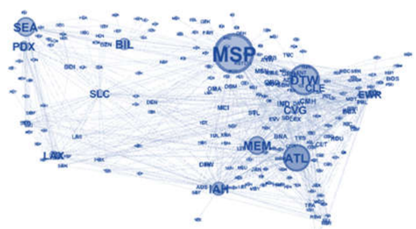
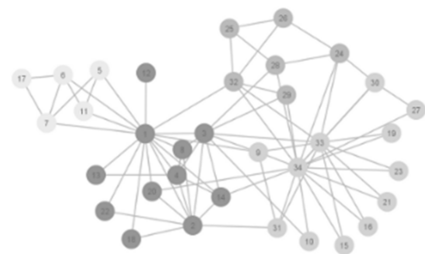
Prof. Ralucca Gera, rgera@nps.edu
Applied Mathematics Department,
Naval Postgraduate School



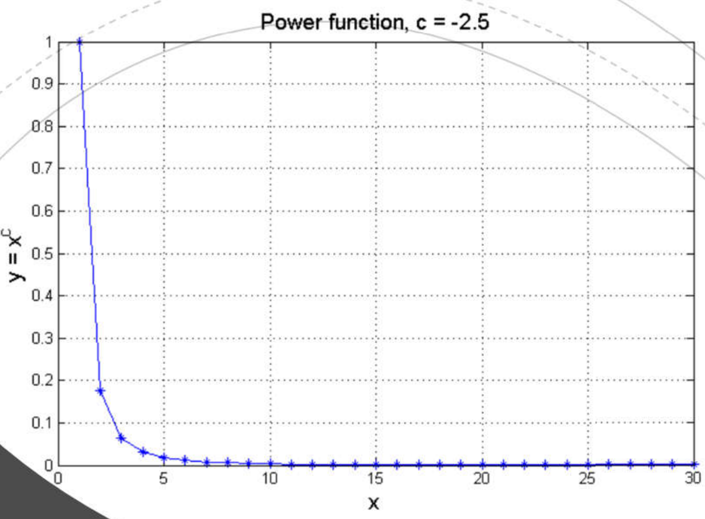


MA4404 Complex Networks

Power Law and Scale Free Networks

Learning Outcomes

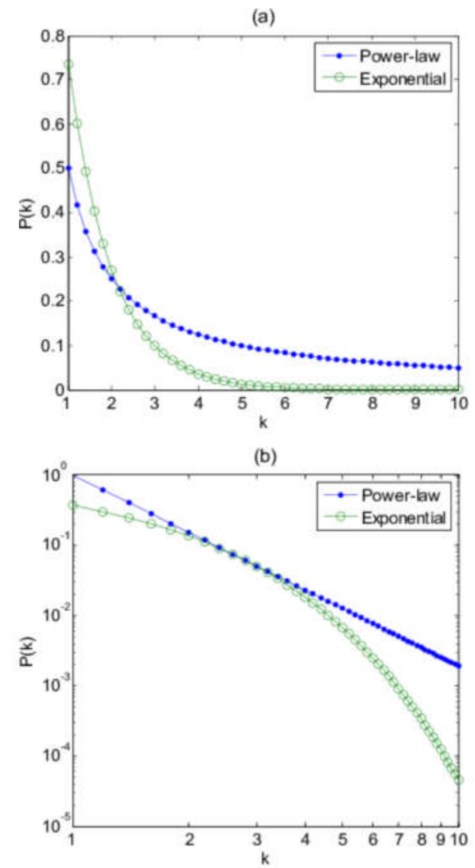
- Understand the different models of presenting distributions.
- Understand what is a power law distribution
- Understand what is a scale free network



