

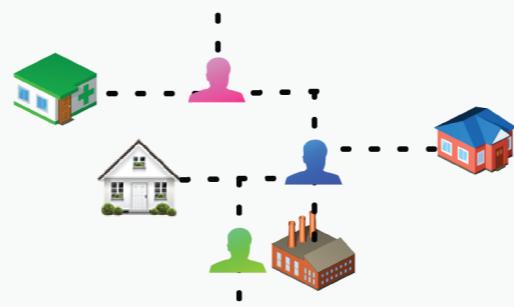
Bruno Gonçalves
www.bgoncalves.com

Geolocated Human
Behavior Through
Social Media



**DATA
SCIENCE**

Mobility



Airline Flights

NATS

Commuting



@bgoncalves

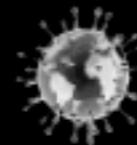
<http://youtube.com/watch?v=FqkinE8khZs>

Subscribe:
[youtube.com/tanvideo11](https://www.youtube.com/tanvideo11)

www.bgoncalves.com

Realistic Epidemic Spreading

PNAS 106, 21484 (2009)



GLEaMviz.org



Jun 23, 2009

<https://www.youtube.com/watch?v=YAf1aXCvwU>

 MoBS

GPS-enabled Smartphones



JAN
2016

GLOBAL DIGITAL SNAPSHOT

A SNAPSHOT OF THE WORLD'S KEY DIGITAL STATISTICAL INDICATORS



TOTAL
POPULATION



7.395
BILLION

URBANISATION: 54%

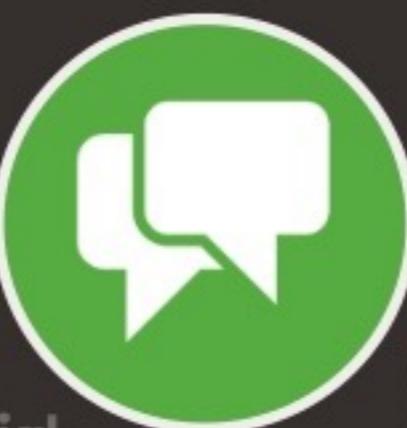
INTERNET
USERS



3.419
BILLION

PENETRATION: 46%

ACTIVE SOCIAL
MEDIA USERS



2.307
BILLION

PENETRATION: 31%

UNIQUE
MOBILE USERS



3.790
BILLION

PENETRATION: 51%

ACTIVE MOBILE
SOCIAL USERS



1.968
BILLION

PENETRATION: 27%

FIGURE REPRESENTS TOTAL GLOBAL POPULATION, INCLUDING CHILDREN

FIGURE INCLUDES ACCESS VIA FIXED AND MOBILE CONNECTIONS

FIGURE BASED ON ACTIVE USER ACCOUNTS, NOT UNIQUE INDIVIDUALS

FIGURE REPRESENTS UNIQUE MOBILE PHONE USERS

FIGURE BASED ON ACTIVE USER ACCOUNTS, NOT UNIQUE INDIVIDUALS

All this technology is making us antisocial



Social Media

SOCIAL MEOWDIA EXPLAINED



Geolocated Tweets



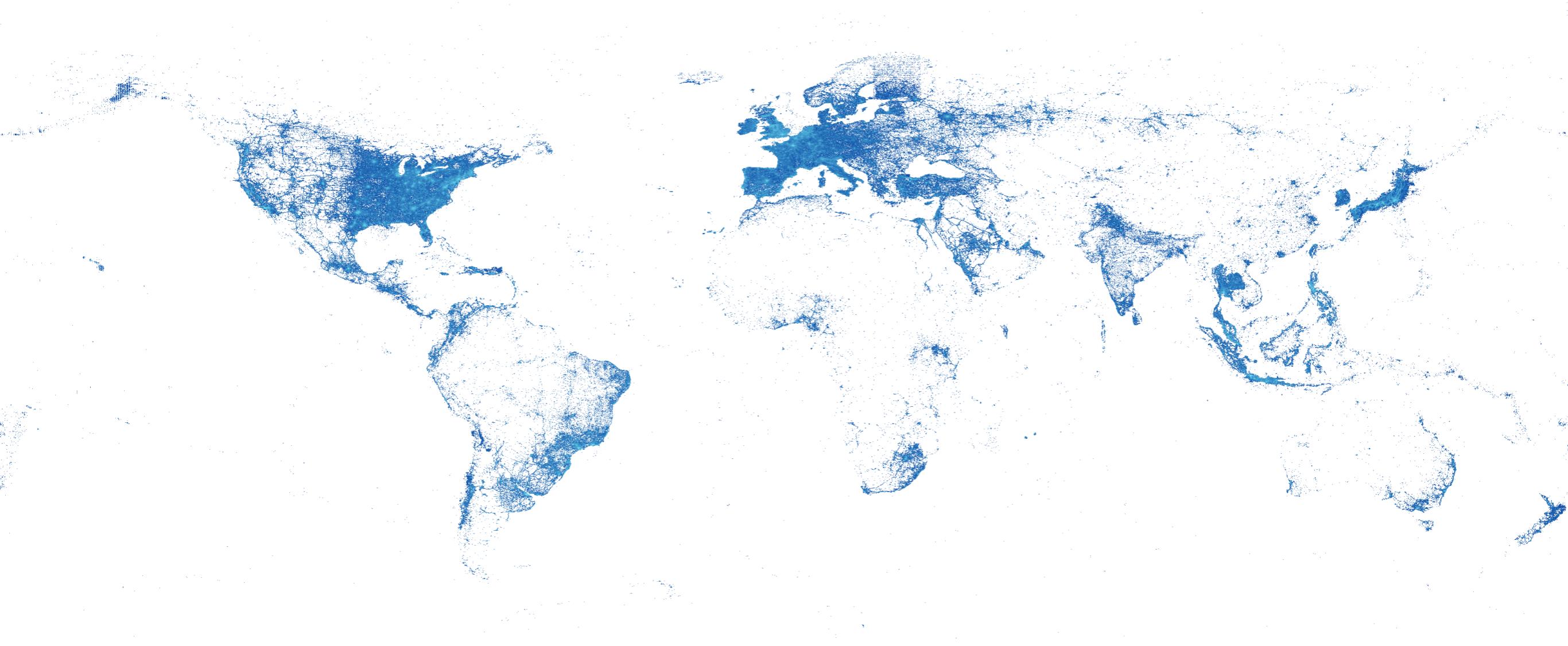
Geolocated Tweets

PLoS One 8, E61981 (2013)

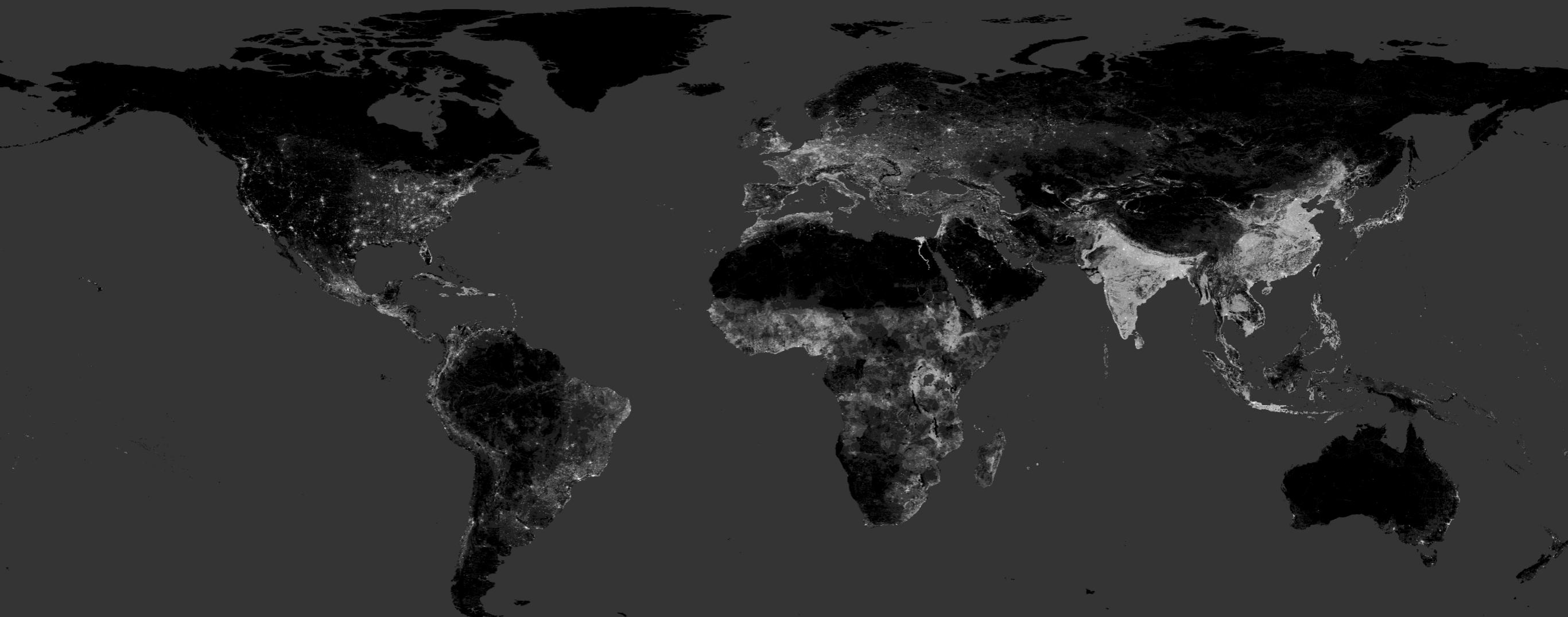


Bias is a term used in statistics and psychology to describe a tendency to favor one outcome over another, often due to preexisting beliefs or information. It can manifest in various ways, such as confirmation bias, where individuals seek out information that supports their existing beliefs and ignore information that contradicts them. Another form of bias is selection bias, where researchers choose participants or data points that are more likely to support their hypothesis. In statistics, bias refers to the difference between the expected value of an estimator and the true value of the parameter being estimated. A biased estimator is one that consistently provides estimates that are not equal to the true value. This can occur due to factors such as sampling error, measurement error, or unrepresentative samples. In psychology, bias is often studied in the context of social cognition, where individuals may have unconscious biases based on gender, race, or other demographic factors. These biases can influence how people perceive and interpret information, leading to unfair treatment or discrimination. To combat bias, it is important to be aware of one's own biases and to strive for objectivity and fairness in all interactions. This requires a commitment to evidence-based decision making and a willingness to consider multiple perspectives and viewpoints.

GPS Coordinates

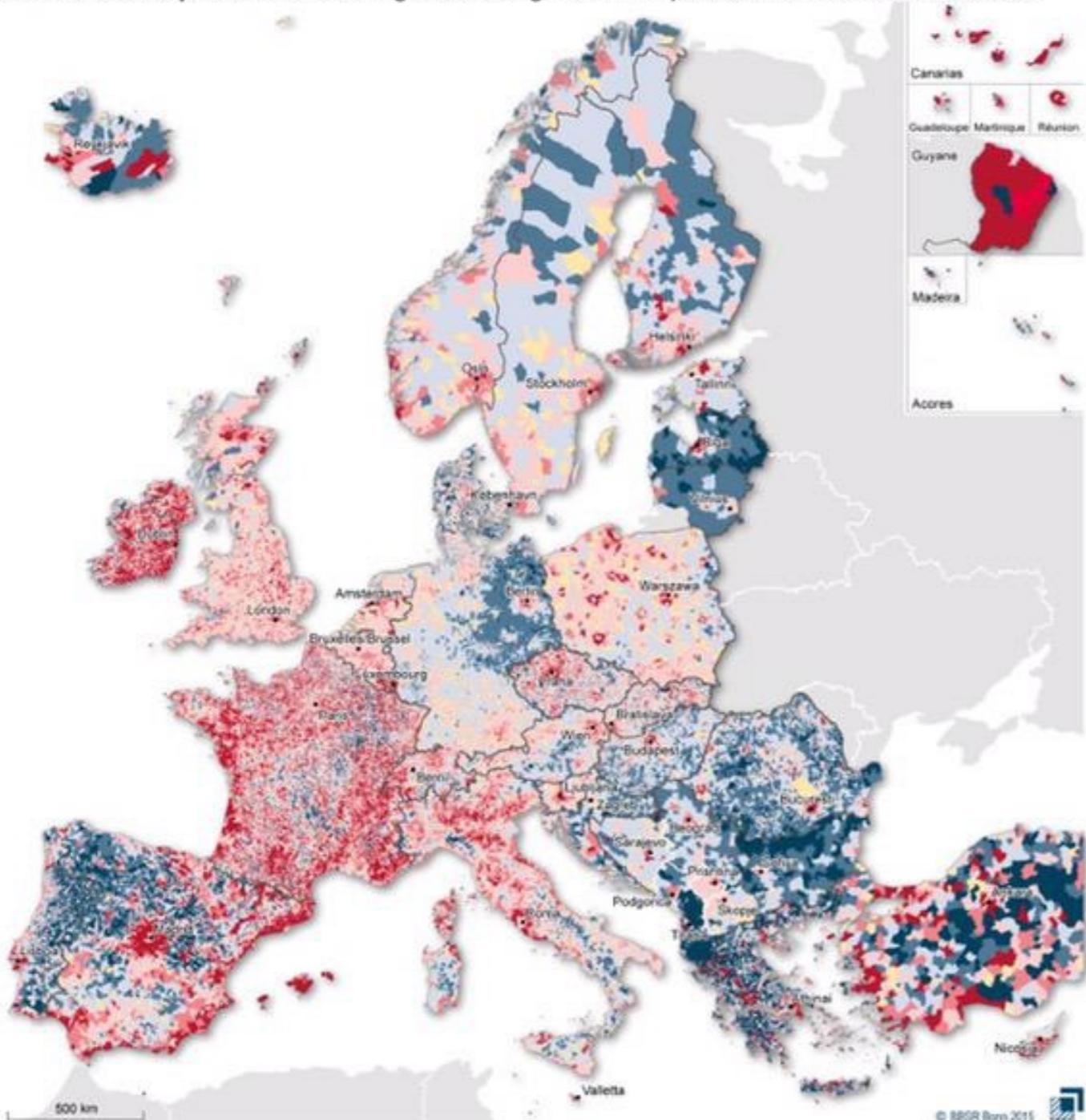


World Population



Biases

Durchschnittliche jährliche Bevölkerungsentwicklung in den Europäischen Lokalen Gebietseinheiten



Durchschnittliche jährliche Bevölkerungsentwicklung von 2001-2011* in % in den Gemeinden (LAU2)**

| | |
|---|---------------------|
| ■ | bis unter -2,0 |
| ■ | -2,0 bis unter -1,0 |
| ■ | -1,0 bis unter -0,1 |
| ■ | -0,1 bis unter +0,1 |
| ■ | 0,1 bis unter 1,0 |
| ■ | 1,0 bis unter 2,0 |
| ■ | 2,0 und mehr |

*Bevölkerungsdaten: Dezember 2001, 2011.
BG: 2004, 2011; FR: 1998, 2009; EL, NL, PL, SL: 2002, 2011;
BA: 2007, 2010; ME: 2003, 2011; BE: 2007, 2011
Zugemeldungen CH: 2007, 2013; AT: 2011, 2012
**Lokale Gebietseinheiten LAU2: BG, LT, ME, AM, TR: LAU1
Autonome Gebietseinheiten: LAU2: Äquivalente AL, PT, GL
LAU1 Äquivalente BA, KR, RU

Stadtstaat: Lokale Gebietseinheiten Europa
Support REUD
Geometrische Grundlage: GRI-Erfassung
Regionen LAU2
Bewohner: R. Blöme, L. Bödker, N. Körber-Bürgin,
T. Pernerstorfer, V. Schmid-Schwarz

Smartphone Ownership Highest Among Young Adults, Those With High Income/Education Levels

% of U.S. adults in each group who own a smartphone

| All adults | 64% |
|-----------------------|-----|
| Male | 66 |
| Female | 63 |
| 18-29 | 85 |
| 30-49 | 79 |
| 50-64 | 54 |
| 65+ | 27 |
| White, non-Hispanic | 61 |
| Black, non-Hispanic | 70 |
| Hispanic | 71 |
| HS grad or less | 52 |
| Some college | 69 |
| College+ | 78 |
| Less than \$30,000/yr | 50 |
| \$30,000-\$49,999 | 71 |
| \$50,000-\$74,999 | 72 |
| \$75,000 or more | 84 |
| Urban | 68 |
| Suburban | 66 |
| Rural | 52 |

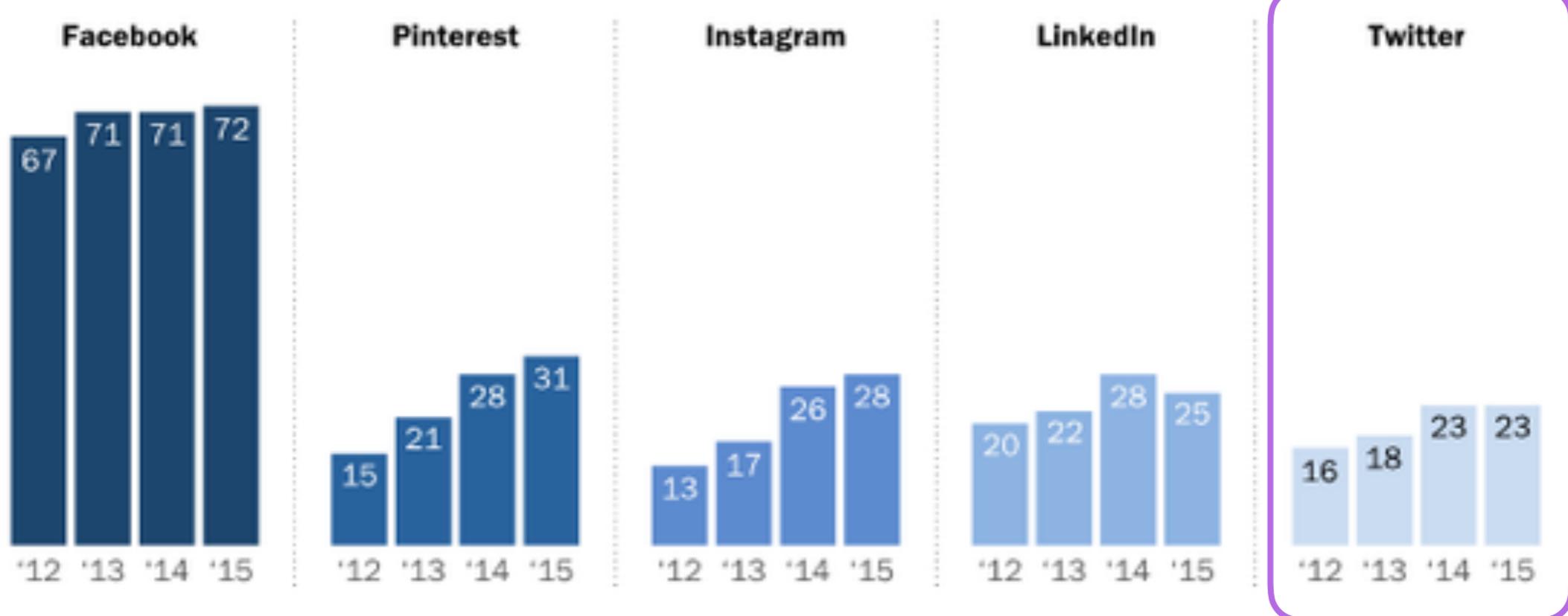
Combined analysis of Pew Research Center surveys conducted Dec. 4-7 and Dec. 18-21, 2014.

