Are human interactivity times **Lognormal?**

Norbert Blenn & Piet Van Mieghem

Conference on Complex Systems, Amsterdam, 19-22 September 2016

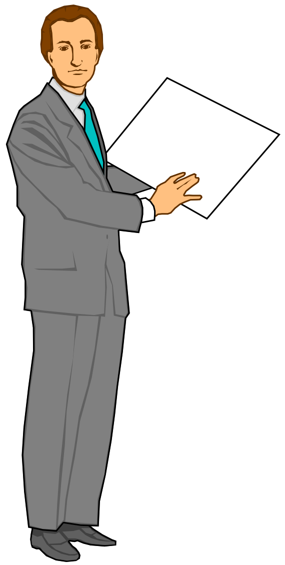
Workshop on Burstiness in Human Behavior and Other Natural Phenomena

COMPLEX SYSTEMS SOCIETY

NAS

TUDelft

Outline



Human interactivity times

Lognormal distribution

Binning and σ

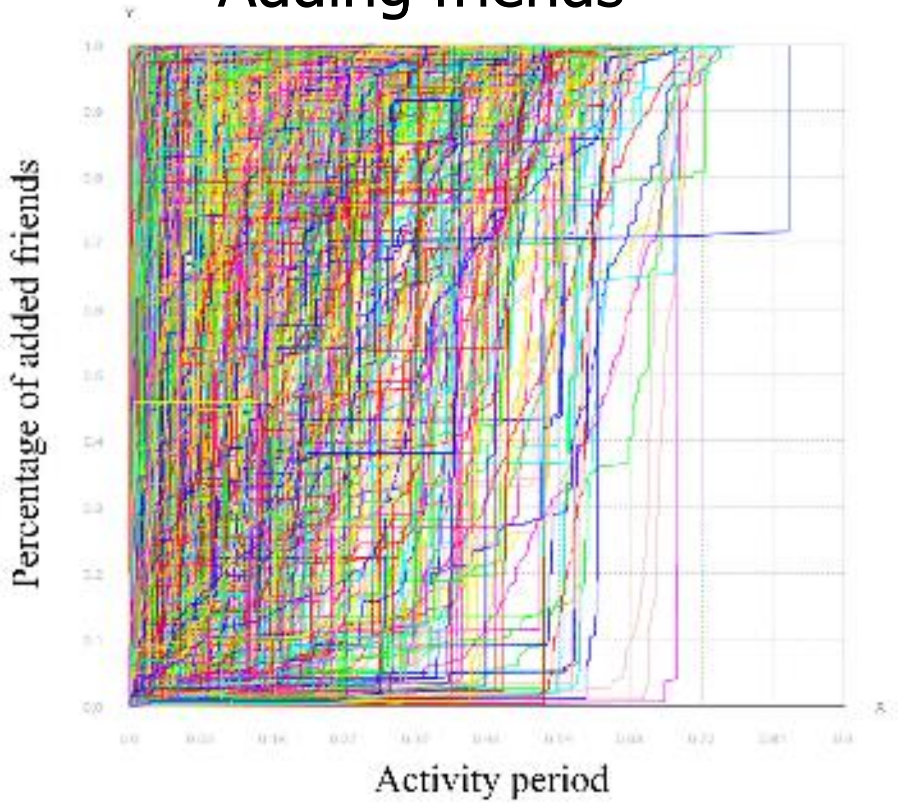
Human interactivity times

- Duration between two human, consecutive talks or events:
  - Sending emails, receiving emails
  - Collecting friends and followers
  - Writing comments in online social networks (OSN)
- Measurability constraints:
  - All human activities that happen, when “logged in” (being connected to the Internet)
- Any measured time has a minimum value (due to the constant speed of light)

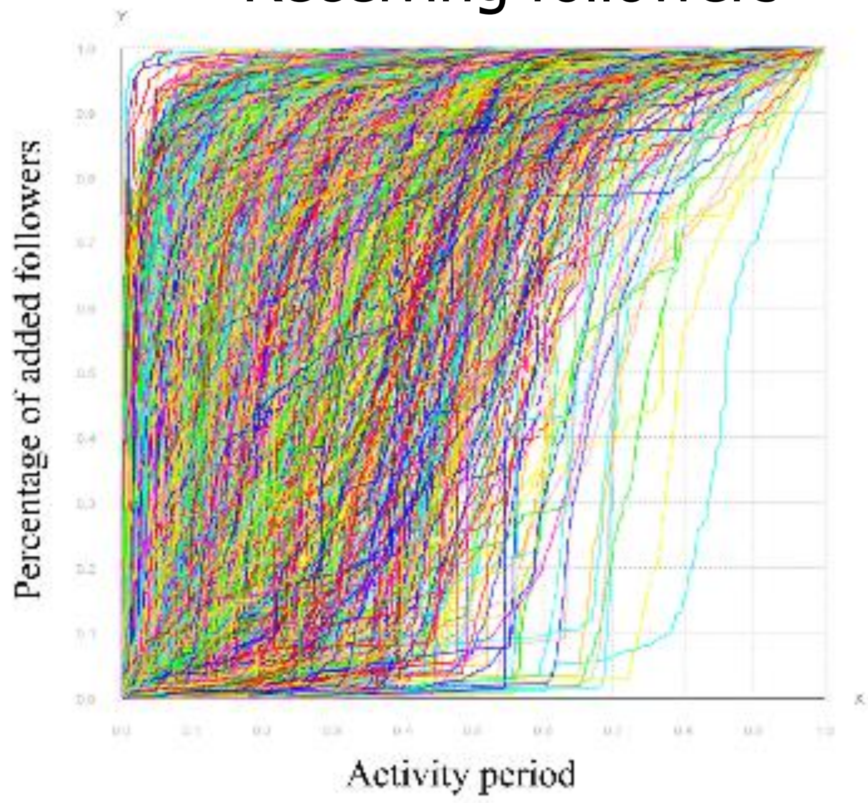
Blenn, N. and P. Van Mieghem, 2016, "Are human interactivity times lognormal?", Delft University of Technology, report20160711 (http://arxiv.org/abs/1607.02952).

