Differential Deep Learning on Graphs and its Applications

Chengxi Zang and Fei Wang
Weill Cornell Medicine
www.calvinzang.com

This Tutorial

- www.calvinzang.com/DDLG_AAAI_2020.html
- AAAI-2020
- Friday, February 7, 2020, 2:00 PM -6:00 PM
- Sutton North, Hilton New York Midtown, NYC

This Tutorial

- **Molecular Graph Generation**: to generate novel molecules with optimized properties
  - Graph generation
  - Graph property prediction
  - Graph optimization
- **Learning Dynamics on Graphs**: to predict temporal change or final states of complex systems
  - Continuous-time network dynamics prediction
  - Structured sequence prediction
  - Node classification/regression
- ✓ **Mechanism Discovery**: to find dynamical laws of complex systems
  - Density Estimation vs. Mechanism Discovery
  - Data-driven discovery of differential equations

Part 3:
Dynamical Origins of Distribution Functions

Chengxi Zang and Fei Wang
Weill Cornell Medicine
www.calvinzang.com

A Quote from Prof. Judea Pearl

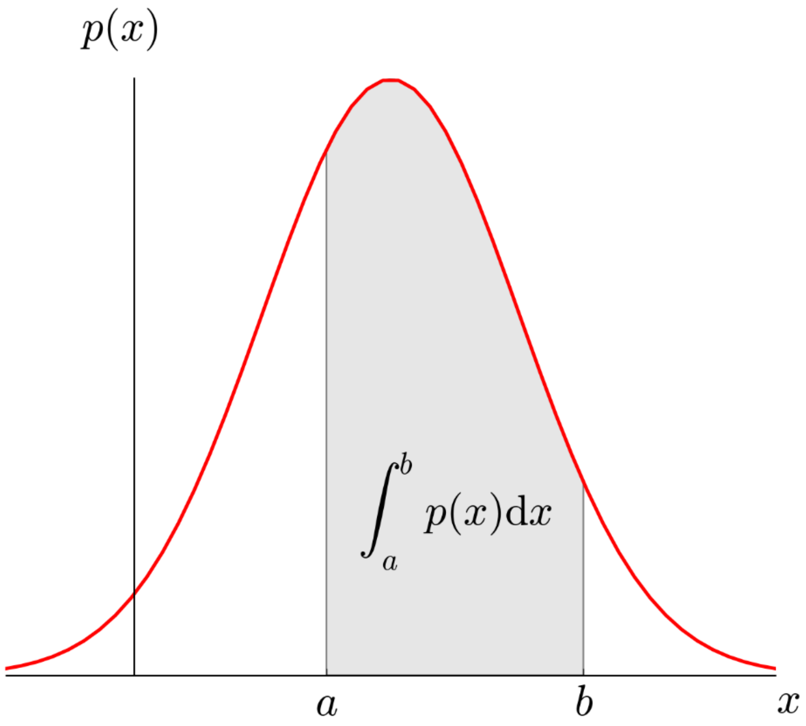
Data Science[1] lacking a model of reality[3] may be statistics[2] but hardly a science.

---- Judea Pearl

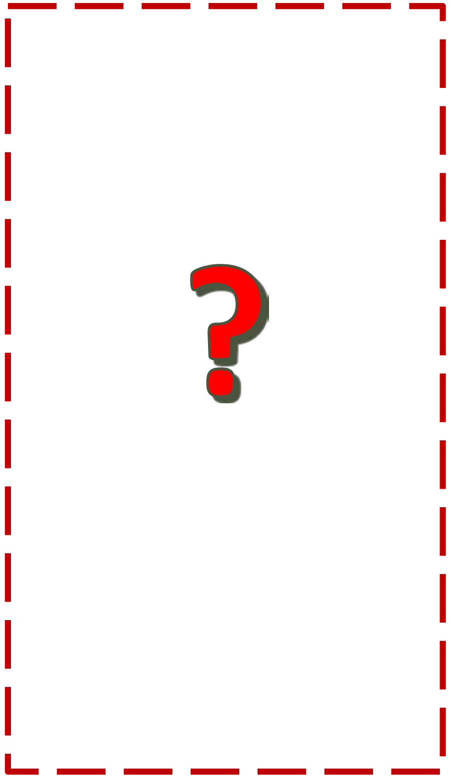
Data Science, Statistics, and Reality Models

Data: x

Statistical Models: $f(x)$

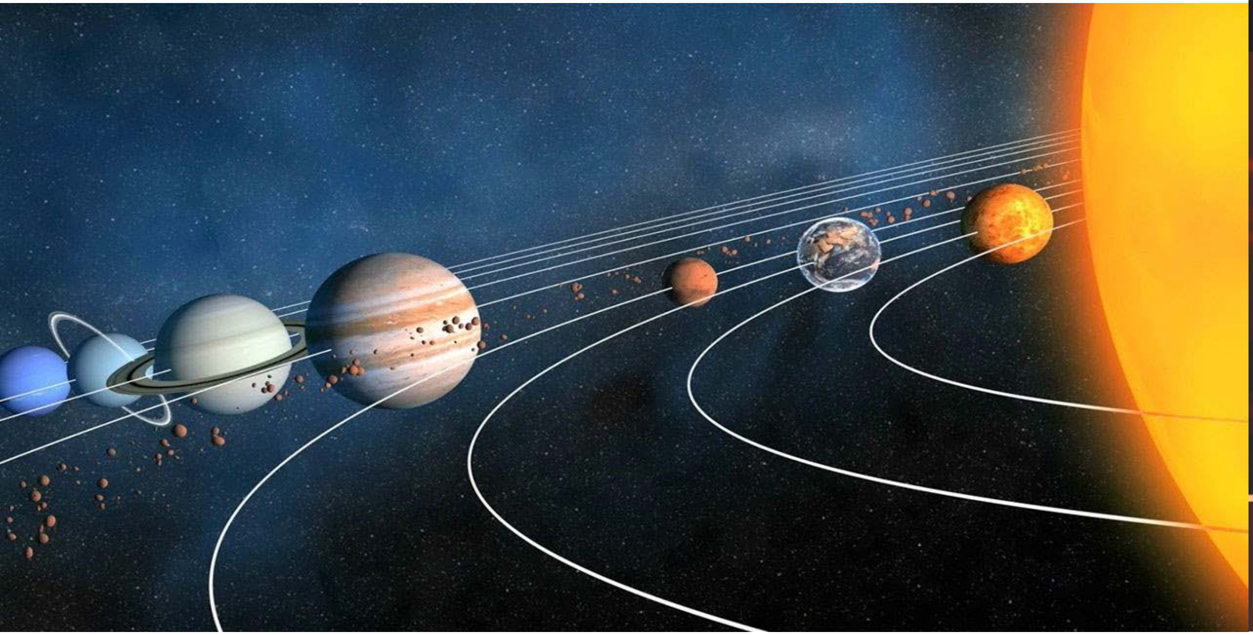


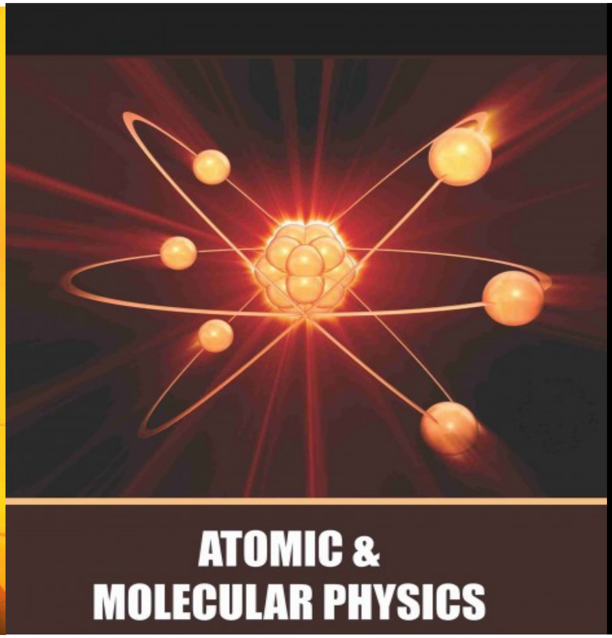
Reality Models?

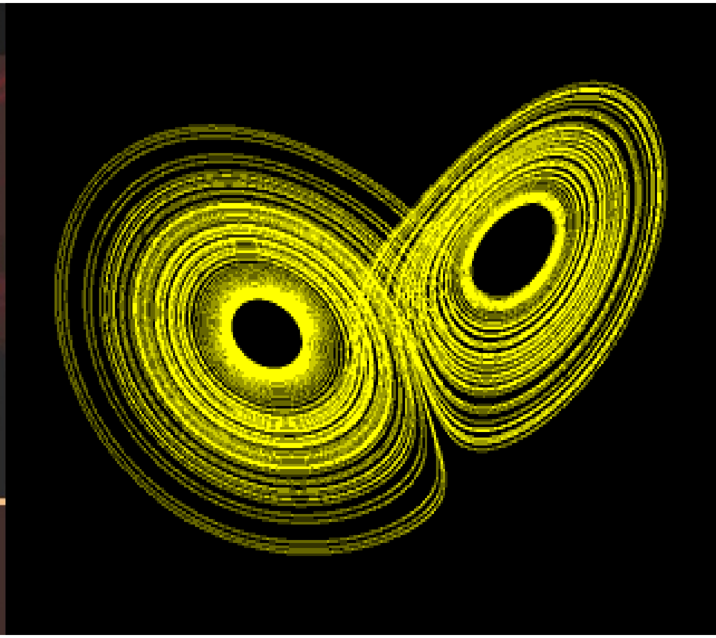


What are Reality Models?

