SCALING & THE UNIVERSALITY OF FUNCTION DIVERSITY ACROSS HUMAN ORGANIZATIONS

Vicky C Yang

With
Chris Kempes
Hyejin Youn
Sid Redner
Geoffrey West

MIT MANAGEMENT
SLOAN SCHOOL

Mathematics is the art of giving the same name to different things

Henri Poincaré

OUTLINE

- Introduction: Organizations, function diversity
- Method: Scaling analysis
- Data sets
- Empirical results: Commonalities & exceptions
- Ideas for models to explain empirical patterns

  & what challenges remains

ORGANIZATIONS DO
WHAT ANY ONE INDIVIDUAL CANNOT

Organizations are complex systems that enable individuals with diverse knowledge and abilities to work to together

In line with Durkheim, Smith, & Weber’s view, organizations as facilitating divisions of labor

MORE RECENTLY: IMPORTANCE OF AGGREGATING KNOWLEDGE

- Ability to aggregate diverse knowledge is connected with solving complex problems/produce prosperous economic outcomes for companies, cities, nations

- In society, collective possesses large volumes of knowledge, individual can only maintain a relatively small subset. Integrate across functions is essential to success of modern human society.

Aminpour et al (2021) *PNAS*, Hidalgo & Hausmann (2009), *PNAS*

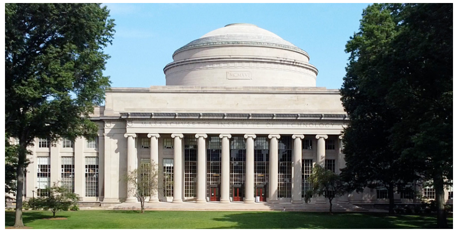
Hausmann et al (2014) *The Atlas of Economic Complexity*

ORGANIZATIONS INTERPRETED BROADLY

Top-down

Companies
e.g., Google

Government agencies
e.g., Federal reserve



Universities
e.g., MIT

Bottom-up

Cities
e.g., Boston

Online communities
e.g., Wikipedia

Participants curated by those in the organization

Participants join as they wish

DEFINE FUNCTION DIVERSITY

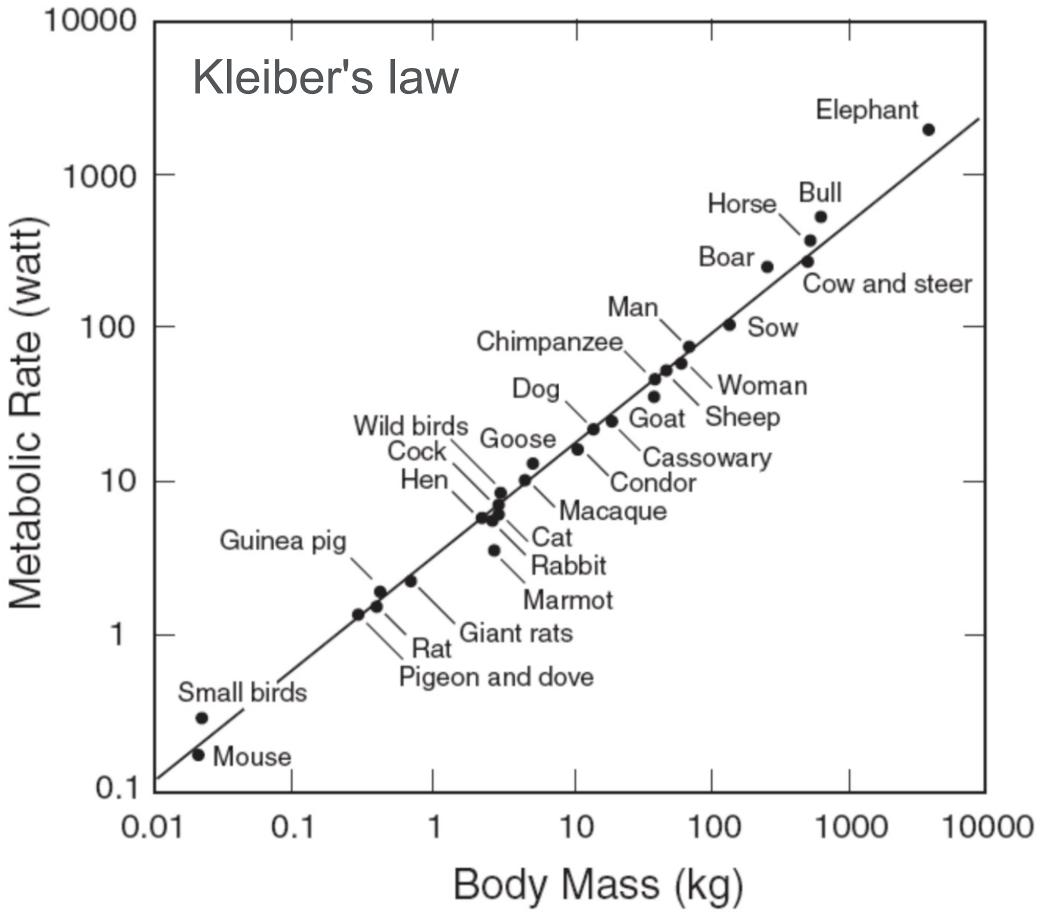
How many different knowledge/skills are aggregated in an organization?

COMMON PATTERNS FOR FUNCTION DIVERSITY ACROSS ORGANIZATIONS?

- Organizations seem very different from each other
- Organization share the same role of coordinating different functions. Perhaps common pattern exist
- Previous studies for diversity in other contexts found commonalities through scaling analysis

SCALING ANALYSIS: EXAMPLE IN BIOLOGY

Scaling analysis: how aggregated properties change with system size



Power law: $y = ax^{\beta}$

On log scale: $\log y = \beta \log x + \log a$

Exponent $\beta \approx$ 3/4

Regularities that transcend individuality

(e.g., evolutionary history, geography)

Empirical findings lead to understanding for underlying mechanisms

Plot from West (2017) *Scale*

SCALING OBSERVATIONS REVEAL NEW UNDERSTANDING FOR STRUCTURES IN THE BODY

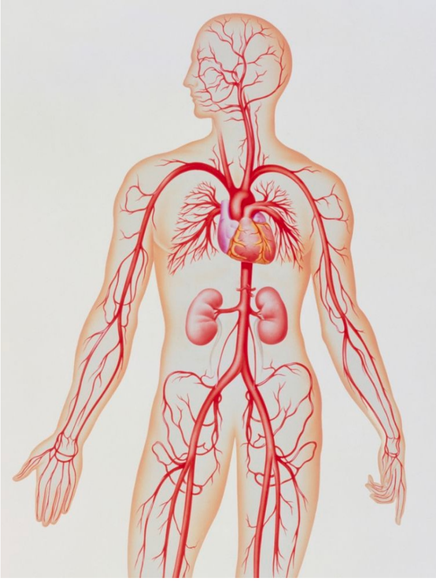
Naively: Metabolism determined by thermal exchange with environment

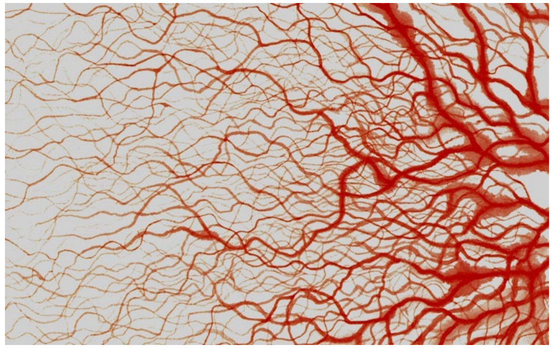
if  $\approx$ 

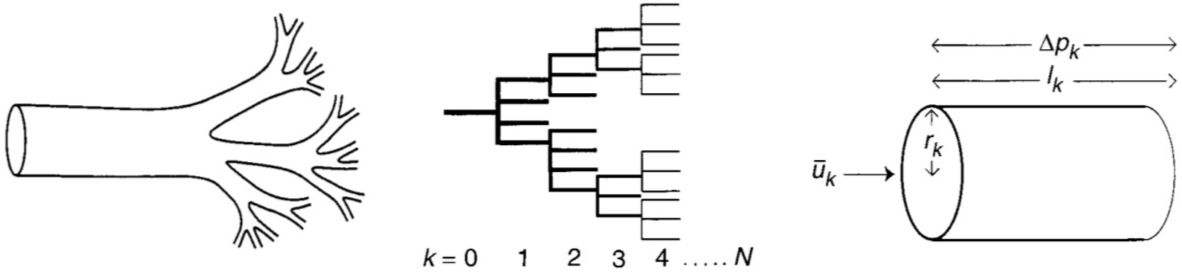
Metabolic rate ~ Surface area ~ $\mathrm{Volume}^{2/3} \sim \mathrm{Mass}^{2/3}$

