wide side of the Internet

why 1/x distribution is so prevalent
in Internet data?

the spotter team

- S. Laki
- P. Mátray
- I. Csabai
- T. Sebők
- Peter Hága
- Gabor Vattay

Physics of Complex Systems
Eötvös University
Budapest
Hungary

Spotter
INNOVATION IN GEOLOCATION

INTERNET DATA AND CHECK FRAUD

The table lists the checks that a manager in the office of the Arizona State Treasurer wrote to divert funds for his own use. The vendors to whom the checks were issued were fictitious.

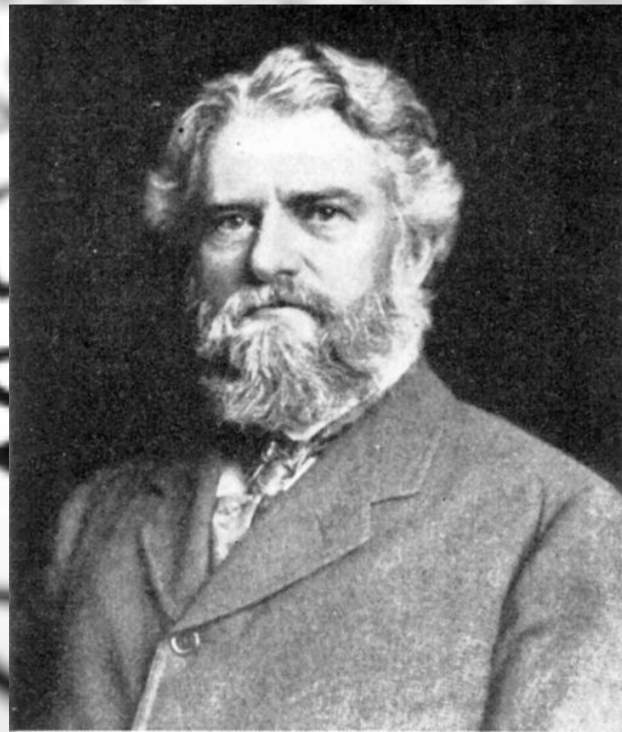
Date of Check	Amount
October 9, 1992	$ 1,927.48
↓	27,902.31
October 14, 1992	86,241.90
↓	72,117.46
	81,321.75
	97,473.96
October 19, 1992	93,249.11
↓	89,658.17
	87,776.89
	92,105.83
	79,949.16
	87,602.93
	96,879.27
	91,806.47
	84,991.67
	90,831.83
	93,766.67
	88,338.72
	94,639.49
	83,709.28
	96,412.21
	88,432.86
	71,552.16
TOTAL	**$ 1,878,687.58**

M. J. Nigrini , J. Accountancy 1999

Table 11-1 Benford's Law: Distribution of Leading Digits

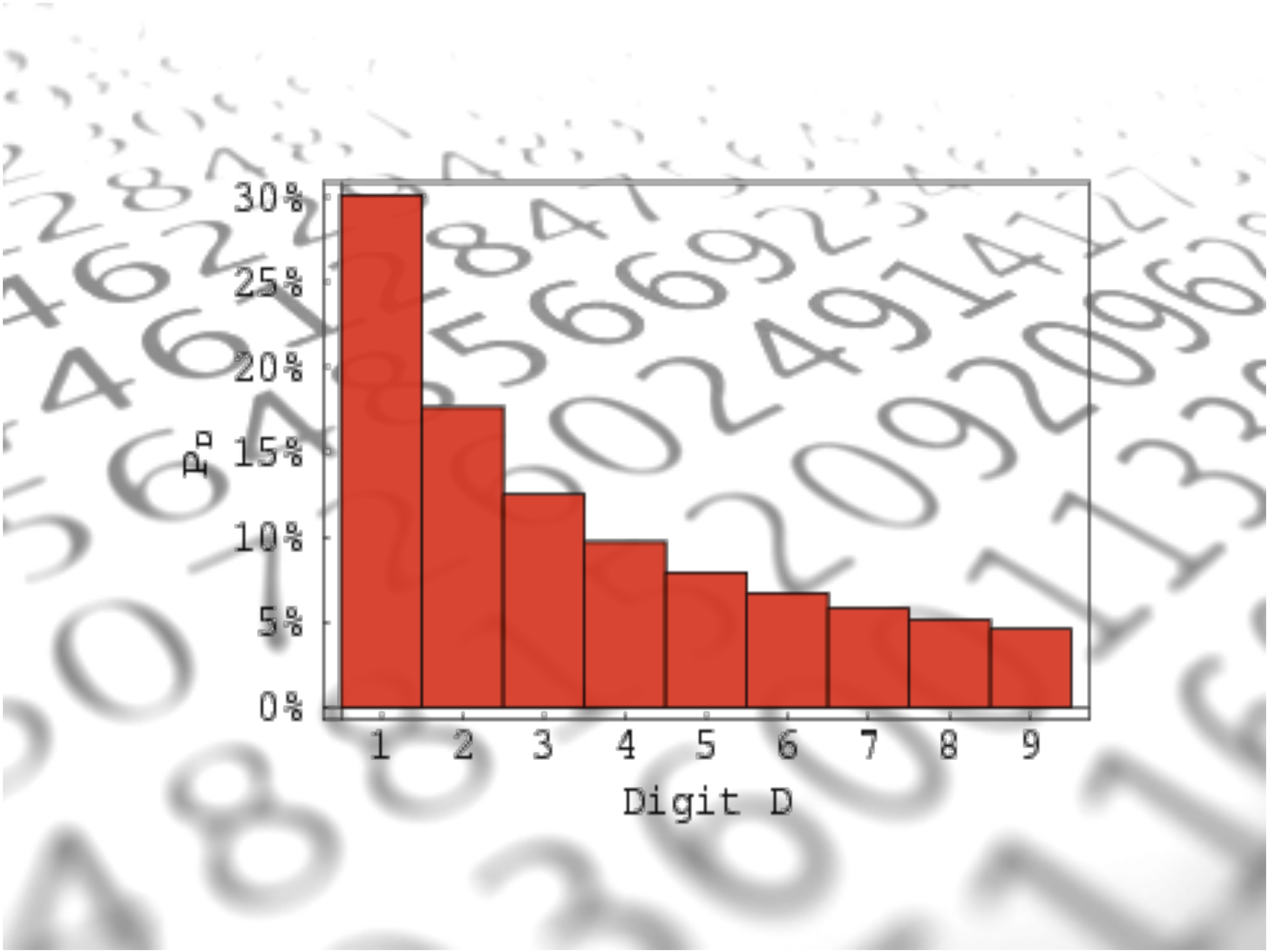
Leading Digit	1	2	3	4	5	6	7	8	9
Benford's law: frequency distribution of leading digits	30.1%	17.6%	12.5%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%
Expected frequencies of leading digits from 784 checks following Benford's law	235.984	137.984	98.000	76.048	61.936	52.528	45.472	39.984	36.064
Observed leading digits of 784 actual checks analyzed for fraud	0	15	0	76	479	183	8	23	0

