CMS/CS/EE 144

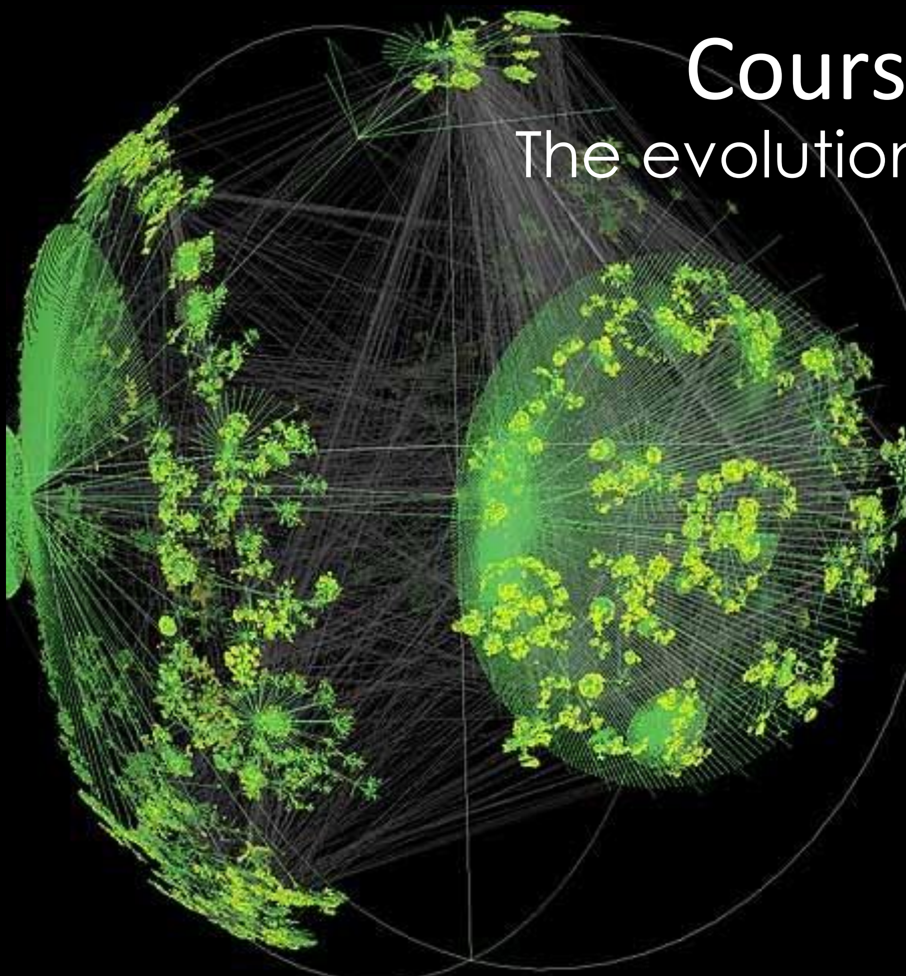
Networks: Structure & Economics

Administrivia

1) Register for Piazza if you haven't already
2) Buy your textbook!
3) HW1 is due Thursday
4) Office hours:
   -- me: After classes on Thursday
   -- TAs: Tue/Wed 7-9pm (107 ANB)
5) Don't forget about your blog posts!
6) Background poll – fill out if you haven't already