$$\frac{2}{3} \neq \frac{3}{4}$$

SCALING OBSERVATIONS REVEAL NEW UNDERSTANDING FOR STRUCTURES IN THE BODY







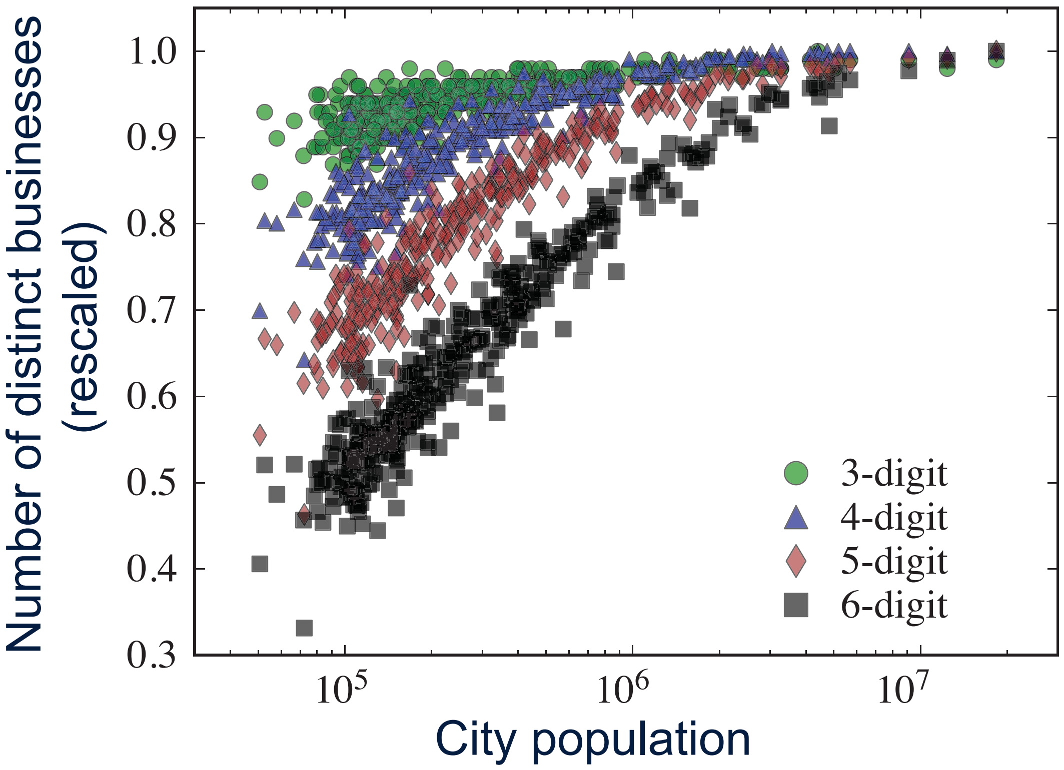
1. Space-filling, fractal-like branching networks
2. Final branch (capillary) in size-invariant
3. Energy required to distribute resources is minimized

→ Metabolic rate ~ $\mathrm{Mass}^{3/4}$

West et al. (1997) *Science*

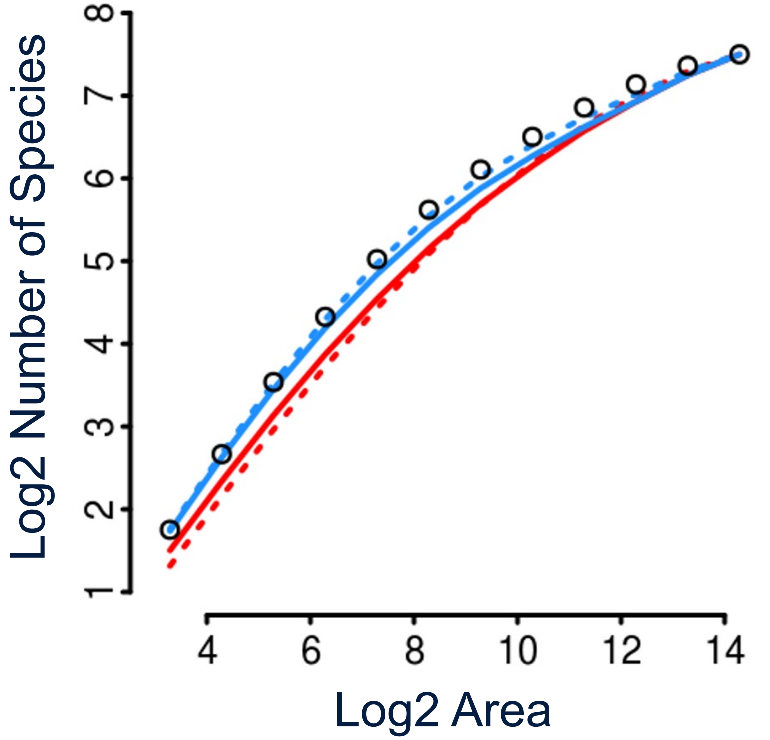
SCALING ANALYSIS FOR DIVERSITY IN OTHER CONTEXTS SHOW REGULARITY

Businesses in US cities



Youn et al (2016) *J. R. Soc. Interface*

Species in ecosystems



McGlinn, Xiao & White (2013) *PeerJ*

Key research questions

Are there common empirical patterns for function diversity in organizations?

What common mechanisms may be governing function diversity in organizations?

Our process

Use scaling framework as point of departure for empirical analysis.

Will develop models to explain these empirical findings.

DATA: TOP-DOWN ORGANIZATIONS

	US Federal agencies	Norwegian companies	US Universities

DATA: TOP-DOWN ORGANIZATIONS

	US Federal agencies	Norwegian companies	US Universities
Data source	FedScope		
Observations	125		
Size range	2 – 390K employees		
Function measure	Occupations classified by Office of Personnel Management		
Number of possible functions	677		

DATA: TOP-DOWN ORGANIZATIONS

	US Federal agencies	Norwegian companies	US Universities
Data source	FedScope	Statistics Norway	
Observations	125	3,191 (1 obs = average of 5 companies)	
Size range	2 – 390K employees	5 – 10K employees	
Function measure	Occupations classified by Office of Personnel Management	Occupations classified by International Standard Classification of Occupations (ISCO-08)	
Number of possible functions	677	436	

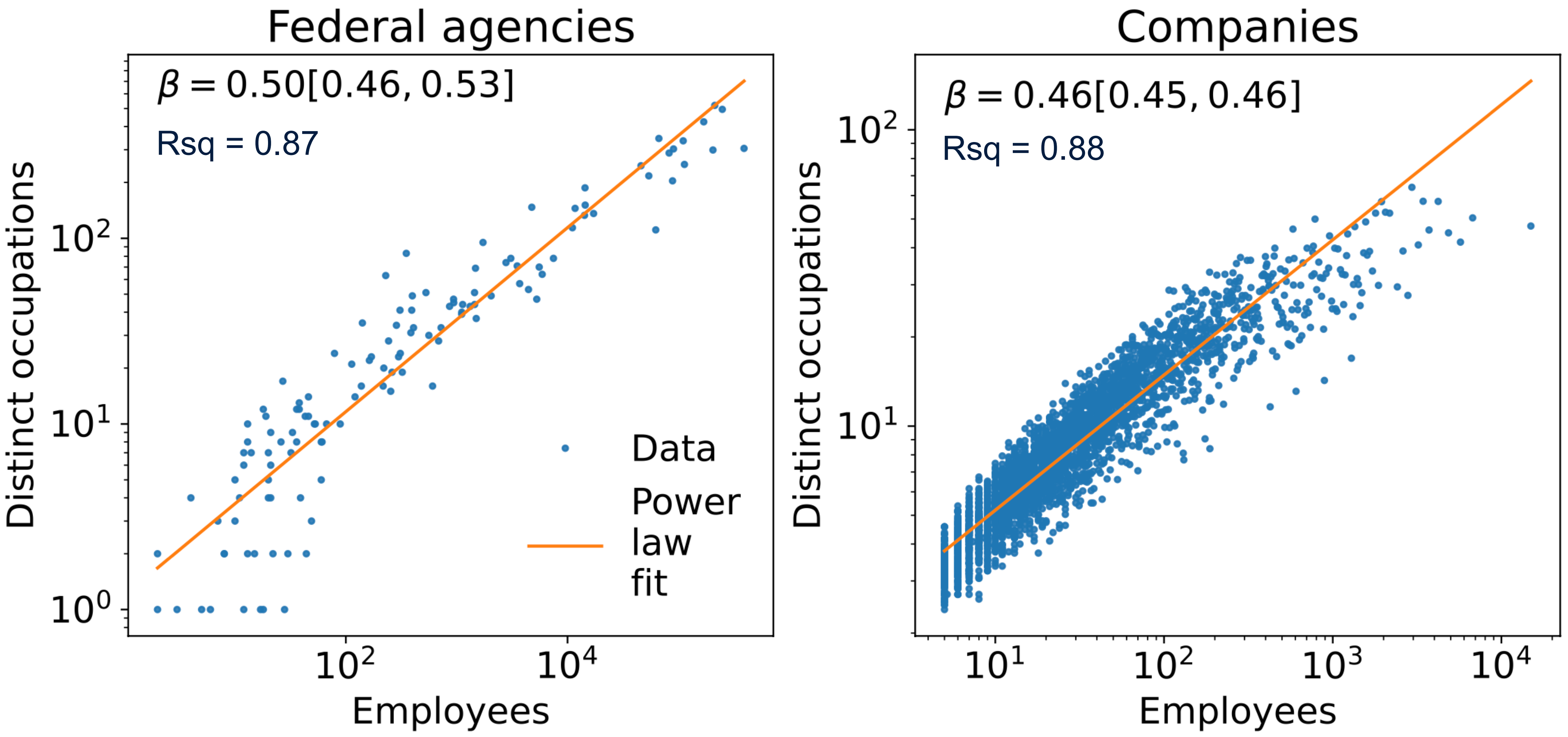
DATA: TOP-DOWN ORGANIZATIONS

	US Federal agencies	Norwegian companies	US Universities
Data source	FedScope	Statistics Norway	Integrated Postsecondary Education Data System (IPEDS)
Observations	125	3,191 (1 obs = average of 5 companies)	1,592 (bachelor's level+) + 950 (associate level)
Size range	2 – 390K employees	5 – 10K employees	5 – 10K graduates
Function measure	Occupations classified by Office of Personnel Management	Occupations classified by International Standard Classification of Occupations (ISCO-08)	Degrees awarded classified by Classification of Instructional Programs (CIP) code
Number of possible functions	677	436	459 CIP codes & 4 degree levels