- **Dynamical Systems by Differential Equations**







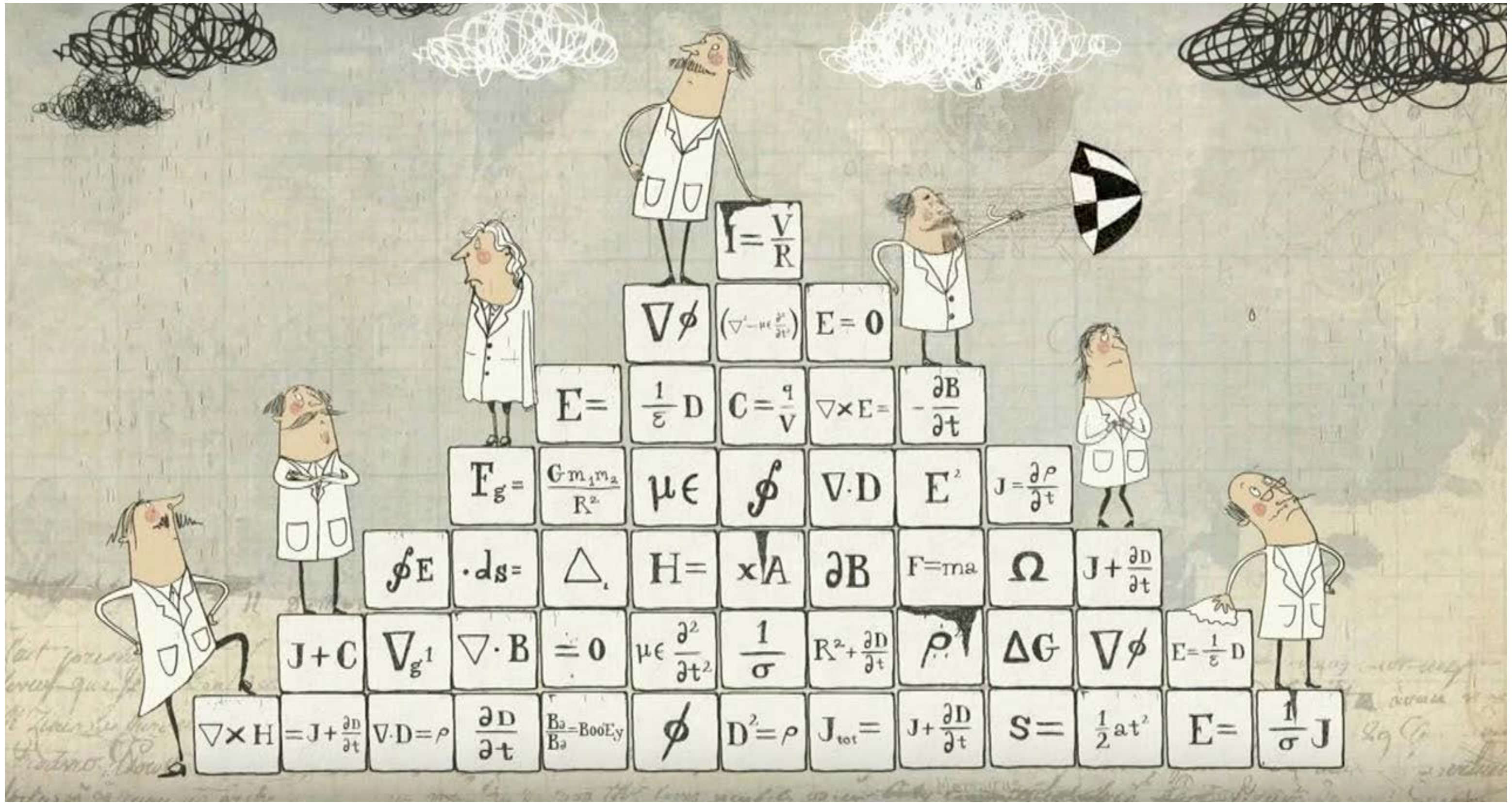
Macro: $\boldsymbol{F} = \dfrac{d(mv)}{dt}$

Micro: $i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r},t) = \left[\frac{-\hbar^2}{2m}\nabla^2 + V(\mathbf{r},t)\right]\Psi(\mathbf{r},t)$

Chaos:
$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y-x),$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y,$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z.$$

What are Reality Models?

❑Dynamical Systems by Differential Equations



Force

$$\sum \mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d(m\mathbf{v})}{dt}$$

$$\sum \mathbf{F} = m\mathbf{a} \quad \text{(Constant Mass)}$$

Acceleration

$$\mathbf{a}_{\text{average}} = \frac{\Delta \mathbf{v}}{\Delta t}$$

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{s}}{dt^2}$$

Variance

$$s^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2$$

Impulse

$$\mathbf{J} = \Delta\mathbf{p} = \int \mathbf{F}\,dt$$

$$\mathbf{J} = \mathbf{F}\Delta t \quad \text{if F is constant}$$

Velocity

$$\mathbf{v}_{\text{average}} = \frac{\Delta \mathbf{d}}{\Delta t}$$

$$\mathbf{v} = \frac{d\mathbf{s}}{dt}$$

Kinetic Energy

$$T = \frac{1}{2}mv^2$$

Gravity

$$F = \frac{Gm_1m_2}{r^2}$$

Mass Energy

$$E = mc^2$$

Density

$$\rho = \frac{m}{v}$$

Motion

$$v = v_0 + at$$

$$s = \frac{1}{2}(v_0 + v)t$$

$$s = v_0t + \frac{1}{2}at^2$$

$$v^2 = v_0^2 + 2as$$

Torque

$$\sum \boldsymbol{\tau} = \frac{d\mathbf{L}}{dt}$$

$$\sum \boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$$

Drude Law

$$\alpha = \frac{k}{\lambda^2 - \lambda_0^2}$$

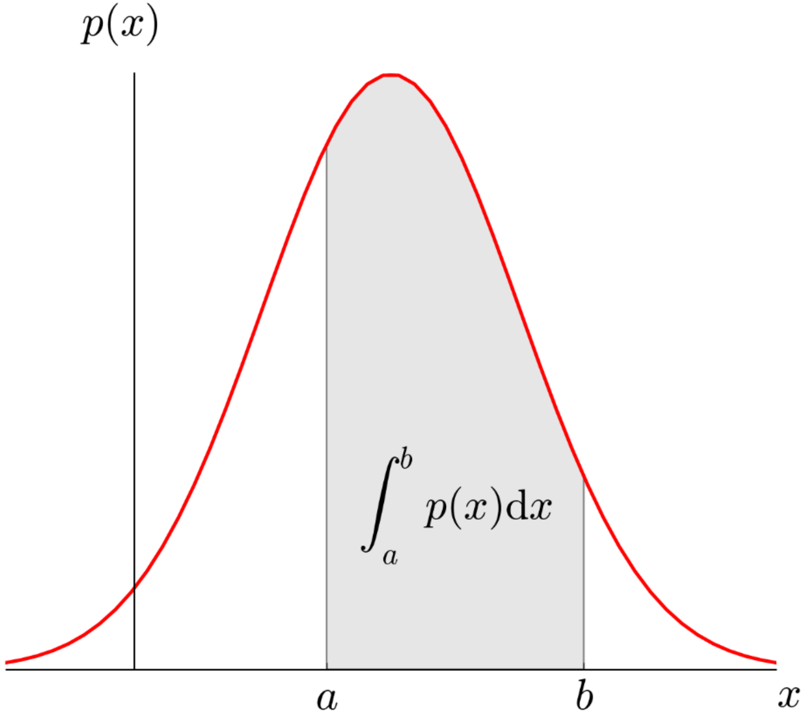
Charge

$$Q = It$$

Data Science, Statistics, and Reality Models

Data: x

Statistical Models: $f(x)$

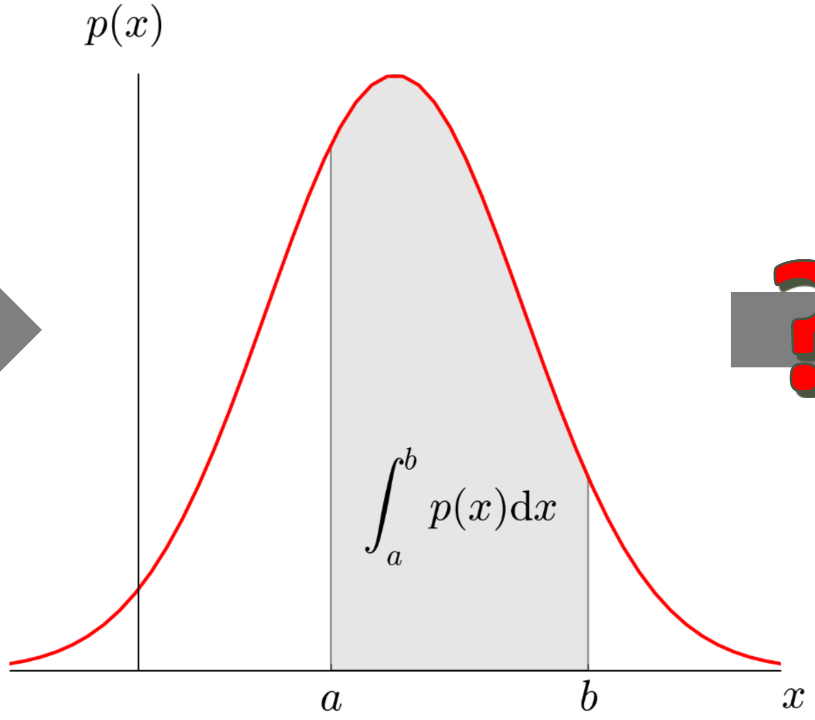


Reality Models $\frac{dx}{dt}$

Data Science Path: Data → Statistical Models → Reality Models

Data: x

Statistical Models: $f(x)$



Reality Models: $\frac{dx}{dt}$

Problem Definition

- **Data X → Statistical Model $f(X)$ → DEs $\frac{dX}{dt}$**
- **Problem 1 (Density Estimation): What are the distributions which generate the observed data samples?**
- **Problem 2 (Mechanism Discovery): What are the dynamic systems described by Differential Equations (DEs) which generate the observed distribution?**