Frank Benford 1938



Simon Newcomb 1881

$$P(d) = \frac{\log_{10}(1+1/d)}{\log_{10} B}$$



30%
25%
20%
15%
10%
5%
0%
P_D
1
2
3
4
5
6
7
8
9
Digit D

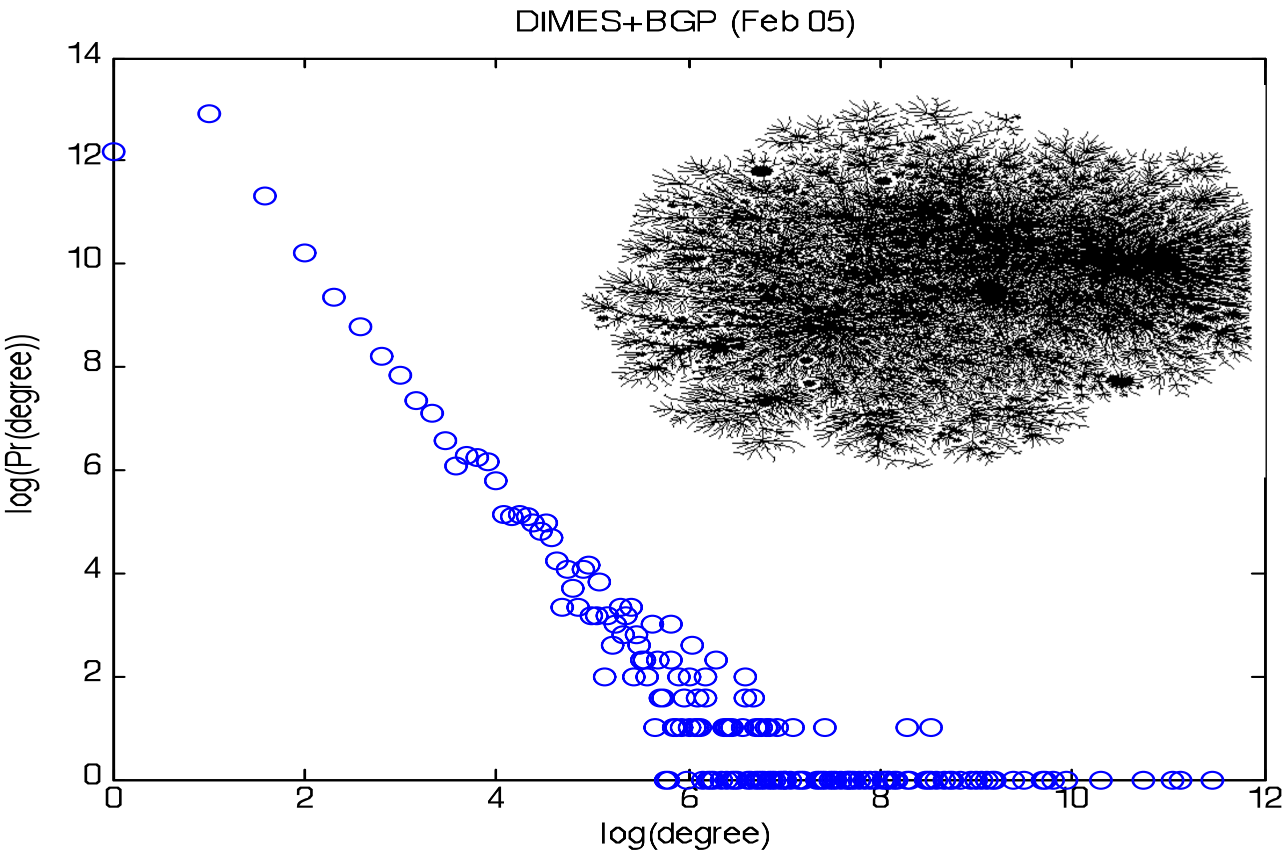
- density 1/x

- cummulative distribution log(x)

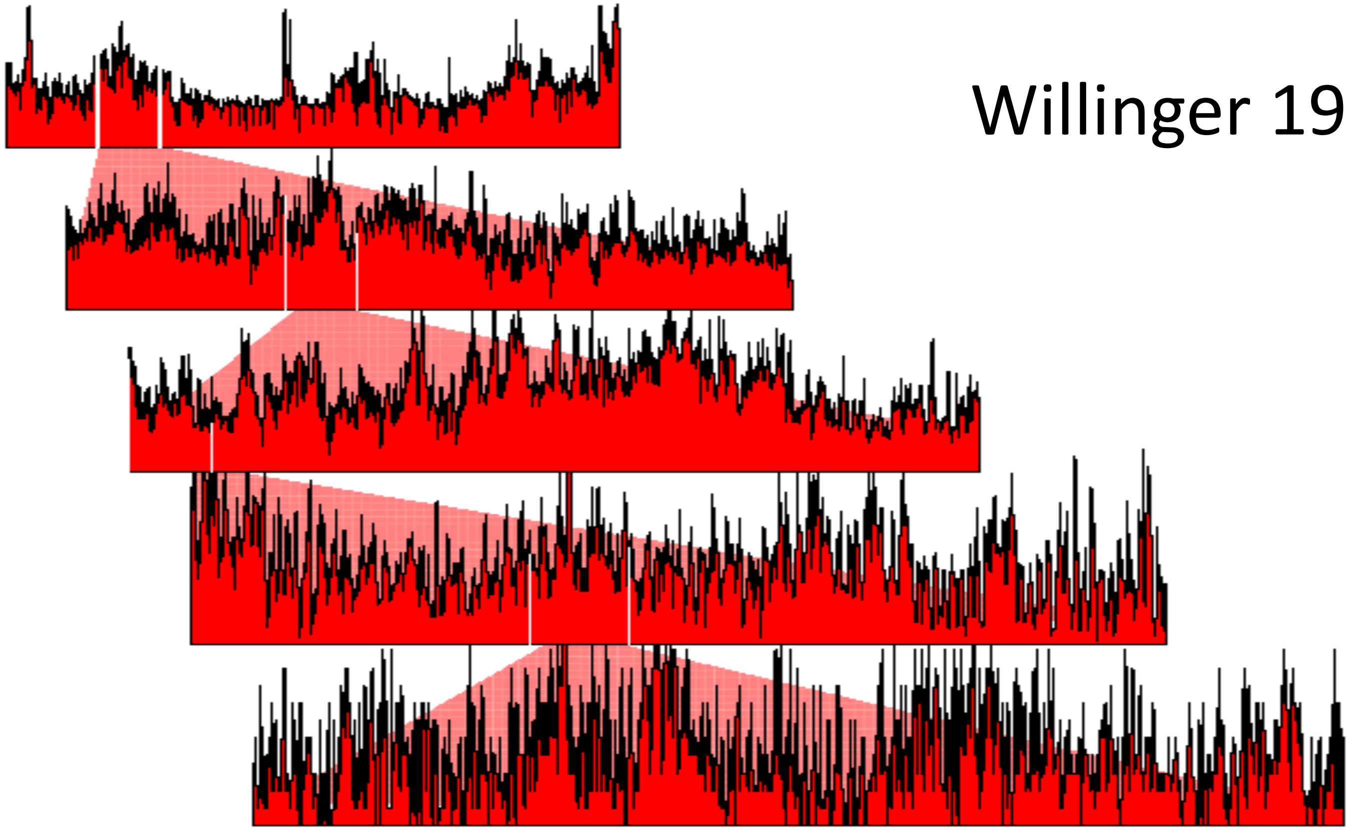
$$P(d) = \log_{10}(d+1) - \log_{10}(d) = \log_{10}\left(1 + \frac{1}{d}\right).$$

