# Finding patterns in large, real networks

Christos Faloutsos
*CMU*
www.cs.cmu.edu/~christos/TALKS/UCLA-05



## Thanks to

- Deepayan Chakrabarti (CMU)

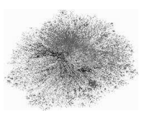
- Michalis Faloutsos (UCR)

- George Siganos (UCR)

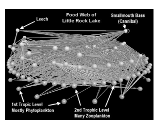




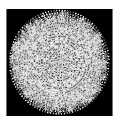
## Introduction



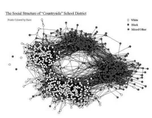
Internet Map
[lumeta.com]



Food Web
[Martinez '91]



Protein Interactions
[genomebiology.com]



Friendship Network
[Moody '01]

Graphs are everywhere!





## Physical graphs

- Physical networks
- Physical Internet
- Telephone lines
- Commodity distribution networks





## Networks derived from "behavior"

- Telephone call patterns
- Email, Blogs, Web, Databases, XML
- Language processing
- Web of trust, epinions.com





## Outline

Topology, 'laws' and generators
- 'Laws' and patterns
- Generators
- Tools

## Motivating questions

- What do real graphs look like?
  - What properties of nodes, edges are important to model?
  - What local and global properties are important to measure?
- How to generate realistic graphs?





## Why should we care?

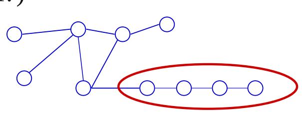
- **A1: extrapolations**: how will the Internet/Web look like next year?
- **A2: algorithm design**: what is a realistic network topology,
  - to try a new routing protocol?
  - to study virus/rumor propagation, and immunization?





## Why should we care? (cont'd)

- **A3: Sampling**: How to get a 'good' sample of a network?
- **A4: Abnormalities**: is this sub-graph / sub-community / sub-network 'normal'? (what **is** normal?)







## Virus propagation

- **Who is the best person/computer to immunize against a virus?**





## Outline

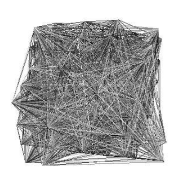
Topology, 'laws' and generators
- 'Laws' and patterns
- Generators
- Tools





## Topology

How does the Internet look like? Any rules?



(Looks random – right?)

Carnegie Mellon

## Laws and patterns

Real graphs are NOT random!!

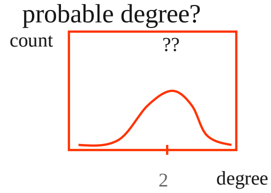
- Diameter
- in- and out- degree distributions
- other (surprising) patterns



Carnegie Mellon

## Laws – degree distributions

- Q: avg degree is ~2 - what is the most probable degree?

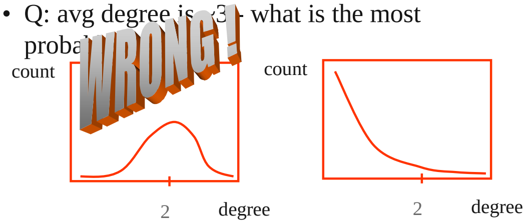





Carnegie Mellon

## Laws – degree distributions

- Q: avg degree is ~3 - what is the most probable degree?

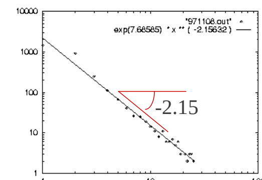





Carnegie Mellon

## I.Power-law: outdegree *O*

Frequency




Outdegree

Exponent = slope

$O = -2.15$

Nov'97

The plot is linear in log-log scale [FFF'99]

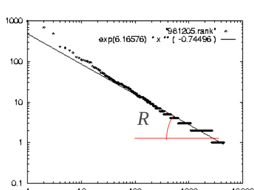
$$freq = degree^{(-2.15)}$$



Carnegie Mellon

## II.Power-law: rank *R*

outdegree




Rank: nodes in decreasing outdegree order

Exponent = slope

$R = -0.74$

Dec'98

- The plot is a line in log-log scale



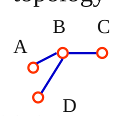
Carnegie Mellon

## III. Eigenvalues

- Let $A$ be the adjacency matrix of graph
- $\lambda$ and $v$ is an eigenvalue/eigenvector pair if:

$$A\, v = \lambda\, v$$

- Eigenvalues are strongly related to graph topology

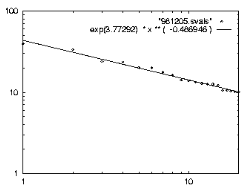



| | A | B | C | D |
|---|---|---|---|---|
| A | | 1 | | |
| B | 1 | | 1 | 1 |
| C | | 1 | | |
| D | | 1 | | |

Carnegie Mellon

## III.Power-law: eigen *E*

Eigenvalue



Exponent = slope

*E = -0.48*

Dec'98

Rank of decreasing eigenvalue

- Eigenvalues in decreasing order (first 20)
- [Mihail+, 02]: $R = 2 * E$

UCLA-IPAM 05 (c) 2005 C. Faloutsos 19

Carnegie Mellon

## IV. The Node Neighborhood

- $N(h)$ = # of pairs of nodes within $h$ hops

UCLA-IPAM 05 (c) 2005 C. Faloutsos 20

Carnegie Mellon

## IV. The Node Neighborhood

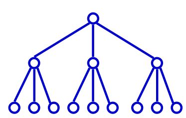
- Q: average degree = 3 - how many neighbors should I expect within 1,2,… *h* hops?
- Potential answer:

  1 hop -> 3 neighbors

  2 hops -> 3 * 3

  …

  *h* hops -> $3^h$



UCLA-IPAM 05 (c) 2005 C. Faloutsos 21

Carnegie Mellon

## IV. The Node Neighborhood

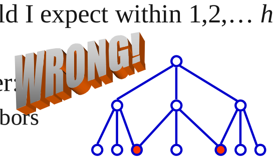
- Q: average degree = 3 - how many neighbors should I expect within 1,2,… *h* hops?
- Potential answer:

  1 hop -> 3 neighbors

  2 hops -> 3 * 3

  …

  h hops -> $3^h$




**WE HAVE DUPLICATES!**

UCLA-IPAM 05 (c) 2005 C. Faloutsos 22

Carnegie Mellon

## IV. The Node Neighborhood

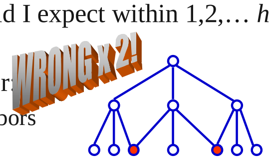
- Q: average degree = 3 - how many neighbors should I expect within 1,2,… *h* hops?
- Potential answer:

  1 hop -> 3 neighbors

  2 hops -> 3 * 3

  …

  h hops -> $3^h$



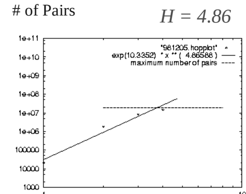

**'avg' degree: meaningless!**

UCLA-IPAM 05 (c) 2005 C. Faloutsos 23

Carnegie Mellon

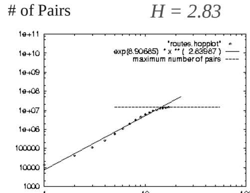
## IV. Power-law: hopplot *H*

# of Pairs *H = 4.86*



Hops Dec 98

# of Pairs *H = 2.83*



Hops Router level '95

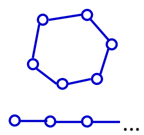
Pairs of nodes as a function of hops $N(h)= h^H$

UCLA-IPAM 05 (c) 2005 C. Faloutsos 24

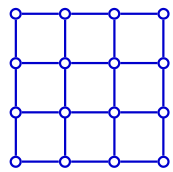
Carnegie Mellon

## Observation

- Q: Intuition behind ‘hop exponent’?
- A: ‘intrinsic=fractal dimensionality’ of the network



N($h$) ~ $h^1$



N($h$) ~ $h^2$

UCLA-IPAM 05 (c) 2005 C. Faloutsos 25

Carnegie Mellon

## Hop plots

- More on fractal/intrinsic dimensionalities: very soon

UCLA-IPAM 05 (c) 2005 C. Faloutsos 26

Carnegie Mellon

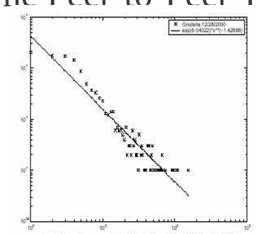
## But:

- Q1: How about graphs from other domains?
- Q2: How about temporal evolution?

UCLA-IPAM 05 (c) 2005 C. Faloutsos 27

Carnegie Mellon

## The Peer-to-Peer Topology




[Jovanovic+]

- Frequency versus degree
- Number of adjacent peers follows a power-law

UCLA-IPAM 05 (c) 2005 C. Faloutsos 28

Carnegie Mellon

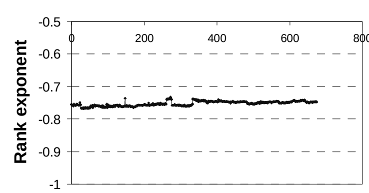
## More Power laws

- Also hold for other web graphs [Barabasi+, ‘99], [Kumar+, ‘99]
- citation graphs (see later)
- and many more

UCLA-IPAM 05 (c) 2005 C. Faloutsos 29

Carnegie Mellon

## Time Evolution: rank *R*




***Domain level***

**#days since Nov. ‘97**

The rank exponent has not changed! [Siganos+, ‘03]

UCLA-IPAM 05 (c) 2005 C. Faloutsos 30

## Outline

Part 1: Topology, ‘laws’ and generators
- ‘Laws’ and patterns
  - Power laws for degree, eigenvalues, hop-plot
  - ???
- Generators
- Tools

Part 2: PageRank, HITS and eigenvalues





## Any other ‘laws’?

Yes!





