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

Esteban Moro

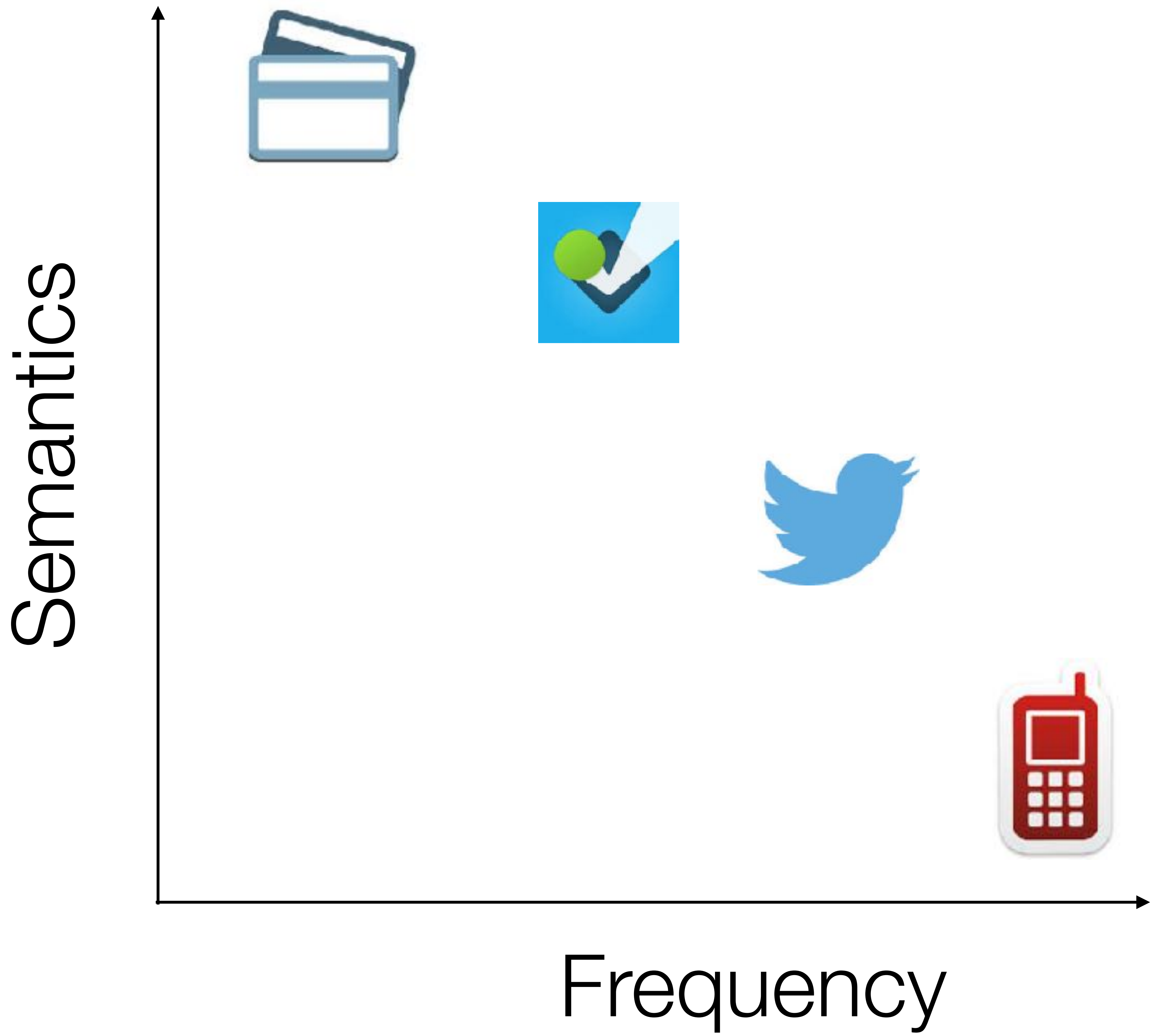
mit media lab

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



Dong, L., Chen, S., Cheng, Y., Wu, Z., Li, C., & Wu, H. (2017). Measuring economic activity in China with mobile big data

Why alternative data?

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

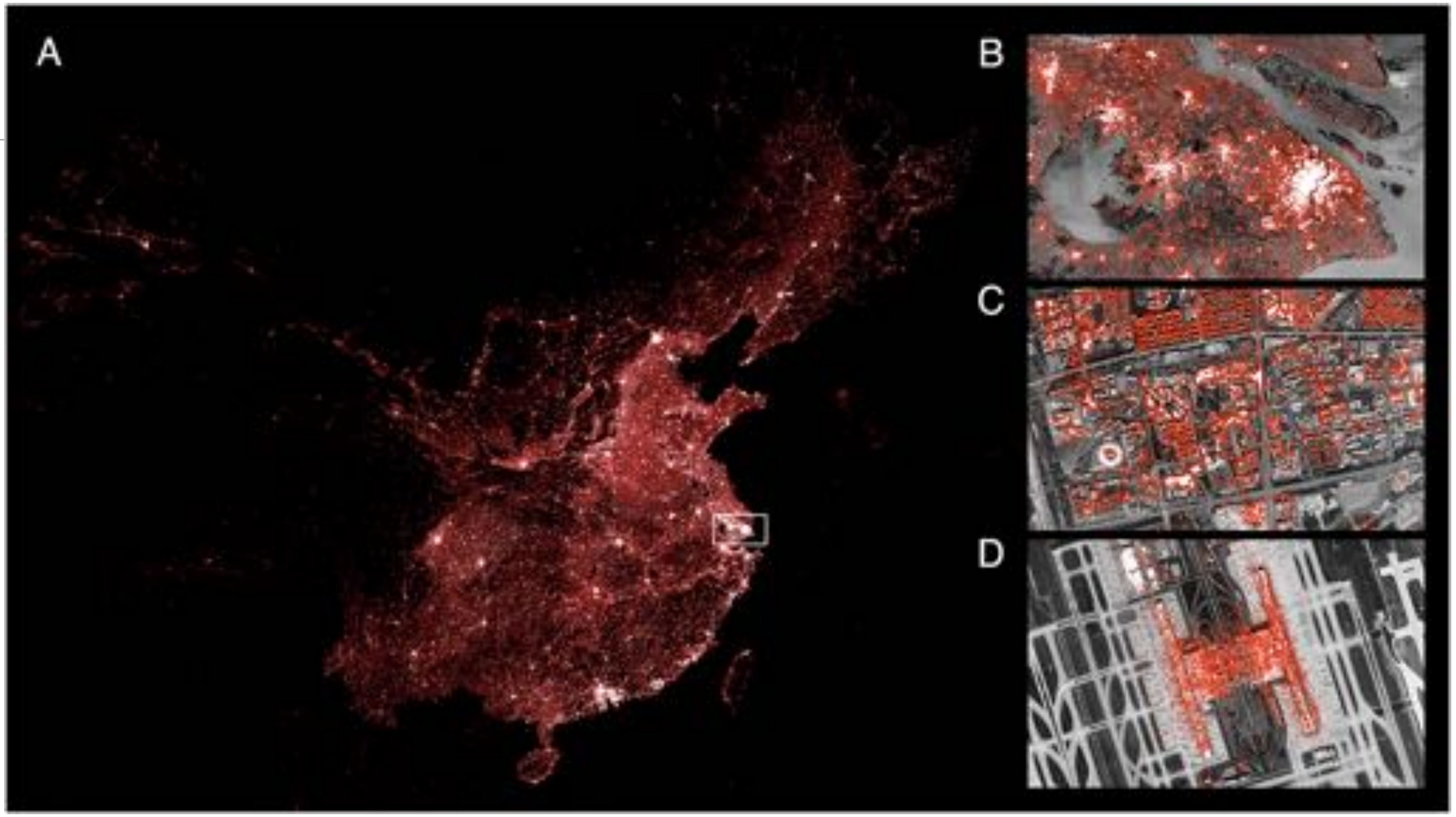
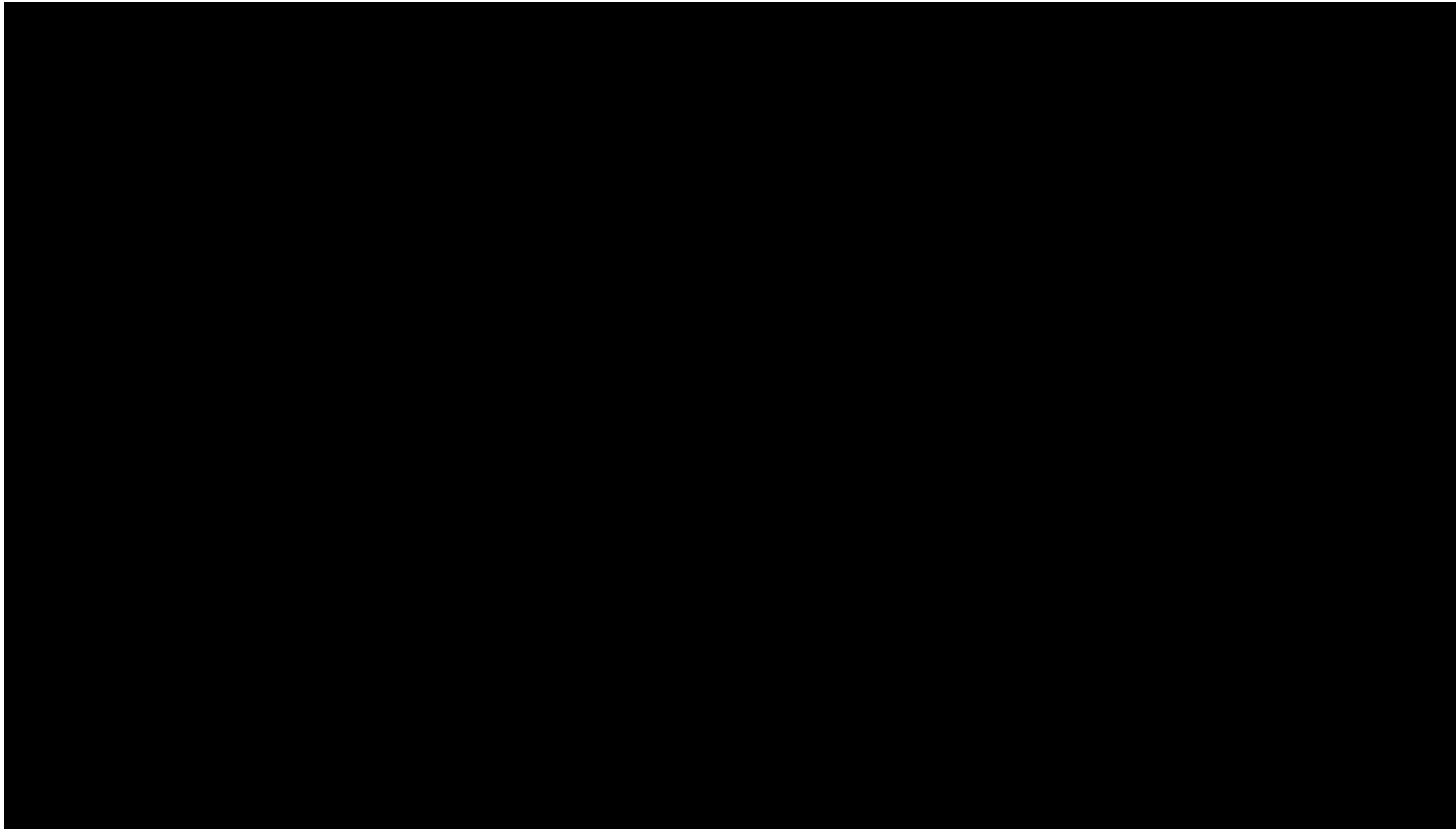


Figure 1 Spatial-temporal big data reflects human activity at different scales. (A) At the national level: Data points depict the fact that most of China's population is concentrated in large cities in the east. As the brightness of the spot increases, the aggregation of the data points (population) 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