Power Law Distribution

Power law distributions

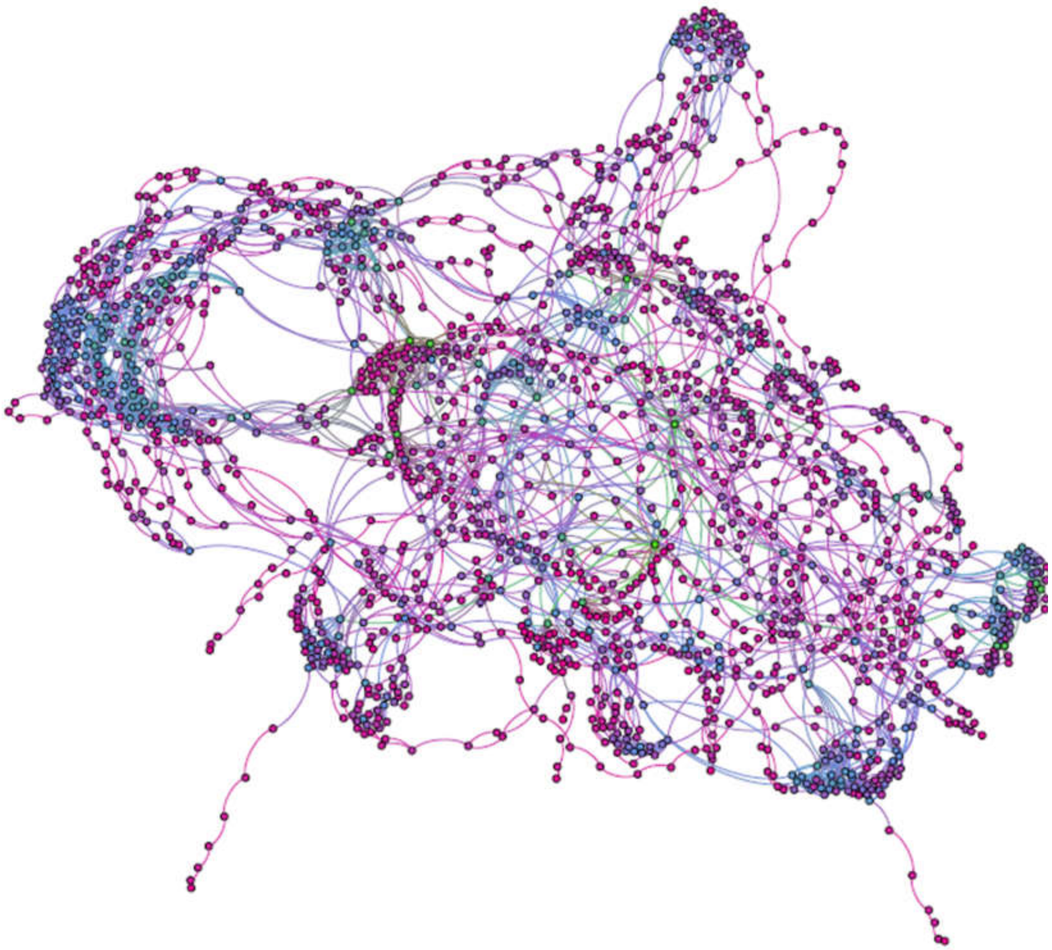
- Power law distribution: $p_k = \alpha\, k^{-c}, c > 0$ also called Pareto distributions (Vilfredo Pareto)
- Differentiating power-law from non power-law distributions is not trivial
  - Simplest (but not accurate) strategy → visual inspection plots
  - In 2006 - fit a distribution over the observed data, and test the goodness of the fit. This analysis revealed that none of them fits the theoretical distribution
- There are alternatives as we will see

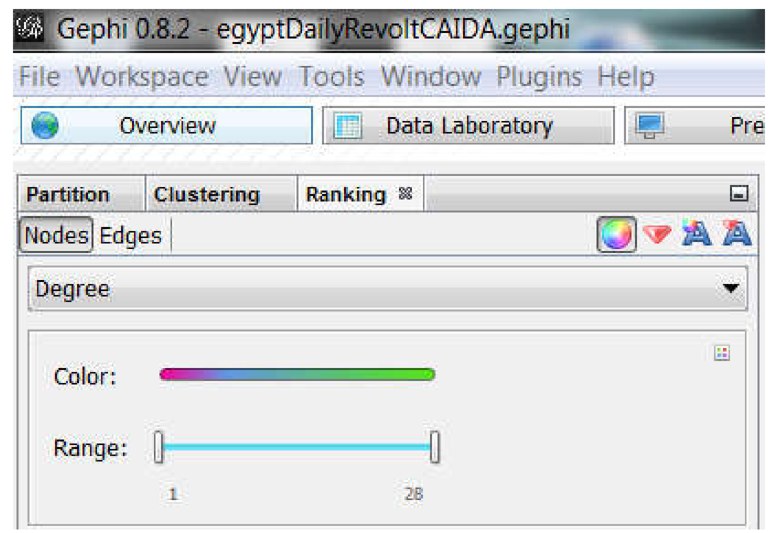


Power-law vs Exponential
(a) normal scale, (b) log-log scale)

https://www.researchgate.net/figure/Power-law-and-exponential-distributions-a-Normal-scale-b-Log-log-scale_fig1_228879251