Humans act in bursts in Digg.com

Adding friends



Receiving followers

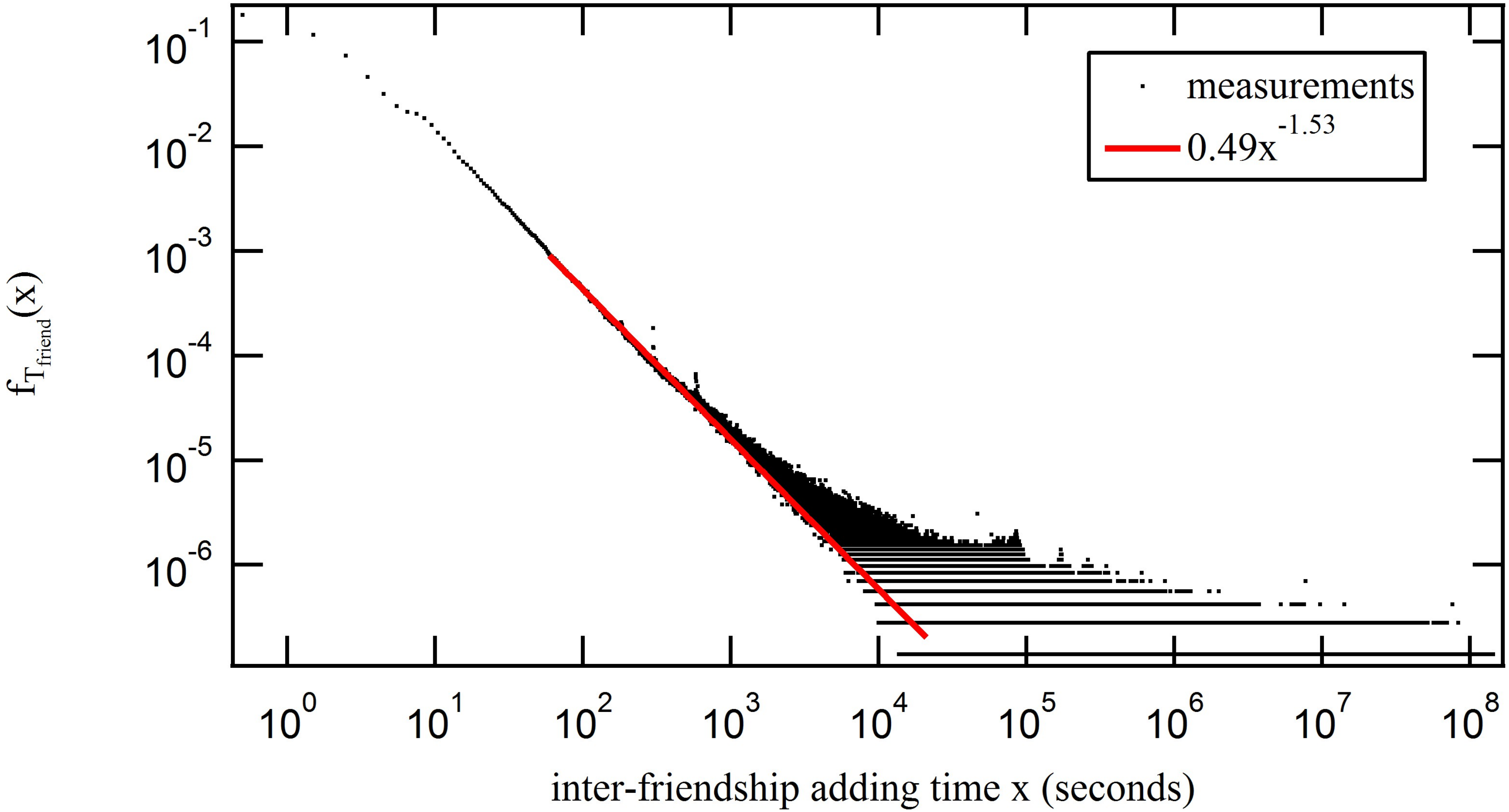


Steep “stair-like” shapes: high activity in short time intervals, followed by long durations of in-activity