Course warmup…

The evolution and structure of the web

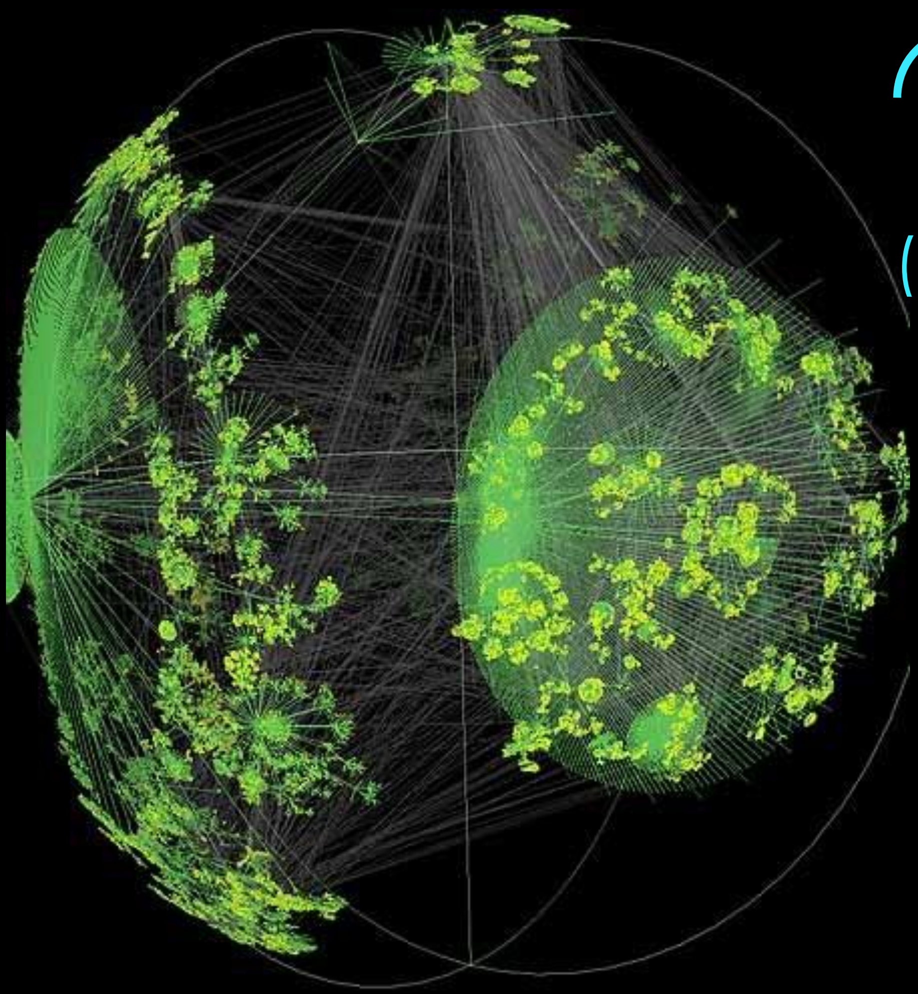


Our goal: What "structure" does the web graph have?