PEW RESEARCH CENTER

Biases

Pinterest and Instagram Usage Doubles Since 2012, Growth on Other Platforms is Slower

% of online adults who say they use the following social media platform, by year

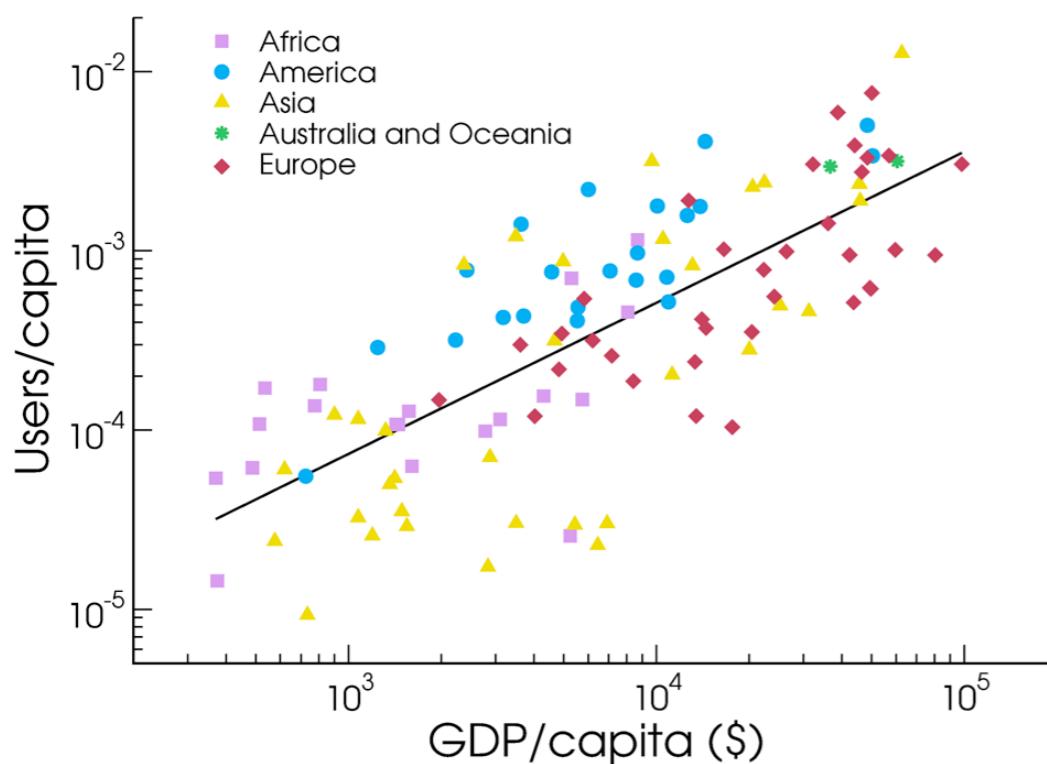
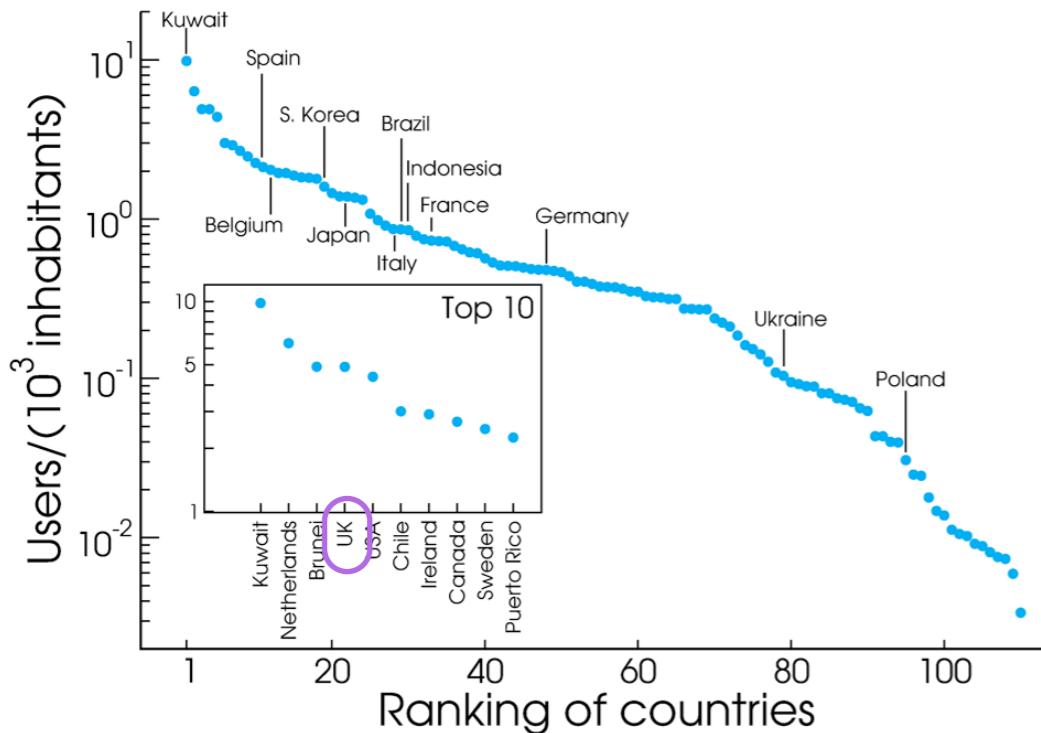


Pew Research Center Survey, March 17-April 12, 2015.

PEW RESEARCH CENTER

Market Penetration

PLoS One 8, E61981 (2013)



Age Distribution

PLoS One 10, e0115545 (2015)

Twitter users

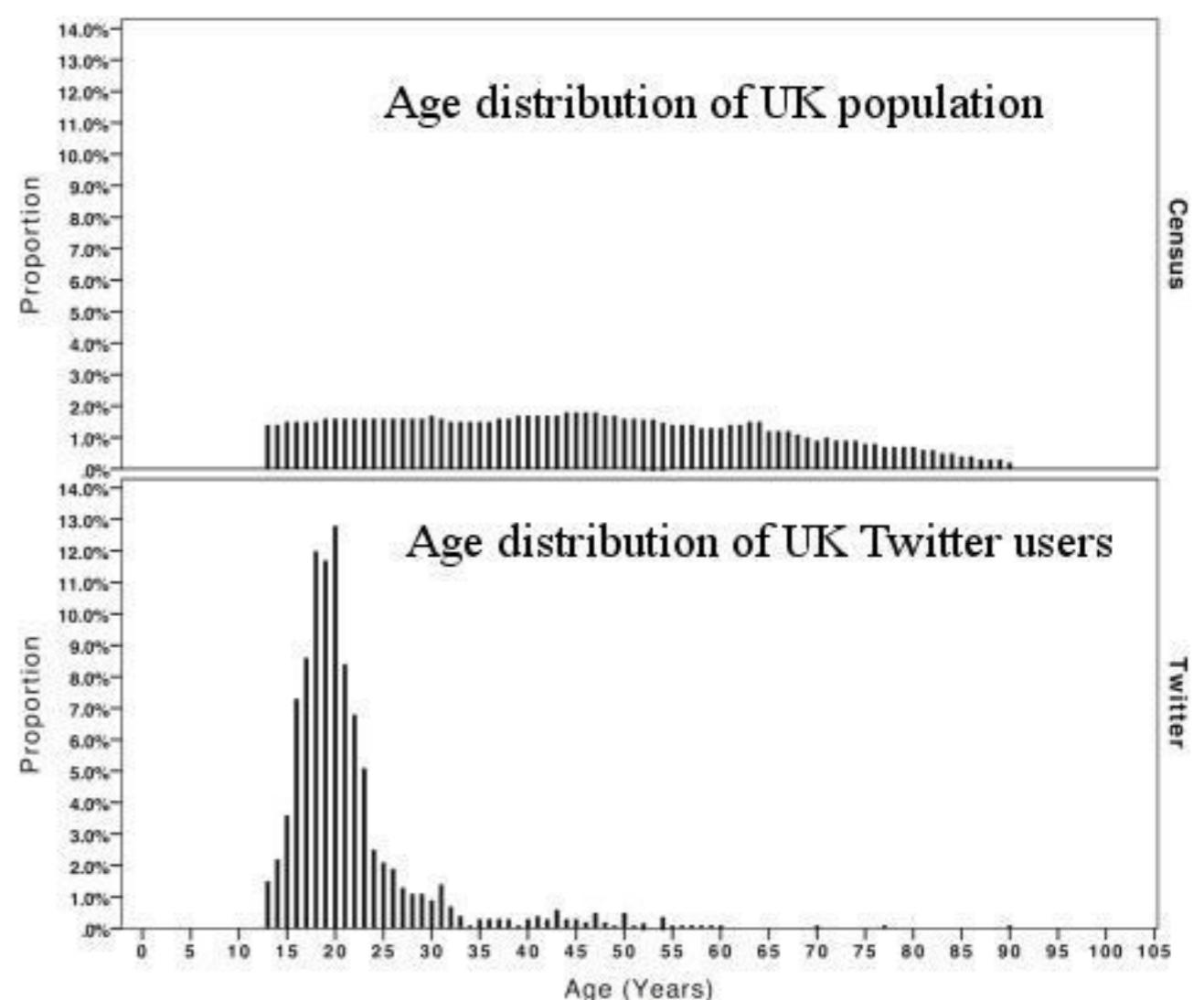
Among online adults, the % who use Twitter

| | 2013 | 2014 |
|--------------------------|------|------|
| All internet users | 18% | 23%* |
| Men | 17 | 24* |
| Women | 18 | 21 |
| White, Non-Hispanic | 16 | 21 * |
| Black, Non-Hispanic | 29 | 27 |
| Hispanic | 16 | 25 |
| 18-29 | 31 | 37 |
| 30-49 | 19 | 25 |
| 50-64 | 9 | 12 |
| 65+ | 5 | 10* |
| High school grad or less | 17 | 16 |
| Some college | 18 | 24 |
| College+ (n= 685) | 18 | 30* |
| Less than \$30,000/yr | 17 | 20 |
| \$30,000-\$49,999 | 18 | 21 |
| \$50,000-\$74,999 | 15 | 27* |
| \$75,000+ | 19 | 27* |
| Urban | 18 | 25* |
| Suburban | 19 | 23 |
| Rural | 11 | 17 |

Source: Pew Research Center's Internet Project September Combined Omnibus Survey, September 11-14 & September 18-21, 2014. N=1,597 internet users ages 18+. The margin of error for all internet users is +/- 2.9 percentage points. 2013 data from Pew Internet August Tracking Survey, August 07 - September 16, 2013, n= 1,445 internet users ages 18+.

Note: Percentages marked with an asterisk (*) represent a significant change from 2013.
Results are significant at the 95% confidence level using an independent z-test.

PEW RESEARCH CENTER



Age Distribution

Twitter users

Among online adults, the % who use Twitter

| | 2013 | 2014 |
|--------------------------|------|------|
| All internet users | 18% | 23%* |
| Men | 17 | 24* |
| Women | 18 | 21 |
| White, Non-Hispanic | 16 | 21 * |
| Black, Non-Hispanic | 29 | 27 |
| Hispanic | 16 | 25 |
| 18-29 | 31 | 37 |
| 30-49 | 19 | 25 |
| 50-64 | 9 | 12 |
| 65+ | 5 | 10* |
| High school grad or less | 17 | 16 |
| Some college | 18 | 24 |
| College+ (n= 685) | 18 | 30* |
| Less than \$30,000/yr | 17 | 20 |
| \$30,000-\$49,999 | 18 | 21 |
| \$50,000-\$74,999 | 15 | 27* |
| \$75,000+ | 19 | 27* |
| Urban | 18 | 25* |
| Suburban | 19 | 23 |
| Rural | 11 | 17 |

Source: Pew Research Center's Internet Project September Combined Omnibus Survey, September 11-14 & September 18-21, 2014. N=1,597 internet users ages 18+. The margin of error for all internet users is +/- 2.9 percentage points. 2013 data from Pew Internet August Tracking Survey, August 07 - September 16, 2013, n= 1,445 internet users ages 18+.

Note: Percentages marked with an asterisk (*) represent a significant change from 2013. Results are significant at the 95% confidence level using an independent z-test.

PEW RESEARCH CENTER

Facebook Demographics

Among internet users, the % who use Facebook

| | Internet users |
|----------------------------|----------------|
| Total | 72% |
| Men | 66 |
| Women | 77 |
| White, Non-Hispanic | 70 |
| Black, Non-Hispanic (n=85) | 67 |
| Hispanic | 75 |
| 18-29 | 82 |
| 30-49 | 79 |
| 50-64 | 64 |
| 65+ | 48 |
| High school grad or less | 71 |
| Some college | 72 |
| College+ | 72 |
| Less than \$30,000/yr | 73 |
| \$30,000-\$49,999 | 72 |
| \$50,000-\$74,999 | 66 |
| \$75,000+ | 78 |
| Urban | 74 |
| Suburban | 72 |
| Rural | 67 |

Source: Pew Research Center, March 17-April 12, 2015.

PEW RESEARCH CENTER

Instagram Demographics

Among internet users, the % who use Instagram

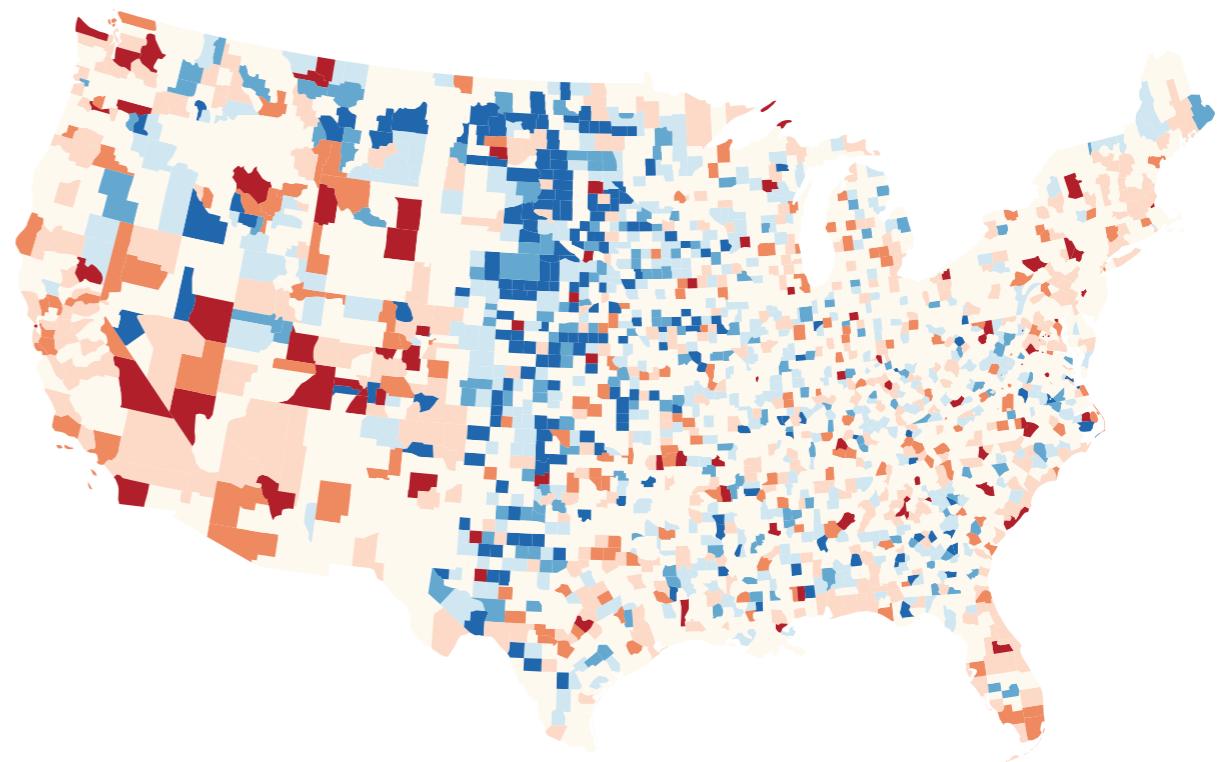
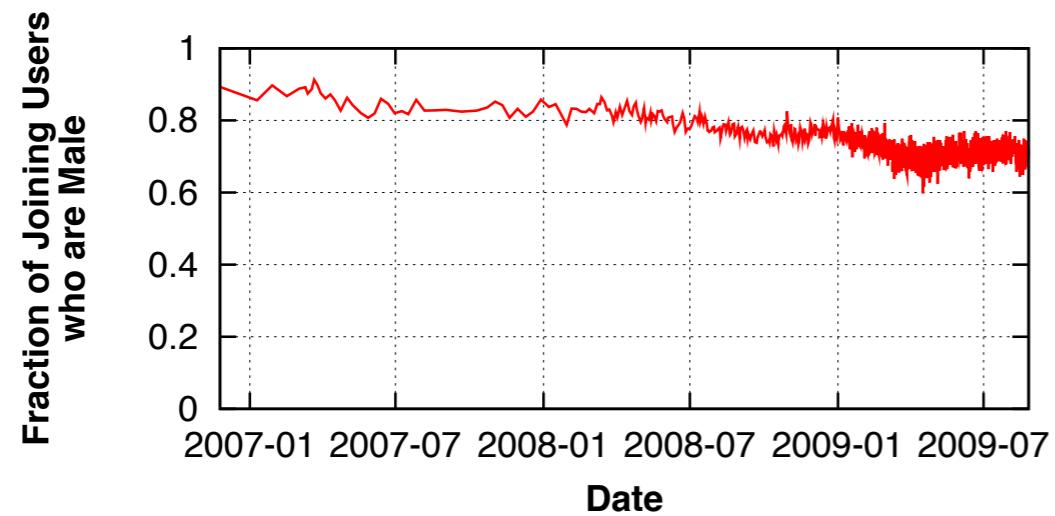
| | Internet user |
|----------------------------|---------------|
| Total | 28% |
| Men | 24 |
| Women | 31 |
| White, Non-Hispanic | 21 |
| Black, Non-Hispanic (n=85) | 47 |
| Hispanic | 38 |
| 18-29 | 55 |
| 30-49 | 28 |
| 50-64 | 11 |
| 65+ | 4 |
| High school grad or less | 25 |
| Some college | 32 |
| College+ | 26 |
| Less than \$30,000/yr | 26 |
| \$30,000-\$49,999 | 27 |
| \$50,000-\$74,999 | 30 |
| \$75,000+ | 26 |
| Urban | 32 |
| Suburban | 28 |
| Rural | 18 |

Source: Pew Research Center, March 17-April 12, 2015.

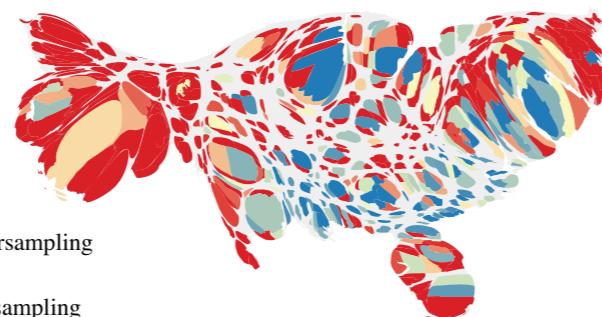
PEW RESEARCH CENTER

Demographics

ICWSM'11, 375 (2011)



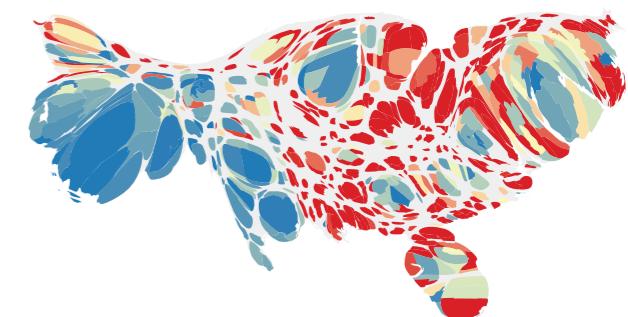
(a) Caucasian (non-hispanic)