Problem Definition

- **Data X → Statistical Model $f(X)$ → DEs $\frac{dX}{dt}$**
- **Problem 1 (Density Estimation): What are the distributions which generate the observed data samples?**
- **Problem 2 (Mechanism Discovery): What are the dynamic systems described by Differential Equations (DEs) which generate the observed distribution?**

Density Estimation: Data → Statistical Models

- **Goal: To fit complex data distribution, and to generate them**
  - Probability density function: $f_X(x)$, Cumulative density function: $F_X(x) = \int_{-\infty}^{x} f_X(r)dr$
  - Hazard function: $\lambda_X(x) = \frac{f_X(x)}{S_X(x)}$, Survival function: $S_X(x) = 1 - F_X(x)$ (Survival analysis)

Density Estimation: Data → Statistical Models

- **Non-parametric v.s. Parametric distribution model**
  - Kernel density estimation, Kaplan–Meier estimator (Survival analysis)
  - PDF/Hazard function (+ Mixture Model) + Maximum Likelihood Estimation
- **Statistical frameworks + computational power from machine learning**
  - **Autoregressive model:** $f_{X,\theta}(\boldsymbol{x}) = \prod_{i=1}^{n} f_{X,\theta}(x_i|\boldsymbol{x}_{<i})$ **(chain-rule)**
  - **Variational autoencoder** (VAE): $f_{X,\theta}(\boldsymbol{x}) = \int f_{X,\theta}(\boldsymbol{x},\boldsymbol{z})\,\mathrm{dz} = \int f_{X|Z,\theta}(\boldsymbol{x}|\boldsymbol{z}) f_{Z,\theta}(\boldsymbol{z})\,\mathrm{dz}$ **(marginal probability)**
  - **Normalizing flow:** $f_{X,\theta}(\boldsymbol{x}) = f_{Z,\theta}(Z(\boldsymbol{x}))|\mathbf{det}\frac{\partial \boldsymbol{z}_\theta}{\partial \boldsymbol{x}}|$ **(change of variable in integration)**
    - **MoFlow in tutorial part I: Generating Molecular Graphs**
  - **Generative Adversarial Networks (GAN):** likelihood-free
- **We are familiar with this step.**

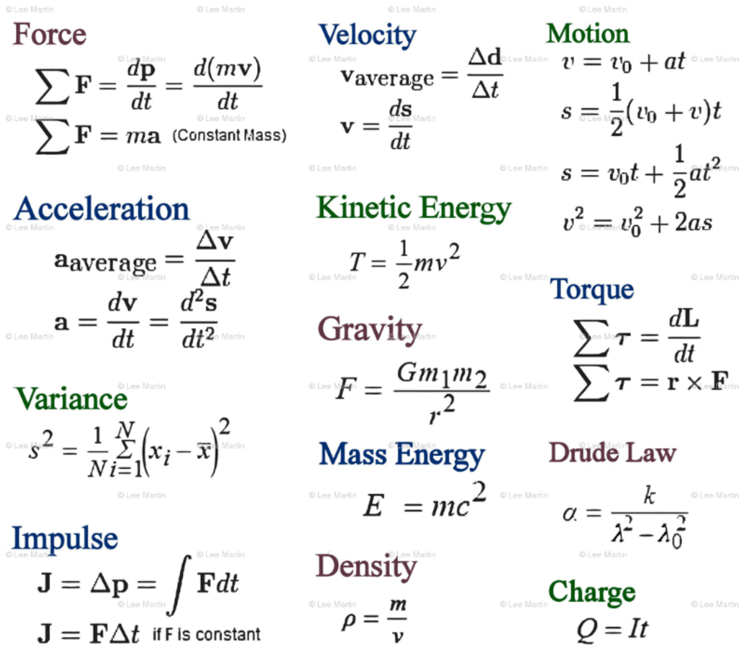
Problem Definition

- Data X → Statistical Model $f(X)$ → DEs $\frac{dX}{dt}$
- **Problem 1 (Density Estimation): What are the distributions which generate the observed data samples?**
- **Problem 2 (Mechanism Discovery): What are the dynamic systems described by Differential Equations (DEs) which generate the observed distribution?**
  - Unfamiliar in CS community

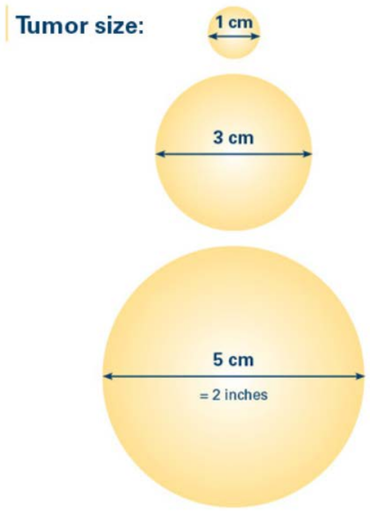
?

Why Is It Matter?

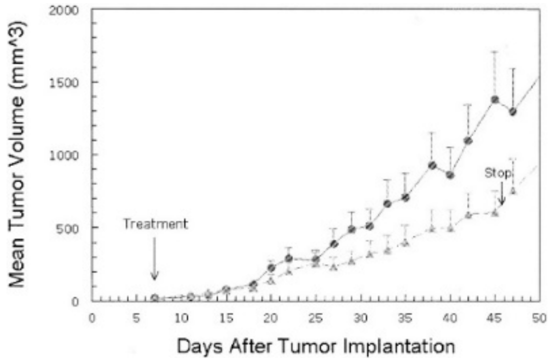
- **To understand, predict, and control real-world dynamic systems in engineering and science.**



Tumor Size



Growth Dynamics



A Major Topic in Network Science

- **Mechanism Discovery Problem: What are the dynamic systems described by Differential Equations which generate the observed distribution?**
  - E.g. Narrow-tailed distribution vs. Heavy-tailed distribution
  - How to generate power law degree distribution in random graphs

$$\frac{dX}{dt}\ ? \longrightarrow$$

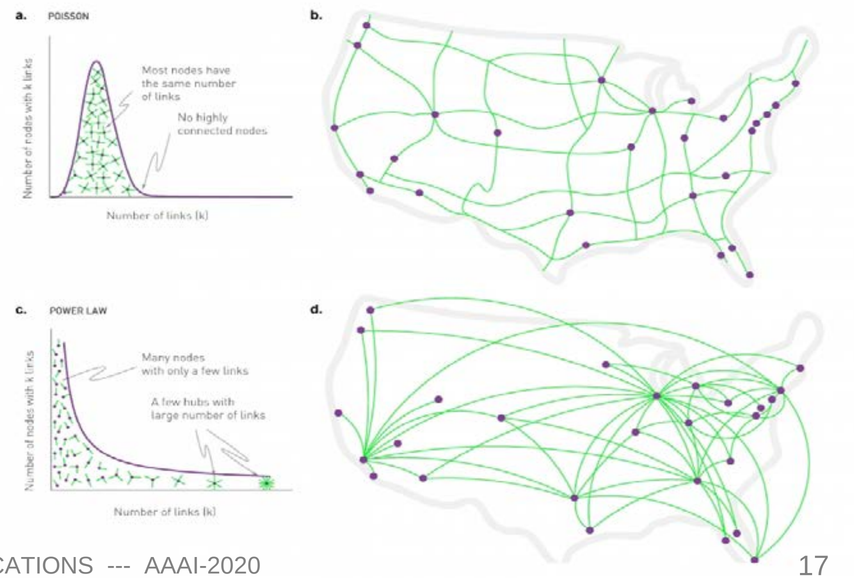
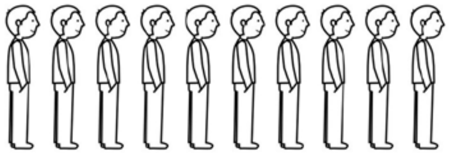


Image from http://networksciencebook.com/chapter/4#hubs

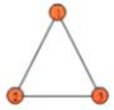
Power-law Degree Distribution in Random Graphs

[HTML] **Emergence of scaling in random networks**
AL Barabási, R Albert - science, 1999 - science.sciencemag.org
Systems as diverse as genetic networks or the World Wide Web are best described as networks with complex topology. A common property of many large networks is that the vertex connectivities follow a scale-free power-law distribution. This feature was found to be ...
Cited by 35561 Related articles All 67 versions Import into BibTeX

