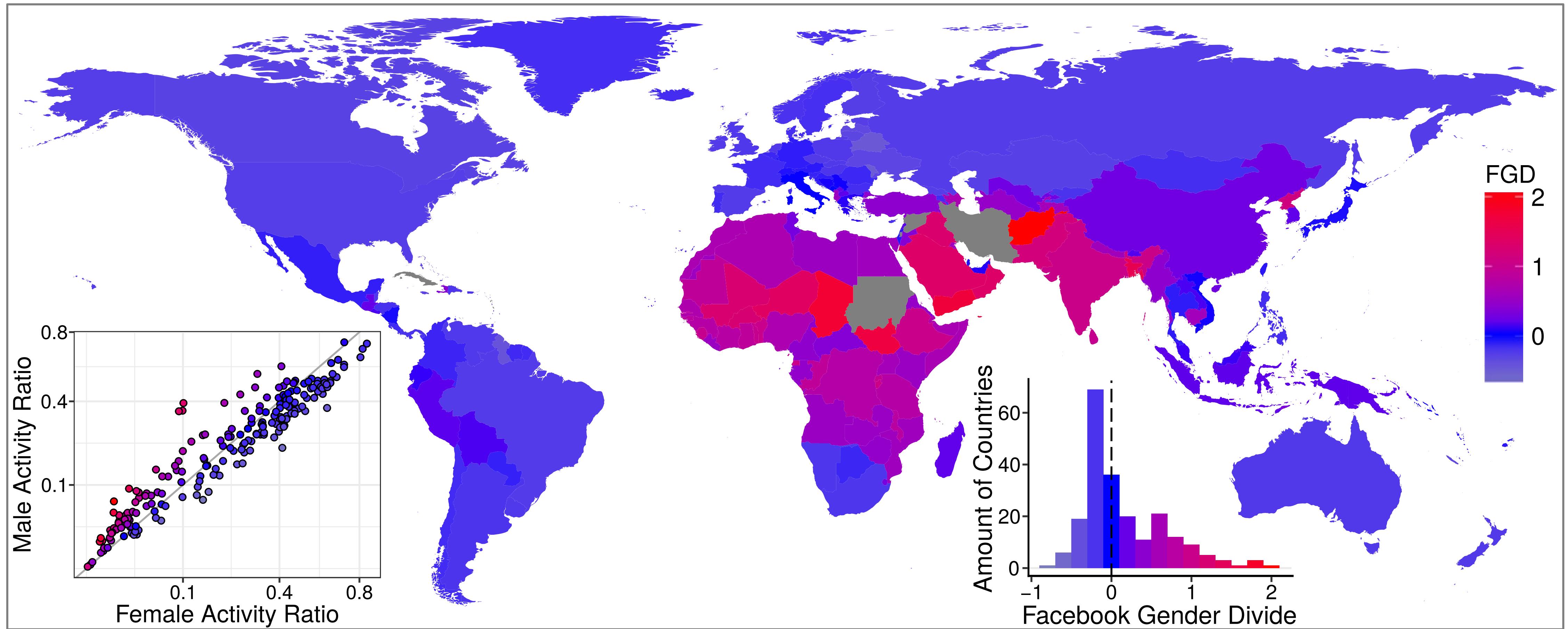
Languages
Enter language

Approximate reach: 48,500
I want to reach people between 24 and 60 years old in the United States who like Marketing.