(b) African-American



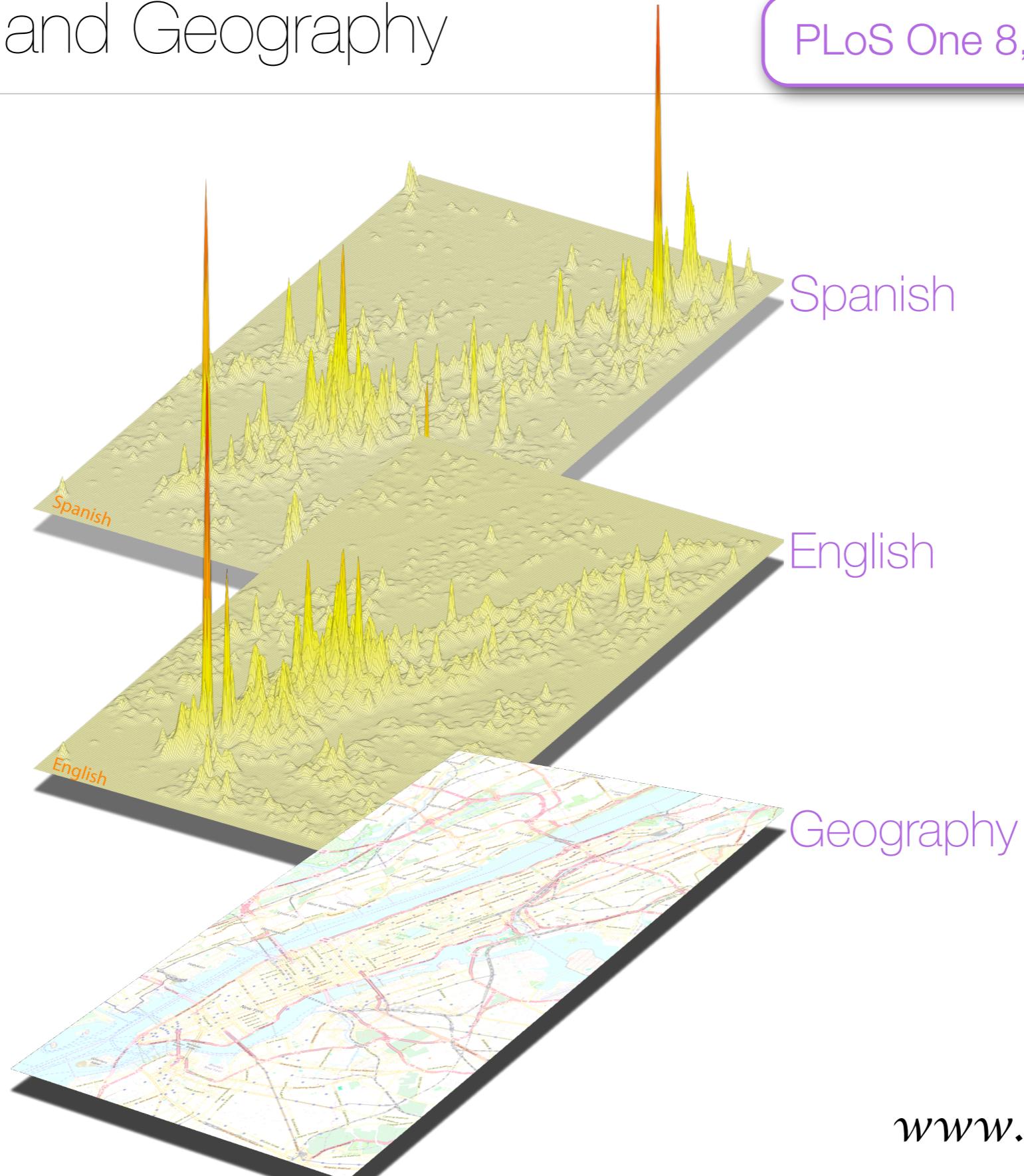
(c) Asian or Pacific Islander



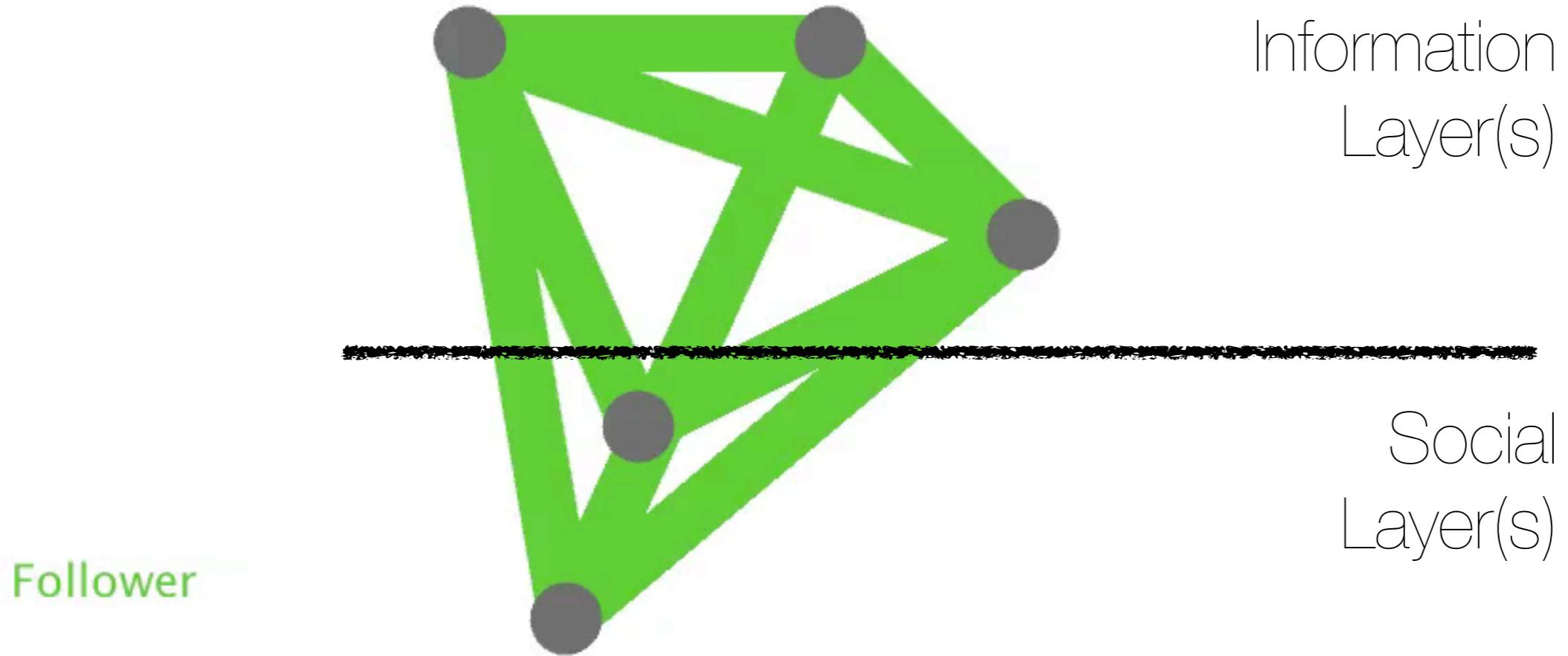
(d) Hispanic

Language and Geography

PLoS One 8, E61981 (2013)

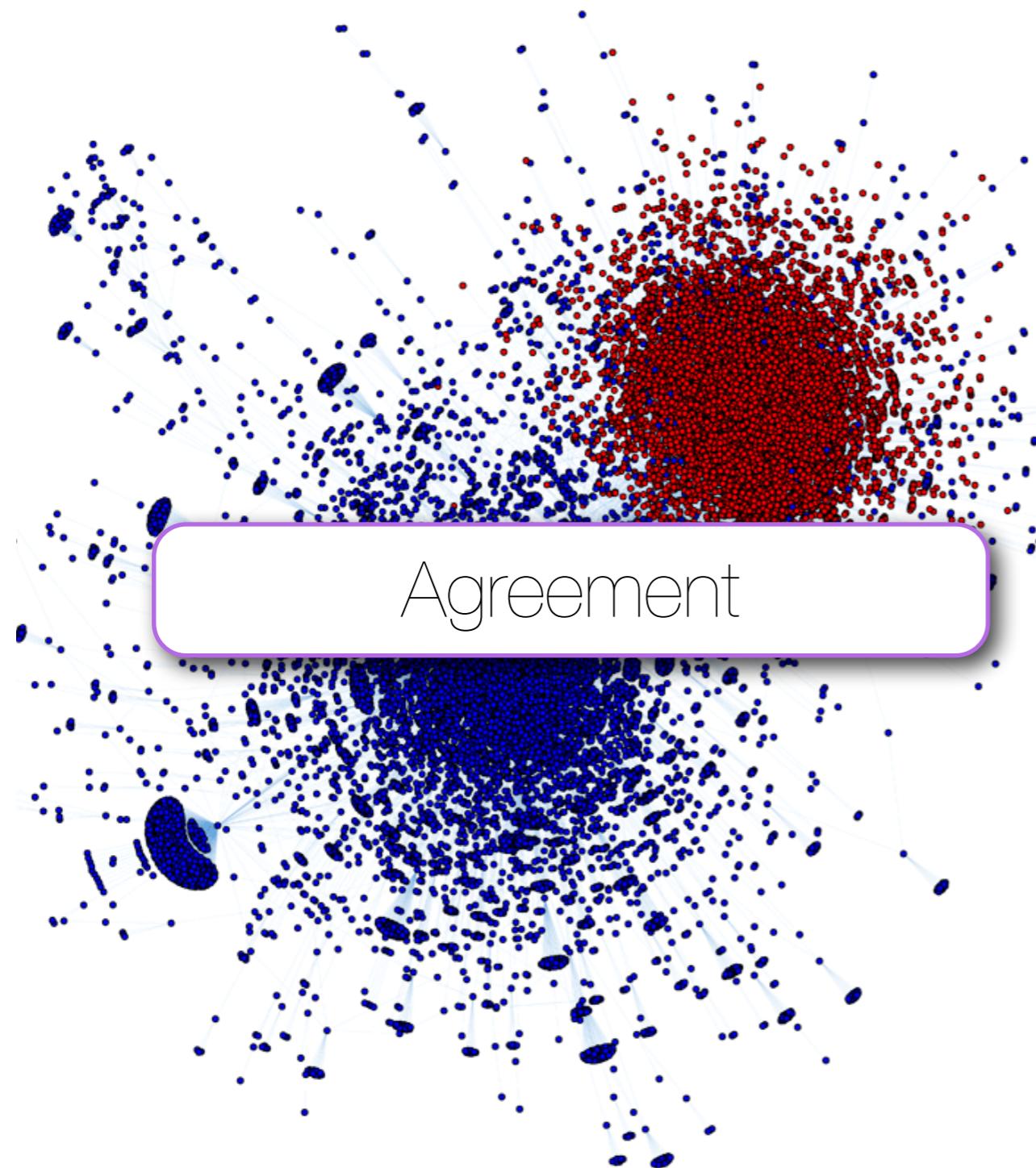


Multilayer Network

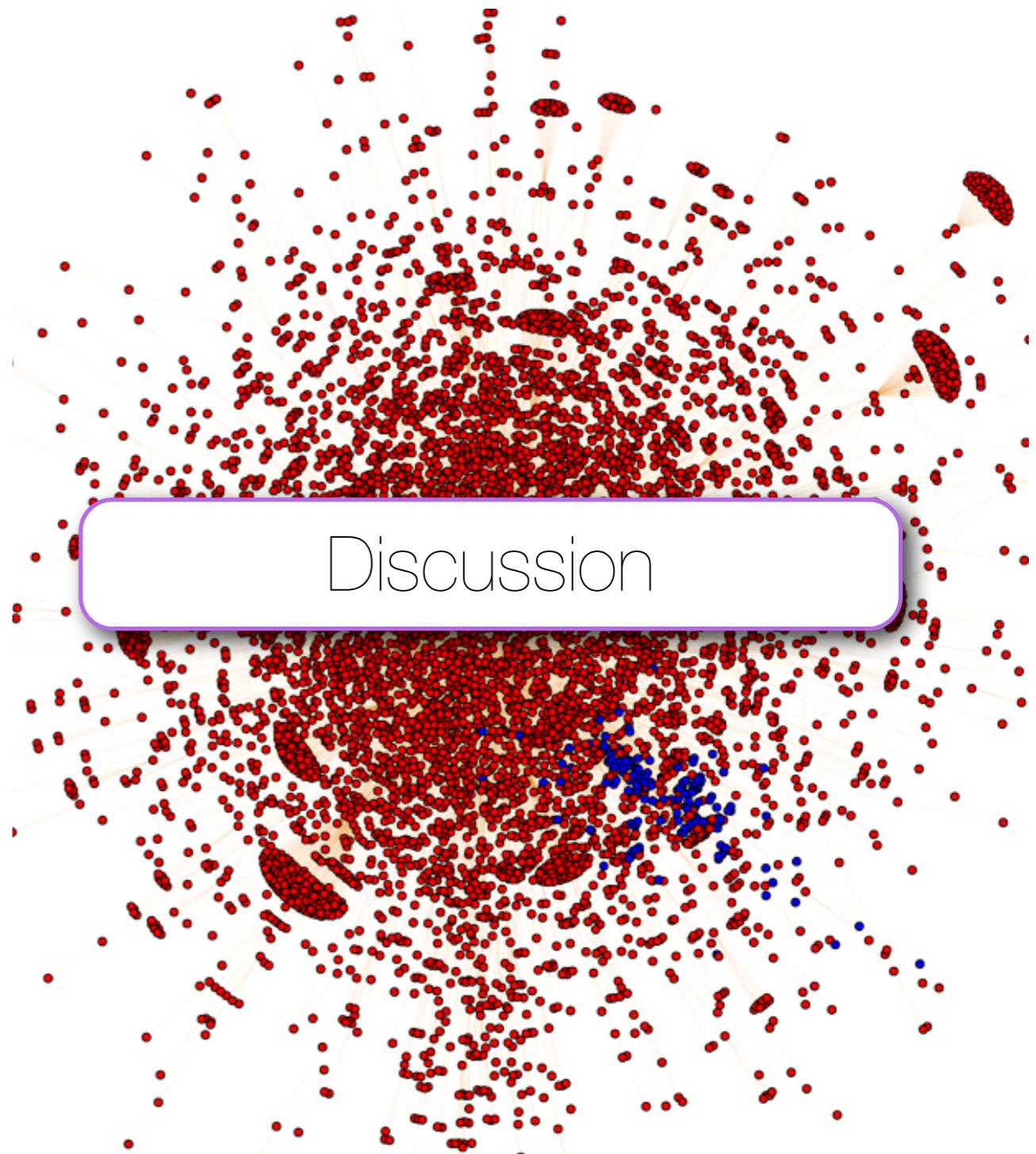


Link Function

ICWSM'11, 89 (2011)



Agreement



Discussion

The Strength of Weak Ties (1973)

The Strength of Weak Ties¹

Mark S. Granovetter
Johns Hopkins University

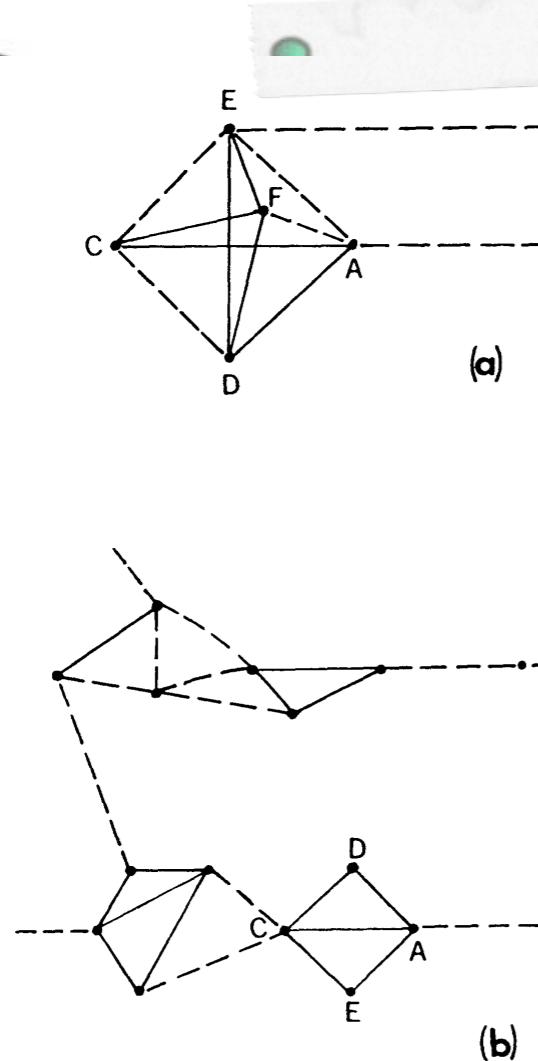
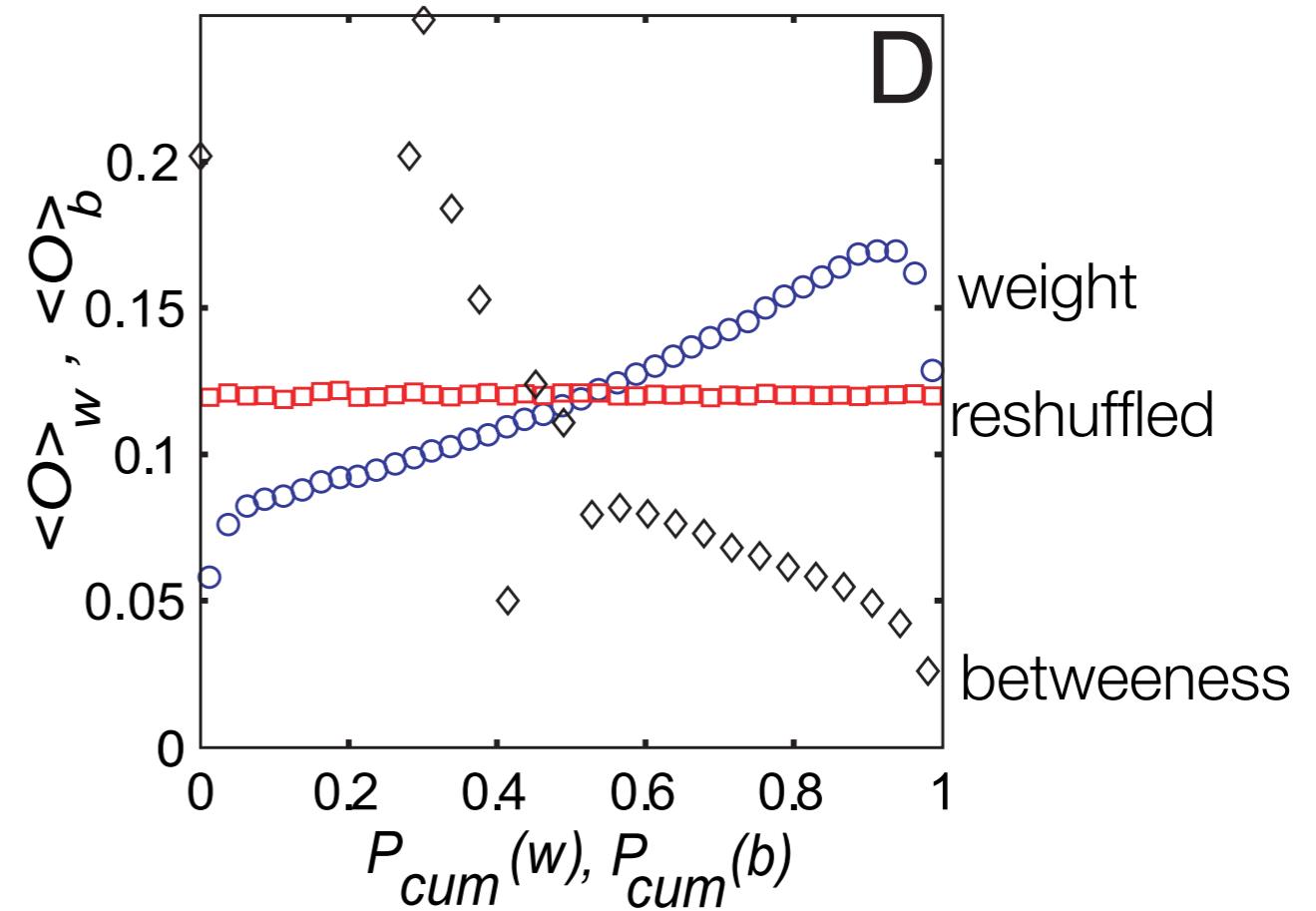
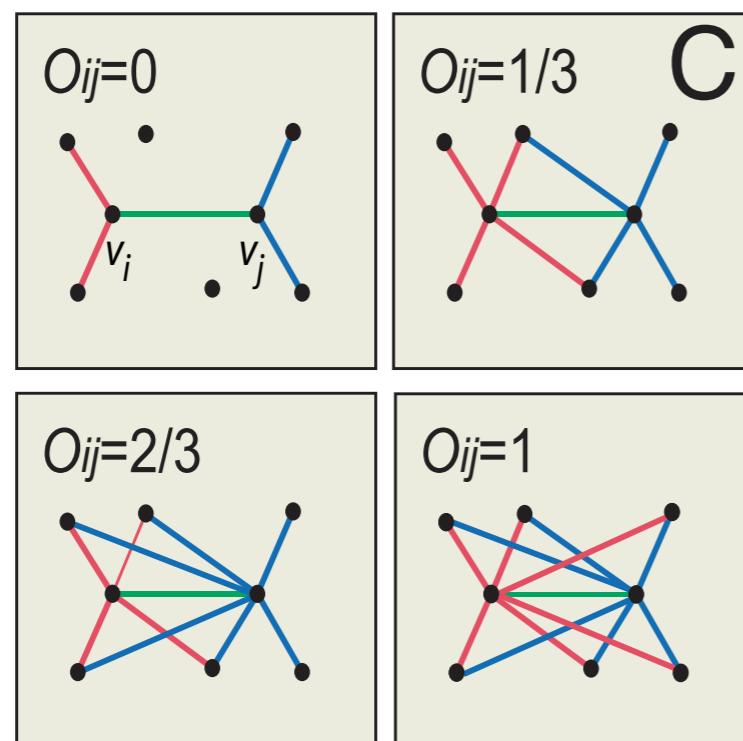


FIG. 2.—Local bridges. *a*, Degree 3; *b*, Degree 13. —— strong tie, - - - weak tie.