Constant Rate Arriving



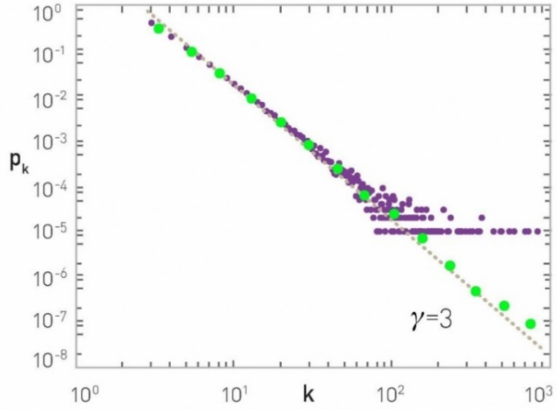
A Dynamic System



Richer-get-richer
Preferential attachment
Matthew effect

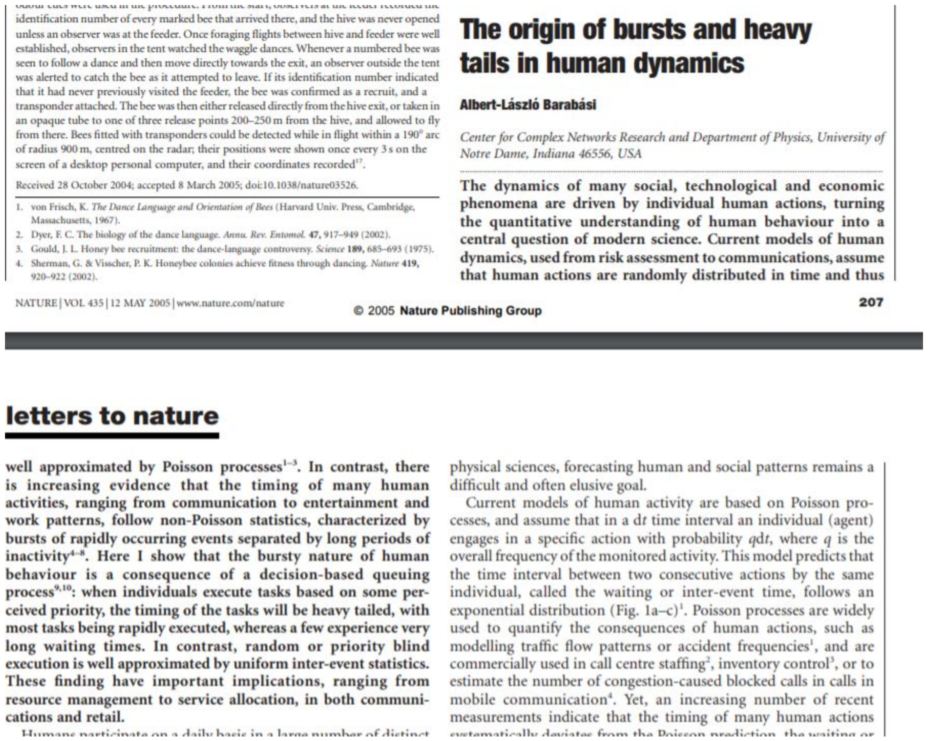
$$\frac{dk_i}{dt} = \frac{k_i}{2t-1}$$

Power-Law Distribution In Random Graph



Power-law Degree Distribution in Human Dynamics

- Burst Human behavior
- Power-law Distributions of Inter-Event times
- Human Behavior Dynamics modeled by Differential Equations


identification number of every marked bee that arrived there, and the hive was never opened unless an observer was at the feeder. Once foraging flights between hive and feeder were well established, observers in the tent watched the waggle dances. Whenever a numbered bee was seen to follow a dance and then move directly towards the exit, an observer outside the tent was alerted to catch the bee as it attempted to leave. If its identification number indicated that it had never previously visited the feeder, the bee was confirmed as a recruit, and a transponder attached. The bee was then either released directly from the hive exit, or taken in an opaque tube to one of three release points 200–250 m from the hive, and allowed to fly from there. Bees fitted with transponders could be detected while in flight within a 190° arc of radius 900 m, centred on the radar; their positions were shown once every 3 s on the screen of a desktop personal computer, and their coordinates recorded[17].

Received 28 October 2004; accepted 8 March 2005; doi:10.1038/nature03526.

1. von Frisch, K. *The Dance Language and Orientation of Bees* (Harvard Univ. Press, Cambridge, Massachusetts, 1967).
2. Dyer, F. C. The biology of the dance language. *Annu. Rev. Entomol.* **47**, 917–949 (2002).
3. Gould, J. L. Honey bee recruitment: the dance-language controversy. *Science* **189**, 685–693 (1975).
4. Sherman, G. & Visscher, P. K. Honeybee colonies achieve fitness through dancing. *Nature* **419**, 920–922 (2002).

The origin of bursts and heavy tails in human dynamics

Albert-László Barabási

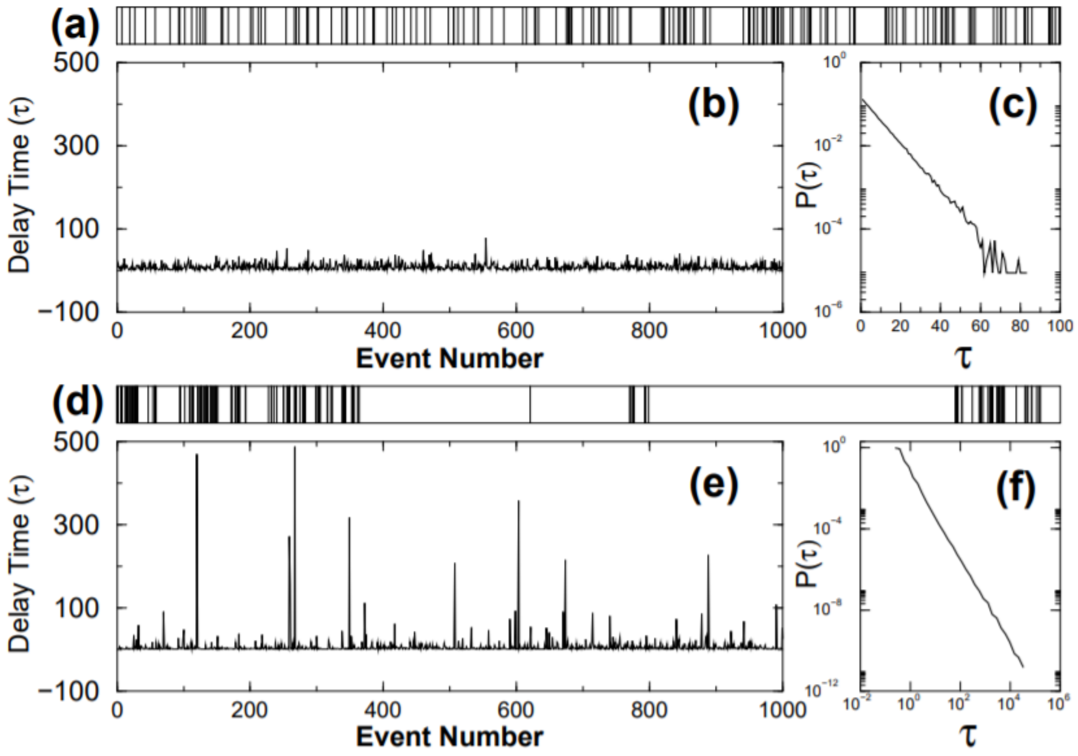
Center for Complex Networks Research and Department of Physics, University of Notre Dame, Indiana 46556, USA

The dynamics of many social, technological and economic phenomena are driven by individual human actions, turning the quantitative understanding of human behaviour into a central question of modern science. Current models of human dynamics, used from risk assessment to communications, assume that human actions are randomly distributed in time and thus

well approximated by Poisson processes[1–3]. In contrast, there is increasing evidence that the timing of many human activities, ranging from communication to entertainment and work patterns, follow non-Poisson statistics, characterized by bursts of rapidly occurring events separated by long periods of inactivity[4–8]. Here I show that the bursty nature of human behaviour is a consequence of a decision-based queuing process[9,10]: when individuals execute tasks based on some perceived priority, the timing of the tasks will be heavy tailed, with most tasks being rapidly executed, whereas a few experience very long waiting times. In contrast, random or priority blind execution is well approximated by uniform inter-event statistics. These finding have important implications, ranging from resource management to service allocation, in both communications and retail.

physical sciences, forecasting human and social patterns remains a difficult and often elusive goal.

Current models of human activity are based on Poisson processes, and assume that in a dt time interval an individual (agent) engages in a specific action with probability qdt, where q is the overall frequency of the monitored activity. This model predicts that the time interval between two consecutive actions by the same individual, called the waiting or inter-event time, follows an exponential distribution (Fig. 1a–c)[1]. Poisson processes are widely used to quantify the consequences of human actions, such as modelling traffic flow patterns or accident frequencies[1], and are commercially used in call centre staffing[2], inventory control[3], or to estimate the number of congestion-caused blocked calls in calls in mobile communication[4]. Yet, an increasing number of recent measurements indicate that the timing of many human actions



Limitations

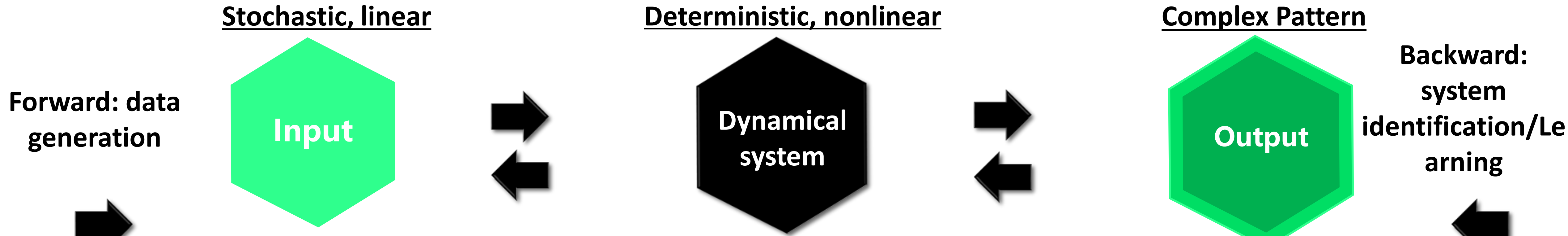
- **Network Science:**
  - Discovery: Power-law (heavy-tailed distributions) in both structure and dynamics of complex systems.
  - Case by case study: a distribution → a dynamical system with differential equations
  - Lacking a general method to find differential equations from distributions
  - Lacking a data-driven way

Our Contribution

- **A math theorem bridges any distribution and its generative dynamic system.**
  - A Dynamic System (N > 1 agents)
  - Simple and interpretable!

Dynamical Origins of Distribution Theorem

- Given **a dynamic system** $\boldsymbol{\mathcal{D}(t) = \{x_i(t) > 0 \mid \frac{dx_i(t)}{dt}, x_i(t_i) = x_0, i = 1, 2, \dots\}}$, consisting of agent i who arrives in the system at time t_i according to the **Poisson process** $\boldsymbol{\mathcal{P}(t|\lambda_P) = \{0 < t_1 \leq \cdots \leq t_i \leq \cdots \leq \mathrm{t}\}}$, the state of the i^{th} agent changes according to **differential equation** $\frac{\mathrm{dx_i(t)}}{\mathrm{dt}}$ **with initial value** $\boldsymbol{x_0}$,
- and the **cross-sectional state** of $\boldsymbol{\mathcal{D}(t)}$ at time point t, namely $\boldsymbol{x(t) = \{x_1(t), \dots, x_i(t), \dots\}}$ follows distributions $\boldsymbol{F(x(t))}$ if and only if $\boldsymbol{\frac{dx_i(t)}{dt}\Big|_{x_0} = \frac{dF^{-1}(1 - \frac{t_i}{t})}{dt}}$.