power laws in Internet data

- Falutsos Brothers 1999

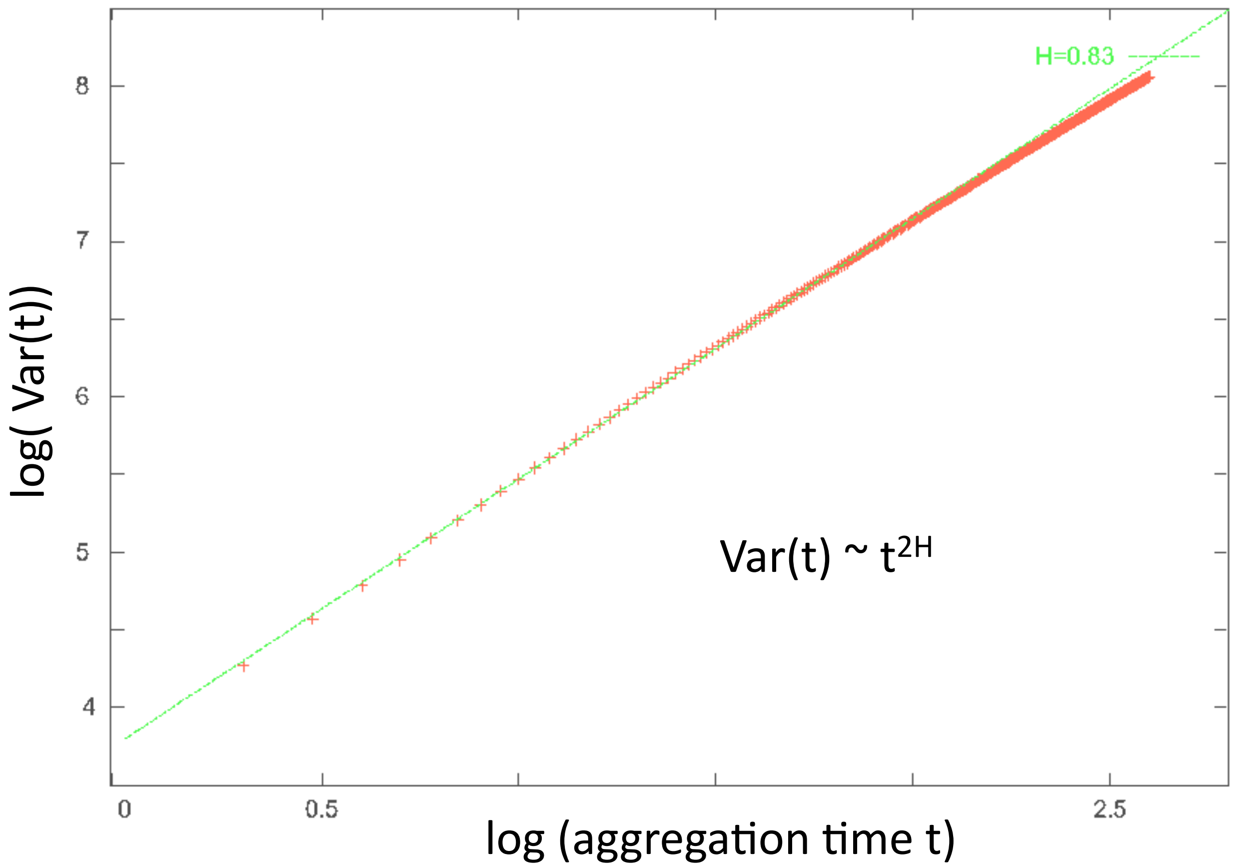


power laws in Internet traffic



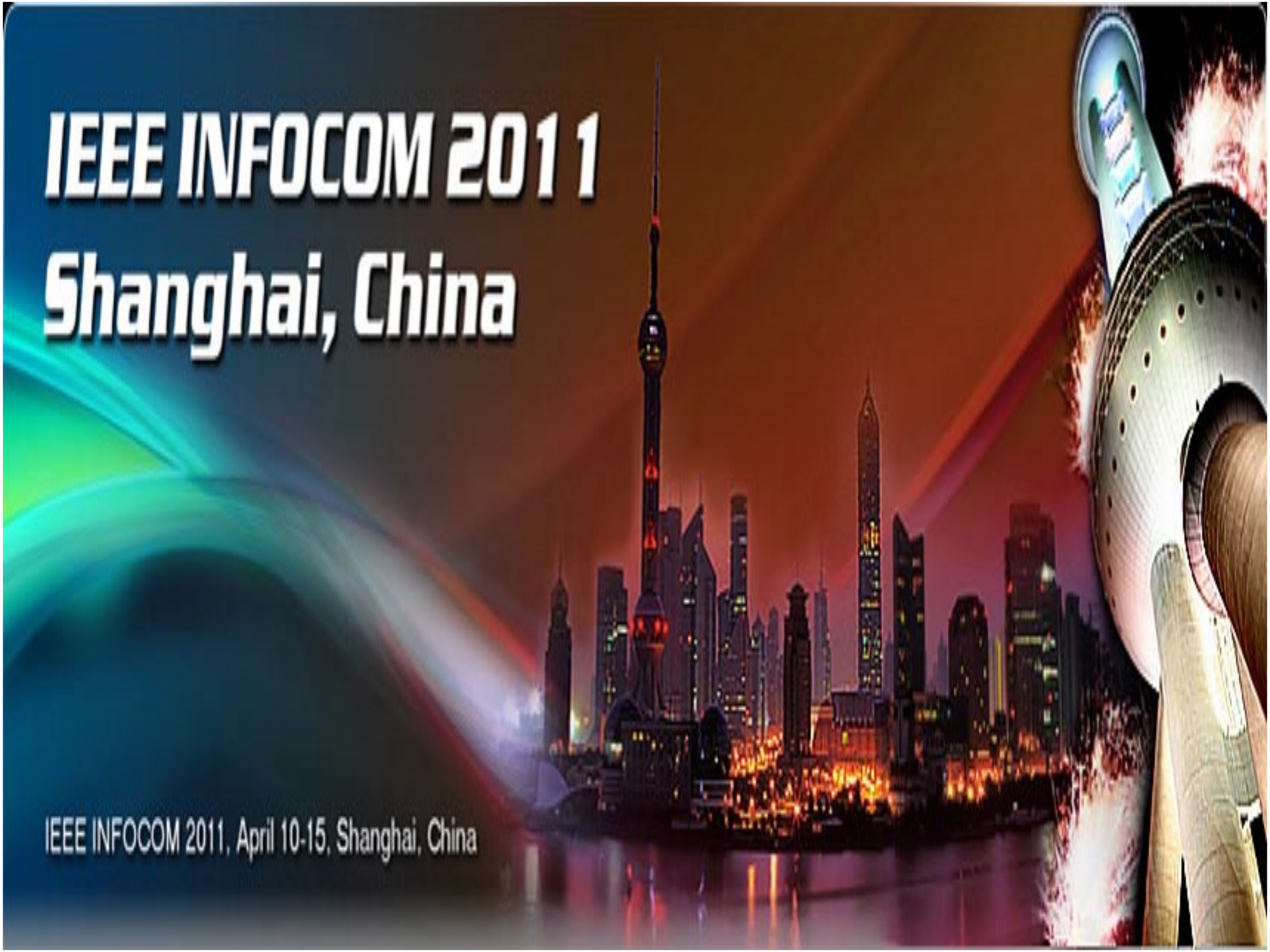