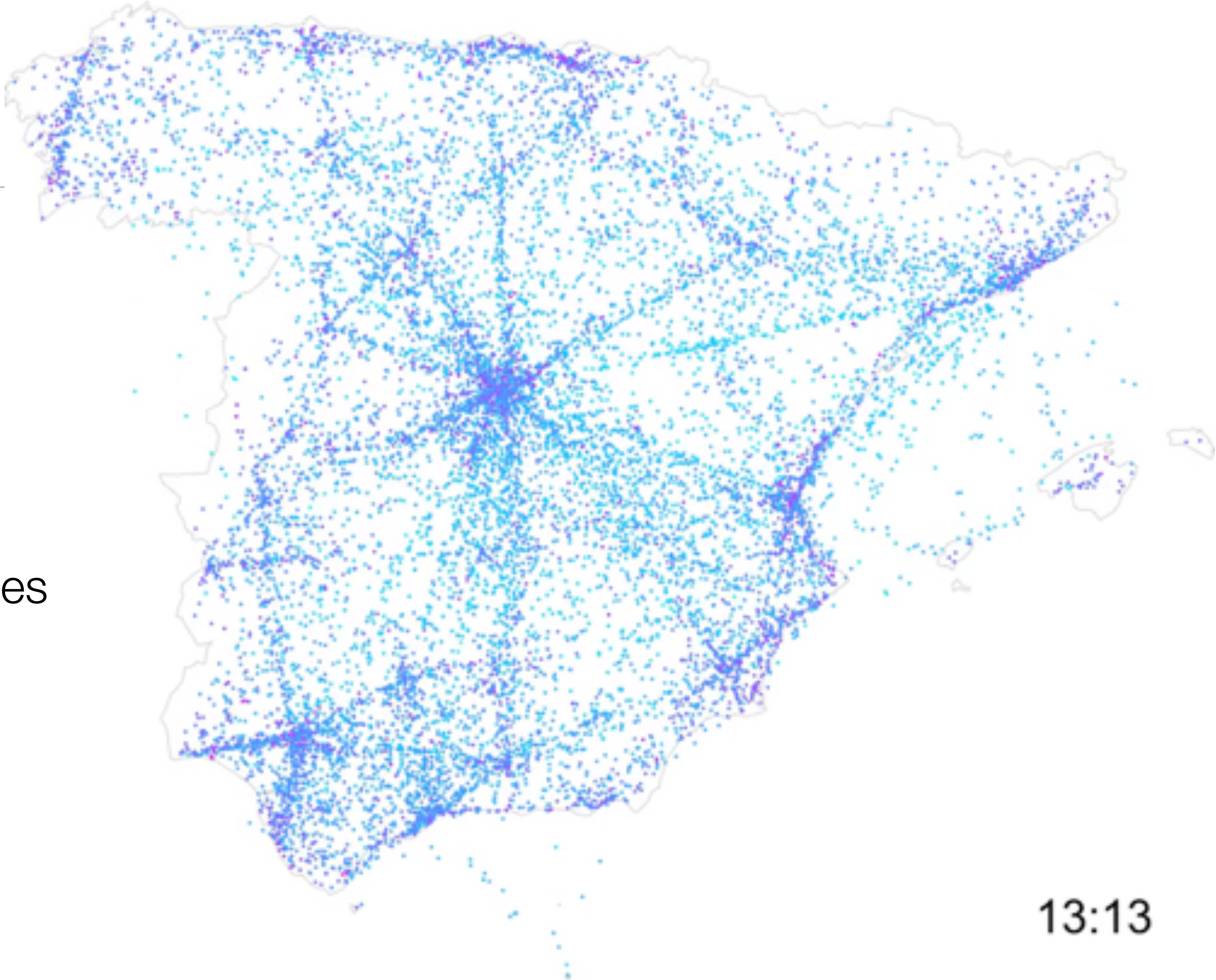
Unemployment

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

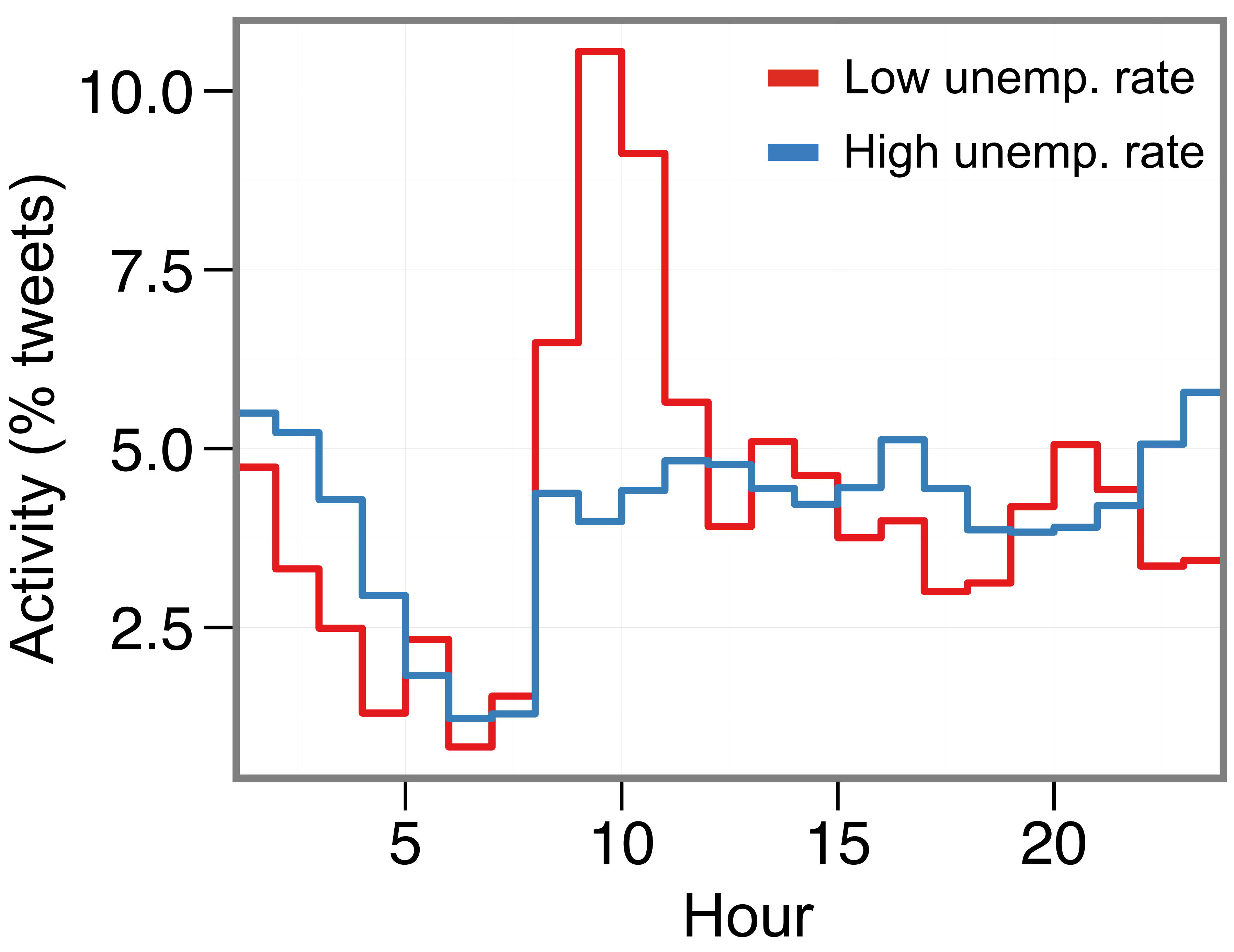


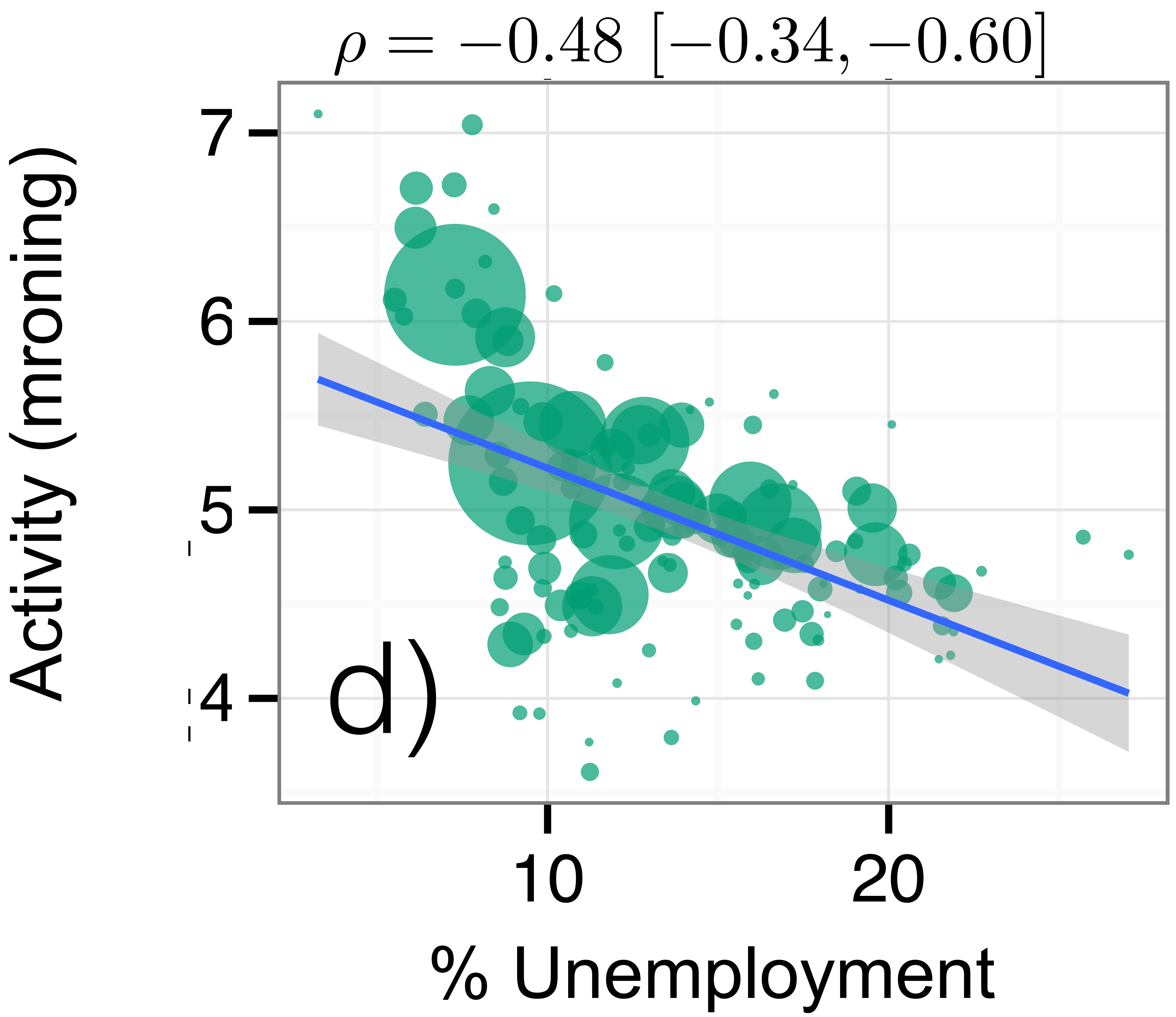
Llorente, A., Garcia-Herranz, M., Cebrian, M., & Moro, E. (2015). Social media fingerprints of unemployment. PLoS ONE, 10(5), e0128692. http://doi.org/10.1371/journal.pone.0128692

Unemployment

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

Just arrived to work, mondays are too hard…
Tweet 10:43 - 2 de jun. de 2014



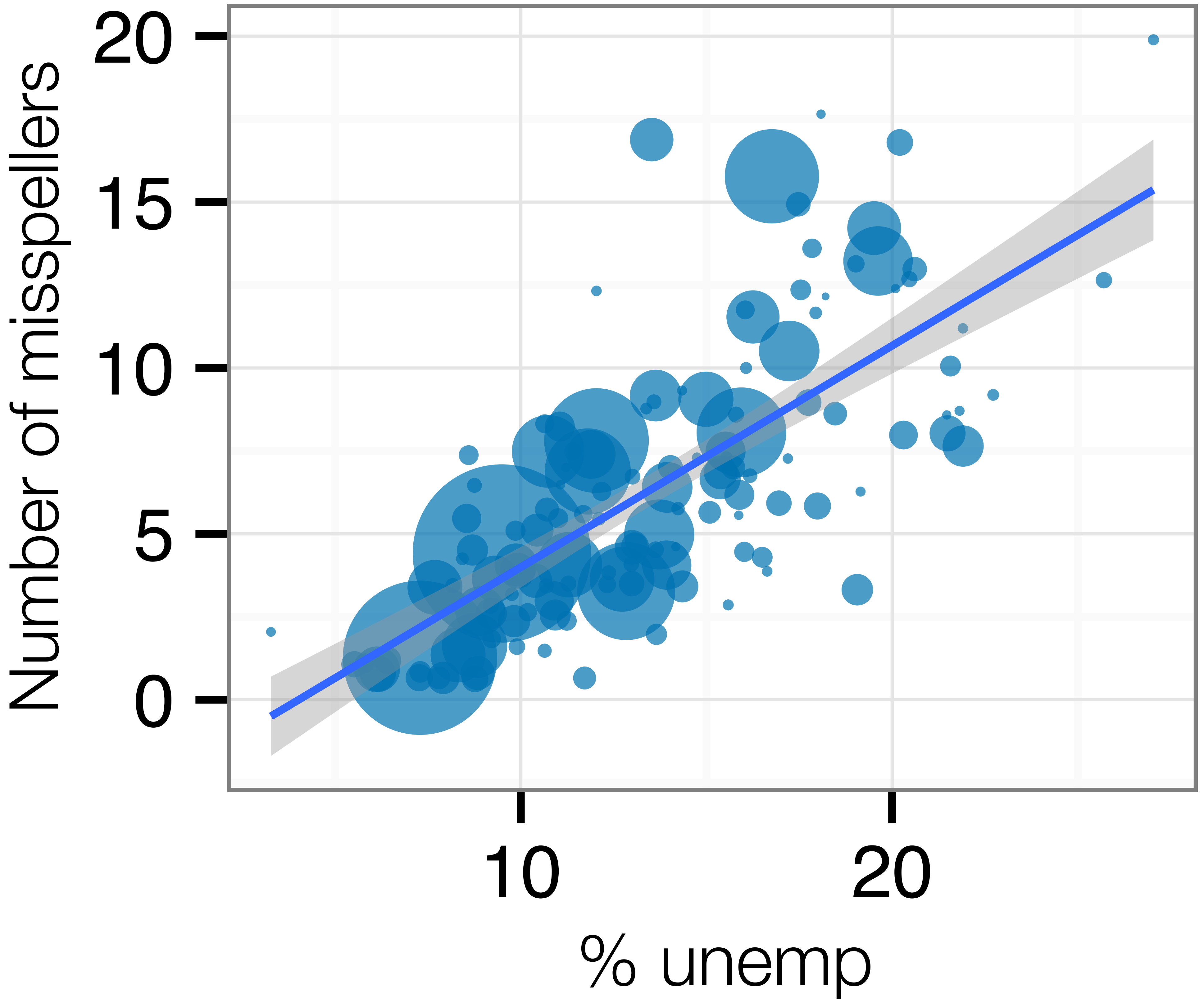


Llorente, A., Garcia-Herranz, M., Cebrian, M., & Moro, E. (2015). Social media fingerprints of unemployment. PLoS ONE, 10(5), e0128692. http://doi.org/10.1371/journal.pone.0128692

Unemployment

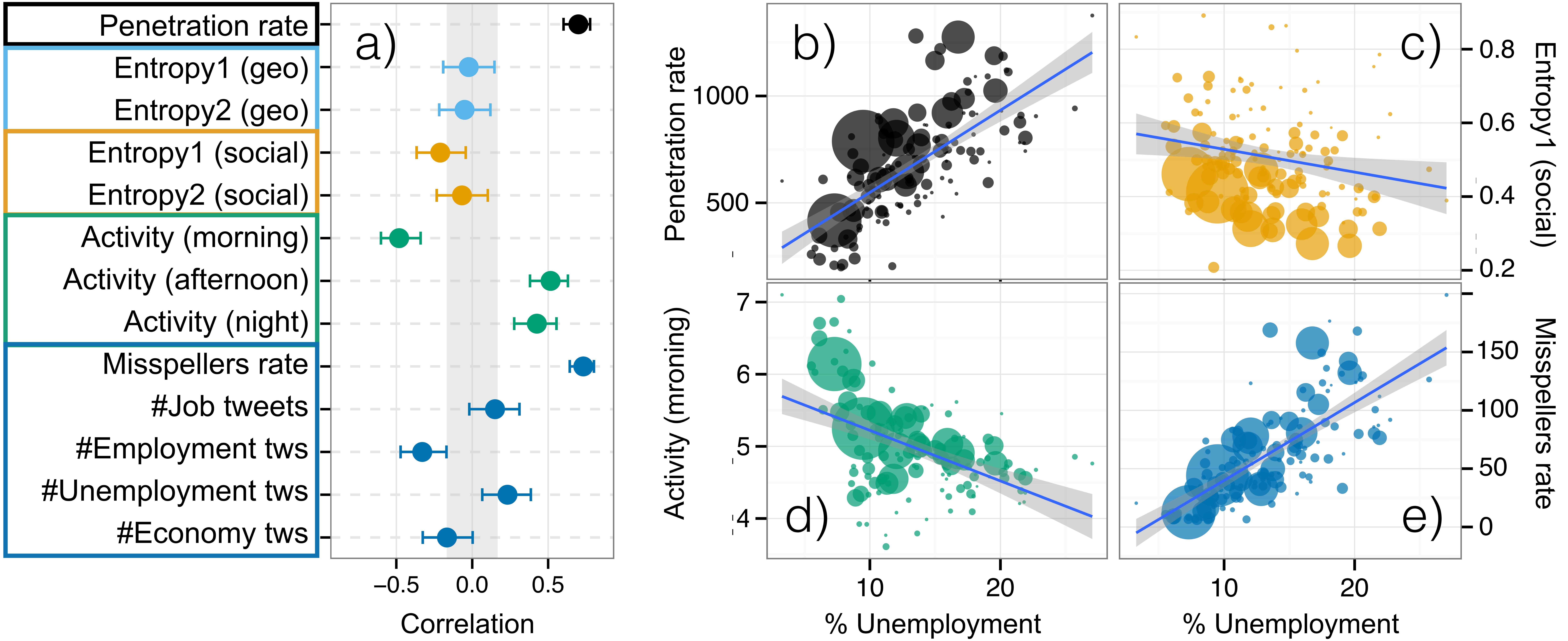
- Tweet's content:
  - Misspelling is highly correlated with unemployment