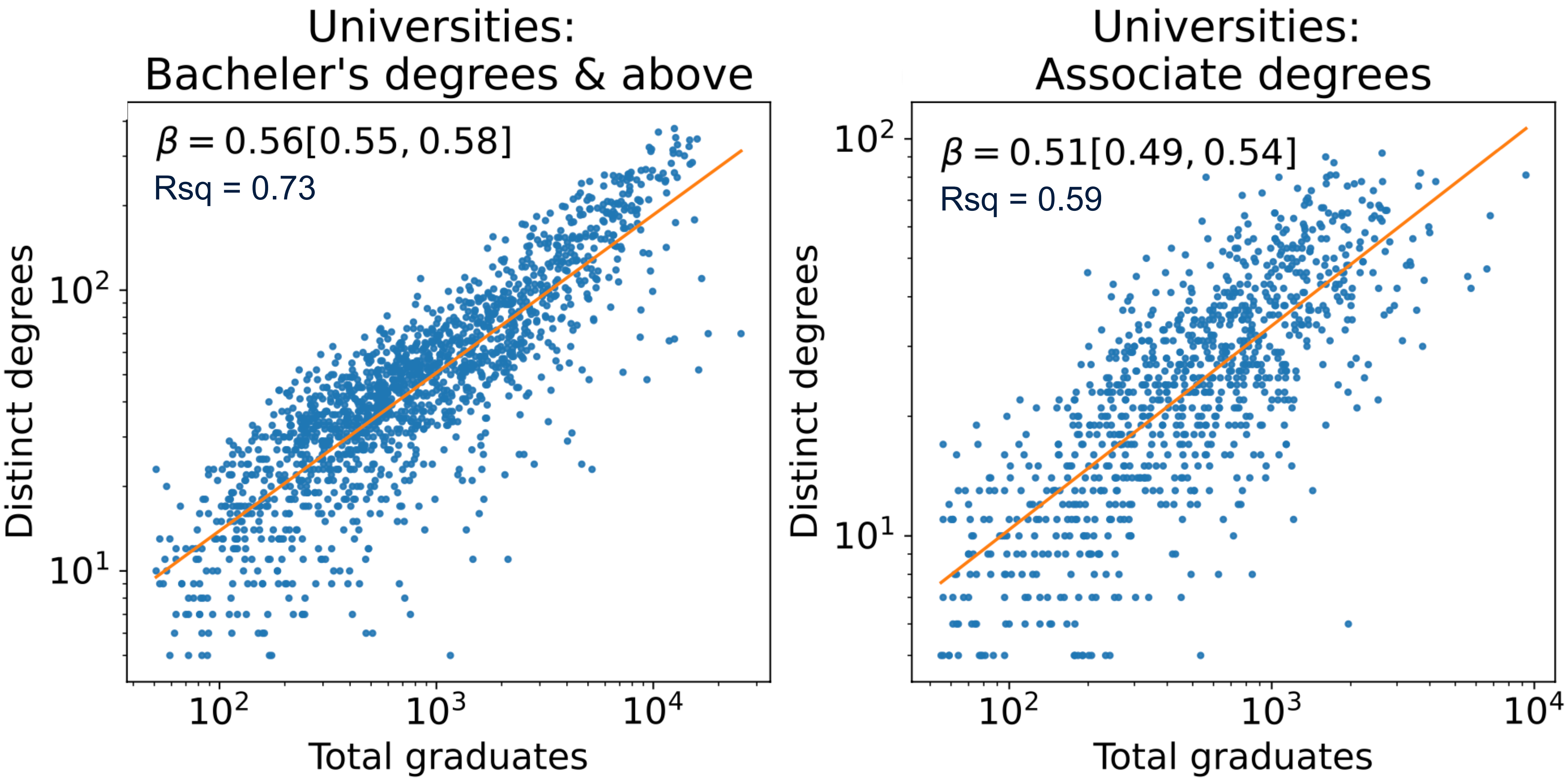
DATA: BOTTOM-UP ORGANIZATION

	US urban areas (cities)
Data source	Bureau of Labor Statistics
Observations	422
Size range	10K –10M individuals
Function measure	Occupations classified by 2010 Standard Occupational Classification system
Number of possible functions	798