An example: Egypt IP-layer data

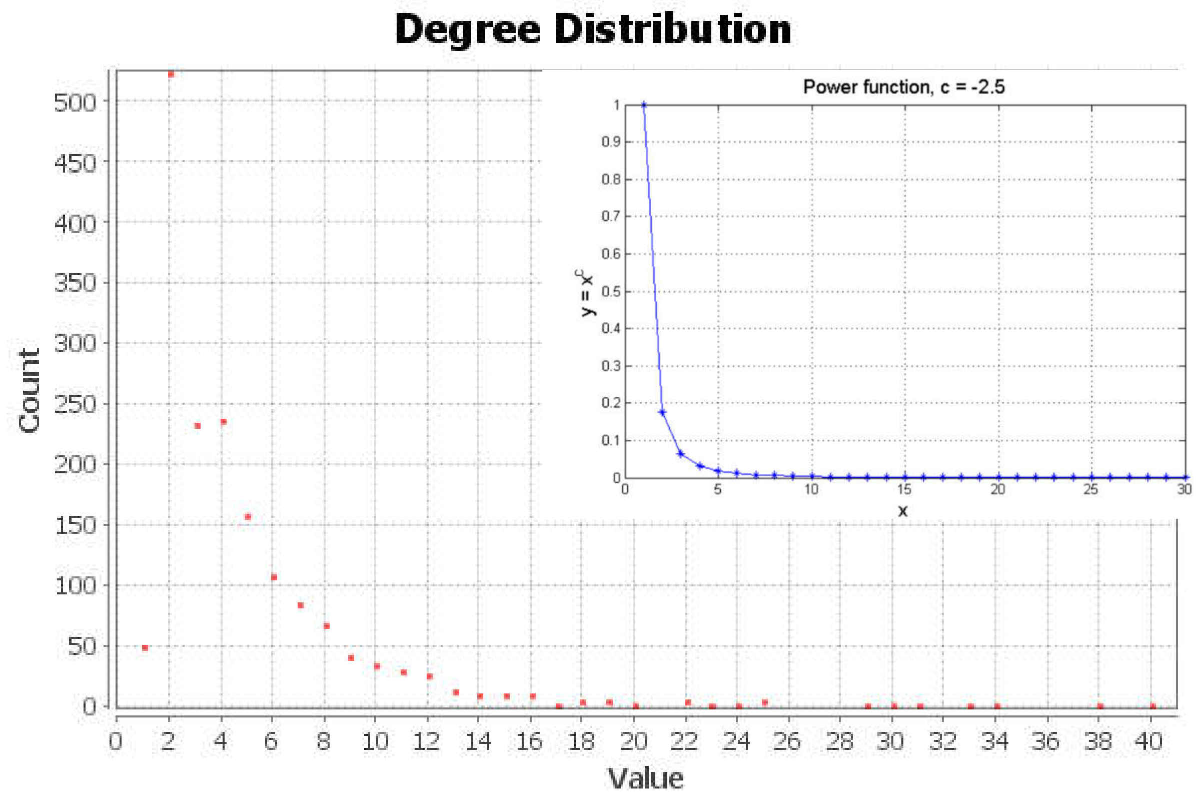




And its degree distribution

Results:

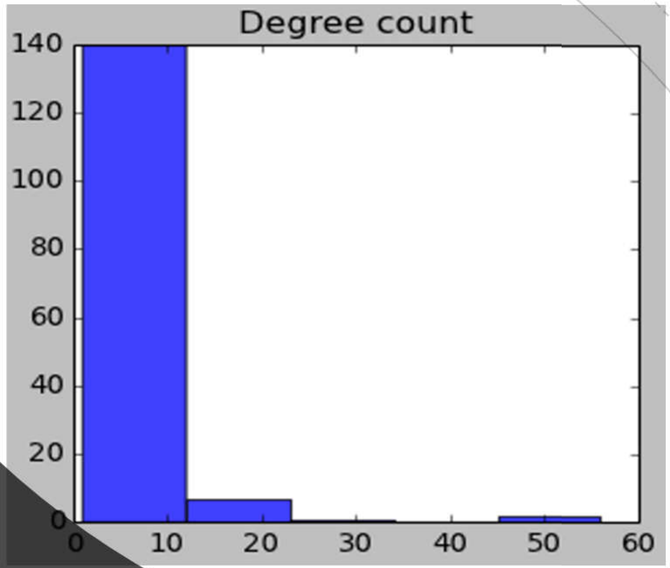
Average Weighted Degree: 4.812



Notice the long tail.
It looks like a
power law
distribution:

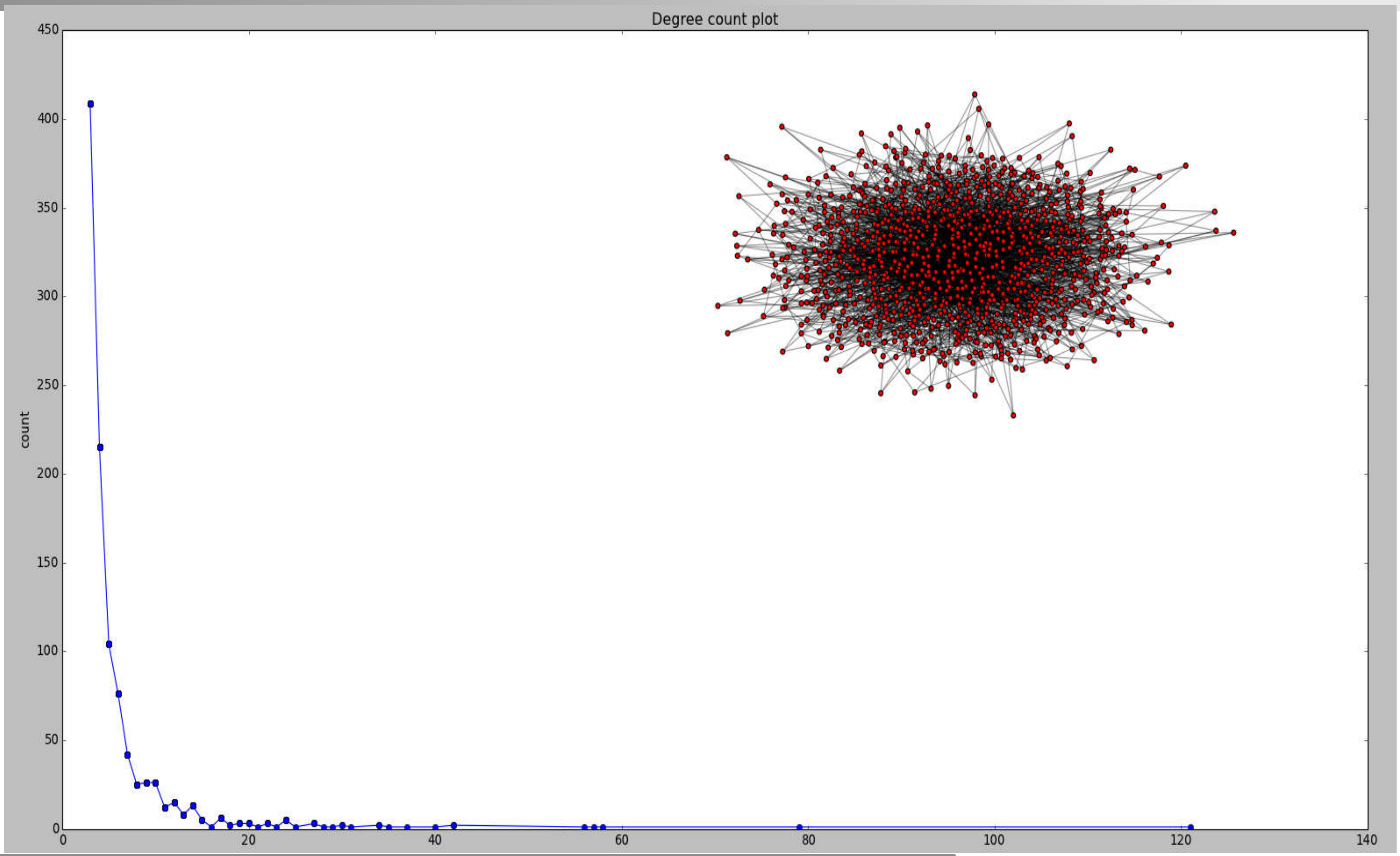
$$p_k = \alpha\, k^{-c}\,, \qquad c > 0$$

Hard to differentiate the values → can this visualization be improved to be useful? How?



