

クレジット:

UTokyo Online Education 数理手法Ⅷ 2019 島田尚

ライセンス:

利用者は、本講義資料を、教育的な目的に限ってページ単位で利用することができます。特に記載のない限り、本講義資料はページ単位でクリエイティブ・コモンズ 表示-非営利-改変禁止 ライセンスの下に提供されています。

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

本講義資料内には、東京大学が第三者より許諾を得て利用している画像等や、各種ライセンスによって提供されている画像等が含まれています。個々の画像等を本講義資料から切り離して利用することはできません。個々の画像等の利用については、それぞれの権利者の定めるところに従ってください。



数理手法VIII

数理・情報教育研究センター 数学基礎教育部門
工学系研究科システム創成学専攻

島田 尚

講義内容予定

- A. グーグル検索、国の競争力などの「重要性評価手法」関連の数理
 - ・ペロン・フロベニウスの定理、power method、特異値分解
- B. 統計解析とデータの縮約（データサイエンス、機械学習）
 - ・特異値分解、最小2乗問題、主成分分析、スパースモデリング
- C. 株価や為替変動、人の通信や移動におけるべき分布、都市成長に見られるスケーリング則等の普遍的統計則
 - ・大数の法則、中心極限定理
 - ・べき分布を生むメカニズム、自己組織的臨界性
- D. 複雑ネットワーク
 - ・ネットワーク構造の特徴づけ
 - ・ネットワークの頑健性、脆弱性、伝搬の数理

1. 確率分布（頻度分布）の扱いと、 正規分布（指数分布／ポワソン）の重要性



Photo by Deutsche Bundesbank, Frankfurt am Main, Germany, from Wikipedia
https://upload.wikimedia.org/wikipedia/commons/0/0d/10_DM_Serie4_Vorderseite.jpg
Public domain

$$N(\mu, \sigma^2) : f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}$$

中心極限定理：

サンプル数 $n \rightarrow \infty$ につれ、

$$Z_n = \frac{1}{\sqrt{n}} \left(\sum_i \frac{X_i - \mu_i}{\sigma_i} \right) \sim N(0,1) \left(= \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \right)$$

2. 正規分布的でない場合に何を考えるか

X_i の分布が幅広（モーメントが発散）

- 1) → 各種特徴量、相関係数など解析に注意が必要
- 2) → その由来は？

A. 乗算ノイズや成長（Yule-Simon）過程

B. $\{X_i\}$ が互いに相関している（臨界現象）

C. 次元性や最適化の結果

ほぼ扱わないこと：
検定・回帰

社会・経済系での普遍分布・ベキ分布

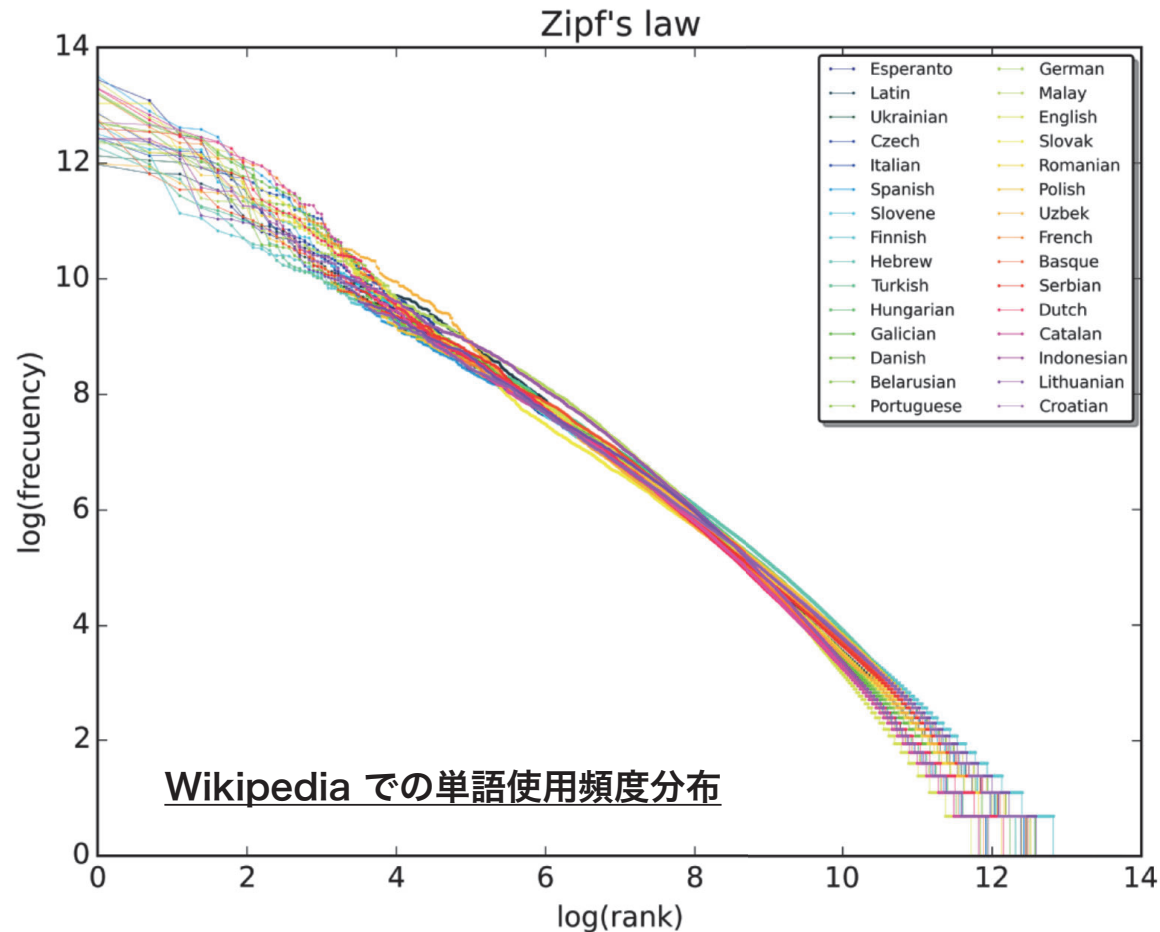
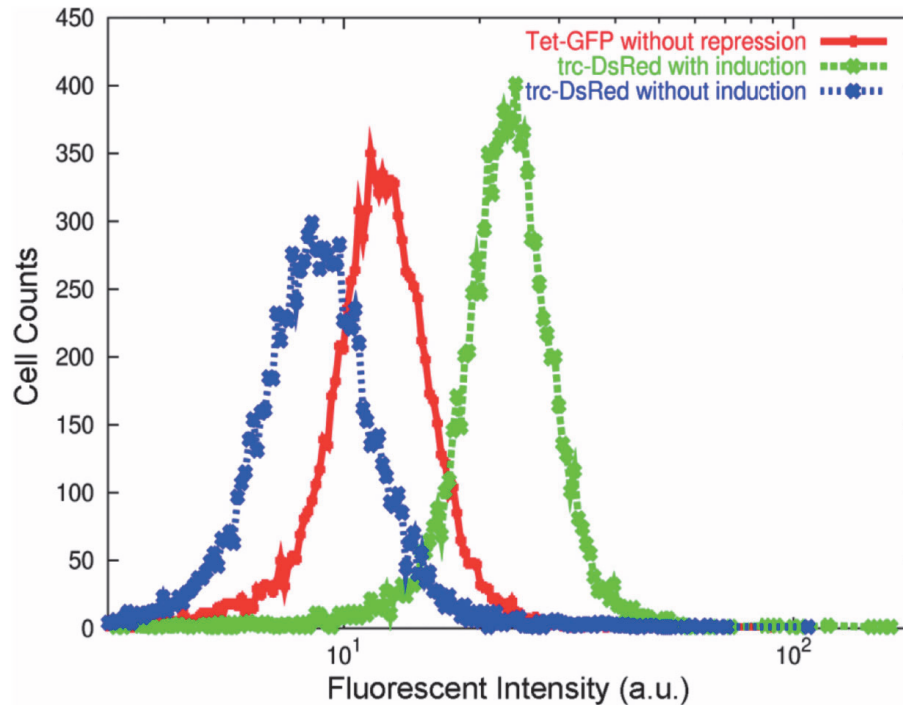