EMPIRICAL RESULTS

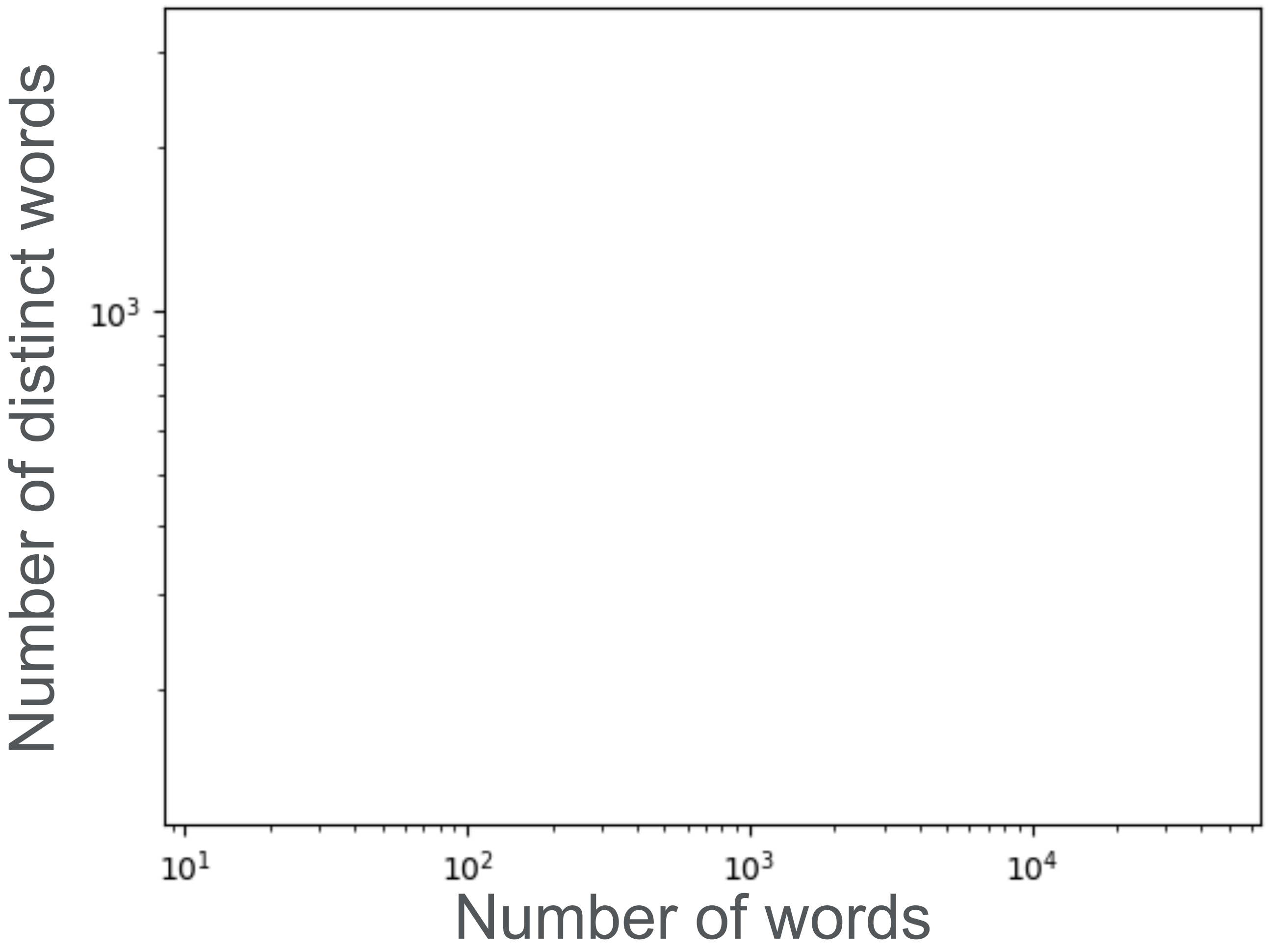


EMPIRICAL RESULTS

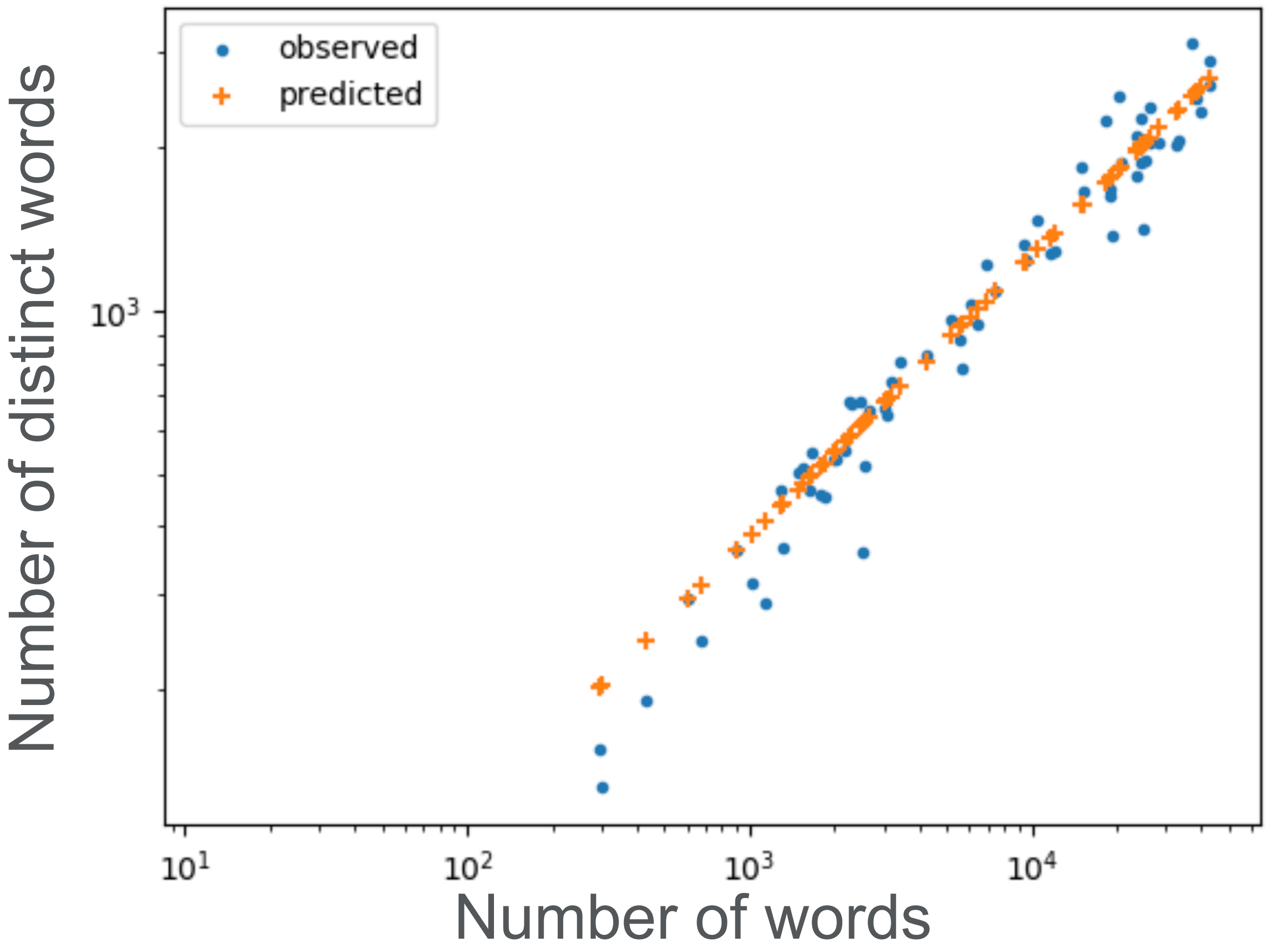


THESE PATTERNS RESEMBLE HEAPS' LAW

HEAPS' LAW: BOOKS



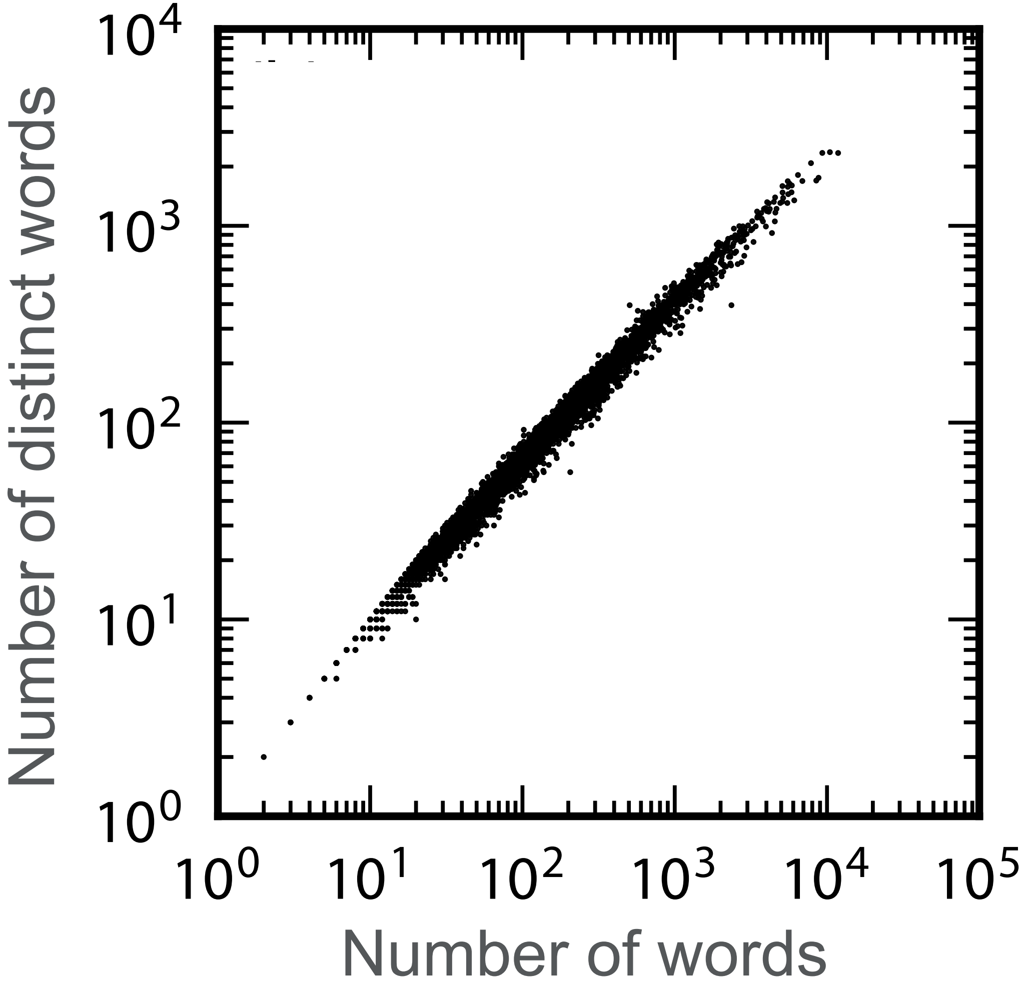
HEAPS' LAW: BOOKS



Books in the bible
(Genesis, Exodus, etc.)
Elements = words
β = 0.52

Cook (2019) Estimating vocabulary size with Heaps' law

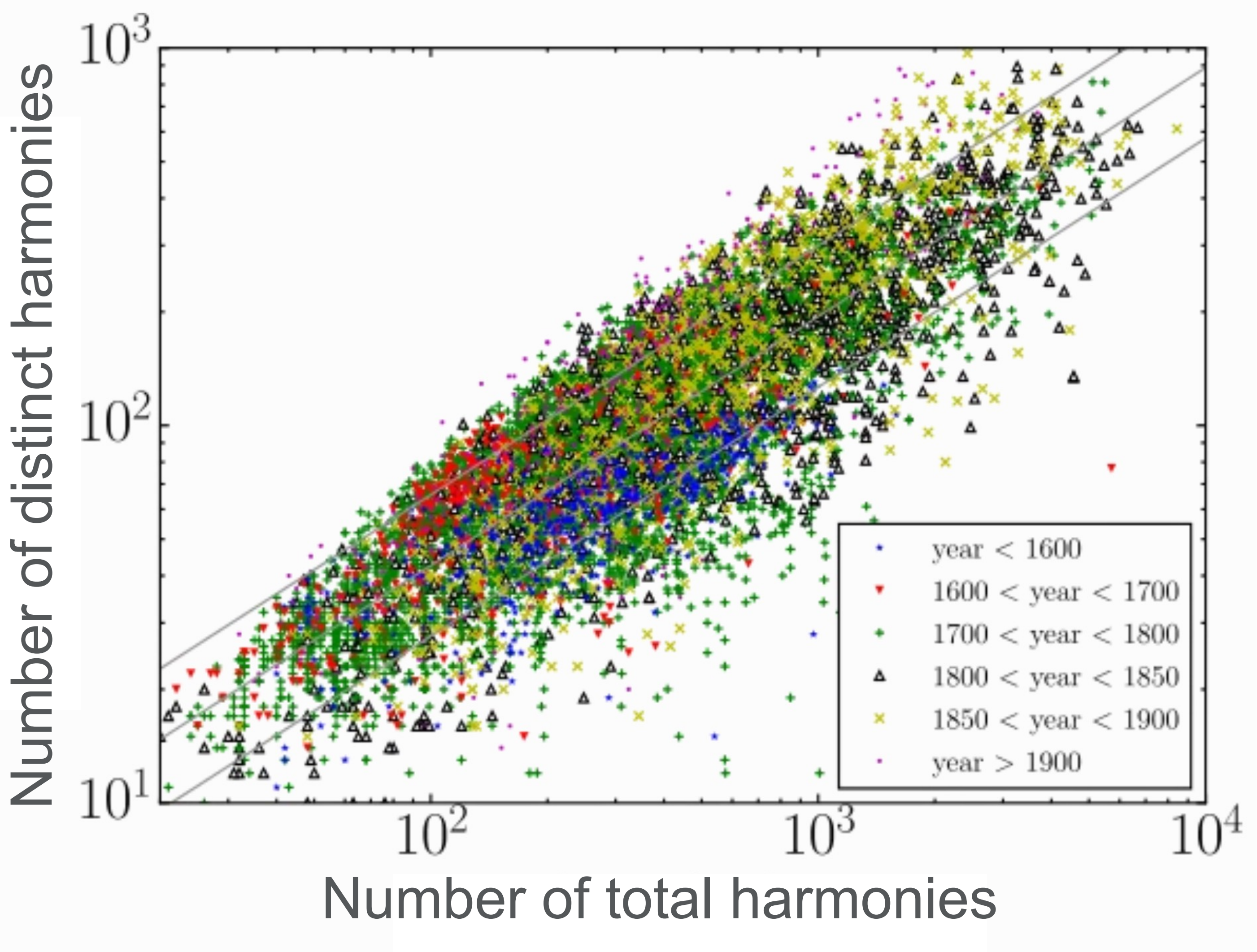
HEAPS' LAW: WIKIPEDIA ARTICLES



Wikipedia articles
Elements = words
$\beta = 0.7$

Gerlach & Altmann (2014) *New Journal of Physics*

HEAPS' LAW: MUSIC



9K+ Classical music pieces

