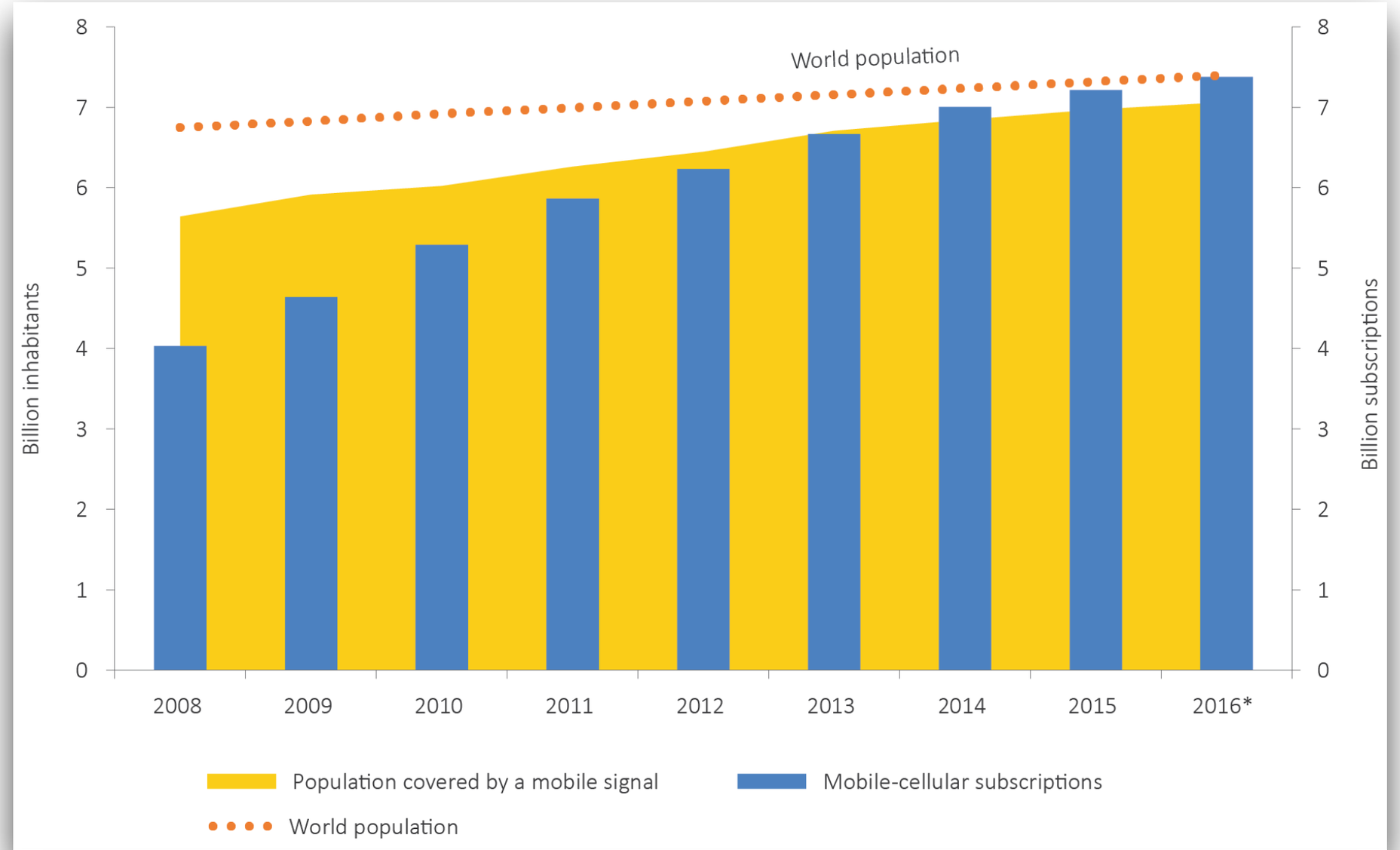
Private companies as sources of big data for research: an example from mobile phone networks

Paolo Del Giudice
ISS, Rome

mobile phone networks: a massive, global source of data on human mobility

There are now more mobile phones in use than there are people in the world to use them



Source: ITU

mobile phones: excellent potential source of information about population distributions and movements

mobile phone networks: potential and opportunities

- **Natural disasters, disease outburst, terrorist attacks:**
  knowing where people are, and when, is key for planning response.