Nodes = pages

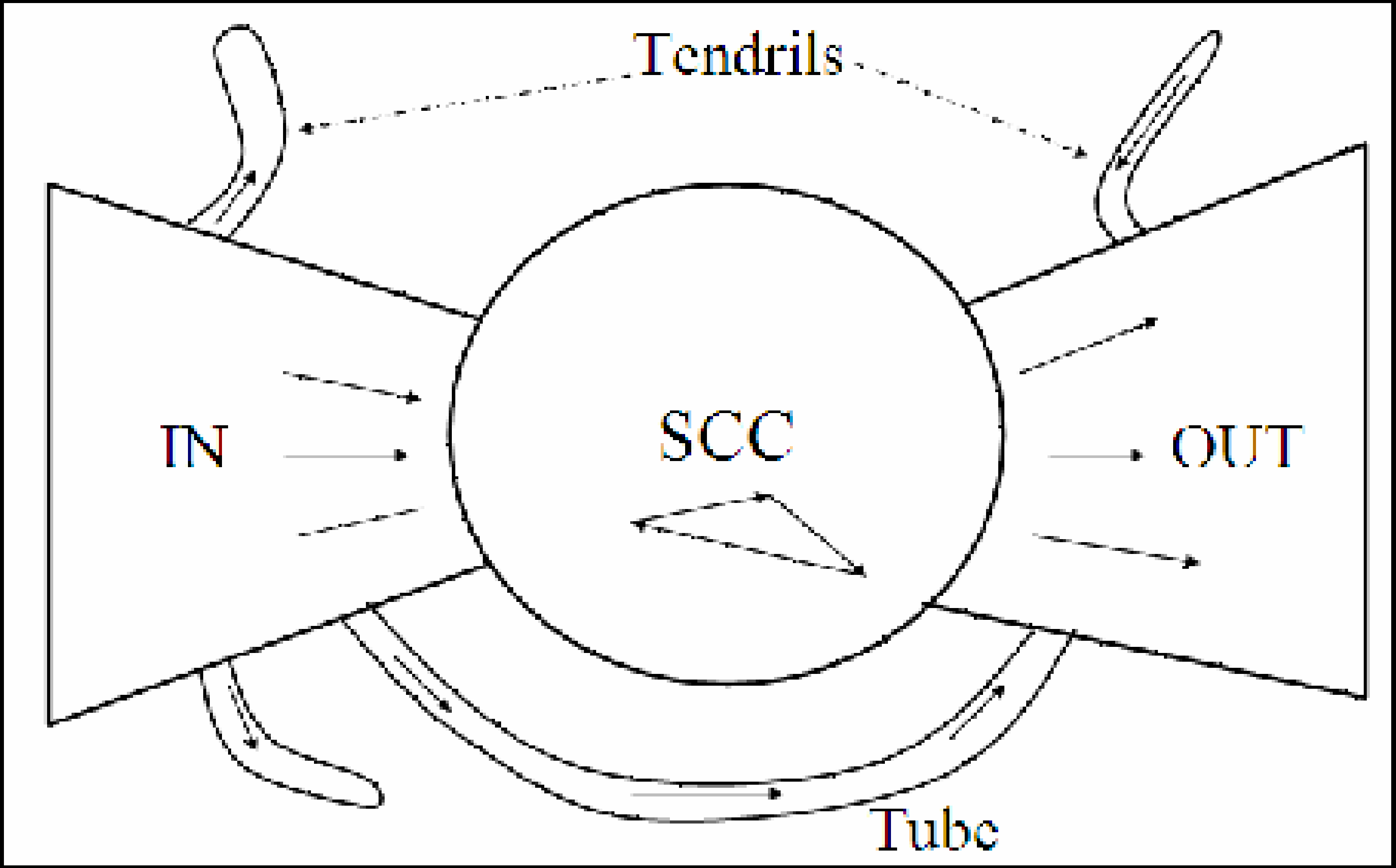
Edges = links

(edges are directed)



Our goal: What "structure" does the web graph have?

1) Connectivity?
2) Diameter?
3) Degree?
4) Clustering?

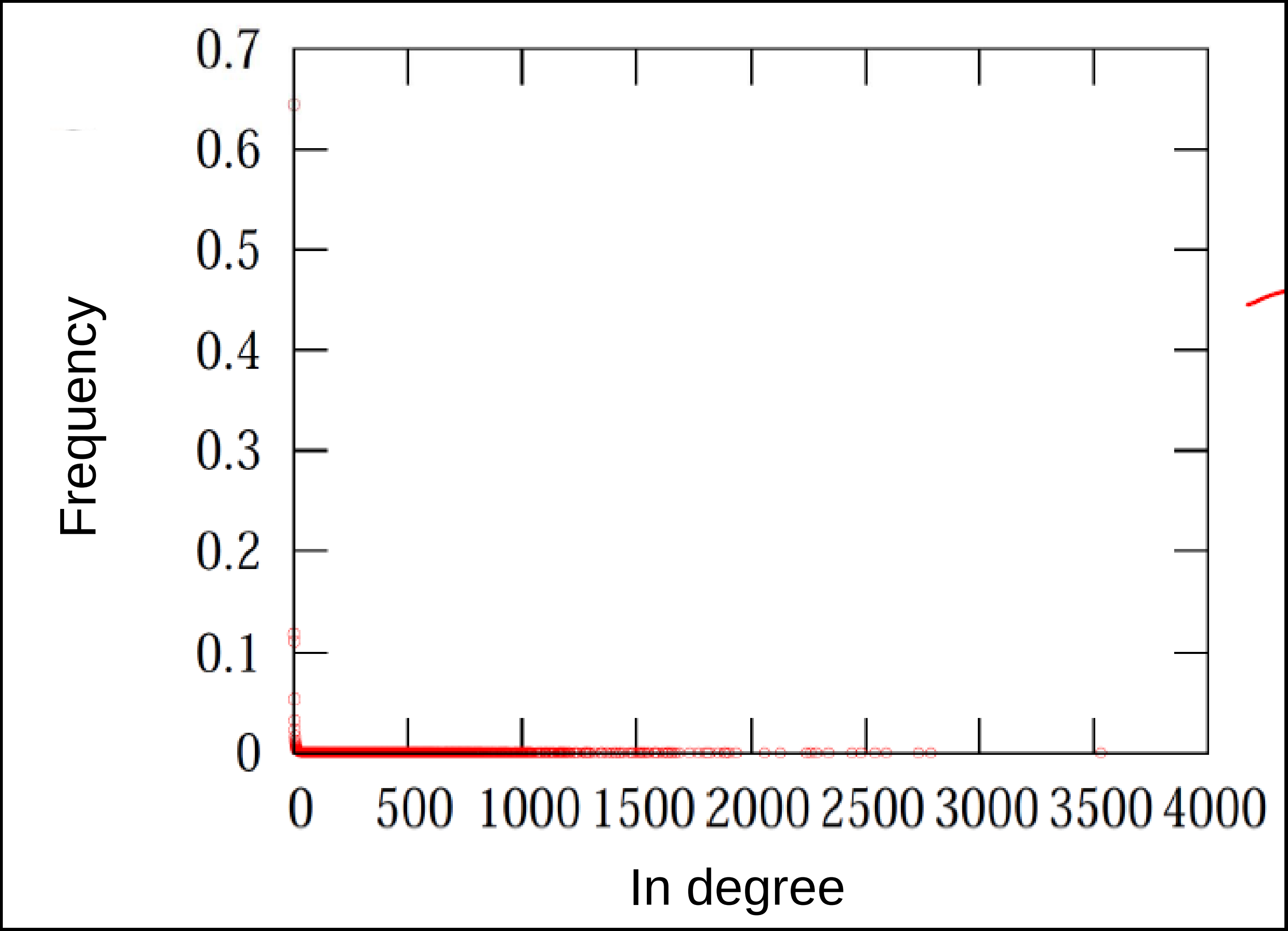

Tendrils
IN
SCC
OUT
Tube

Our goal: What "structure" does the web graph have?