Elements = harmonies

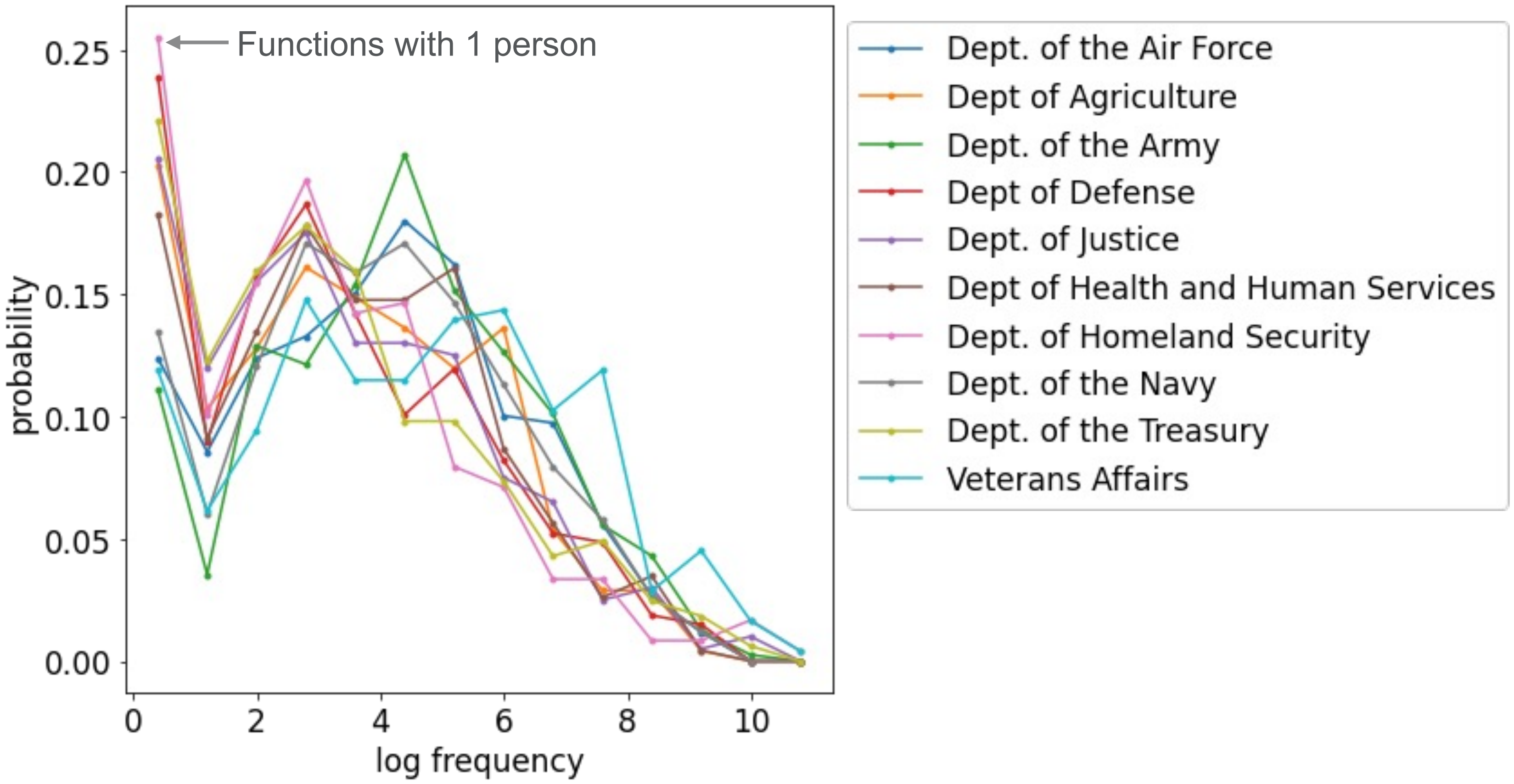
$\beta = 0.66$

Serra-Peralta, Serra, Corral (2021) *EPJ Data Science*

IN INDIVIDUAL ORGANIZATIONS

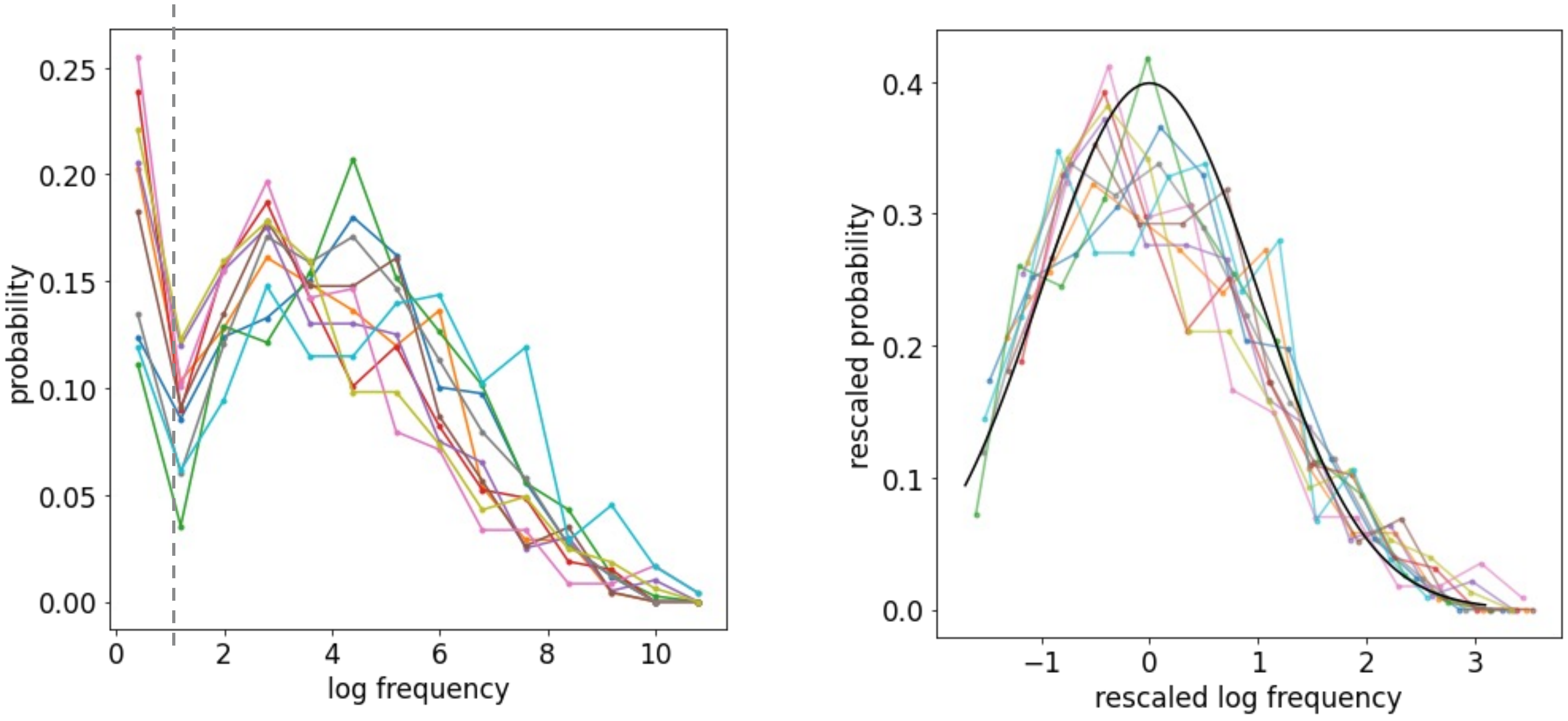
ABUNDANCE DISTRIBUTION OF FUNCTIONS

10 largest federal agencies: ~ log normal distribution



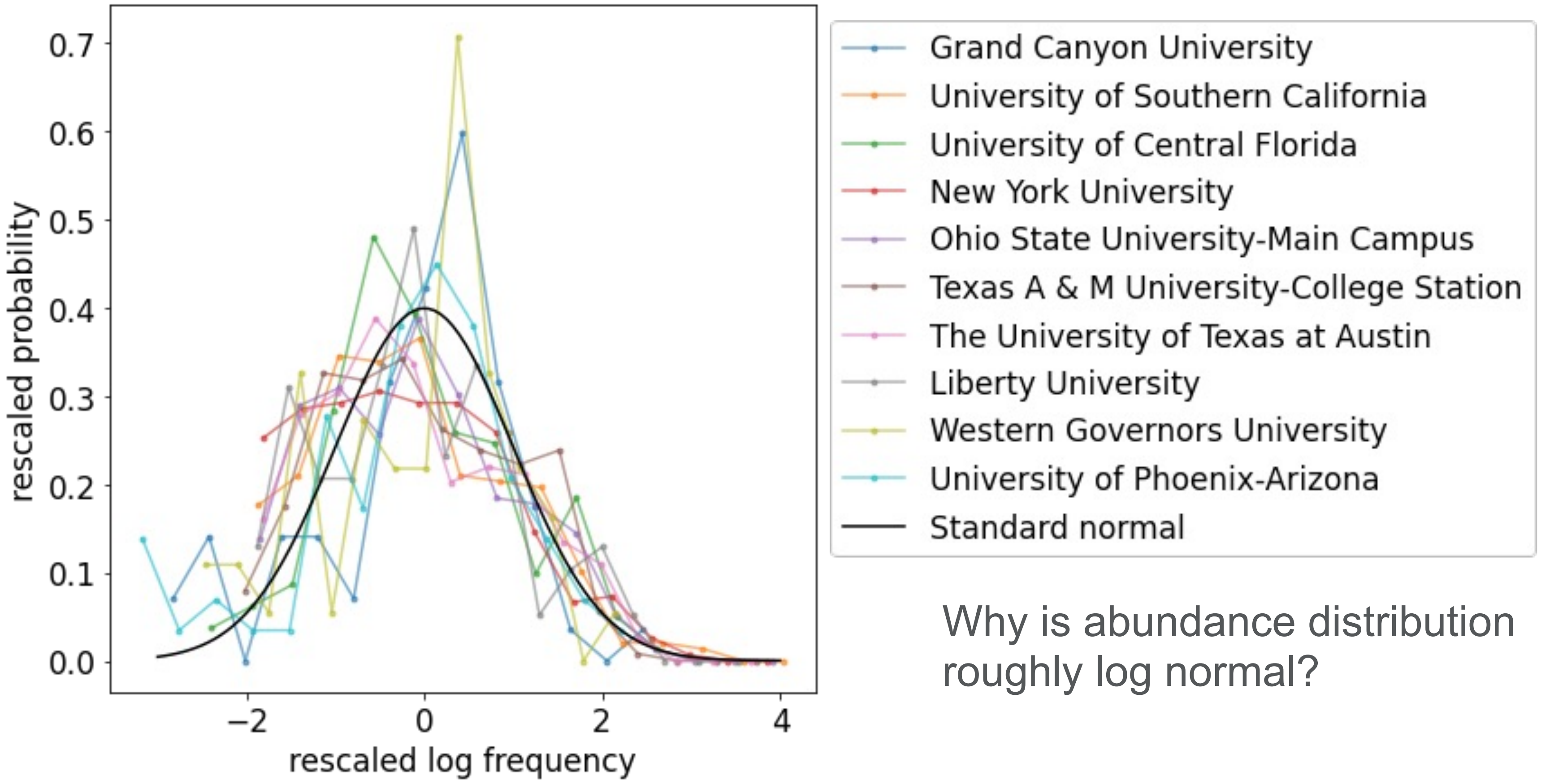
ABUNDANCE DISTRIBUTION OF FUNCTIONS

10 largest federal agencies: ~ log normal distribution



ABUNDANCE DISTRIBUTION OF FUNCTIONS

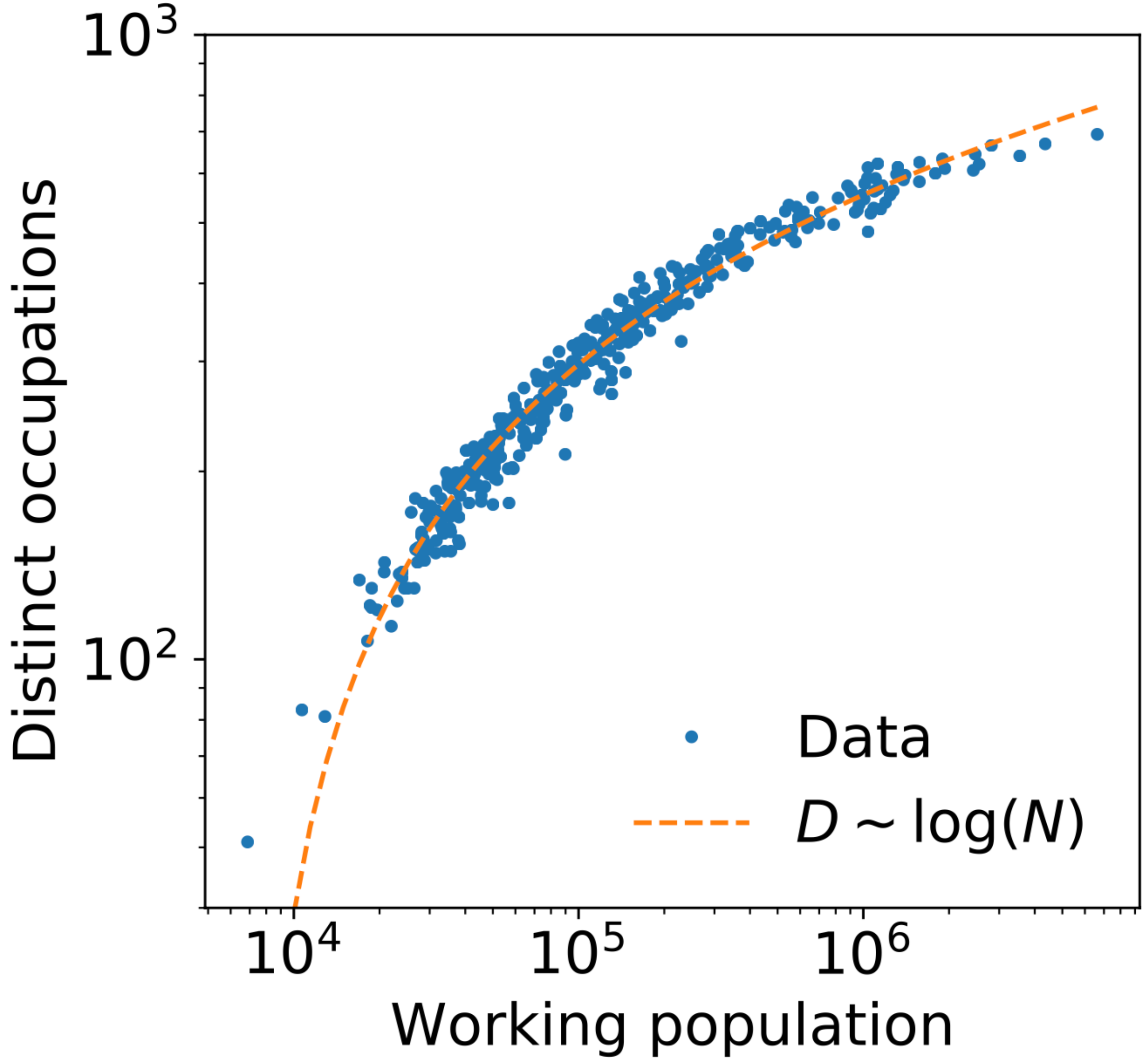
10 largest universities granting bachelor's degrees



