Potsdam Institute for Climate Impact Research

About growth and correlations in cities and social communities

Diego Rybski

H.D. Rozenfeld, J.S. Andrade, Jr.,
M. Batty, H.E. Stanley, H.A. Makse

S.V. Buldyrev, S. Havlin, F. Liljeros

X. Gabaix

9.12.2009 16:15 - 17:30
Martin-Luther-Universität Halle-Wittenberg
Kolloquium der Fachgruppe Theoretische Physik

the City College of New York
of the City University of New York

Outline

1. **Motivation**

2. **New laws of city growth**
   (a) City Clustering Algorithm (CCA)
   (b) City size (Zipf's law)
   (c) City growth (Gibrat's law)

3. **Human activity, long-term correlations, and Gibrat's law**
   (a) Online community data
   (b) Growth process
   (c) Temporal correlations
   (d) Missing link

4. **Summary, conclusions, and outlook**

Motivation

Scaling behaviour in the growth of companies

Michael H. R. Stanley*, Luís A. N. Amaral*, Sergey V. Buldyrev*, Shlomo Havlin*†, Heiko Leschhorn*, Philipp Maass*, Michael A. Salinger‡ & H. Eugene Stanley*

* Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA
† Minerva Center and Department of Physics, Bar-Ilan University, Raman Gan, Israel
‡ Department of Finance and Economics, School of Management, Boston University, Boston, Massachusetts 02215, USA

A SUCCESSFUL theory of corporate growth should include both the external and internal factors that affect the growth of a company[1–18]. Whereas traditional models emphasize production-related influences such as investment in physical capital and in research and development[18], recent models[10–20] recognize the equal importance of organizational infrastructure. Unfortunately, no exhaustive empirical account of the growth of companies exists by which these models can be tested. Here we present a broad, phenomenological picture of the dependence of growth on company size, derived from data for all publicly traded US manufacturing companies between 1975 and 1991. We find that, for firms with similar sales, the distribution of annual (logarithmic) growth rates has an exponential form; the spread in the distribution of rates decreases with increasing sales as a power law over seven orders of magnitude. A model wherein the probability of a company's growth depends on its past as well as present sales accounts for the former observation. As the latter observation applies to companies that manufacture products of all kinds, organizational structures common to all firms might well be stronger determinants of growth than production-related

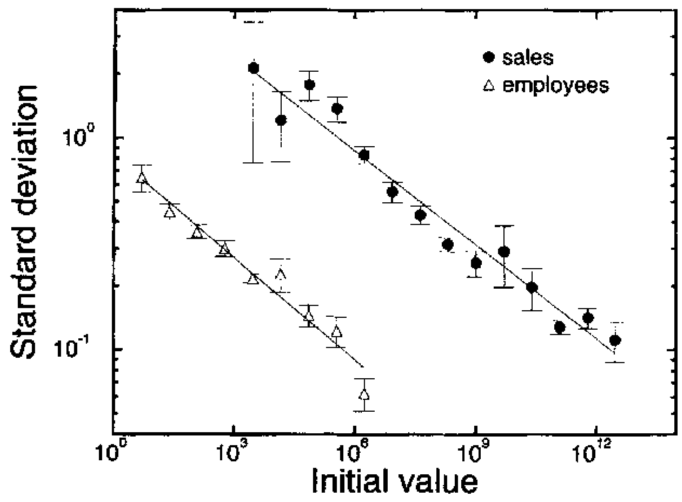


FIG. 2 Standard deviation of the one-year growth rates of the sales (circles) and of the one-year growth rates of the number of employees (triangles) as a function of the initial values. The solid lines are least-square fits to the data with slopes $\beta = 0.15 \pm 0.03$ for the sales and $\beta = 0.16 \pm 0.03$ for the number of employees. We also show error bars of one standard deviation about each data point. These error bars appear asymmetric as the ordinate is a log scale.

NATURE · VOL 379 · 29 FEBRUARY 1996

Motivation

logarithmic growth rate:

$$r(S_0) \equiv \ln R(S_0) = \ln(S_1/S_0)$$

conditional average growth rate:

$$\langle r(S_0) \rangle \sim S_0^{-\alpha}$$

conditional standard deviation:

$$\sigma(S_0) \sim S_0^{-\beta}$$

Motivation

	$\sigma(S_0) \sim S_0^{-\beta}$	
firms	0.15 - 0.18	Stanley et al. 1996
countries GDP	0.15+/-0.03	Canning et al. 1998
research costs at universities	0.25	Plerou et al. 1999
voluntary organizations	0.19	Liljeros et al. 2003
scientific output	0.28 - 0.40	Matia et al. 2005
religious activity	0.16	Picoli et al. 2008
...		

which are the processes behind this
non-trivial emergent scaling?

Motivation: Gibrat's law

1. What is the origin of broad distributions?
   (such as of firm sizes)

2. R. Gibrat proposed 1931 a
   multiplicative process (law of proportional effect)

3. Unit size is iteratively multiplied with
   random variable

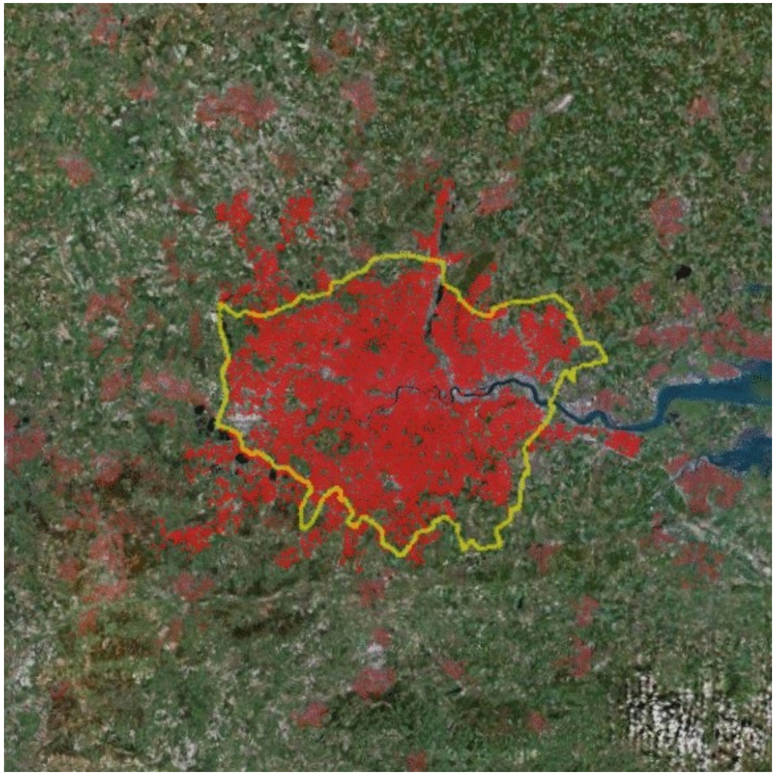
4. Central Limit Theorem: summing up logs
   results in normal distribution,
   and log-normal distribution of original size

5. Assumption: iid random variable
   (growth rate independent of unit size, aka Gibrat's law)

New laws of city growth

Measurement problem: what is a city?

Examples:
- New York City / Jersey City
- Boston / Cambridge
- Greater London



"administrative" or "legal" definitions may lead to spurious statistical properties

"Metropolitan Statistical Areas" (MSA) built by the US Census are indeed agglomerations (use socio-economic factors)

Only top cities (time-consuming, hardly transferable to other countries)

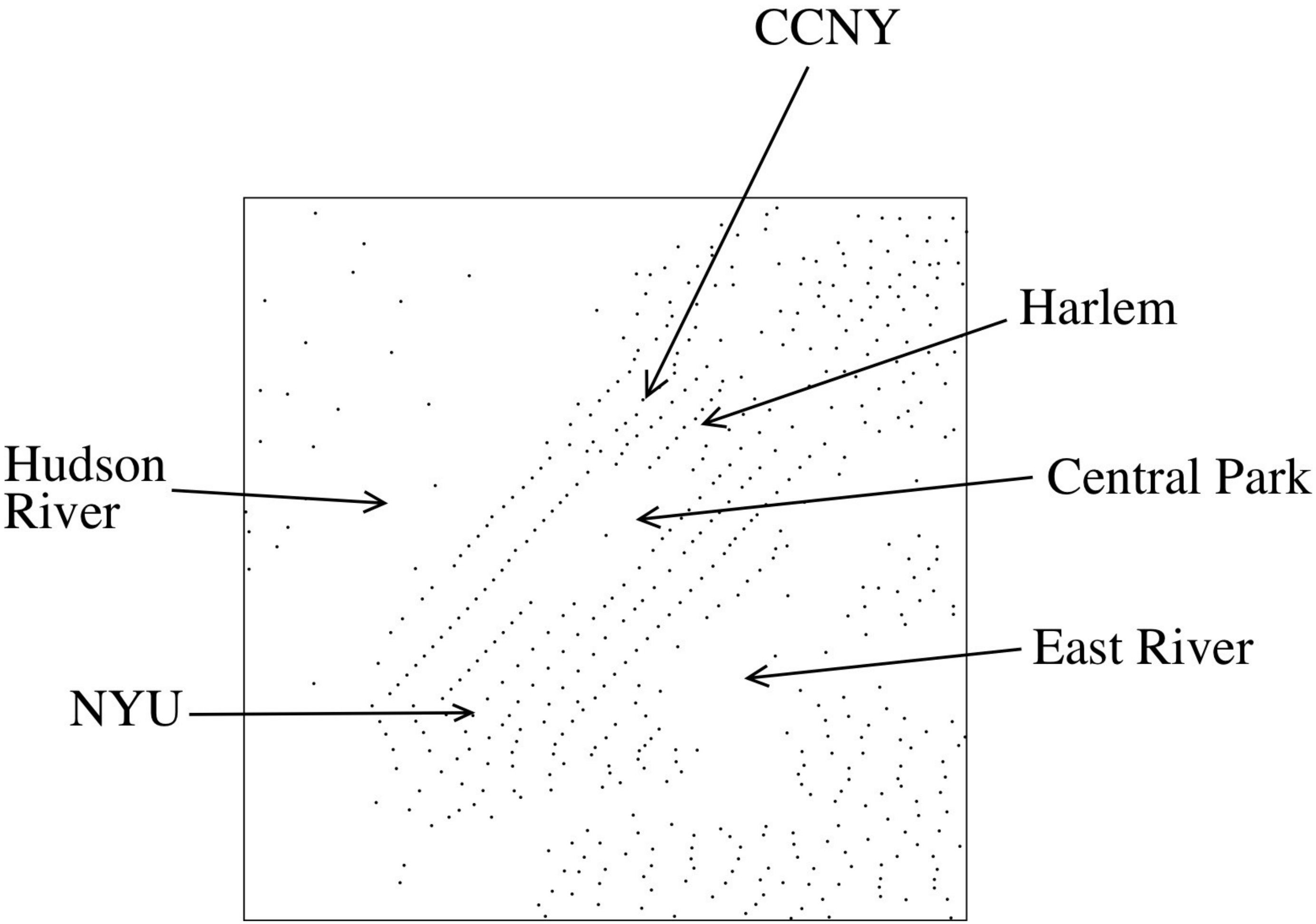
Population data

Two data sets

1. USA:
- FIPS-code (Federal Information Processing Standard)
- partition the area into units with approx. 1500-8000 inhabitants (average ~4000)
- total population: ~300 million

2. GB:
- gridded population (cell size 200m)
- total population: ~59 million

Example: USA, NYC, Manhattan



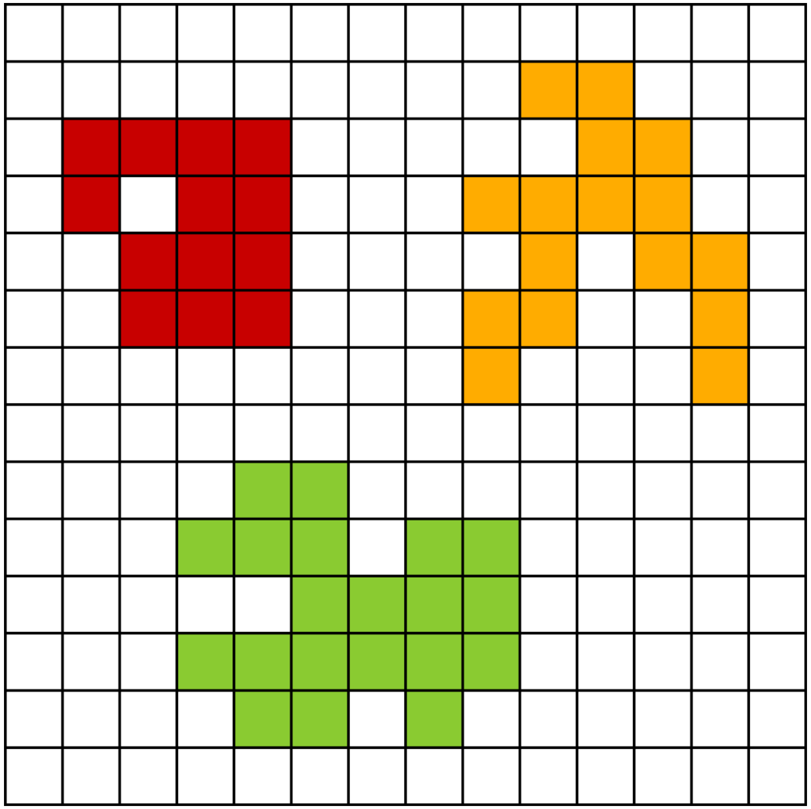
City Clustering Algorithm (CCA)