Smooth “S-shaped” curves

TU Delft

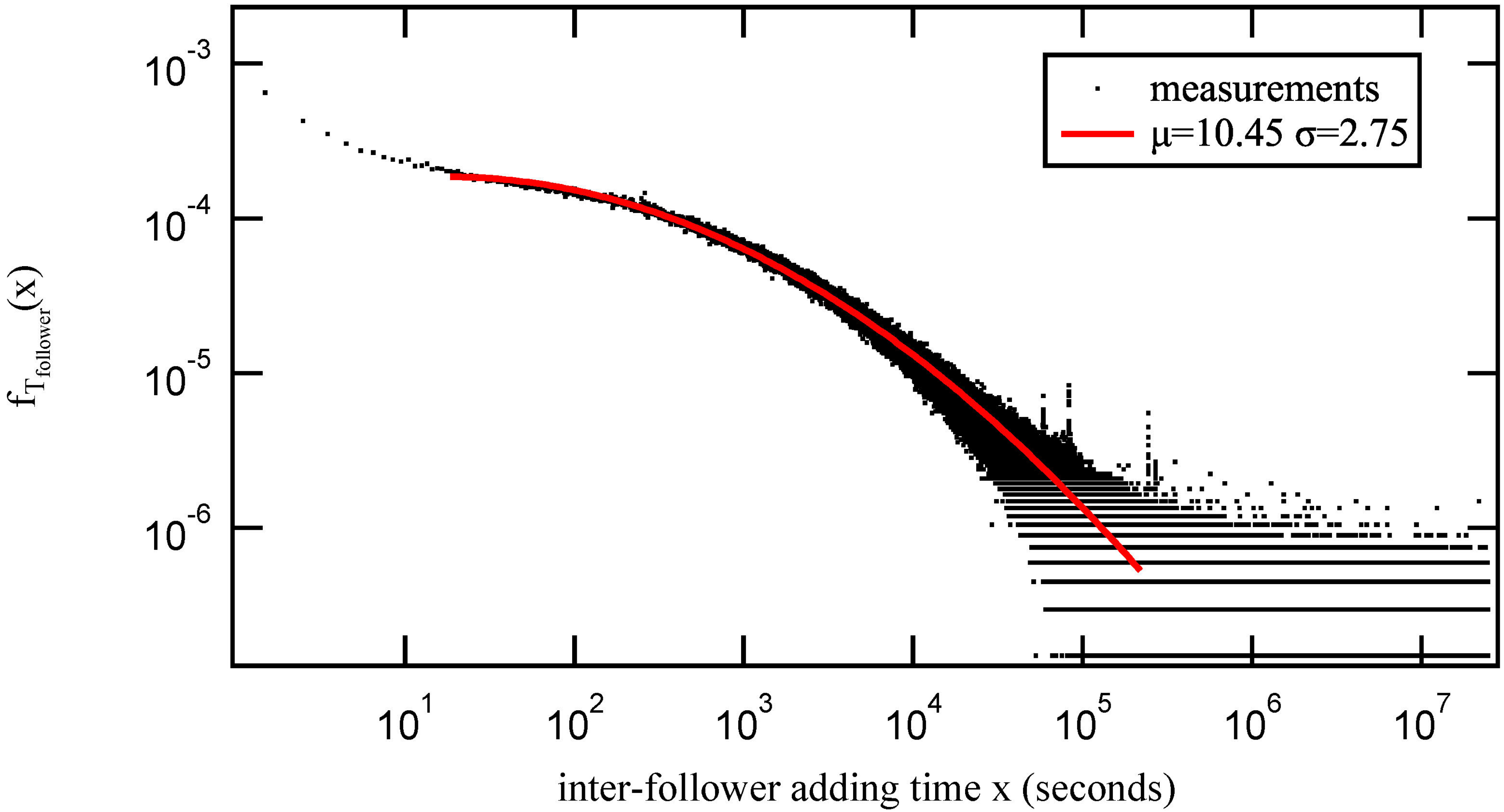
Time difference between adding friends



Digg.com

TU Delft

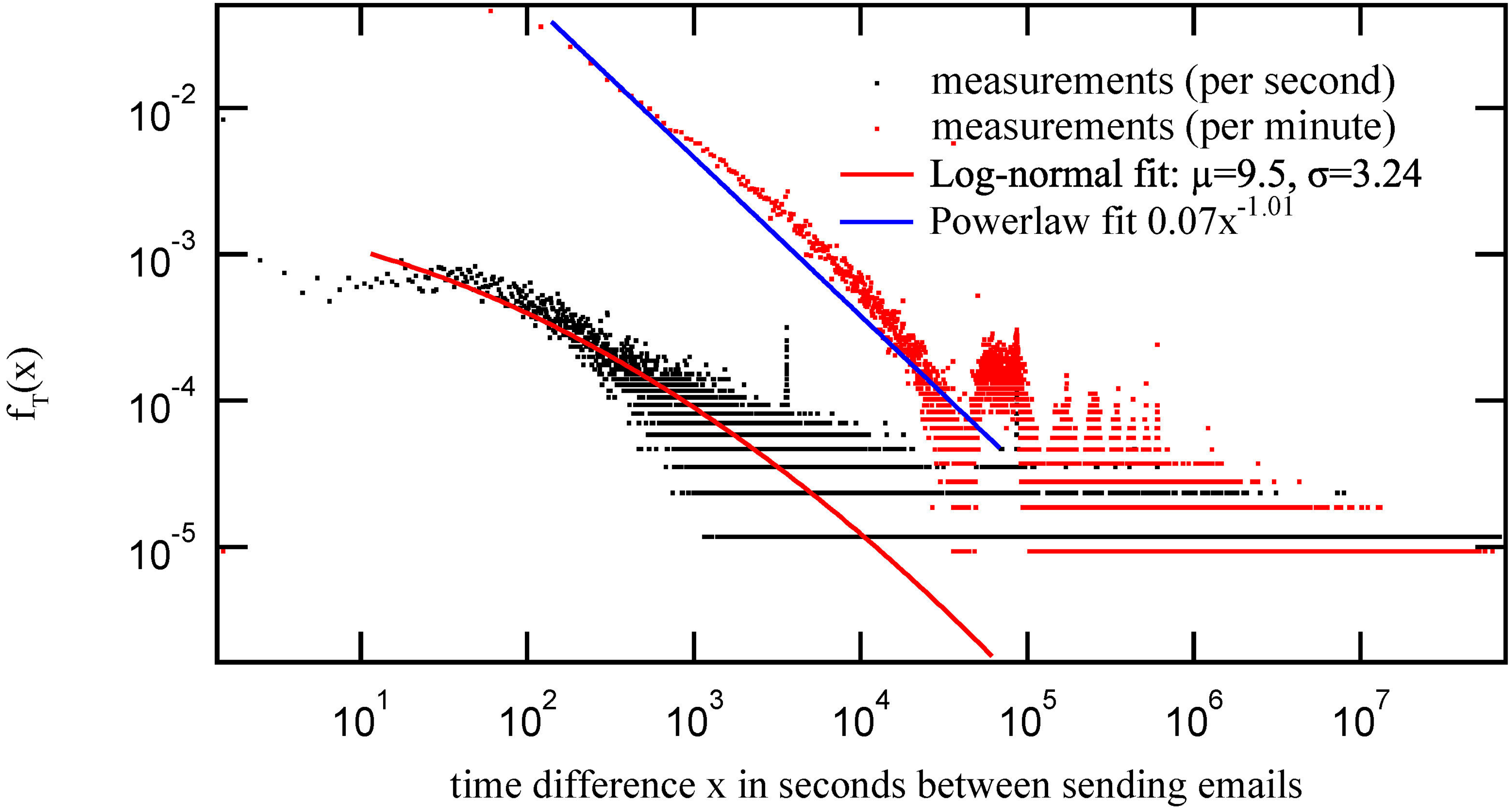
Time difference between receiving followers