Why is abundance distribution roughly log normal?

AN EXCEPTION: CITIES

EMPIRICAL RESULTS: CITIES



Cities show logarithmic scaling

GDP (and other socio-economic metrics) is *exponentially* associated with diversity

Why are cities different?

MODELING THE MECHANISMS DETERMINING FUNCTION DIVERSITY

WHAT IS GOOD MODEL

- Power-law scaling and logarithmic scaling

  Ideally predict both by altering one parameter

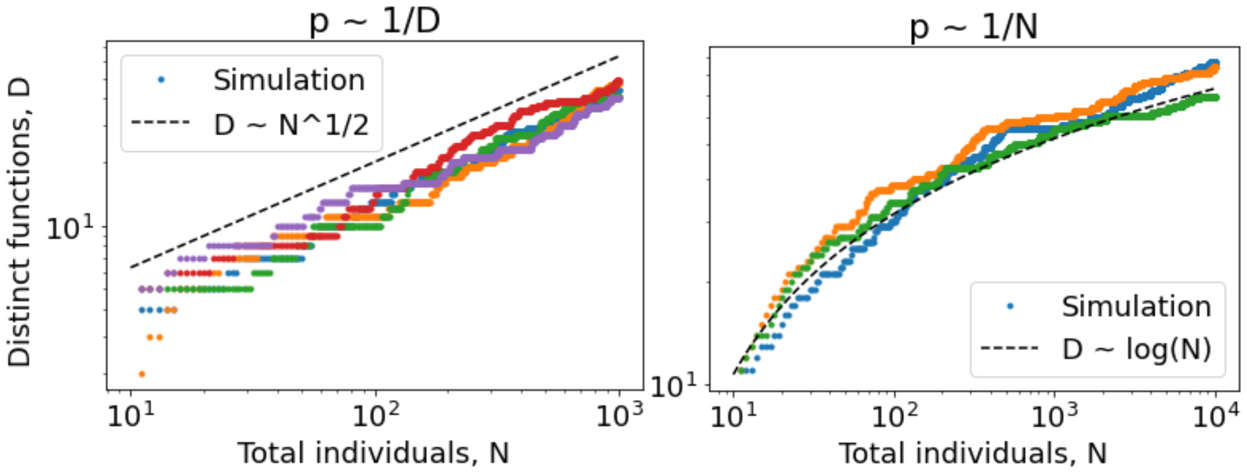
- Recovers abundance distributions in organizations
- The mechanisms considered in the model faithful to actual processes in organizations