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



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

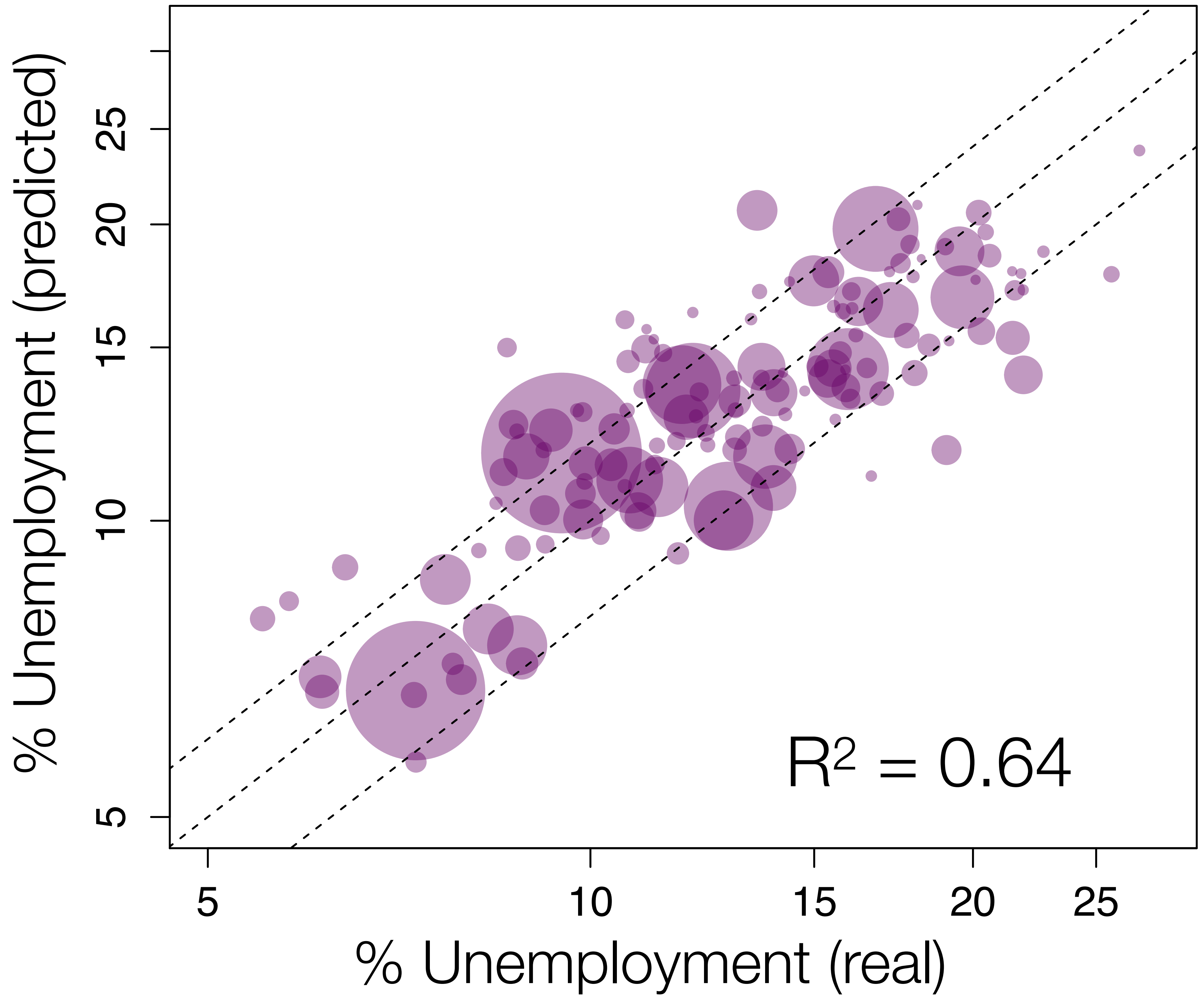
Unemployment

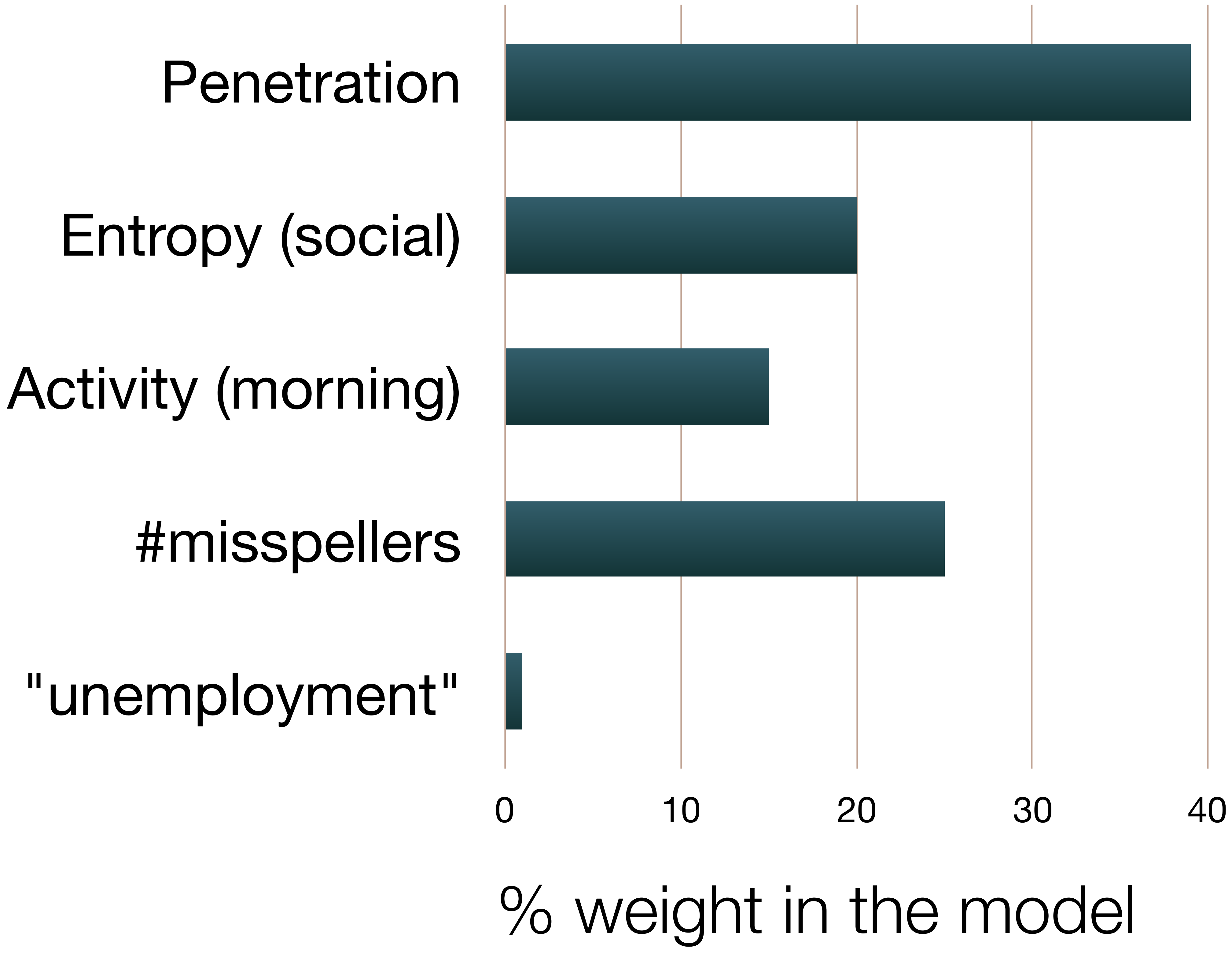


Llorente, A., Garcia-Herranz, M., Cebrian, M., & Moro, E. (2015). Social media fingerprints of unemployment. PLoS ONE, 10(5), e0128692. http://doi.org/10.1371/journal.pone.0128692

Unemployment

- Explanatory power of the social media fingerprints





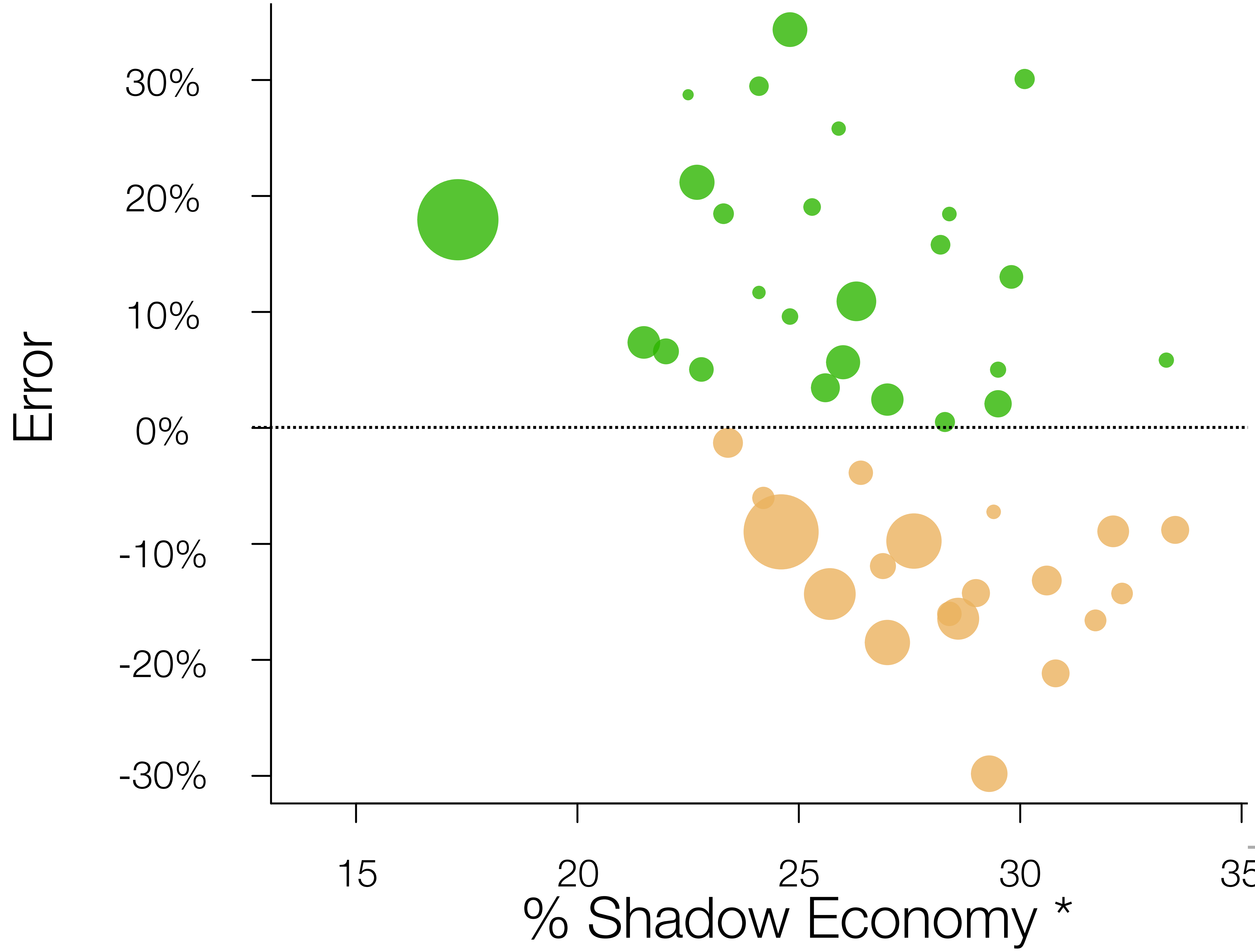
@estebanmoro

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

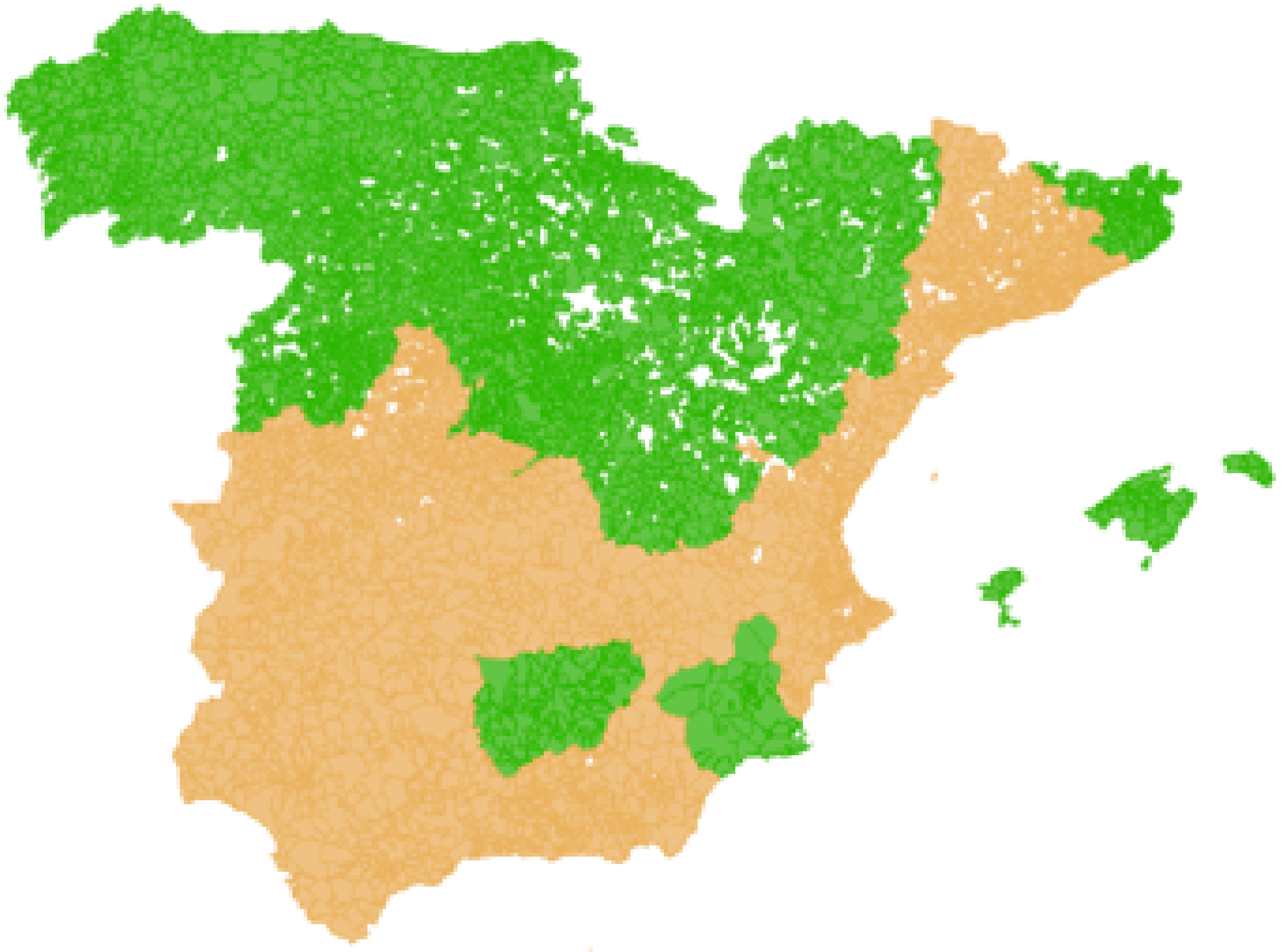
Unemployment

- Are we really wrong?

