DEEP LEARNING INSTITUTE

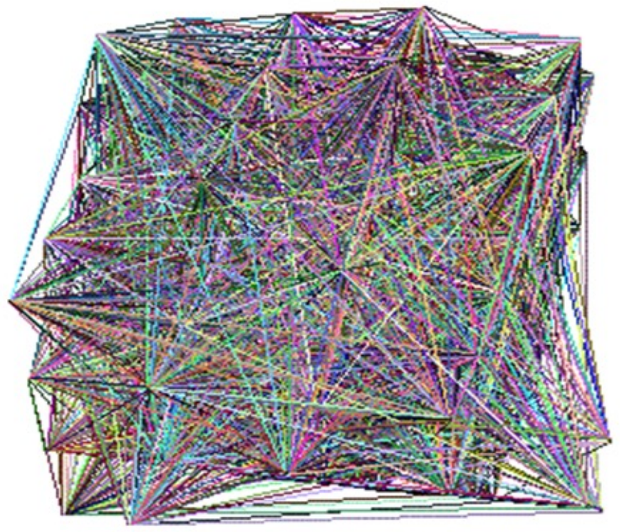
DLI Accelerated Data Science Teaching Kit

Lecture 17.2 - Graph Power Laws

I have a graph. Now what?

Analyze it! Do “data mining” or “graph mining”.

And visualize it if it’s small.

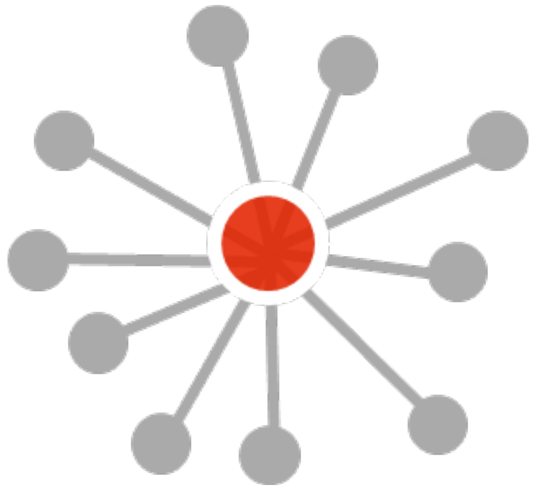


Yuck.

Does it follow any expected patterns? Or does it not follow patterns (outliers)?

Why does this matter?

- If we find patterns (models), we can do prediction, recommendation,
  e.g., is Alice going to “friend” Bob on Facebook?
- Outliers often give us new insights
  e.g., telemarketer’s friends don’t know each other

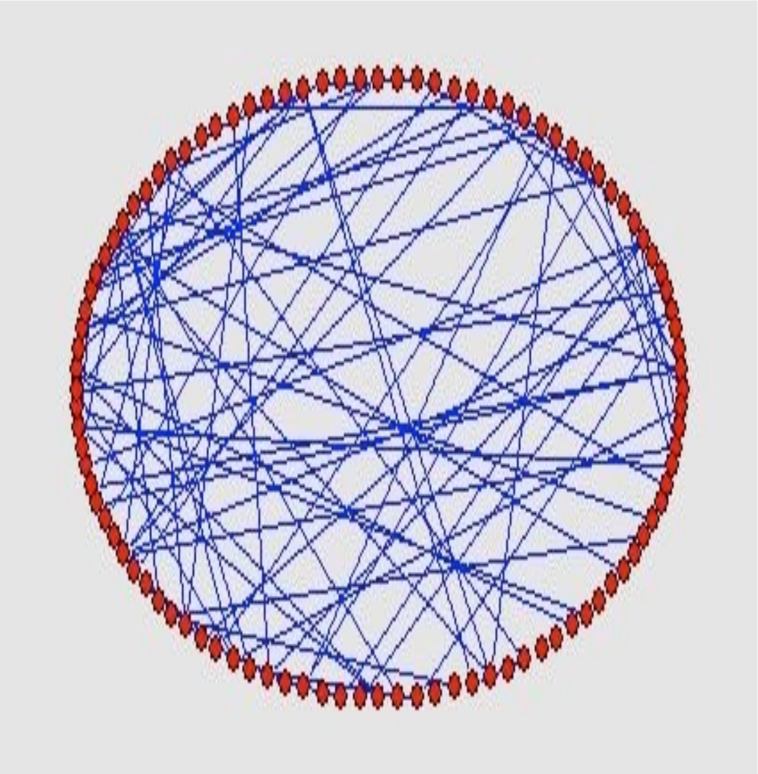


Are Real Graphs Random?