Facebook gender divide

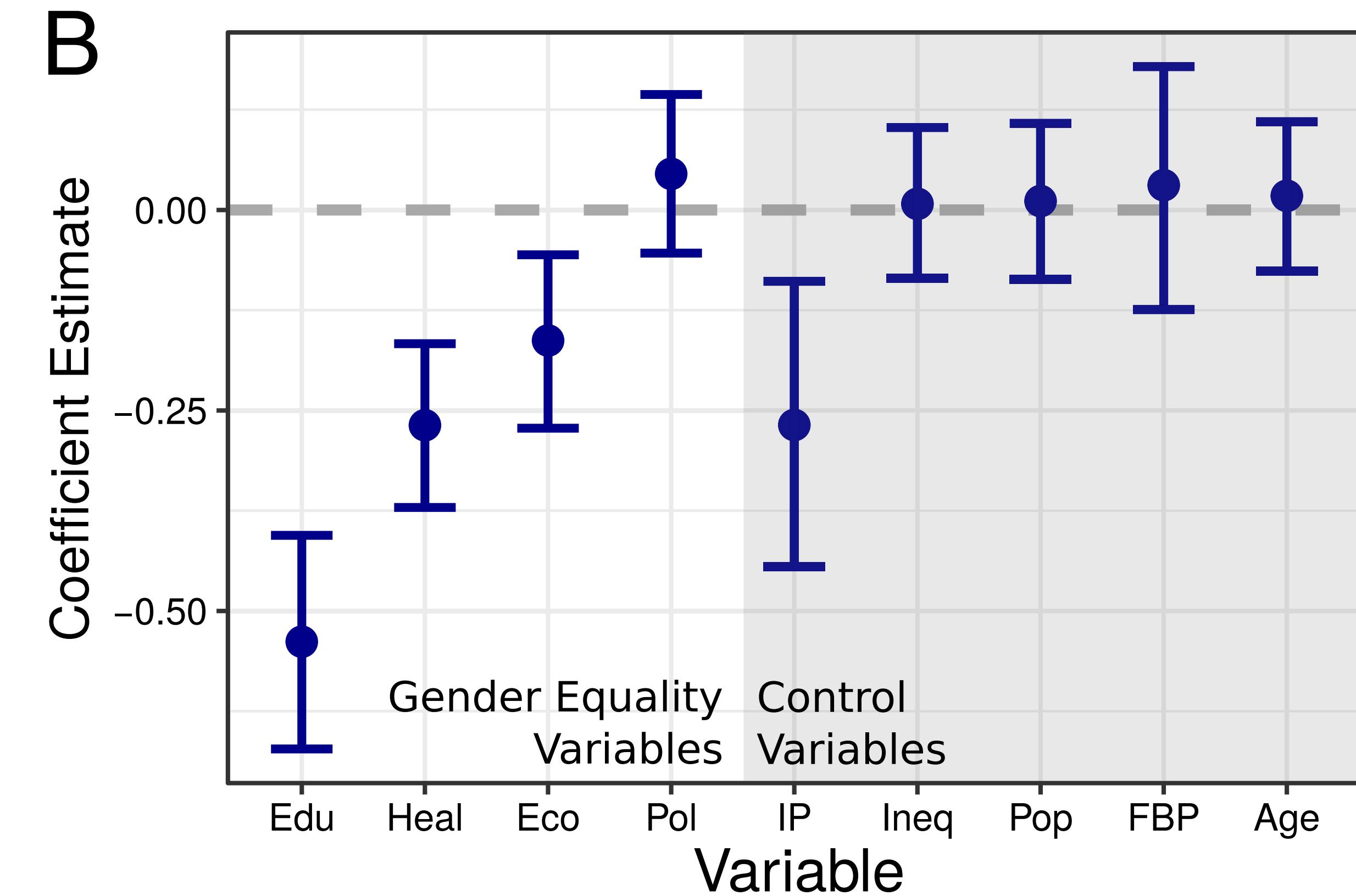
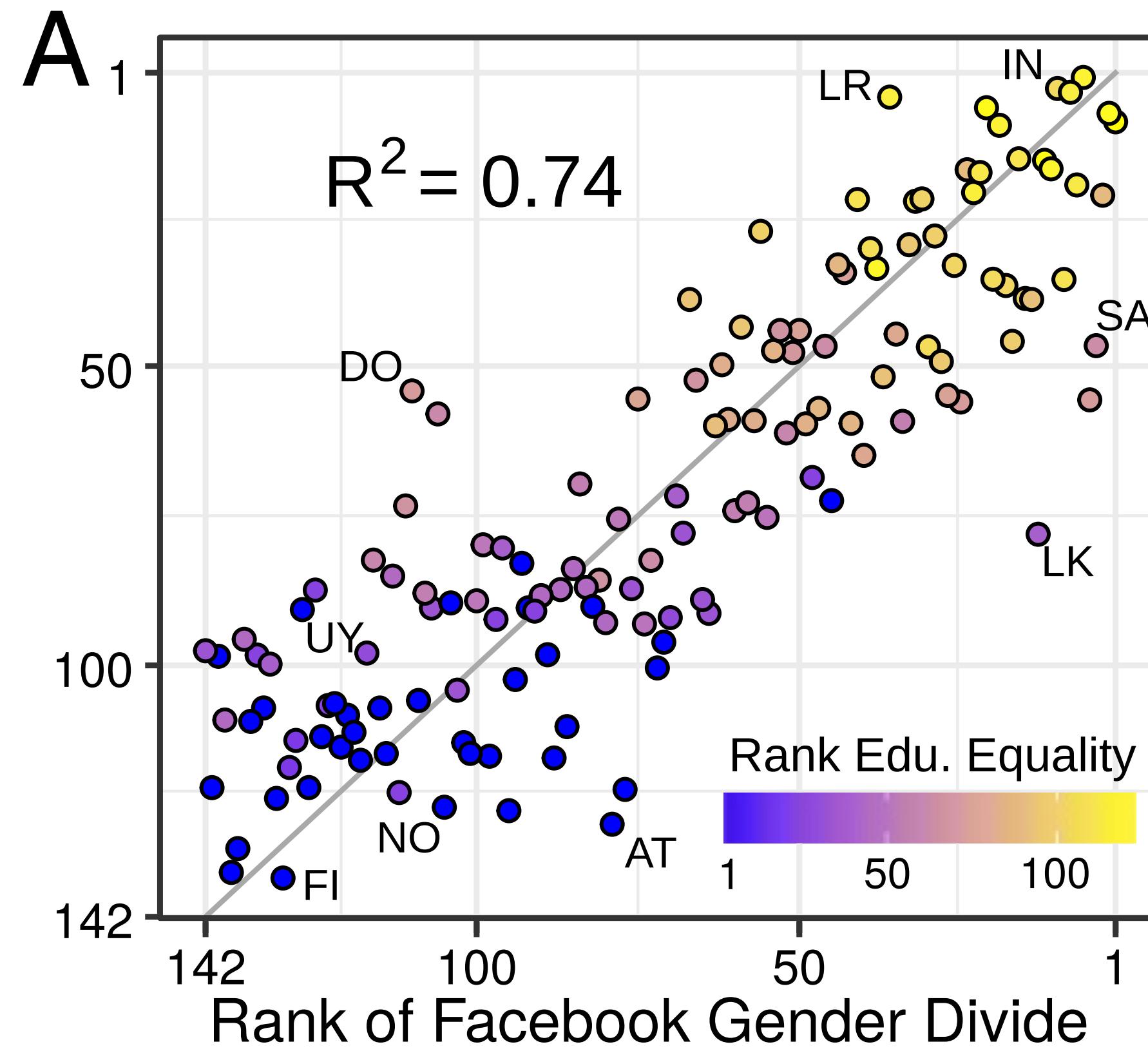
Garcia, D., Kassa, Y. M., Cuevas, A., Cebrian, M., Moro, E., Rahwan, I., & Cuevas, R. (2018). PNAS

- 217 countries, around 1.4 billion users



Facebook gender divide

- What are the explanatory variables for the Facebook gender divide?



Can we use networks to mobilize people?

- Yes, information travels very fast in social networks, but can we use it to mobilize people?