## Any other ‘laws’?

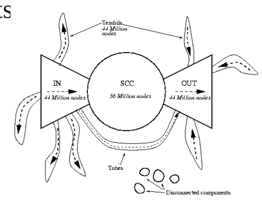
Yes!
- Small diameter
  – six degrees of separation / ‘Kevin Bacon’
  – small worlds [Watts and Strogatz]





## Any other ‘laws’?

- Bow-tie, for the web [Kumar+ ‘99]
- IN, SCC, OUT, ‘tendrils’
- disconnected components

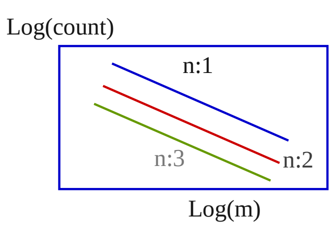


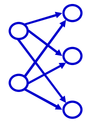





## Any other ‘laws’?

- power-laws in communities (bi-partite cores) [Kumar+, ‘99]



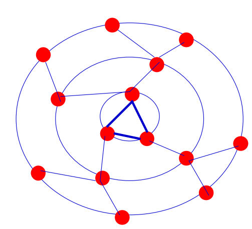
2:3 core
(m:n core)





## Any other ‘laws’?

- “Jellyfish” for Internet [Tauro+ ’01]
- core: ~clique
- ~5 concentric layers
- many 1-degree nodes



Carnegie Mellon

## How do graphs evolve?

- degree-exponent seems constant - anything else?



Carnegie Mellon

## Evolution of diameter?

- Prior analysis, on power-law-like graphs, hints that

  diameter ~ O(log(N)) or

  diameter ~ O( log(log(N)))
- i.e.., slowly increasing with network size
- Q: What is happening, in reality?



Carnegie Mellon

## Evolution of diameter?

- Prior analysis, on power-law-like graphs, hints that

  diameter ~ O(log(N)) or

  diameter ~ O( log(log(N)))
- i.e.., slowly increasing with network size
- Q: What is happening, in reality?
- A: It **shrinks**(!!), towards a constant value

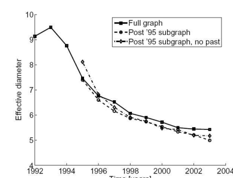


Carnegie Mellon

## Shrinking diameter

ArXiv physics papers and their citations

[Leskovec+05a]



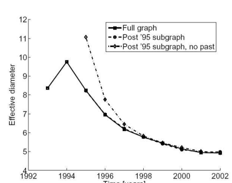

(a) arXiv citation graph



Carnegie Mellon

## Shrinking diameter

ArXiv: who co-authored with whom



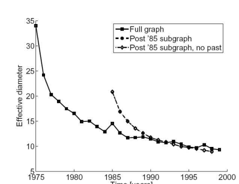

(b) Affiliation network



Carnegie Mellon

## Shrinking diameter

U.S. patents citing each other

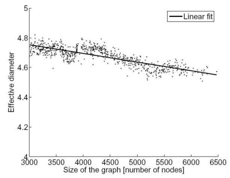



(c) Patents

Carnegie Mellon

## Shrinking diameter

Autonomous systems




(d) AS

UCLA-IPAM 05 (c) 2005 C. Faloutsos 43

Carnegie Mellon

## Temporal evolution of graphs

- N(t) nodes; E(t) edges at time t
- suppose that

  N(t+1) = 2 * N(t)
- Q: what is your guess for

  E(t+1) =? ...* E(t)

UCLA-IPAM 05 (c) 2005 C. Faloutsos 44

Carnegie Mellon

## Temporal evolution of graphs

- N(t) nodes; E(t) edges at time t
- suppose that

  N(t+1) = 2 * N(t)
- Q: what is your guess for

  E(t+1) =? ...* E(t)
- A: over-doubled!

UCLA-IPAM 05 (c) 2005 C. Faloutsos 45

Carnegie Mellon

## Temporal evolution of graphs

- A: over-doubled - but obeying:

  $\boxed{E(t) \sim N(t)^a \quad \text{for all } t}$

  where $1<a<2$

  a=1: constant avg degree

  a=2: ~full clique
- Real graphs **densify** over time [Leskovec+05a]

UCLA-IPAM 05 (c) 2005 C. Faloutsos 46

Carnegie Mellon

## Temporal evolution of graphs

- A: over-doubled - but obeying:

  $\boxed{E(t) \sim N(t)^a \quad \text{for all } t}$
- Identically:

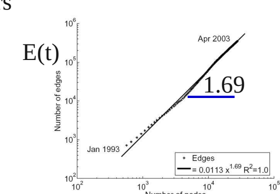
  $\boxed{\log(E(t)) / \log(N(t)) = \text{constant}}$ for all $t$

UCLA-IPAM 05 (c) 2005 C. Faloutsos 47

Carnegie Mellon

## Densification Power Law

ArXiv: Physics papers and their citations




(a) arXiv

N(t)

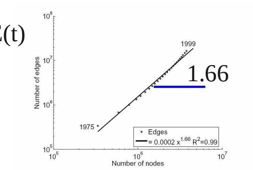
UCLA-IPAM 05 (c) 2005 C. Faloutsos 48

Carnegie Mellon

## Densification Power Law

U.S. Patents, citing each other

E(t)




(b) Patents

N(t)

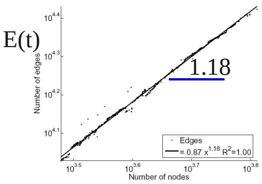
UCLA-IPAM 05 (c) 2005 C. Faloutsos 49

Carnegie Mellon

## Densification Power Law

Autonomous Systems

E(t)




(c) Autonomous Systems

N(t)

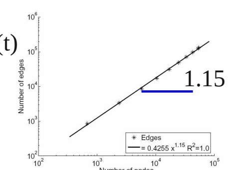
UCLA-IPAM 05 (c) 2005 C. Faloutsos 50

Carnegie Mellon

## Densification Power Law

ArXiv: who co-authored with whom

E(t)