We define a new way to construct cities:

- unbiased
- automated
- fast
- can be easily used in any country
- based only on location of population
- allows studying cities at different levels of observation

City Clustering Algorithm

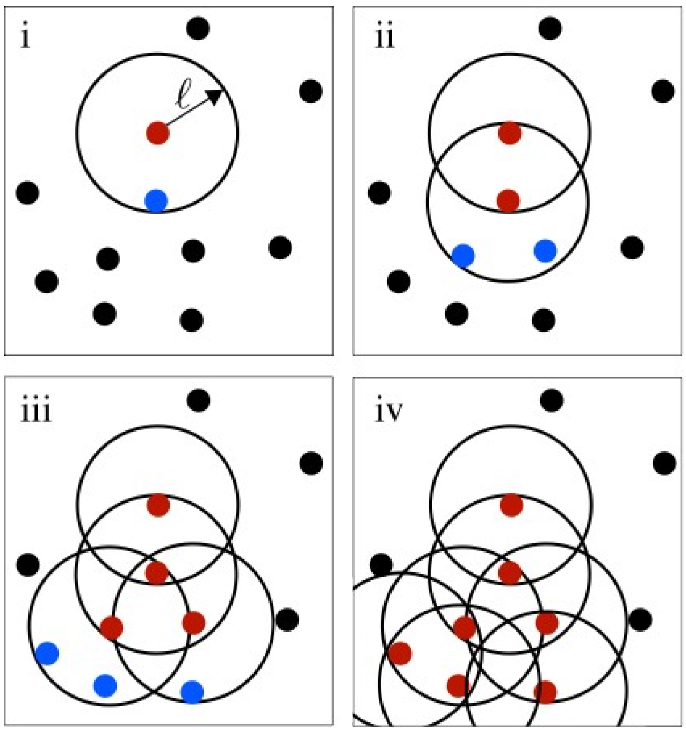
Idea of CCA:



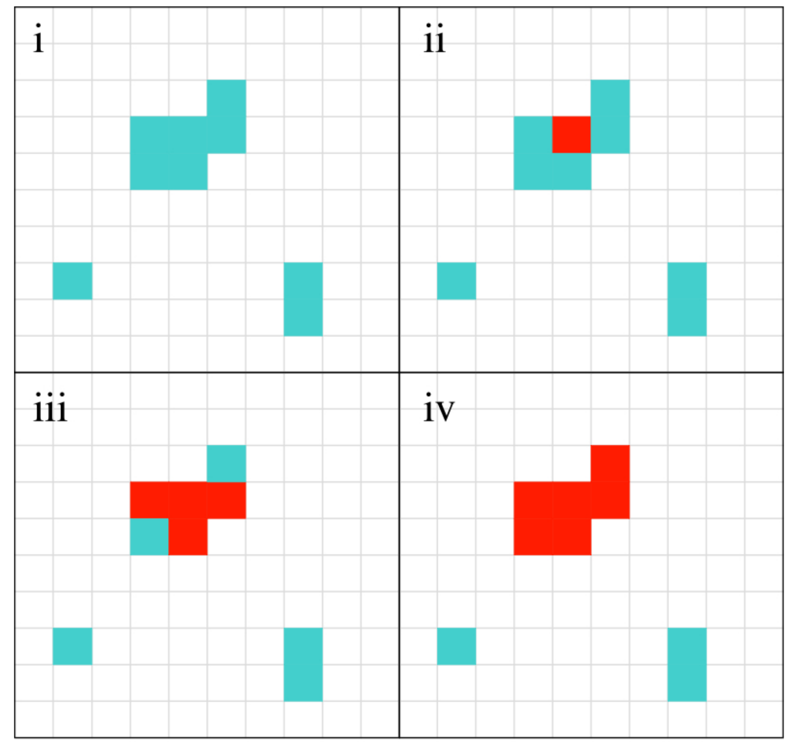
City as maximal size cluster of connected subunits

City Clustering Algorithm

USA



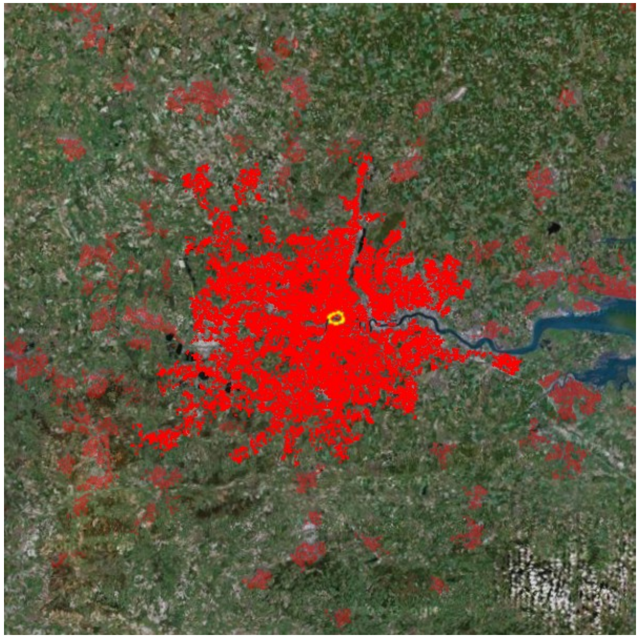
GB



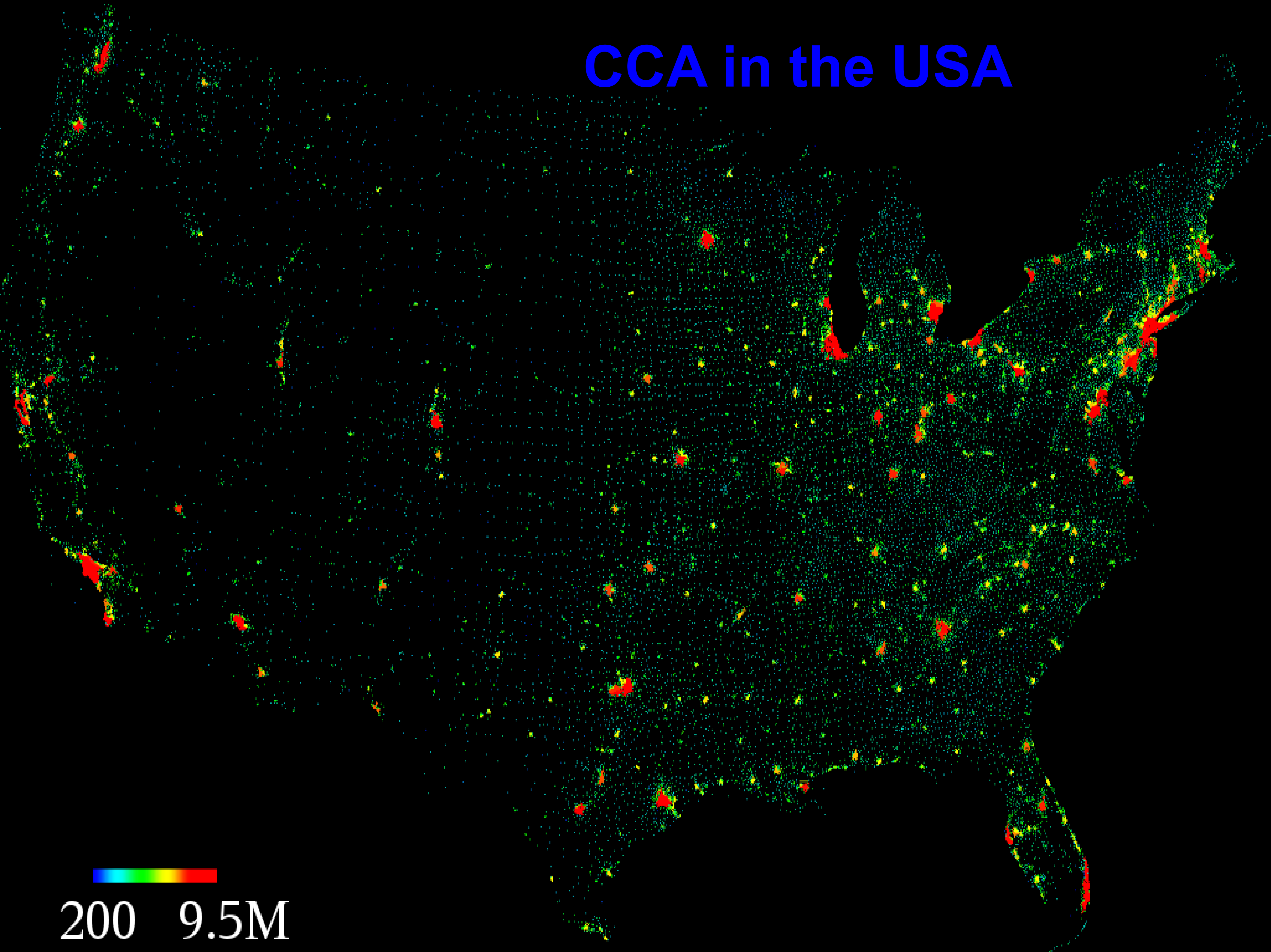
2 parameters: coarse-graining level ℓ

threshold density D_* (set to 0)

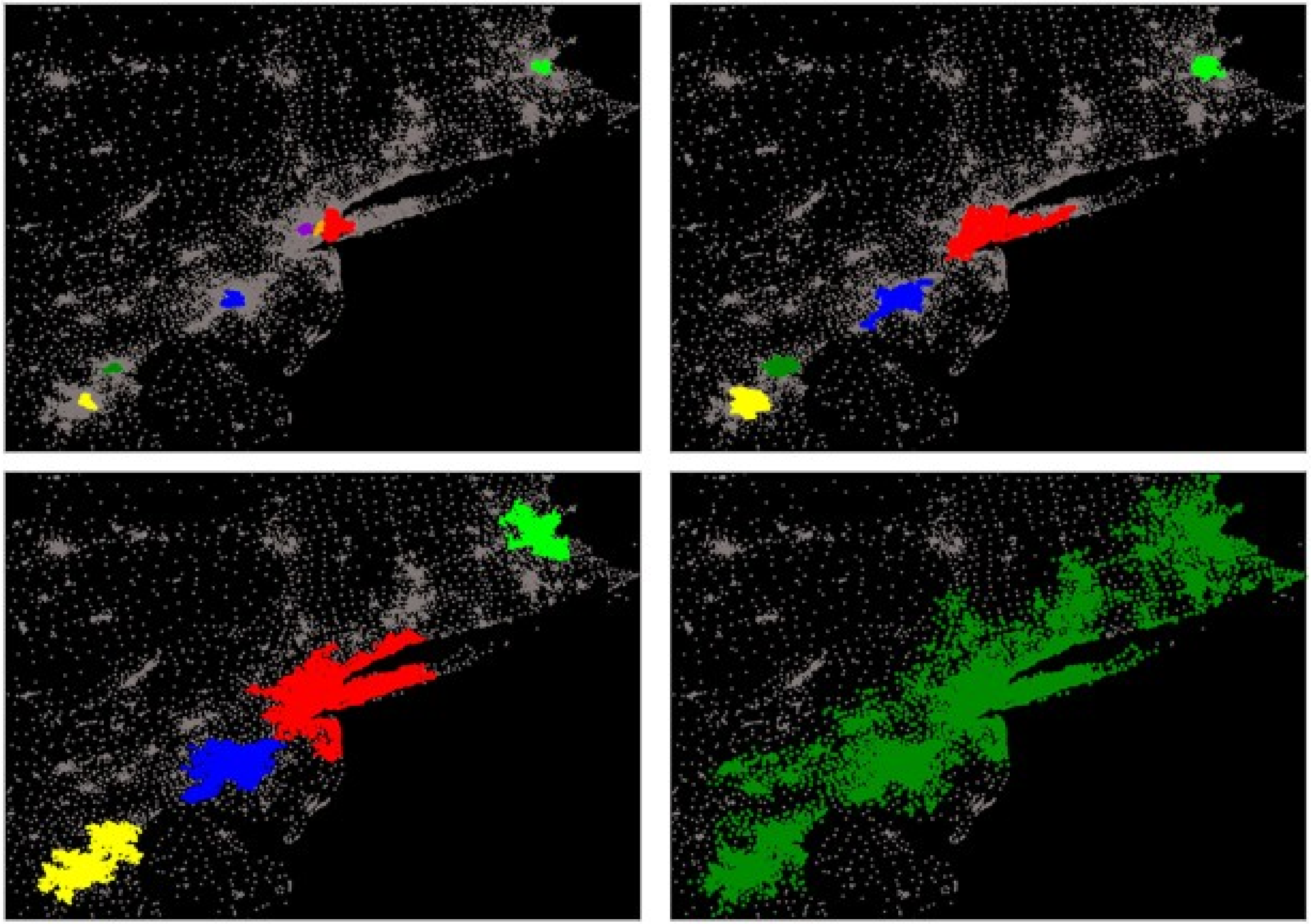
CCA in Great Britain



CCA applied
to Greater London


CCA in the USA
200
9.5M

CCA in the northeastern USA



Zipf's law

$$P(S) \sim S^{-\zeta-1}, \quad \zeta = 1$$

The distribution of sizes follows
a power-law with $\zeta = 1$

Zipf's law has been documented for
words, firms, size of exports,
and many more

Does the city size distribution obey Zipf’s law?