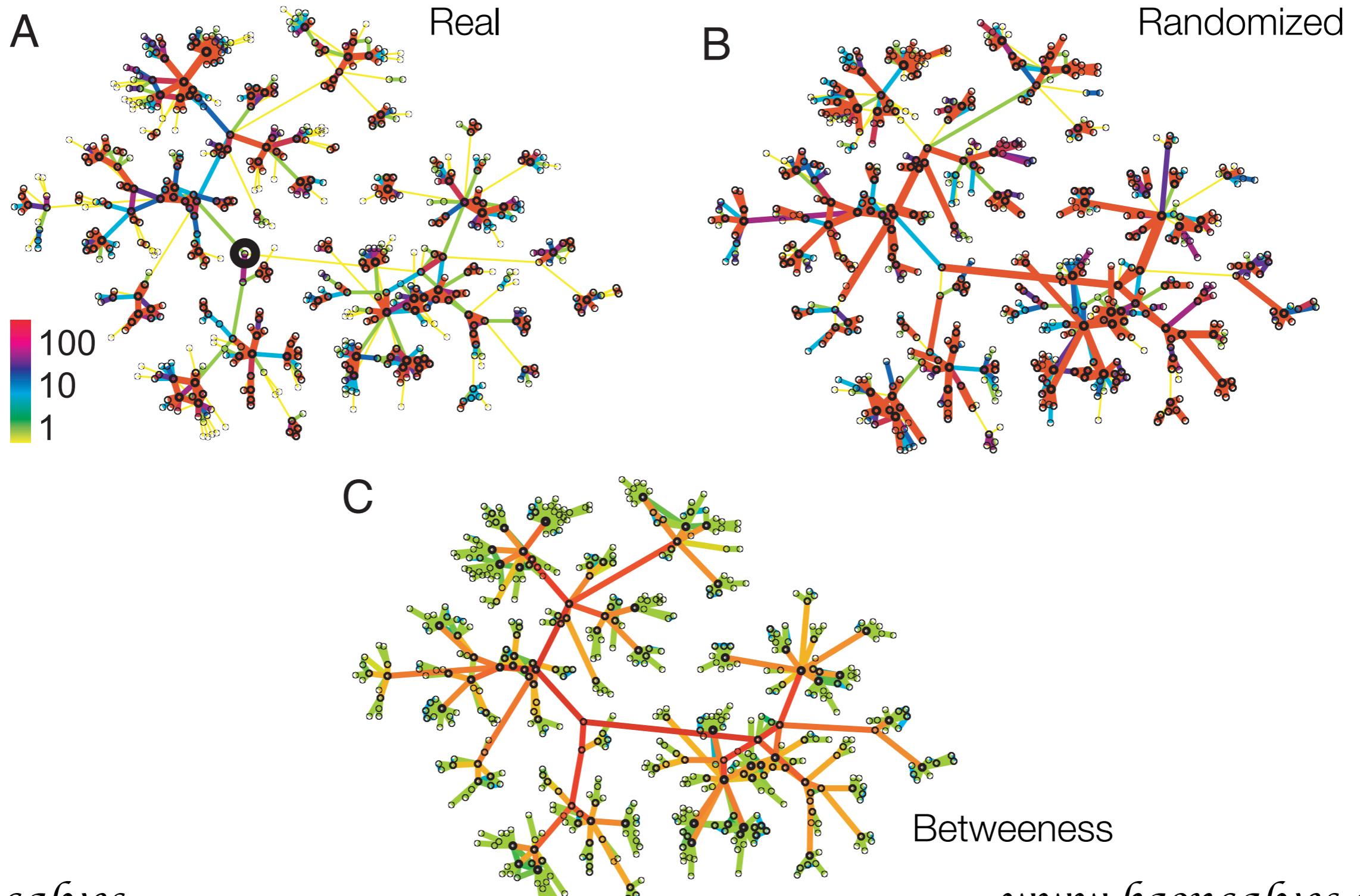
Neighborhood Overlap

PNAS 104, 7333 (2007)



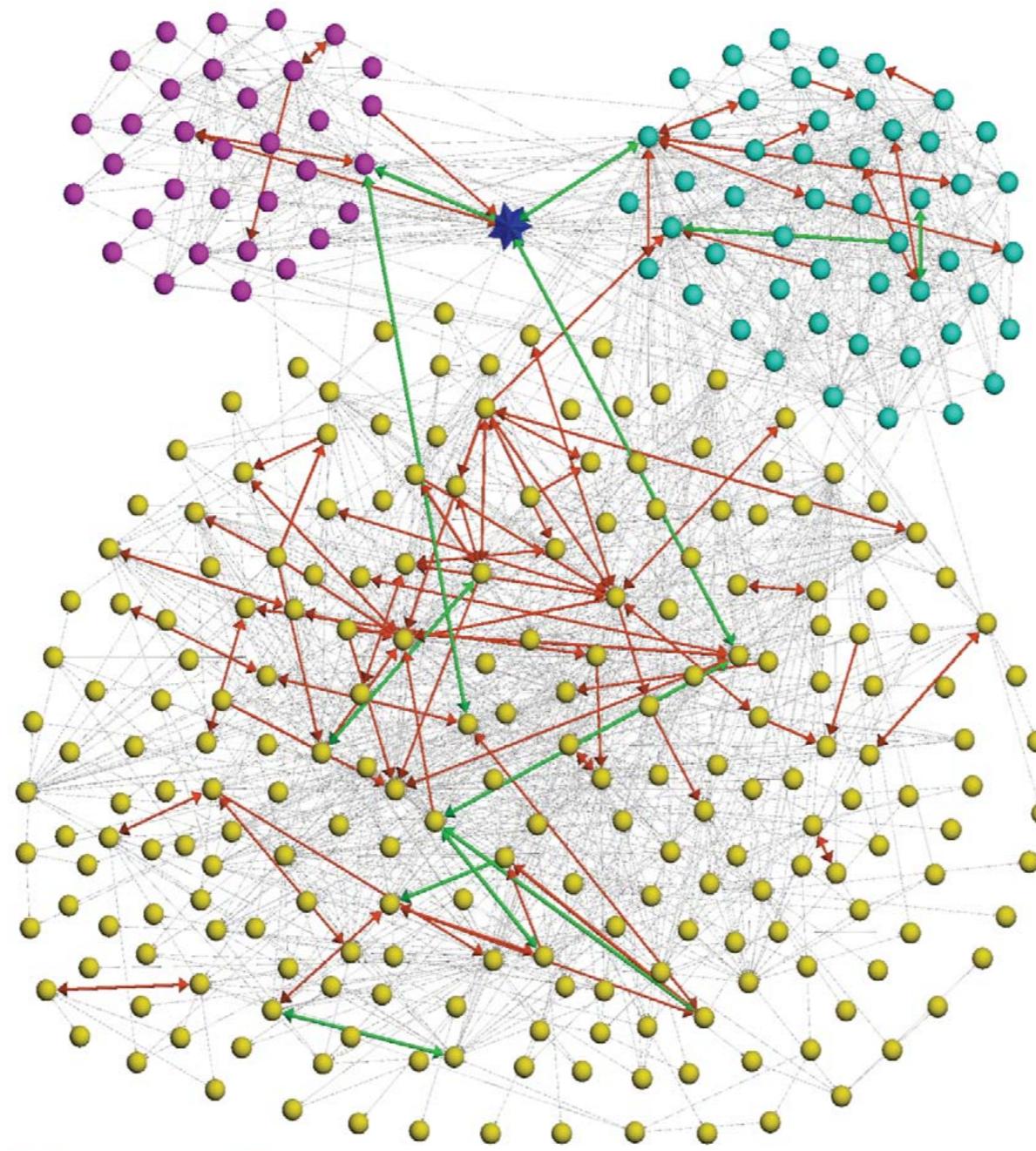
Strong Ties have higher overlaps

PNAS 104, 7333 (2007)



Network Structure

PLoS One 7, e29358 (2012)



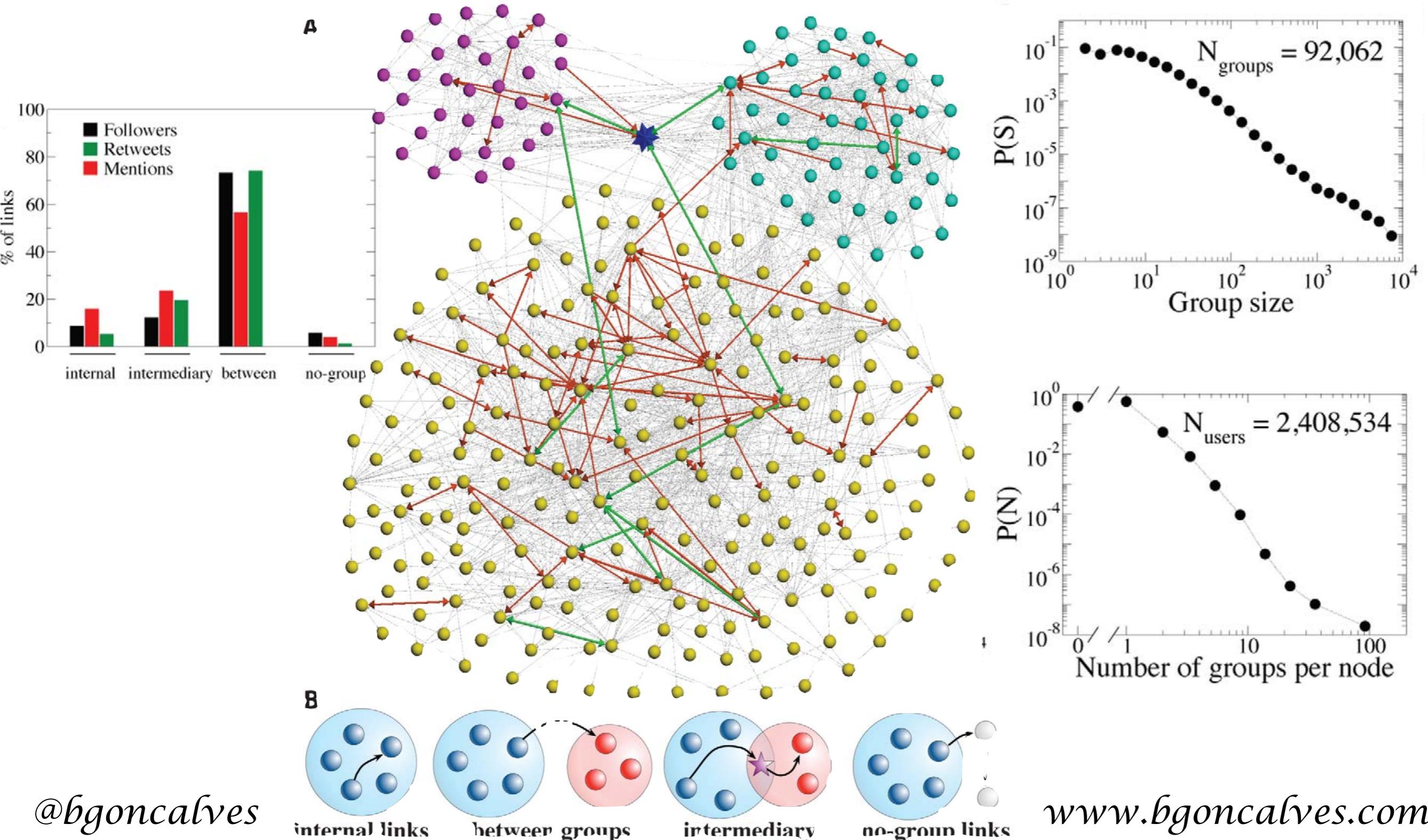
Structural Holes and Good Ideas¹

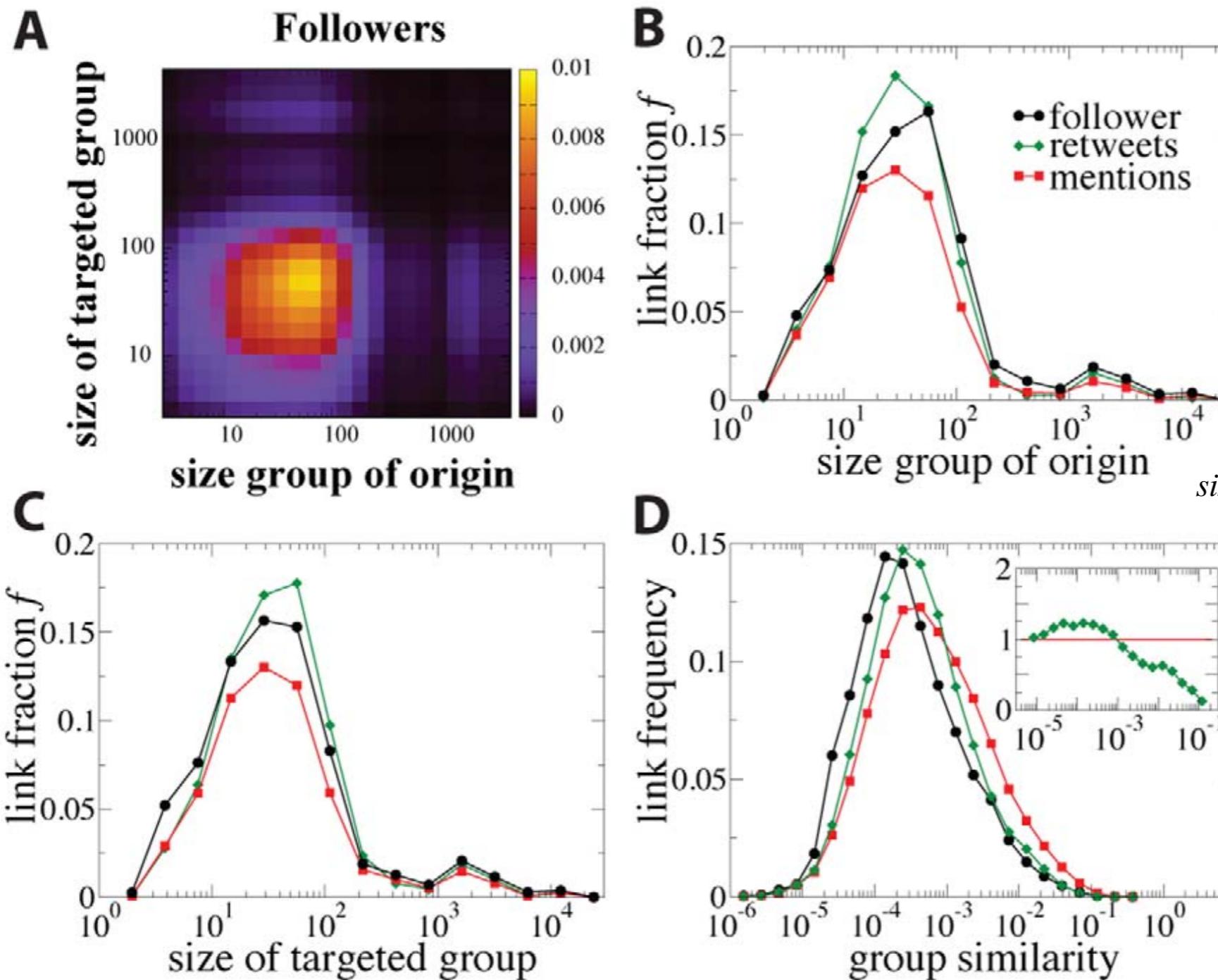
Ronald S. Burt
University of Chicago

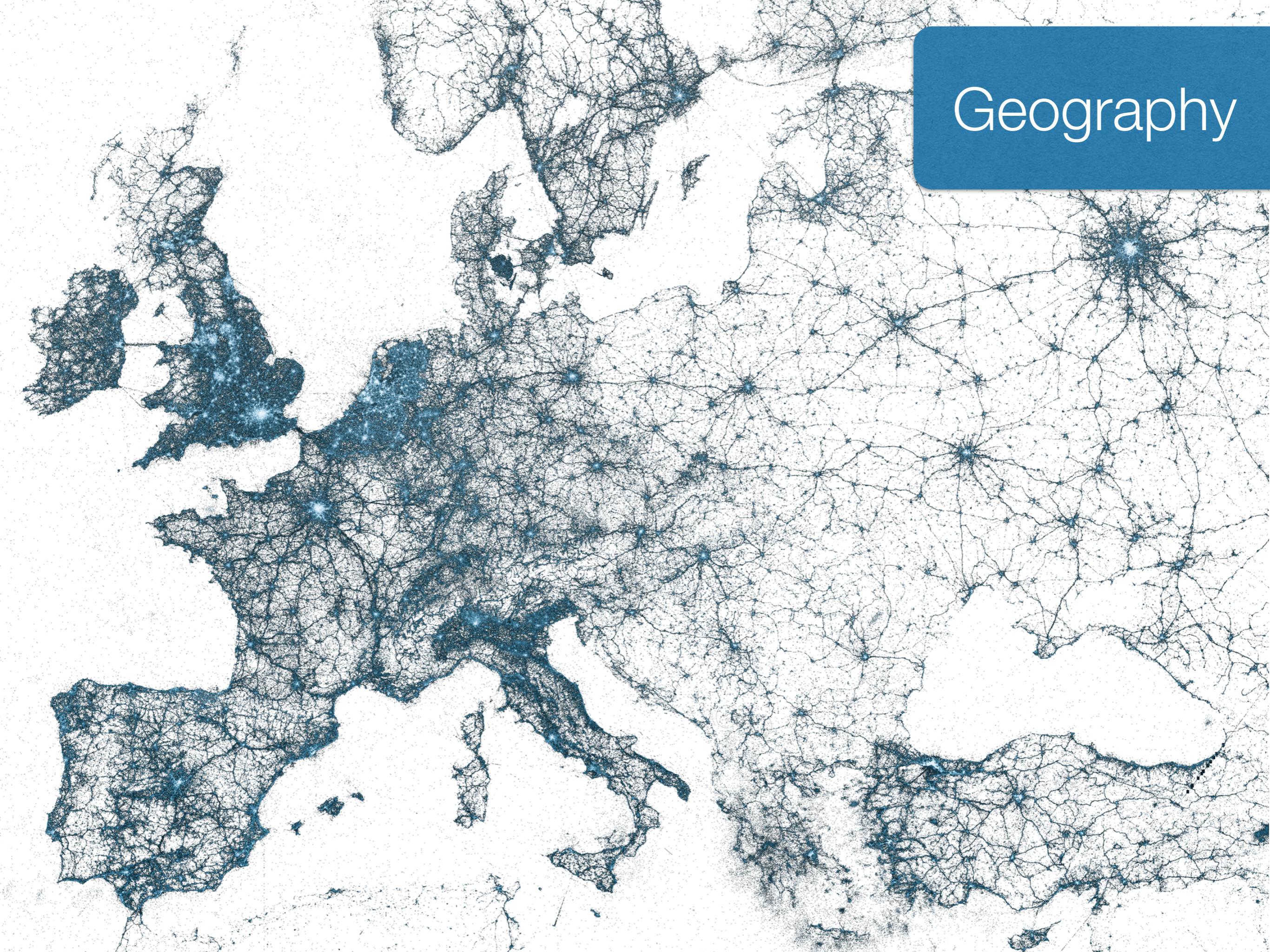
“People whose networks bridge the structural holes between groups have an advantage in detecting and developing rewarding opportunities. Information arbitrage is their advantage. They are able to see early, see more broadly, and translate information across groups.”

Network Structure

PLoS One 7, e29358 (2012)



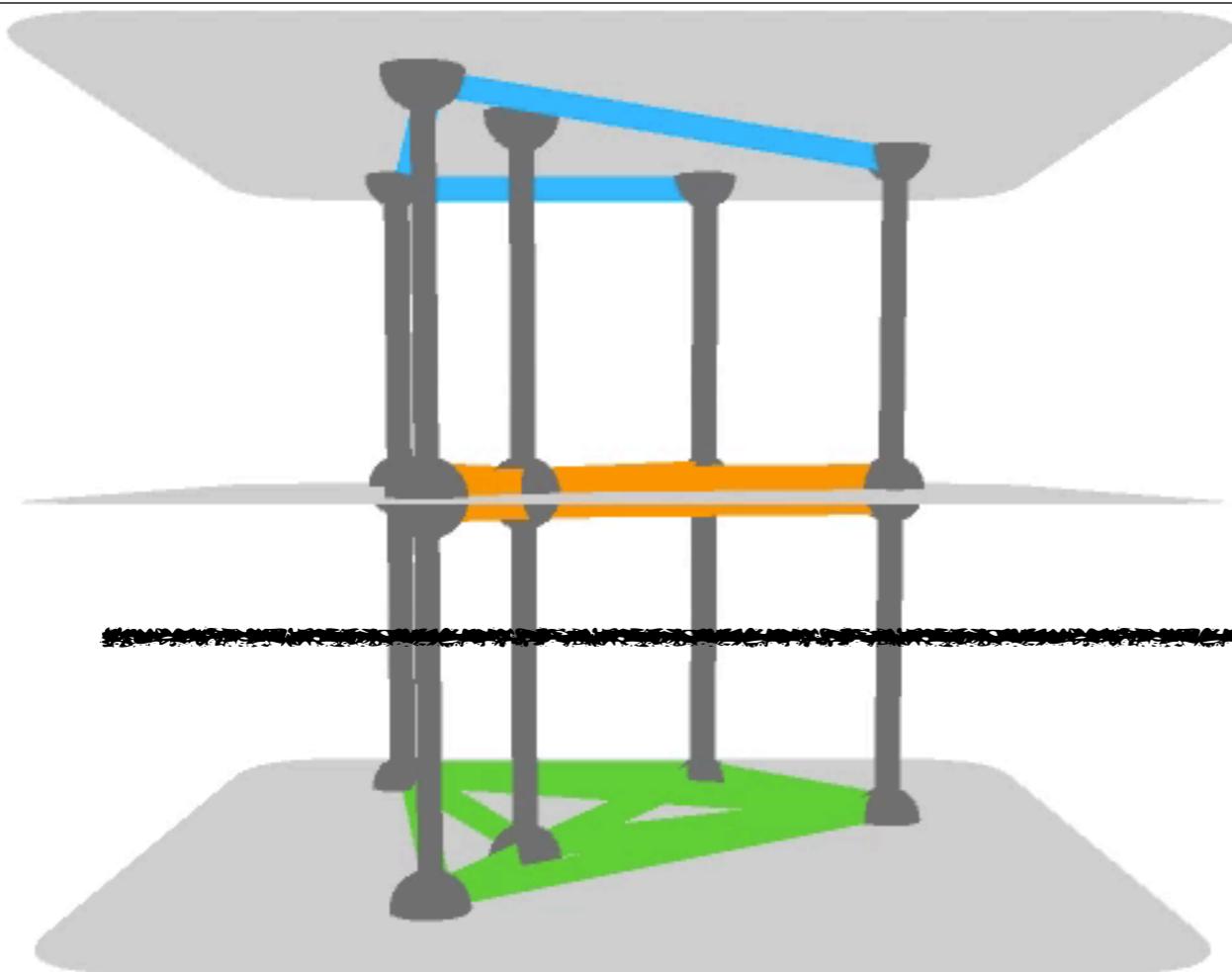


A world map where every country's borders are represented by a dense web of blue lines and small dots. The density of the network varies by country, with higher connectivity shown in darker shades of blue. Major cities are marked with larger, glowing blue dots. The overall pattern highlights the interconnectedness of the world's economies and networks.