- **Need for accurate, spatially-explicit, high resolution maps of population distributions through time**

- **Population distribution details are usually drawn from a census.**
  - Typically every several years - details quickly become inaccurate or incomplete
  - In many poorer countries census data can be especially outdated
  - Anyway, a census only provides static information

- **Mobile phones can provide invaluable data**
  - The call connection records are updated at high speed
  - The anonymous ID of the user, the time and the location of the routing tower are potentially available

- **Obvious (solvable) issues of privacy and commercial interests.**

✺ **Use of data from mobile networks could become a routine practice, in collaboration with governments, providing better understanding of population dynamics and response to crisis.**

✺ **In a wider sense, mobile phone data are driving a new "computational social science"**

first examples

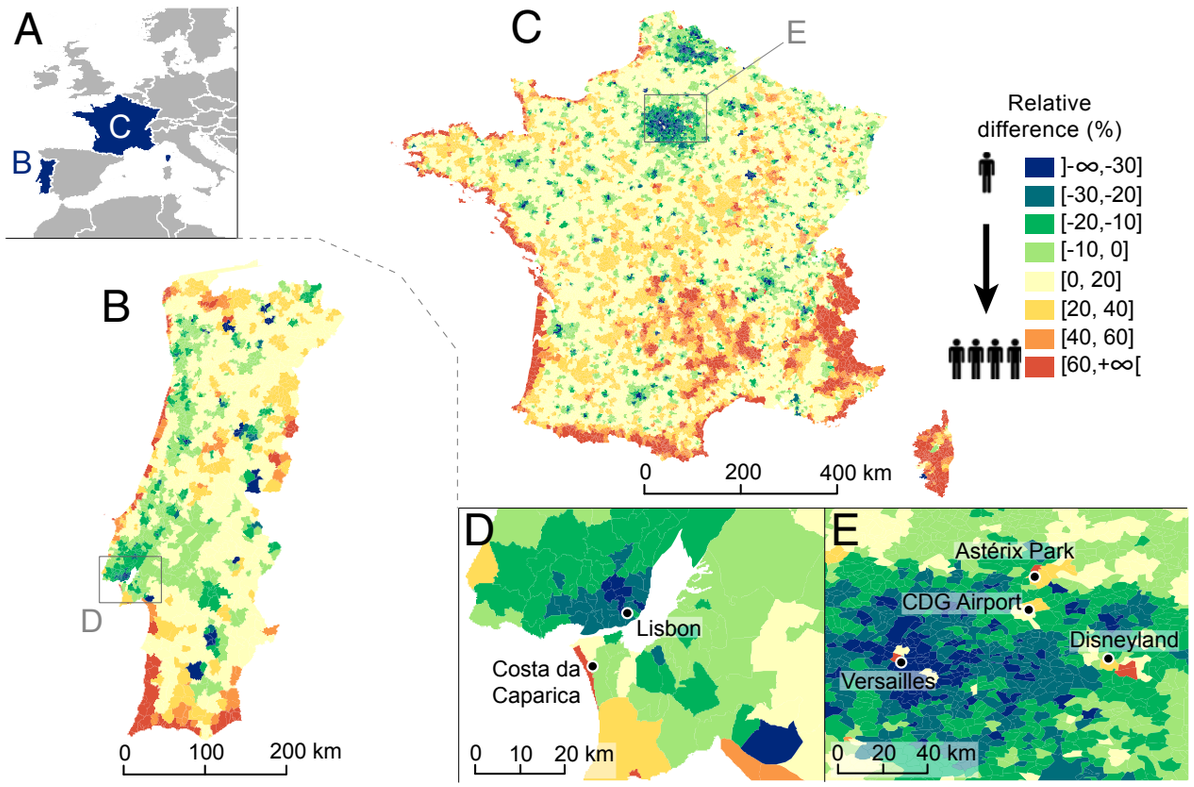
Dynamic population mapping using mobile phone data

Pierre Deville[a,b,c,1], Catherine Linard[c,d,1,2], Samuel Martin[e], Marius Gilbert[c,d], Forrest R. Stevens[f], Andrea E. Gaughan[f], Vincent D. Blondel[a], and Andrew J. Tatem[g,h,i]

PNAS | **November 11, 2014** | vol. 111 | no. 45

- One billion mobile calls database from France and Portugal
- Map population distribution at the country scale
- Reveal spatio-temporal patterns of population distribution

Seasonal changes in population distribution: relative difference of population density (July and August) - (September to June)