Zipf's law

Understanding the origin of this regularity is an ongoing task.

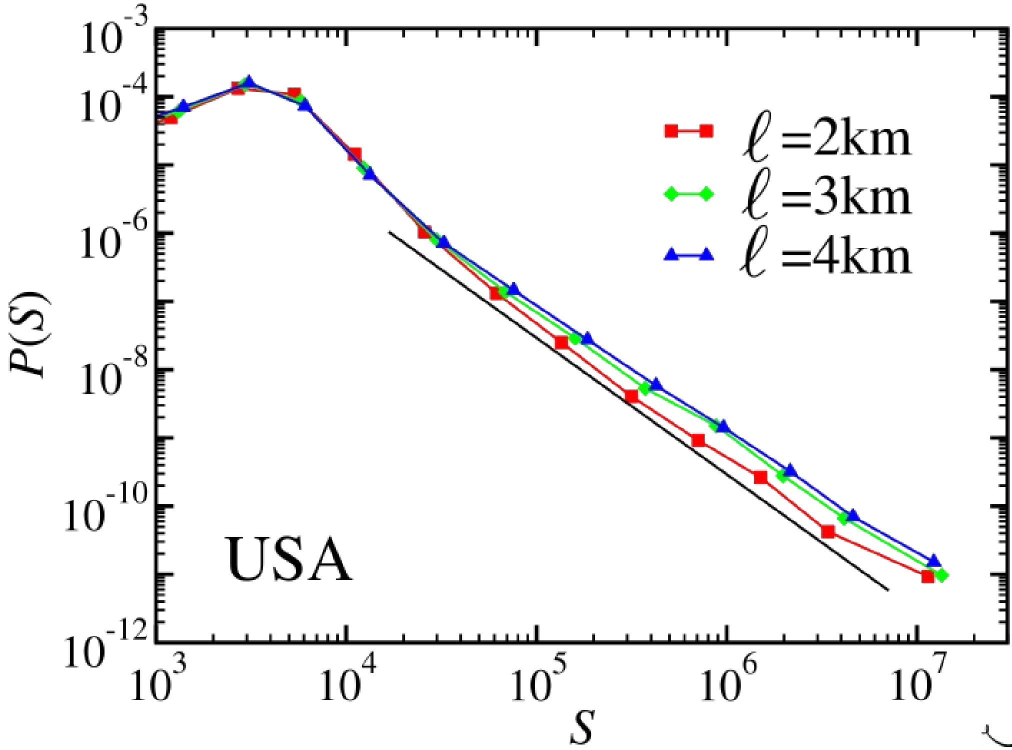
Typically, studies use MSAs for the top 200 cities, i.e. Eeckhout ('07)

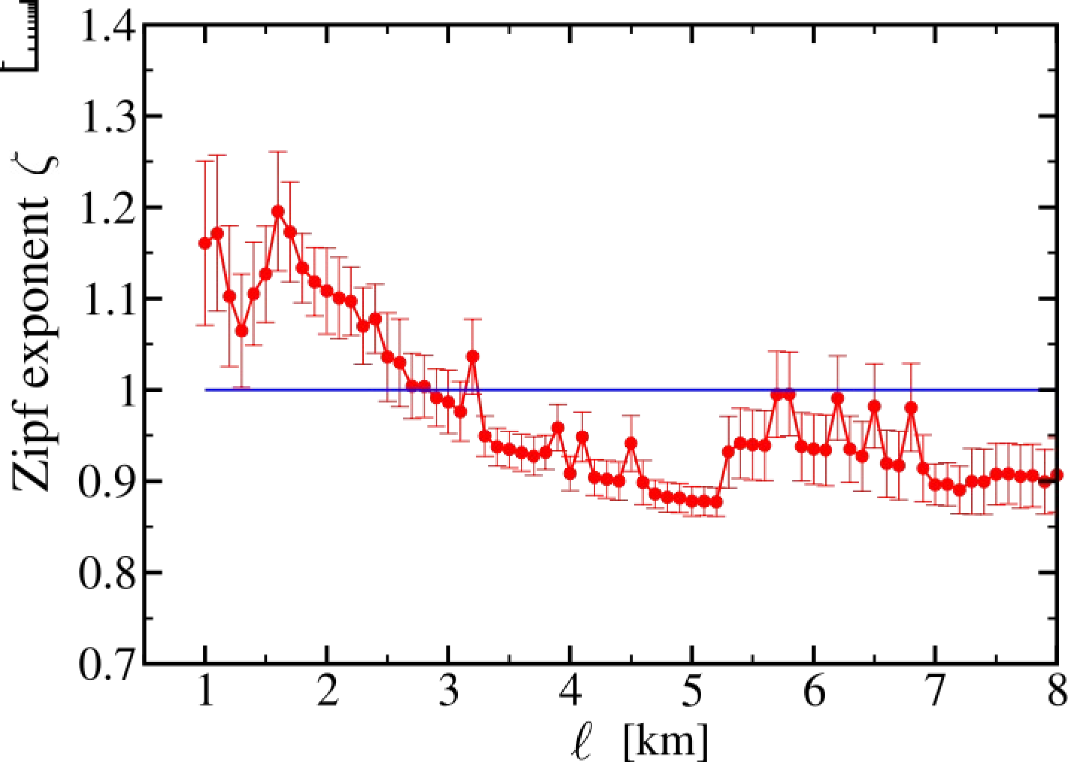
Eeckhout ('07)

- Uses data on all administrative cities
- Finds a very good log-normal fit

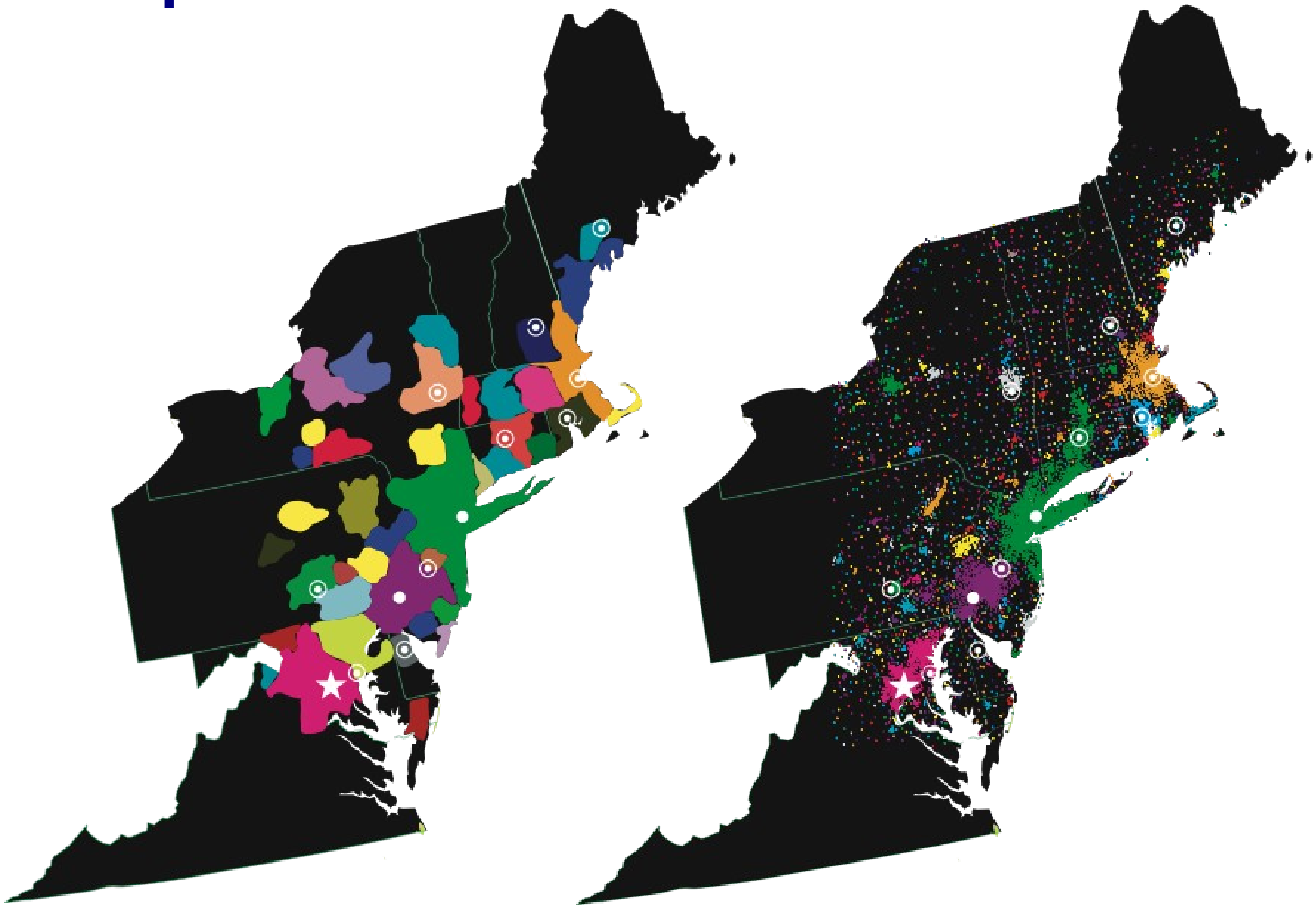
Distribution of city size using CCA?