Model Error = Model[variables] - Official unemployment



(* GESTHA report 2012)



Use twitter to estimate the Human Development Index (HDI)

UNICEF INNOVATION

BRAZIL

COSTA RICA

MEXICO

NIGERIA

POLAND

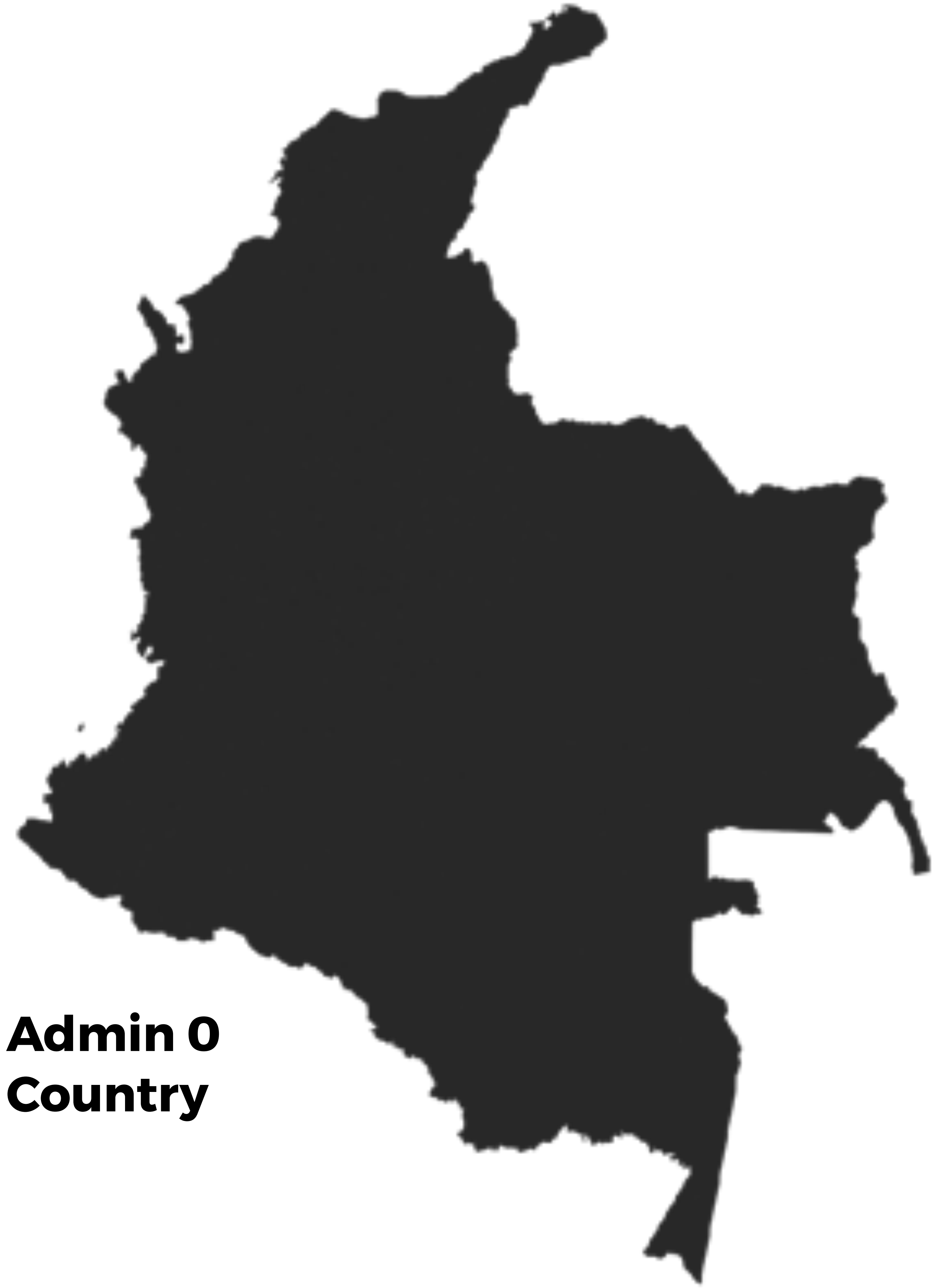
COLOMBIA

INDONESIA

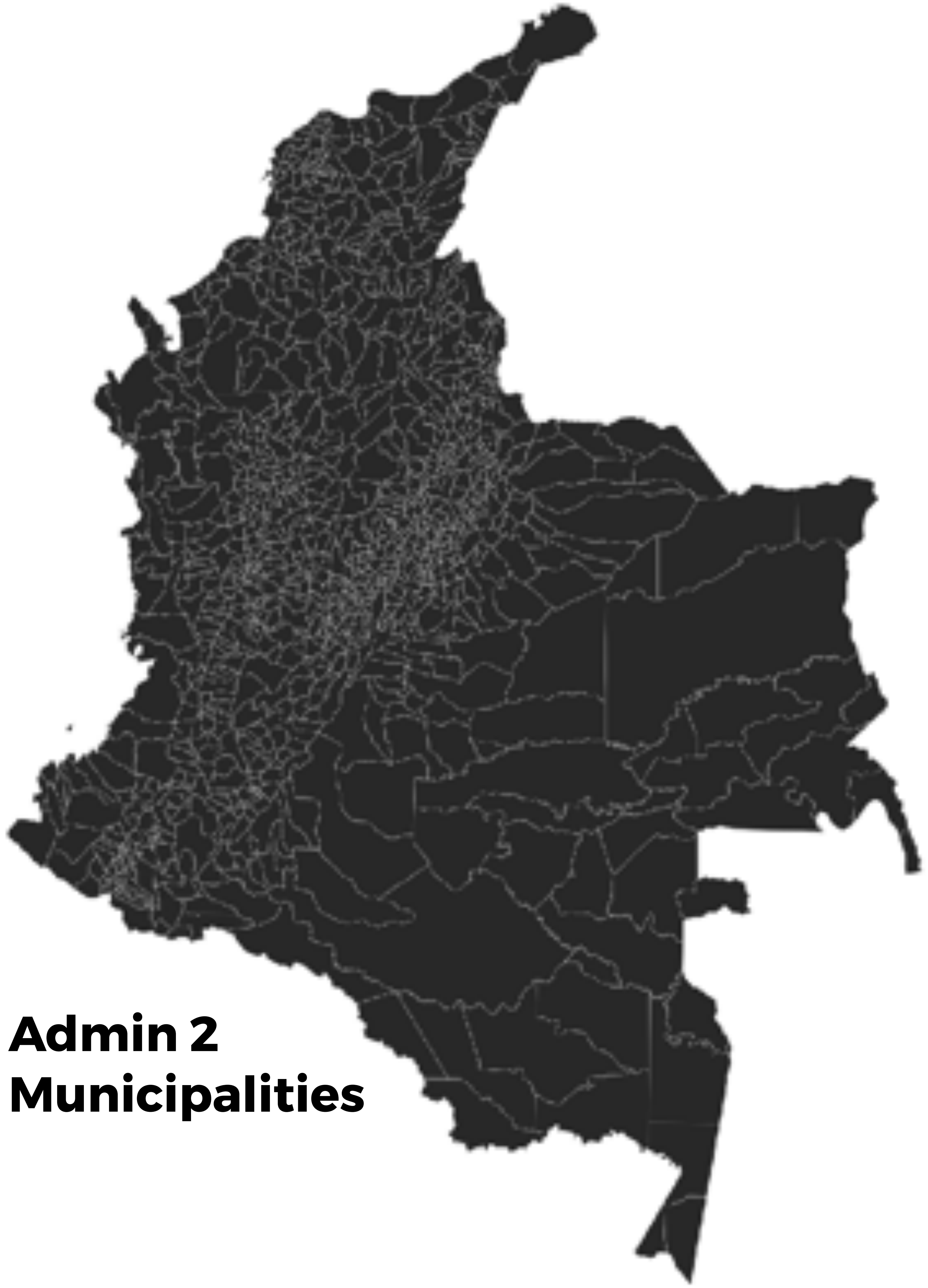
NEPAL

PAKISTAN

Use twitter to estimate HDI on sub-national levels

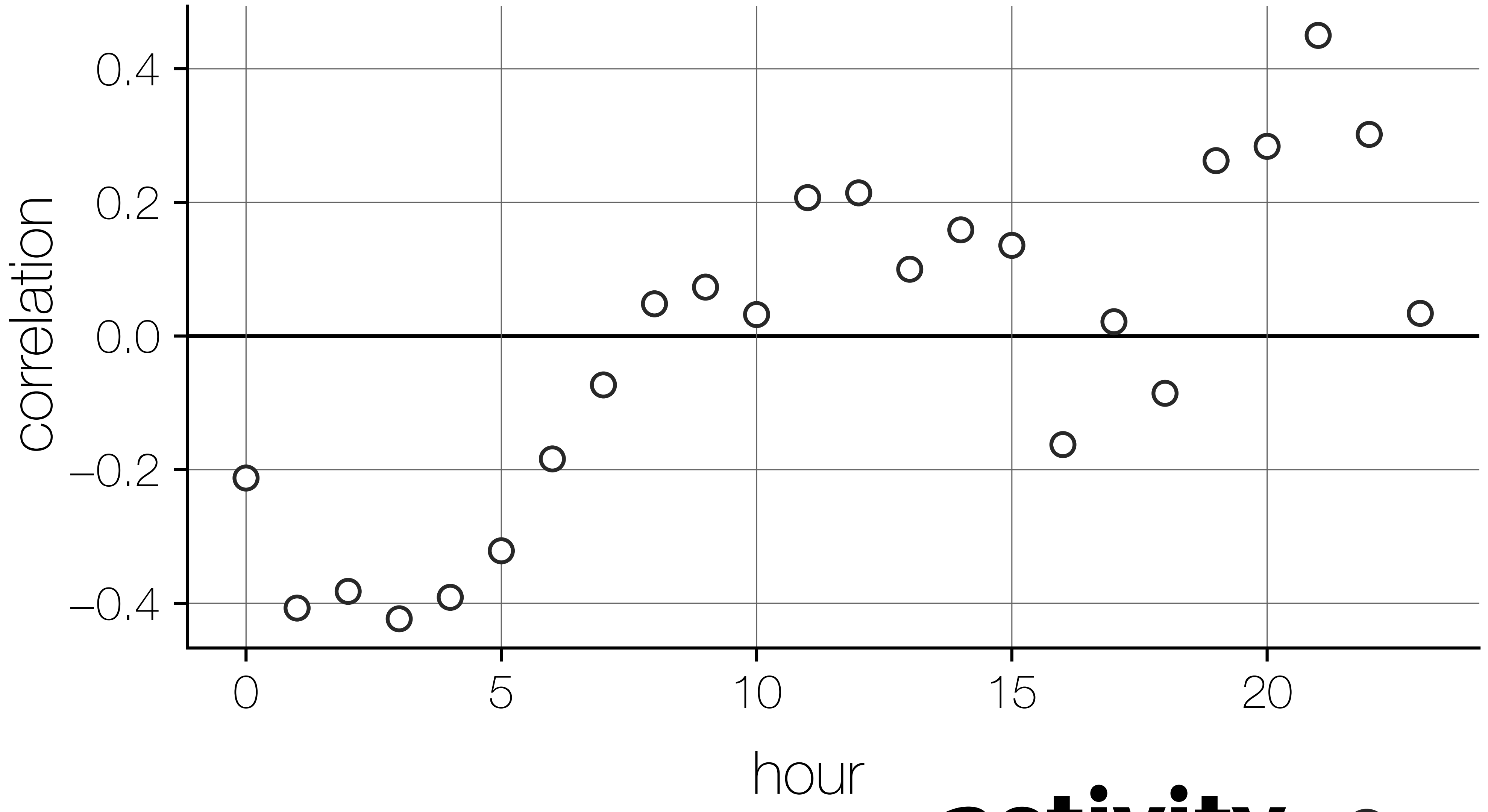


Admin 0
Country



Admin 2
Municipalities