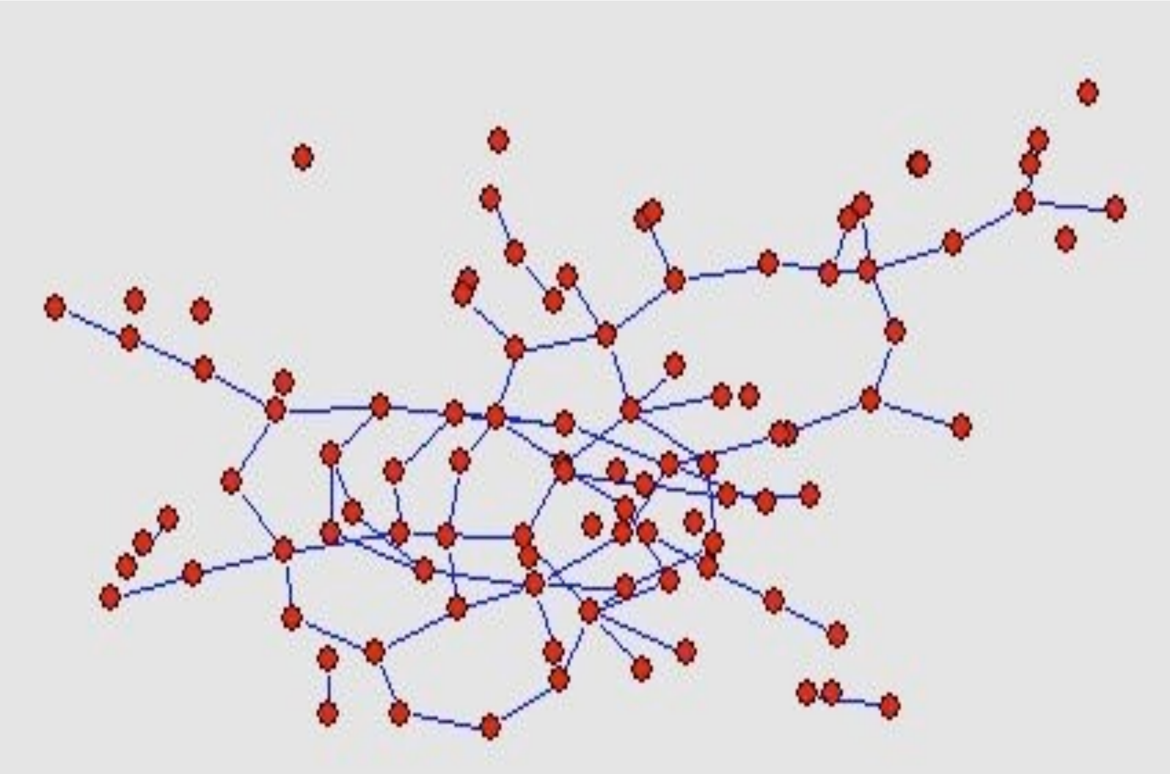
Random graph (Erdos-Renyi): edges are independent, and have equal chance in forming → no obvious patterns

Are real-world graphs (e.g., social networks) random?

Before layout



After layout



Graph and layout generated with pajek

http://vlado.fmf.unilj.si/pub/networks/pajek/

Laws and Patterns

Nope! Real graphs are NOT random.

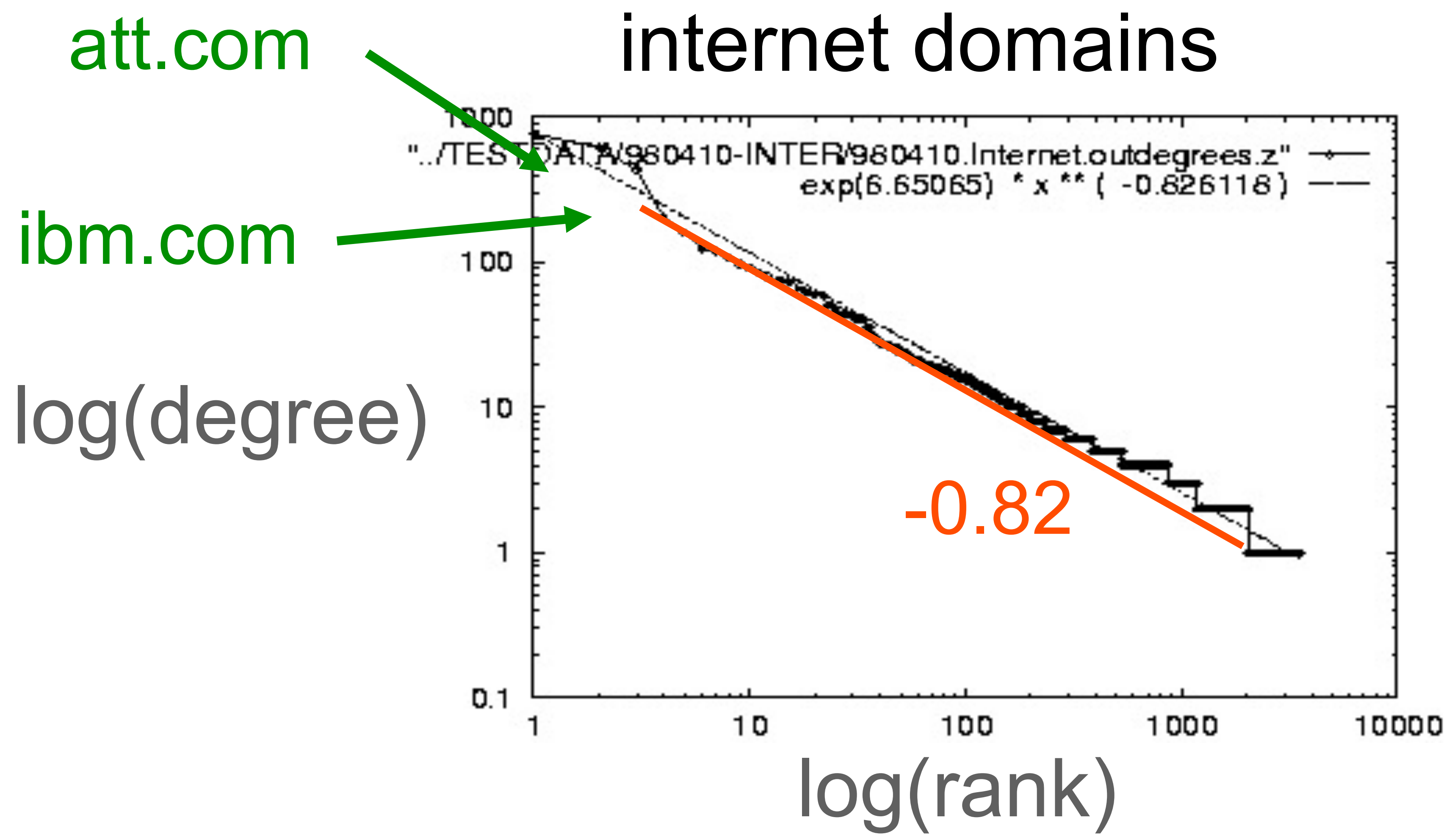
- Diameter (longest shortest path)
- In-degree and out-degree distributions
- Other (surprising) patterns
- So, let's look at the data