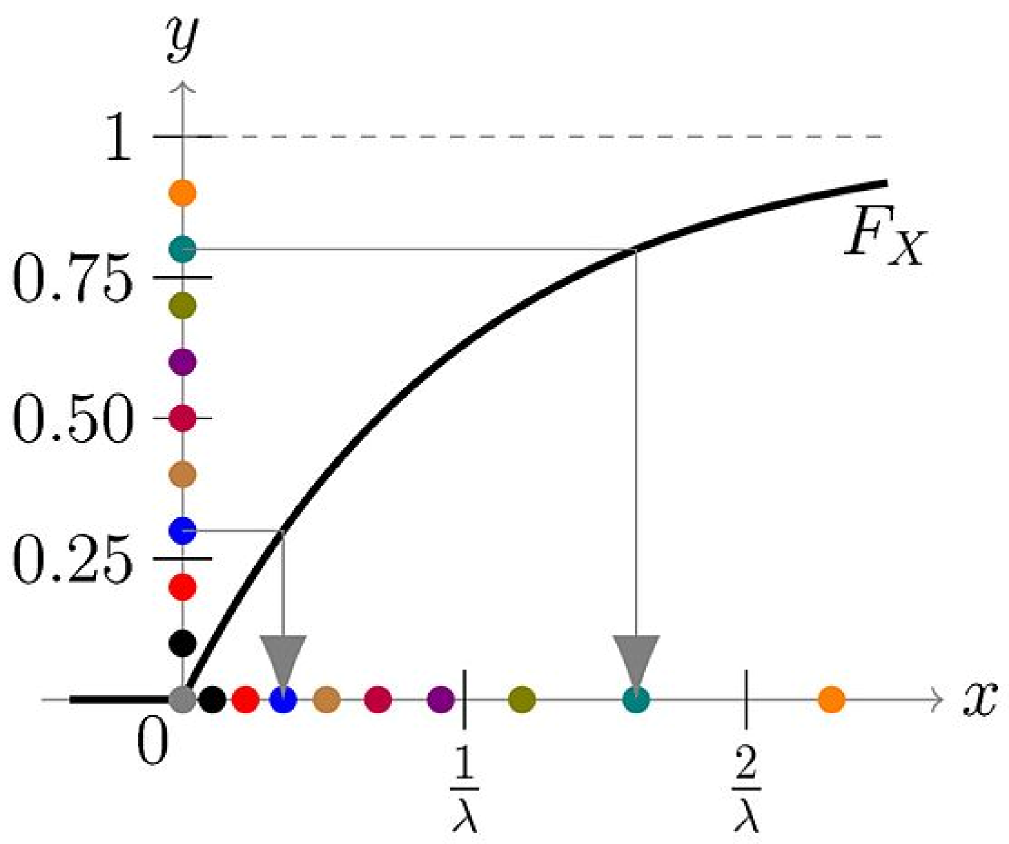
Dynamical Origins of Distribution Corollary

- **A Survival Analysis Version: e.g. for biostatistics**
  - Survival function: $S(x) = \int_{x_0}^{x} f(s)ds = 1 - F(x)$
  - Hazard function: $\lambda(x) = \lim_{\Delta x \to 0} \frac{\Pr(x \le X < x + \Delta x \mid X \ge x)}{\Delta x} = \frac{f(x)}{S(x)}$
  - Cumulative hazard function: $\Lambda(x) = \int_{x_0}^{x} \lambda(s)ds$
- **Under the same conditions, the cross-sectional state of $\mathcal{D}(t)$ at time point t, namely $\boldsymbol{x}(t) = \{\boldsymbol{x}_1(t), \ldots, \boldsymbol{x}_i(t), \ldots\}$ follows distribution $\boldsymbol{F}(\boldsymbol{x}(t))$ if and only if $\frac{d\boldsymbol{x}_i(t)}{dt} = \frac{d\Lambda^{-1}(\ln\frac{t}{t_i})}{dt}$ with initial value $\boldsymbol{x}_0$**

Proof Sketch

- **Inverse transform sampling in statistics:**

1. Generate a random number u from the standard uniform distributions $Unif\,[0,1]$.
2. Find the inverse of the desired CDF, e.g. $F_X^{-1}(x)$.
3. Compute $u = F_X^{-1}(x)$. The computed random variable X has distribution $F_X(x)$



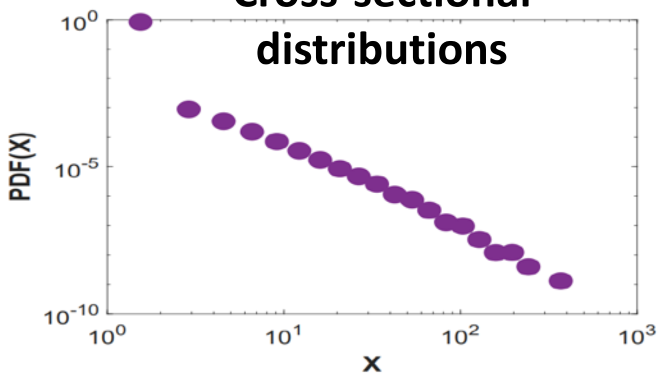
Intuition: CDF transforms complex data into uniform probability from 0 to 1. Reversing the process from distribution to data samples.

Examples

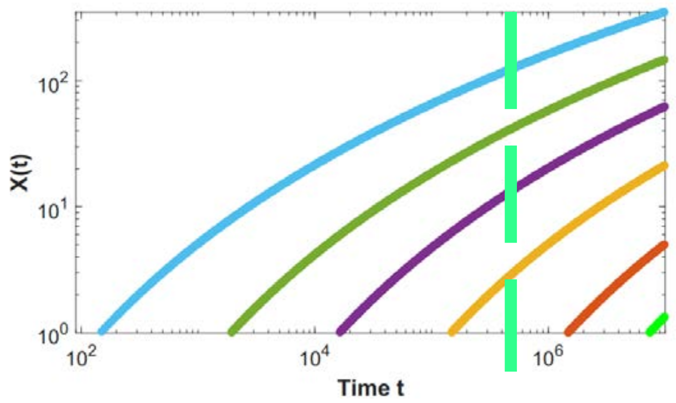
- **Power law:** $f(x) = \alpha x_0^{\alpha} x^{-(\alpha+1)}, \lambda(x) = \frac{\alpha}{x}$
- **Exponential:** $f(x) = \alpha e^{-\alpha x}, \lambda(x) = \alpha$
- **Sigmoid:** $f(x) = \frac{e^x}{(1+e^x)^2}, \lambda(x) = \frac{e^x}{1+e^x}$ (x > 0)
- **Weibull:** $f(x) = \alpha \lambda^{\alpha} x^{\alpha-1} e^{-(\lambda x)^{\alpha}}, \lambda(x) = \alpha \lambda^{\alpha} x^{\alpha-1}$

- $\frac{dx_i(t)}{dt} = \frac{x_i(t)}{\alpha t}$
- $\frac{dx_i(t)}{dt} = \frac{1}{\alpha t}$
- $\frac{dx_i(t)}{dt} = \frac{1}{t-t_0}$
- $\frac{dx_i(t)}{dt} = \frac{x_i^{1-\alpha}(t)}{\lambda^{\alpha} \alpha t}$

Cross-sectional distributions



Cross-sectional states



Deterministic dynamics

Poisson arriving

App1: Discovering Distributions'Governing ODEs

Interpretable Mechanisms: Preferential Attachment, Growth Competition, Env. Limits, Non-liearity etc.

	Data Distribution		Survival Analysis		Dynamic System		
	$f(x)$	$F(x)$	$\lambda(x)$	$\Lambda(x)$	$x_i(t)$	Dynamics $\frac{dx_i(t)}{dt}$	Interpretation
Exponential	$\alpha e^{-\alpha x}$	$1 - e^{-\alpha x}$	α	αx	$\frac{\ln(\frac{t}{t_i})}{\alpha}$	$\frac{1}{\alpha t}$	GC
Power law	$\alpha x_0^{\alpha} x^{-(\alpha+1)}$	$1 - (\frac{x_0}{x})^{\alpha}$	$\frac{\alpha}{x}$	$a \ln \frac{x}{x_0}$	$x_0(\frac{t}{t_i})^{\frac{1}{\alpha}}$	$\frac{x_i(t)}{\alpha t}$	PA + GC
Stretched exponential	$\frac{\alpha}{x^{\theta}} e^{-\frac{\alpha(x^{1-\theta}-x_0^{1-\theta})}{1-\theta}}$	$1 - e^{-\frac{\alpha(x^{1-\theta}-x_0^{1-\theta})}{1-\theta}}$	$\frac{\alpha}{x^{\theta}}$	$\frac{\alpha}{1-\theta}(x^{1-\theta} - x_0^{1-\theta})$	$[\ln(\frac{t}{t_i})\frac{1-\theta}{\alpha} + x_0^{1-\theta}]^{\frac{1}{1-\theta}}$	$\frac{x_i^{\theta}(t)}{\alpha t}$	Non-linear PA + GC
Weibull	$\alpha \lambda^{\alpha} x^{\alpha-1} e^{-(\lambda x)^{\alpha}}$	$1 - e^{-(\lambda x)^{\alpha}}$	$\alpha \lambda^{\alpha} x^{\alpha-1}$	$(\lambda x)^{\alpha}$	$\frac{(\ln \frac{t}{t_i})^{\frac{1}{\alpha}}}{\lambda}$	$\frac{x_i^{1-\alpha}(t)}{\lambda^{\alpha} \alpha t}$	Non-linear PA + GC
Log-logistic	$\frac{\lambda\alpha(\lambda x)^{\alpha-1}}{[1+(\lambda x)^{\alpha}]^2}$	$1 - \frac{1}{1+(\lambda x)^{\alpha}}$	$\frac{\lambda\alpha(\lambda x)^{\alpha-1}}{1+(\lambda x)^{\alpha}}$	$\ln[1 + (\lambda x)^{\alpha}]$	$\frac{(\frac{t}{t_i}-1)^{\frac{1}{\alpha}}}{\lambda}$	$\frac{x_i(t)}{\alpha(t-t_i)}$	PA + Since then GC
Sigmoid	$\frac{e^x}{(1+e^x)^2}$	$1 - \frac{1}{1+e^x}$	$\frac{e^x}{1+e^x}$	$\ln(1 + e^x)$	$\ln(\frac{t}{t_i} - 1)$	$\frac{1}{t-t_i}$	Since then GC
Log-normal *	$\frac{1}{x\sqrt{2\pi}} e^{-\frac{(\ln x)^2}{2}}$	$\Phi(\ln x)$	$\frac{f(x)}{1-\Phi(\ln x)}$	$-\ln[1 - \Phi(\ln x)]$	$e^{\Phi^{-1}(1-\frac{t_i}{t})}$	$x_i \frac{d\Phi^{-1}(z)}{dz} \frac{t_i}{t^2}$	PA + Square GC
Normal *	$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	$\Phi(x)$	$\frac{f(x)}{1-\Phi(x)}$	$-\ln[1 - \Phi(x)]$	$\Phi^{-1}(1 - \frac{t_i}{t})$	$\frac{d\Phi^{-1}(z)}{dz} \frac{t_i}{t^2}$	Square GC
Uniform	$\frac{1}{b-a}$	$\frac{x-a}{b-a}$	$\frac{1}{b-x}$	$\ln \frac{b-a}{b-x}$	$b - (b - a)\frac{t_i}{t}$	$\frac{b-x_i(t)}{t}$	EL + GC