Willinger 1993

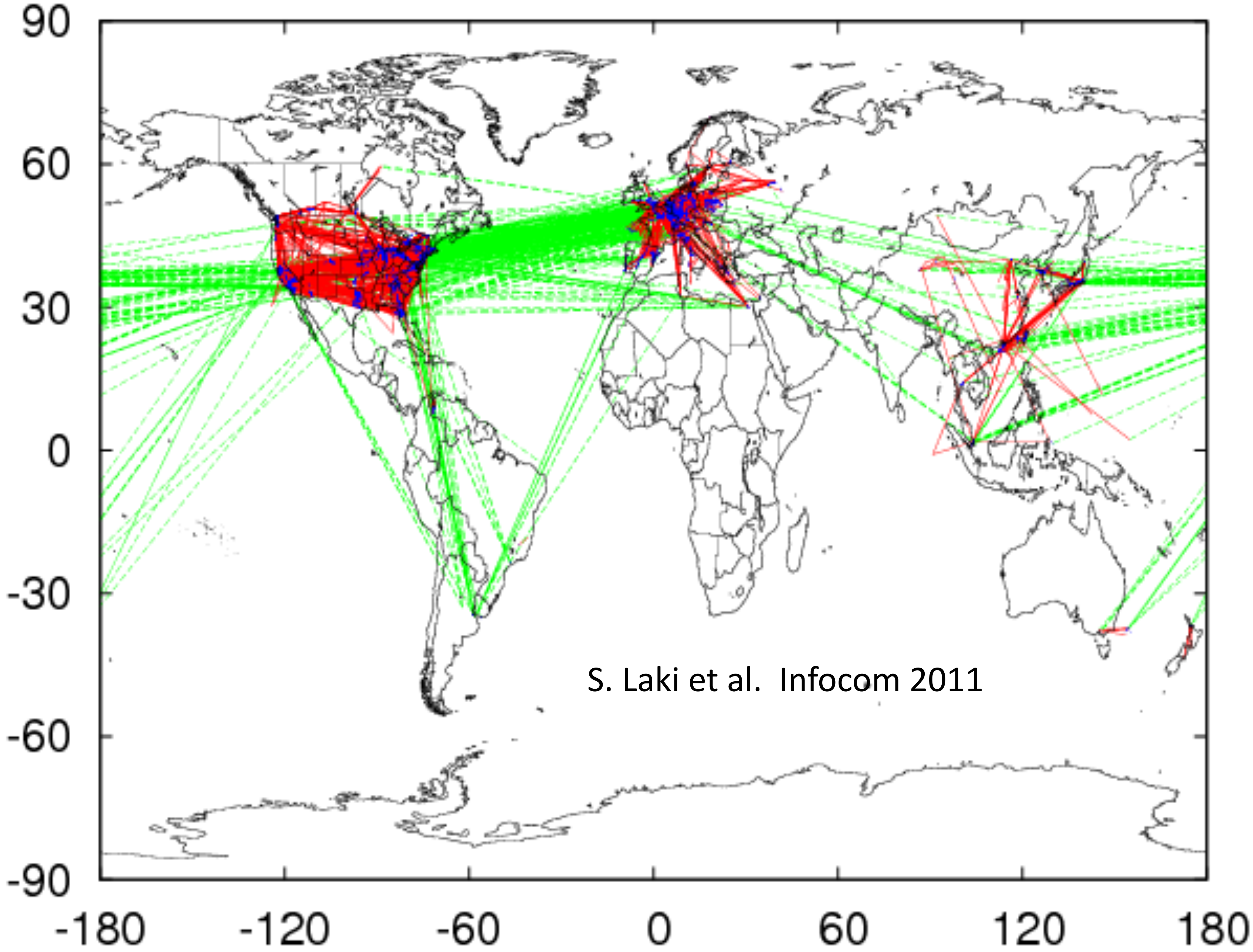
power laws in Internet traffic

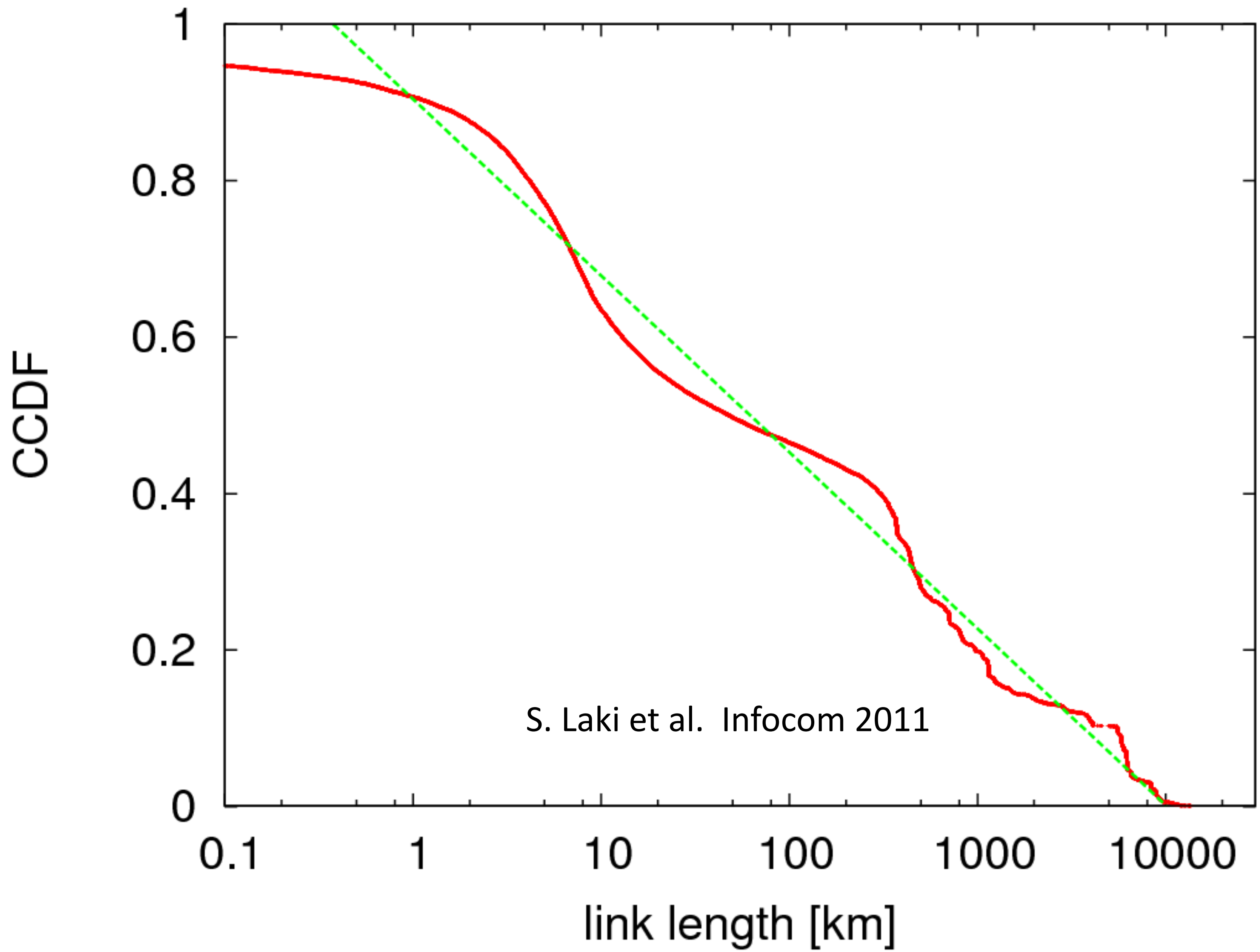


1/x

- has infinite variance
- has infinite expectation