Zipf's law for the USA

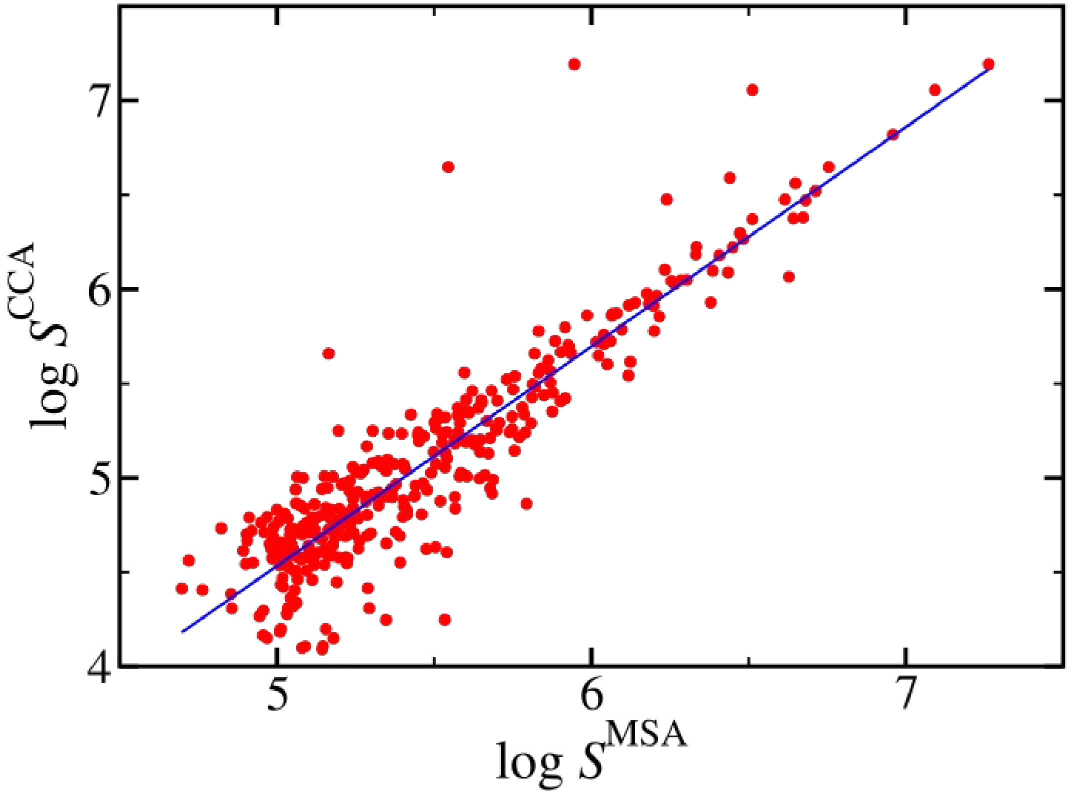


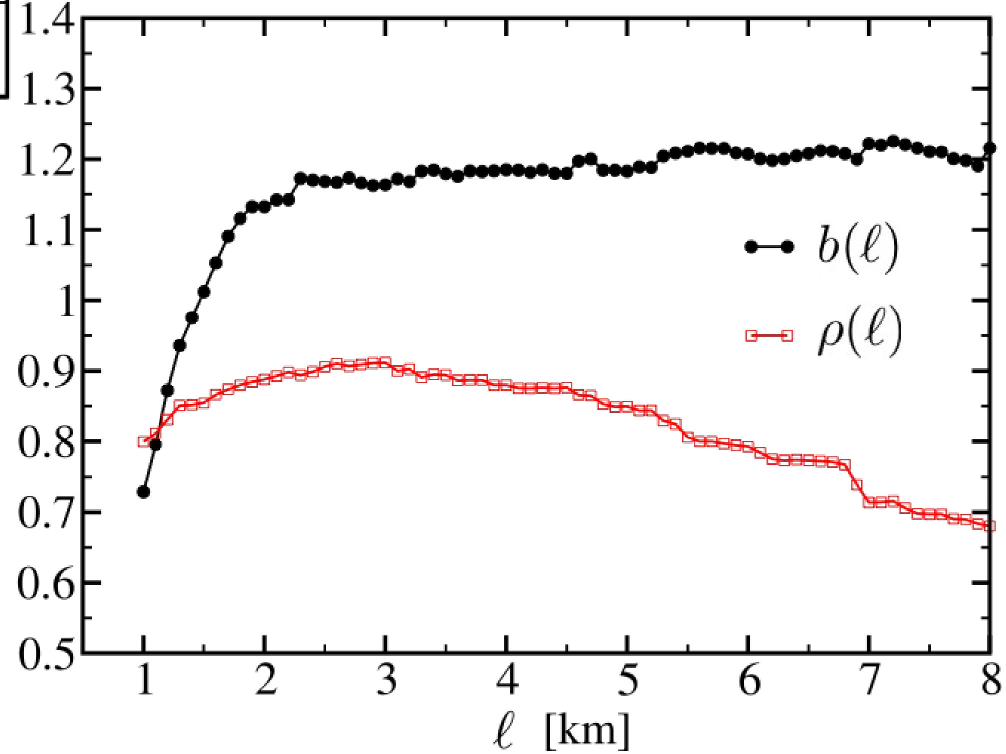


Comparison with MSA: Northeastern USA

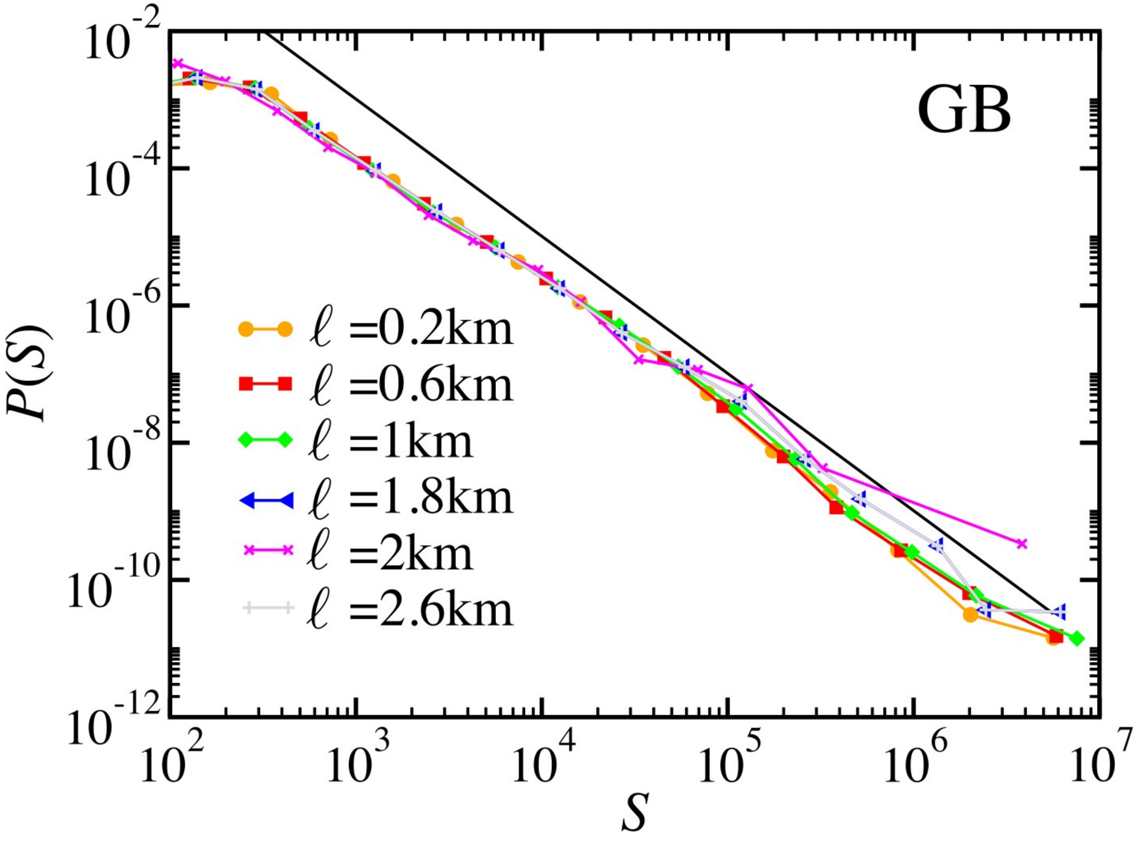


Correlations between MSA and CCA

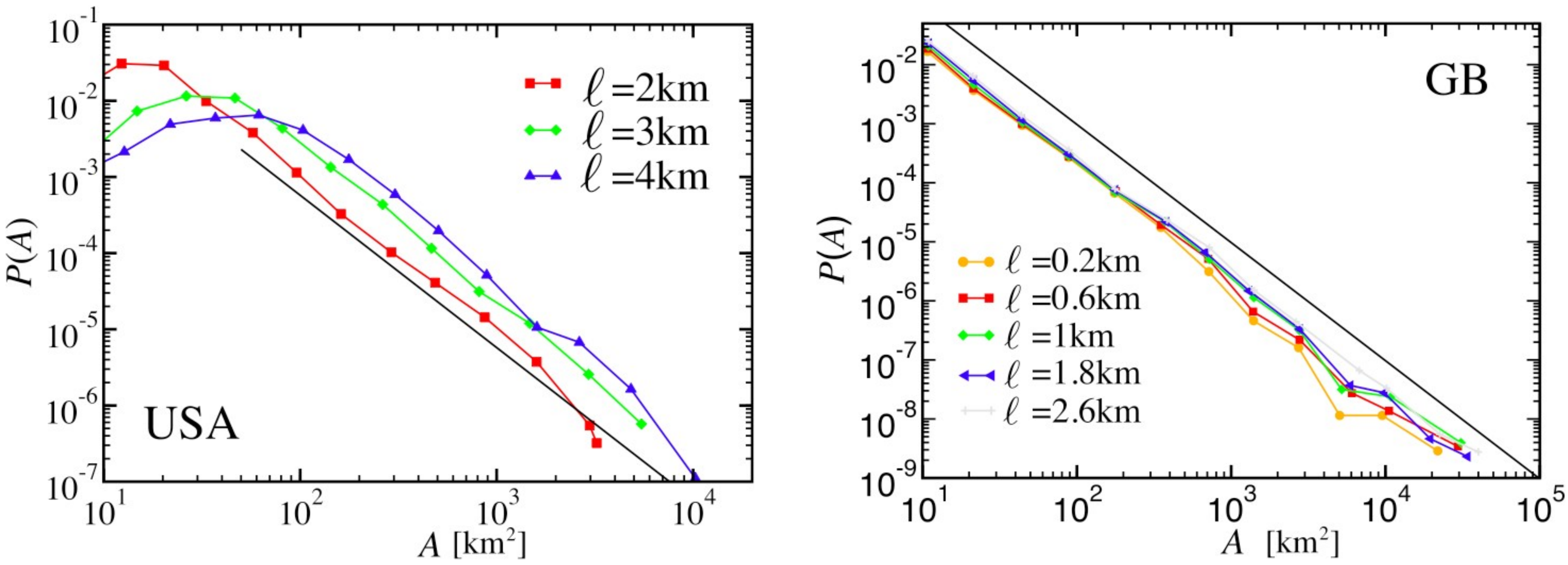




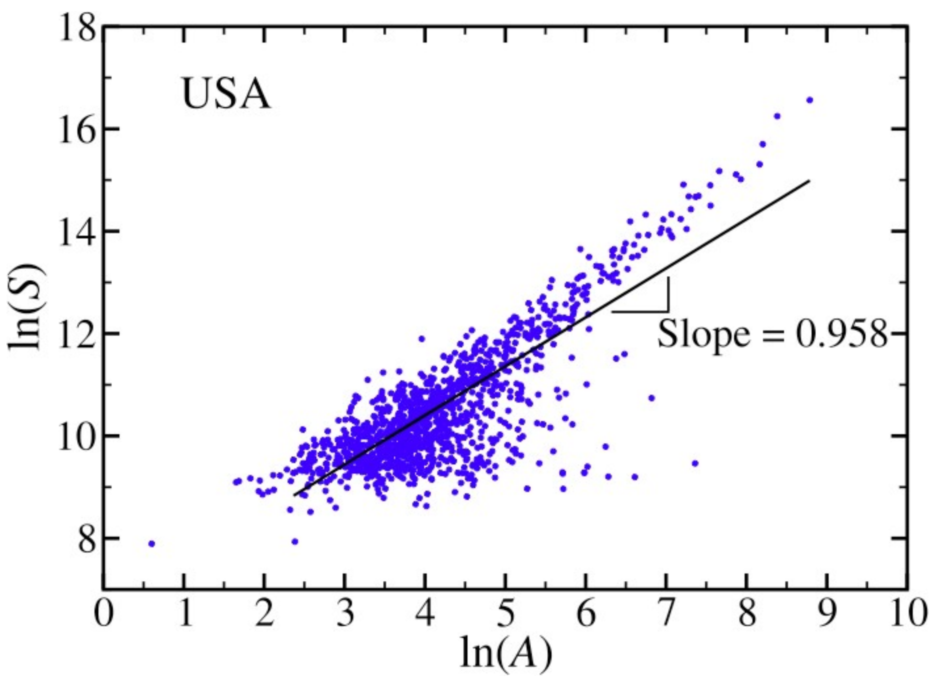
Zipf's law for GB

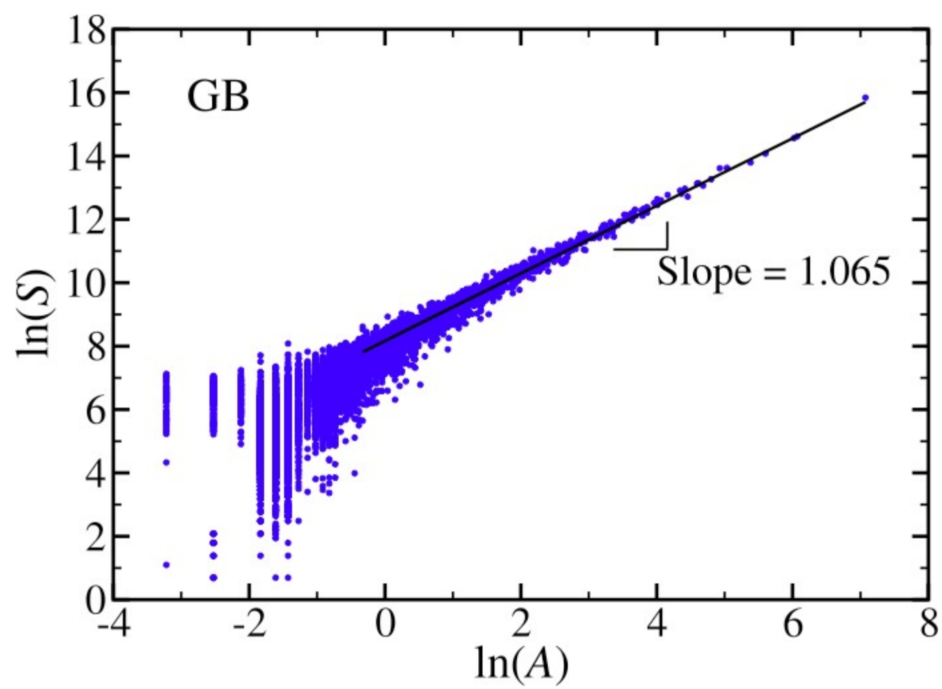


Zipf's law for areas



Population vs area





City *size* take home message

- Zipf's law holds pretty well for size above 12000 (USA) or 5000 (GB) inhabitants
- Zipf's law for areas
- Population is proportional to area
- Density is roughly independent of city size
- How about other countries?

City growth

S_0 Population of a city at time 0.