Digg.com

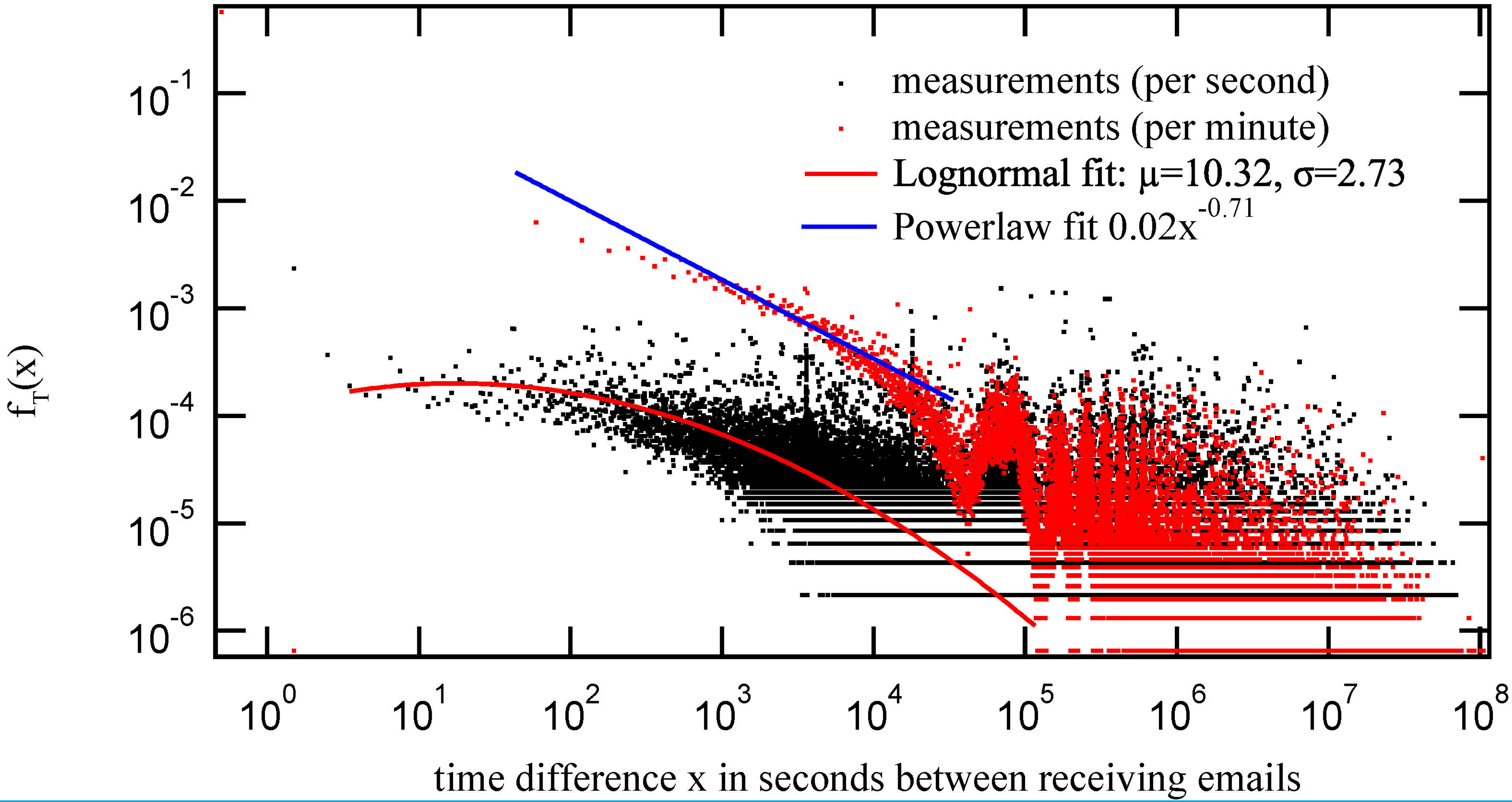
TU Delft

Time between emails sent (Enron)



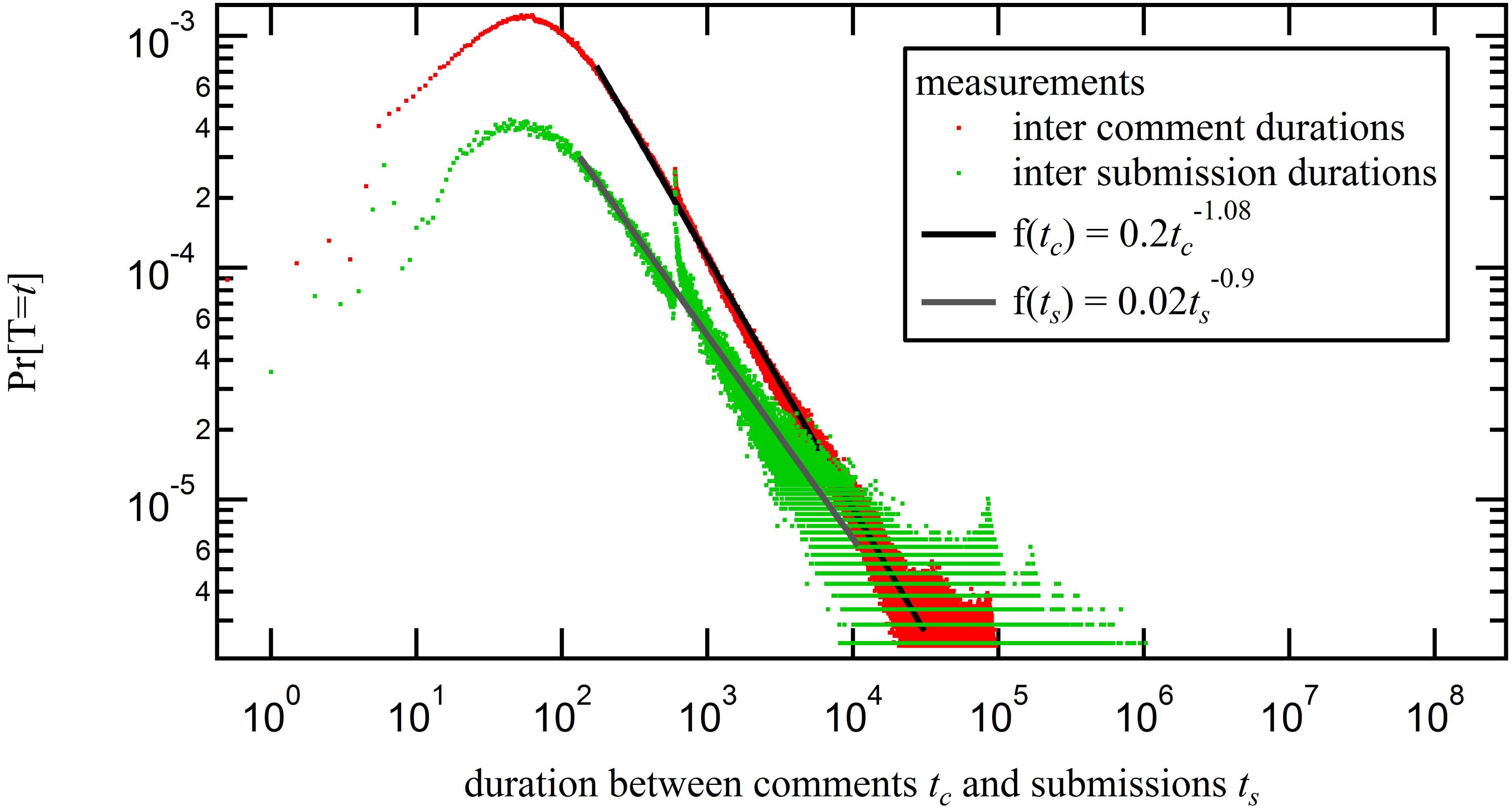
Red: binsize one minute
Black: binsize one second