The collage illustrates the DARPA Network Challenge, a competition that utilized social networks to mobilize people across the United States. The top right image shows the official website with a banner reading "DARPA NETWORK CHALLENGE" and a map of the US with ten locations marked. The bottom right image is a map of the US with numbered pins at various locations. The bottom left image shows several red balloons numbered 1 through 6.

DARPA NETWORK CHALLENGE

Home Gallery FAQ Rules DARPA

“Impossible by conventional intelligence”

9 - Waterfront Park
Portland, OR

1 - Union Square
San Francisco, CA

4 - Chase Palm Park
Santa Barbara, CA

2 - Chaparral Park
Scottsdale, AZ

5 - Lee Park
Memphis, TN

7 - Glasgow Park
Christiana, DE

3 - Tonsler Park
Charlottesville, VA

10 - Centennial Park
Atlanta, GA

8 - Katy Park
Katy, TX

6 - Collins Avenue
Miami, FL

1

2

3

4

5

6

Can we use networks to mobilize people?

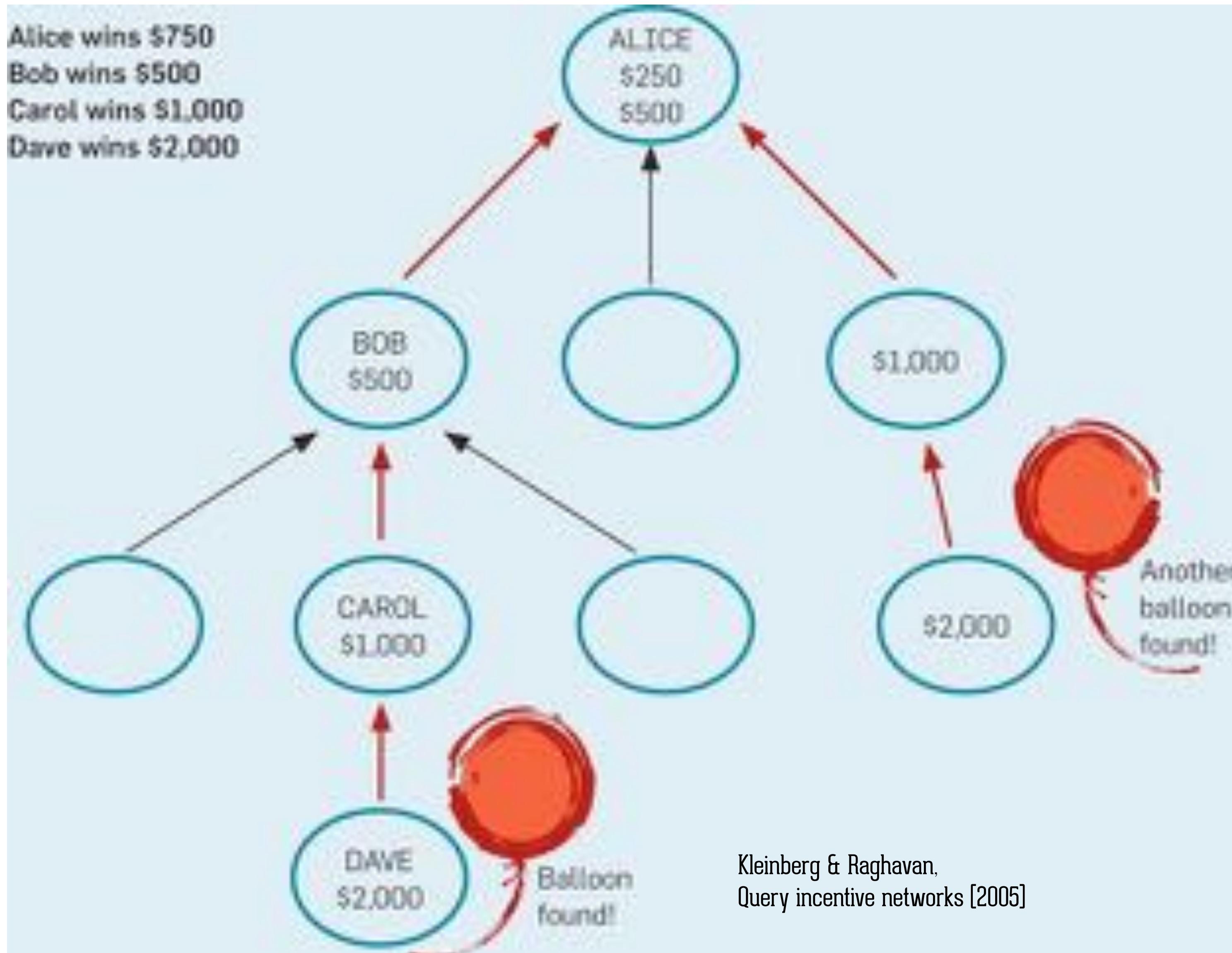


Time-Critical Social Mobilization

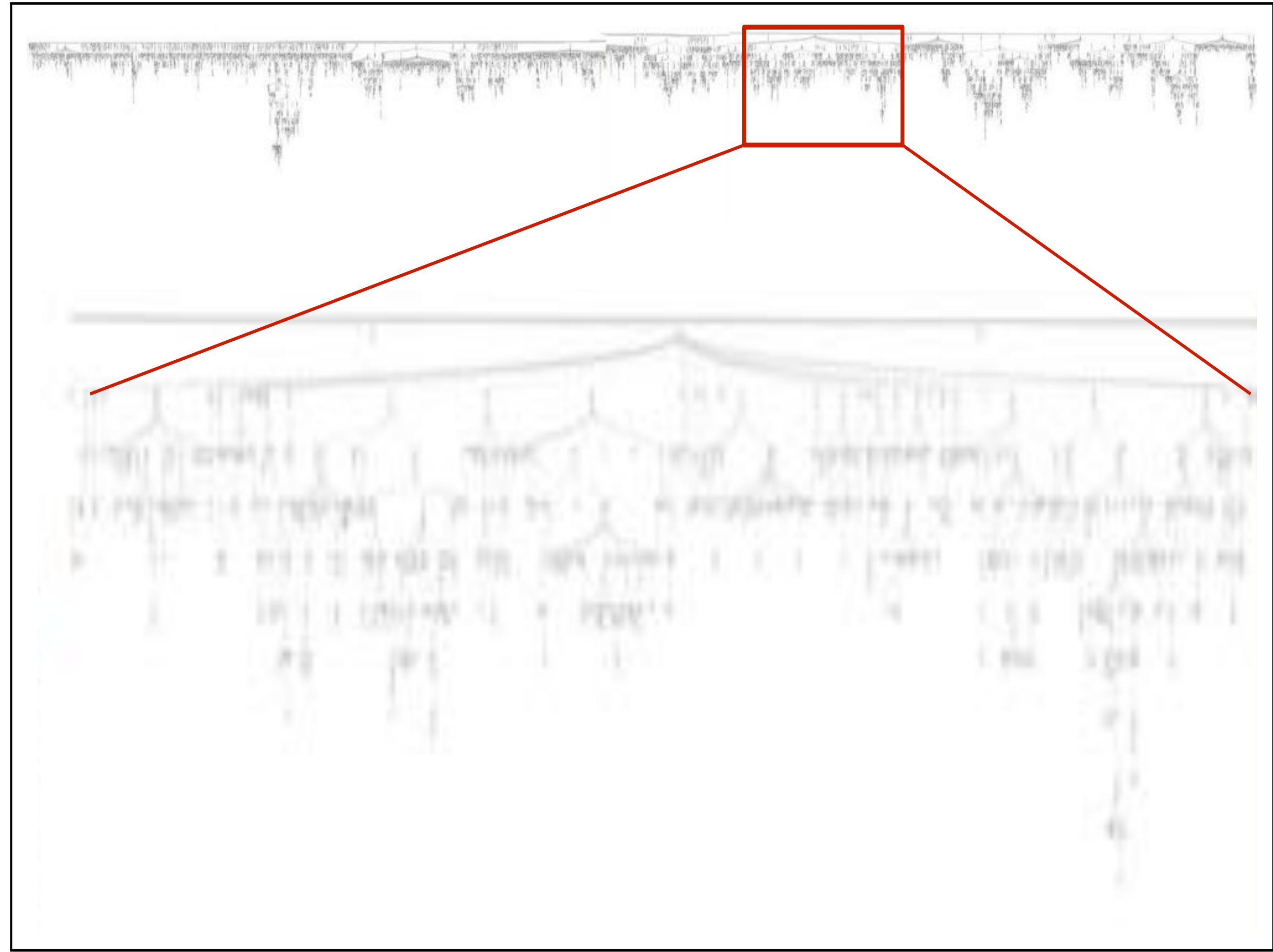