Power Law in Degree Distribution

Faloutsos, Faloutsos, Faloutsos [SIGCOMM99]
Seminal paper, by 3 brothers. Must read!

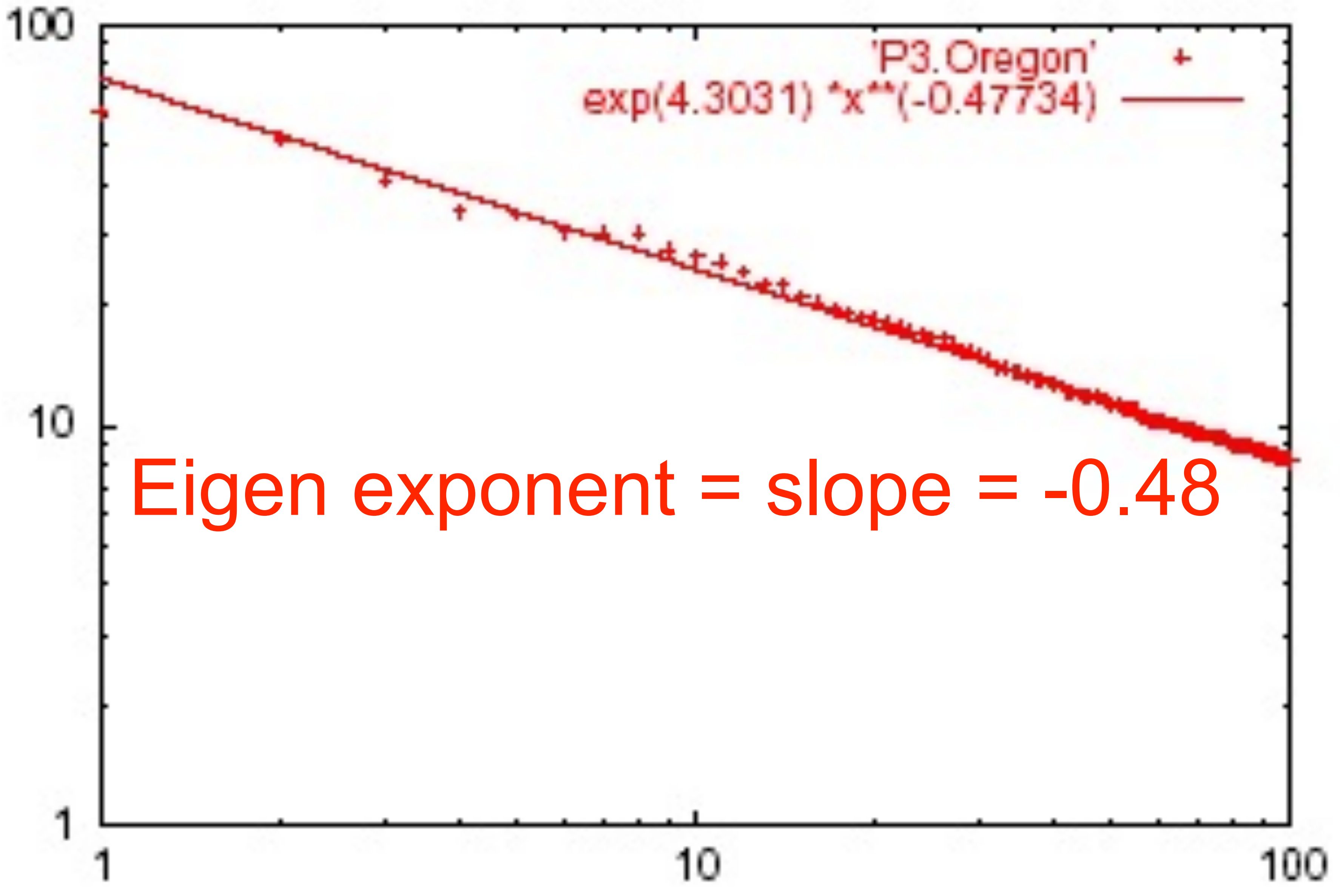
Christos Faloutsos was Polo's advisor

internet domains



Power Law of Eigenvalues

Eigenvalue



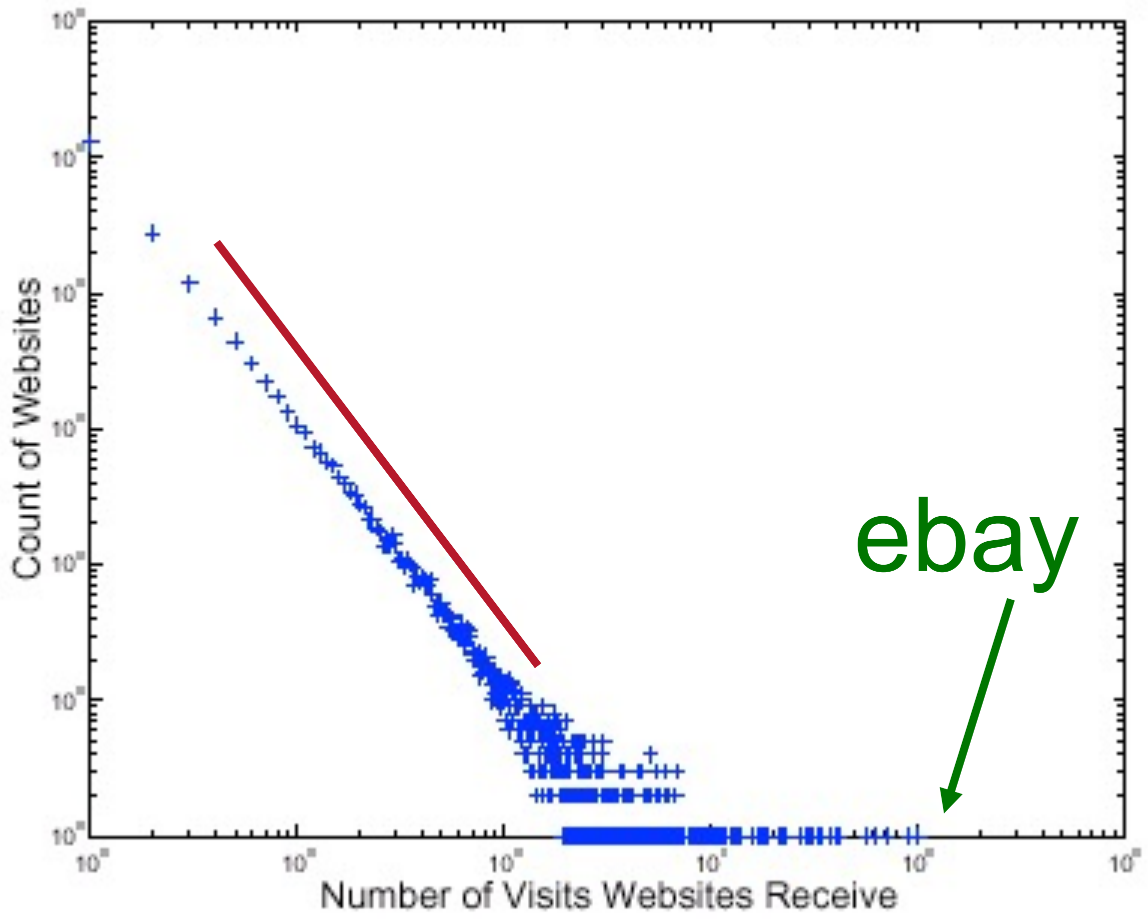
Rank of decreasing eigenvalue

Power Laws in Many Domains

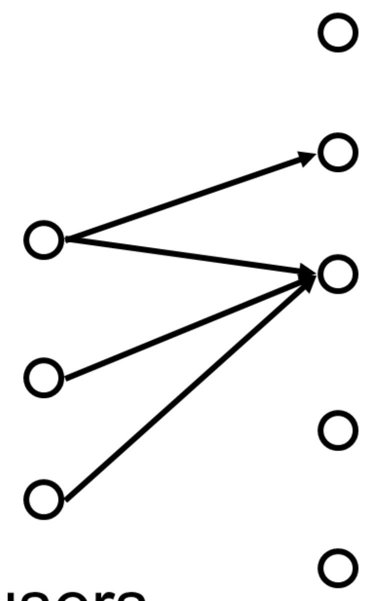
Web site traffic

[Alan L. Montgomery and Christos Faloutsos]

log(#website)



log(#website visit)



users

sites

- # of sexual contacts
- Income [Pareto]:
  80-20 distribution
- Duration of downloads
  [Bestavros+]
- Duration of UNIX jobs
- File sizes
- …

Any other ‘laws’?