S_1 Population of a city at time 1.

$$S_1 = R(S_0)S_0$$ → R growth factor

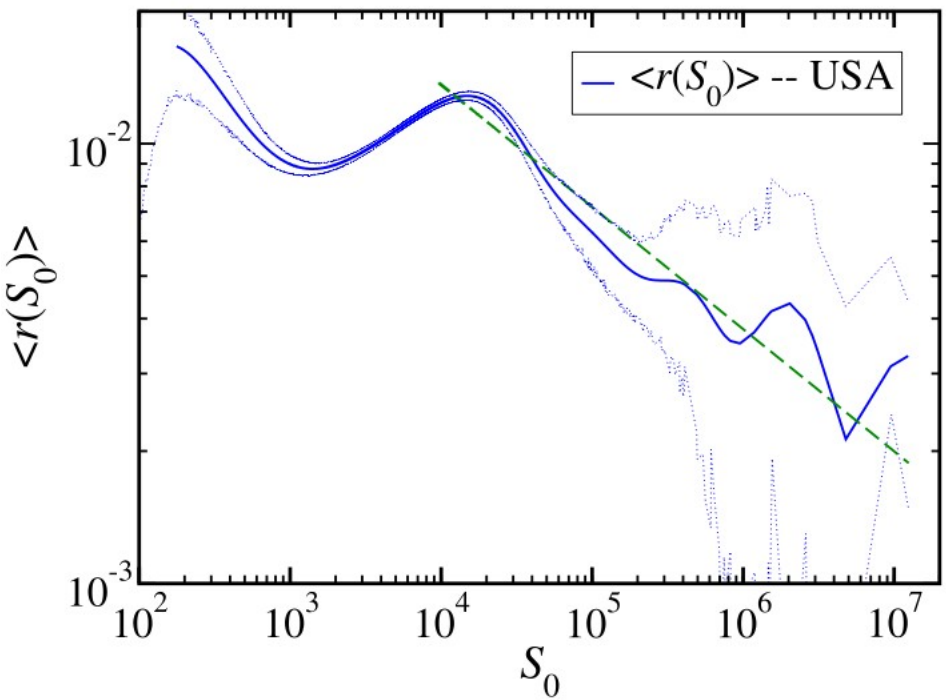
$$r(S_0) \equiv \ln R(S_0) = \ln(S_1/S_0)$$ → r growth rate

$$\langle r(S_0) \rangle \sim S_0^{-\alpha}$$

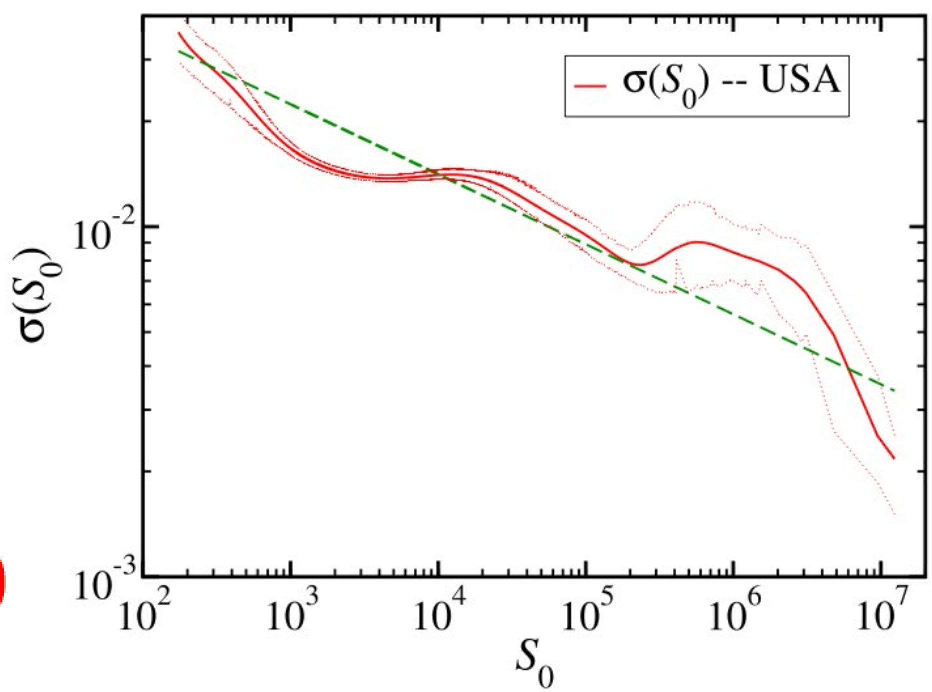
$$\sigma(S_0) = \sqrt{\langle r(S_0)^2 \rangle - \langle r(S_0) \rangle^2}$$

$$\sigma(S_0) \sim S_0^{-\beta}$$

City growth in the USA (1990-2000)



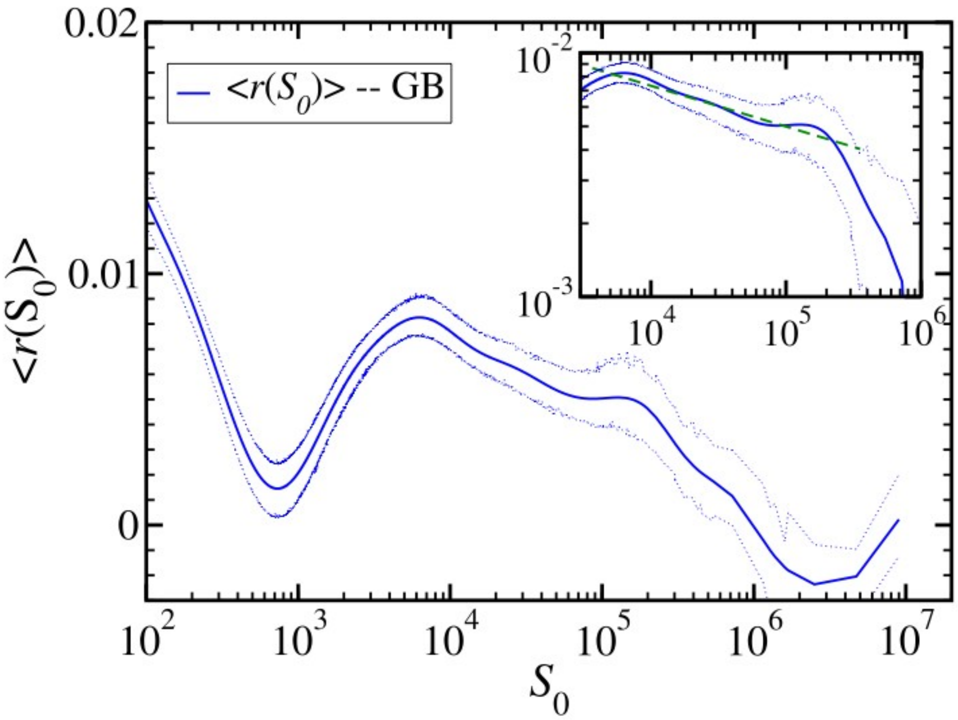
$$\langle r(S_0)\rangle \sim S_0^{-\alpha}\,, \quad \alpha = 0.28$$



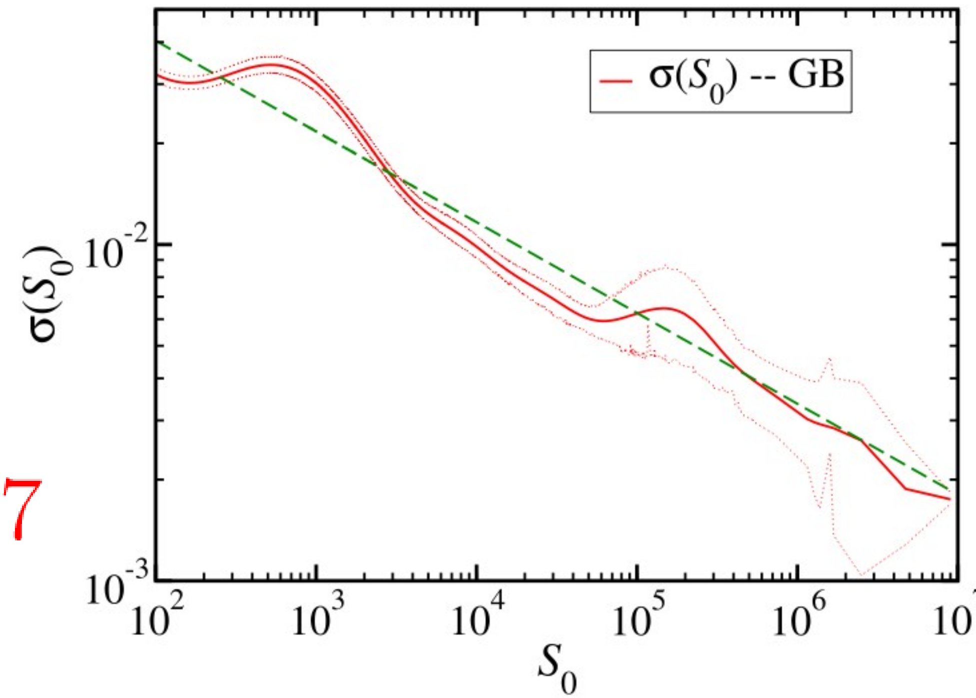
$$\sigma(S_0) \sim S_0^{-\beta}\,, \quad \beta = 0.20$$

Are not in agreement with Gibrat's Law (stating that average growth rate and standard deviation are constant)

City growth in the GB (1981-1991)

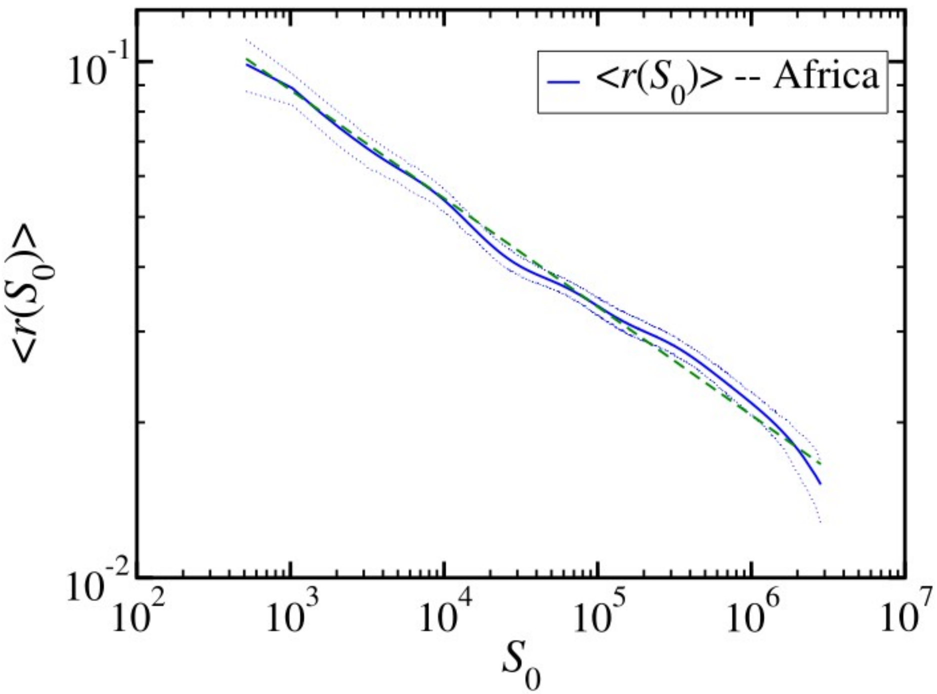


$$\langle r(S_0)\rangle \sim S_0^{-\alpha}\,, \quad \alpha = 0.17$$

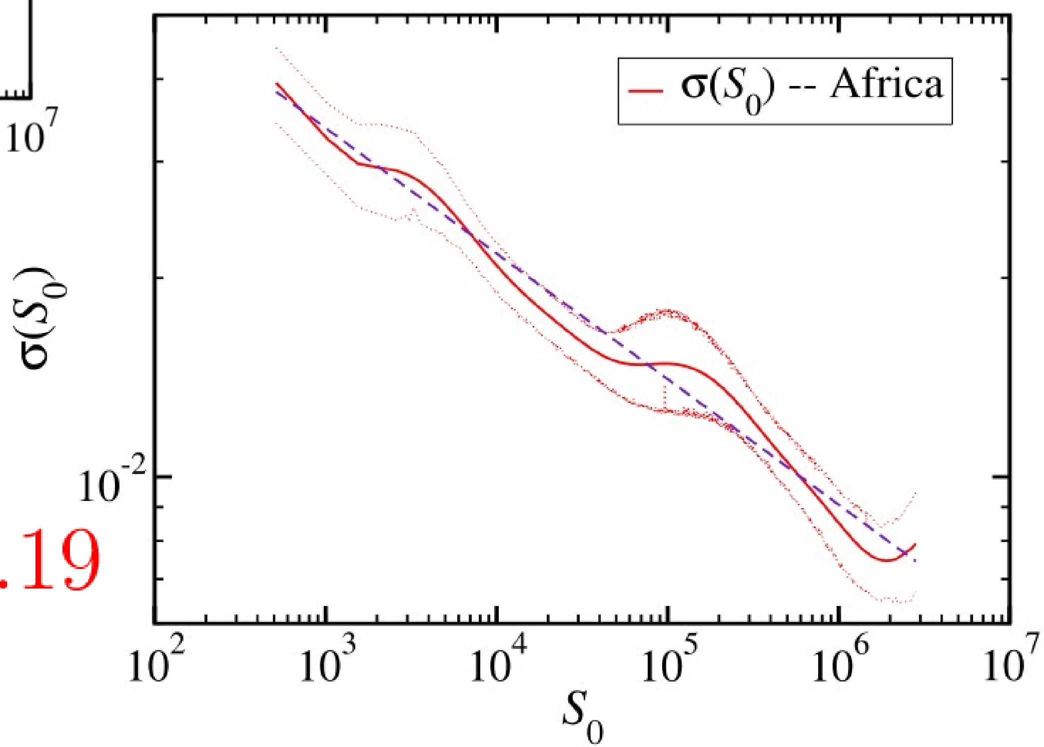


$$\sigma(S_0) \sim S_0^{-\beta}\,, \quad \beta = 0.27$$

City growth in Africa (1960-1990)