Binning

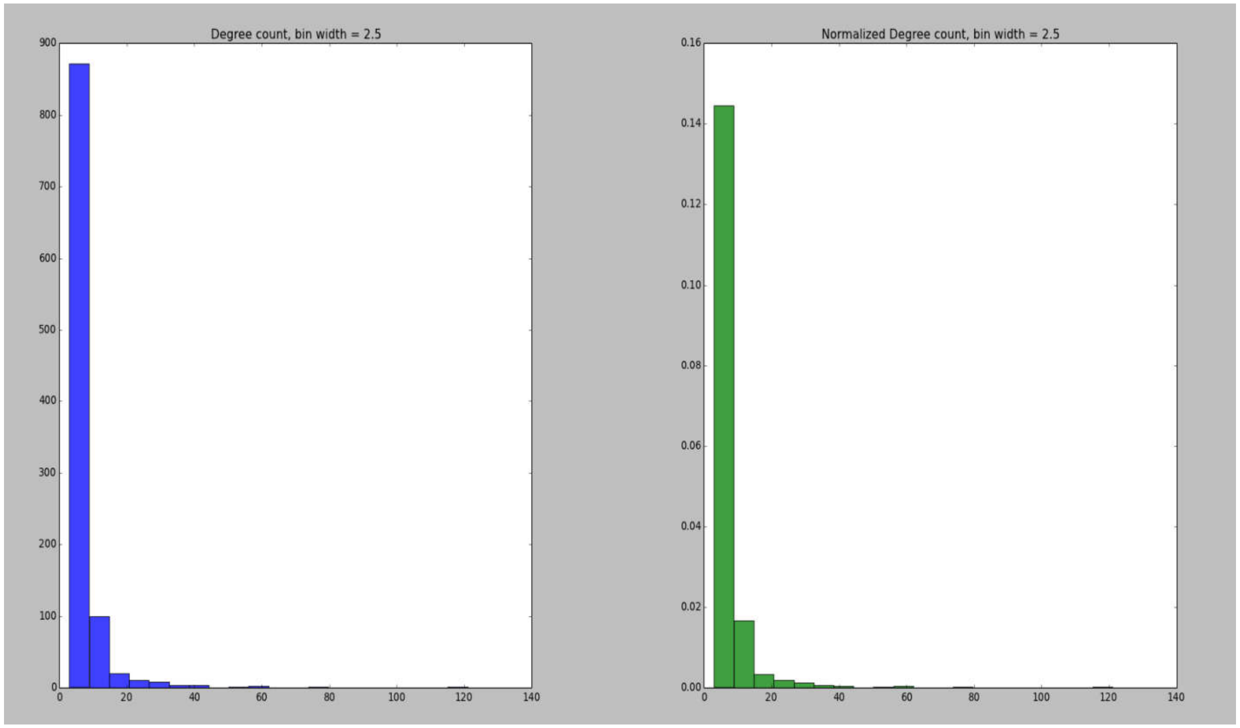
BA Example (n=1000)



Binning

A binned degree histogram will give poor statistics at the tail of the distribution