Yes! Small diameter (~ constant!)

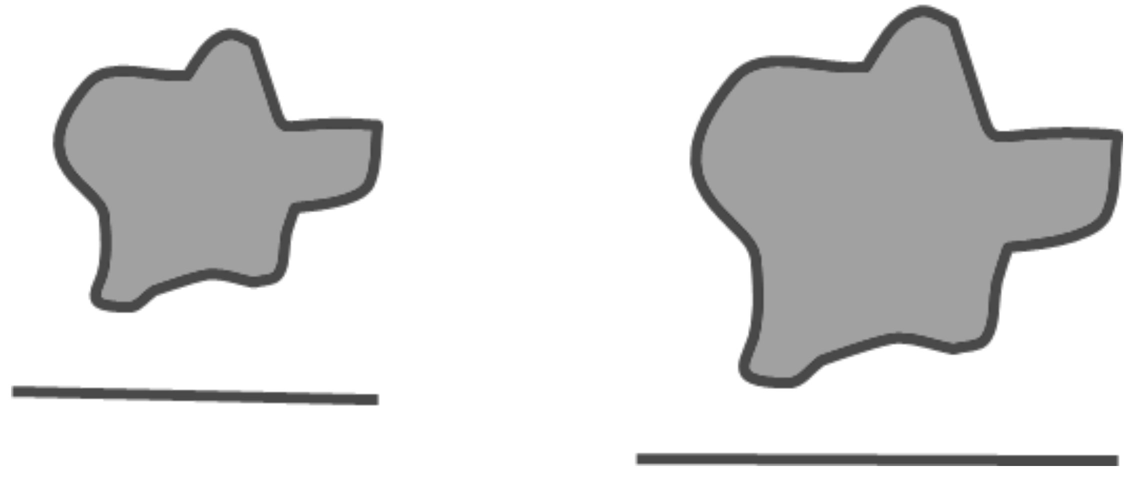
- Six degrees of separation / ‘Kevin Bacon’
- Small worlds [Watts and Strogatz]

But how does the graph diameter change over time?

Evolution of the Diameter

By Leskovec, Kleinberg, Faloutsos

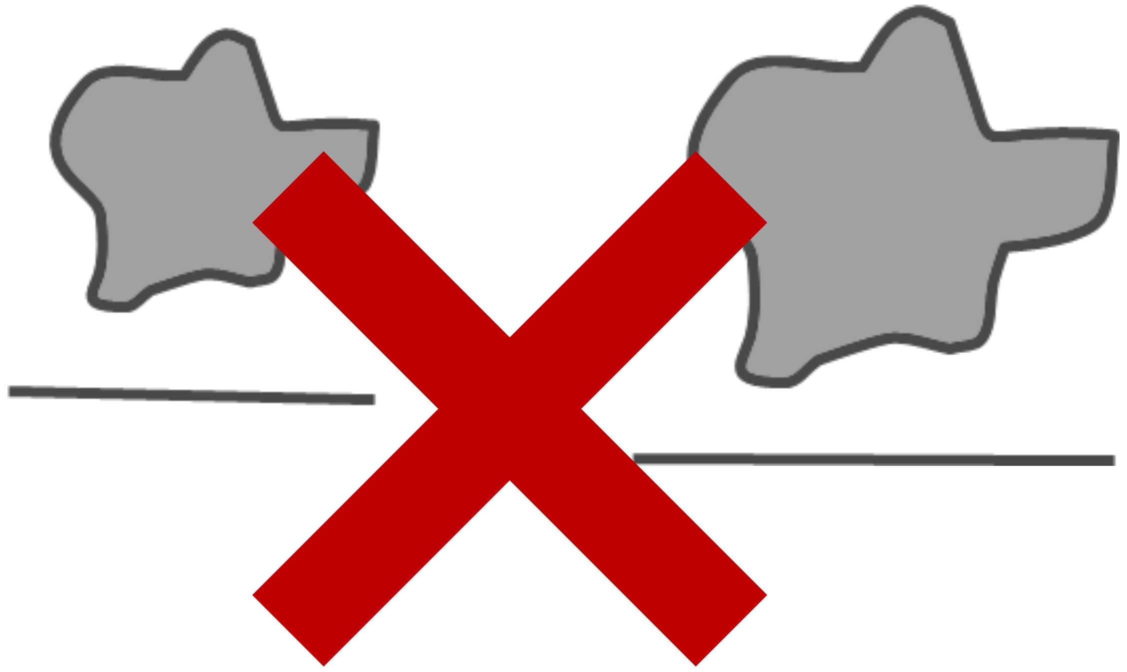
- Prior work on Power Law graphs hints at slowly growing diameter:
  - diameter ~ O(log N)
  - diameter ~ O(log log N)
- What is happening in real data?