App1: Dynamics for Heavy-tailed Distributions

Shared Mechanisms: Preferential Attachment for all heavy-tailed distributions

	Data Distribution		Survival Analysis		Dynamic System		
	$f(x)$	$F(x)$	$\lambda(x)$	$\Lambda(x)$	$x_i(t)$	Dynamics $\frac{dx_i(t)}{dt}$	Interpretation
Exponential	$\alpha e^{-\alpha x}$	$1-e^{-\alpha x}$	α	αx	$\frac{\ln(\frac{t}{t_i})}{\alpha}$	$\frac{1}{\alpha t}$	GC
Power law	$\alpha x_0^{\alpha} x^{-(\alpha+1)}$	$1-(\frac{x_0}{x})^{\alpha}$	$\frac{\alpha}{x}$	$a \ln \frac{x}{x_0}$	$x_0(\frac{t}{t_i})^{\frac{1}{\alpha}}$	$\frac{x_i(t)}{\alpha t}$	PA + GC
Stretched exponential	$\frac{\alpha}{x^{\theta}} e^{-\frac{\alpha(x^{1-\theta}-x_0^{1-\theta})}{1-\theta}}$	$1-e^{-\frac{\alpha(x^{1-\theta}-x_0^{1-\theta})}{1-\theta}}$	$\frac{\alpha}{x^{\theta}}$	$\frac{\alpha}{1-\theta}(x^{1-\theta}-x_0^{1-\theta})$	$[\ln(\frac{t}{t_i})\frac{1-\theta}{\alpha}+x_0^{1-\theta}]^{\frac{1}{1-\theta}}$	$\frac{x_i^{\theta}(t)}{\alpha t}$	Non-linear PA + GC
Weibull	$\alpha\lambda^{\alpha}x^{\alpha-1}e^{-(\lambda x)^{\alpha}}$	$1-e^{-(\lambda x)^{\alpha}}$	$\alpha\lambda^{\alpha}x^{\alpha-1}$	$(\lambda x)^{\alpha}$	$\frac{(\ln\frac{t}{t_i})^{\frac{1}{\alpha}}}{\lambda}$	$\frac{x_i^{1-\alpha}(t)}{\lambda^{\alpha}\alpha t}$	Non-linear PA + GC
Log-logistic	$\frac{\lambda\alpha(\lambda x)^{\alpha-1}}{[1+(\lambda x)^{\alpha}]^2}$	$1-\frac{1}{1+(\lambda x)^{\alpha}}$	$\frac{\lambda\alpha(\lambda x)^{\alpha-1}}{1+(\lambda x)^{\alpha}}$	$\ln[1+(\lambda x)^{\alpha}]$	$\frac{(\frac{t}{t_i}-1)^{\frac{1}{\alpha}}}{\lambda}$	$\frac{x_i(t)}{\alpha(t-t_i)}$	PA + Since then GC
Sigmoid	$\frac{e^x}{(1+e^x)^2}$	$1-\frac{1}{1+e^x}$	$\frac{e^x}{1+e^x}$	$\ln(1+e^x)$	$\ln(\frac{t}{t_i}-1)$	$\frac{1}{t-t_i}$	Since then GC
Log-normal *	$\frac{1}{x\sqrt{2\pi}}e^{-\frac{(\ln x)^2}{2}}$	$\Phi(\ln x)$	$\frac{f(x)}{1-\Phi(\ln x)}$	$-\ln[1-\Phi(\ln x)]$	$e^{\Phi^{-1}(1-\frac{t_i}{t})}$	$x_i\frac{d\Phi^{-1}(z)}{dz}\frac{t_i}{t^2}$	PA + Square GC
Normal *	$\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	$\Phi(x)$	$\frac{f(x)}{1-\Phi(x)}$	$-\ln[1-\Phi(x)]$	$\Phi^{-1}(1-\frac{t_i}{t})$	$\frac{d\Phi^{-1}(z)}{dz}\frac{t_i}{t^2}$	Square GC
Uniform	$\frac{1}{b-a}$	$\frac{x-a}{b-a}$	$\frac{1}{b-x}$	$\ln\frac{b-a}{b-x}$	$b-(b-a)\frac{t_i}{t}$	$\frac{b-x_i(t)}{t}$	EL + GC

App2: Discovering New Distributions from DEs

New distributions

Interpretable ODE

Generated by Exponential Dynamics	$\frac{\frac{\alpha}{t_i}}{x(\frac{\alpha}{t_i}\ln\frac{x}{x_0}+1)^2}$	$1-\frac{1}{1+\frac{\alpha}{t_i}\ln\frac{x}{x_0}}$	$\frac{\frac{\alpha}{t_i}}{x(\frac{\alpha}{t_i}\ln\frac{x}{x_0}+1)}$	$\ln[\frac{\alpha}{t_i}\ln\frac{x}{x_0}+1]$	$x_0 e^{\frac{t-t_i}{\alpha}}$	$\frac{x_i(t)}{\alpha}$	PA
Generated by Stretched Exponential Dynamics	$\frac{\frac{\alpha}{t_i^{1-\theta}}}{x[\frac{\alpha(1-\theta)}{t_i^{1-\theta}}\ln\frac{x}{x_0}+1]^{\frac{2-\theta}{1-\theta}}}$	$1-[1+\frac{\alpha(1-\theta)}{t_i^{1-\theta}}\ln\frac{x}{x_0}]^{\frac{-1}{1-\theta}}$	$\frac{\alpha}{\alpha(1-\theta)x\ln\frac{x}{x_0}+t_i^{1-\theta}x}$	$\frac{1}{1-\theta}\ln[\frac{\alpha(1-\theta)}{t_i^{1-\theta}}\ln\frac{x}{x_0}+1]$	$x_0 e^{\frac{t^{1-\theta}-t_i^{1-\theta}}{\alpha(1-\theta)}}$	$\frac{x_i(t)}{\alpha t^{\theta}}$	PA+ Non-linear GC
Generated by Sigmoid Dynamics*	$\frac{d}{dx}(\frac{\frac{\alpha}{Nt_i}\ln A}{1+\frac{\alpha}{Nt_i}\ln A})$	$1-\frac{1}{1+\frac{\alpha}{Nt_i}\ln A}$	$\frac{d}{dx}(\ln[\frac{\alpha}{Nt_i}\ln A+1])$	$\ln[\frac{\alpha}{Nt_i}\ln A+1]$	$N\frac{Be^{\frac{N(t-t_i)}{\alpha}}}{1+Be^{\frac{N(t-t_i)}{\alpha}}}$	$\frac{x_i(t)[N-x_i(t)]}{\alpha}$	PA + EL
Generated by Log Logistic Dynamics*	$\frac{\alpha(\frac{N-x_0}{x_0})^{\frac{-\alpha}{N}}x^{-\frac{\alpha}{N}-1}}{N(N-x)^{\frac{-\alpha}{N}+1}}$	$1-A^{-\frac{\alpha}{N}}$	$\frac{\frac{\alpha}{N}}{x(N-x)}$	$\frac{\alpha}{N}\ln A$	$N\frac{B(\frac{t}{t_i})^{\frac{N}{\alpha}}}{1+B(\frac{t}{t_i})^{\frac{N}{\alpha}}}$	$\frac{x_i(t)[N-x_i(t)]}{\alpha t}$	PA + EL + GC
Generated by stretched Logistic Dynamics*	$-\frac{d}{dx}[1+\frac{\alpha(1-\theta)}{Nt_i^{1-\theta}}\ln A]^{\frac{-1}{1-\theta}}$	$1-[1+\frac{\alpha(1-\theta)}{Nt_i^{1-\theta}}\ln A]^{\frac{-1}{1-\theta}}$	$\frac{d}{dx}\frac{\ln[1+\frac{\alpha(1-\theta)}{Nt_i^{1-\theta}}\ln A]}{1-\theta}$	$\frac{\ln[1+\frac{\alpha(1-\theta)}{Nt_i^{1-\theta}}\ln A]}{1-\theta}$	$\frac{NBe^{\frac{N}{\alpha}\frac{t^{1-\theta}-t_i^{1-\theta}}{1-\theta}}}{1+Be^{\frac{N}{\alpha}\frac{t^{1-\theta}-t_i^{1-\theta}}{1-\theta}}}$	$\frac{x_i(t)[N-x_i(t)]}{\alpha t^{\theta}}$	PA + EL + Non-linear GC
Generated by Confined Exponential	$\frac{\frac{\alpha}{t_i}\frac{1}{N-x}}{(1-\frac{\alpha}{t_i}\ln\frac{N-x}{N-x_0})^2}$	$1-\frac{1}{1-\frac{\alpha}{t_i}\ln\frac{N-x}{N-x_0}}$	$\frac{\frac{\alpha}{t_i}\frac{1}{N-x}}{1-\frac{\alpha}{t_i}\ln\frac{N-x}{N-x_0}}$	$\ln[1-\frac{\alpha}{t_i}\ln\frac{N-x}{N-x_0}]$	$N-\frac{N-x_0}{e^{\frac{(t-t_i)}{\alpha}}}$	$\frac{N-x_i(t)}{\alpha}$	EL
Generated by Confined Power law	$\frac{\alpha(N-x_0)^{-\alpha}}{(N-x)^{1-\alpha}}$	$1-(\frac{N-x}{N-x_0})^{\alpha}$	$\frac{\alpha}{N-x}$	$-\alpha\ln\frac{N-x}{N-x_0}$	$N-(N-x_0)(\frac{t}{t_i})^{\frac{-1}{\alpha}}$	$\frac{N-x_i(t)}{\alpha t}$	GC + EL
Generated by Confined Stretched Exponentail	$\frac{\frac{\alpha}{t_i^{1-\theta}(N-x)}}{[1-\frac{\alpha(1-\theta)}{t_i^{1-\theta}}\ln\frac{N-x}{N-x_0}]^{\frac{2-\theta}{1-\theta}}}$	$1-[1-\frac{\alpha(1-\theta)}{t_i^{1-\theta}}\ln\frac{N-x}{N-x_0}]^{\frac{-1}{1-\theta}}$	$\frac{\frac{\alpha}{t_i^{1-\theta}(N-x)}}{1-\frac{\alpha(1-\theta)}{t_i^{1-\theta}}\ln\frac{N-x}{N-x_0}}$	$\frac{\ln[1-\frac{\alpha(1-\theta)}{t_i^{1-\theta}}\ln\frac{N-x}{N-x_0}]}{1-\theta}$	$N-\frac{(N-x_0)}{e^{\frac{t^{1-\theta}-t_i^{1-\theta}}{\alpha(1-\theta)}}}$	$\frac{N-x_i(t)}{\alpha t^{\theta}}$	Non-linear GC + EL
Generated by Linear Dynamics	$\frac{\frac{\alpha}{t_i}}{[\frac{\alpha}{t_i}(x-x_0)+1]^2}$	$1-\frac{1}{\frac{\alpha}{t_i}(x-x_0)+1}$	$\frac{1}{(x-x_0)+\frac{t_i}{\alpha}}$	$\ln[\frac{\alpha}{t_i}(x-x_0)+1]$	$x_0+\frac{t-t_i}{\alpha}$	$\frac{1}{\alpha}$	Constant rate