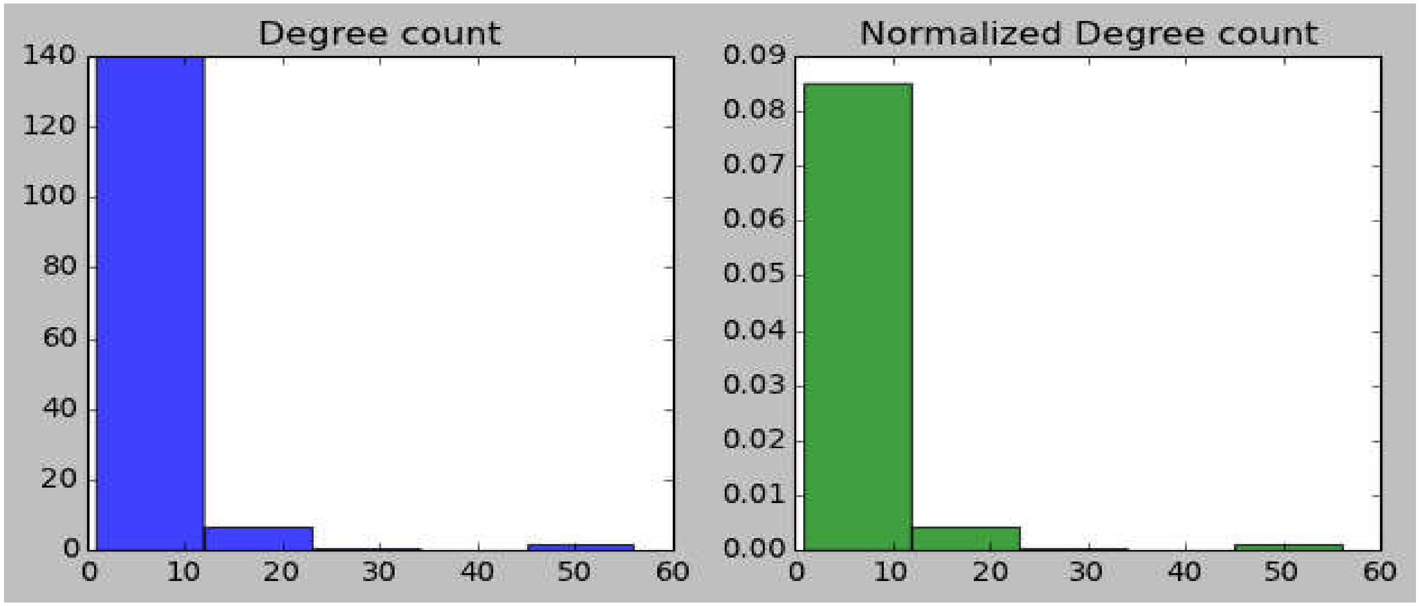
- As k gets large, in every bin there will be only a few samples → large statistical fluctuations in the number of samples from bin to bin → which makes the tail end noisy and difficult to decide if it follows a power-law

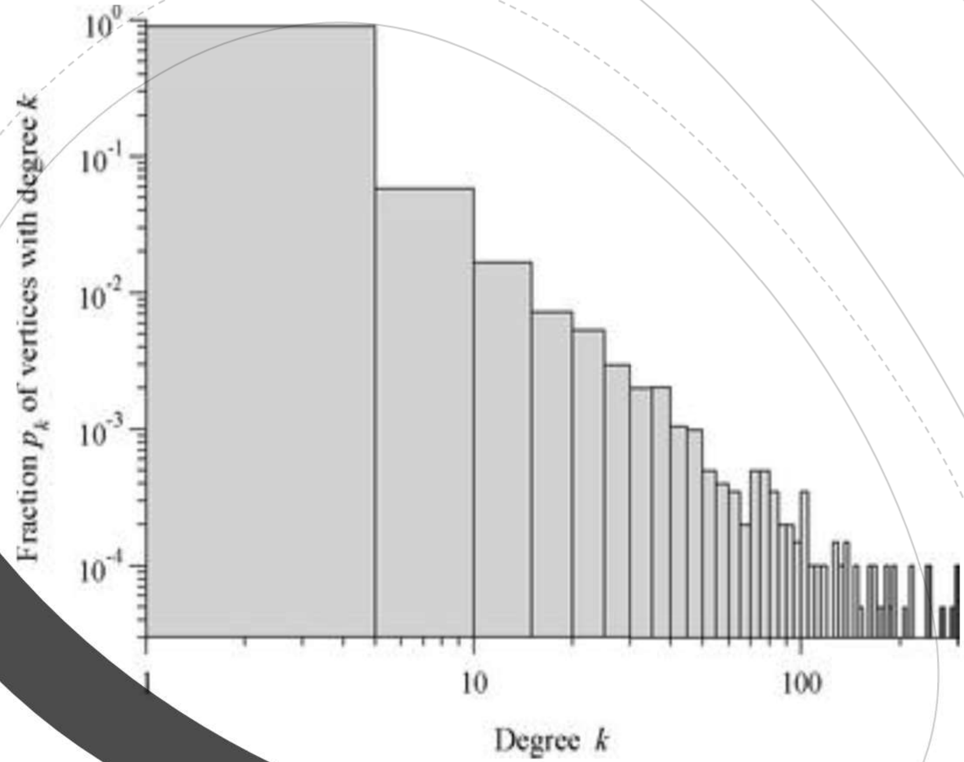


Detecting and visualizing power laws

Alternatives:

- We could use larger bins to reduce the noise at the tail since more samples fall in the same bin
  - However, this reduces the detail captured from the histogram since we end up with fewer bins





Logarithmic Binning

Detecting and visualizing power laws

Alternatives:

Try to get the best of both worlds → different bin sizes in different parts of the histogram