Geography

Multilayer Network

Retweet



Information
Layer(s)

Mention

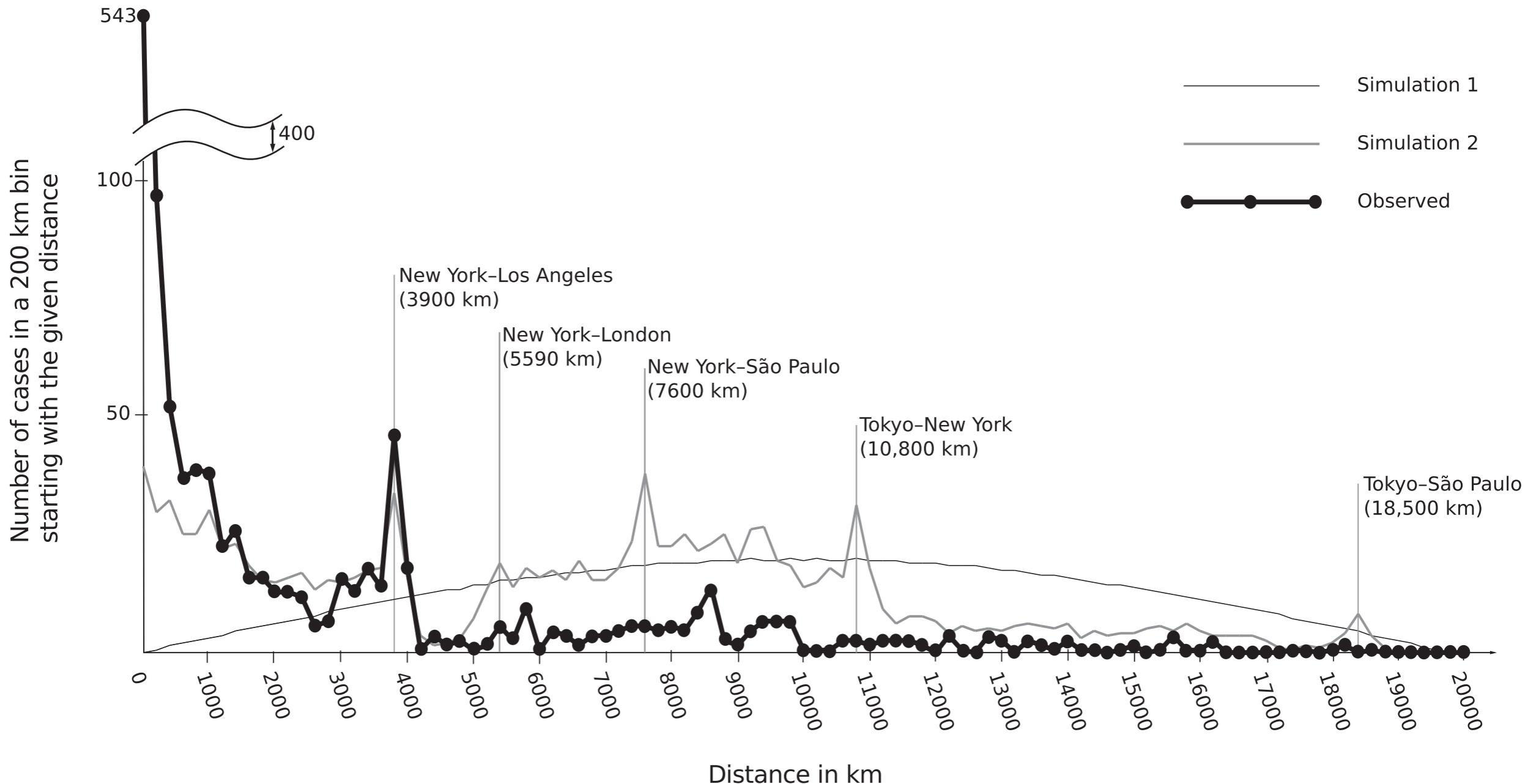
Social
Layer(s)

Follower

Geographical
Layer(s)

Twitter Follower Distance

Social Networks 34, 73 (2012)



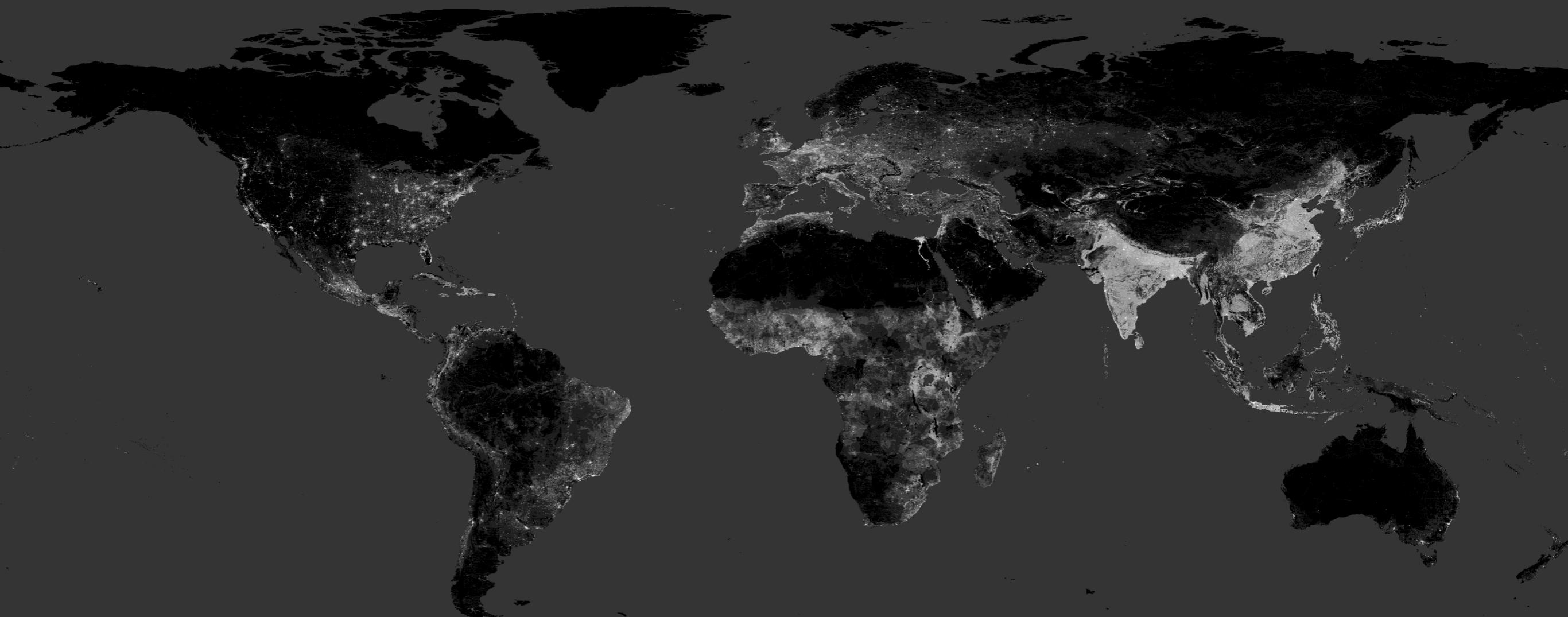
| | Share of egos (%) ^a | Share of egos (%) for egos in dyads ^b | Share of alters (%) ^c | Percentage of domestic ties ^d | Percentage of domestic ties among non-local ties ^d |
|-------------|-----------------------------------|--|-------------------------------------|---|---|
| USA | 48.5 | 45.7 | 54.5 | 91.6 | 89.3 |
| Brazil | 10.6 | 12.1 | 10.5 | 83.5 | 72.5 |
| UK | 7.6 | 8.3 | 7.6 | 50.6 | 33.3 |
| Japan | 5.5 | 6.5 | 6.3 | 92.1 | 86.0 |
| Canada | 3.7 | 3.8 | 2.9 | 33.3 | 23.1 |
| Australia | 2.7 | 2.7 | 1.9 | 50.0 | 32.0 |
| Indonesia | 2.6 | 1.8 | 1.2 | 60.0 | 25.0 |
| Germany | 2.1 | 1.8 | 1.3 | 62.9 | 58.8 |
| Netherlands | 1.4 | 1.4 | 1.2 | 66.7 | 22.2 |
| Mexico | 1.2 | 1.3 | 0.7 | 44.0 | 8.3 |

^a Out of the 2852 egos located at the level of country or better.^b Out of the egos included in 1953 dyads with both parties located at the level of country or better.^c Out of the 1953 alters located at the level of country or better.^d The number of ties with the ego and the alter in the given country as a share of all ties for egos in that country.

| Rank | Cluster ^a | Share of egos (%) ^b | Share of egos (%) for egos in dyads ^c | Share of alters (%) ^d | Locality ^e |
|------|---------------------------|-----------------------------------|--|-------------------------------------|-----------------------|
| 1 | "New York" | 8.5 | 8.3 | 10.2 | 54.3 |
| 2 | "Los Angeles, CA" | 5.1 | 5.6 | 10.4 | 53.3 |
| 3 | "東京都" (Tokyo) | 4.1 | 4.8 | 5.0 | 62.9 |
| 4 | "London" | 3.6 | 3.3 | 4.9 | 48.8 |
| 5 | "São Paulo" | 3.5 | 3.0 | 3.6 | 78.4 |
| 6 | "San Francisco" | 2.8 | 2.7 | 4.1 | 41.2 |
| 7 | "New Jersey" ^f | 2.5 | 2.8 | 2.1 | 20.0 |
| 8 | "Chicago" | 2.2 | 2.0 | 1.7 | 32.0 |
| 9 | "Washington, DC" | 2.1 | 2.8 | 2.6 | 34.3 |
| 10 | "Manchester, UK" | 1.9 | 2.0 | 1.1 | 30.8 |
| 11 | "Atlanta" | 1.7 | 2.1 | 2.1 | 46.2 |
| 12 | "San Diego" | 1.5 | 1.5 | 1.1 | 26.3 |
| 13 | "Toronto, Canada" | 1.3 | 1.1 | 1.5 | 42.9 |
| 14 | "Seattle" | 1.3 | 1.4 | 1.2 | 58.8 |
| 15 | "Houston" | 1.2 | 1.2 | 1.0 | 40.0 |
| 16 | "Dallas, Texas" | 1.2 | 1.0 | 1.4 | 61.5 |
| 17 | "Rio de Janeiro" | 1.2 | 1.0 | 1.1 | 30.8 |
| 18 | "Boston, MA" | 1.2 | 1.2 | 1.1 | 20.0 |
| 19 | "Amsterdam" | 1.1 | 1.1 | 0.9 | 50.0 |
| 20 | "Jakarta, Indonesia" | 1.1 | 0.6 | 0.3 | 42.9 |
| 21 | "Austin, TX" | 1.0 | 1.0 | 1.3 | 50.0 |
| 22 | "Sydney" | 0.9 | 1.0 | 0.8 | 38.5 |
| 23 | "Orlando, Florida" | 0.9 | 1.0 | 0.6 | 16.7 |
| 24 | "Phoenix, AZ" | 0.8 | 0.7 | 0.6 | 11.1 |
| 25 | "兵庫" (Hyōgo) ^g | 0.8 | 1.0 | 1.0 | 25.0 |

^a Each cluster is labeled with the name most frequently used for locations assigned to the cluster.^b Out of the 2167 egos located with precision of <25,000 km².^c Out of the 1259 egos included in dyads with both parties located with precision of <25,000 km².^d Out of the 1259 alters included in dyads with both parties located with precision of <25,000 km².^e Defined as the share of local of ties among all ties for egos in a cluster.^f Centered between Philadelphia and Trenton, NJ and includes all locations identified as just "New Jersey".^g Centered near the boundary between Hyōgo and Osaka prefectures, in the Kansai region of Japan.

World Population



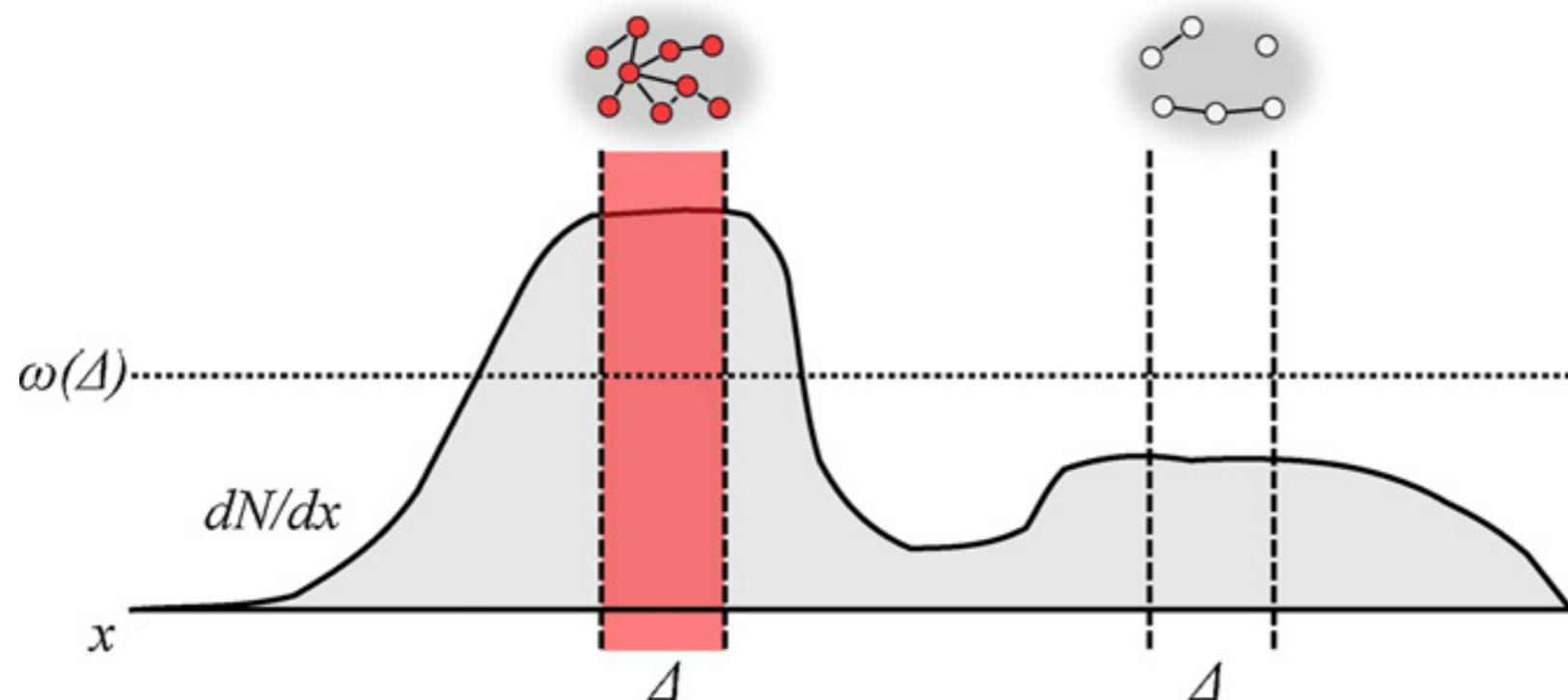
Population Heterogeneity

Social Networks 34, 82 (2012)

- Bernoulli process to generate adjacency matrix given a distance matrix between nodes

$$P(A = a|D) = \prod_{\{i,j\}} B(A_{ij} = a_{ij} | \mathcal{F}(D_{ij}, \theta))$$

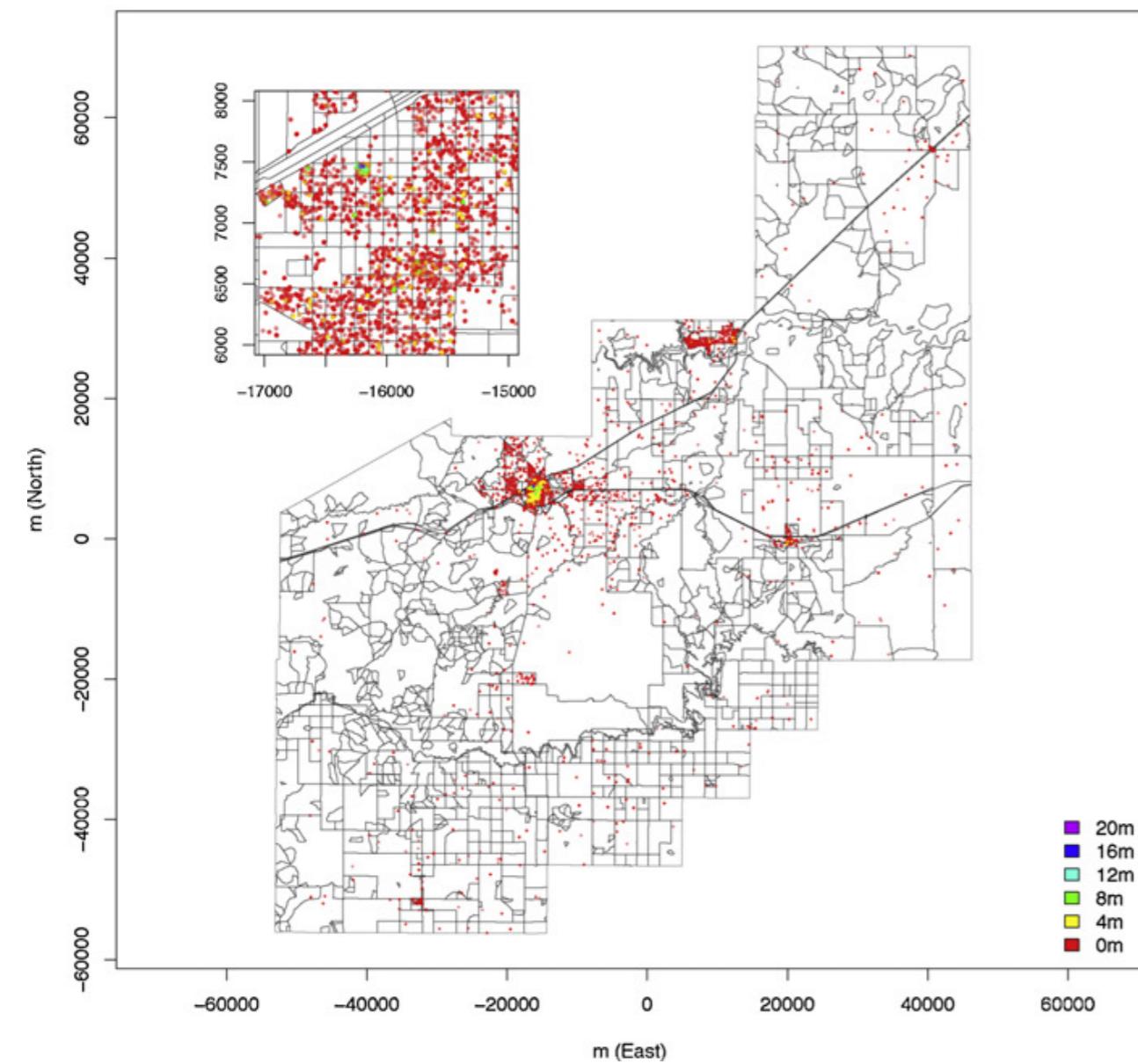
- Above some density threshold, $\omega(\Delta)$ networks is naturally connected.