App3: New Statistical Models: One DE, Many Distributions

- ❑ **A showcased example: A new statistical tool to generate/fit many complex multiscale distributions:**

ODE

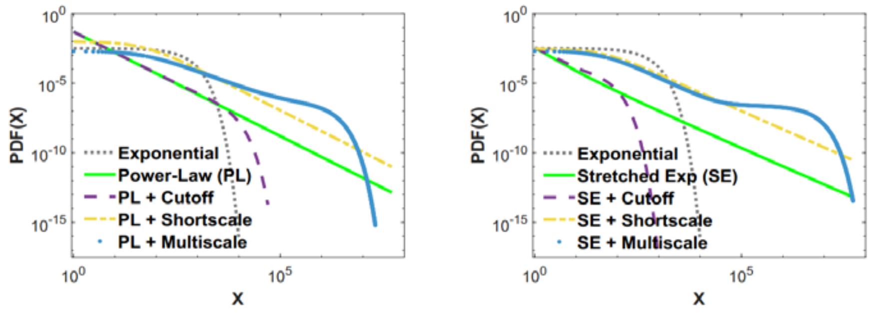
$$\frac{dx_i(t)}{dt} = \frac{(x_i(t) + \Delta)^{\theta}}{\beta (x_i(t) + \Delta)^{\theta} t + \alpha t}$$

Hazard func.

$$\lambda(x) = \beta + \alpha (x + \Delta)^{-\theta}$$

Distribution Family

Complex multiscale distributions



(a) $\theta = 1$ (b) $\theta \neq 1$

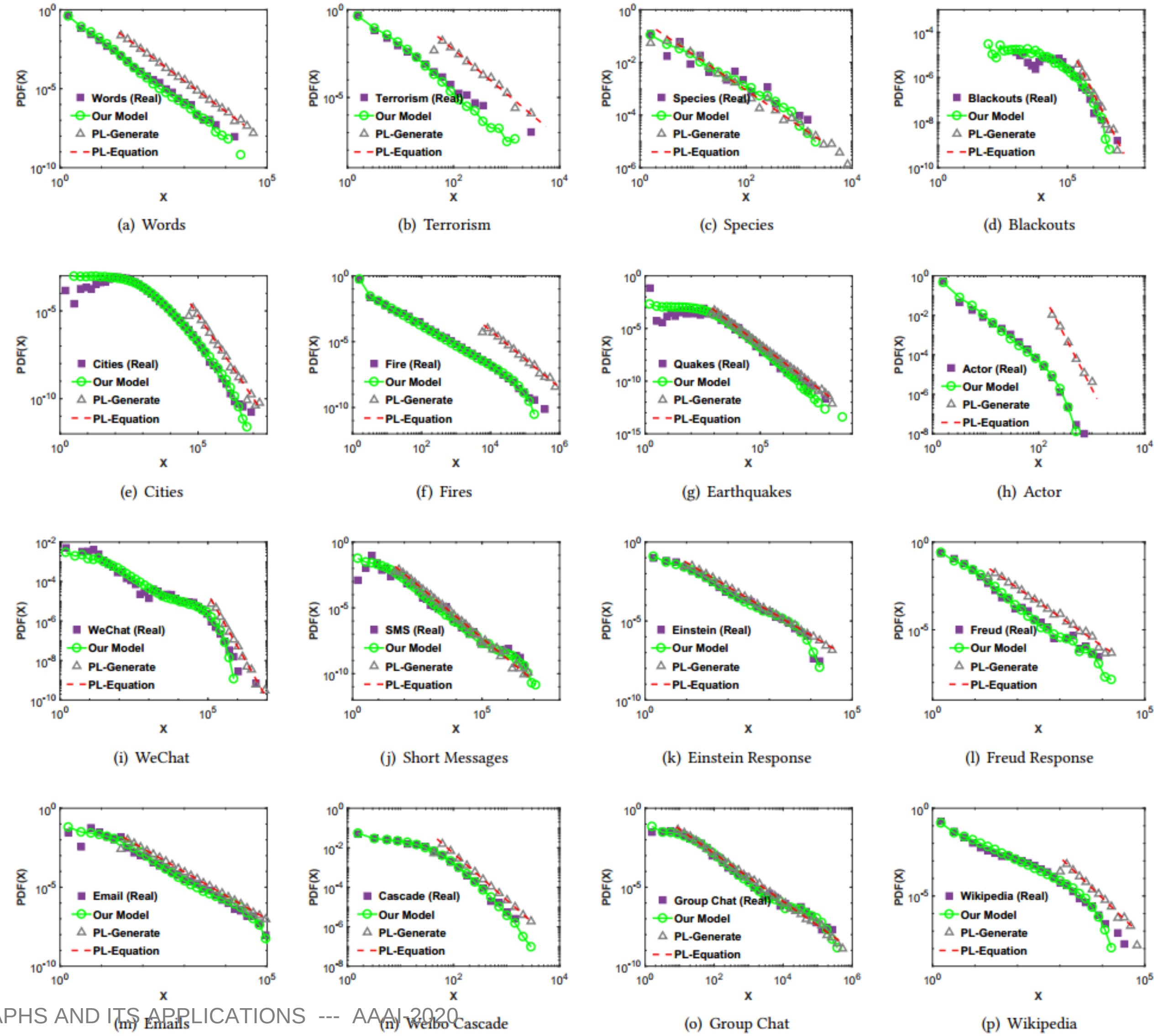
Capability	Exponential	Power law	Power law + cutoff *	Power law + Shortscale	Power law + Multiscale
PDF ($\theta = 1$)	$\beta e^{-\beta x}$	$\alpha \Delta^{\alpha} x^{-(\alpha+1)}$	$\alpha \Delta^{\alpha} x^{-(\alpha+1)} e^{-\beta x}$	$\alpha \Delta^{\alpha} (x + \Delta)^{-(\alpha+1)}$	$(\beta + \frac{\alpha}{x+\Delta})(\frac{x}{\Delta} + 1)^{-\alpha} e^{-\beta x}$
Hazard rate	β	$\frac{\alpha}{x}$	$\beta + \frac{\alpha}{x}$	$\frac{\alpha}{x+\Delta}$	$\beta + \frac{\alpha}{x+\Delta}$
Our model	✓	✓	✓	✓	✓
Capability	**Exponential**	**Stretched exponential ****	**Stretched exponential + Cutoff ***	**Stretched exponential + Shortscale**	**Stretched exponential +Multiscale**
PDF ($\theta \neq 1$)	$\alpha e^{-\alpha x}$	$\alpha x^{-\theta} e^{-\frac{\alpha}{1-\theta} x^{1-\theta}}$	$\alpha x^{-\theta} e^{-\frac{\alpha}{1-\theta} x^{1-\theta} - \beta x}$	$\alpha (x + \Delta)^{-\theta} e^{-\frac{\alpha}{1-\theta}[(x+\Delta)^{1-\theta} - \Delta^{1-\theta}]}$	$[\beta + \alpha (x + \Delta)^{-\theta}] e^{-\beta x - \frac{\alpha}{1-\theta}[(x+\Delta)^{1-\theta} - \Delta^{1-\theta}]}$
Hazard rate	α	$\frac{\alpha}{x^{\theta}}$	$\beta + \frac{\alpha}{x^{\theta}}$	$\frac{\alpha}{(x+\Delta)^{\theta}}$	$\beta + \frac{\alpha}{(x+\Delta)^{\theta}}$
Our model	✓	✓	✓	✓	✓

* For the Power law distribtuion with cutoff case and Stretched exponential distribution with cutoff case, the probability density functions of which are derived approximately by the harzard rates. Refer to the Model Section.