- has infinite integral – not even a distribution

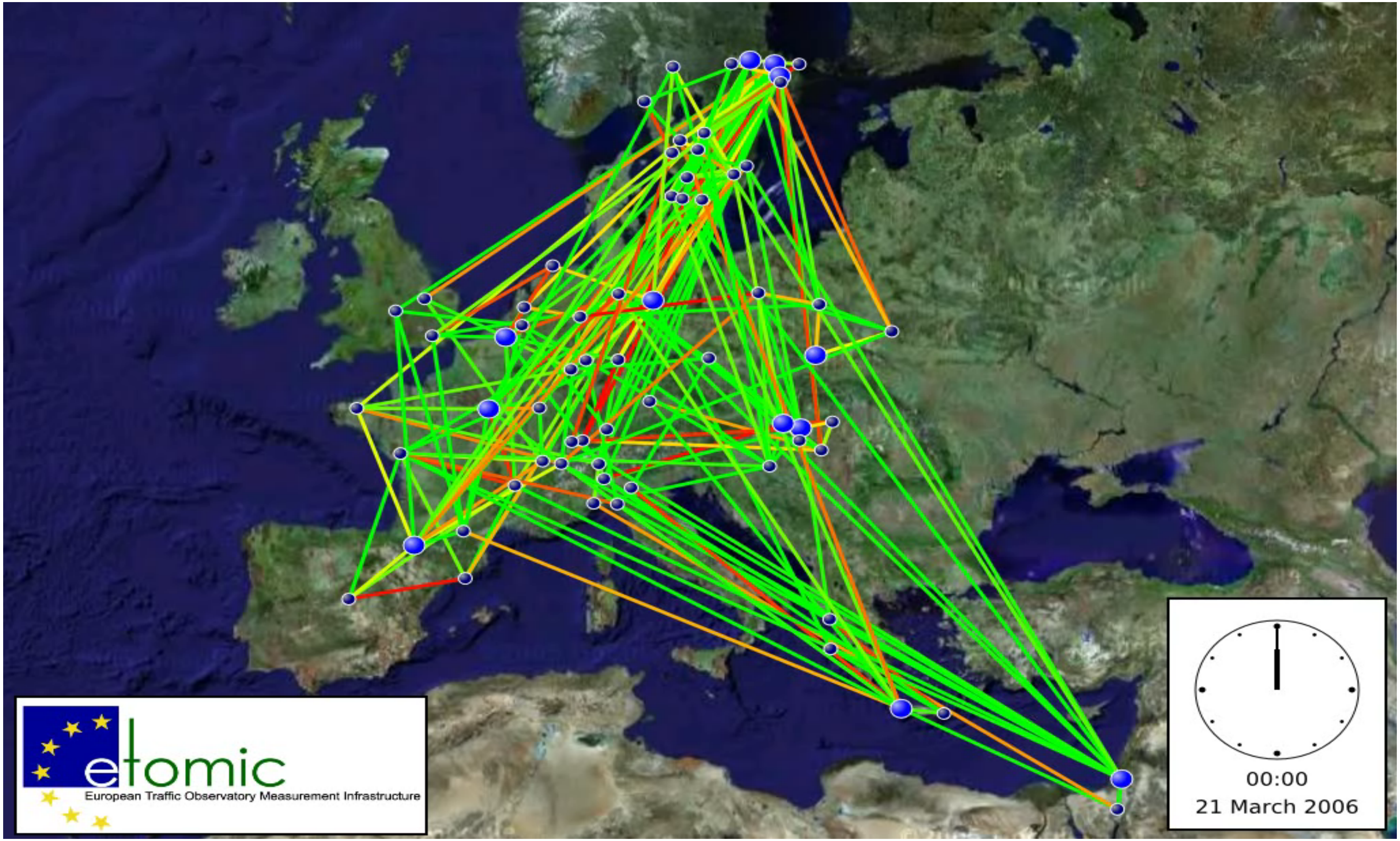

IEEE INFOCOM 2011
Shanghai, China
IEEE INFOCOM 2011, April 10-15, Shanghai, China