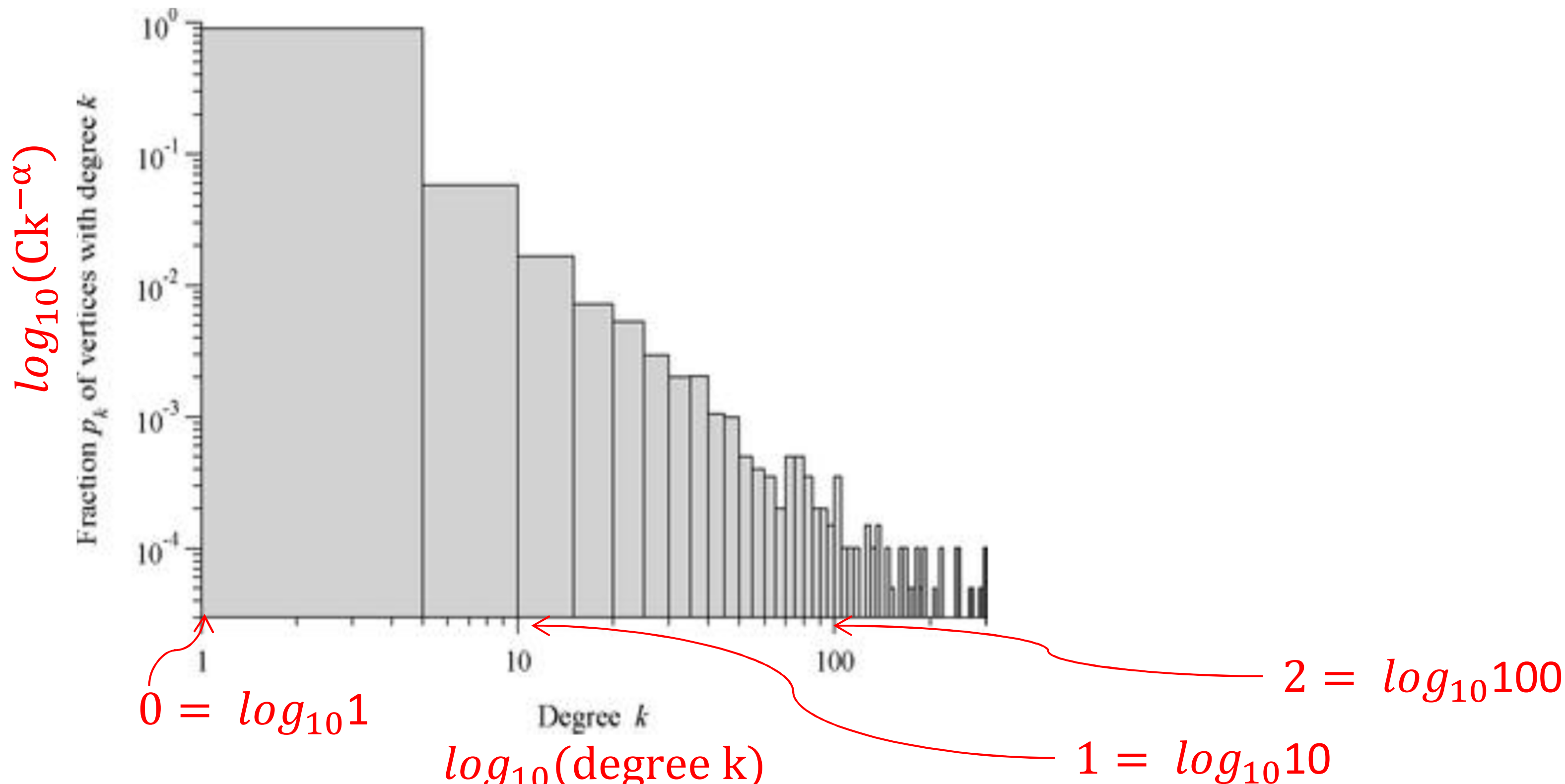
- Careful at normalizing the bins correctly: a bin of width 10 will get 10 times more samples compared to the previous bin of width 1 → divide the count number by 10 → logarithmic view (log-log plots)

In this scheme each bin is made wider than its predecessor by a constant factor *a*