Photo by SergioJimenez, from Wikipedia
https://en.wikipedia.org/wiki/Zipf%27s_law#/media/File:Zipf_30wiki_en_labels.png
CC BY-SA 4.0

- ・ 情報圧縮などの最適化の結果？
- ・ 「成長し続ける系であるせい」かも（Yule-Simon 過程、BAモデル）

対数正規分布（べき分布）の例： 遺伝子発現量

Chikara Furusawa, Takao Suzuki, Akiko Kashiwagi, Tetsuya Yomo and Kunihiro Kaneko,
"Ubiquity of log-normal distributions in intra-cellular reaction dynamics",
biophysics Vol.1, pp.25-31(2003) より引用 Fig.3.
https://www.jstage.jst.go.jp/article/biophysics/1/0/1_0_25/_article



C. Furusawa et al., doi: 10.2142/biophysics.1.25

著作権等の都合により、ここに挿入されていた画像を削除しました。

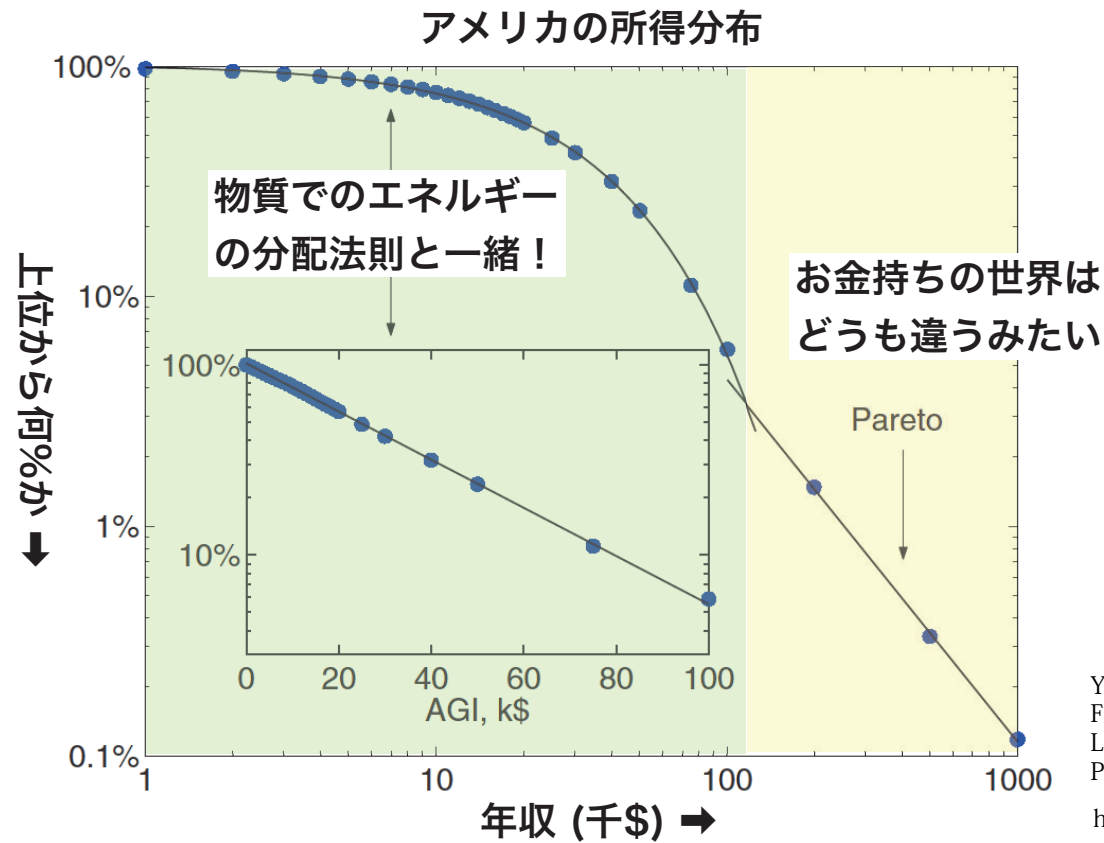
ランク別遺伝子発現量グラフ

Chikara Furusawa and Kunihiro
Kaneko, Zipf's Law in Gene
Expression, Physical Review Letters
90, 088102-Published 26 February
2003 Figure1

[https://journals.aps.org/prl/
abstract/10.1103/
PhysRevLett.90.088102](https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.90.088102)

$$\text{乗算ノイズ過程} : x_{t+1} = \eta_t x_t + \xi_t$$

社会・経済系での普遍分布・ベキ分布



収入上位数%層をのぞけば、物質で使う統計物理の法則が収入にも当てはまる!