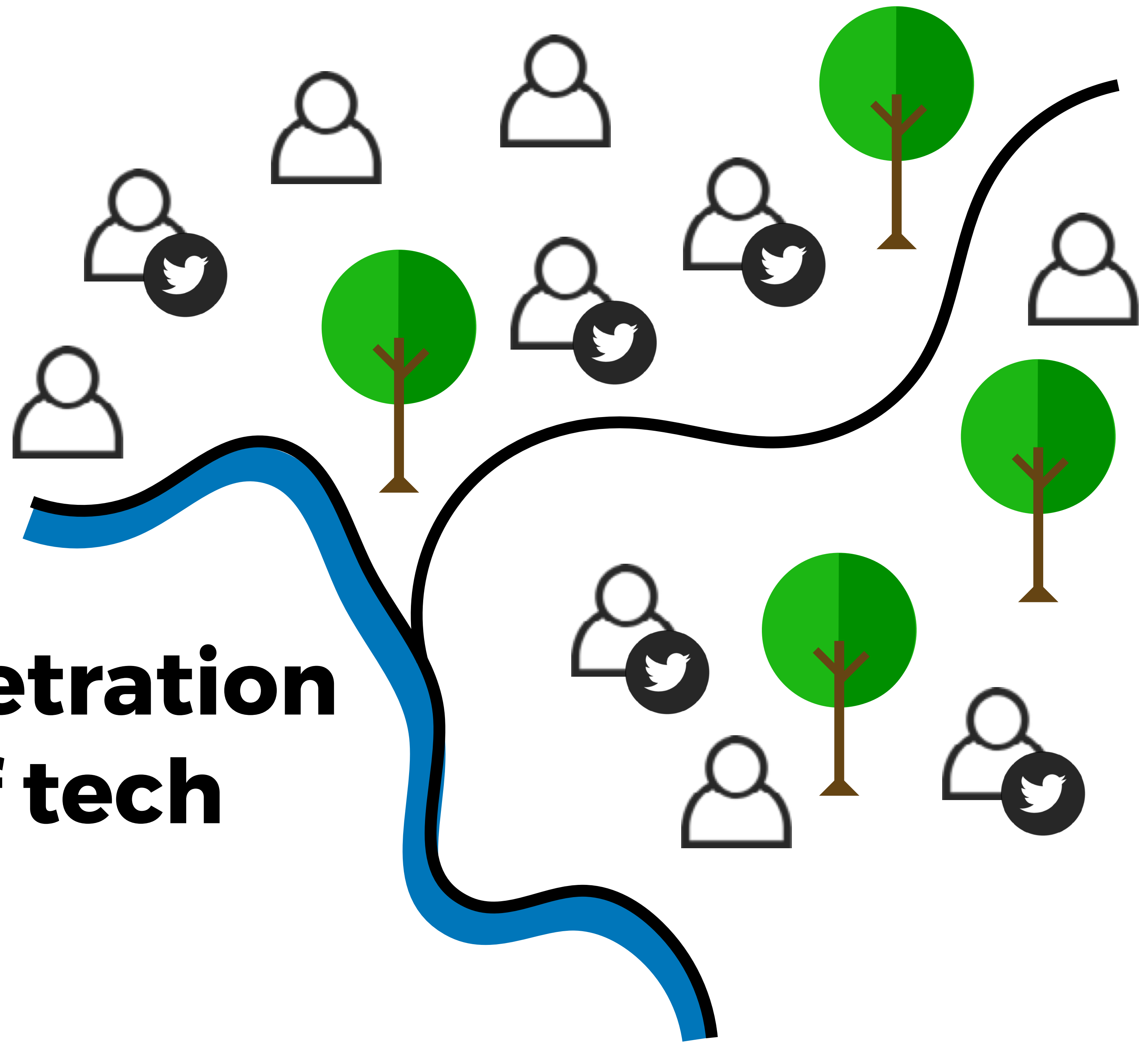
features



activity

- no. tweets
- morning
- daily
- night

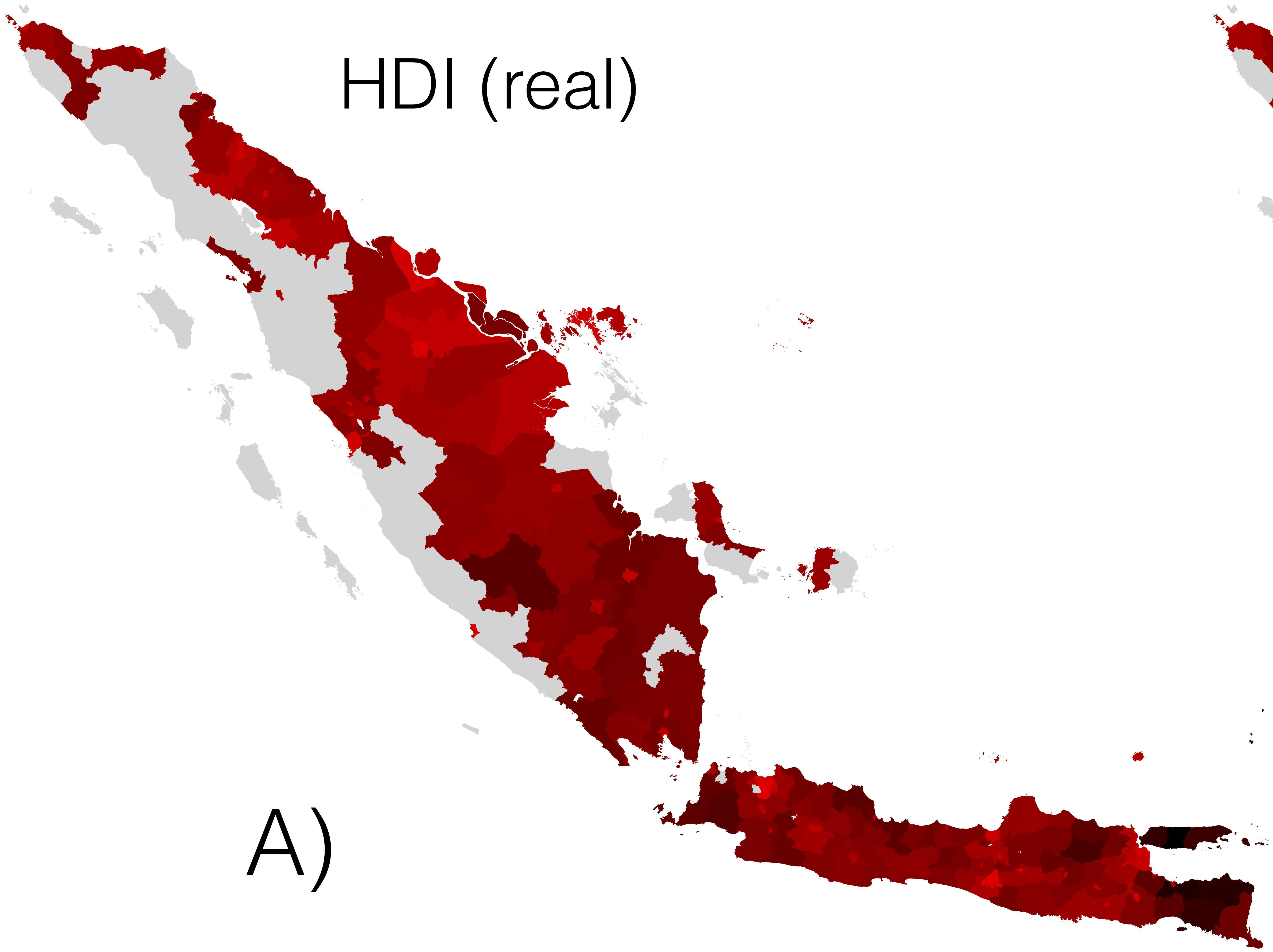
penetration of tech

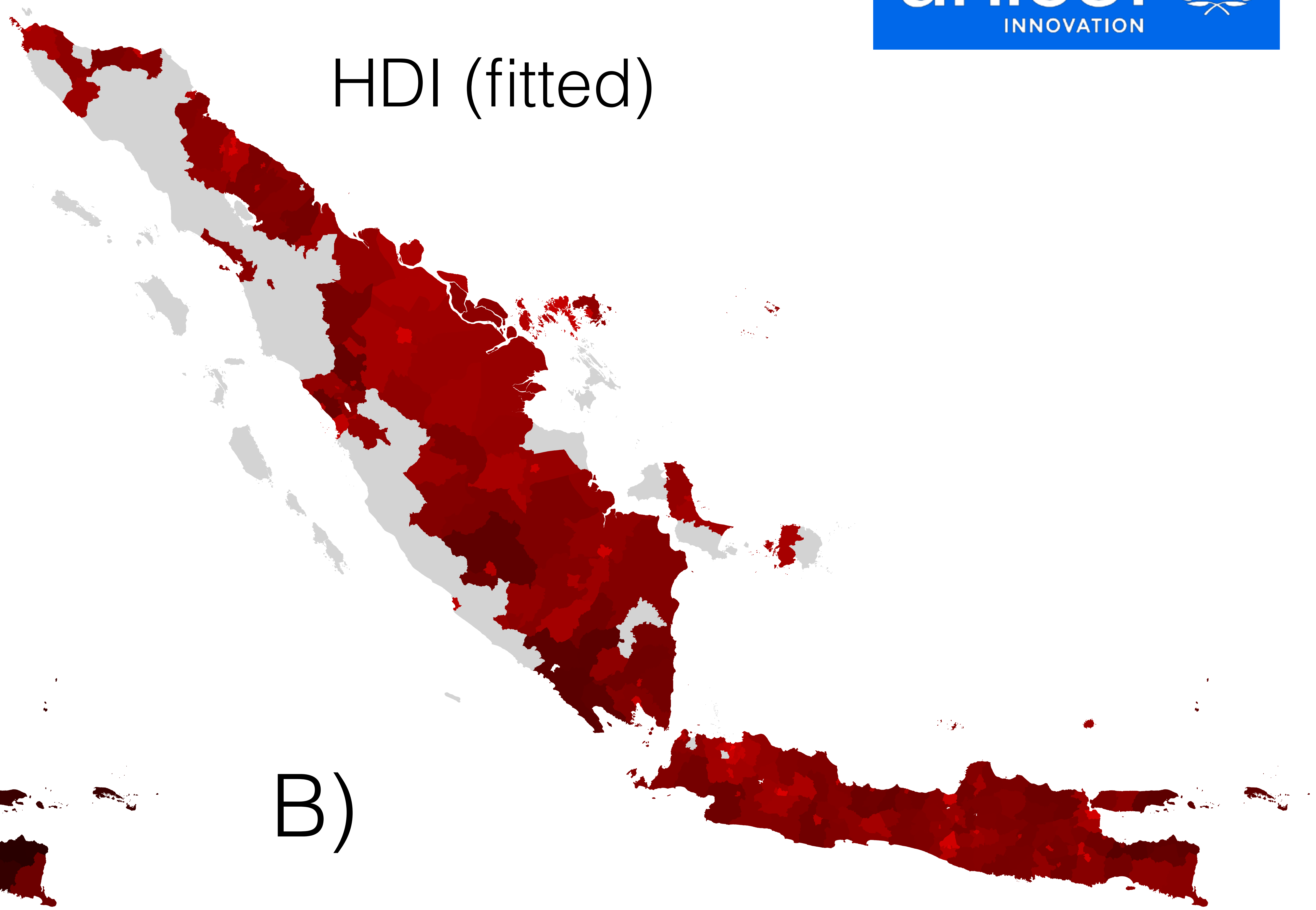


mobility

Human development

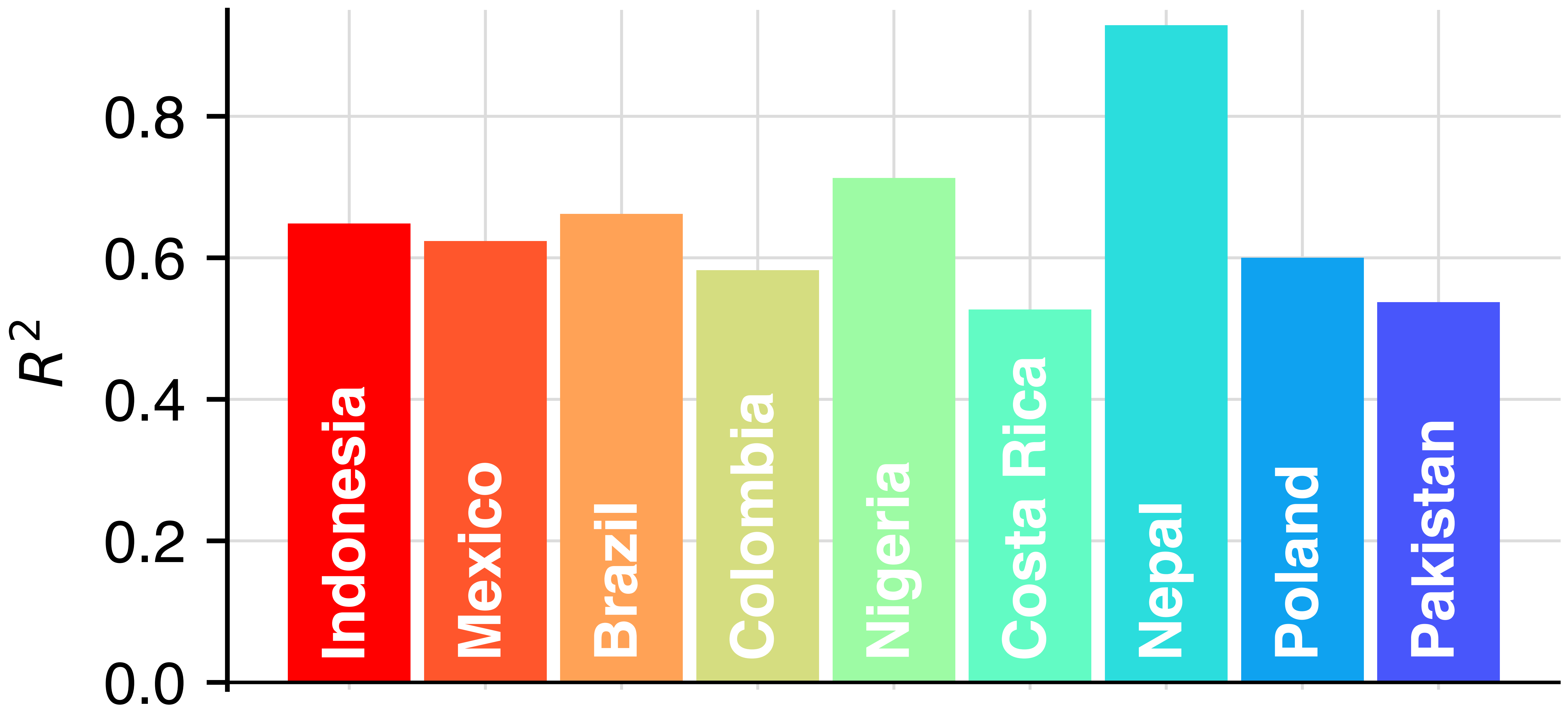
- HDI = Human Development index (sub-national)

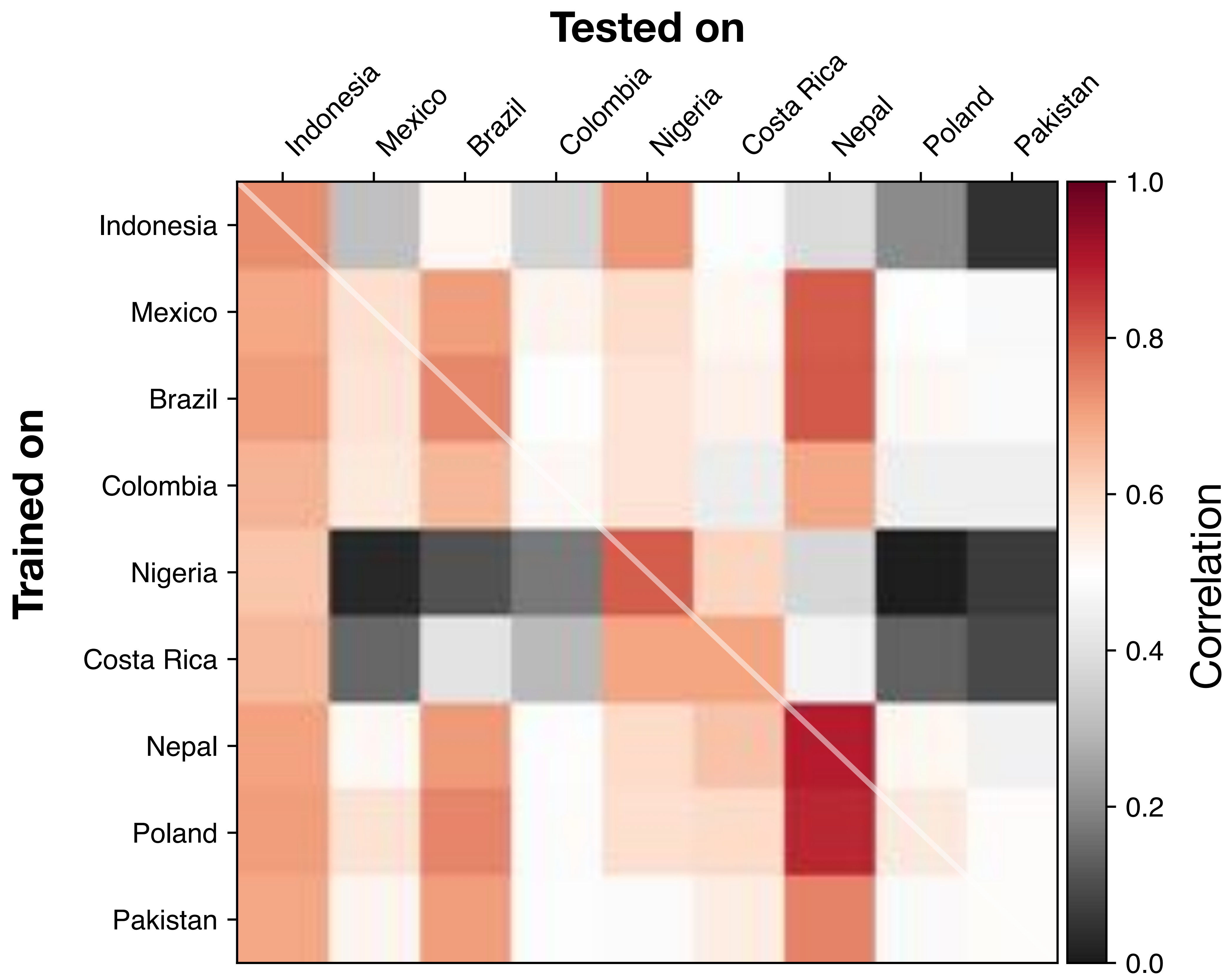




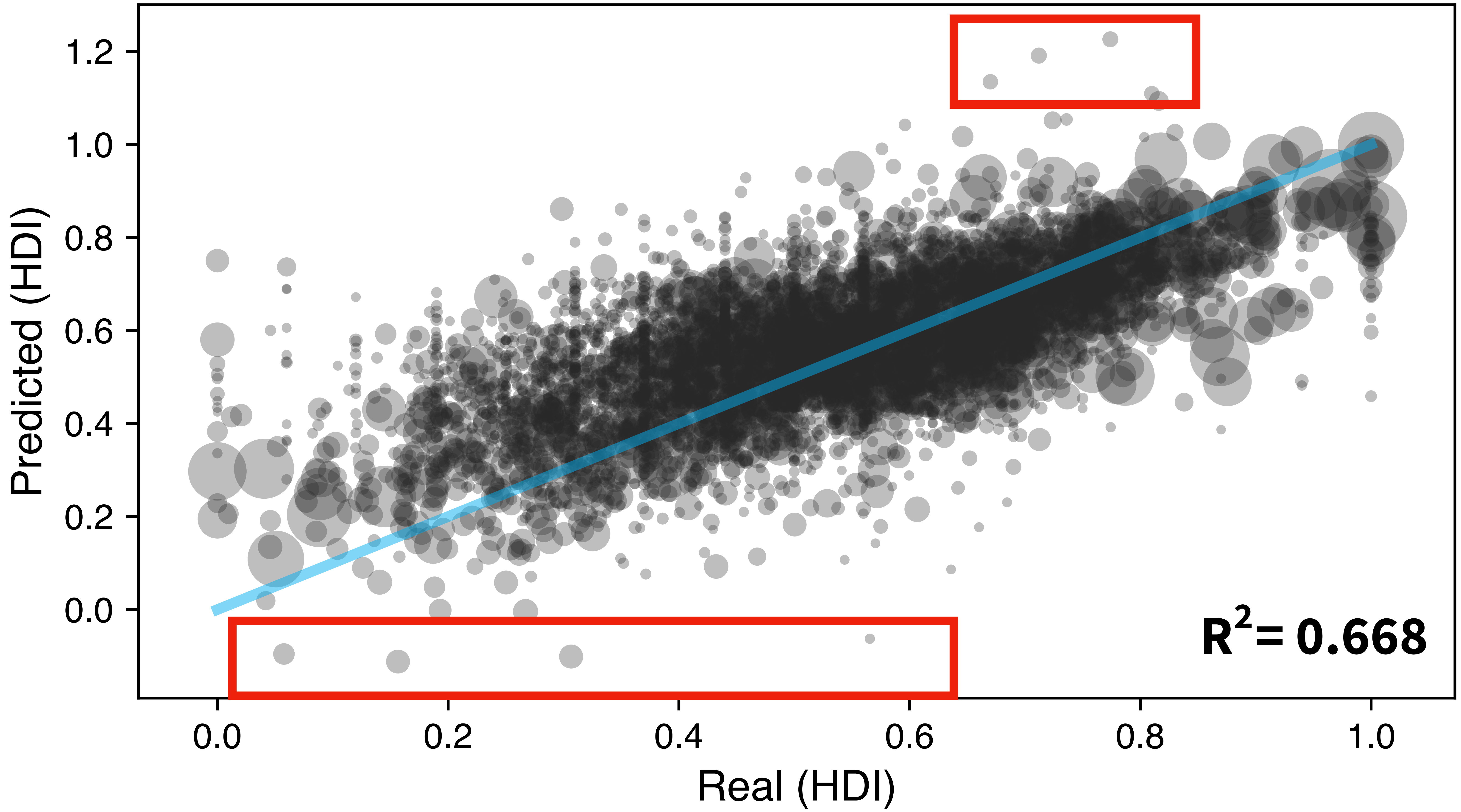
Human development

- Can be re-use these models?