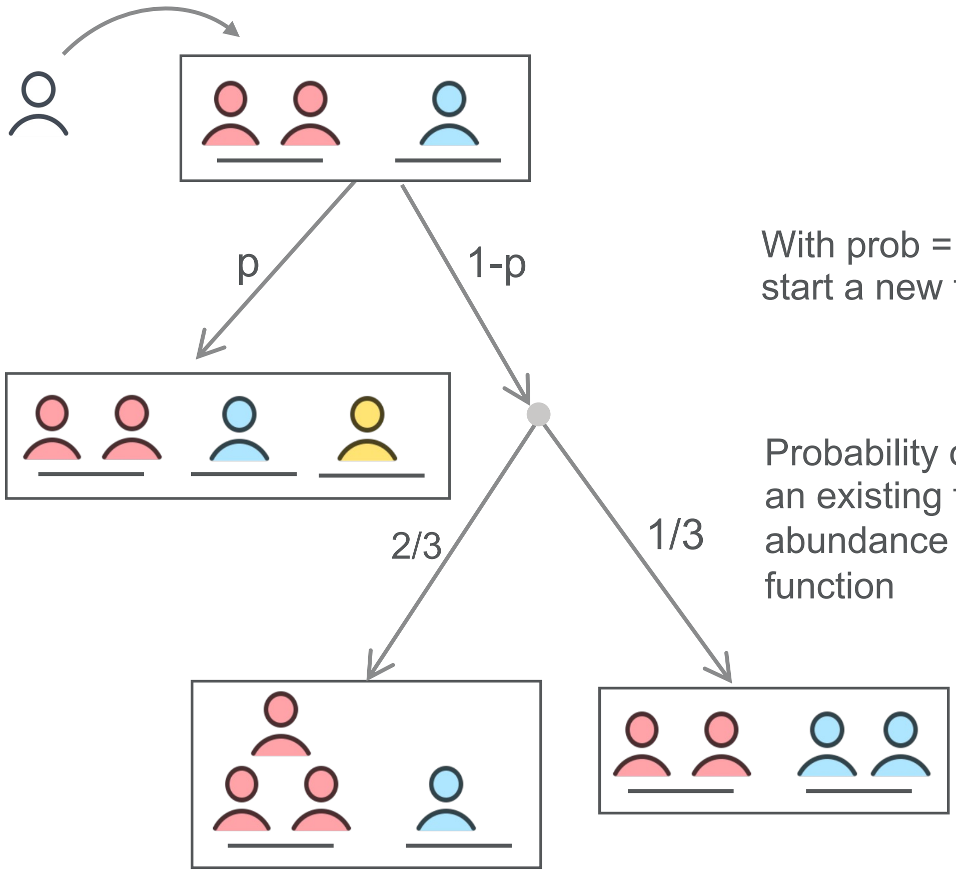
PROMISING MODELING FRAMEWORKS

Rich-get-richer process

E.g., Yule-Simon model

Various p can generate different scaling relationship
If $p \sim 1/D \rightarrow D \sim N^{1/2}$ $(dD/dN = c/N)$
If $p \sim 1/N \rightarrow D \sim \log N$ $(dD/dN = c/D)$





With prob = p, start a new function

Probability of joining an existing function ~ abundance of function

But:
How realistic is it for organizations?
Can we measure these parameters in real organizations?

PROMISING MODELING FRAMEWORKS

Sample Space Reduction/Expansion

As organizations grow

- Functions makes certain new functions possible
- The number of options available gets more limited

Perhaps the degree of the reduction/expression explains the difference of top-down vs. bottom-up organizations

But:

- Explains ½ scaling, but not yet logarithms.
- What is the structure for one function to stimulate/prohibit another function?

MORE THOUGHTS ON MODELING CHALLENGES

- A common challenge for models inspired by scaling analysis is that multiple models can explain the same exponent, making it hard to compare across models.
  - This is happening in the modeling of Heaps' law, species-area curve, and models about urban areas
- We have identified two modes of behavior. This should help narrow down the valid models
- We have also identified some common behavior with the abundance distribution, which could further help to select models

WHAT'S NEXT

- Connect with organizational literature on what are realistic mechanisms for functions to depend on/inhibit each other
- Measure some parameters empirically?
- Evaluate models based on the selection criteria we developed

CONCLUSIONS

- Commonalities across organizations: number of distinct functions ~ size to ½ power (Federal agencies, companies, universities)
- Differences: cities (bottom-up organizations) are different from other top-down examples. Considerable variance across many organizations.
- Identified model goals & promising frameworks. But challenges remain

THANK YOU!

LET'S AGGREGATE SOME DIVERSE KNOWLEDGE

Vicky Chuqiao Yang

vcyang@mit.edu| www.vcyang.com | @VickyCYang

Preprint: Yang, Kempes, Youn, Redner, West (2022) arXiv: 2208.06487
Scaling and the Universality of Function Diversity Across Human Organizations

BONUS SLIDES

MIT SLOAN SCHOOL OF MANAGEMENT

MIT MANAGEMENT SLOAN SCHOOL

RANK-FREQUENCY DISTRIBUTION

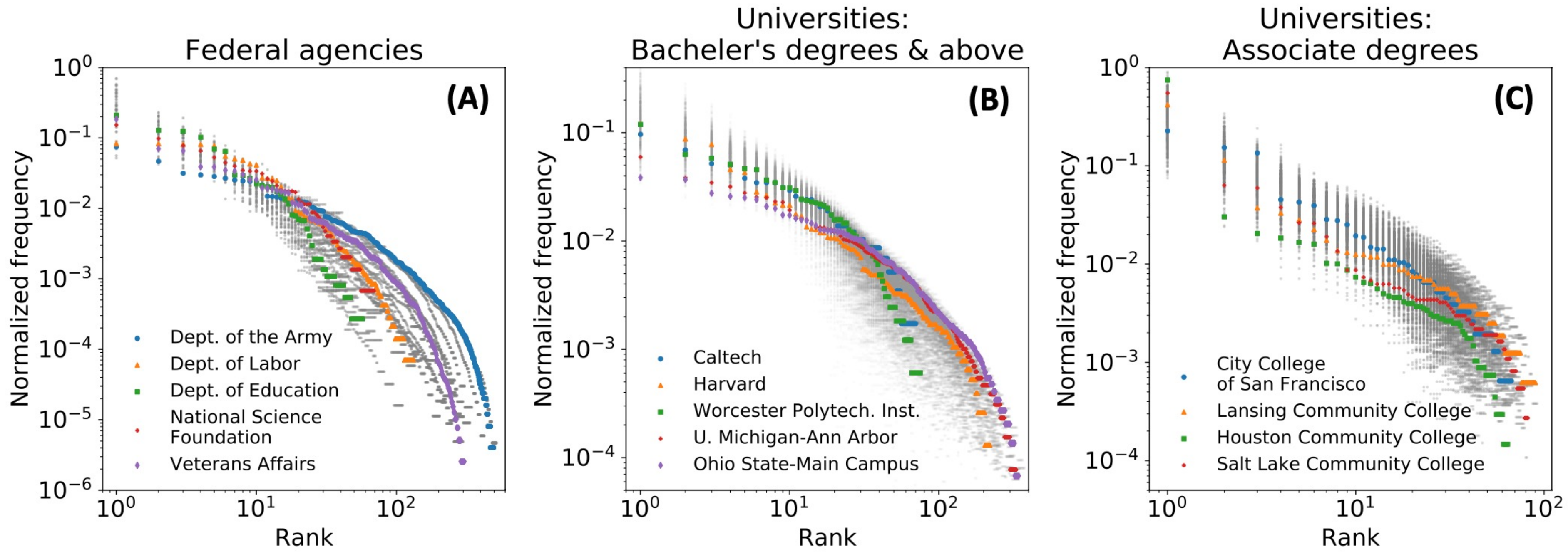
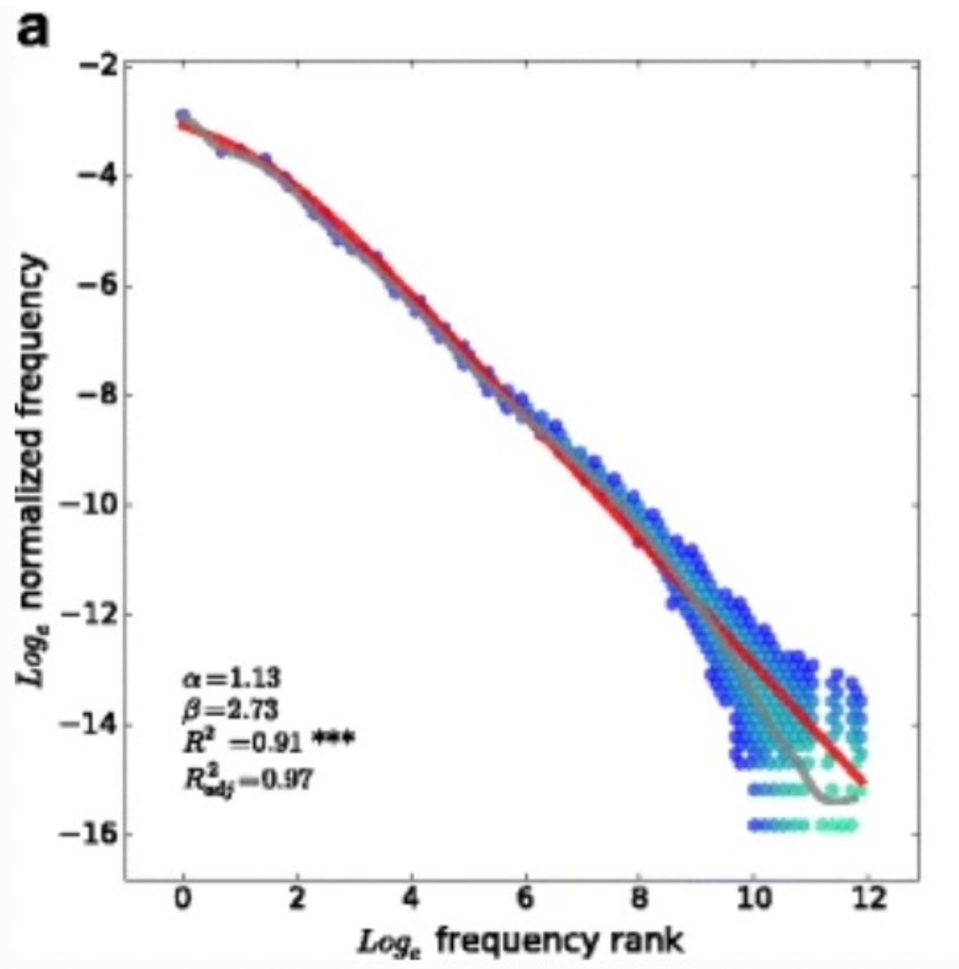