社会・経済系での普遍分布・ベキ分布

株価や売り上げの変動はガウス分布でない（「経済物理学」の端緒）

著作権等の都合により、ここに挿入されていた画像を削除しました。

Rosario N. Mantegna & H. Eugene Stanley, "Scaling behaviour in the dynamics of an economic index",
Nature volume 376,
pages 46–49 (1995), FIG. 1

<https://www.nature.com/articles/376046a0>

株価の変動の分布

Mantagna & Stanley, Nature (1995)

著作権等の都合により、ここに挿入されていた画像を削除しました。

Michael H. R. Stanley, Luís A. N. Amaral, Sergey V. Buldyrev, Shlomo Havlin, Heiko Leschhorn, Philipp Maass, Michael A. Salinger & H. Eugene Stanley, "Scaling behaviour in the growth of companies",
Nature volume 379, pages 804–806
(1996), FIG. 3

<https://www.nature.com/articles/379804a0>

企業の成長率の分布

Stanley *et al.*, Nature (1996)

アメリカ株価変動の相関構造

GICS sort (産業分類)

Energy (エネルギー)

Materials (素材)

Industrials (資本財・サービス・輸送)

Consumer Discretionary
(一般消費財・サービス、車、衣料、etc)

Consumer Staples (生活必需品)

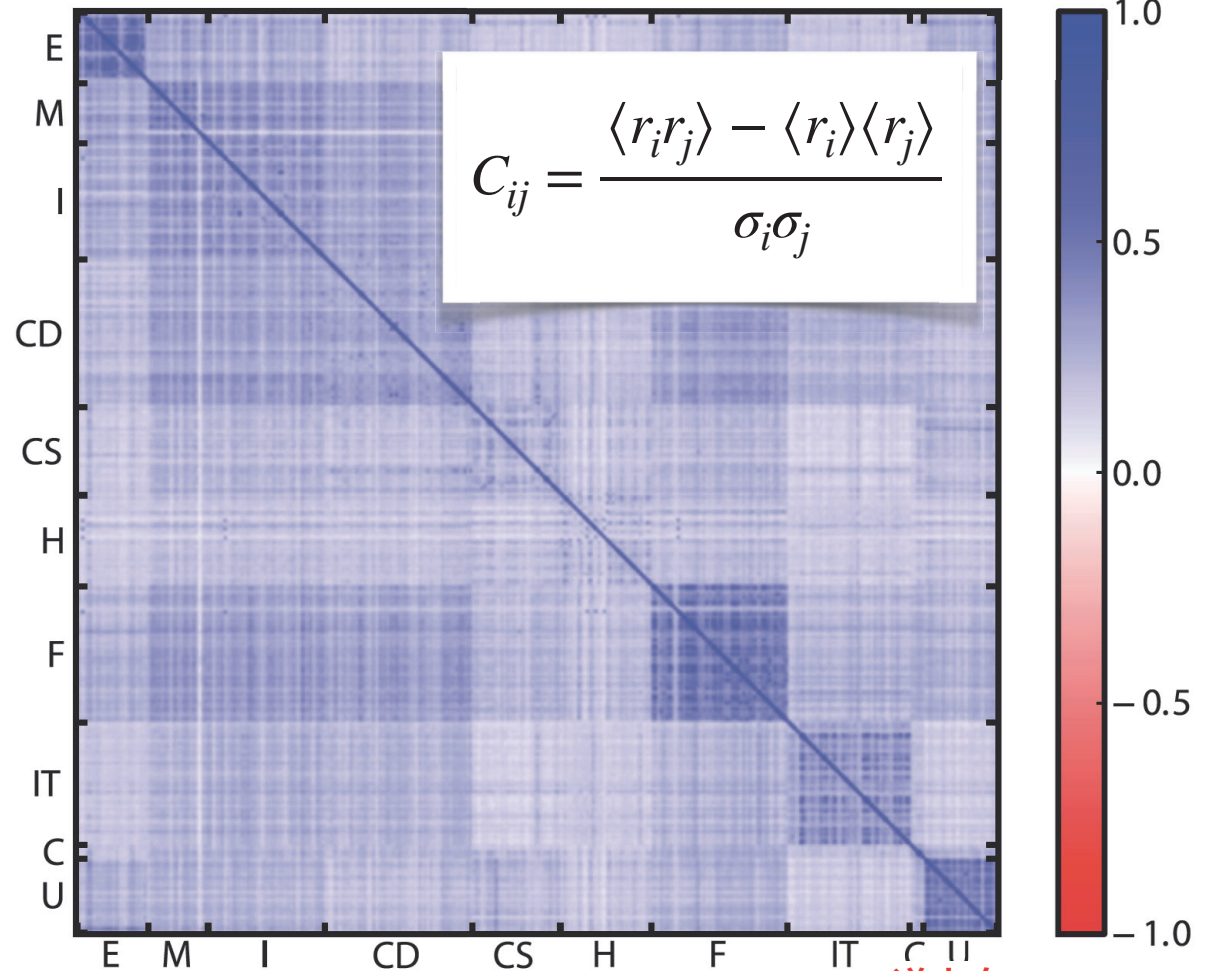
Health Care (ヘルスケア)

Financials (金融)

Information Technology (情報技術)

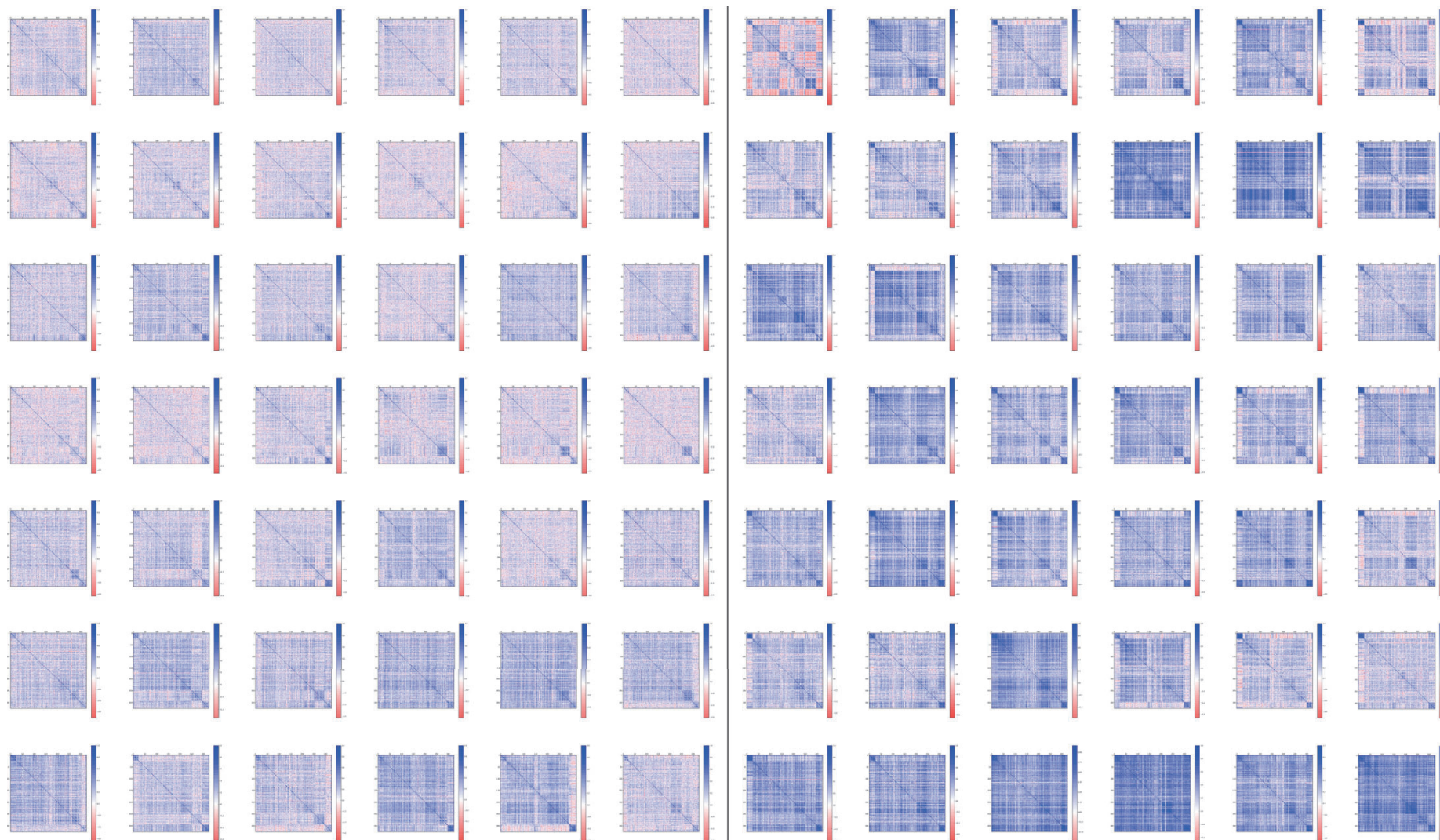
Telecommunication Services (通信)