Time between emails received (Enron)



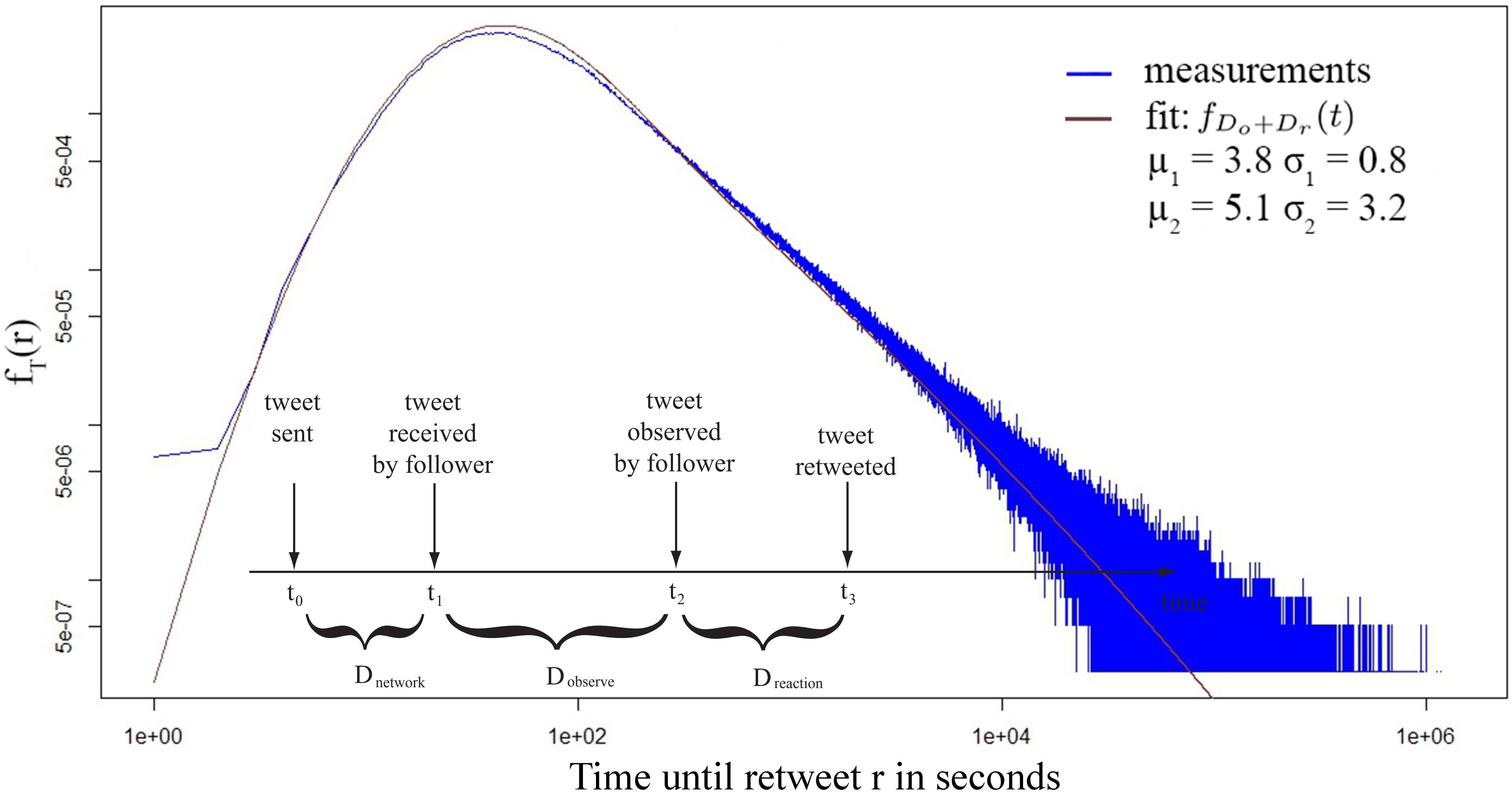
Red: binsize one minute
Black: binsize one second

Reddit



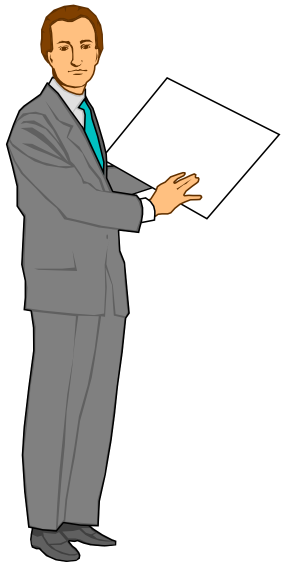
TU Delft

Twitter spreading times are lognormal-like



C. Doerr, N. Blenn and P. Van Mieghem, "Lognormal infection times of Online information spread", PLOS ONE, Vol. 8, No. 5, p. e64349, 2013

Outline



Human interactivity times

Lognormal distribution

Binning and σ

Lognormal distribution

A lognormal random variable X is defined as $X = \exp(Y)$, where $Y = N\left(\mu,\sigma^2\right)$ is a Gaussian random variable:

$$F_X(t) = \Pr[X \le t] = \Pr\left[e^Y \le t\right] = \Pr[Y \le \log t]$$

Thus

$$F_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\log t} e^{-\left(\frac{(u-\mu)^2}{2\sigma^2}\right)} du$$

with mean and variance

$$E[X] = e^{\mu} e^{\frac{\sigma^2}{2}} \qquad Var[X] = e^{2\mu} e^{\frac{\sigma^2}{2}} \left(e^{\frac{\sigma^2}{2}} - 1\right)$$