Galen Pickard,^{1,2*} Wei Pan,^{1*} Iyad Rahwan,^{1,3*} Manuel Cebrian,^{1*} Riley Crane,¹ Anmol Madan,¹ Alex Pentland^{1†}

The World Wide Web is commonly seen as a platform that can harness the collective abilities of large numbers of people to accomplish tasks with unprecedented speed, accuracy, and scale. To explore the Web's ability for social mobilization, the Defense Advanced Research Projects Agency (DARPA) held the DARPA Network Challenge, in which competing teams were asked to locate 10 red weather balloons placed at locations around the continental United States. Using a recursive incentive mechanism that both spread information about the task and incentivized individuals to act, our team was able to find all 10 balloons in less than 9 hours, thus winning the Challenge. We analyzed the theoretical and practical properties of this mechanism and compared it with other approaches.

Can we use networks to mobilize people?

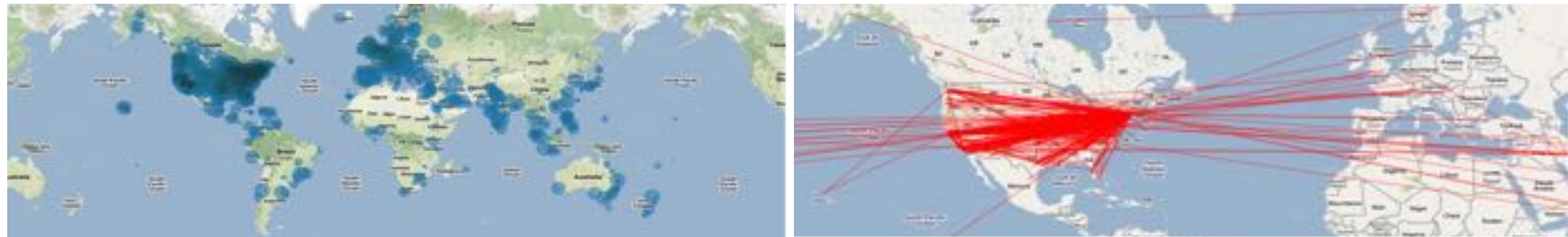


Can we use networks to mobilize people?



Can we use networks to mobilize people?

Global reach in **36 hours**



Can we use networks to mobilize people?

Can we model it?