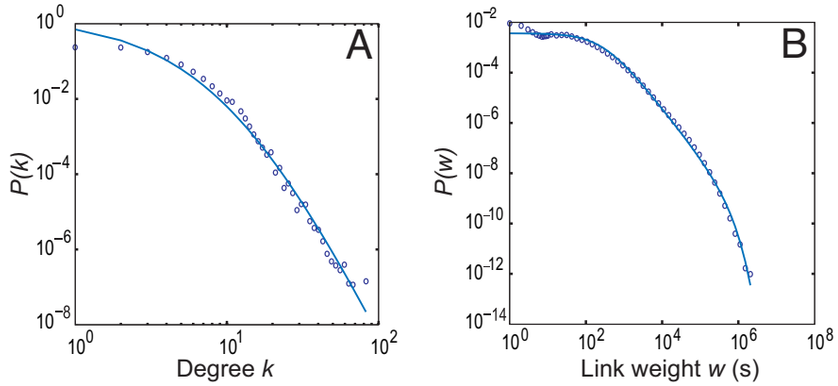
https://www.youtube.com/watch?v=qsUDH5dUnvY

Structure and tie strengths in mobile communication networks

PNAS | **May 1, 2007** | vol. 104 | no. 18

J.-P. Onnela*†‡, J. Saramäki*, J. Hyvönen*, G. Szabó§¶, D. Lazer‖, K. Kaski*, J. Kertész*,**, and A.-L. Barabási§¶

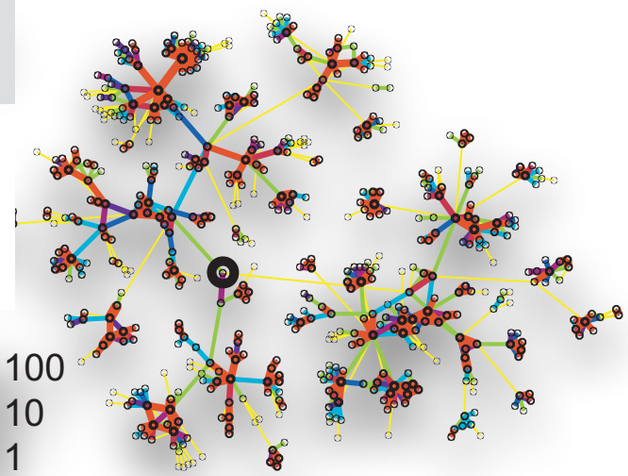
- Data accumulated over 18 weeks (only voice calls) from ~20% of the population
- Undirected links (social associations) are established only for reciprocal calls (no telemarketing)
- Links are weighted by the average call duration
- The resulting 'Mobile Call Graph' (MCG): a proxy of the communication networkk between the users
- MCG has a 'fat tail' (many talk to few, few talk to many) and few 'hubs'
- Weight distribution is broad (many user pairs talk for few minutes, some for hours)



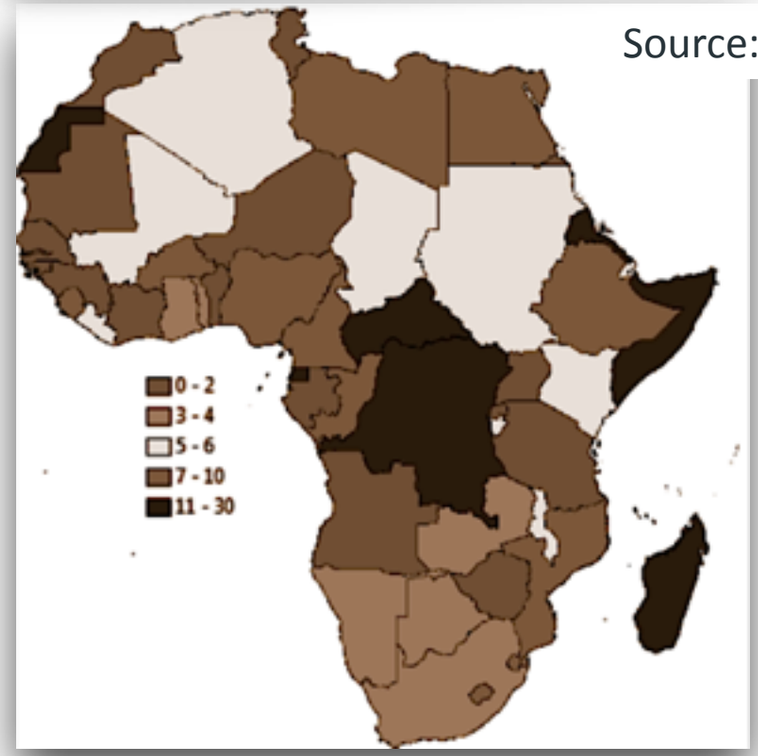
The structure of the MCG around a randomly chosen individual