Utilities (公益事業)

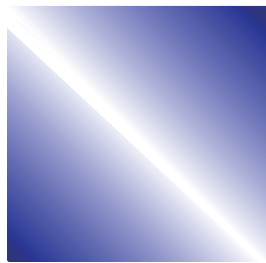


(S&P500 のうち310銘柄)

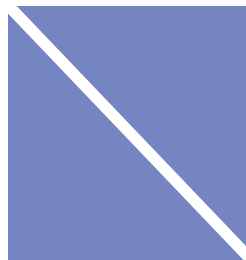
相関構造は毎月のように変わる



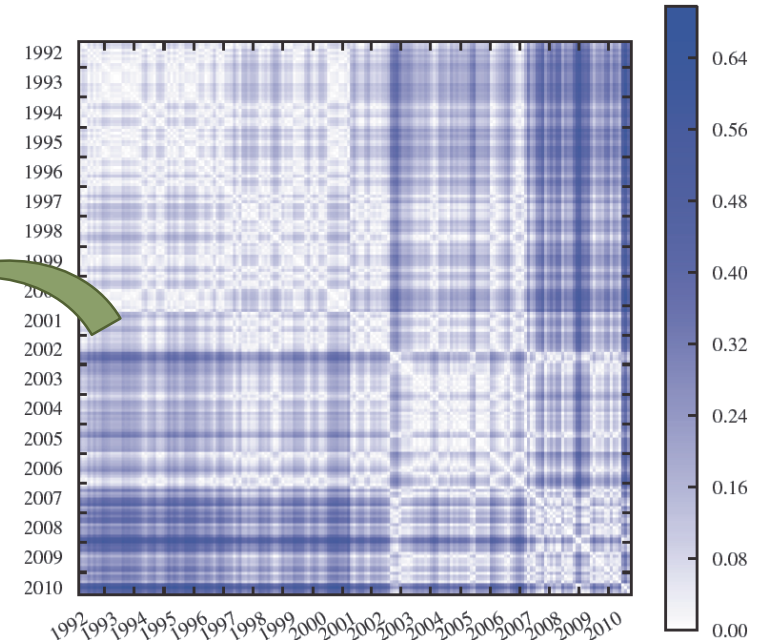
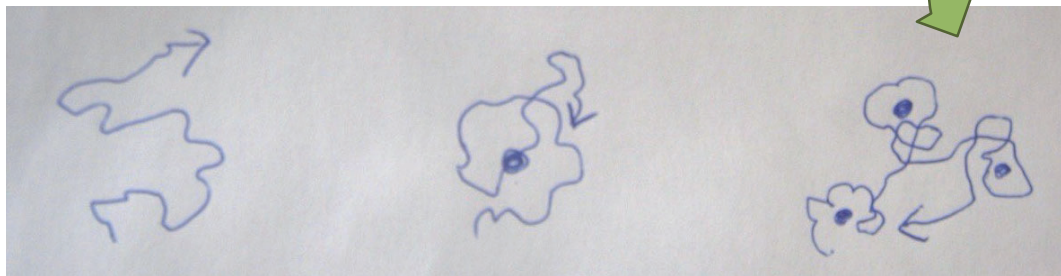
異時刻間類似度の“タータンチェック模様” とその意味