A model of social geographical mobilization

Geographical spreading
of social networks

Temporal dynamics of
message propagation



Branching dynamics
of recruitment



Limits of social mobilization

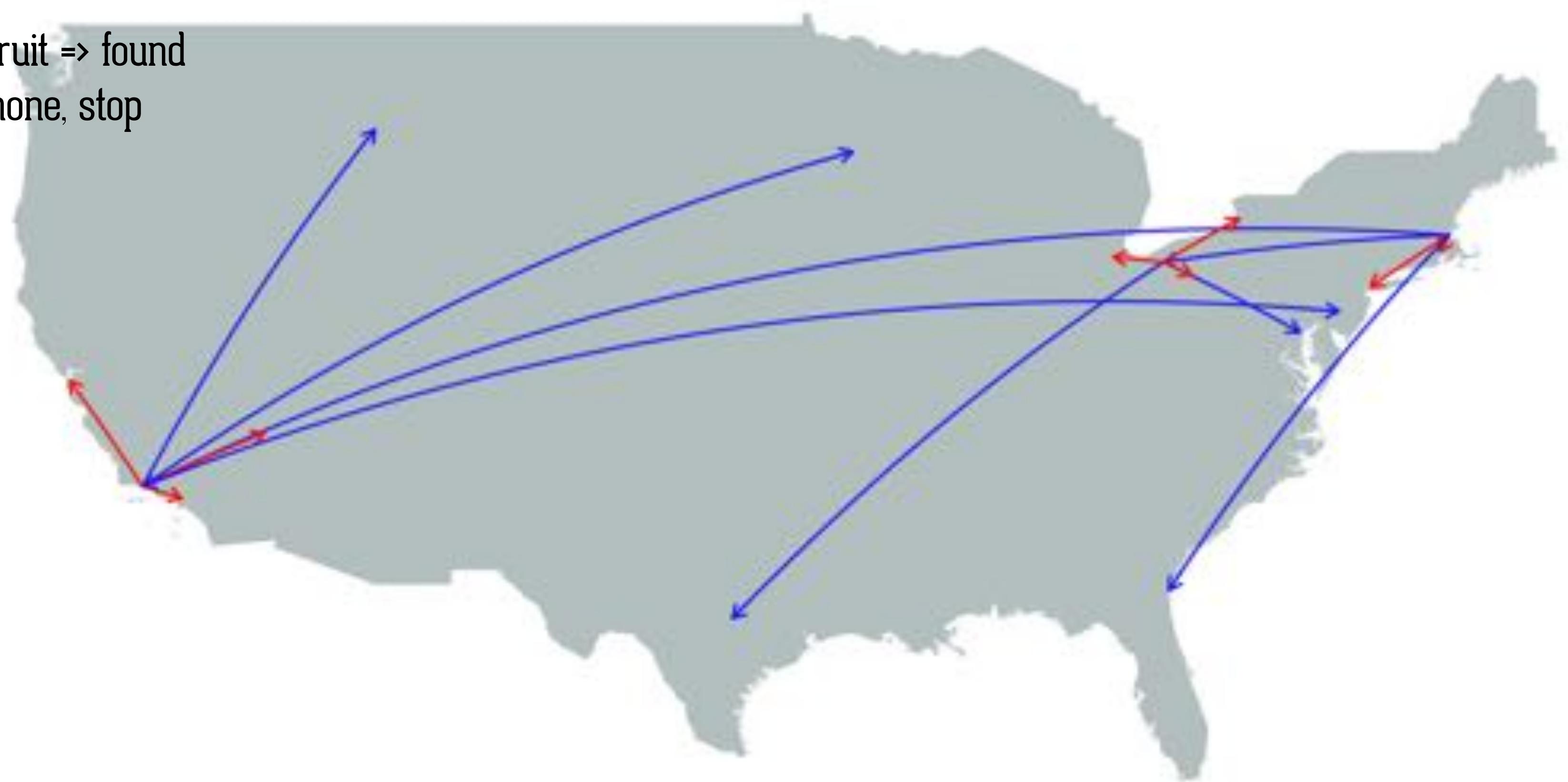
Alex Rutherford^a, Manuel Cebrian^{b,c}, Sohan Dsouza^a, Esteban Moro^{d,e}, Alex Pentland^f, and Iyad Rahwan^{a,g,1}

^aComputing and Information Science, Masdar Institute of Science and Technology, Abu Dhabi 54224, United Arab Emirates; ^bDepartment of Computer Science and Engineering, University of California at San Diego, La Jolla, CA 92093; ^cNational Information and Communications Technology Australia, Melbourne, VIC 3010, Australia; ^dDepartamento de Matemáticas and Grupo Interdisciplinar de Sistemas Complejos, Universidad Carlos III de Madrid, 28911 Madrid, Spain; ^eInstituto de Ingeniería del Conocimiento, Universidad Autónoma de Madrid, 28049 Madrid, Spain; ^fMedia Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139; and ^gSchool of Informatics, University of Edinburgh, Edinburgh EH8 9AB, United Kingdom

Edited by James Minihane, Cornell University, Ithaca, NY, and approved March 1, 2013 (received for review September 20, 2012)