(d) Affiliation network

N(t)

UCLA-IPAM 05 (c) 2005 C. Faloutsos 51

Carnegie Mellon

## Summary of ‘laws’

- Power laws for degree distributions
- …………... for eigenvalues, bi-partite cores
- Small & **shrinking** diameter (‘6 degrees’)
- ‘Bow-tie’ for web; ‘jelly-fish’ for internet
- ``Densification Power Law’’, over time

UCLA-IPAM 05 (c) 2005 C. Faloutsos 52

Carnegie Mellon

## Outline

Part 1: Topology, ‘laws’ and generators
- ‘Laws’ and patterns
- ➡ Generators
- Tools

UCLA-IPAM 05 (c) 2005 C. Faloutsos 53

Carnegie Mellon

## Generators

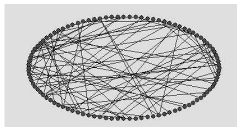
- How to generate random, realistic graphs?
  - Erdos-Renyi model: beautiful, but unrealistic
  - process-based generators
  - recursive generators

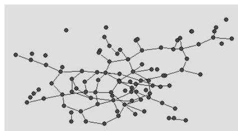
UCLA-IPAM 05 (c) 2005 C. Faloutsos 54

Carnegie Mellon

## Erdos-Renyi

- random graph – 100 nodes, avg degree = 2
- Fascinating properties (phase transition)
- But: unrealistic (Poisson degree distribution != power law)







Carnegie Mellon

## Process-based

- Barabasi; Barabasi-Albert: Preferential attachment -> power-law tails!
  - ‘rich get richer’
- [Kumar+]: preferential attachment + mimic
  - Create ‘communities’



Carnegie Mellon

## Process-based (cont’d)

- [Fabrikant+, ‘02]: H.O.T.: connect to closest, high connectivity neighbor
- [Pennock+, ‘02]: Winner does NOT take all
- ... and many more



Carnegie Mellon

## Recursive generators - intuition

- recursion <-> self-similarity <-> power laws
  (see details later)
- Recursion -> communities within communities within communities



Carnegie Mellon

## Wish list for a generator:

- Power-law-tail in- and out-degrees
- Power-law-tail scree plots
- **shrinking/constant** diameter
- Densification Power Law
- communities-within-communities

Q: how to achieve all of them?



Carnegie Mellon

## Wish list for a generator:

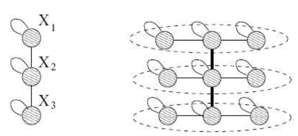
- Power-law-tail in- and out-degrees
- Power-law-tail scree plots
- shrinking/constant diameter
- Densification Power Law
- communities-within-communities

Q: how to achieve all of them?

A: Kronecker matrix product [Leskovec+05b]

**Carnegie Mellon**

# Kronecker product

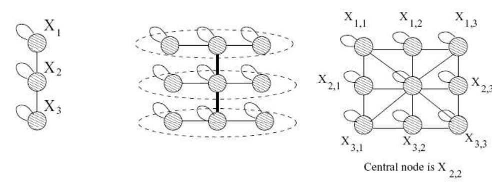



(a) Graph $G_1$ (b) Intermediate stage



**Carnegie Mellon**

# Kronecker product

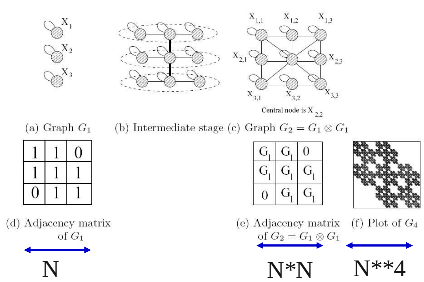



(a) Graph $G_1$ (b) Intermediate stage (c) Graph $G_2 = G_1 \otimes G_1$



**Carnegie Mellon**

# Kronecker product






**Carnegie Mellon**

# Properties of Kronecker graphs:

- Power-law-tail in- and out-degrees
- Power-law-tail scree plots
- **constant** diameter
- perfect Densification Power Law
- communities-within-communities



**Carnegie Mellon**

# Properties of Kronecker graphs:

- Power-law-tail in- and out-degrees
- Power-law-tail scree plots
- **constant** diameter
- perfect Densification Power Law
- communities-within-communities

**and we can <u>prove all</u> of the above**

(first and only generator that does that)



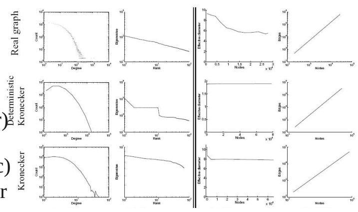
**Carnegie Mellon**

# Properties of Kronecker graphs:

- ‘stochastic’ version gives even better results and
  - Includes Erdos-Renyi as special case
  - Includes ‘RMAT’ as special case [Chakrabarti+,’04]
- (stochastic version: generate Kronecker matrix; decimate edges with some probability)