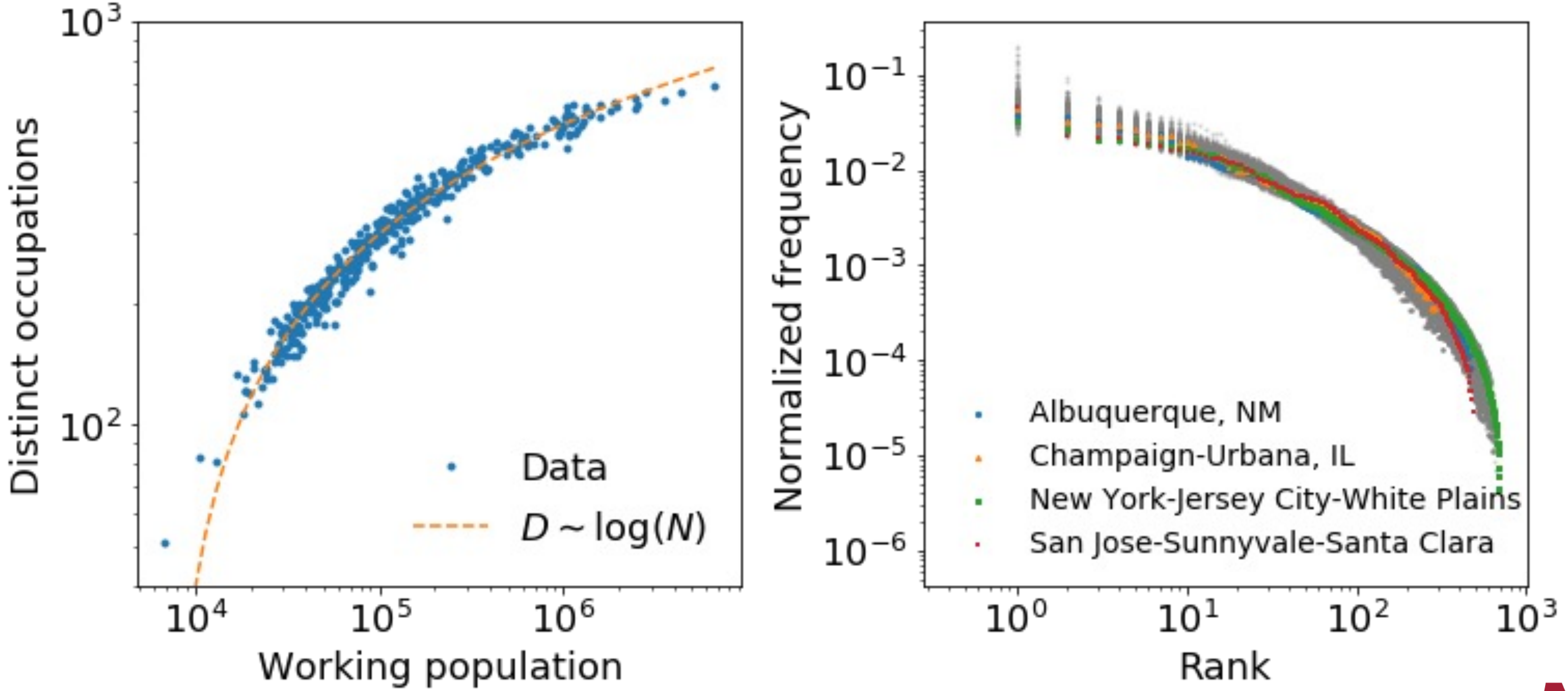
Figure 2. The abundance of functions versus rank of the function (one denoting the most frequent) for (A) US Federal agencies, (B) US universities granting bachelor's degrees and above, and (C) US universities granting associate degrees. Grey points show data for all organizations, while colored points highlight several selected organizations for illustration.

RANK-FREQUENCY DISTRIBUTION: ENGLISH



Piantadosi (2014) *Psychonomic Bulletin & Review*

CITIES – FREQUENCY DISTRIBUTION

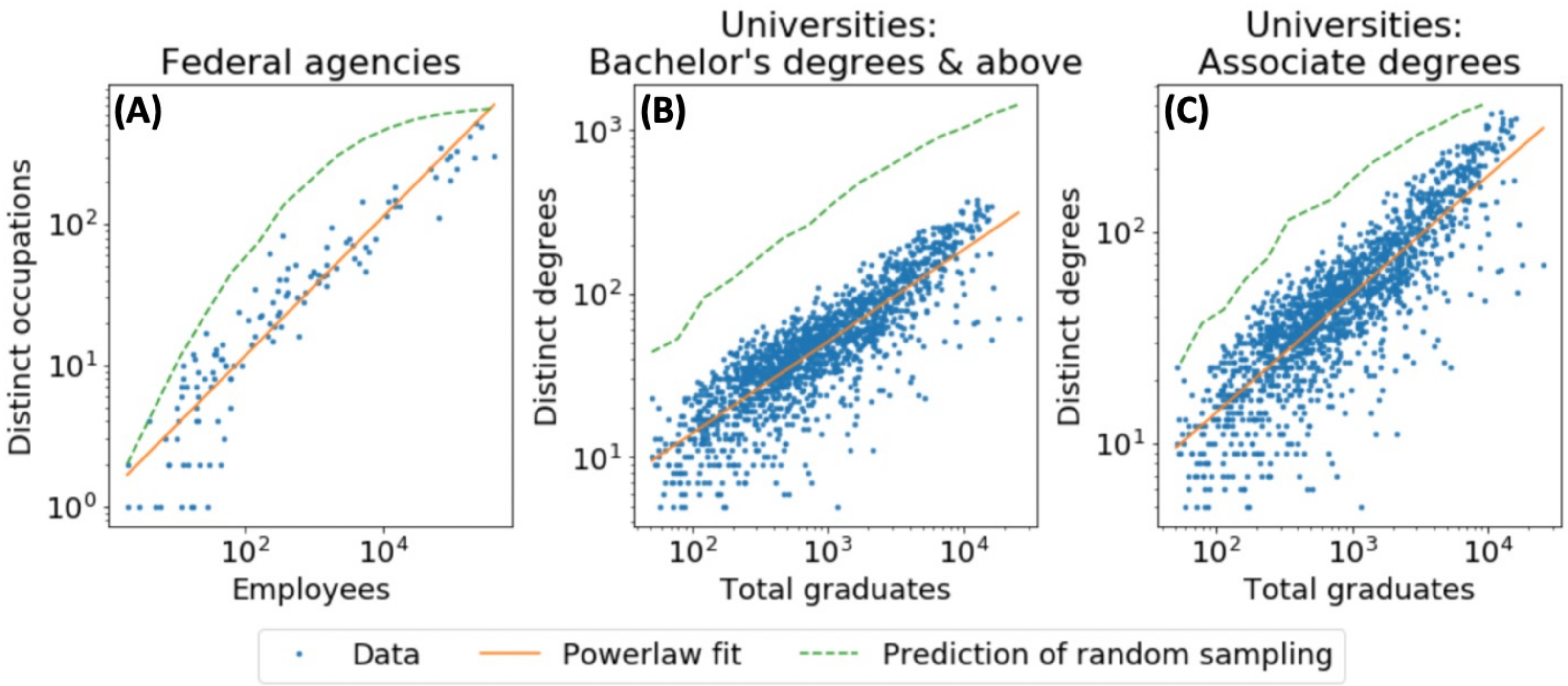


SUMMARY OF MODELS THAT EXPLAIN HEAPS LAW

1. Random sample from Zifp’s rank-frequency distribution.
2. Rich-gets-richer (e.g., Yule-Simon process)
3. Sample space reducing
4. Sample space expanding (e.g., Polya process with triggering of new possibilities)

(1) RANDOM SAMPLING FROM ZIPF'S RANK-FREQUENCY DISTRIBUTION

But: rank-frequency in human organizations do not follow Zipf's law. Random sampling leads to over estimates.



(2) RICH-GET-RICHER DYNAMICS

e.g., Polya process, Yule-Simon model, and their extensions

For each new person entering the organization, the likelihood they join a function increases with frequency in the function

But:

How realistic is it for organizations?

Predicts power-law scaling in its original form

(3) SAMPLE SPACE REDUCING

- Models aging process where the number of options gets more limited as one ages.
- Plan going forward: Sample space reducing (top-down) vs. expansion (bottom-up)

(4) SAMPLE SPACE EXPANSION

- Polya model with triggering dynamics.

REVISED YULE-SIMON COULD EXPLAINS BOTH SCALING

As an individual join an organization, they have probability p be a function that did not exist in the organization before, and 1-p to join an existing function.

If $p \sim 1/D \rightarrow D \sim N^{1/2}$ $(dD/dN = 1/N)$

If $p \sim 1/N \rightarrow D \sim \log N$ $(dD/dN = 1/D)$

The likelihood of joining an existing function is proportional to the number of people in that function, then the abundance distribution should be log normal in both scenarios.