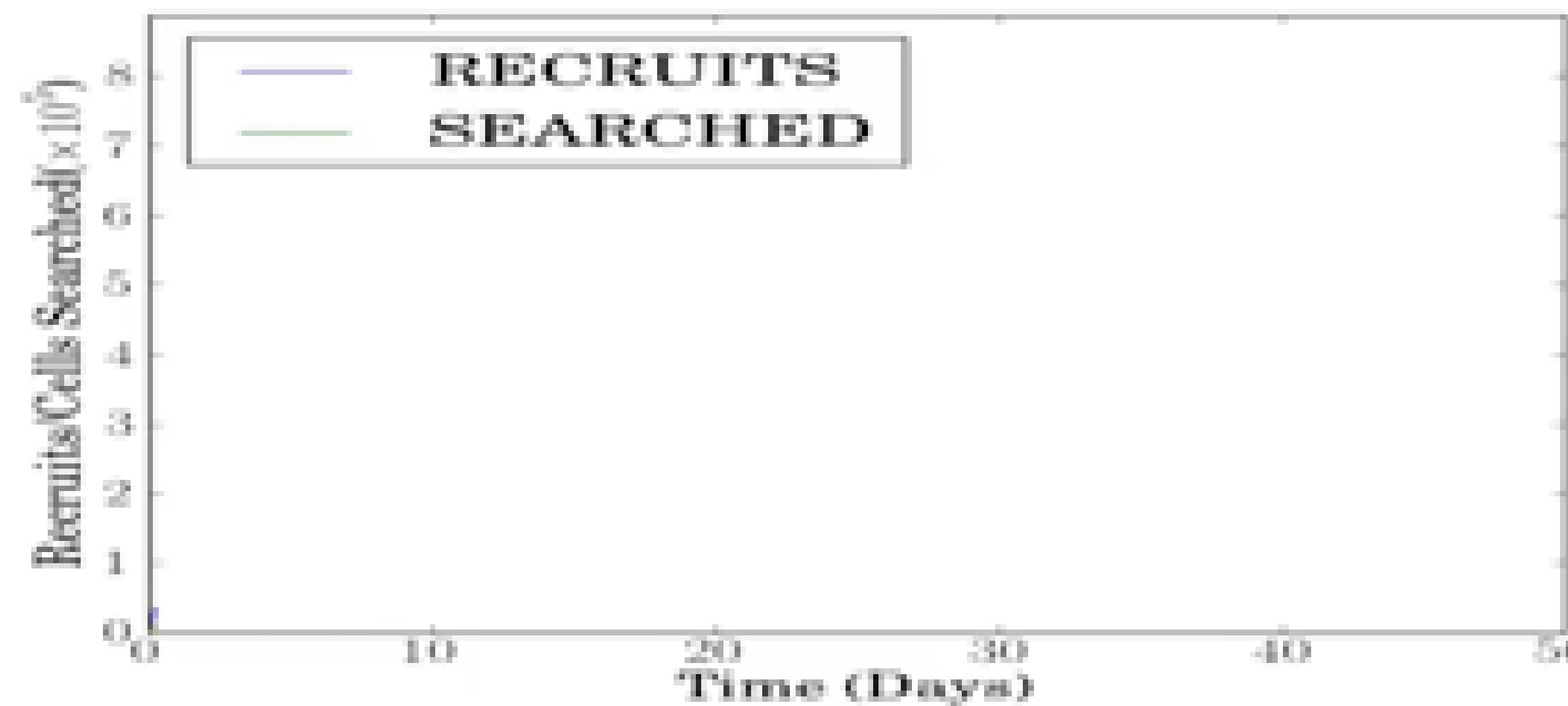
Can we use networks to mobilize people?

1. Select a seed [@MIT]
2. Wait for a response time
3. Recruit a number of active/passive new members
4. Choose them on short/large distances
5. If balloon is in the search area of the recruit => found
6. Proceed to 2 with the active recruits. If none, stop