Carnegie Mellon

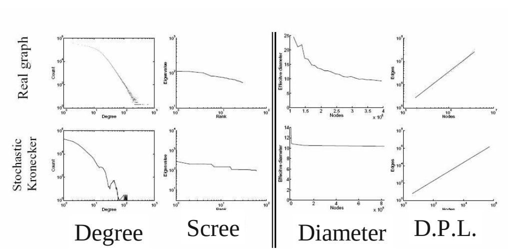
# Kronecker - ArXiv




UCLA-IPAM 05 (c) 2005 C. Faloutsos 67

Carnegie Mellon

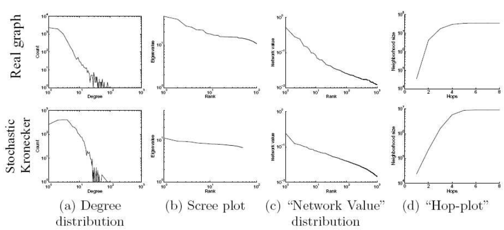
# Kronecker - patents




UCLA-IPAM 05 (c) 2005 C. Faloutsos 68

Carnegie Mellon

# Kronecker - A.S.




UCLA-IPAM 05 (c) 2005 C. Faloutsos 69

Carnegie Mellon

# Conclusions

'Laws' and patterns:

- Power laws for degrees, eigenvalues, 'communities'/cores
- Small / Shrinking diameter
- Bow-tie; jelly-fish

UCLA-IPAM 05 (c) 2005 C. Faloutsos 70

Carnegie Mellon

# Conclusions, cont'd

Generators

- Preferential attachment (Barabasi)
- Variations
- Recursion – Kronecker product & RMAT

UCLA-IPAM 05 (c) 2005 C. Faloutsos 71

Carnegie Mellon

# Outline

Topology, 'laws' and generators

- 'Laws' and patterns
- Generators
- ➡ Tools

UCLA-IPAM 05 (c) 2005 C. Faloutsos 72

Carnegie Mellon

# Outline

Part 1: Topology, ʻlawsʼ and generators

- ʻLawsʼ and patterns
- Generators
- Tools: power laws and fractals
  - Why so many power laws?
  - Self-similarity, power laws, fractal dimension



Carnegie Mellon

# Power laws

- Q1: Are they only in graph-related settings?
- A1:
- Q2: Why so many?
- A2:



Carnegie Mellon

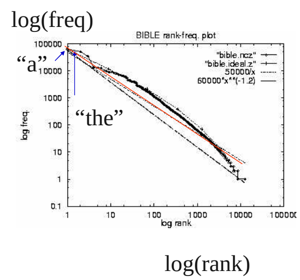
# Power laws

- Q1: Are they only in graph-related settings?
- A1: NO!
- Q2: Why so many?
- A2: self-similarity; ‘rich-get-richer’



Carnegie Mellon

# A famous power law: Zipf’s law




- Bible - **rank** vs **frequency** (log-log)



Carnegie Mellon

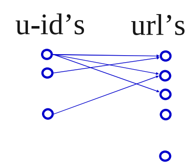
# Power laws, cont’ed

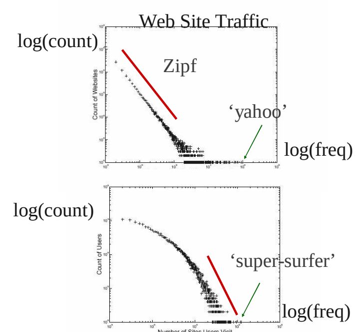
- length of file transfers [Bestavros+]
- web hit counts [Huberman]
- magnitude of earthquakes (Guttenberg-Richter law)
- sizes of lakes/islands (Korcak’s law)
- Income distribution (Pareto’s law)



# Click-stream data

Carnegie Mellon

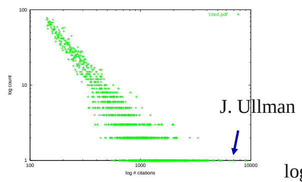


## Lotka's law

(Lotka's law of publication count); and citation counts: (*citeseer.nj.nec.com* 6/2001)

log(count)




log(#citations)





## Power laws

- Q1: Are they only in graph-related settings?
- A1: NO!
- Q2: Why so many?
- A2: self-similarity; 'rich-get-richer'





## Fractals and power laws

- Power laws and fractals are closely related
- And fractals appear in **MANY** cases
  - coast-lines: 1.1-1.5
  - brain-surface: 2.6
  - rain-patches: 1.3
  - tree-bark: ~2.1
  - stock prices / random walks: 1.5
  - ... [see Mandelbrot; or Schroeder]





## Digression: intro to fractals

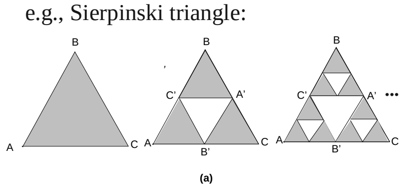
- Fractals: sets of points that are self similar





## A famous fractal

e.g., Sierpinski triangle:




**zero area;**
**infinite length!**

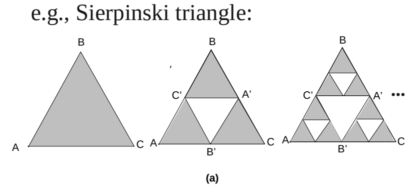
dimensionality = ??





## A famous fractal

e.g., Sierpinski triangle:

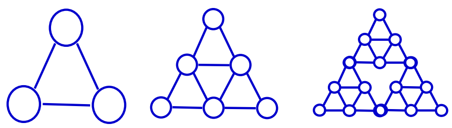



**zero area;**
**infinite length!**

dimensionality = log(3)/log(2) = 1.58

## A famous fractal

equivalent graph:







## Intrinsic ('fractal') dimension

How to estimate it?