Lognormal probability density

$$F_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\log t} e^{-\left(\frac{(u-\mu)^2}{2\sigma^2}\right)} du$$

$$f_X(t) = \frac{dF_X(t)}{dt} = \frac{1}{\sigma t\sqrt{2\pi}} e^{-\left(\frac{(\log t-\mu)^2}{2\sigma^2}\right)}$$

Rewritten in “power-law” form:

$$f_X(t) = \frac{e^{-\left(\frac{\mu}{2\sigma^2}\right)^2}}{\sigma\sqrt{2\pi}} t^{-\alpha(t)}$$ with $$\alpha(t) = 1 + \frac{\log t - 2\mu}{2\sigma^2}$$

TU Delft

Power law "exponent" of Lognormal

$$f_X(t) = \frac{e^{-\left(\frac{\mu}{2\sigma^2}\right)^2}}{\sigma\sqrt{2\pi}} t^{-\alpha(t)}$$

with

$$\alpha(t) = 1 + \frac{\log t - 2\mu}{2\sigma^2}$$

We observe a right-hand side tail, when t > t_{max} and $t_{max} = e^{\mu-\sigma^2}$

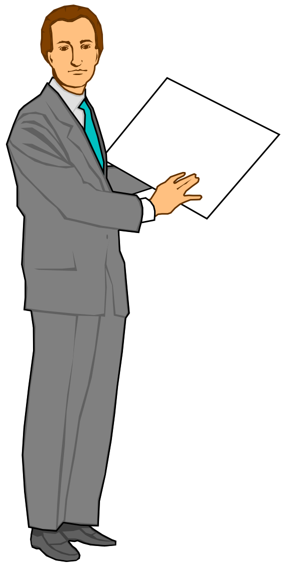
Thus, if $\left|\frac{\log t - 2\mu}{2\sigma^2}\right| \le \varepsilon$ equivalent to $t \in \left[e^{2\mu-2\sigma^2\varepsilon}, e^{2\mu+2\sigma^2\varepsilon}\right]$

then

$$\alpha(t) \le 1 + \varepsilon$$

→ Hardly distinguishable from a "real" power law

Outline



Human interactivity times

Lognormal distribution

Binning and σ

TU Delft

Scaling of Lognormal random variable

Solving the parameters (μ,σ) from the mean and variance:

$$\mu = \log E[X] - \frac{\sigma^2}{2} \qquad \sigma^2 = \log\left(1 + \frac{Var[X]}{(E[X])^2}\right)$$

Scaling the lognormal random variable Z = b X yields

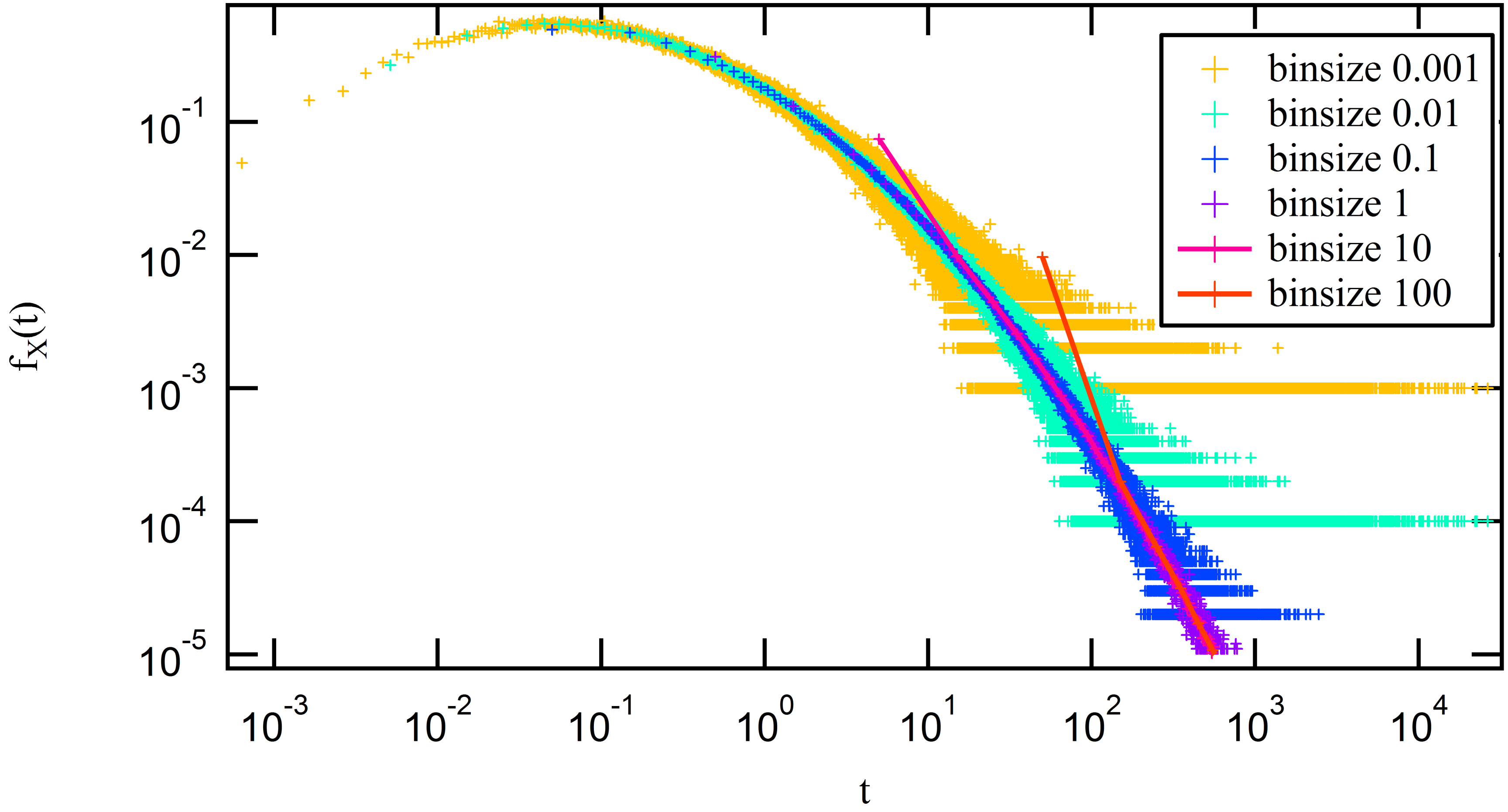
$$\mu_Z = \log b + \mu_X \qquad \sigma_Z = \sigma_X$$