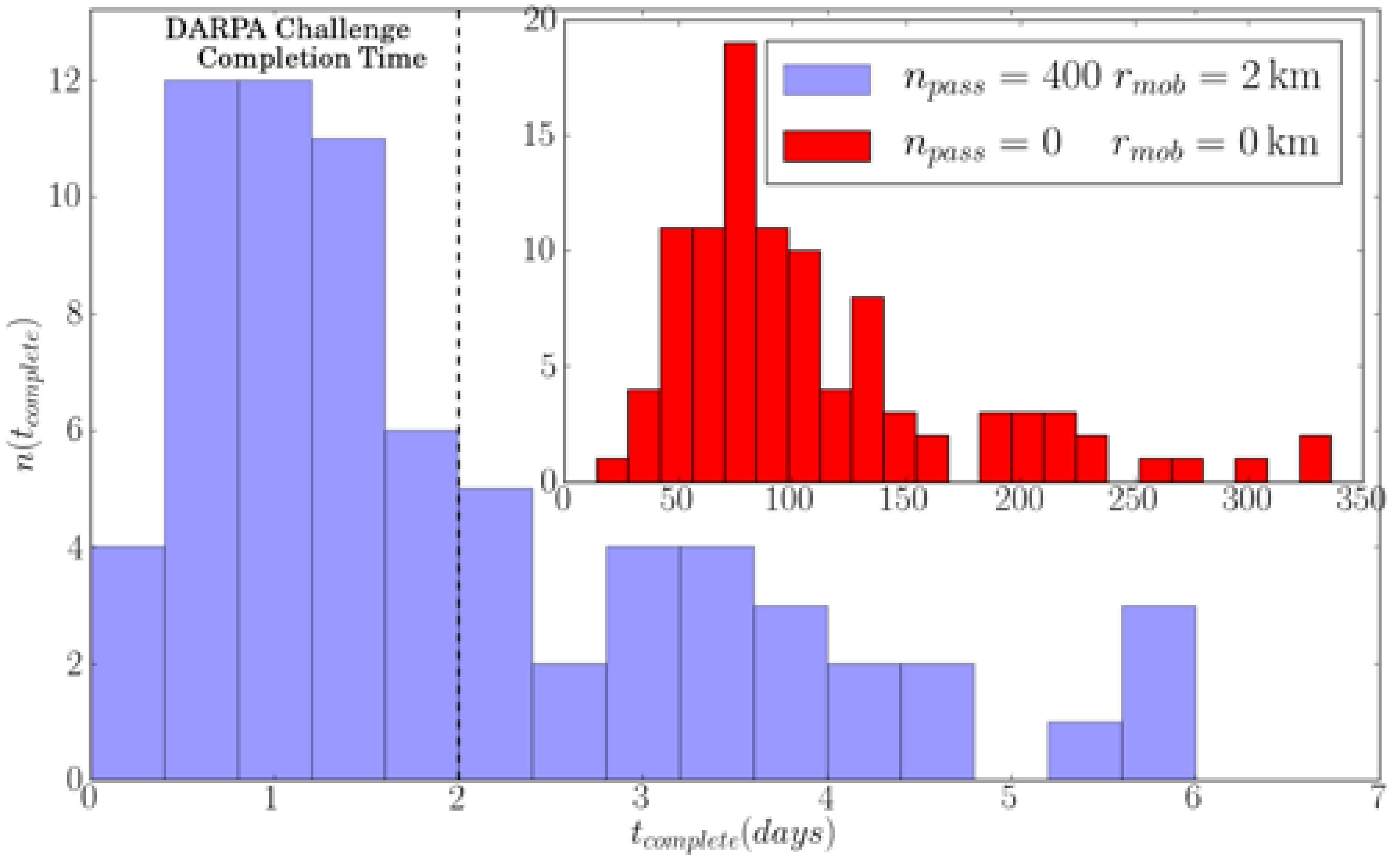
0.25 DAYS

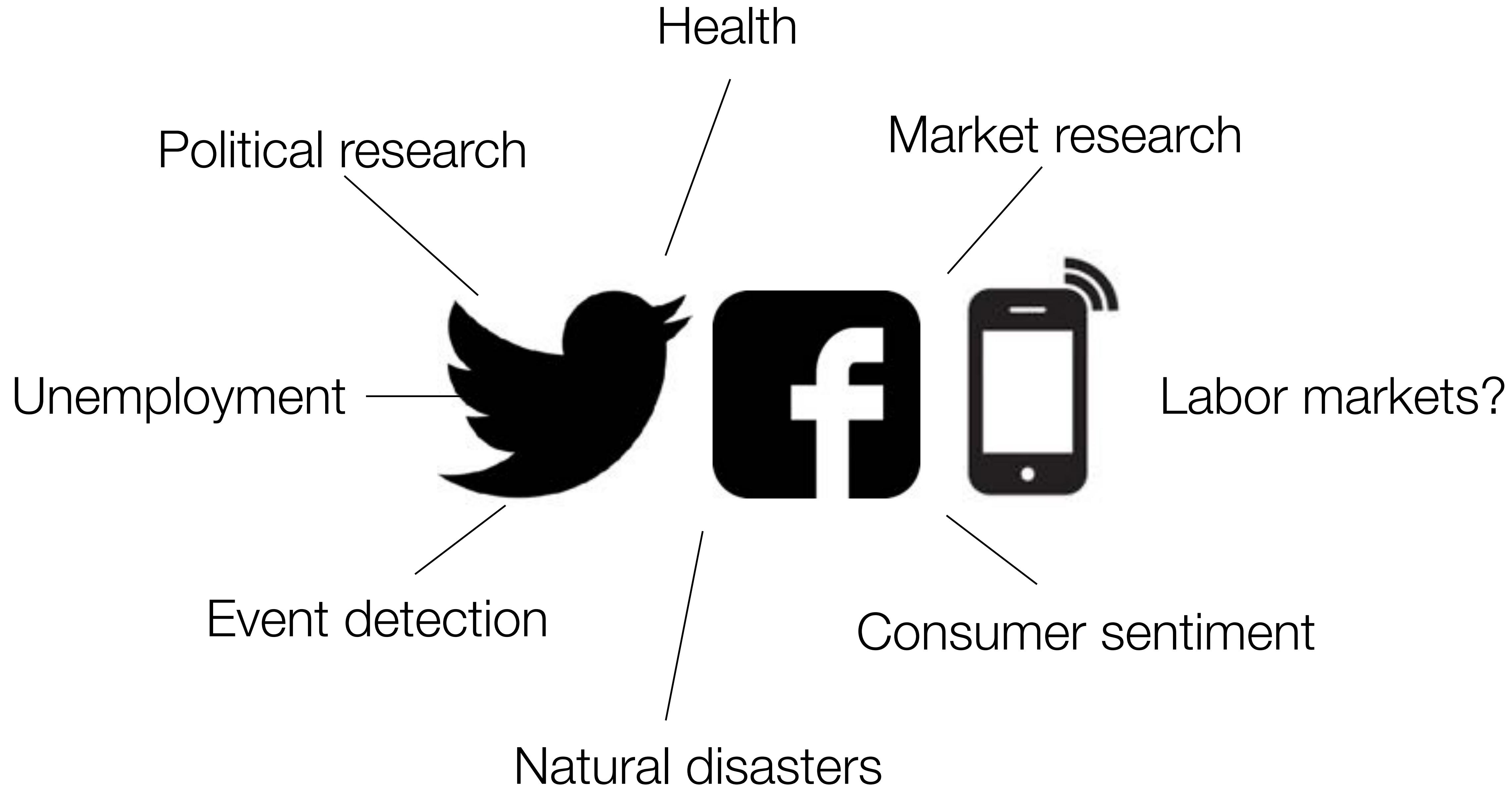


Can we use networks to mobilize people?

Time to completion
(8 balloons found)

Yes, it can (risky).
But its success depends
on:
* Incentives to
participate/search.
* Incentives to recruit
* Use of geographical
heuristics







Societies:

- Spread of diseases
- Social influence
- Privacy
- Product adoption
- Marketing

Economies:

- Loan Repayment
- Food consumption and poverty indices
- Microcredit approval
- Labor market

Crowds:

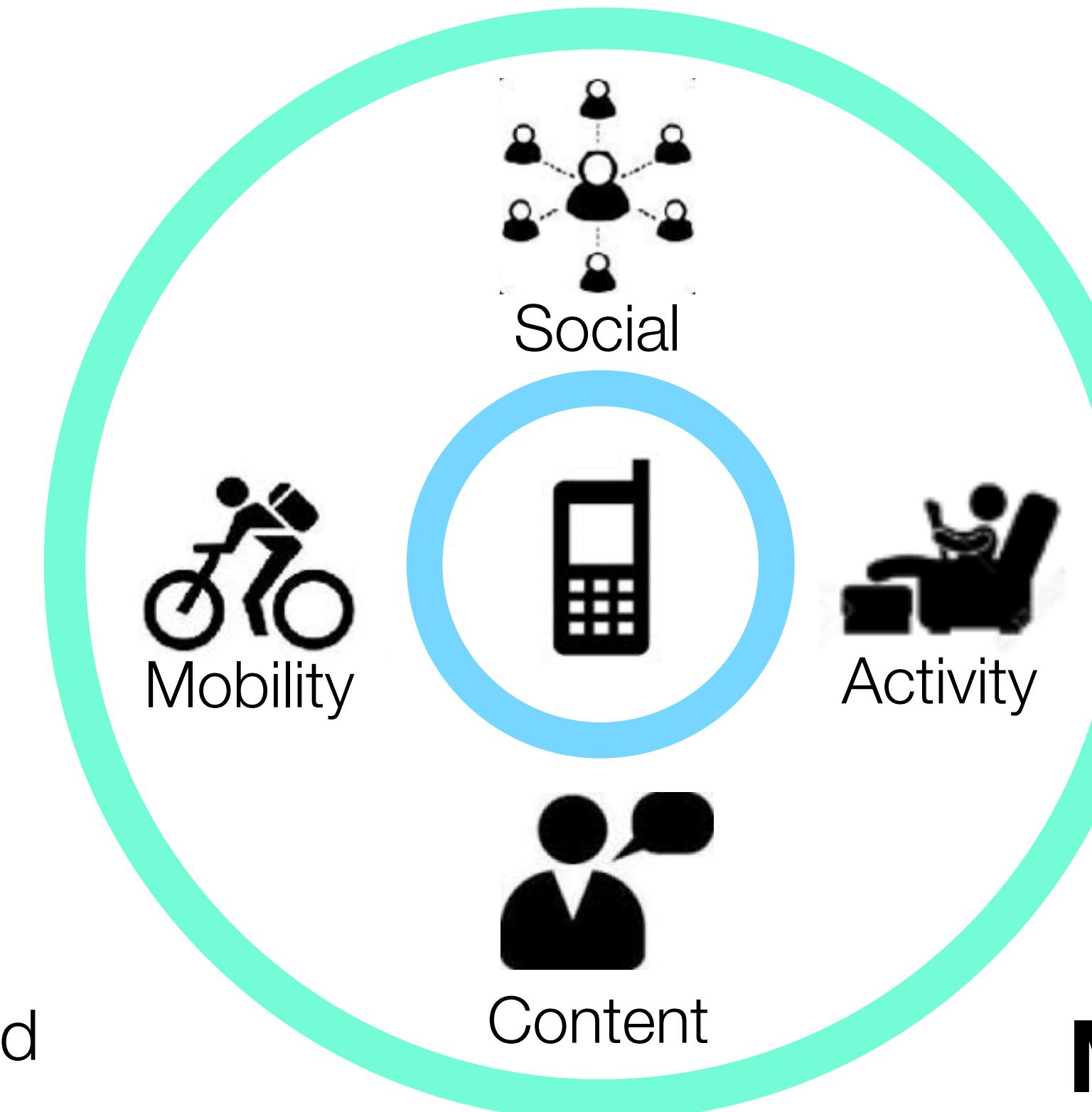
- Real time event detection in cities
- Estimating attendance of events

Cities:

- Energy consumption
- Predicting crime hotspots
- Health catchment areas
- Census estimation

Mobility:

- Mobility prediction
- Impact of Sharing Economy
- Optimization of public transportation



References

- **Facebook Gender Divide**
 - Garcia, D., Kassa, Y. M., Cuevas, A., Cebrian, M., Moro, E., Rahwan, I., & Cuevas, R. (2018). Analyzing gender inequality through large-scale Facebook advertising data. *Proceedings of the National Academy of Sciences*, 303, 201717781. <http://doi.org/10.1073/pnas.1717781115>
- Social media and **extreme weather**
 - Baylis, P., Obradovich, N., Kryvasheyeu, Y., Chen, H., Coviello, L., Moro, E., et al. (2018). Weather impacts expressed sentiment. *PLoS ONE*, 13(4), e0195750. <http://doi.org/10.1371/journal.pone.0195750>
- Social media and **disasters**
 - Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., & Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. *Science Advances*, 2(3), e1500779–e1500779. <http://doi.org/10.1126/sciadv.1500779>
 - Kryvasheyeu, Y., Chen, H., Moro, E., Van Hentenryck, P., & Cebrian, M. (2015). Performance of Social Network Sensors during Hurricane Sandy. *PLoS ONE*, 10(2), e0117288. <http://doi.org/10.1371/journal.pone.0117288>

References

- Social media and **unemployment**
 - Llorente, A., Garcia-Herranz, M., Cebrian, M., & Moro, E. (2015). Social media fingerprints of unemployment. PLoS ONE, 10(5), e0128692. <http://doi.org/10.1371/journal.pone.0128692>
- Social media and **viral spreading**
 - Garcia-Herranz, M., Moro, E., Cebrian, M., Christakis, N. A., & Fowler, J. H. (2014). Using Friends as Sensors to Detect Global-Scale Contagious Outbreaks. PLoS ONE, 9(4), e92413. <http://doi.org/10.1371/journal.pone.0092413>
 - Grabowicz, P. A., Ramasco, J. J., Moro, E., Pujol, J. M., & Eguiluz, V. M. (2012). Social Features of Online Networks: The Strength of Intermediary Ties in Online Social Media. PLoS ONE, 7(1), e29358. <http://doi.org/10.1371/journal.pone.0029358>
- Social networks and **mobilization**
 - Rutherford, A., Cebrian, M., D'souza, S., Moro, E., Pentland, A., & Rahwan, I. (2013). Limits of social mobilization., 110(16), 6281–6286. <http://doi.org/10.1073/pnas.1216338110>