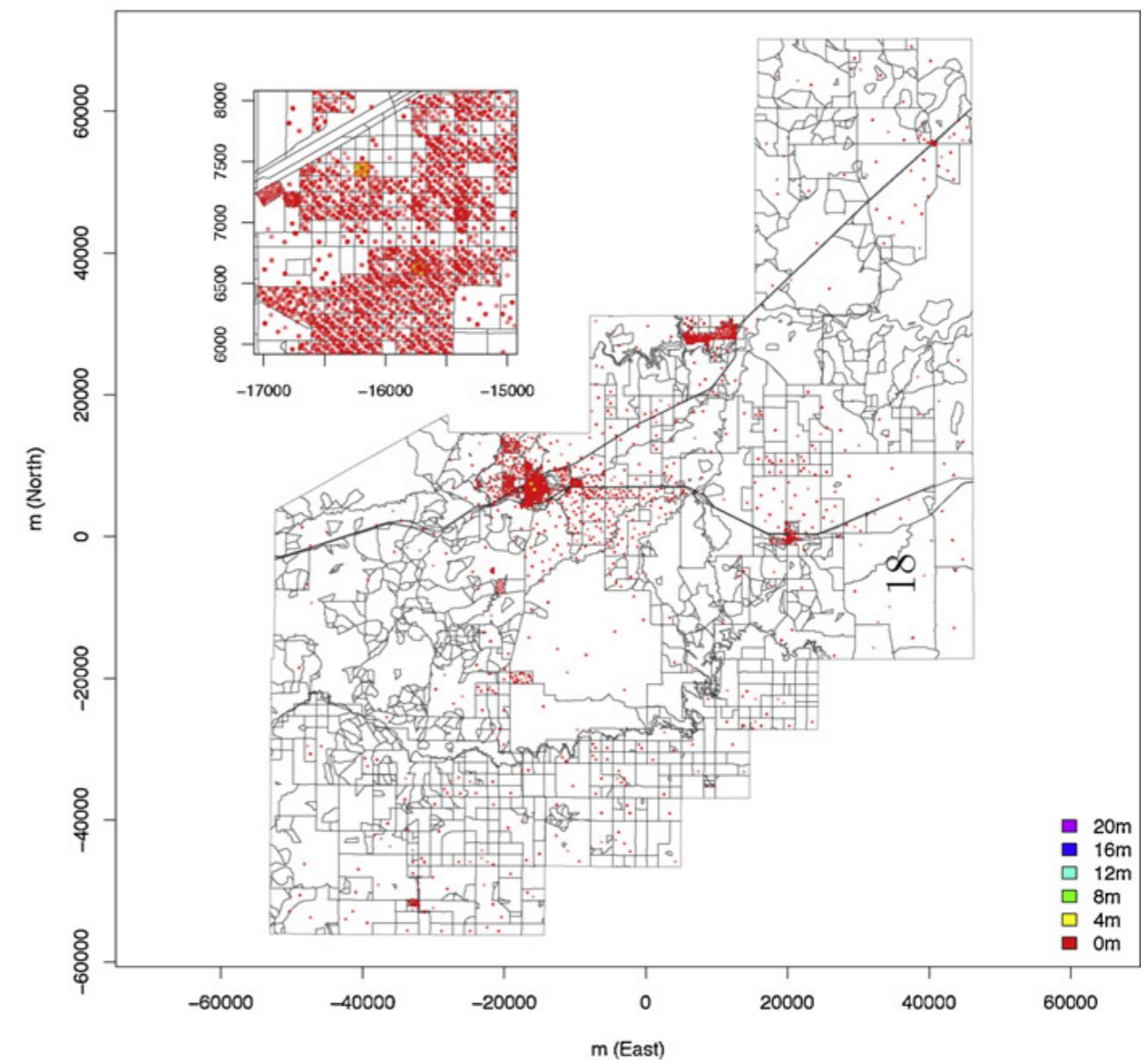
Vertex Placement

Social Networks 34, 82 (2012)

Quay, NM, Uniform Vertex Placement



Quay, NM, Quasi-Random Vertex Placement



Friendship probability

Social Networks 34, 82 (2012)

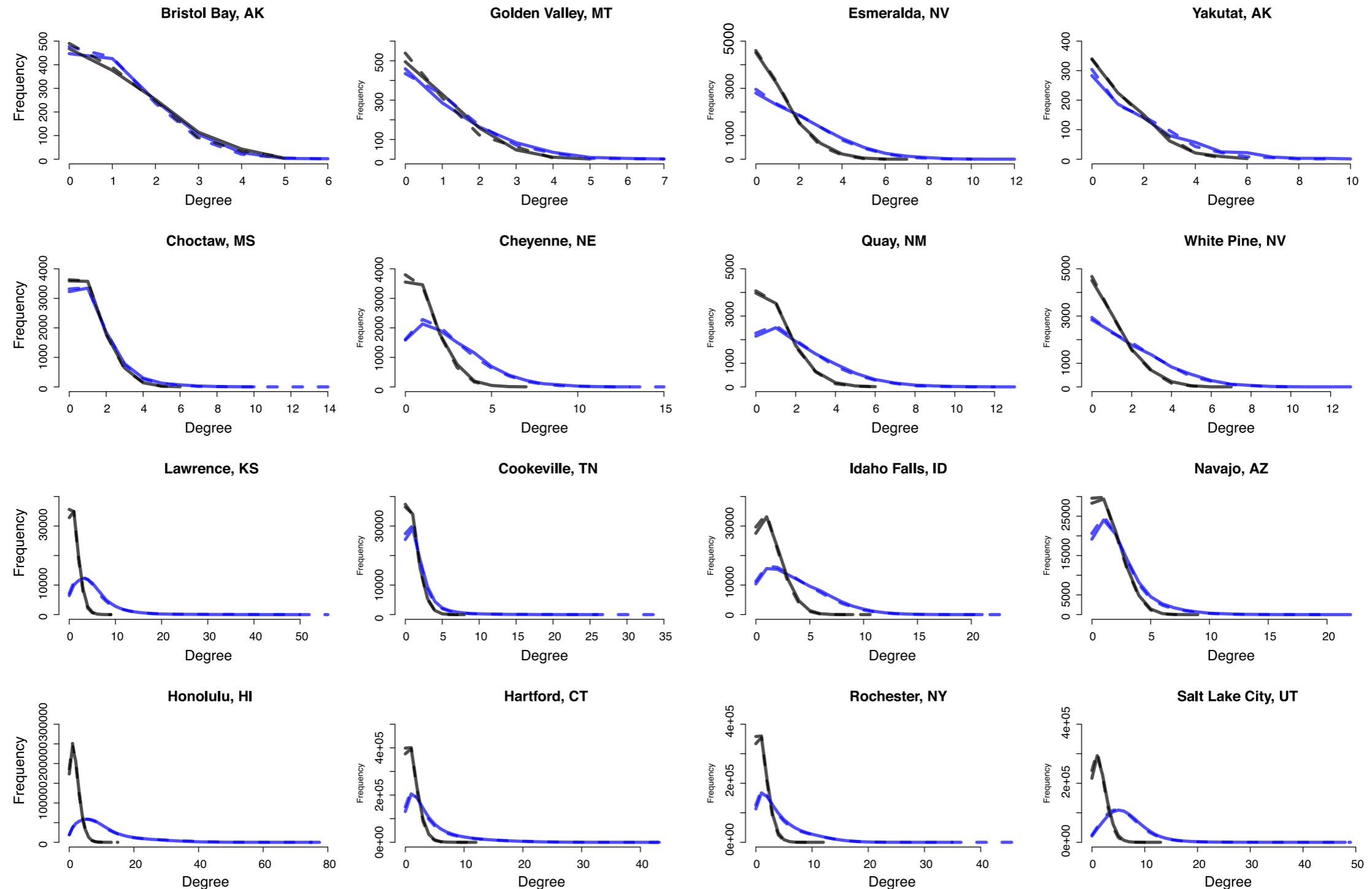
- Probability that two people are friends as a function of distance:

$$\mathcal{F}(d) = \frac{\theta_1}{(1 + \theta_2 d)^{\theta_3}}$$

- with **(0.533, 0.032, 2.788)** for "social friendships" and **(0.859, 0.035, 6.437)** for "face-to-face interactions".

Social Network Properties

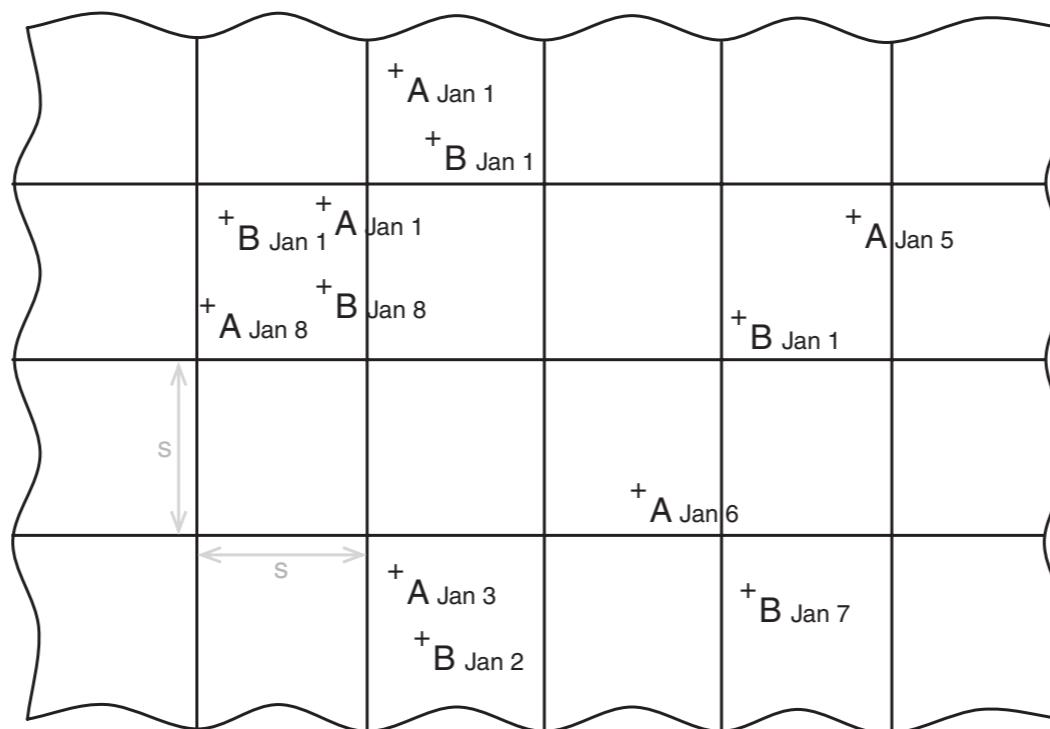
Social Networks 34, 82 (2012)



Co-occurrences and Social Ties

PNAS 107, 22436 (2010)

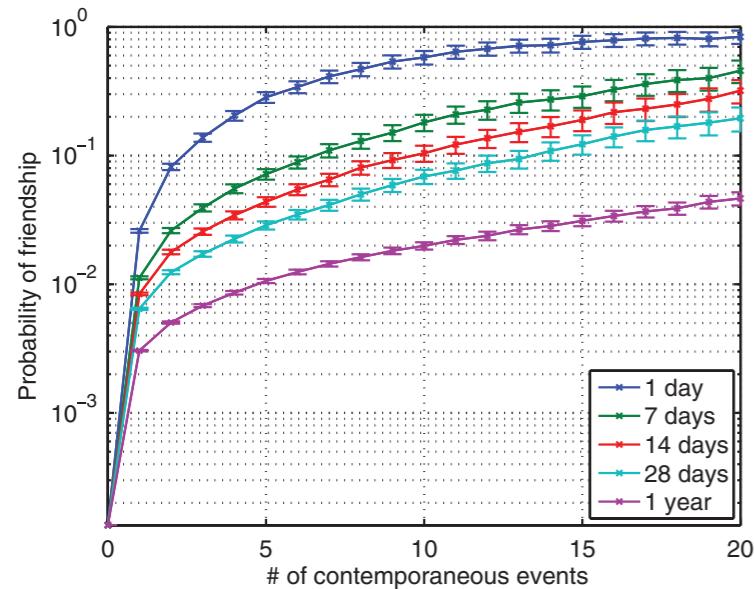
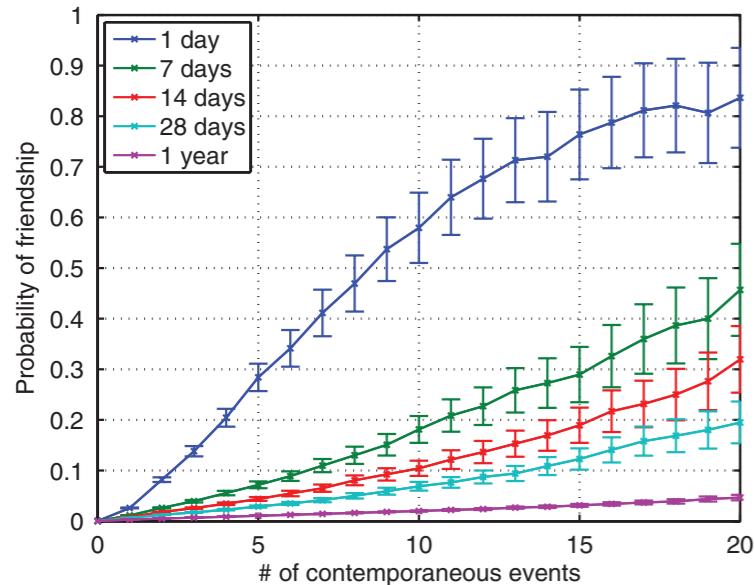
- Geotagged Flickr Photos
- Divide the world into a grid



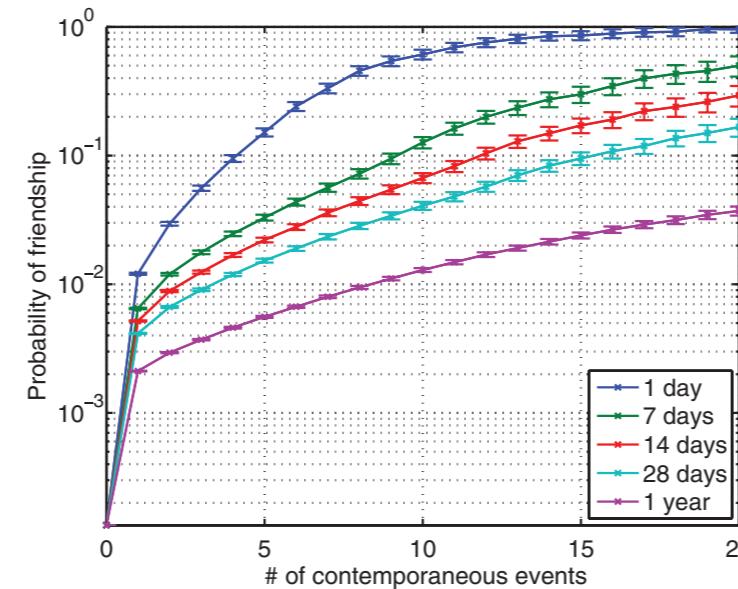
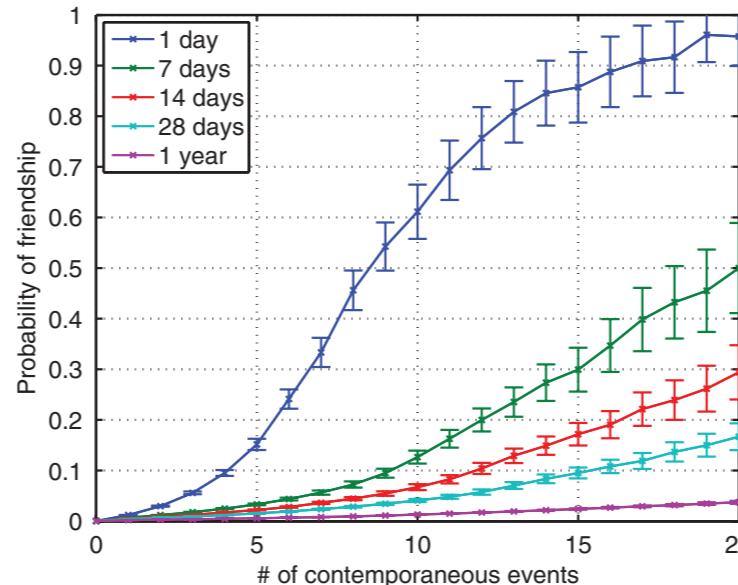
Count number of cells on which two individuals were within a given interval

Co-occurrences and Social Ties

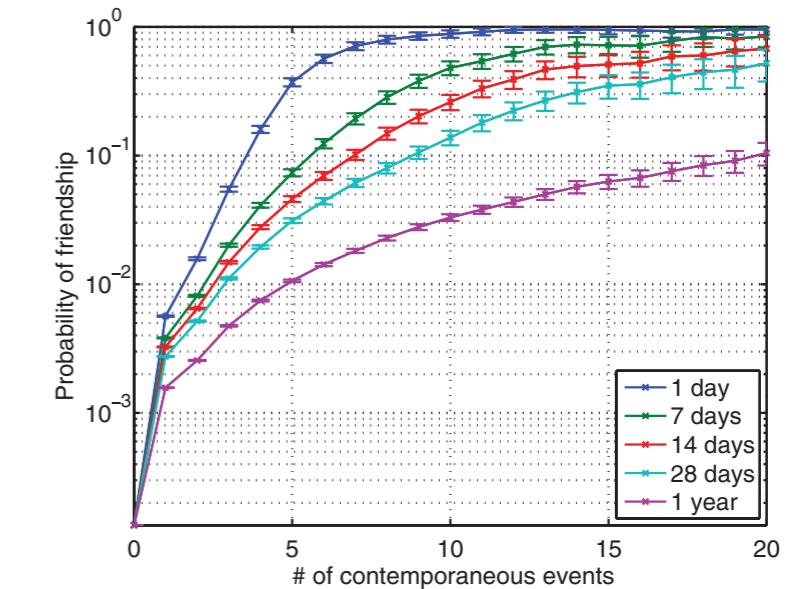
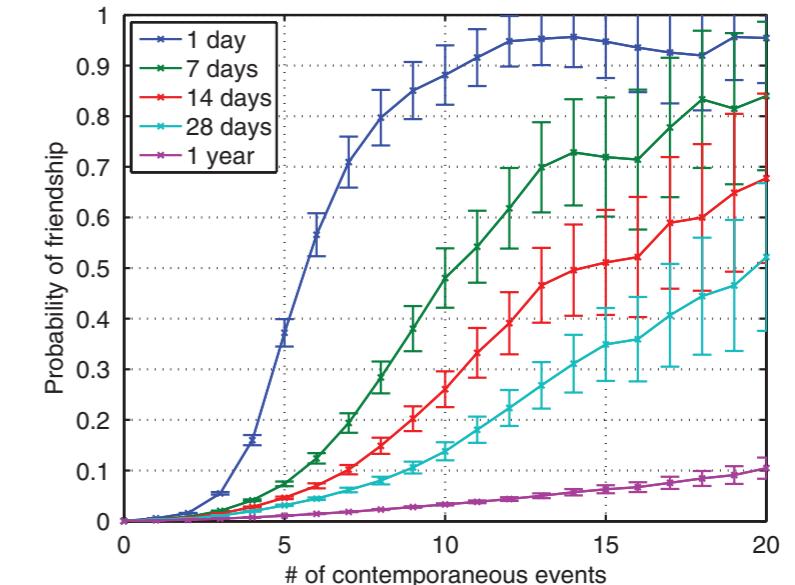
PNAS 107, 22436 (2010)