Scaling by a factor b does not change σ

Measuring times in units of *second* or in units of *hours* does not change the parameter σ only the parameter μ

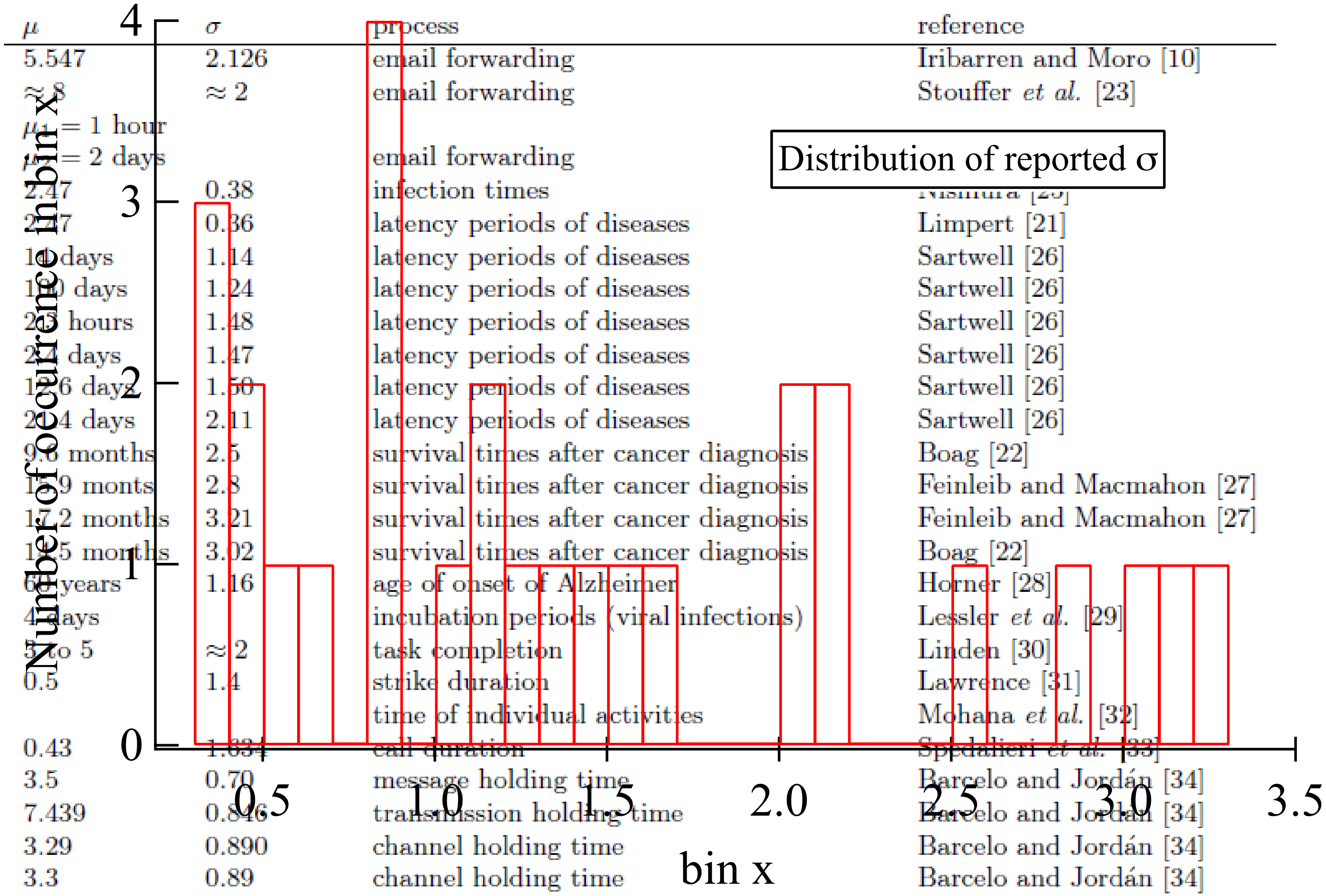
P. Van Mieghem, 2014, ***Performance Analysis of Complex Networks and Systems***, Cambridge University Press, U. K.

Effect of Binning



TU Delft

Reported lognormals



μ	σ	process	reference
5.547	2.126	email forwarding	Iribarren and Moro [10]
≈ 8	≈ 2	email forwarding	Stouffer *et al.* [23]
$\mu_1 = 1$ hour $\mu_2 = 2$ days		email forwarding	[illegible]
2.47	0.38	infection times	Nishiura [25]
2.47	0.36	latency periods of diseases	Limpert [21]
1[illegible] days	1.14	latency periods of diseases	Sartwell [26]
1[illegible]0 days	1.24	latency periods of diseases	Sartwell [26]
2[illegible] hours	1.48	latency periods of diseases	Sartwell [26]
2[illegible] days	1.47	latency periods of diseases	Sartwell [26]
1[illegible]6 days	1.50	latency periods of diseases	Sartwell [26]
2[illegible]4 days	2.11	latency periods of diseases	Sartwell [26]
9.[illegible] months	2.5	survival times after cancer diagnosis	Boag [22]
1[illegible]9 monts	2.8	survival times after cancer diagnosis	Feinleib and Macmahon [27]
17.2 months	3.21	survival times after cancer diagnosis	Feinleib and Macmahon [27]
1[illegible].5 months	3.02	survival times after cancer diagnosis	Boag [22]
6[illegible] years	1.16	age of onset of Alzheimer	Horner [28]
4 days		incubation periods (viral infections)	Lessler *et al.* [29]
[illegible] to 5	≈ 2	task completion	Linden [30]
0.5	1.4	strike duration	Lawrence [31]
		time of individual activities	Mohana *et al.* [32]
0.43	1.834	call duration	Spedalieri *et al.* [33]
3.5	0.70	message holding time	Barcelo and Jordán [34]
7.439	0.846	transmission holding time	Barcelo and Jordán [34]
3.29	0.890	channel holding time	Barcelo and Jordán [34]
3.3	0.89	channel holding time	Barcelo and Jordán [34]

$$0.35 \le \sigma \le 3.2$$

TU Delft

Which processes generate a lognormal?