じわじわと変化している場合



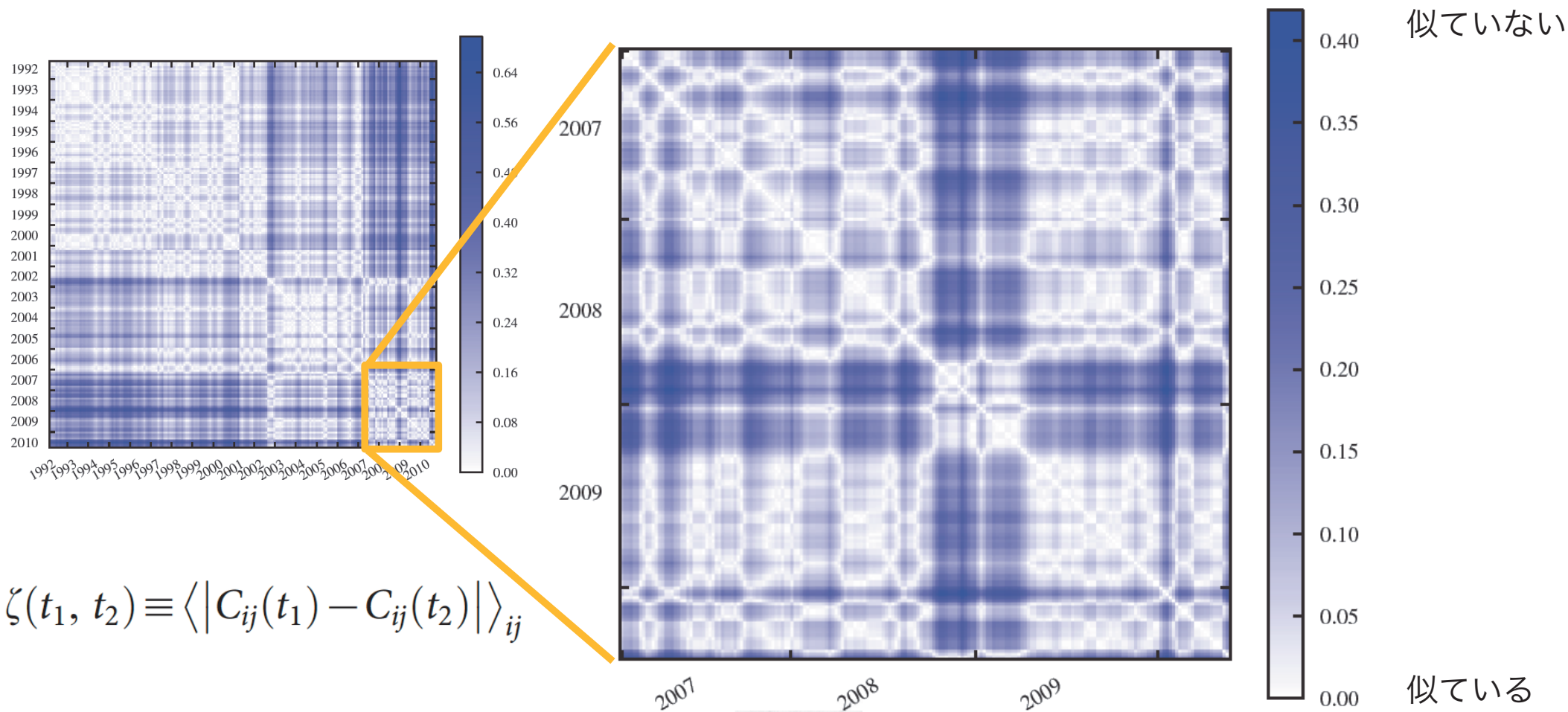
平均のまわりでゆらいている場合



実際の異時刻類似度マップ

- 時間差が小さくても、類似度が急に低くなることがある
 - 時間差の大きいペアにも似たものがある
- ➔ 典型的な相関構造（状態）というのがあるって、その間をぴよんぴよん飛び移っている

より細かい時間スケールで見ても変化は間欠的で急峻（自己相似的）



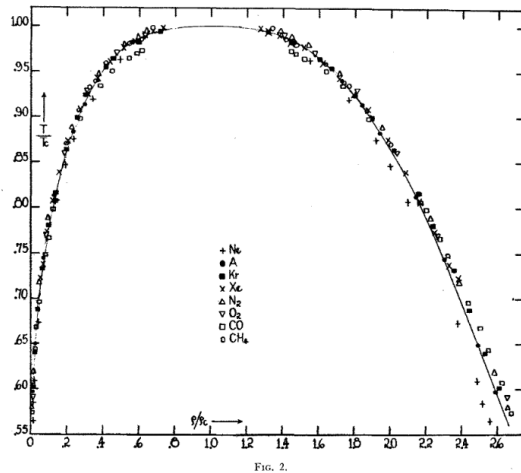
$$\zeta(t_1, t_2) \equiv \langle |C_{ij}(t_1) - C_{ij}(t_2)| \rangle_{ij}$$

M. C. Munix, T. Shimada, R. Shafer, ..., & H.E. Stanley, *Scientific Reports* vol. 2 (2012) 644.

トピックC:

社会・経済系でのフラクタルやベキ分布・普遍分布

臨界現象、カオス、フラクタル、自己組織的臨界性、...



Reproduced from E. A. Guggenheim, "The Principle of Corresponding States", J. Chem. Phys. 13, 253 (1945); Fig.2, with the permission of AIP Publishing.

<https://doi.org/10.1063/1.1724033>

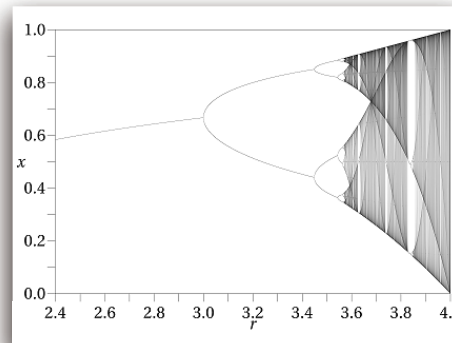


Photo by Jordan Pierce, Germany, from Wikipedia
https://en.wikipedia.org/wiki/Logistic_map#/media/File:Logistic_Bifurcation_map_High_Resolution.png
CCO

著作権等の都合により、ここに挿入されていた画像を削除しました。

砂山

<https://www.dreamstime.com/stock-photo-fist-falling-sand-pile-running-hand-as-symbol-time-running-out-isolated-white-background-image47036102>

著作権等の都合により、ここに挿入されていた画像を削除しました。

アメリカ合衆国ネバダ州ラスベガスのミート湖岸の衛星写真

https://www.flickr.com/photos/buggs_moran/4516938146

著作権等の都合により、ここに挿入されていた画像を削除しました。

Spekoski curve

David Jablonski, "The Future of the Fossil Record", Science, Vol. 284, Issue 5423, pp. 2114-2116, 25 Jun 1999 Figure 1
<https://science.sciencemag.org/content/284/5423/2114/tab-figures-data>

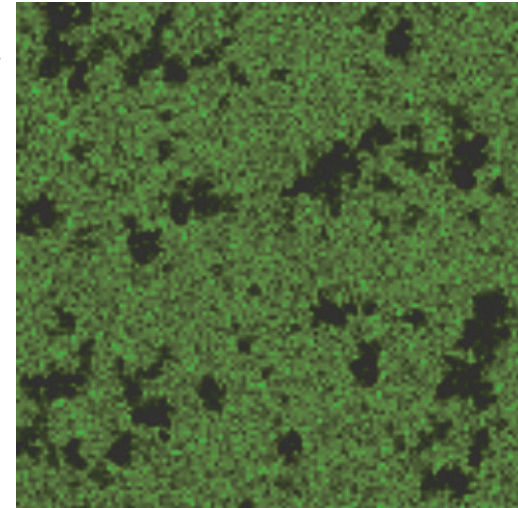


Photo by Claudio Rocchini, from Wikipedia
https://en.wikipedia.org/wiki/Forest-fire_model#/media/File:Forest_fire_model.gif
CC BY 2.5

森林火災

(https://en.wikipedia.org/wiki/Forest-fire_model)

Photo by John McColgan, from Wikipedia
https://en.wikipedia.org/wiki/Wildfire#/media/File:Deerfire_high_res.jpg
Public domain



(火災が起きると燃え広がる先の木が減る → 森林火災が大火災になるかどうかのギリギリ)

株価などの集団的動きの突発性は、自然界でもある
“勝手にギリギリの状態になる現象”（自己組織的臨界性）に似ている

著作権等の都合により、ここに挿入されていた画像を削除しました。

Ole Peters & J. David Neelin, "Critical phenomena in atmospheric precipitation", Nature Physics volume 2, pages 393–396 (2006), Figure 1