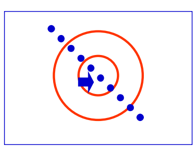
## Intrinsic ('fractal') dimension

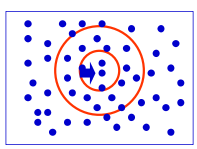
- Q: fractal dimension of a line?
- A: nn ( <= r ) ~ r^1

('power law': y=x^a)

- Q: fd of a plane?
- A: nn ( <= r ) ~ r^2

fd== slope of (log(nn) vs log(r) )





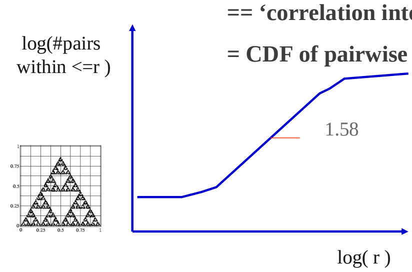




## Sierpinsky triangle

**== 'correlation integral'**

**= CDF of pairwise distances**





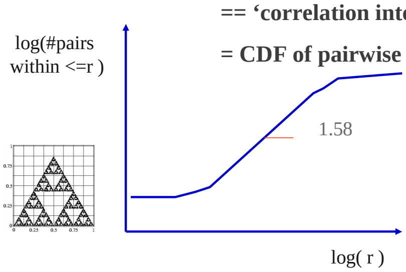



## Sierpinsky triangle

**== hopplot**

**== 'correlation integral'**

**= CDF of pairwise distances**



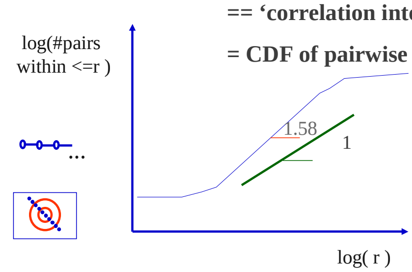





## Line

**== 'correlation integral'**

**= CDF of pairwise distances**

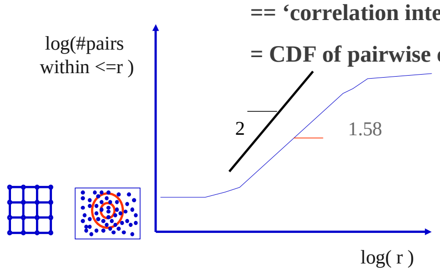


Carnegie Mellon

## 2-d (Plane)

== ‘correlation integral’

= CDF of pairwise distances

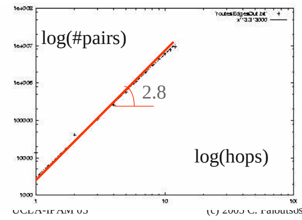



UCLA-IPAM 05 (c) 2005 C. Faloutsos 91

Carnegie Mellon

## Recall: Hop Plot

- Internet routers: how many neighbors within *h* hops? (= **correlation integral**!)




Reachability function: number of neighbors within r hops, vs r (log-log).

Mbone routers, 1995

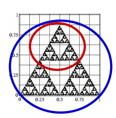
UCLA-IPAM 05 (c) 2005 C. Faloutsos 92

Carnegie Mellon

## Fractals and power laws

They are related concepts:

- fractals <=>
- self-similarity <=>
- scale-free <=>
- power laws ( $y= x^a$ )
  - $F = C r^{(-2)}$

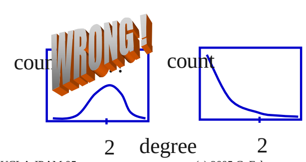


UCLA-IPAM 05 (c) 2005 C. Faloutsos 93

Carnegie Mellon

## Conclusions

- Real settings/graphs: skewed distributions
  - ‘mean’ is meaningless

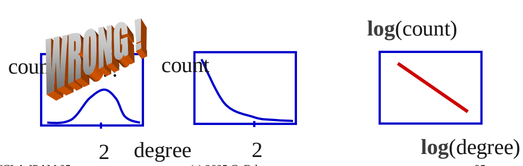



UCLA-IPAM 05 (c) 2005 C. Faloutsos 94

Carnegie Mellon

## Conclusions

- Real settings/graphs: skewed distributions
  - ‘mean’ is meaningless
  - slope of power law, instead




UCLA-IPAM 05 (c) 2005 C. Faloutsos 95

Carnegie Mellon

## Conclusions: Tools:

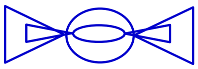
- rank-frequency plot (a’la Zipf)
- Correlation integral (= neighborhood function)

UCLA-IPAM 05 (c) 2005 C. Faloutsos 96

Carnegie Mellon

## Conclusions (cont’d)

- Recursion/self-similarity
  - May reveal non-obvious patterns (e.g., bow-ties within bow-ties within bow-ties) [Dill+, ‘01]



**“To iterate is human, to recurse is divine”**

UCLA-IPAM 05 (c) 2005 C. Faloutsos 97

Carnegie Mellon

## Resources

Generators:

- RMAT (deepay AT cs.cmu.edu)
- Kronecker ({deepay,jure} AT cs.cmu.edu)
- BRITE http://www.cs.bu.edu/brite/
- INET: http://topology.eecs.umich.edu/inet

UCLA-IPAM 05 (c) 2005 C. Faloutsos 98

Carnegie Mellon

## Other resources

Visualization - graph algo’s:

- Graphviz: http://www.graphviz.org/
- pajek: http://vlado.fmf.uni-lj.si/pub/networks/pajek/

Kevin Bacon web site: http://www.cs.virginia.edu/oracle/

UCLA-IPAM 05 (c) 2005 C. Faloutsos 99

Carnegie Mellon

## References


- [Aiello+, '00] William Aiello, Fan R. K. Chung, Linyuan Lu: *A random graph model for massive graphs*. STOC 2000: 171-180
- [Albert+] Reka Albert, Hawoong Jeong, and Albert-Laszlo Barabasi: *Diameter of the World Wide Web*, Nature 401 130-131 (1999)
- [Barabasi, '03] Albert-Laszlo Barabasi *Linked: How Everything Is Connected to Everything Else and What It Means* (Plume, 2003)


UCLA-IPAM 05 (c) 2005 C. Faloutsos 100

Carnegie Mellon

## References, cont’d


- [Barabasi+, '99] Albert-Laszlo Barabasi and Reka Albert. *Emergence of scaling in random networks*. Science, 286:509--512, 1999
- [Broder+, '00] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. *Graph structure in the web*, WWW, 2000


UCLA-IPAM 05 (c) 2005 C. Faloutsos 101

Carnegie Mellon

## References, cont’d


- [Chakrabarti+, ‘04] *RMAT: A recursive graph generator*, D. Chakrabarti, Y. Zhan, C. Faloutsos, SIAM-DM 2004
- [Dill+, '01] Stephen Dill, Ravi Kumar, Kevin S. McCurley, Sridhar Rajagopalan, D. Sivakumar, Andrew Tomkins: *Self-similarity in the Web*. VLDB 2001: 69-78


UCLA-IPAM 05 (c) 2005 C. Faloutsos 102

Carnegie Mellon

## References, cont'd


- [Fabrikant+, '02] A. Fabrikant, E. Koutsoupias, and C.H. Papadimitriou. *Heuristically Optimized Trade-offs: A New Paradigm for Power Laws in the Internet*. ICALP, Malaga, Spain, July 2002
- [FFF, 99] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "*On power-law relationships of the Internet topology*," in SIGCOMM, 1999.


UCLA-IPAM 05 (c) 2005 C. Faloutsos 103

Carnegie Mellon

## References, cont'd


- [Leskovec+05a] Jure Leskovec, Jon Kleinberg and Christos Faloutsos *Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations* KDD 2005, Chicago, IL. **(Best research paper award)**
- [Leskovec+05b] Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg and Christos Faloutsos, *Realistic, Mathematically Tractable Graph Generation and Evolution, Using Kronecker Multiplication*, ECML/PKDD 2005, Porto, Portugal.


UCLA-IPAM 05 (c) 2005 C. Faloutsos 104

Carnegie Mellon

## References, cont'd


- [Jovanovic+, '01] M. Jovanovic, F.S. Annexstein, and K.A. Berman. *Modeling Peer-to-Peer Network Topologies through "Small-World" Models and Power Laws*. In TELFOR, Belgrade, Yugoslavia, November, 2001
- [Kumar+ '99] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, Andrew Tomkins: *Extracting Large-Scale Knowledge Bases from the Web*. VLDB 1999: 639-650


UCLA-IPAM 05 (c) 2005 C. Faloutsos 105

Carnegie Mellon

## References, cont'd


- [Leland+, '94] W. E. Leland, M.S. Taqqu, W. Willinger, D.V. Wilson, *On the Self-Similar Nature of Ethernet Traffic*, IEEE Transactions on Networking, 2, 1, pp 1-15, Feb. 1994.
- [Mihail+, '02] Milena Mihail, Christos H. Papadimitriou: *On the Eigenvalue Power Law*. RANDOM 2002: 254-262


UCLA-IPAM 05 (c) 2005 C. Faloutsos 106

Carnegie Mellon

## References, cont'd


- [Milgram '67] Stanley Milgram: *The Small World Problem*, Psychology Today 1(1), 60-67 (1967)
- [Montgomery+, '01] Alan L. Montgomery, Christos Faloutsos: *Identifying Web Browsing Trends and Patterns*. IEEE Computer 34(7): 94-95 (2001)


UCLA-IPAM 05 (c) 2005 C. Faloutsos 107

Carnegie Mellon

## References, cont'd


- [Palmer+, '01] Chris Palmer, Georgos Siganos, Michalis Faloutsos, Christos Faloutsos and Phil Gibbons *The connectivity and fault-tolerance of the Internet topology* (NRDM 2001), Santa Barbara, CA, May 25, 2001
- [Pennock+, '02] David M. Pennock, Gary William Flake, Steve Lawrence, Eric J. Glover, C. Lee Giles: *Winners don't take all: Characterizing the competition for links on the web* Proc. Natl. Acad. Sci. USA 99(8): 5207-5211 (2002)


UCLA-IPAM 05 (c) 2005 C. Faloutsos 108

Carnegie Mellon

## References, cont'd


- [Schroeder, '91] Manfred Schroeder *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise* W H Freeman & Co., 1991 **(excellent book on fractals)**