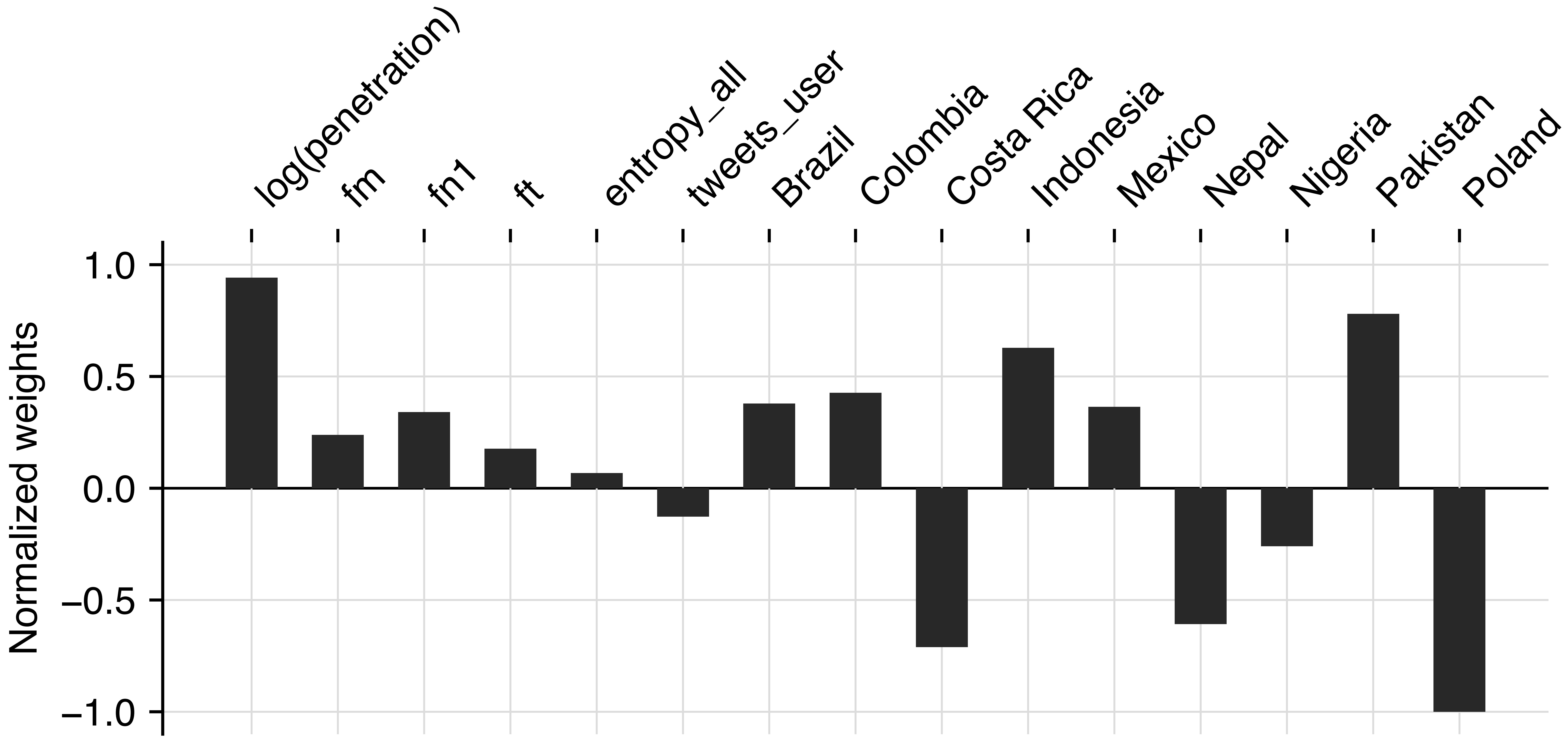




Über model
Building one model for all countries



Über model
Importance of features

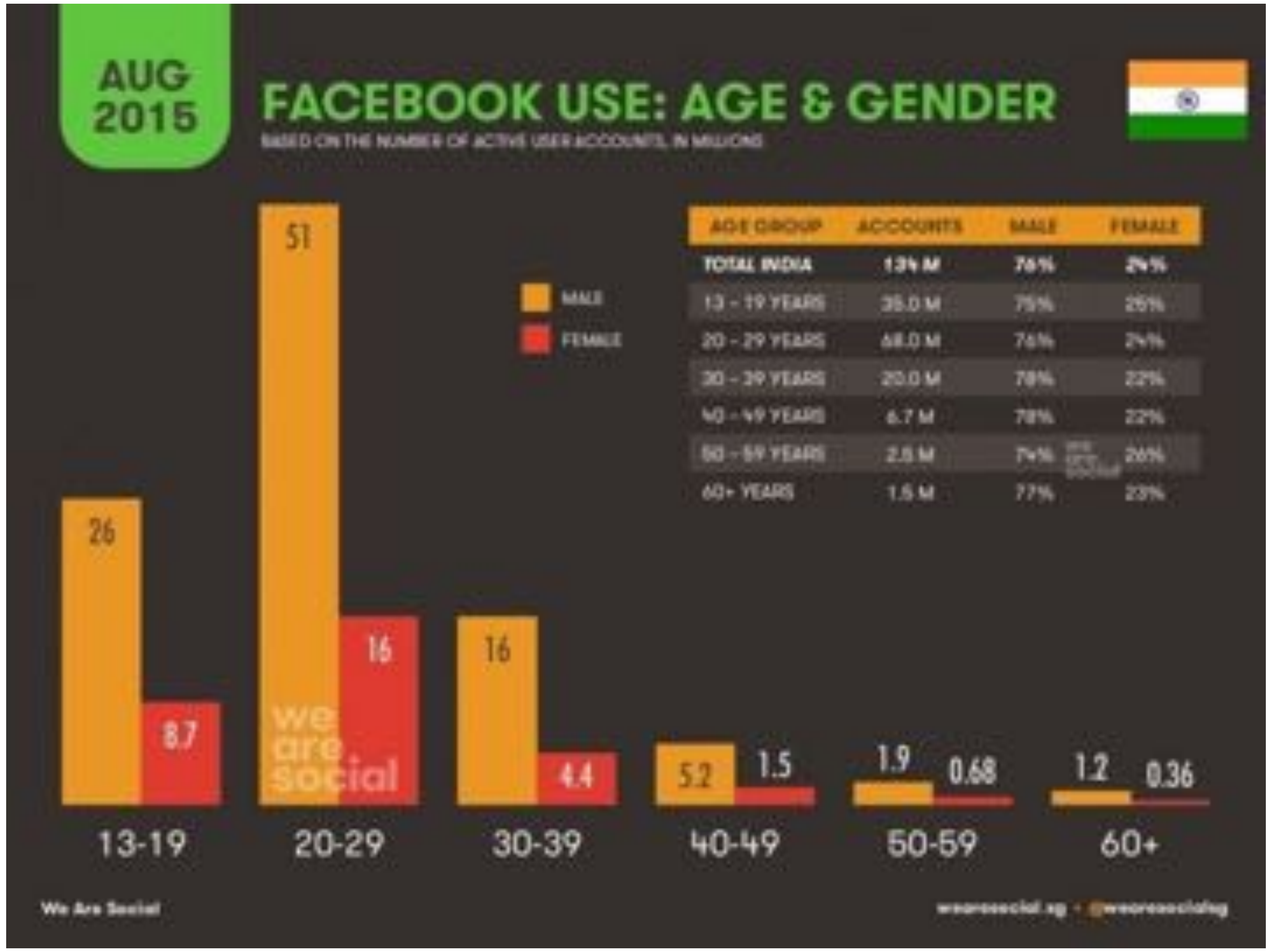


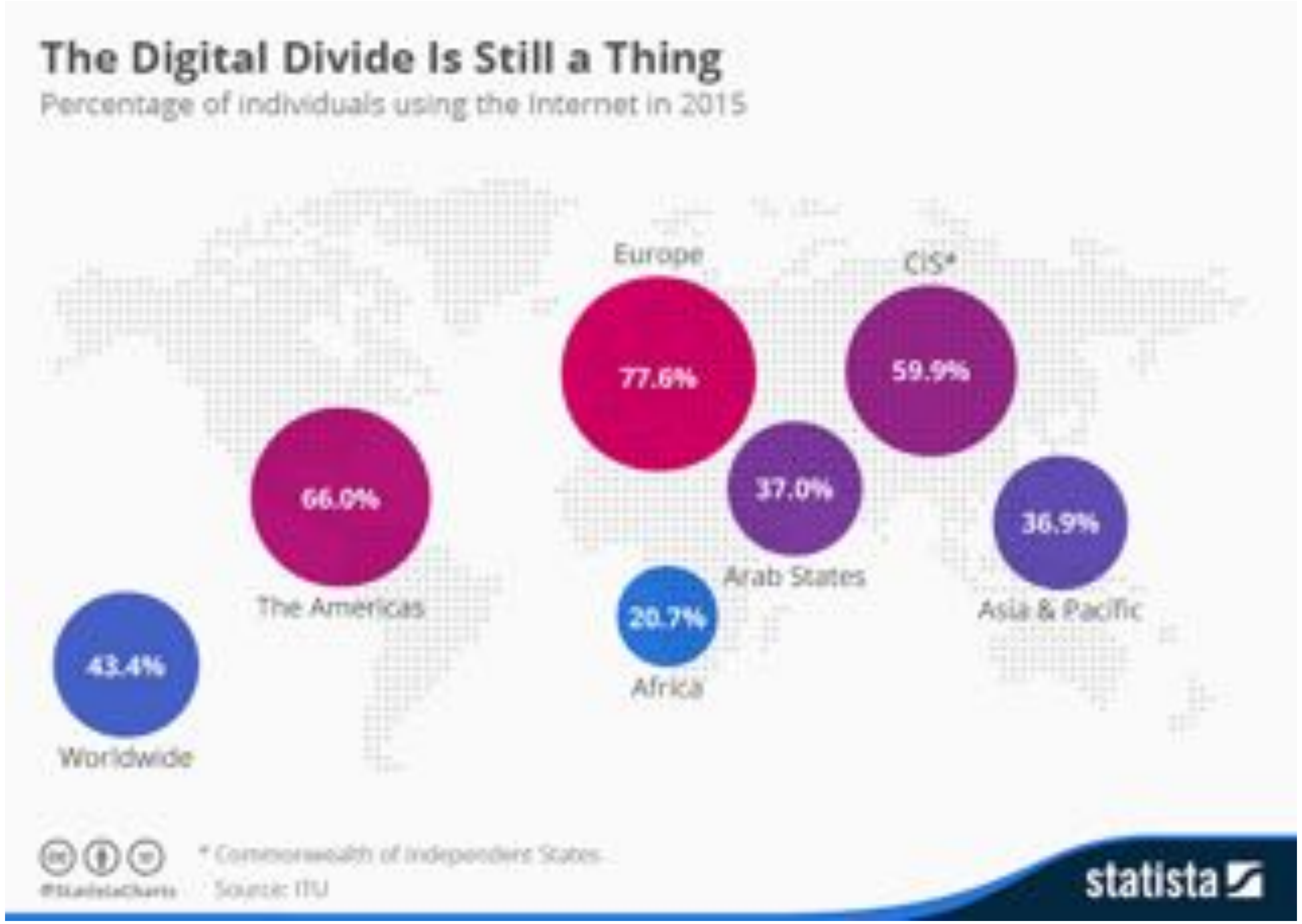
Sekara, Lee, Luengo, Obradovich, García-Herranz and EM, 2018

@e

Facebook gender divide

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





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

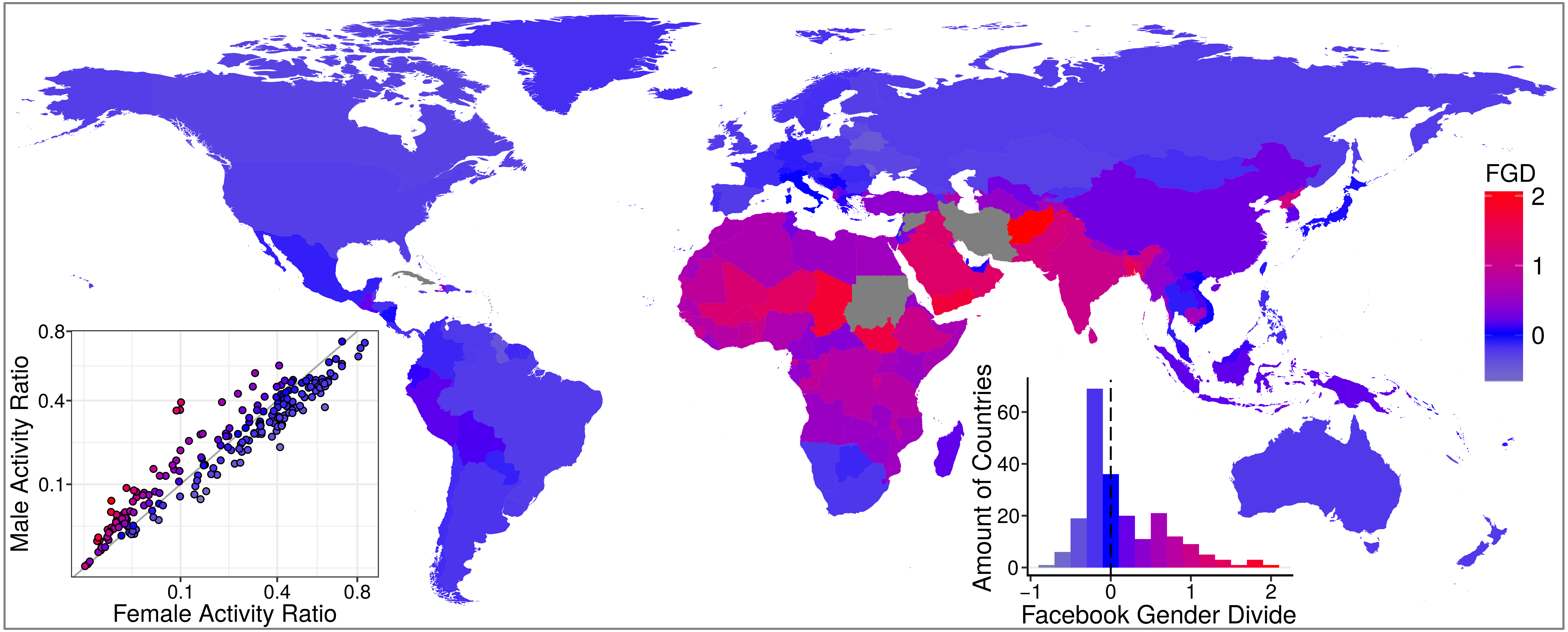
$$FGD_c = \log\left(\frac{R_{Male,c}}{R_{Female,c}}\right),$$