$$\langle r(S_0)\rangle \sim S_0^{-\alpha}\,, \quad \alpha = 0.21$$



$$\sigma(S_0) \sim S_0^{-\beta}\,, \quad \beta = 0.19$$

Correlations

δ_j= population growth of cell j

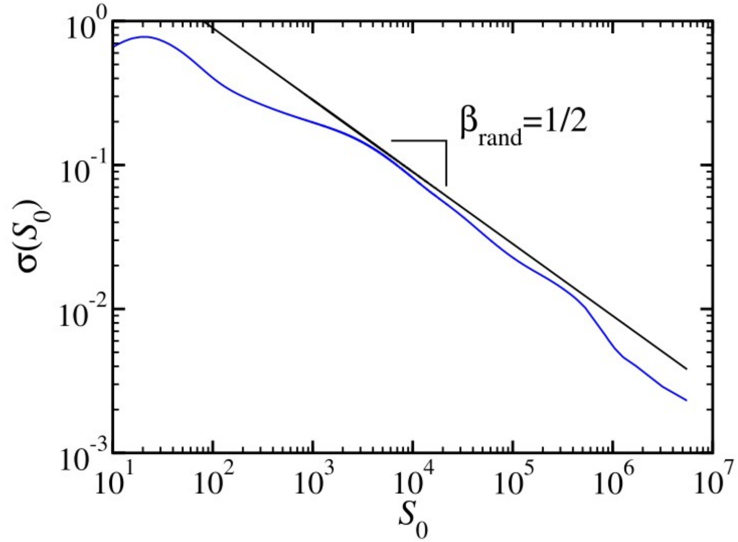
$$S_1 = S_0 + \sum_{j=1}^{N_i} \delta_j$$

$$\langle(\delta_j - \bar{\delta})(\delta_k - \bar{\delta})\rangle \sim \frac{\Delta^2}{|\vec{x}_j - \vec{x}_k|^\gamma}$$

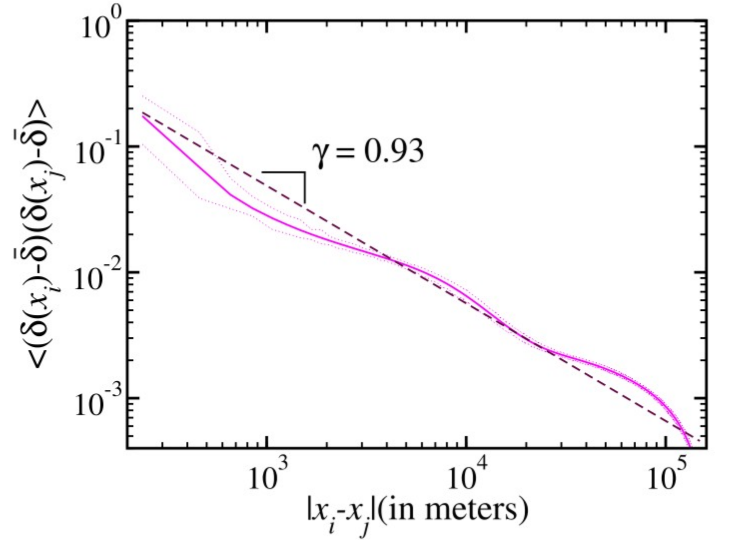
$$\beta = \gamma/4$$

When $\gamma = 2 \longrightarrow \beta = 1/2$

$$\gamma_{GB} = 0.93$$



GB



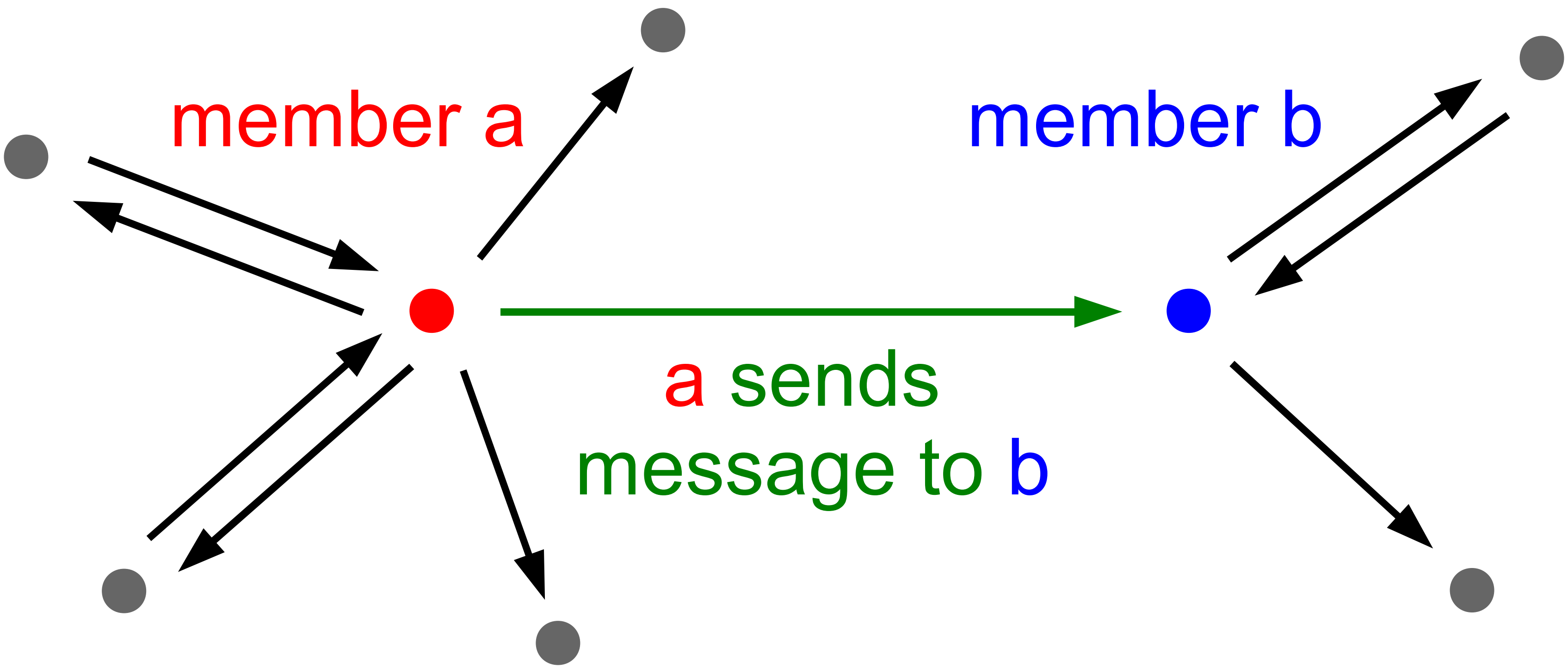
City *growth* take home message

- CCA identifies cities based only on geographical features
- Scale-invariant growth mechanisms at different geographical scales (violation of Gibrat's Law)
- Power-law standard deviation is due to long-range spatial correlations in the growth
- How about other countries?

Human activity,
long-term correlations,
and Gibrat's law

Online community

members sending messages



either following an existing link $m_{\mathrm{a}} \longrightarrow m_{\mathrm{a}} + 1$

or creating a new one $k_{\mathrm{a}}^{\mathrm{out}} \longrightarrow k_{\mathrm{a}}^{\mathrm{out}} + 1$

=> **growth process**

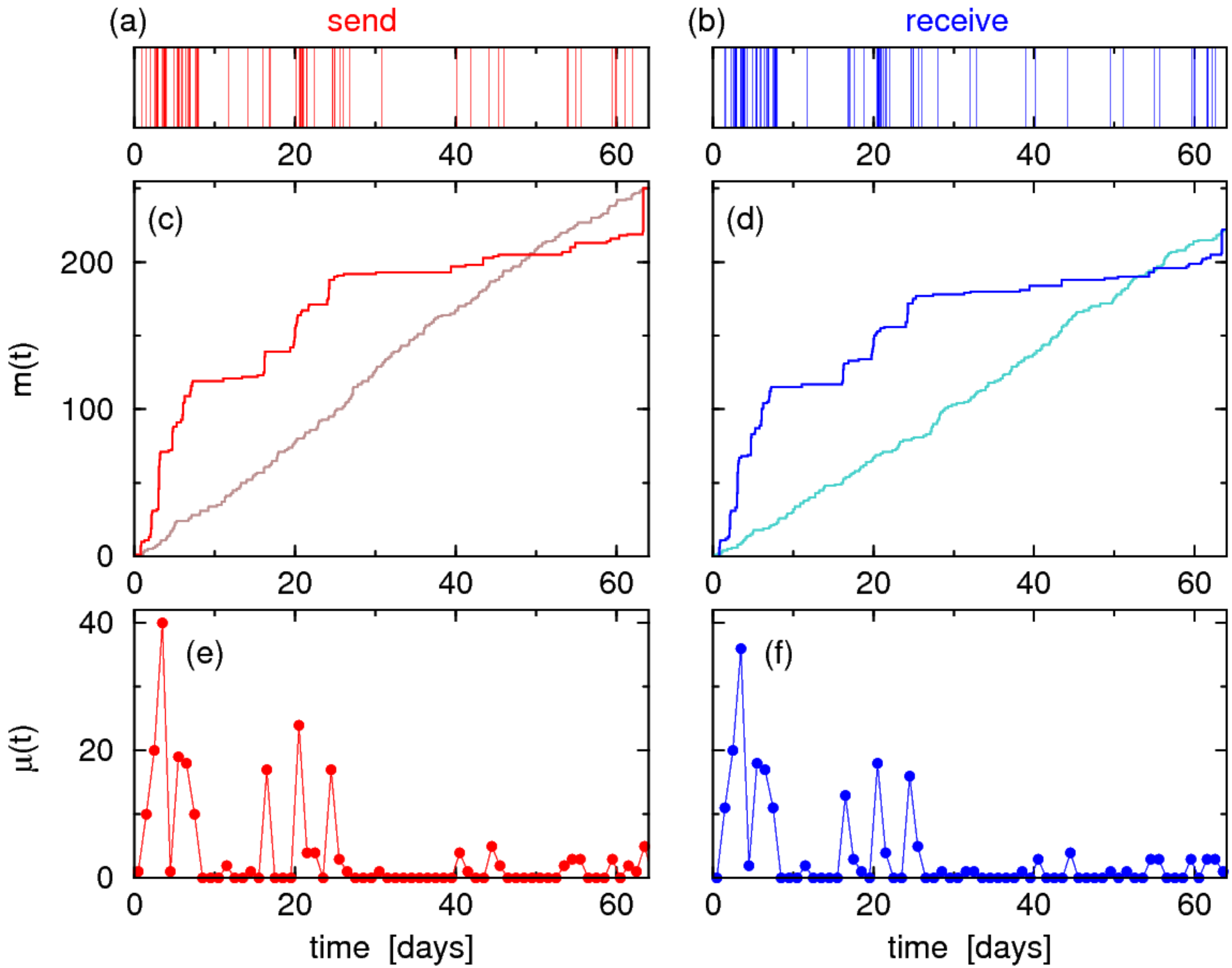
Online community data

online community 1 (OC1):
- 80,000 members
- 12.5 million messages
- 63 days

online community 2 (OC2):
- 30,000 members
- 500,000 messages
- 492 days

both are dating-communities
also used for social interaction in general
completely anonymous

Typical activity (OC1)