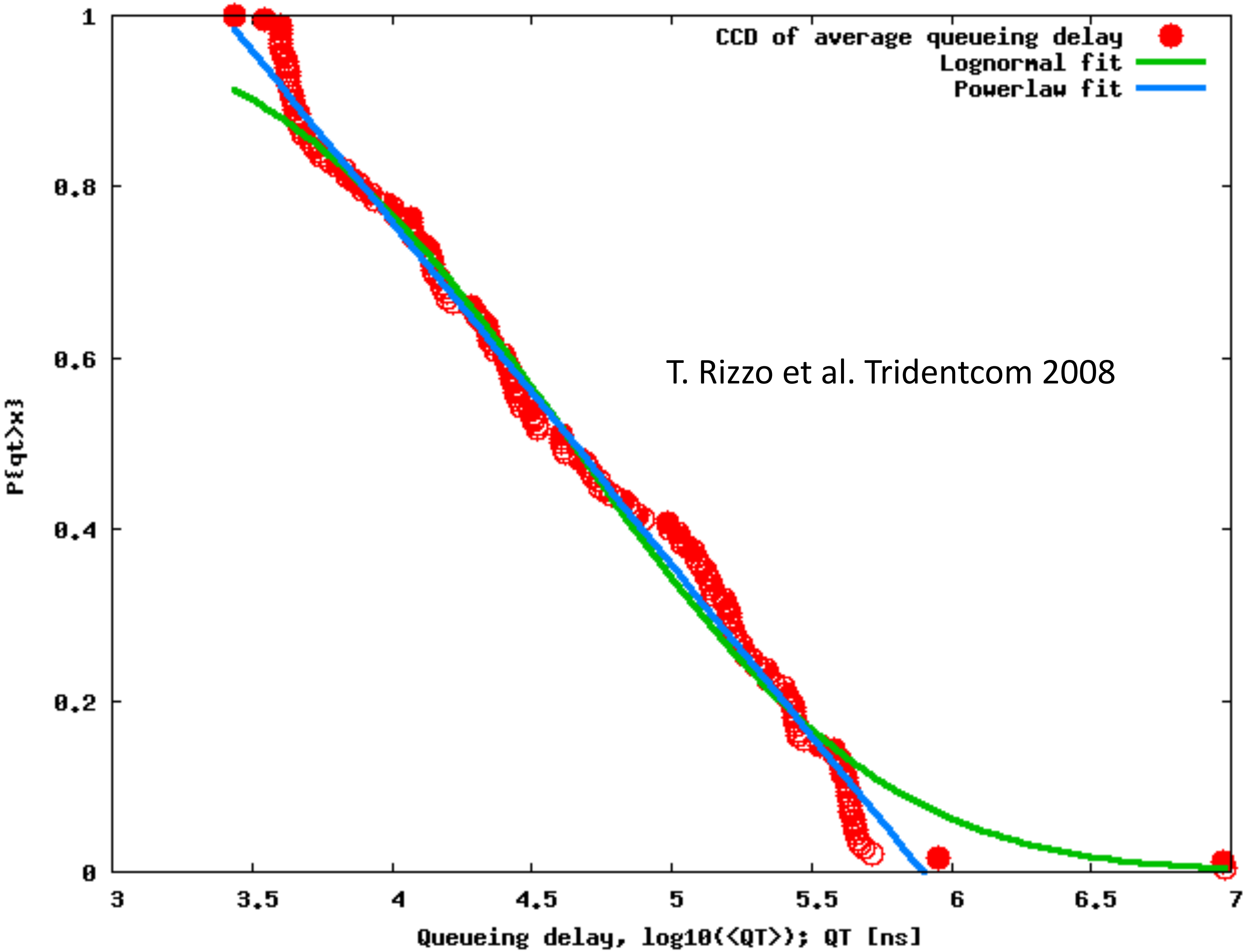

90
60
30
0
-30
-60
-90
-180
-120
-60
0
60
120
180
S. Laki et al. Infocom 2011



1
0.8
0.6
0.4
0.2
0
CCDF
0.1
1
10
100
1000
10000
link length [km]
S. Laki et al. Infocom 2011