  - Find the main explanatory variables for that divide

Facebook gender divide

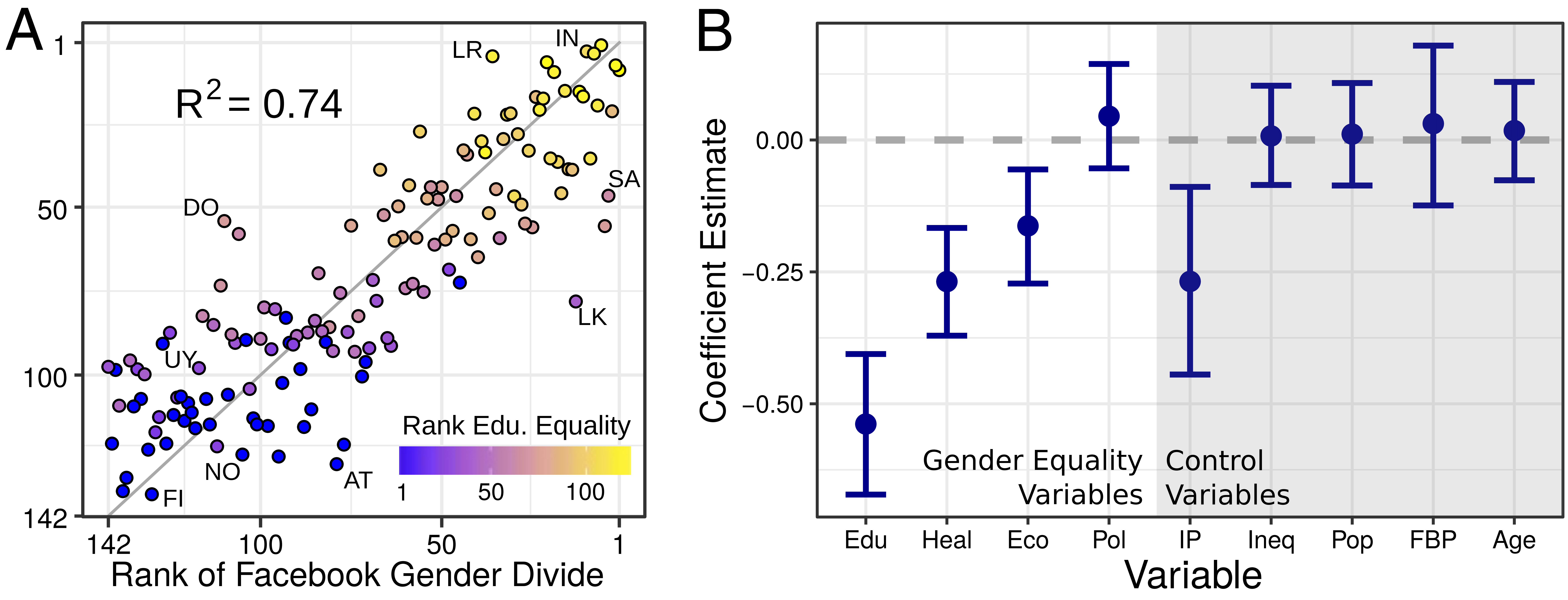
Garcia, D., Kassa, Y. M., Cuevas, A., Cebrian, M., Moro, E., Rahwan, I., & Cuevas, R. (2018). PNAS

- 217 countries, around 1.4 billion users



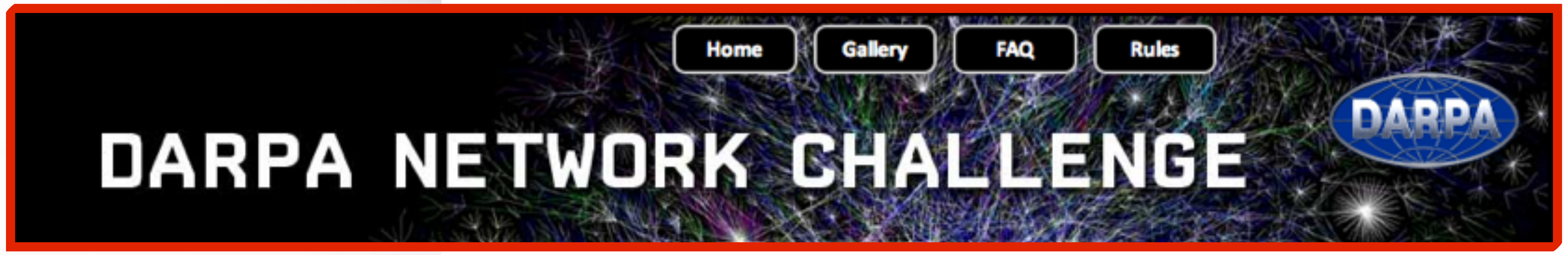
Facebook gender divide

- 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?



"Impossible by conventional intelligence"

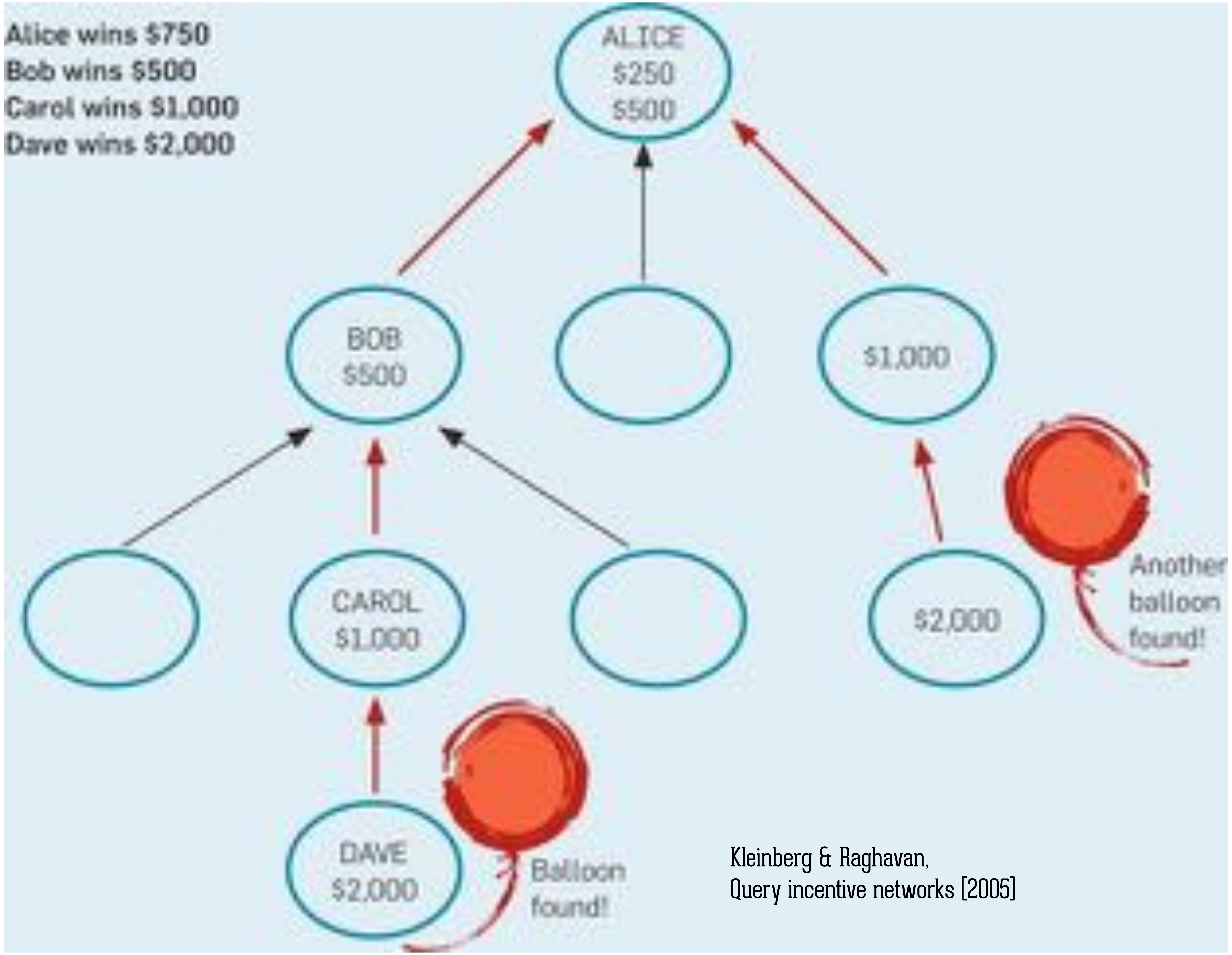
Can we use networks to mobilize people?

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]†

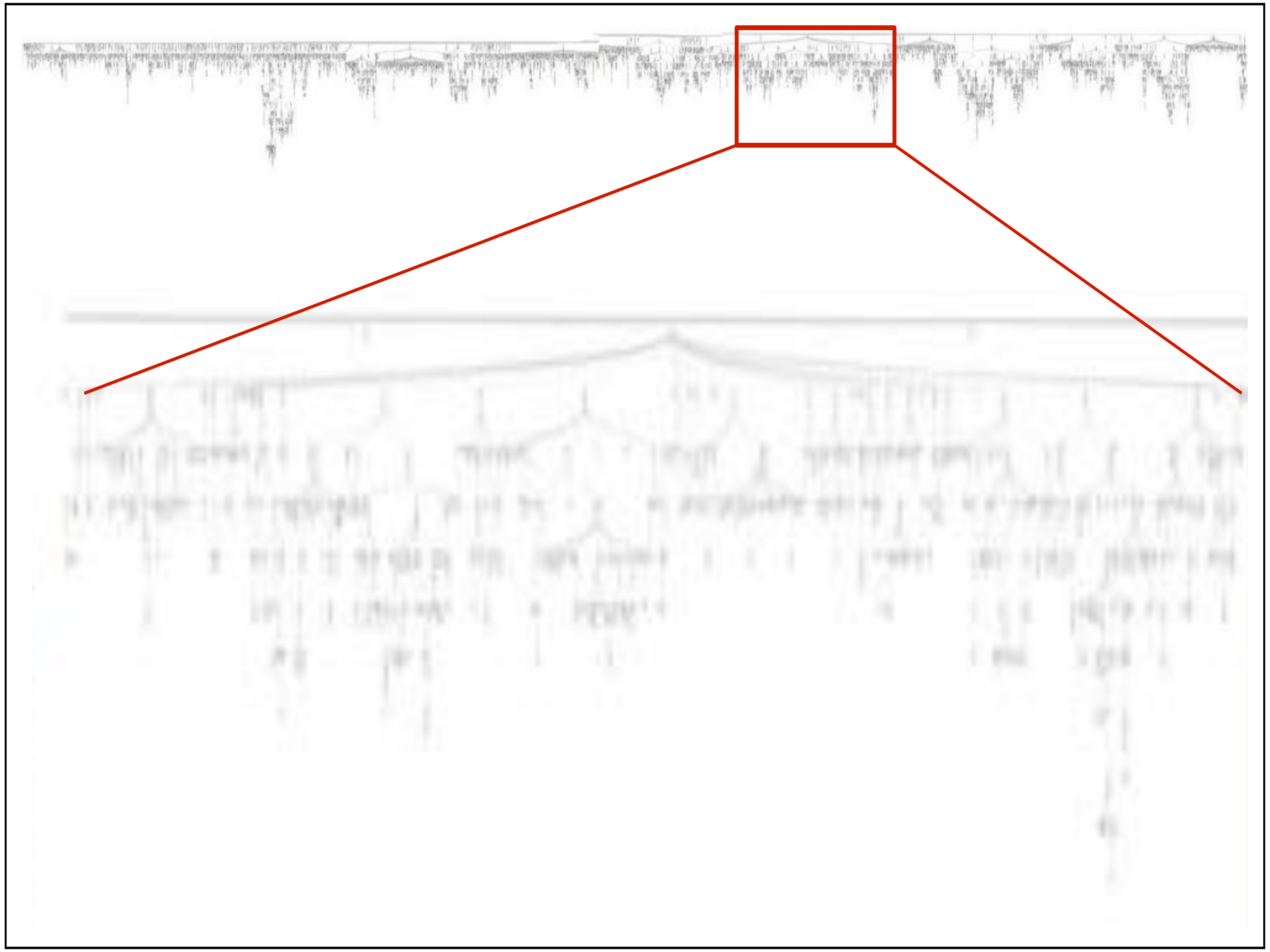
The World Wide Web is commonly seen as a platform that can harness the collective abilities of large numbers of people to accomplish tasks with unprecedented speed, accuracy, and scale. To explore the Web's ability for social mobilization, the Defense Advanced Research Projects Agency (DARPA) held the 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?



Kleinberg & Raghavan,
Query incentive networks [2005]