Evolution of the Diameter

By Leskovec, Kleinberg, Faloutsos

- Prior work on Power Law graphs hints at slowly growing diameter:
  - diameter $\sim O(\log N)$
  - diameter $\sim O(\log \log N)$
- What is happening in real data?

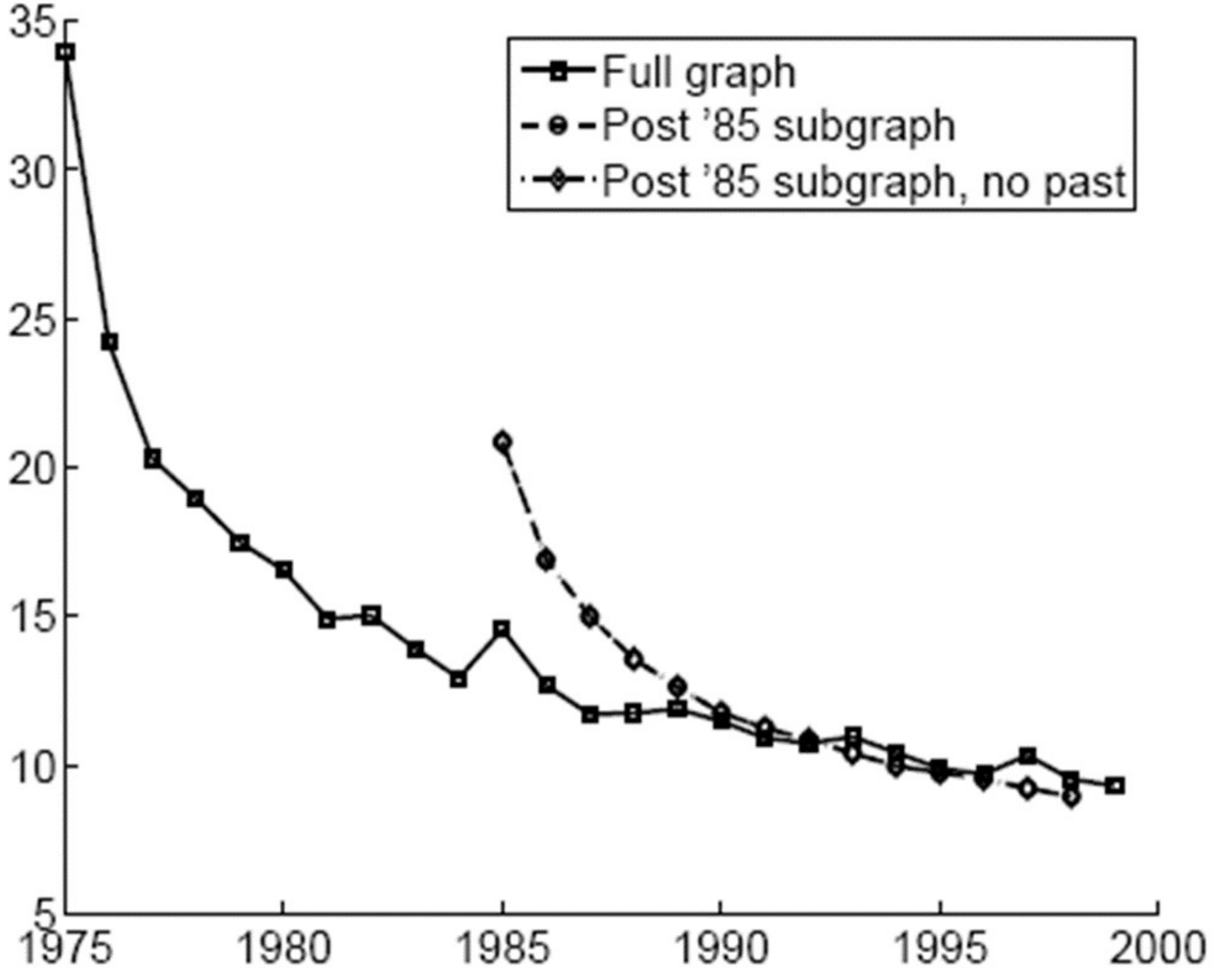


Diameter shrinks over time!

Diameter – Patents Network

- Patent citation network
- 25 years of data
- @1999
  - 2.9 M nodes
  - 16.5 M edges

Effective diameter



Time (year)

NVIDIA DEEP LEARNING INSTITUTE Georgia Tech PRAIRIE VIEW A&M UNIVERSITY

Why Effective Diameter?

The maximum diameter is susceptible to outliers

So, we use effective diameter instead

- Defined as the minimum number of hops in which 90% of connected node pairs can reach each other

Evolution of #Node and #Edge

$N(t)$ = nodes at time t

$E(t)$ = edges at time t

Suppose. $N(t+1) = 2 * N(t)$

Q: What is your guess for $E(t+1)$? $2 * E(t)$?