A $s = 0.001^\circ$



B $s = 0.01^\circ$



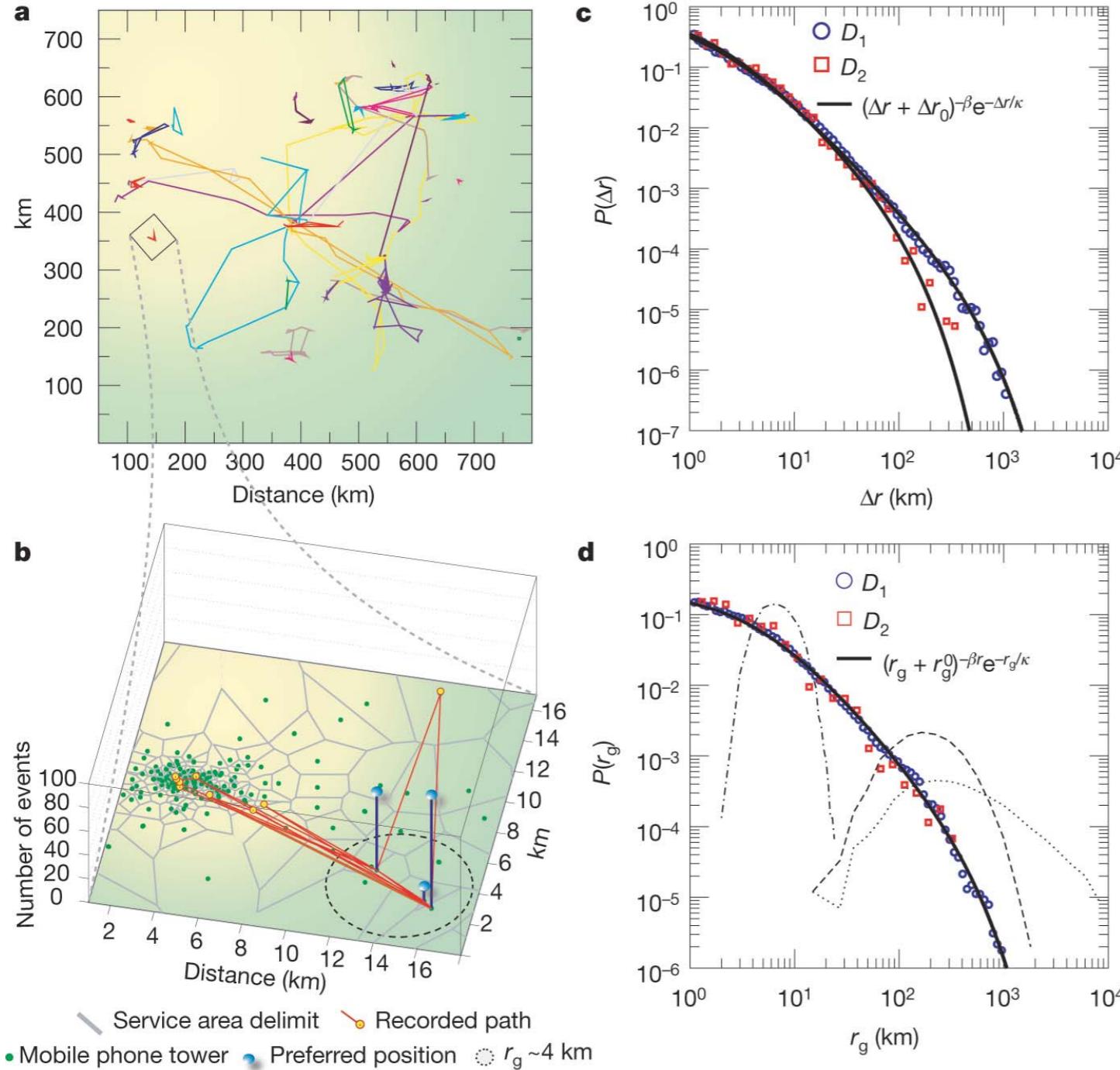
C $s = 0.1^\circ$

Geolocated Tweets



Human Mobility

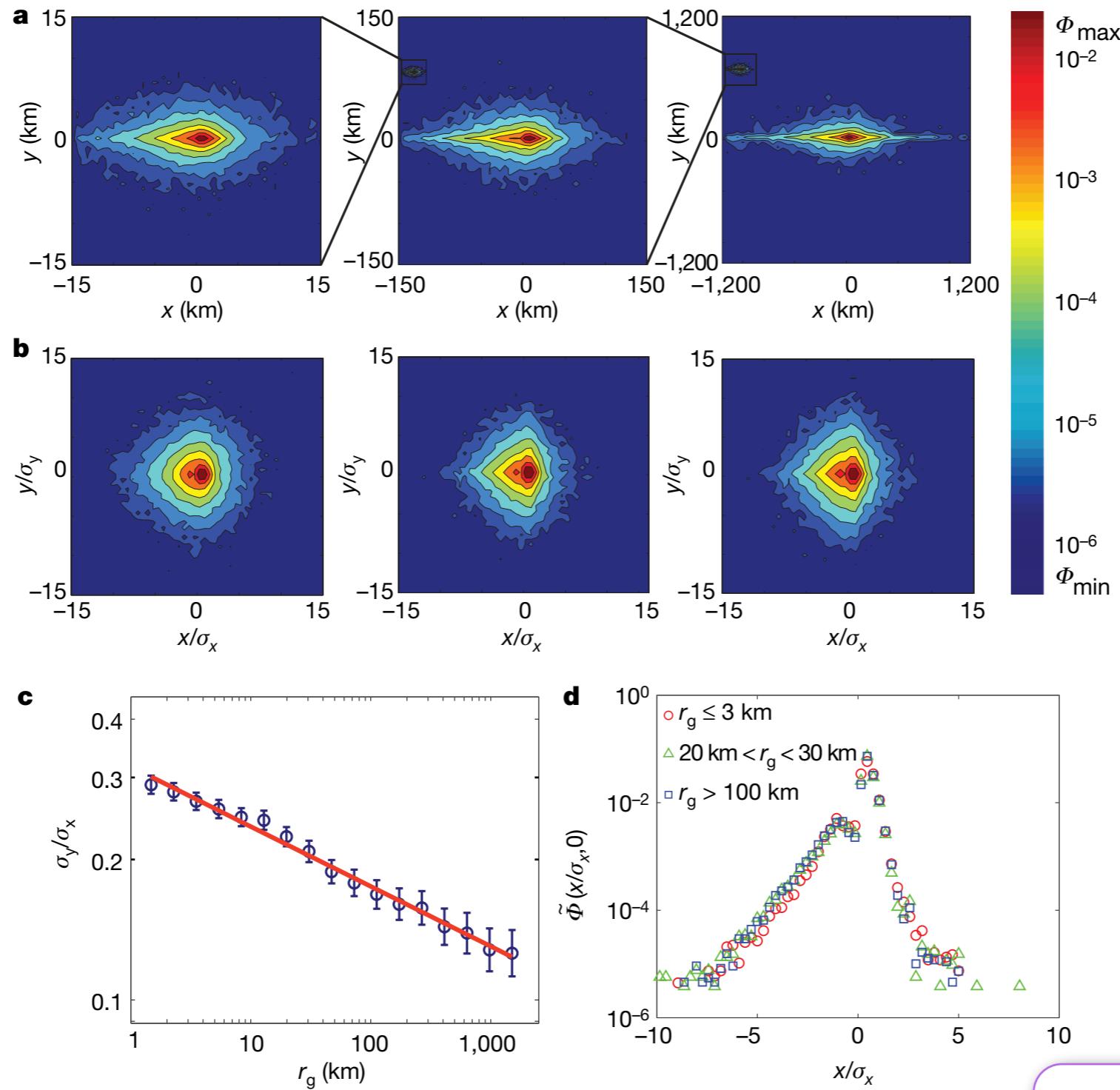
Nature 453, 779 (2008)

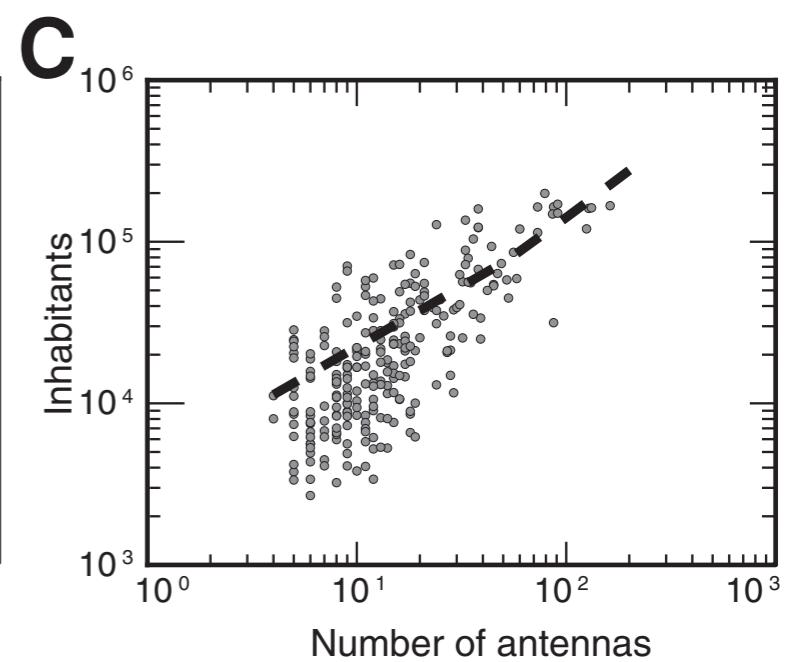
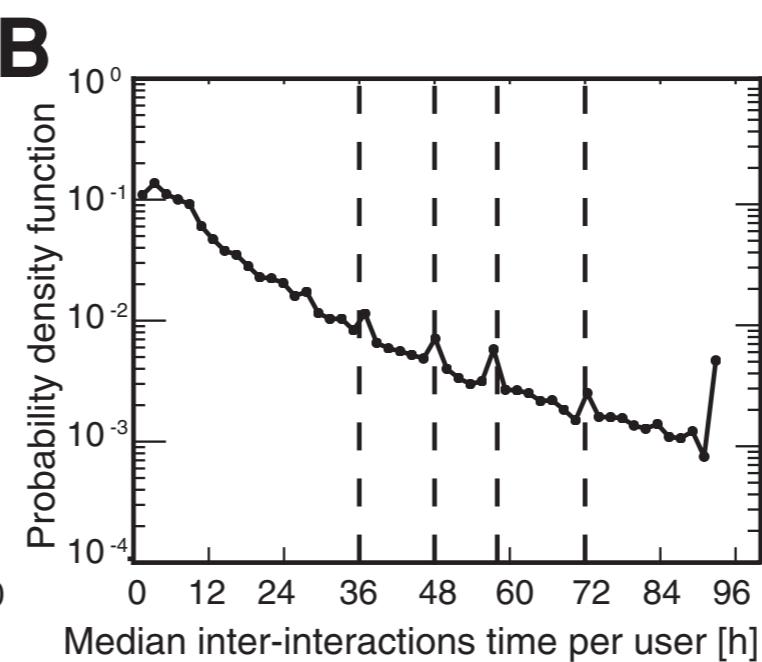
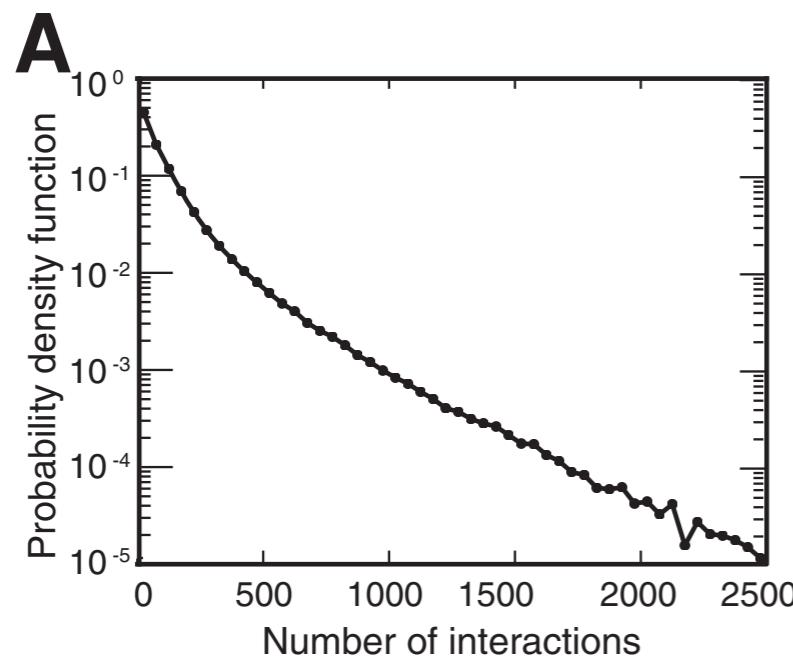
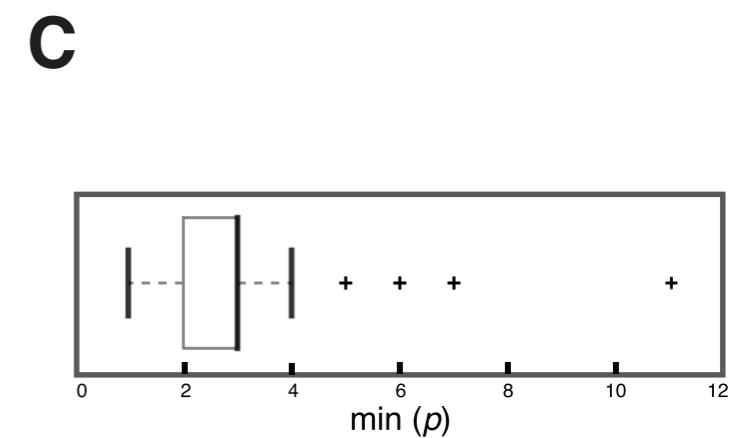
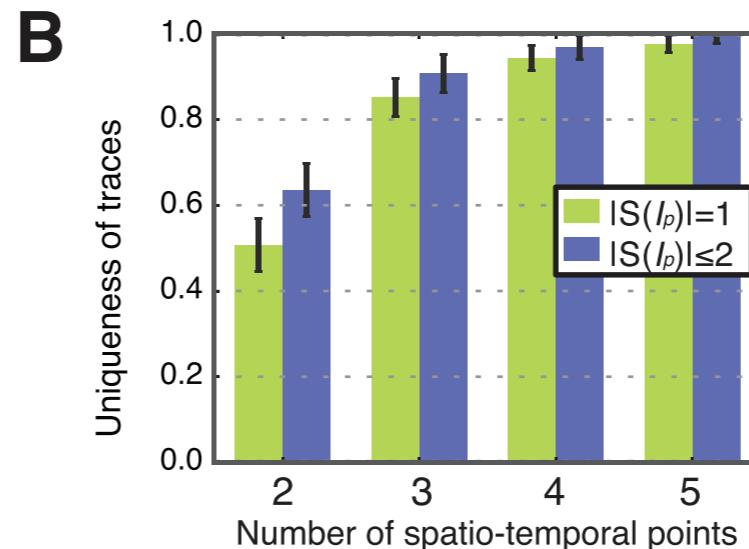
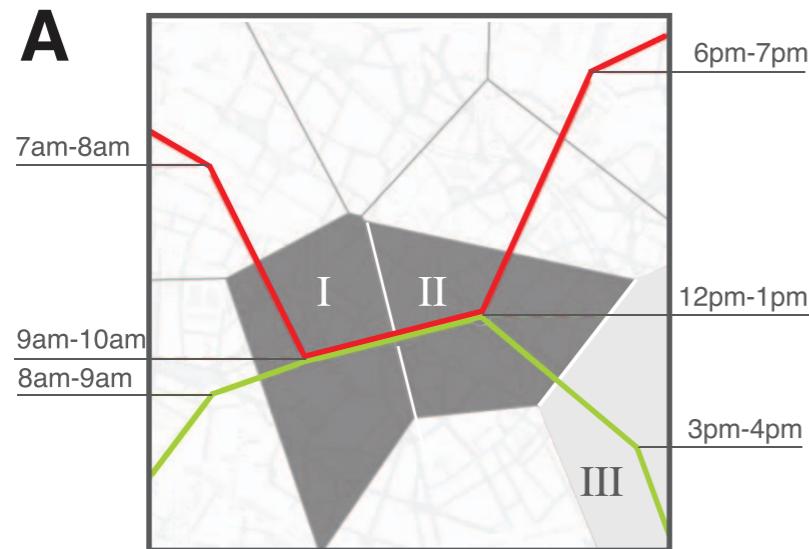


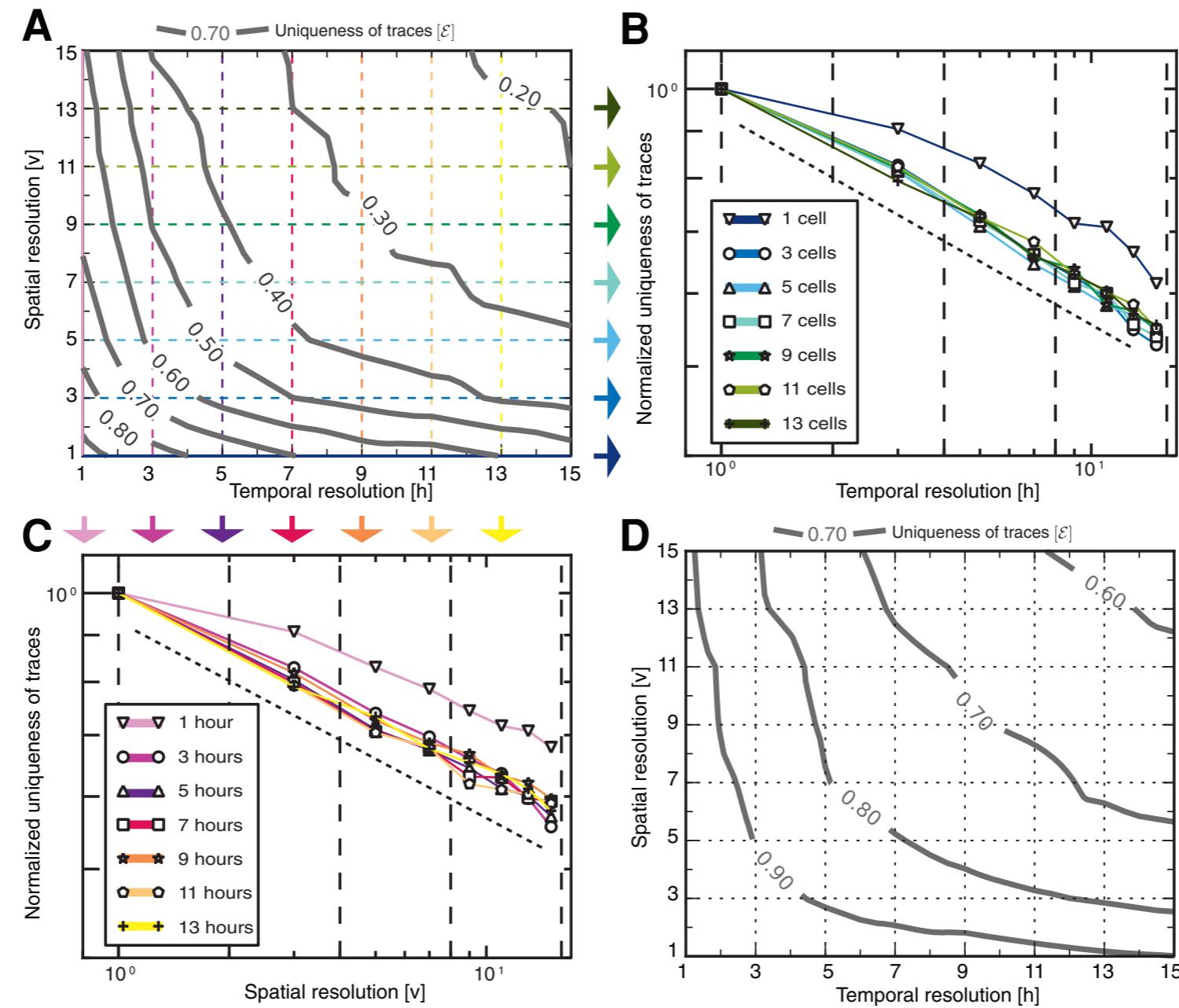
$$P(\Delta r) = (\Delta r + \Delta r_0)^{-\beta} \exp(-r_g/\kappa)$$

Human Mobility

Nature 453, 779 (2008)