** Special case is approximately Normal Distribution when $\theta = -1$. Refer to the Model Section.

App3: New Statistical Models: One ODE, Many Distributions

- ❑ **One Equation, Many Distributions**
- ❑ **More robust statistical tools to fit empirical data, existing methods show large bias**
  - o Baseline: statistical tool to fit heavy-tailed distribution in



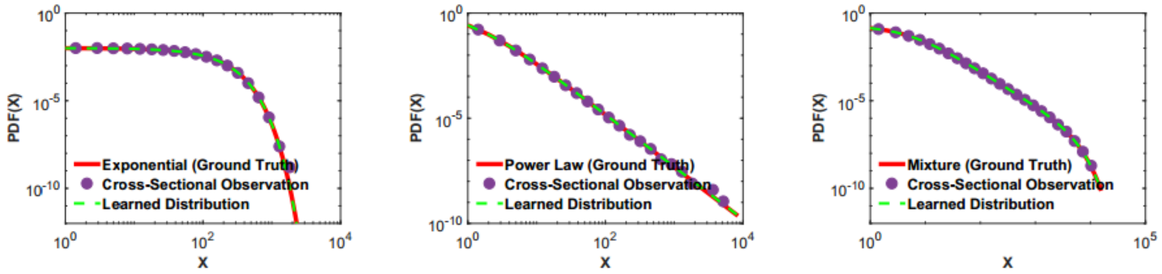
(a) Words (b) Terrorism (c) Species (d) Blackouts (e) Cities (f) Fires (g) Earthquakes (h) Actor (i) WeChat (j) Short Messages (k) Einstein Response (l) Freud Response (m) Emails (n) Weibo Cascade (o) Group Chat (p) Wikipedia

Clauset et al. 2009. [Power-Law Distributions in Empirical Data.] *SIAM Rev.*.

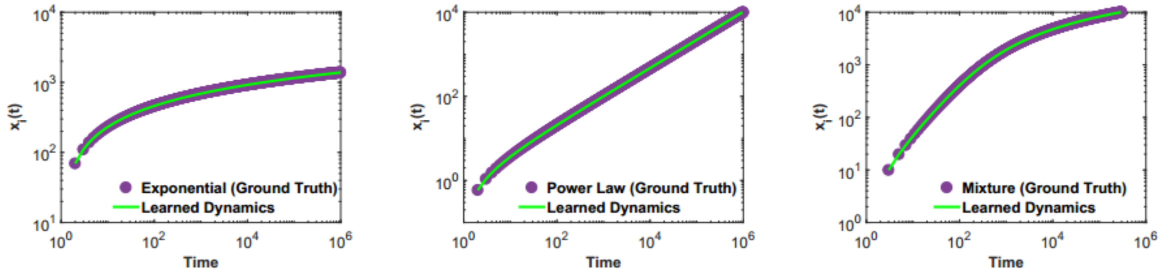
App3: Data-Driven learning ODEs

- ❑ **Fit Distribution from Samples, and Learn Dynamics**
- ❑ **Synthetic Data**

	Dynamics $\frac{dx_i(t)}{dt}\|_{x_0}$	PDF $f(x)$	Parameters
Exponential	$\frac{1}{\beta t}$	$\beta e^{\beta x}$	$\beta = 0.01$
Power Law	$\frac{x_i(t)+\Delta}{\alpha t}$	$\alpha \Delta^{\alpha} x^{-(\alpha+1)}$	$\alpha = 1.5$ $\Delta = 1$
Mix model	$\frac{x_i(t)+\Delta}{\beta(x_i(t)+\Delta)t+\alpha t}$	$\beta e^{-\beta x}(\frac{x}{\Delta}+1)^{-\alpha}$ $+\frac{\alpha}{\Delta}(\frac{x}{\Delta}+1)^{-(\alpha+1)} e^{-\beta x}$	$\beta = 5e{-4}$ $\alpha = 1$ $\Delta = 5$



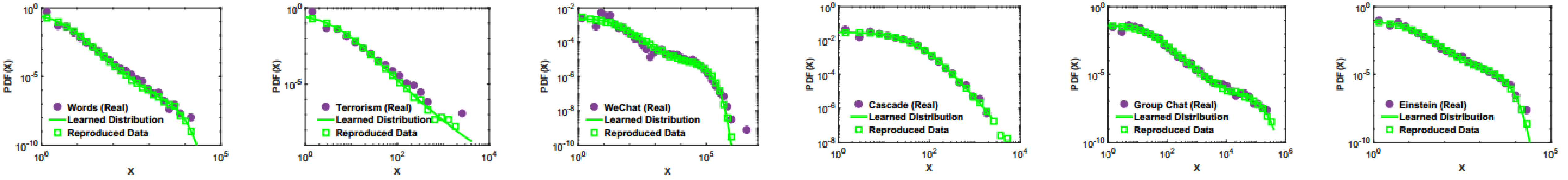
(a) Exp. Distribution (b) PL. Distribution (c) Mix. Distribution



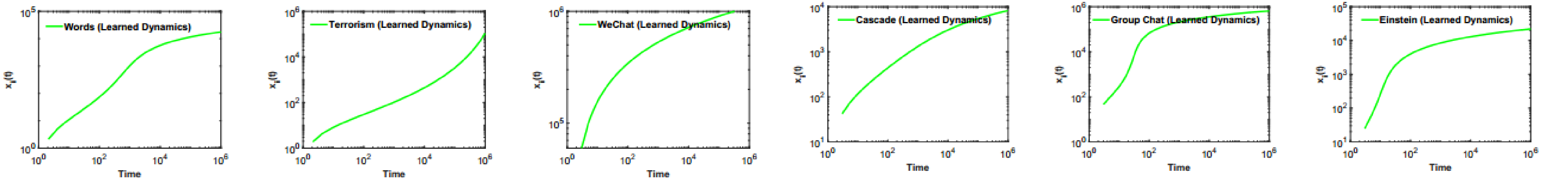
(d) Exp. Generative Dynamics (e) PL. Generative Dynamics (f) Mix. Generative Dynamics

App3: Data-Driven learning ODEs

- Empirical Data → Empirical distributions → Empirical ODEs



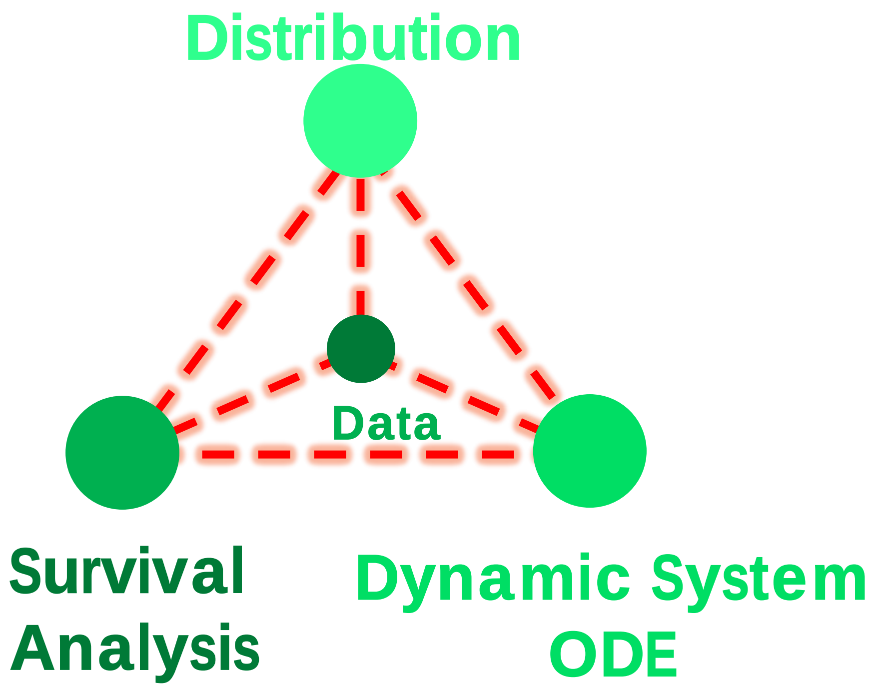
(a) Words $f(x)$ (b) Terrorism $f(x)$ (c) WeChat $f(x)$ (d) Cascade $f(x)$ (e) Group Chat $f(x)$ (f) Einstein $f(x)$



(g) Words $\frac{dx_i(t)}{dt}$ (h) Terrorism $\frac{dx_i(t)}{dt}$ (i) WeChat $\frac{dx_i(t)}{dt}$ (j) Cascade $\frac{dx_i(t)}{dt}$ (k) Group Chat $\frac{dx_i(t)}{dt}$ (l) Einstein $\frac{dx_i(t)}{dt}$

Summary

- ❑ A theorem constructing dynamic systems described by Differential Equations which generate the observed distribution
- ❑ Discovery many new DEs and distributions
- ❑ Learning many multiscale distributions by one differential equation
- ❑ A framework to connect these dots



This Tutorial

- ❑ **Molecular Graph Generation**: to generate novel molecules with optimized properties
  - o Graph generation
  - o Graph property prediction
  - o Graph optimization
- ❑ **Learning Dynamics on Graphs**: to predict temporal change or final states of complex systems
  - o Continuous-time network dynamics prediction
  - o Structured sequence prediction
  - o Node classification/regression
- ☑ **Mechanism Discovery**: to find dynamical laws of complex systems
  - o Density Estimation vs. Mechanism Discovery
  - o Data-driven discovery of differential equations

Part 3: Dynamical Origins of Distribution Functions

Chengxi Zang and Fei Wang

Weill Cornell Medicine

www.calvinzang.com