<https://www.nature.com/articles/nphys314?draft=journal>

雨の降り方

(Nature Physics Vol. 2, 393 (2006))

著作権等の都合により、ここに挿入されていた画像を削除しました。

Ole Peters & J. David Neelin, "Critical phenomena in atmospheric precipitation", Nature Physics volume 2, pages 393–396 (2006), Figure 3

<https://www.nature.com/articles/nphys314?draft=journal>

(雨が降れば水蒸気減る → 大雨が降るか降らないかのギリギリ)

人間社会（と難しい自然現象）での 各つながりの活動の特徴：バースト

著作権等の都合により、ここに挿入されていた
画像を削除しました。

色分けされたアクティビティの図

Márton Karsai, Kimmo Kaski, Albert-László
Barabási & János Kertész,
"Universal features of correlated bursty
behaviour", Scientific Reports volume 2, Article
number: 397 (2012) Fig. 1

<https://www.nature.com/articles/srep00397>

地震の活動

ニューロンの発火

携帯電話の通話

通信の頻度に関するべき分布（バースト性）

著作権等の都合により、ここに挿入されていた画像を削除しました。

Barabási, A.-L. "The origin of bursts and heavy tails in human dynamics", Nature 435 207-211 2005 Fig.2. 図aおよびb

<https://www.nature.com/articles/nature03459>

e-mail の送信間隔（左）と返信までの時間（右）の分布（Barabasi, Nature (2005)）

著作権等の都合により、ここに挿入されていた画像を削除しました。

João Gama Oliveira & Albert-László Barabási, "Darwin and Einstein correspondence patterns", Nature 437, 1251 (2005) Figure 1

<https://www.nature.com/articles/4371251a?foxtrotcallback=true>

ダーウィンとアインシュタインの手紙の頻度（Oliveira & Barabasi, Nature (2005)）

Priority queuing model

著作権等の都合により、
ここに挿入されていた画像を削除しました。

emailの送信間隔

Fig.2. (左)
Barabási, A.-L. The origin of bursts and
heavy tails in human dynamics, Nature
435 207-211 2005

[https://www.nature.com/articles/
nature03459](https://www.nature.com/articles/nature03459)

著作権等の都合により、
ここに挿入されていた画像を削除しました。

emailの送信時間と質

[https://www.slideshare.net/HangHyunJo1/
statistical-physics-of-complex-dynamics](https://www.slideshare.net/HangHyunJo1/statistical-physics-of-complex-dynamics)

Barabasi, Nature (2005)

著作権等の都合により、
ここに挿入されていた画像を削除しました。

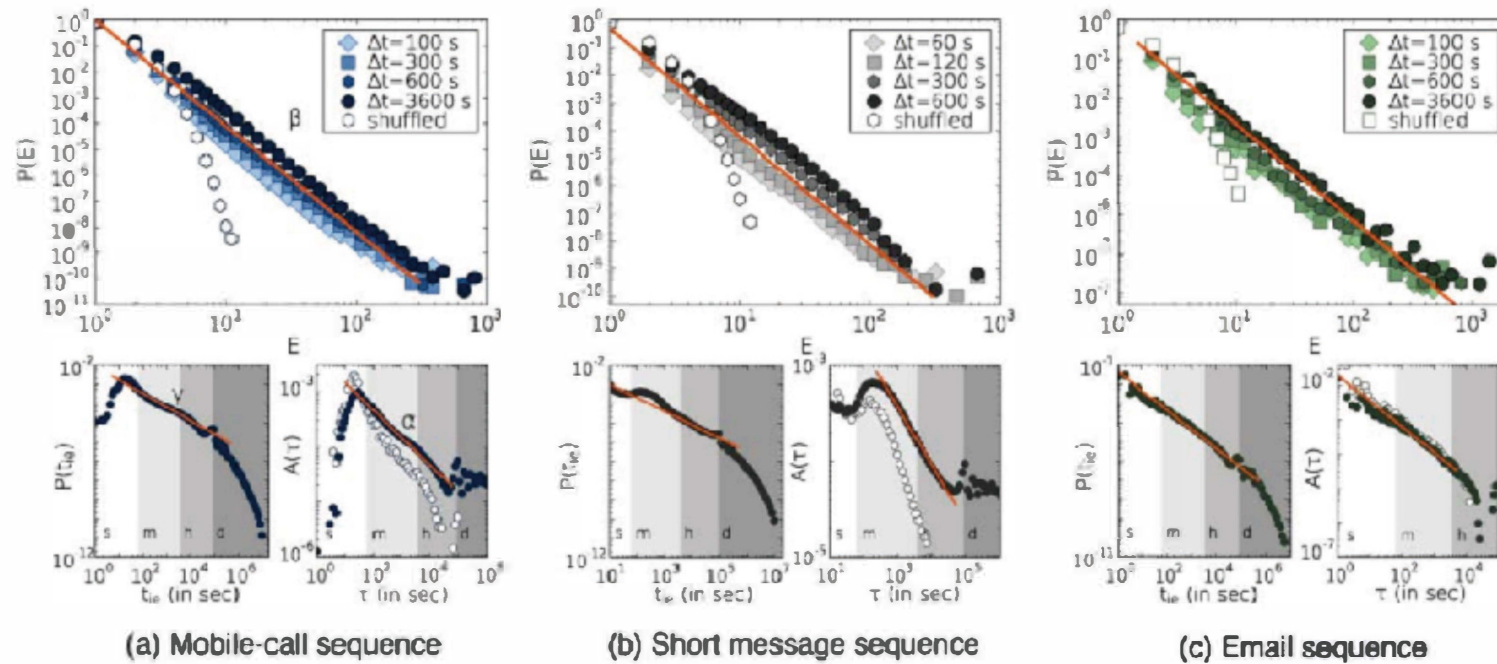
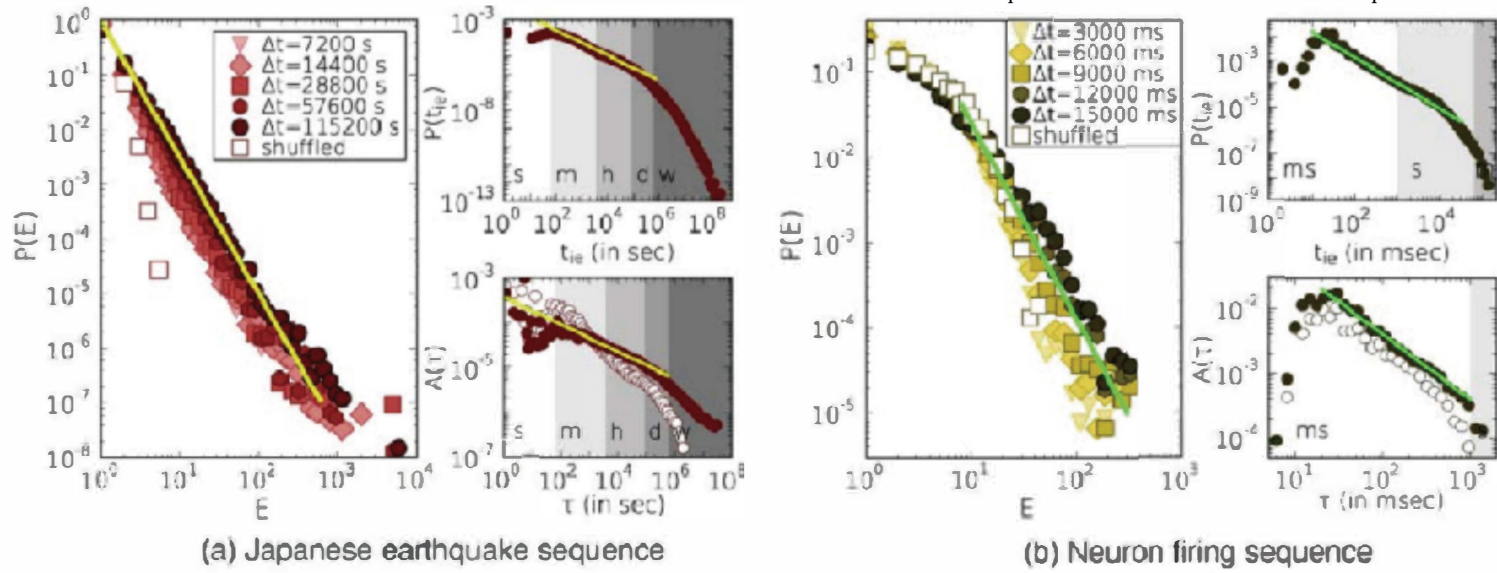
Interactive-driven collective activities:
face-to-face, mobile phone call, email, web forum イベント間の
時間分布 $P(\tau)$ の例