- The *1*-st bin will cover the range: $10^0 \le k < 10^1$
- The *2*-nd bin will cover the range: $10^1 \le k < 10^2$
- The *3*-rd bin will cover the range: $10^2 \le k < 10^3$
  - The most common choices for *a* are 2 and 10 since larger values of *a* give ranges that are too big, and smaller *a* values give non-integer ranges limits
- The *n*-th bin will cover the range: $a^{n-1} \le k < a^n$

Logarithmic binning

- Plot the degree distribution for the Internet in log-log scale: width of the bins is fixed, the display of the width is different



Statistics for real *networks*

α = power in the power law distribution

	Network	Type	n	m	c	S	ℓ	α	C	C_{WS}	
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980	3.48	2.3	0.20	0.78	0.
	Company directors	Undirected	7 673	55 392	14.44	0.876	4.60	–	0.59	0.88	0.
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822	7.57	–	0.15	0.34	0.
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838	6.19	–	0.45	0.56	0.
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918	4.92	–	0.088	0.60	0.
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16			2.1			
	Email messages	Directed	59 812	86 300	1.44	0.952	4.95	1.5/2.0		0.16	
	Email address books	Directed	16 881	57 029	3.38	0.590	5.22	–	0.17	0.13	0.
	Student dating	Undirected	573	477	1.66	0.503	16.01	–	0.005	0.001	−0.
	Sexual contacts	Undirected	2 810					3.2			
Information	WWW nd.edu	Directed	269 504	1 497 135	5.55	1.000	11.27	2.1/2.4	0.11	0.29	−0.
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	0.914	16.18	2.1/2.7			
	Citation network	Directed	783 339	6 716 198	8.57			3.0/–			
	Roget's Thesaurus	Directed	1 022	5 103	4.99	0.977	4.87	–	0.13	0.15	0.
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000		2.7		0.44	
Technological	Internet	Undirected	10 697	31 992	5.98	1.000	3.31	2.5	0.035	0.39	−0.
	Power grid	Undirected	4 941	6 594	2.67	1.000	18.99	–	0.10	0.080	−0.
	Train routes	Undirected	587	19 603	66.79	1.000	2.16	–		0.69	−0.
	Software packages	Directed	1 439	1 723	1.20	0.998	2.42	1.6/1.4	0.070	0.082	−0.
	Software classes	Directed	1 376	2 213	1.61	1.000	5.40	–	0.033	0.012	−0.
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000	11.05	3.0	0.010	0.030	−0.
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805	4.28	2.1	0.012	0.011	−0.
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996	2.56	2.2	0.090	0.67	−0.
	Protein interactions	Undirected	2 115	2 240	2.12	0.689	6.80	2.4	0.072	0.071	−0.
	Marine food web	Directed	134	598	4.46	1.000	2.05	–	0.16	0.23	−0.
	Freshwater food web	Directed	92	997	10.84	1.000	1.90	–	0.20	0.087	−0.
	Neural network	Directed	307	2 359	7.68	0.967	3.97	–	0.18	0.28	−0.

Power laws and scale free networks

- Networks that follow power law degree distribution are referred to as scale-free networks
  - A Scale-free network means that the ratio of very connected nodes to the number of nodes in the rest of the network remains constant as the network changes in size (supports research based on some sampling methods): $\boldsymbol{f(ax) = b\, f(x),\ \forall\, x \in DegSeq(G)}$
- Real networks do not follow power law degree distribution over the whole range of degree *k*
  - That is the degree distribution is not monotonically decreasing over the whole range of degrees
  - Many times we refer to its tail only (high degrees)
    - Deviations from power law can appear for high values of *k* as well

Power laws and scale free networks

- It is important to mention what property of the network is scale-free.
  - Generally it is the degree distribution, and we say that the “network is scale-free” which in reality says “the degree distribution is scale-free”
  - Sometimes the exponent of the degree distribution captures this. Why? The exponent is the slope of the line that fits the data on log-log scale.
  - This has been tested with random subnetworks of scale free models that indeed show scale free degree distribution (but not always the same exponent)

Scale free

- General belief: Scale free networks grow because of preferential attachment
- However:

preferential attachment → scale free

But

preferential attachment $\not\leftarrow$ scale free

(counterexample: the configuration model)

References

- Newman, “The Structure and Function of Complex Networks” http://epubs.siam.org/doi/pdf/10.1137/S003614450342480
- Source: L. Barabasi. Source: http://barabasi.com/networksciencebook/chapter/5#origins
- Source: Watts, DJ; Strogatz, S H. 1998. Collective dynamics of 'small-world' networks, *NATURE* 393 (668).