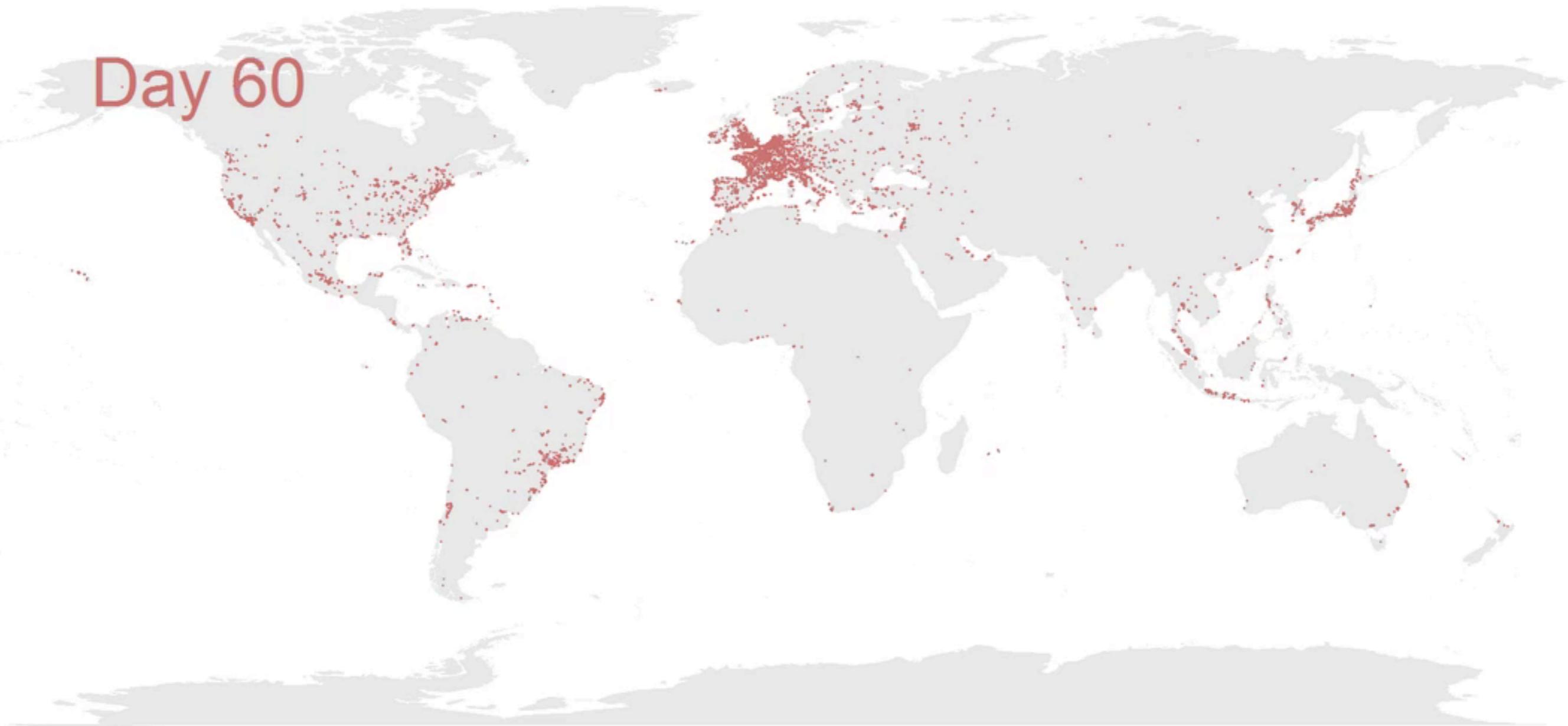




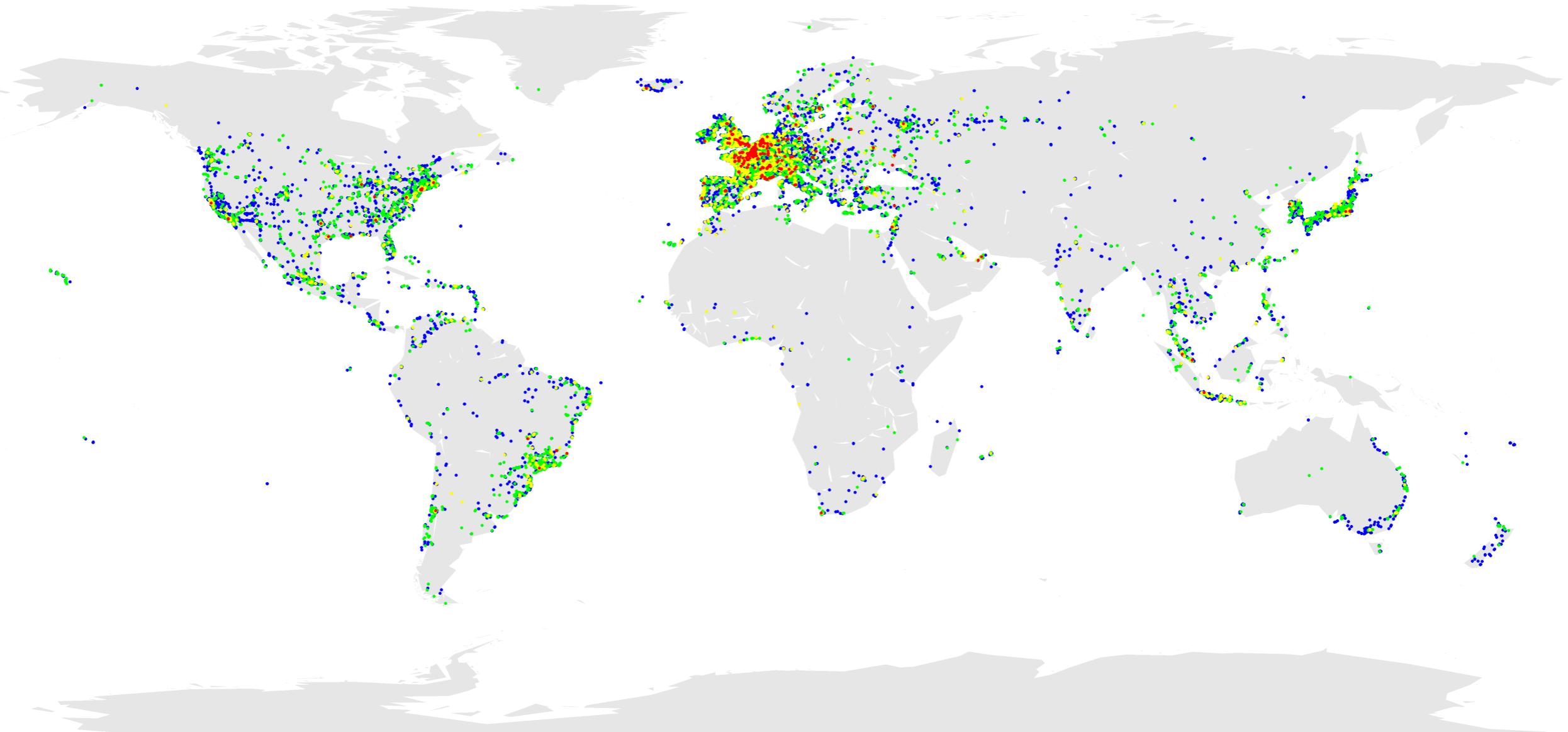
Human Diffusion

J. R. Soc. Interface 12, 20150473 (2015)
EPJ Data Science 5, 12 (2016)

Day 60



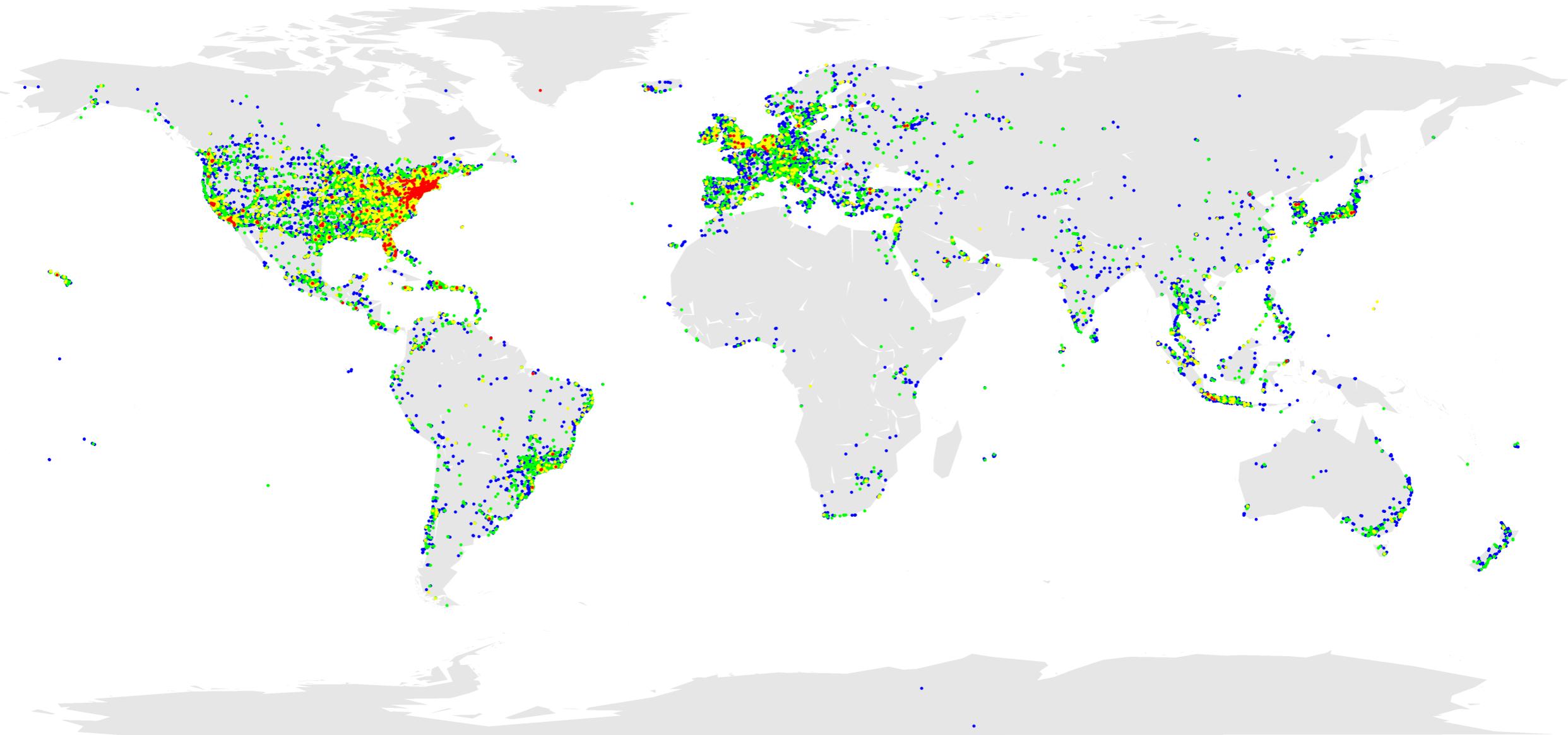
Starting from Paris



Human Diffusion

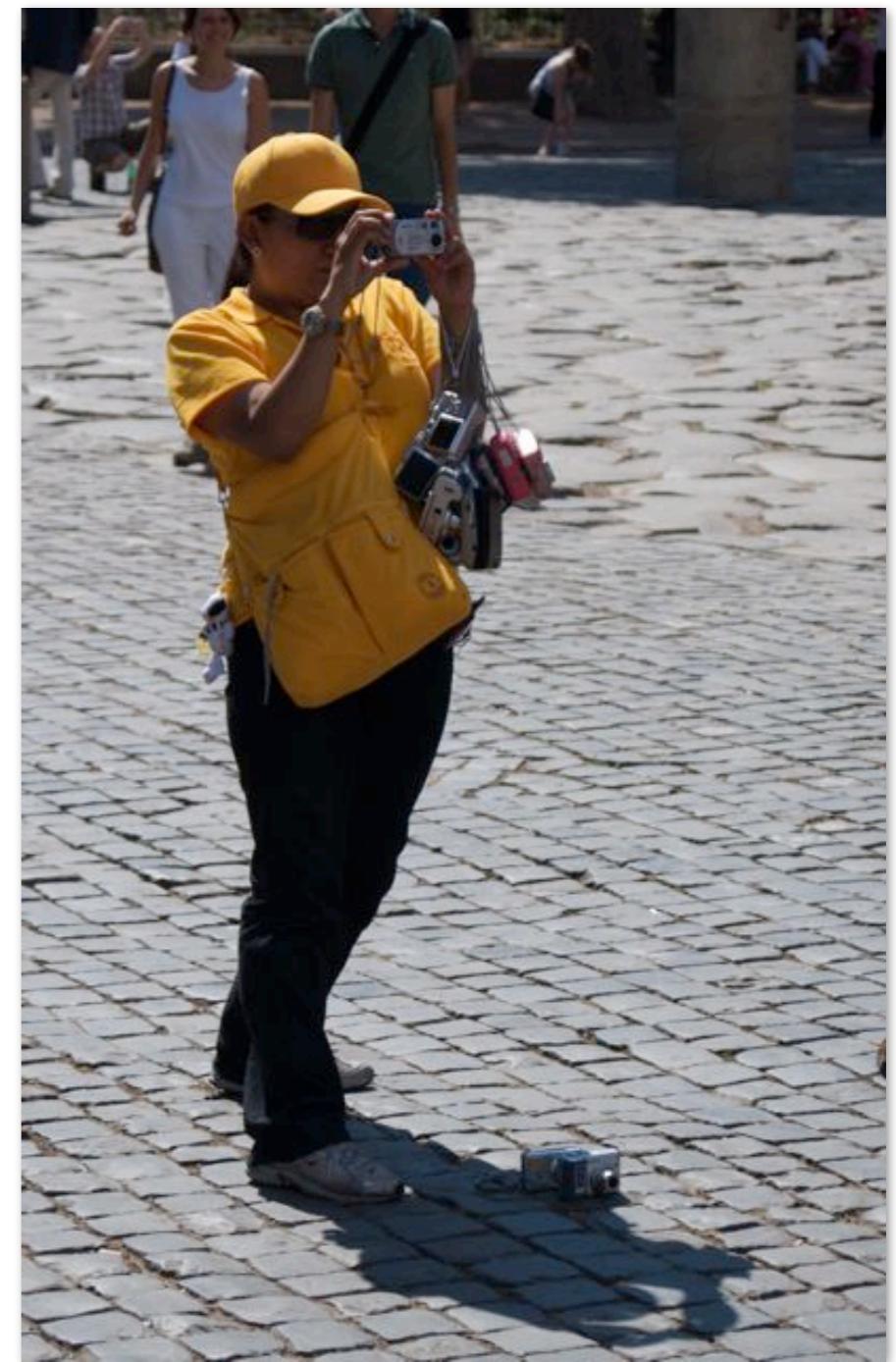
J. R. Soc. Interface 12, 20150473 (2015)
EPJ Data Science 5, 12 (2016)

Starting from New York



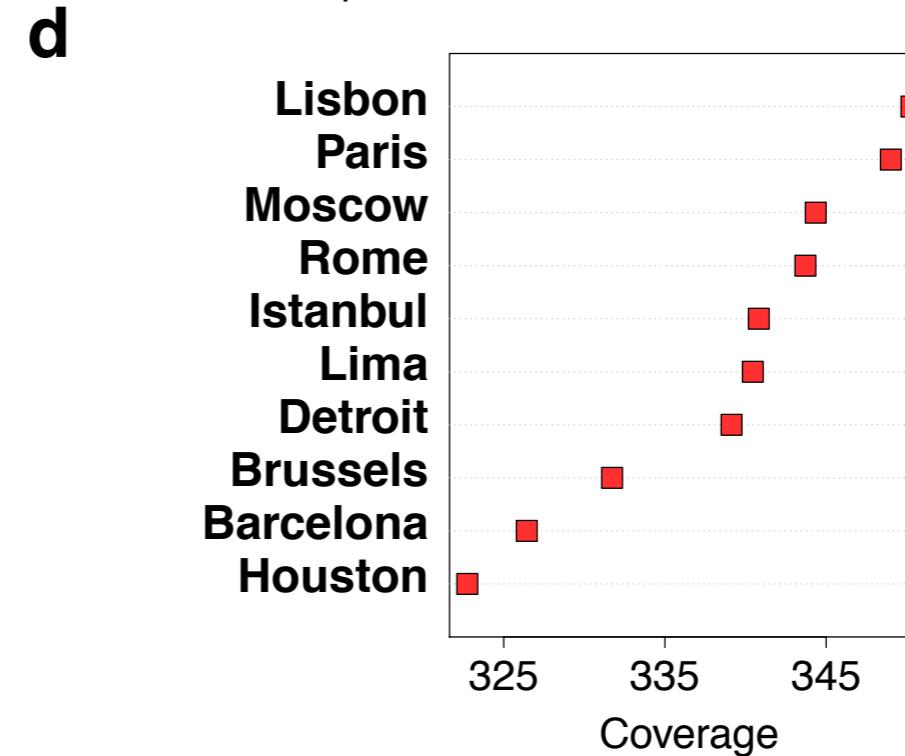
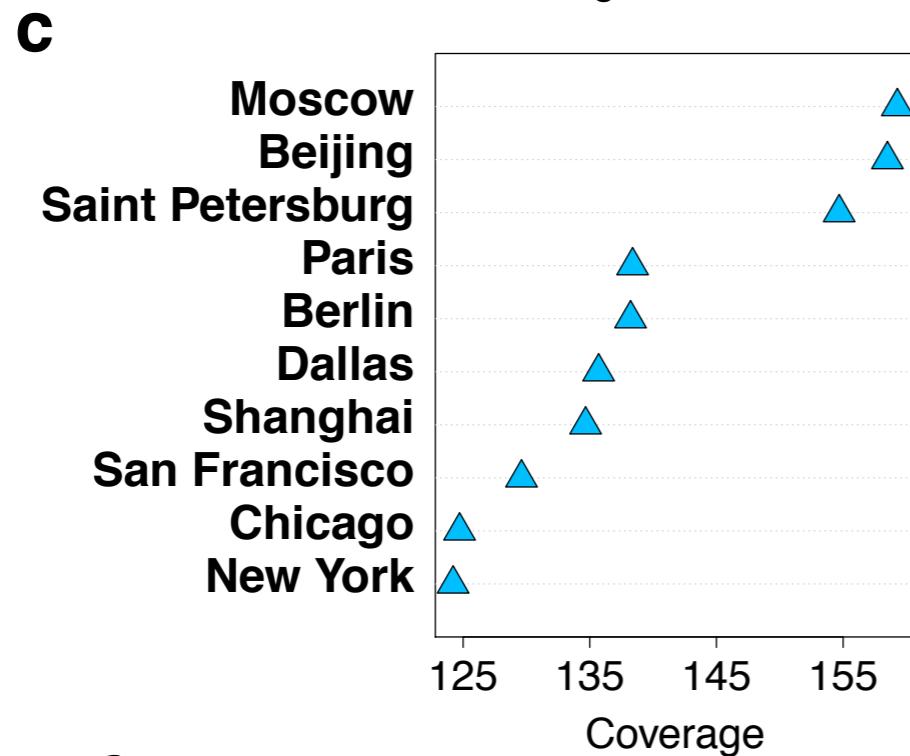
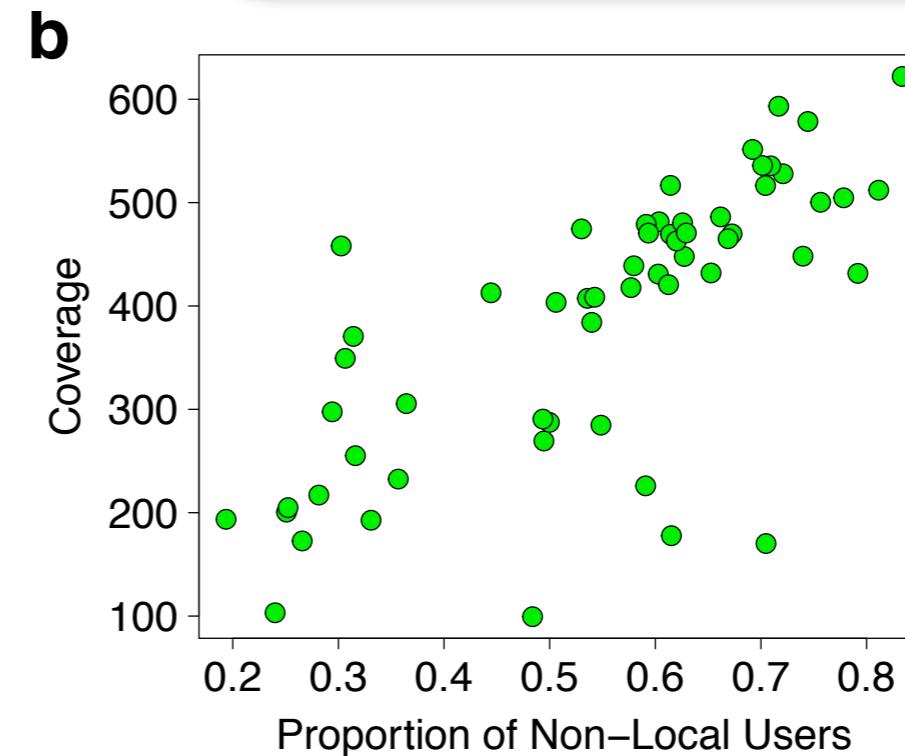
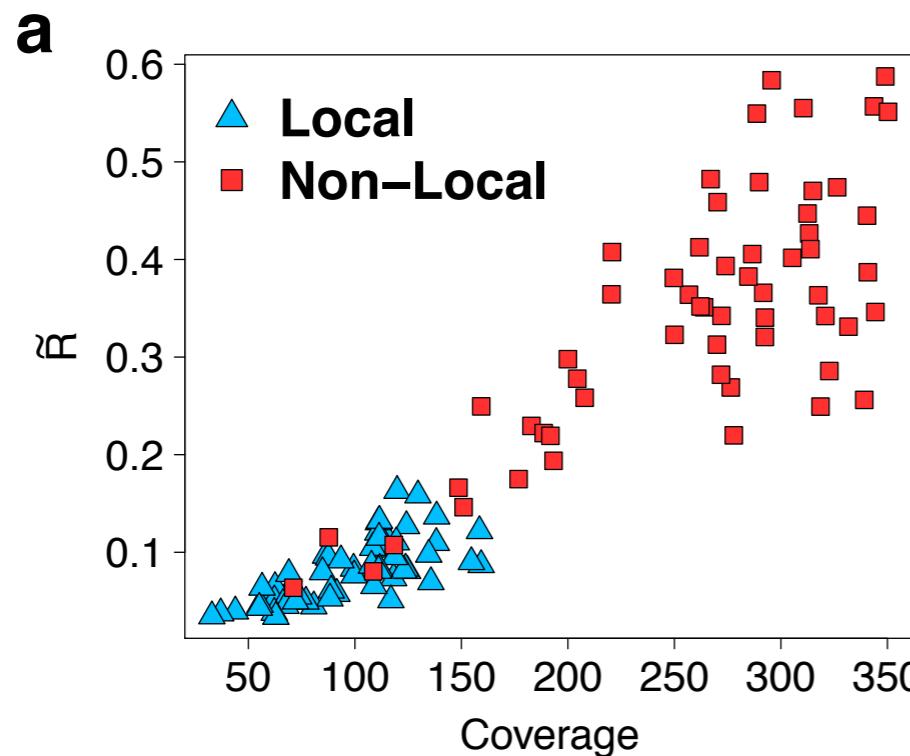
Residents

J. R. Soc. Interface 12, 20150473 (2015)
EPJ Data Science 5, 12 (2016)



Residents and Tourists

J. R. Soc. Interface 12, 20150473 (2015)
EPJ Data Science 5, 12 (2016)



City Communities

J. R. Soc. Interface 12, 20150473 (2015)
EPJ Data Science 5, 12 (2016)