Fig3.2

Márton Karsai Hang-Hyun Jo Kimmo Kaski,
"Bursty Human Dynamics", Springer, Cham

[https://link.springer.com/book/10.1007%
2F978-3-319-68540-3](https://link.springer.com/book/10.1007%2F978-3-319-68540-3)

Fig. 1.
 Márton Karsai, Kimmo Kaski, Albert-László Barabási & János Kertész, "Universal features of correlated bursty behaviour", Scientific Reports volume 2, Article number: 397 (2012)
<https://www.nature.com/articles/srep00397>



Karsai et al., Sci. Rep. (2012)

社会・経済系での普遍分布・ベキ分布

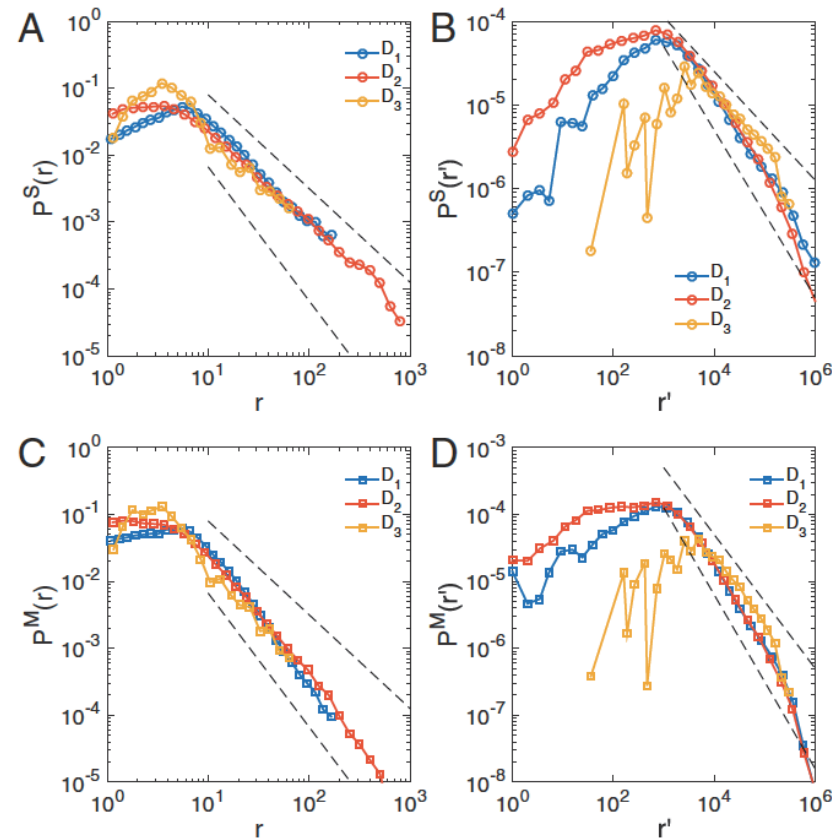


Fig.1.
Pierre Deville, Chaoming Song, Nathan Eagle, Vincent D. Blondel, Albert-László Barabási, and Dashun Wang, "Scaling identity connects human mobility and social interactions.", PNAS June 28, 2016 113 (26) 7047-7052 CC BY-NY-ND
<https://doi.org/10.1073/pnas.1525443113>

(D. Wang et al., PNAS (2016))

人の通信(A)、移動(C)もレヴィ分布 (昔：紙幣の拡散の研究)

都市のサイズと機能のスケーリング則

Luis M. A. Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B. West, Growth, innovation, scaling, and the pace of life in cities, PNAS April 24, 2007 104 (17) 7301-7306 <https://www.pnas.org/content/104/17/7301> Table1. Copyright (2007) National Academy of Sciences

著作権等の都合により、ここに挿入されていた画像を削除しました。

Tony Serino, Ph.D., Biology
Department Misericordia
University, Fuel Homeostasis
General Physiology Biology 346