Growth process

for each member:

cumulative number of messages $m(t)$

logarithmic growth rate $r = \ln \frac{m_1}{m_0}$

between two time-steps t_0, t_1

two quantities:

conditional average growth $\langle r(m_0) \rangle = \langle r | m_0 \rangle$

cond. standard deviation $\sigma(m_0) = \sigma(r | m_0)$

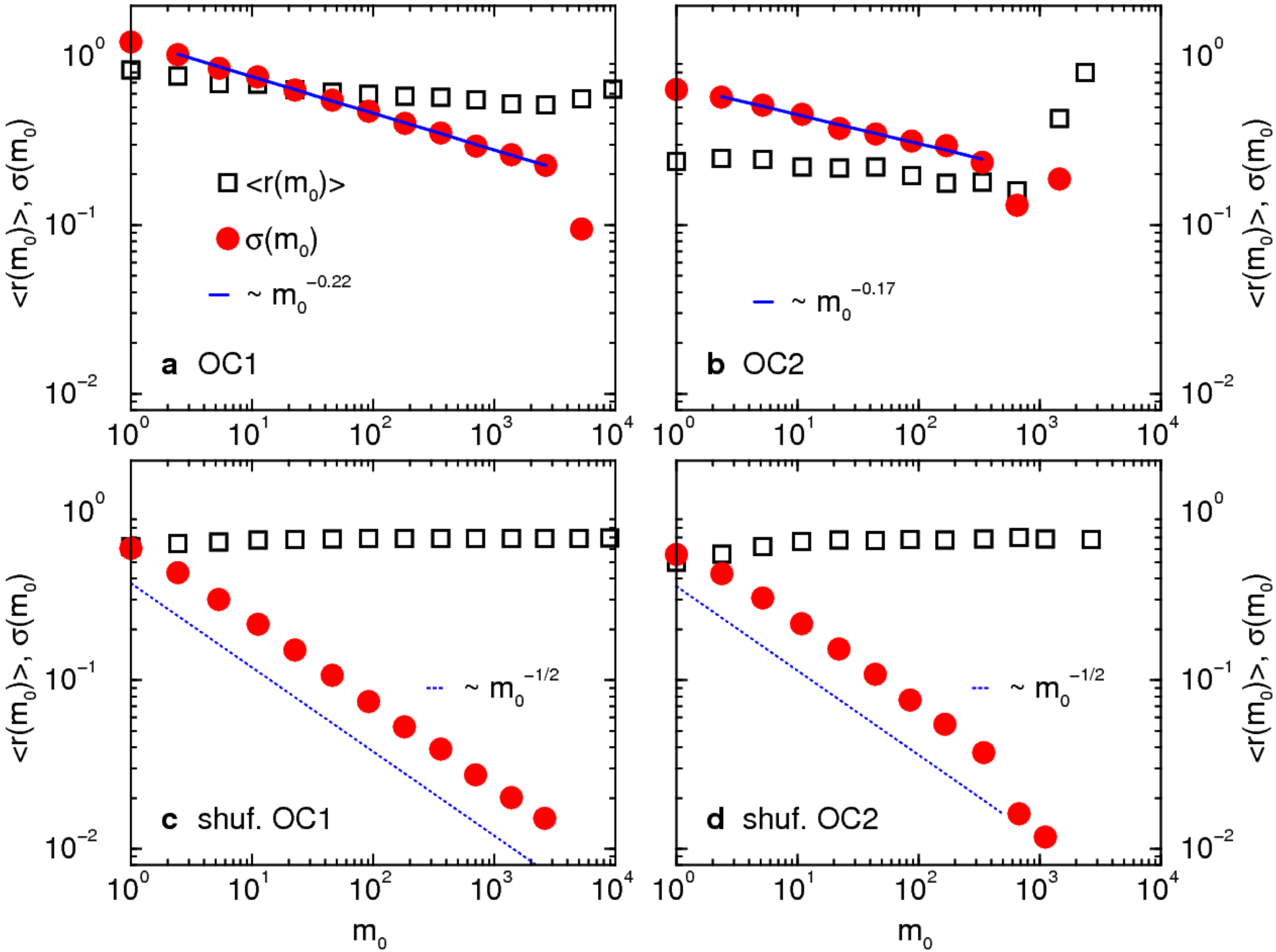
Analogy to other data, such as city growth

(1) The members of a community represent a population similar to the population of a country.

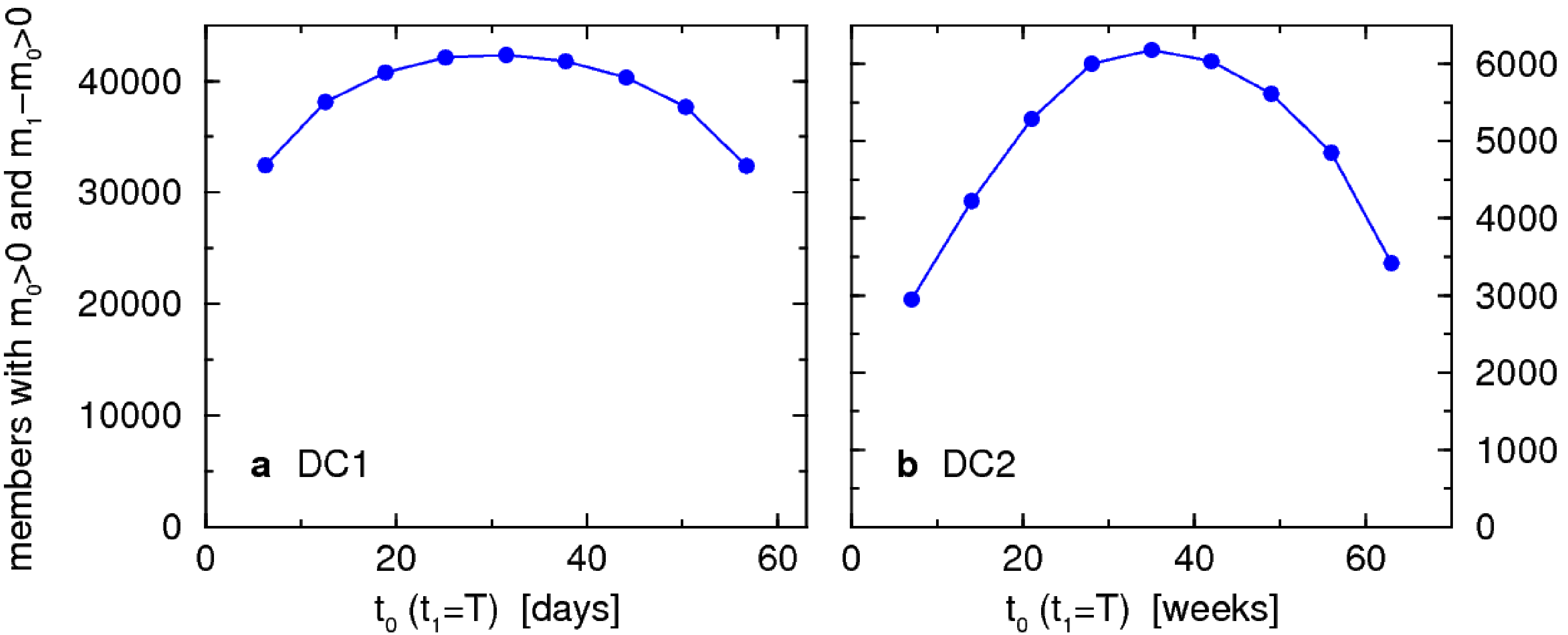
(2) The number of members fluctuates and typically grows analogous to the number of cities of a country.

(3) The activity or number of links of individuals fluctuates and grows similar to the size of cities.

Growth process: results



Optimal times



Growth process: results

$$\sigma(m_0) \sim m_0^{-\beta}$$

OC1: $\beta_{\mathrm{OC1}} = 0.22 \pm 0.01$
OC2: $\beta_{\mathrm{OC1}} = 0.17 \pm 0.03$
shuffled: $\beta_{\mathrm{rnd}} = 1/2$

Gibrat's law of proportionate growth

multiplicative process
to explain broad distributions (log-normal)

involves assumption: $\langle r(m_0) \rangle = \mathrm{const.}$
$\sigma(m_0) = \mathrm{const.}$

=> $\beta_{\mathrm{G}} = 0$

Temporal correlations

- shuffling destroys temporal correlations, leading to $\beta_{\mathrm{rnd}} = 1/2$

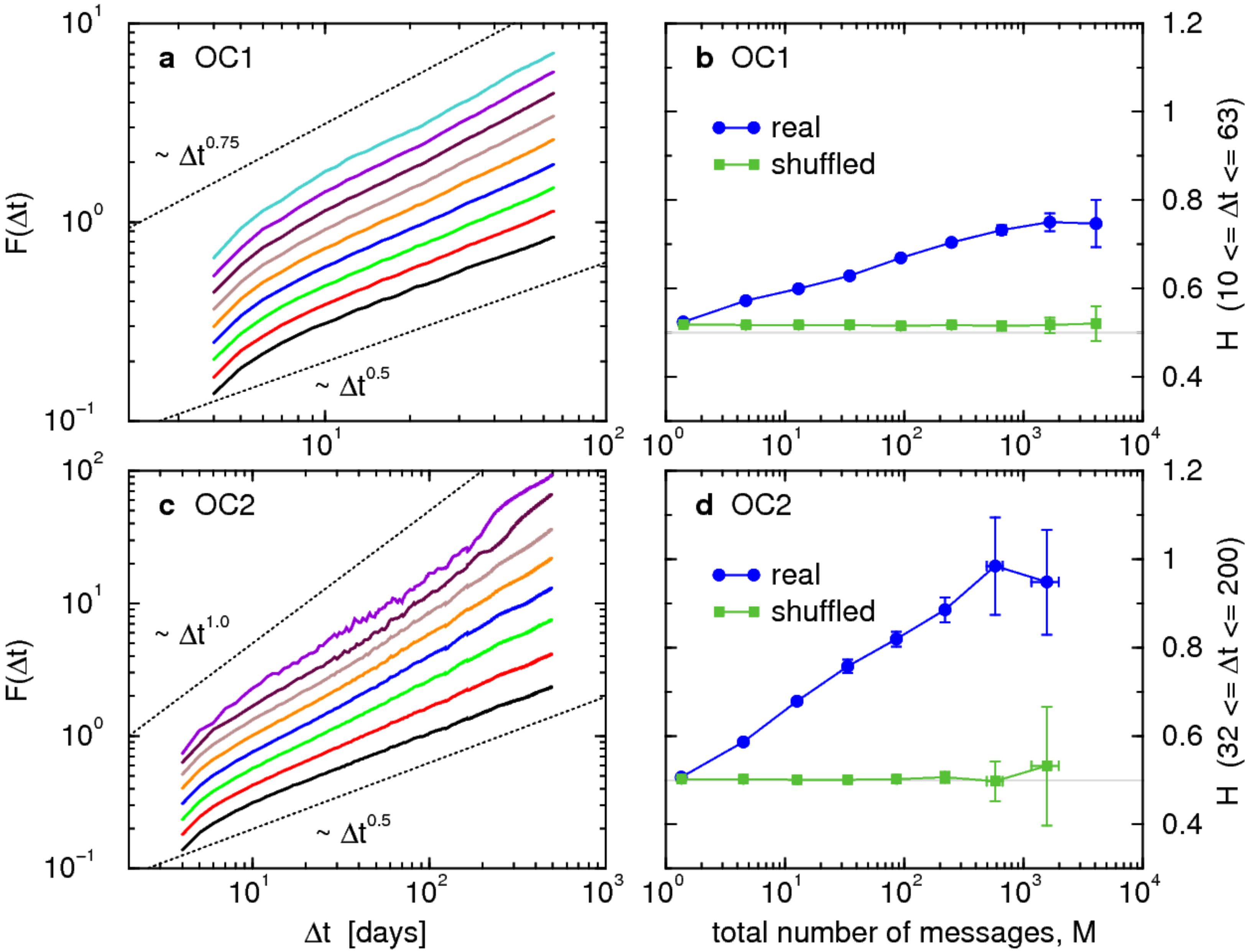
- this suggests $\beta \approx 0.2$ might be due to temporal correlations

- we use Detrended Fluctuation Analysis (DFA) to quantify long-term correlations in the activity (messages per day): $\mu(t)$

fluctuation function: $F(\Delta t) \sim (\Delta t)^H$

$1/2 < H < 1$ => ltc

Temporal correlations: results



Missing link

derivation leads to:

$$\beta = 1 - H$$

accordingly:

$$\beta \approx 0.2 \quad \Rightarrow \quad H \approx 0.8 \qquad \text{OCs}$$

$$\beta_{\mathrm{rnd}} = 1/2 \quad \Rightarrow \quad H_{\mathrm{rnd}} = 1/2 \qquad \text{shuffled}$$

$$\beta_{\mathrm{G}} = 0 \quad \Rightarrow \quad H_{\mathrm{G}} = 1 \qquad \text{Gibrat's law}$$

Derivation

$$r = \ln \frac{m_1}{m_0} \approx \frac{\Delta m}{m_0}$$

$$\Delta m = \sum \mu(t) = m_1 - m_0$$

$$r \approx \frac{1}{m_0} \sum \mu(t)$$

$$\left[r(m_0) - \langle r(m_0) \rangle\right]^2 = \frac{1}{m_0^2} \left(\sum \left(\mu(t) - \langle \mu(t) \rangle\right) \right)^2$$

$$\langle \left[r(m_0) - \langle r(m_0) \rangle\right]^2 \rangle \approx \frac{1}{m_0^2} \sum \sum \sigma_\mu^2 C(j-i)$$