Clustered structure:

strong ties within clusters, weak ties between clusters

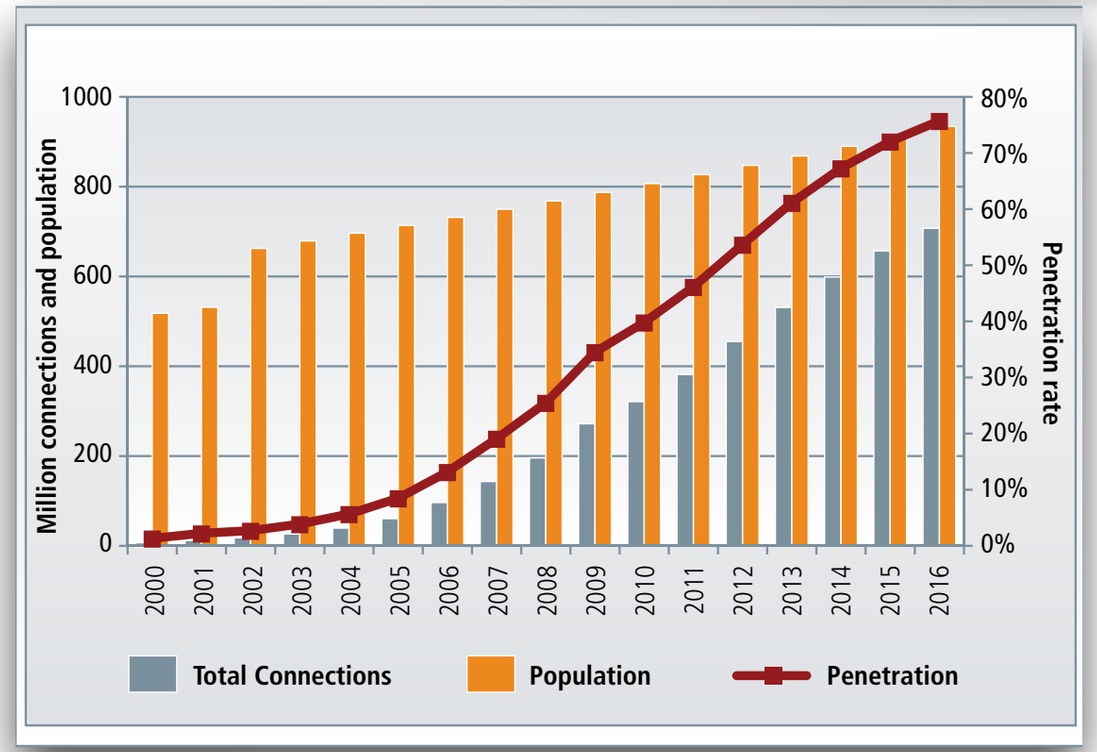


opportunities for low-income countries



Source: flowminder.org

Years since the last census for African countries.



Source: Wireless Intelligence

FLOWMINDER.ORG

- The Flowminder Foundation is a non-proft organisation based in Stockholm,.
- Mission is to improve public health and welfare in low- and middle-income countries.
- Develops methods for estimating mobility patterns and population displacement.
- Works with governments, inter-governmental organizations and NGOs

ITU (International Telecommunication Union) is the United Nations specialized agency for information and communication technologies

world pop

The WorldPop project aims to provide an open access archive of high-resolution spatial demographic datasets for Central and South America, Africa and Asia

Mobile phone data to improve response to natural disasters

OPEN ACCESS Freely available online August 2011 | Volume 8 | Issue 8 | e1001083 PLoS MEDICINE

Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti

Linus Bengtsson[1]*, Xin Lu[1,2], Anna Thorson[1], Richard Garfield[3], Johan von Schreeb[1]

- Time span: from 6 weeks before to 5 months after event
- Data from 282 million calls from 1.9 million individuals
- Estimated: number of individuals displaced; timing of major population movements; areas of the country involved

A retrospective study - what potential for real-time tracking?

- Outflow of individuals from post-earthquake cholera outbreak focus area towards safe areas
- Outcome of analysis available and usable within 12 hours

PLOS | CURRENTS DISASTERS

Rapid and Near Real-Time Assessments of Population Displacement Using Mobile Phone Data Following Disasters: The 2015 Nepal Earthquake

February 24, 2016 · *Research Article*

FLOWMINDER.