<https://slideplayer.com/slide/12677525/>

Table 1. Scaling exponents for urban indicators vs. city size

Y	β	95% CI	Adj- R^2	Observations	Country-year
New patents	1.27	[1.25,1.29]	0.72	331	U.S. 2001
Inventors	1.25	[1.22,1.27]	0.76	331	U.S. 2001
Private R&D employment	1.34	[1.29,1.39]	0.92	266	U.S. 2002
"Supercreative" employment	1.15	[1.11,1.18]	0.89	287	U.S. 2003
R&D establishments	1.19	[1.14,1.22]	0.77	287	U.S. 1997
R&D employment	1.26	[1.18,1.43]	0.93	295	China 2002
Total wages	1.12	[1.09,1.13]	0.96	361	U.S. 2002
Total bank deposits	1.08	[1.03,1.11]	0.91	267	U.S. 1996
GDP	1.15	[1.06,1.23]	0.96	295	China 2002
GDP (superlinear!)	1.26	[1.09,1.46]	0.64	196	EU 1999–2003
GDP	1.13	[1.03,1.23]	0.94	37	Germany 2003
Total electrical consumption	1.07	[1.03,1.11]	0.88	392	Germany 2002
New AIDS cases	1.23	[1.18,1.29]	0.76	93	U.S. 2002–2003
Serious crimes	1.16	[1.11, 1.18]	0.89	287	U.S. 2003
Total housing	1.00	[0.99,1.01]	0.99	316	U.S. 1990
Total employment	1.01	[0.99,1.02]	0.98	331	U.S. 2001
Household electrical consumption	1.00	[0.94,1.06]	0.88	377	Germany 2002
Household electrical consumption	1.05	[0.89,1.22]	0.91	295	China 2002
Household water consumption	1.01	[0.89,1.11]	0.96	295	China 2002
Gasoline stations	0.77	[0.74,0.81]	0.93	318	U.S. 2001
Gasoline sales	0.79	[0.73,0.80]	0.94	318	U.S. 2001
Length of electrical cables	0.87	[0.82,0.92]	0.75	380	Germany 2002
Road surface	0.83	[0.74,0.92]	0.87	29	Germany 2002

Data sources are shown in *SI Text*. CI, confidence interval; Adj- R^2 , adjusted R^2 ; GDP, gross domestic product.

(動物の代謝率のスケーリング)

(Geoffrey B. West et al., PNAS (2007))

人間や社会のつながり方の特徴：“複雑ネットワーク”

Chapter 4 Section 4.5
Network Science by Albert-Laszlo Barabasi
<http://networksciencebook.com/chapter/4#scientific-impact>
Image 4.1 1

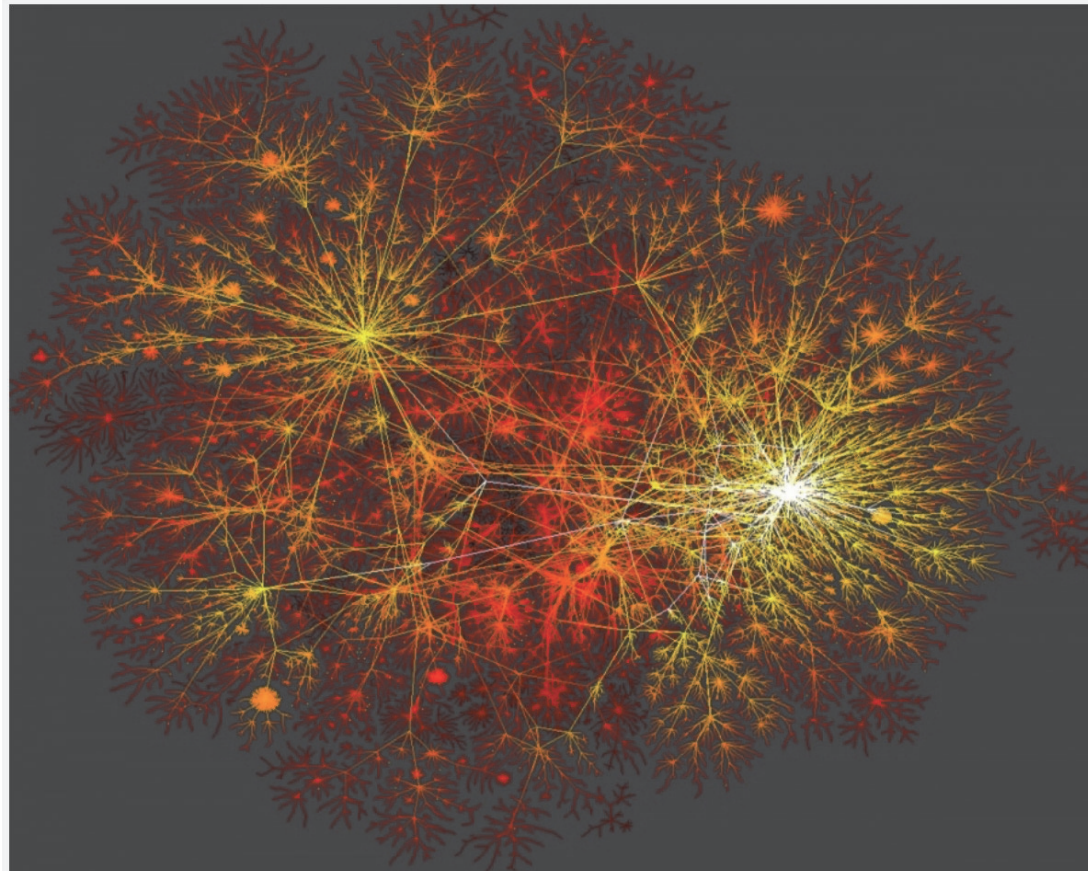
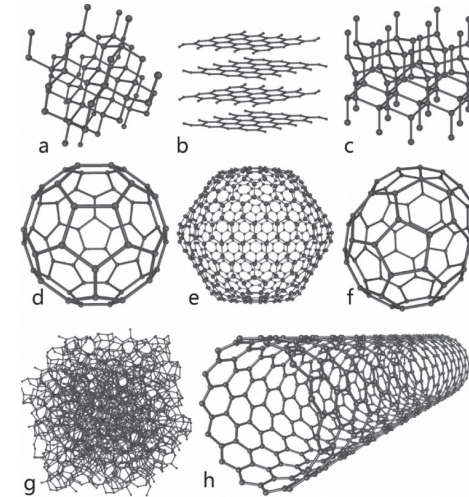


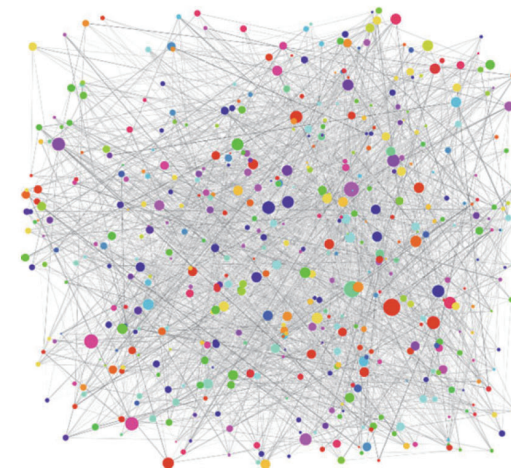
Image 4.9

The topology of the Internet

An iconic representation of the Internet topology at the beginning of the 21st century. The image was produced by CAIDA, an organization based at University of California in San Diego, devoted to collect, analyze, and visualize Internet data. The map illustrates the Internet's scale-free nature: A few highly connected hubs hold together numerous small nodes.



(物質系のネットワーク)



(ランダムネットワーク)

インターネットの構造

Chapter 4 Section 4.5
Network Science by Albert-Laszlo Barabasi
<http://networksciencebook.com/chapter/4#scientific-impact>
Image 4.9