Waiting time

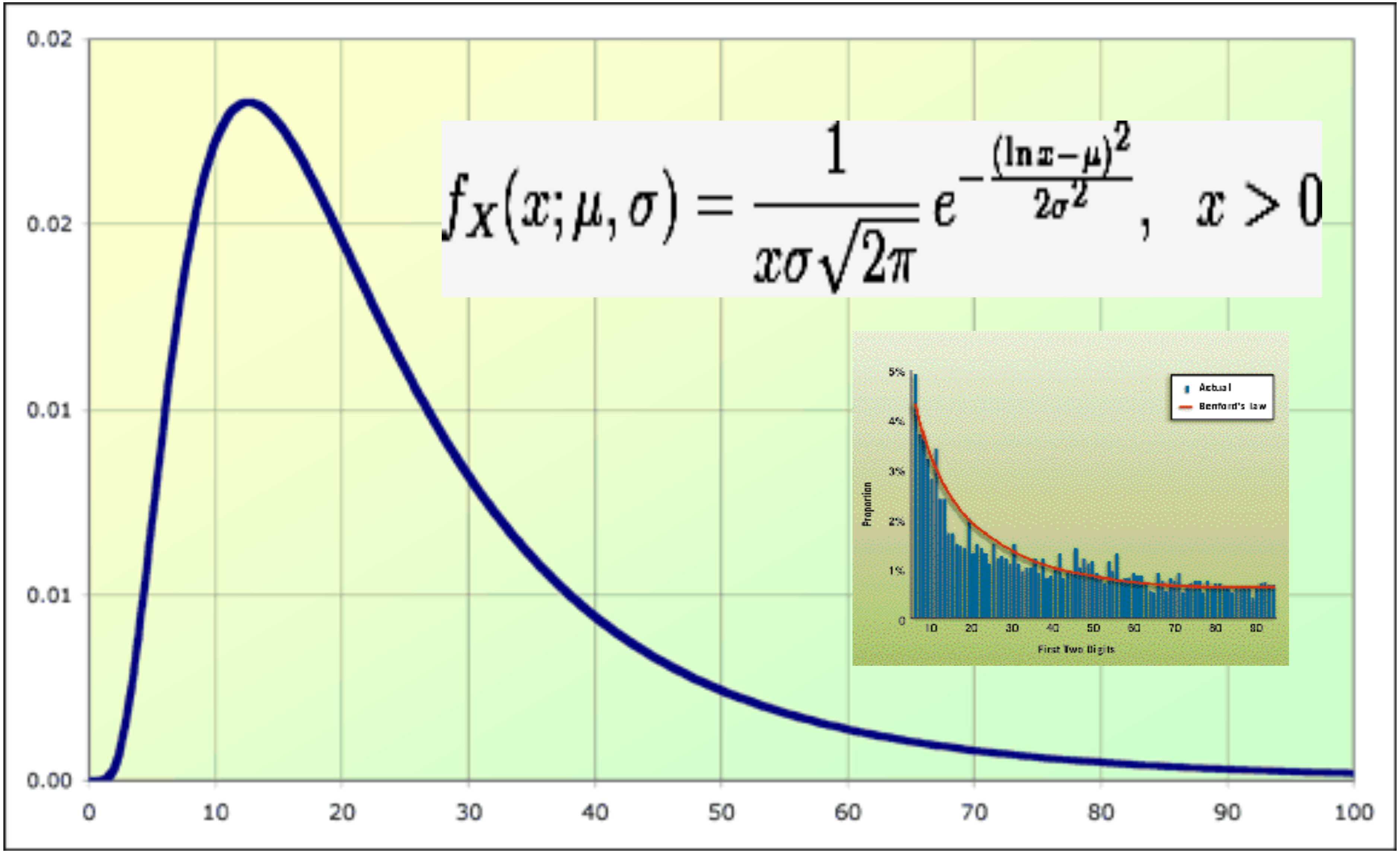




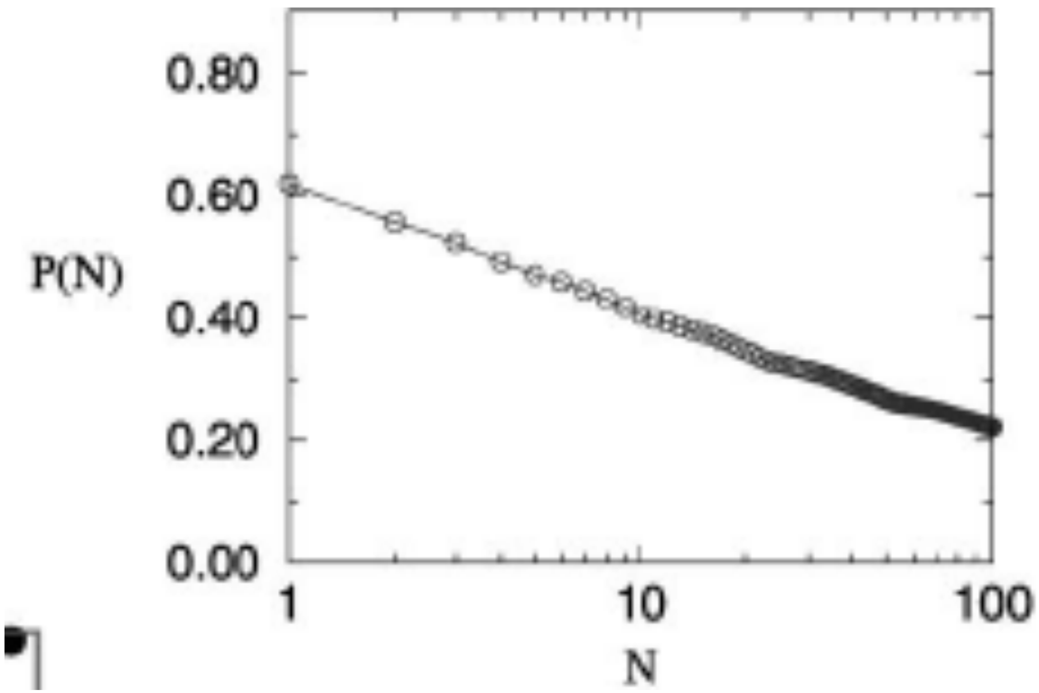
CCD of average queueing delay
Lognormal fit
Powerlaw fit
T. Rizzo et al. Tridentcom 2008
P{qt>x}
Queueing delay, log10(<QT>); QT [ns]
1
0.8
0.6
0.4
0.2
0
3
3.5
4
4.5
5
5.5
6
6.5
7