- Purpose: spatiotemporally detailed estimates of population displacements
- Data from 12 million phones
- Flows normalized to (high) baseline mobility
- Flows scaled based on penetration rates and fine population density data from **www.worldpop.org.uk**
- Estimated: *Transition matrix*: anomalous flow between all location pairs (defined at the administrative boundary level)

Limitations and issues:

- Natural disasters can affect the mobile network and power grids (hence recharging phones)
- Mobile phones towers are much denser in residential areas; difficult tracking in rural areas
- Cross-border flows are hard to track due to multiple network providers
- Age and gender bias in phone ownership (e.g. elderly or - in some countries - women)
- Ensure privacy in the face of multiple apps
- Need for cooperation and coordination between different network providers

Mobile phone data to improve response to epidemic

PLOS | COMPUTATIONAL BIOLOGY

PLOS Computational Biology | DOI:10.1371/journal.pcbi.1004846 April 4, 2016

world pop

FLOWMINDER.

RESEARCH ARTICLE

Identifying Malaria Transmission Foci for Elimination Using Human Mobility Data

Nick W. Ruktanonchai[1,2]*, Patrick DeLeenheer[3], Andrew J. Tatem[1,2], Victor A. Alegana[1,2], T. Trevor Caughlin[4], Elisabeth zu Erbach-Schoenberg[1,2], Christopher Lourenço[1,2,5], Corrine W. Ruktanonchai[1,2], David L. Smith[6,7,8,9]

- Purpose: - estimate human-mediated parasite mobility
  - identify areas acting as sources of parasite
  - better target efforts towards elimination
  - formulate a mathematical model of areas with self-sustaining malaria transmission
- mobile phone traffic data from Namibia
- 9 billion phone communications from 1.19 million SIM cards
- Time span: Oct 2010 to Sep 2011
- Data associated with the 402 settlements across Namibia
- *Connectivity matrix*: mean time resident in patch A spent in patch B (stationary analysis)
- The connectivity matrix enters a multi-patch version of the Ross-Macdonald mathematical model
- Important because importing malaria can derive from visits both to and from areas with malaria

PLOS | CURRENTS OUTBREAKS

Commentary: Containing the Ebola Outbreak – the Potential and Challenge of Mobile Network Data

September 29, 2014 · *Discussion*

FLOWMINDER.

- Purpose: use information on population movements to monitor the progression of the Ebola outbreak and to predict its future spread
- Rapid spread of the virus driven by local and regional travel
- Epidemiological models of Ebola spread rely on estimates of population flows
- Mobile phone data from Ivory Cost, Senegal and Kenya to build models of mobility patterns

BUT not from affected countries:

The Economist

Ebola and big data

Waiting on hold

Mobile-phone records would help combat the Ebola epidemic. But getting to look at them has proved hard

Oct 25th 2014

other case studies

Global Environmental Change 38 (2016) 1–7

Contents lists available at ScienceDirect

Global Environmental Change

journal homepage: www.elsevier.com/locate/gloenvcha

Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh

Xin Lu[a,b,c], David J. Wrathall[d], Pål Roe Sundsøy[e], Md. Nadiruzzaman[f,g], Erik Wetter[a,h], Asif Iqbal[e], Taimur Qureshi[e], Andrew Tatem[a,b,i], Geoffrey Canright[e], Kenth Engø-Monsen[e], Linus Bengtsson[a,b,*]

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
PSYCHOLOGY AND BEHAVIOUR
NETWORK TOPOLOGY
STATISTICS

Received
28 March 2014

Network Structure and Community Evolution on Twitter: Human Behavior Change in Response to the 2011 Japanese Earthquake and Tsunami

Xin Lu[1,2,3,4]* & Christa Brelsford[5,6]*

frontiers in Public Health

TECHNOLOGY REPORT
published: 07 August 2015
doi: 10.3389/fpubh.2015.00189

OPEN ACCESS Freely available online

PLoS MEDICINE

Perspective

Can Mobile Phone Data Improve Emergency Response to Natural Disasters?

Peter W. Gething[1]*, Andrew J. Tatem[2,3,4]

August 2011 | Volume 8 | Issue 8 | e1001085

Mobile network data for public health: opportunities and challenges

Nuria Oliver, Aleksandar Matic and Enrique Frias-Martinez*

Telefónica Research, Barcelona, Spain

2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing

An Agent-Based Model of Epidemic Spread using Human Mobility and Social Network Information

Enrique Frías-Martínez ‡, Graham Williamson [1#], Vanessa Frías-Martínez ‡,

SCIENTIFIC REPORTS

6 February 2014

OPEN

SUBJECT AREAS:
COMPLEX NETWORKS
NONLINEAR PHENOMENA

Quantifying Information Flow During Emergencies

Liang Gao[1,2], Chaoming Song[3], Ziyou Gao[1], Albert-László Barabási[2,4,5], James P. Bagrow[6,7] & Dashun Wang[8]