Carnegie Mellon

## References, cont'd


- [Siganos+, '03] G. Siganos, M. Faloutsos, P. Faloutsos, C. Faloutsos *Power-Laws and the AS-level Internet Topology*, Transactions on Networking, August 2003.
- [Watts+ Strogatz, '98] D. J. Watts and S. H. Strogatz *Collective dynamics of 'small-world' networks*, Nature, 393:440-442 (1998)
- [Watts, '03] Duncan J. Watts *Six Degrees: The Science of a Connected Age* W.W. Norton & Company; (February 2003)




Carnegie Mellon

## Thank you!

www.cs.cmu.edu/~christos

www.db.cs.cmu.edu



Carnegie Mellon

## EXTRA
## Virus propagation



Carnegie Mellon

## Outline

Topology, 'laws' and generators

EXTRA: Virus Propagation



Carnegie Mellon

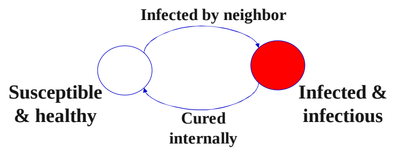
## Problem definition

- Q1: How does a virus spread across an arbitrary network?
- Q2: will it create an epidemic?

Carnegie Mellon

## Framework

- Susceptible-Infected-Susceptible (SIS) model
  - Cured nodes immediately become susceptible

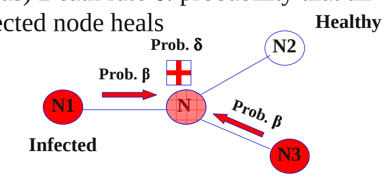



UCLA-IPAM 05 (c) 2005 C. Faloutsos 115

Carnegie Mellon

## The model

- (virus) Birth rate $\beta$: probability than an infected neighbor attacks
- (virus) Death rate $\delta$: probability that an infected node heals

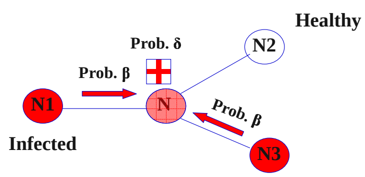



UCLA-IPAM 05 (c) 2005 C. Faloutsos 116

Carnegie Mellon

## The model

- Virus 'strength' s= $\beta/\delta$




UCLA-IPAM 05 (c) 2005 C. Faloutsos 117

Carnegie Mellon

## Epidemic threshold $\tau$

of a graph, defined as the value of $\tau$, such that

if strength $s = \beta / \delta < \tau$

an epidemic can not happen

Thus,

- given a graph
- compute its epidemic threshold

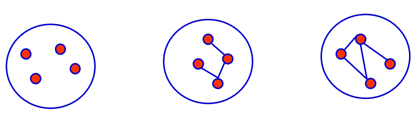
UCLA-IPAM 05 (c) 2005 C. Faloutsos 118

Carnegie Mellon

## Epidemic threshold $\tau$

What should $\tau$ depend on?

- avg. degree? and/or highest degree?
- and/or variance of degree?
- and/or third moment of degree?



UCLA-IPAM 05 (c) 2005 C. Faloutsos 119

Carnegie Mellon

## Epidemic threshold

- [Theorem] We have no epidemic, if

$$\boxed{\beta/\delta < \tau = 1/\lambda_{1,A}}$$

UCLA-IPAM 05 (c) 2005 C. Faloutsos 120

Carnegie Mellon

## Epidemic threshold

- [Theorem] We have no epidemic, if

recovery prob. epidemic threshold

$$\beta/\delta < \tau = 1/\lambda_{1,A}$$

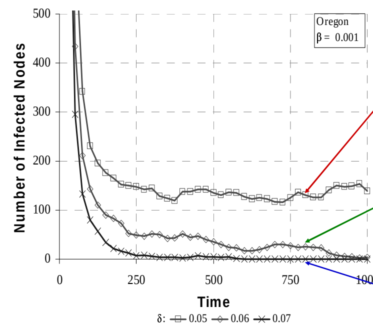
attack prob. largest eigenvalue of adj. matrix ***A***

Proof: [Wang+03]

UCLA-IPAM 05 (c) 2005 C. Faloutsos 121

Carnegie Mellon

## Experiments (Oregon)




**β/δ > τ (above threshold)**

**β/δ = τ (at the threshold)**

**β/δ < τ (below threshold)**

UCLA-IPAM 05 (c) 2005 C. Faloutsos 122

Carnegie Mellon

## Our result:

- Holds for **any** graph
- includes older results as special cases

UCLA-IPAM 05 (c) 2005 C. Faloutsos 123

Carnegie Mellon

## Reference

- [Wang+03] Yang Wang, Deepayan Chakrabarti, Chenxi Wang and Christos Faloutsos: *Epidemic Spreading in Real Networks: an Eigenvalue Viewpoint*, SRDS 2003, Florence, Italy.

UCLA-IPAM 05 (c) 2005 C. Faloutsos 124

Carnegie Mellon

## Thank you!

www.cs.cmu.edu/~christos

www.db.cs.cmu.edu

(really done this time ☺ )

UCLA-IPAM 05 (c) 2005 C. Faloutsos 125