The definition: A lognormal random variable $X = \exp(Y)$, where Y is a Gaussian random variable

Central Limit Theorem: a scaled sum of i.i.d rv. converges to a Gaussian (replace r.v. X by log X, then convergence to a Lognormal)

Law of proportionate effect (is similar in nature)

Many time related human sensing/actions scale logarithmically rather than in actual time units (e.g. hearing, optical light intensities)

Van Mieghem, P., N. Blenn and C. Doerr, 2011,"Lognormal Distribution in the Digg Online Social Network", The European Physical Journal B, Vol. 83, No. 2, pp. 251-261.

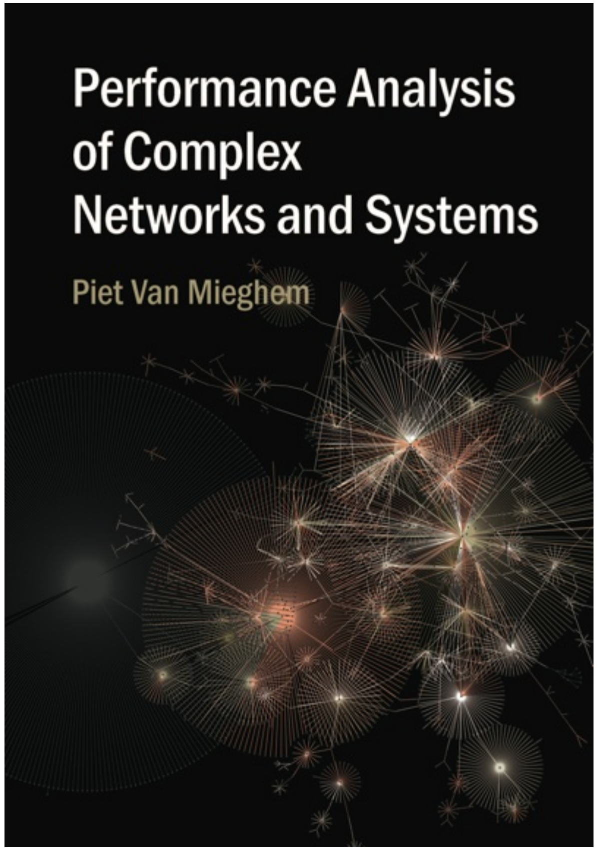
Conclusion

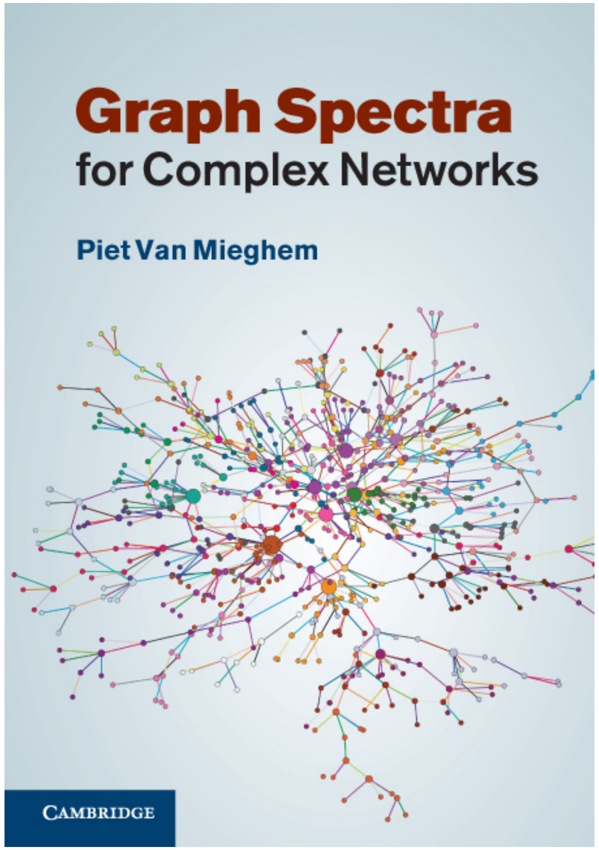
- Binning changes the shape of the distribution
  but not σ, the sigma value of the fitted log-normal

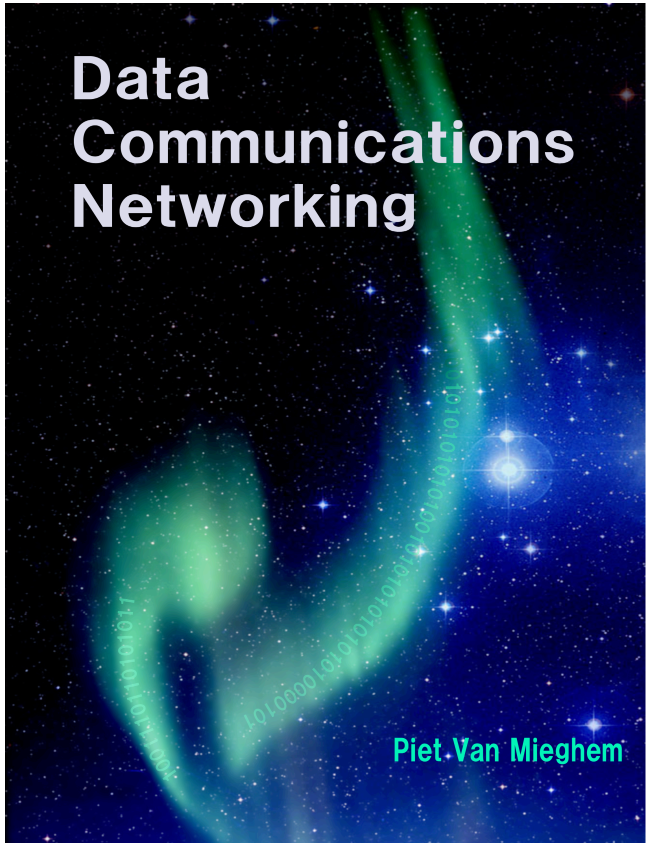
- Distributions of interactivity, which are approximated by a power law with exponent ~ 1, are possibly lognormal

- Precise underlying processes that generate lognormal-like distributions are largely unknown
  (apart from the Central Limit Theorem for a sum i.i.d. logX rv's)

Books







Articles: **http://www.nas.ewi.tudelft.nl**

Thank You

Piet Van Mieghem
NAS, TUDelft
P.F.A.VanMieghem@tudelft.nl