Evolution of #Node and #Edge

N(t) = nodes at time t

E(t) = edges at time t

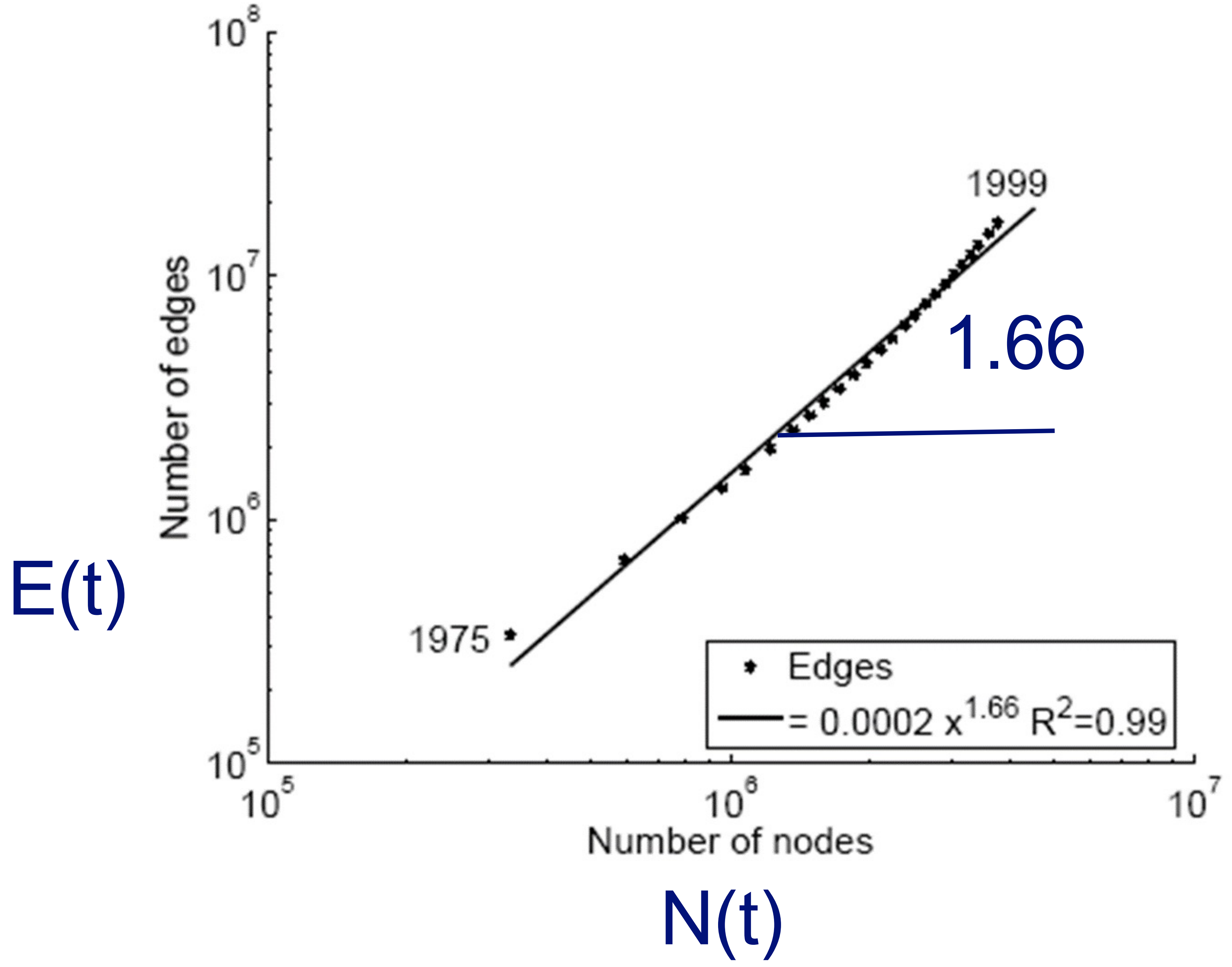
Suppose. N(t+1) = 2 * N(t)

Q: What is your guess for E(t+1)? 2 * E(t) ?

A: Over-doubled! And obeying the "Densification Power Law"

Diameter – Patents Network

- Patent citation network
- 25 years of data
- @1999
    - 2.9 M nodes
    - 16.5 M edges



So Many Laws!

There will be more to come...

To date, there are 11 (or more) laws

RTG: A Recursive Realistic Graph Generator using Random Typing [Akoglu, Faloutsos]

L01 *Power-law degree distribution*: the degree distibution should follow a power-law in the form of $f(d) \propto d^{\gamma}$, with the exponent $\gamma < 0$ [5, 11, 16, 24]

L02 *Densification Power Law (DPL)*: the number of nodes N and the number of edges E should follow a power-law in the form of $E(t) \propto N(t)^{\alpha}$, with $\alpha > 1$, over time [20].

L03 *Weigth Power Law (WPL)*: the total weight of the edges W and the number of edges E should follow a power-law in the form of $W(t) \propto E(t)^{\beta}$, with $\beta > 1$, over time [22].

L04 *Snapshot Power Law (SPL)*: the total weight of the edges W_n attached to each node and the number of such edges, that is, the degree d_n should follow a power-law in the form of $W_n \propto d_n^{\theta}$, with $\theta > 1$ [22].