MULTIPLICATIVE RANDOMNESS

multiplicative central limit theorem

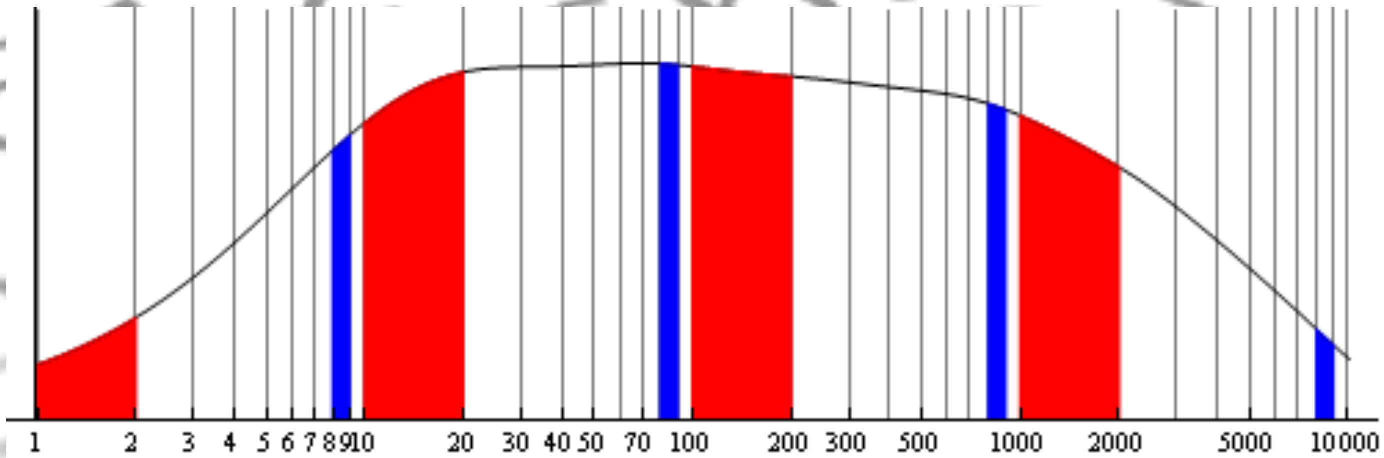


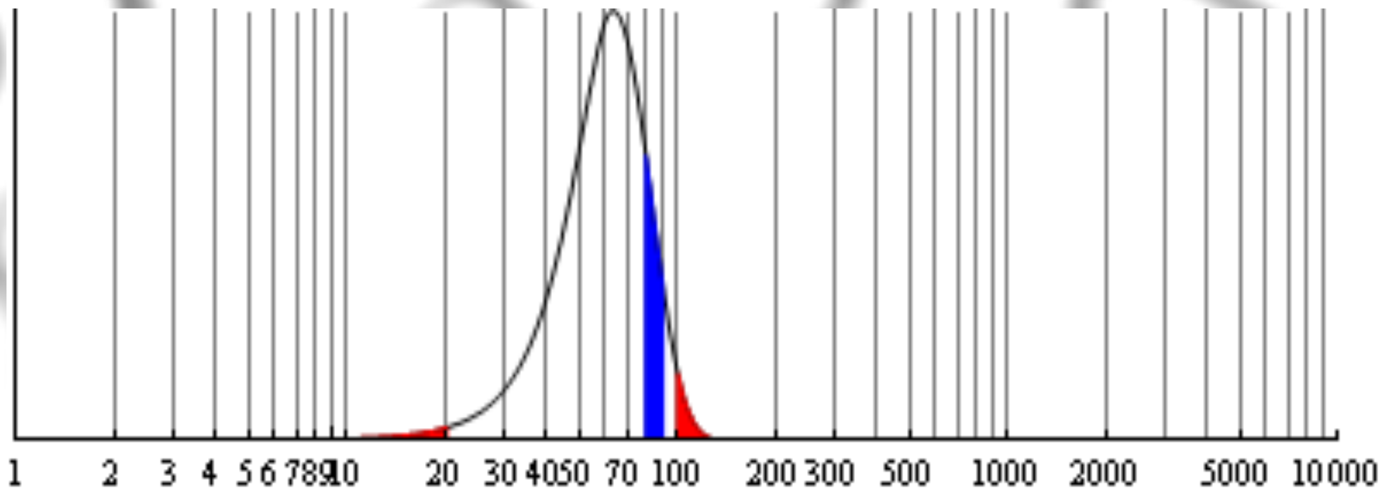
multiplicative noise



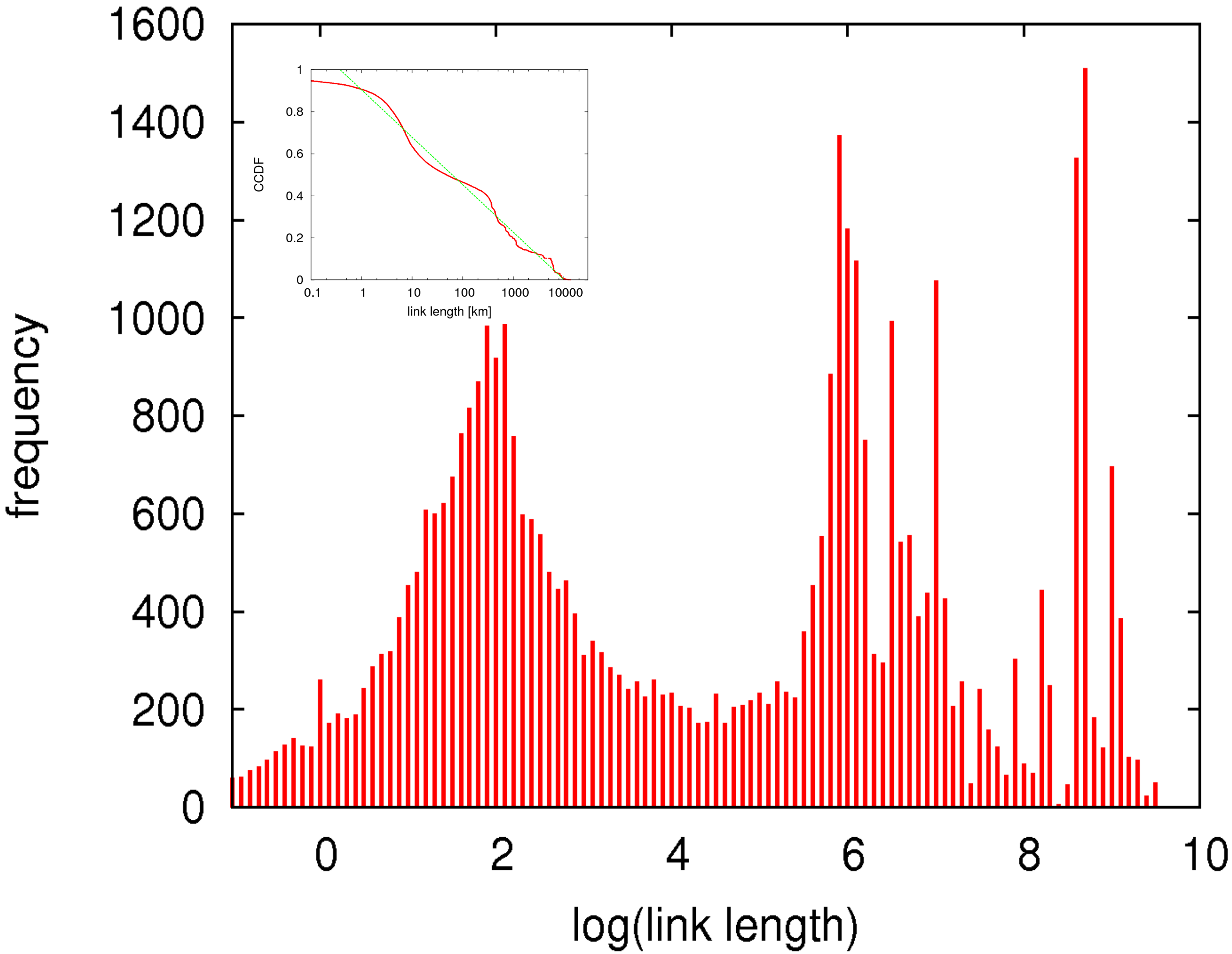
Pietronero et al. Physica A 2001

wide and narrow





wide distributions



1/x

- ‘super’ scale invariant
- describes well various spatial distributions in the Internet network
- multiplicative process is behind
- log(x) has wide distribution

thank you!

vattay@elte.hu