Derivation ...

$$\langle [r(m_0) - \langle r(m_0)\rangle]^2 \rangle \approx \frac{1}{m_0^2} \sum \sum \sigma_\mu^2 C(j-i)$$

$$C(\Delta t) = \frac{1}{\sigma_\mu^2 (T - \Delta t)} \sum^{T-\Delta t} \mu(t)\mu(t+\Delta t) \quad \langle \mu \rangle = 0$$

$$C(\Delta T) \sim (\Delta T)^{-\gamma}$$

$$\langle [r(m_0) - \langle r(m_0)\rangle]^2 \rangle \sim \frac{1}{m_0^2} \sigma_\mu^2 (\Delta t)^{2-\gamma}$$

$$\Delta t = x\, t_0 \quad m_0 \sim t_0$$

Derivation

$$\sigma(m_0) \sim \sigma_\mu m_0^{-\gamma/2}$$

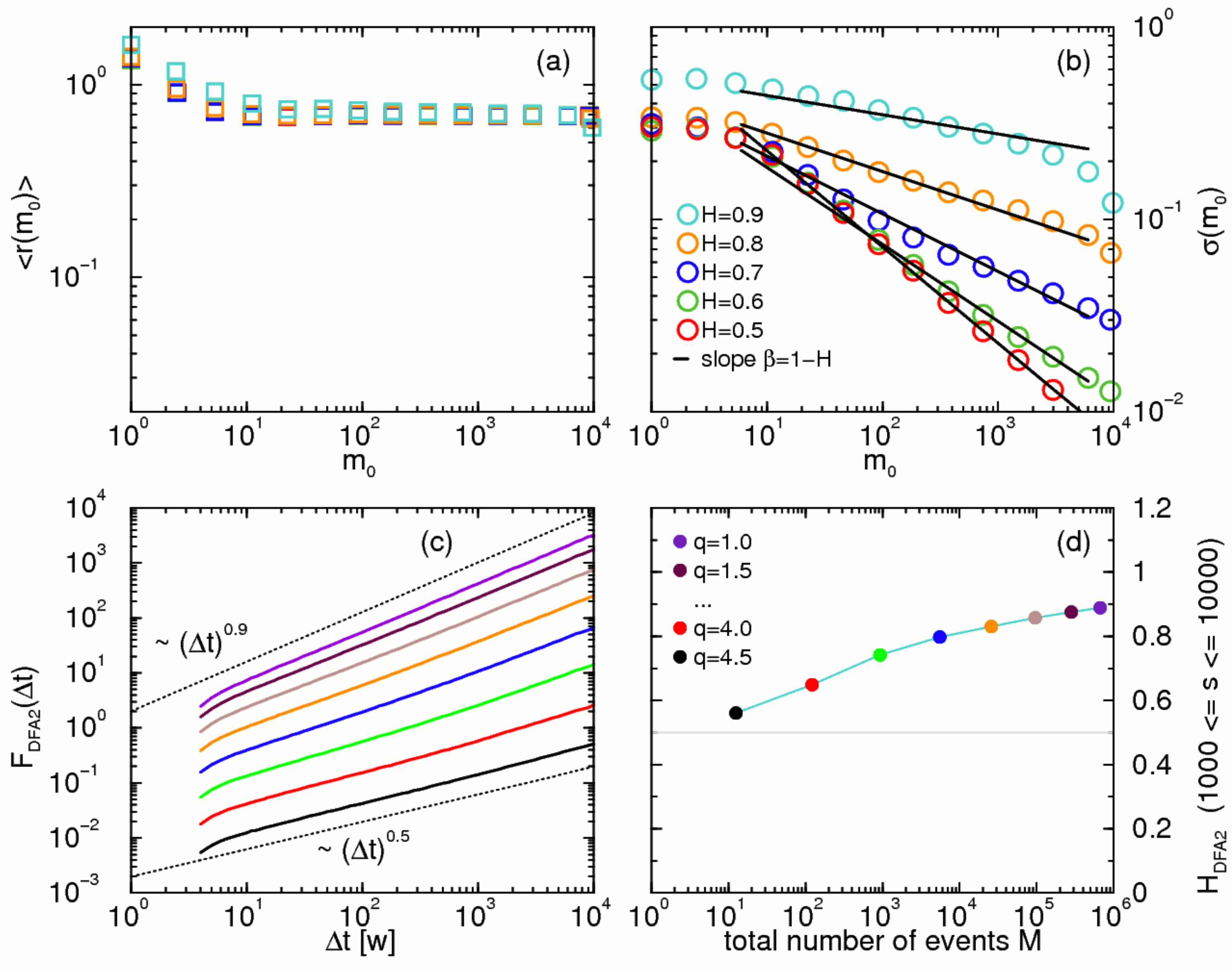
$$\sigma(m_0) \sim m_0^{-\beta}$$

$$\beta = \gamma/2$$

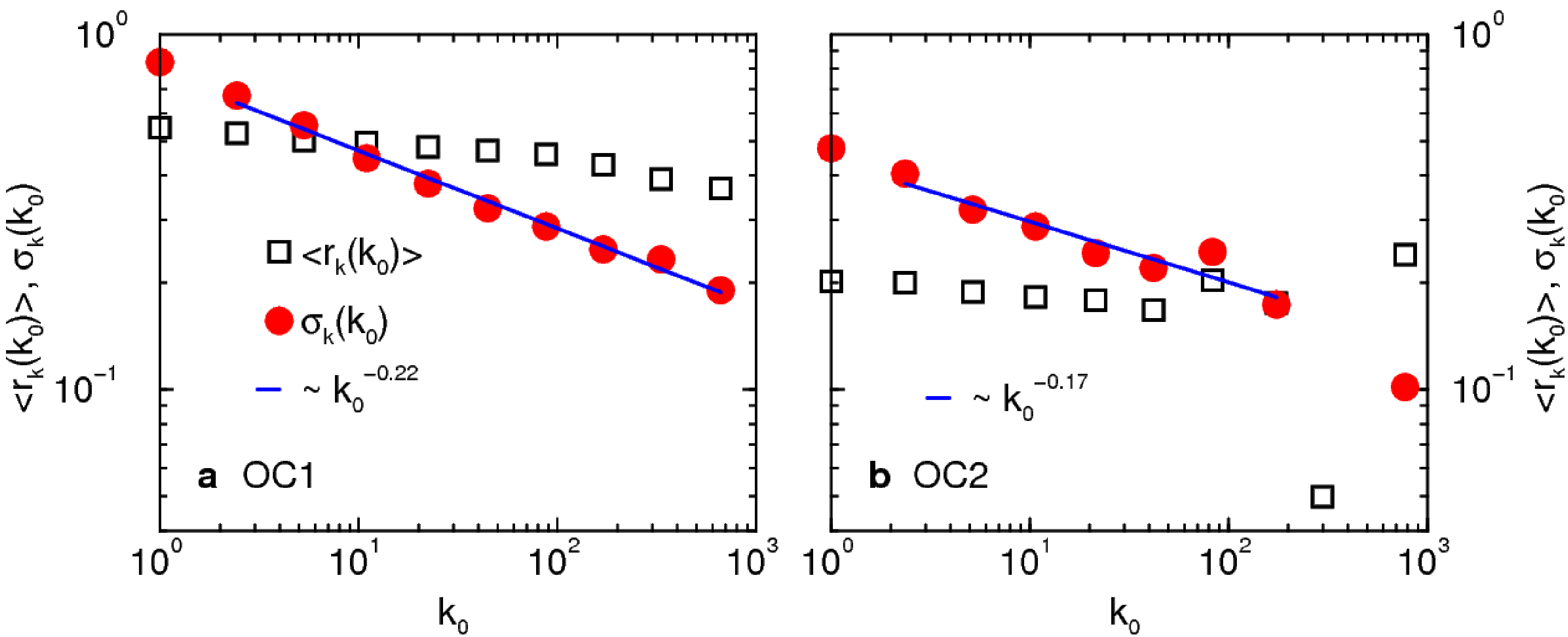
$$\gamma = 2 - 2H$$

$$\boxed{\beta = 1 - H}$$

Simulations

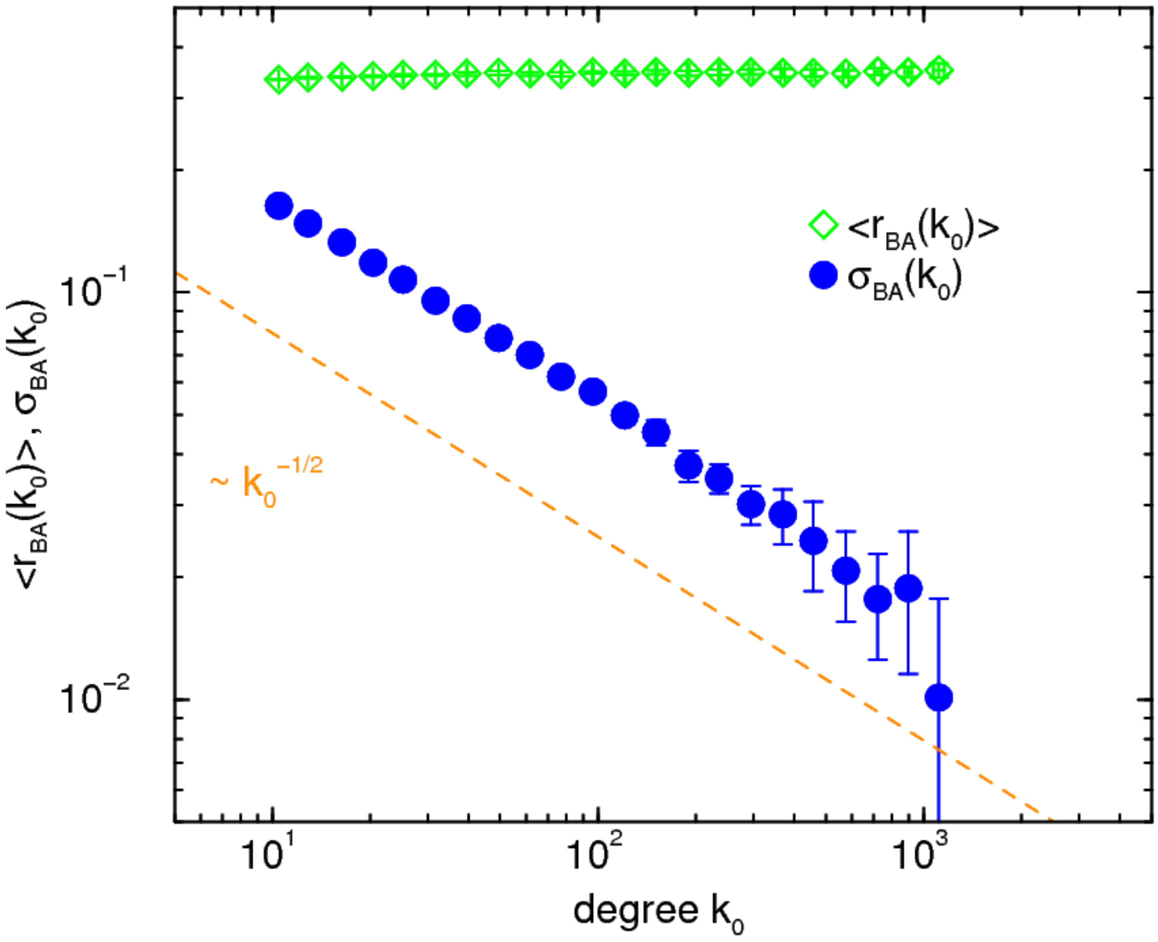


Growth process: out-degree



see also: Maillart T, et al., arXiv 0807.0014, 2008

Growth process: preferential attachment



see also: Barabasi AL and Albert R, Science 286, 1999

Human activity take home message

- scaling in growth of number of messages or out-degree implies that active members are better predictable than less active ones

- human activity sending messages is long-term correlated

- scaling in growth is due to long-term correlations

$$\sigma(m_0) \sim m_0^{-\beta}$$

=> this may also be the case for other data

Summary, conclusions, and outlook

1. Growth processes are common in nature, society and technology
2. Most systems comprise complex growth features (generalized Gibrat's law)
3. The growth correlation exponent is related to correlations in the dynamics
4. Original Gibrat's law is a special case corresponding to 1/f-noise

Thank you for your attention.

http://www.rybski.de/diego/

[1] H.D. Rozenfeld, D. Rybski et al., Laws of population growth, *PNAS* 105, 18702 (2008).

[2] D. Rybski et al., Scaling laws of human interaction activity, *PNAS* 106, 12640 (2009).

[3] H.D. Rozenfeld, D. Rybski et al., The area and population of cities: new insights from a different perspective on cities, *submitted to American Economic Review*.

[4] D. Rybski et al., Communication activity: temporal correlations, clustering, and growth, *in preparation* (2010).