L05 *Triangle Power Law (TPL)*: the number of triangles Δ and the number of nodes that participate in Δ number of triangles should follow a power-law in the form of $f(\Delta) \propto \Delta^{\sigma}$, with $\sigma < 0$ [29].

L06 *Eigenvalue Power Law (EPL)*: the eigenvalues of the adjacency matrix of the graph should be power-law distributed [28].

L07 *Principal Eigenvalue Power Law ($\lambda_1 PL$)*: the largest eigenvalue λ_1 of the adjacency matrix of the graph and the number of edges E should follow a power-law in the form of $\lambda_1(t) \propto E(t)^{\delta}$, with $\delta < 0.5$, over time [1].

L08 *small and shrinking diameter*: the (effective) diameter of the graph should be small [2] with a possible spike at the 'gelling point' [22]. It should also shrink over time [20].

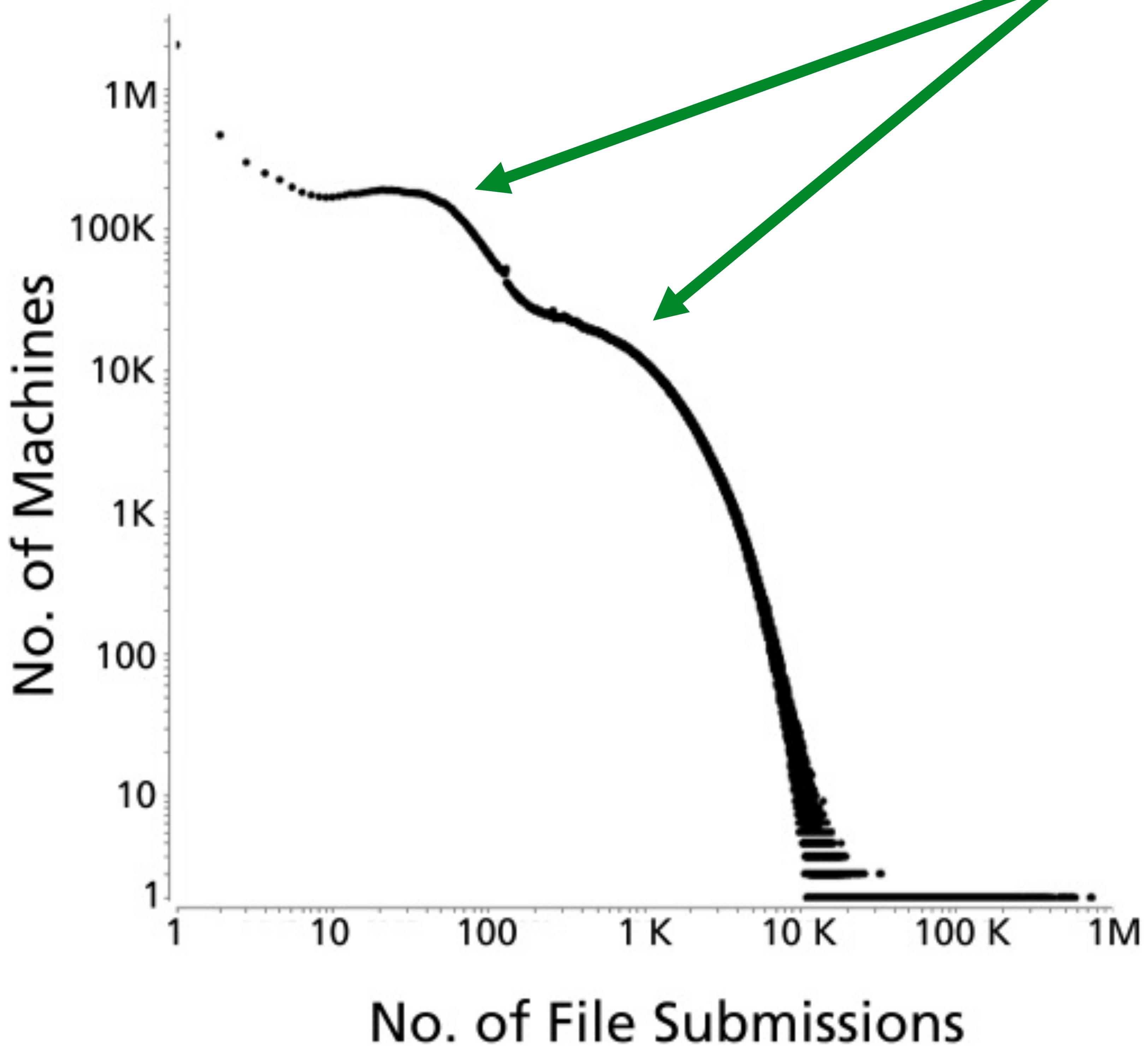
So Many Laws!

What should You Do?

- Try as many distributions as possible and see if your graph fits them.

- If it doesn't, find out the reasons. Sometimes it's due to errors/problems in the data; sometimes, it signifies some new patterns!

What might be the reasons for the “hills”?



Polonium: Tera-Scale Graph Mining and Inference for Malware Detection [Chau, et al]

DLI Accelerated Data Science Teaching Kit

Thank You