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


PNAS

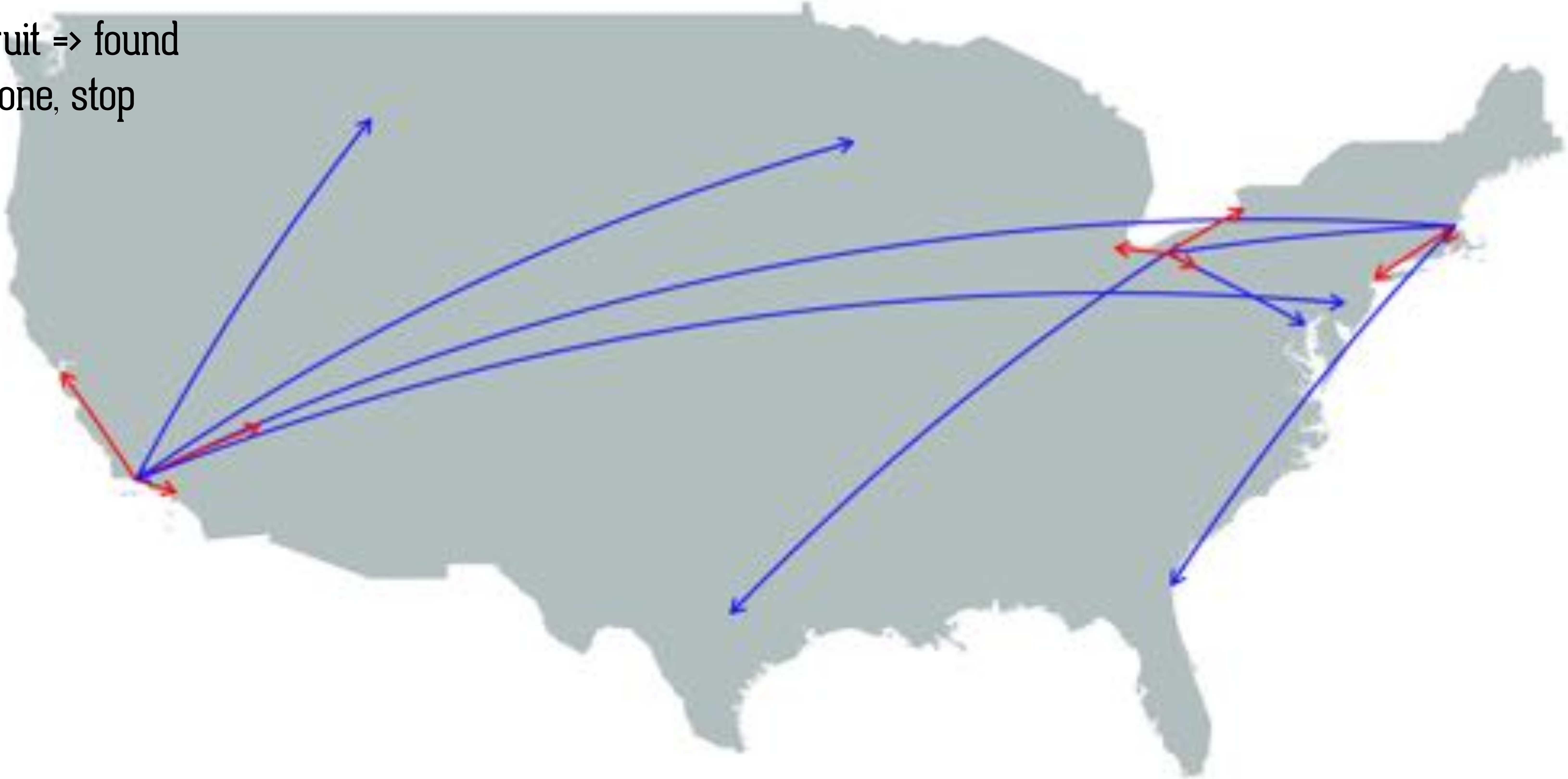
Limits of social mobilization

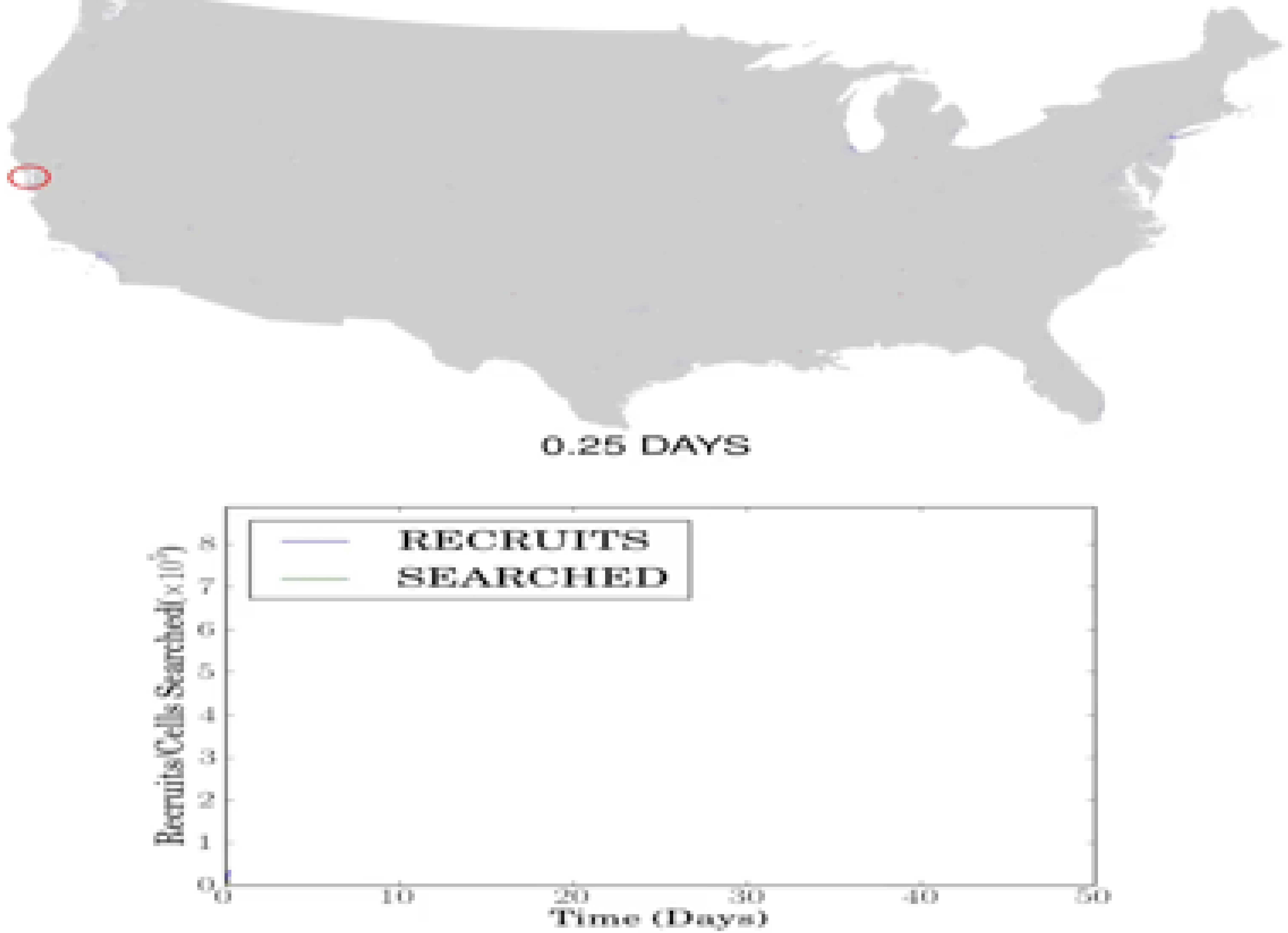
Alex Rutherford[a], Manuel Cebrian[b,c], Sohan Dsouza[a], Esteban Moro[d,e], Alex Pentland[f], and Iyad Rahwan[a,g,1]

[a]Computing and Information Science, Masdar Institute of Science and Technology, Abu Dhabi 54224, United Arab Emirates; [b]Department of Computer Science and Engineering, University of California at San Diego, La Jolla, CA 92093; [c]National Information and Communications Technology Australia, Melbourne, VIC 3010, Australia; [d]Departamento de Matemáticas and Grupo Interdisciplinar de Sistemas Complejos, Universidad Carlos III de Madrid, 28911 Madrid, Spain; [e]Instituto de Ingeniería del Conocimiento, Universidad Autónoma de Madrid, 28049 Madrid, Spain; [f]Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139; and [g]School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, United Kingdom

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 ballon is in the search area of the recruit => found
6. Proceed to 2 with the active recruits. If none, stop



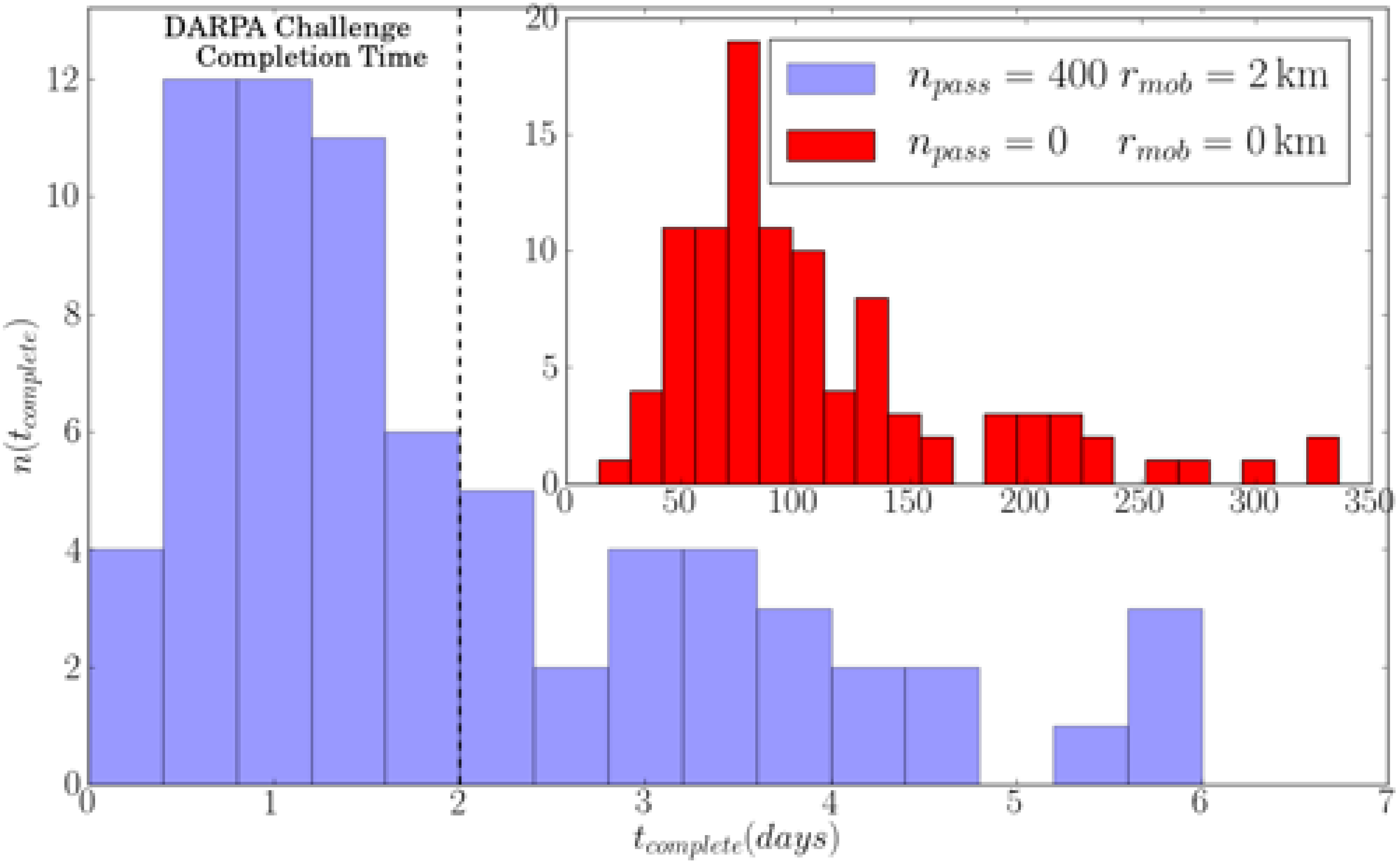


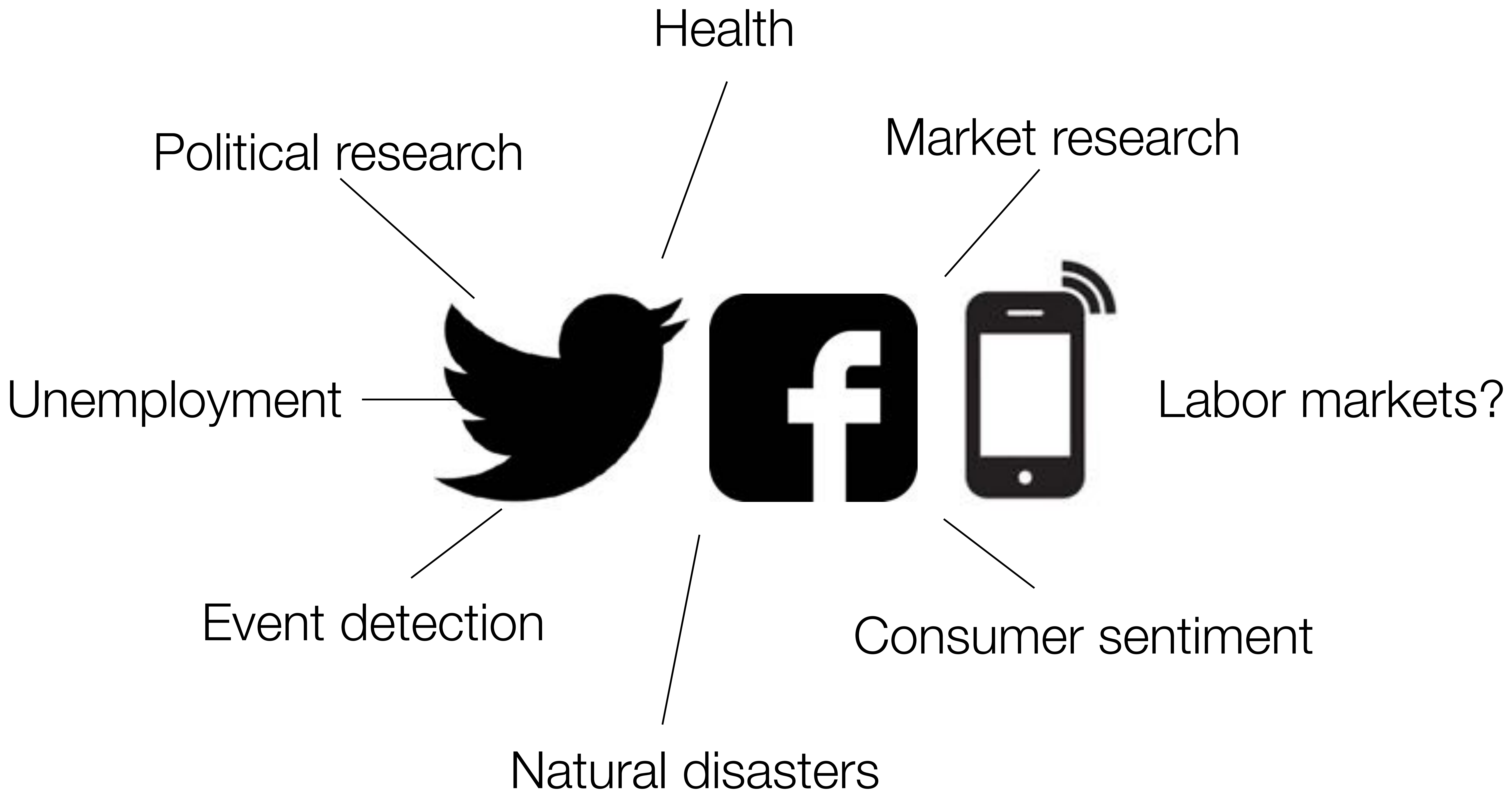
0.25 DAYS
RECRUITS
SEARCHED
Recruits/Cells Searched
0
1
2
3
4
5
6
7
8
0
10
20
30
40
50
Time (Days)

Can we use networks to mobilize people?

Time to completion
(8 balloons found)

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




Health
Political research
Market research
Unemployment
Labor markets?
Event detection
Consumer sentiment
Natural disasters

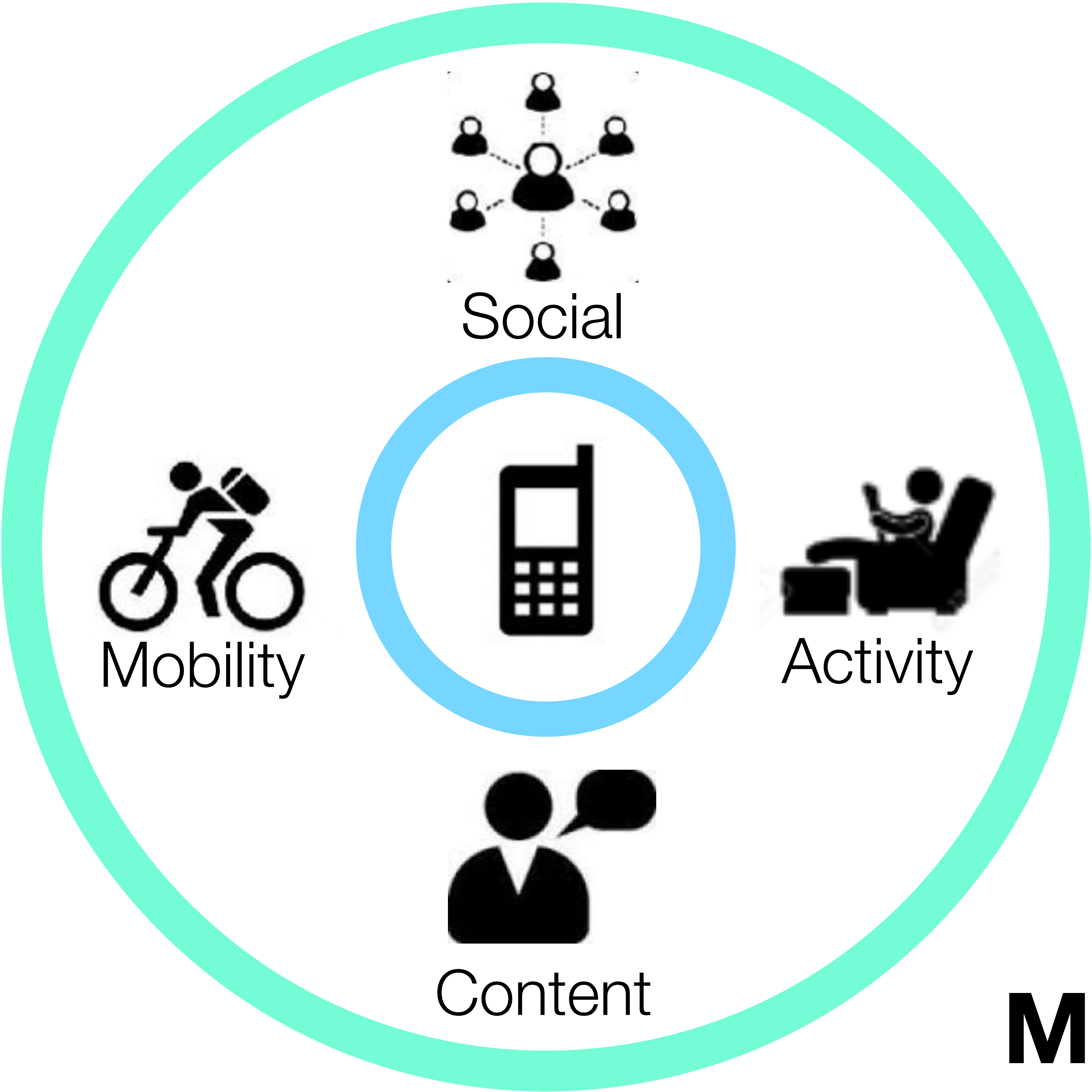
NetMob

Crowds:

Real time event detection in cities
Estimating attendance of events

Societies:

Spread of diseases
Social influence
Privacy
Product adoption
Marketing



Cities:

Energy consumption
Predicting crime hotspots
Health catchment areas
Census estimation

Economies:

Loan Repayment
Food consumption and poverty indices
Microcredit approval
Labor market

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