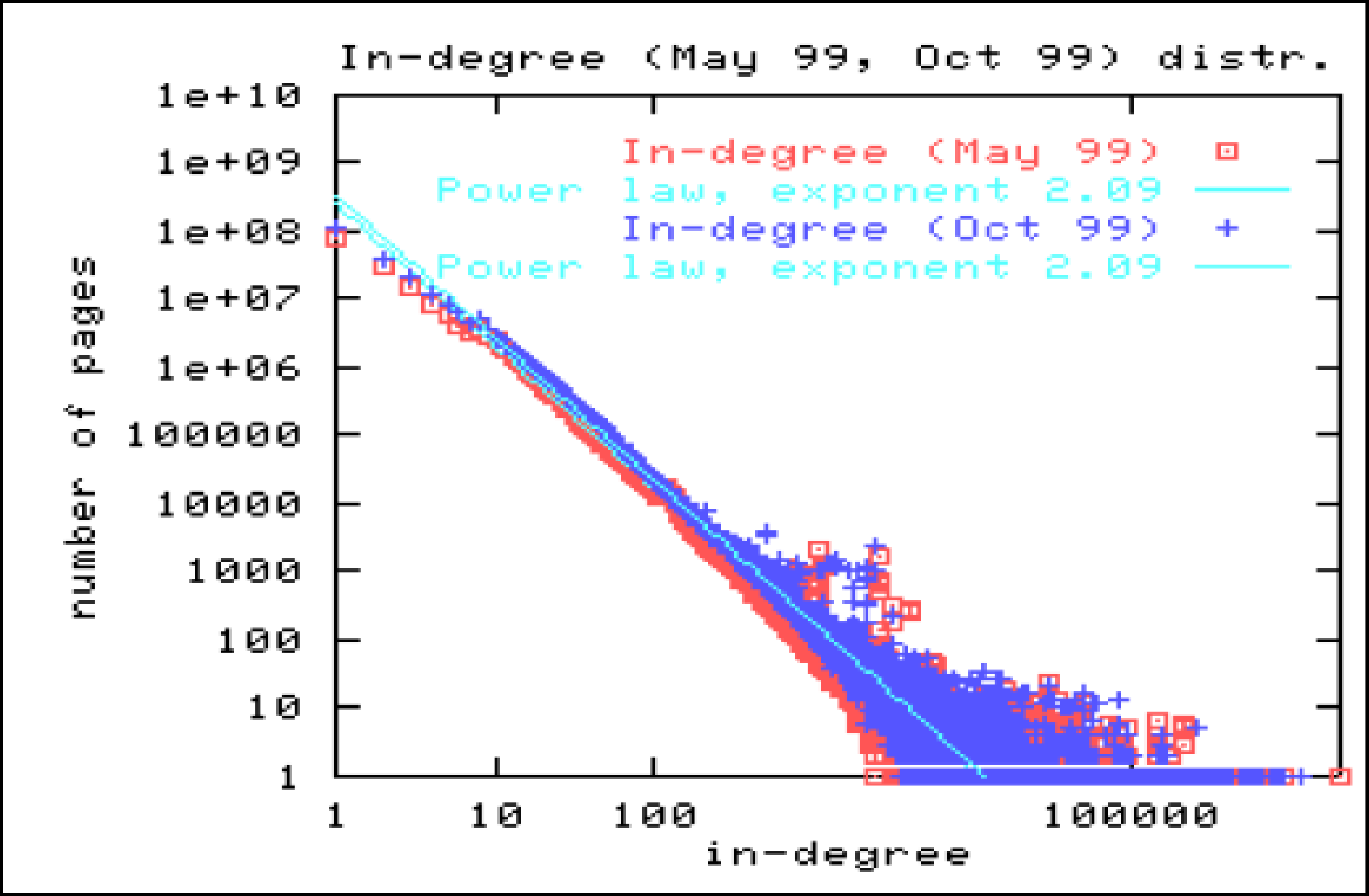
1) Connectivity? There is a giant connected component
2) Diameter?
3) Degree?
4) Clustering?

Our goal: What "structure" does the web graph have?

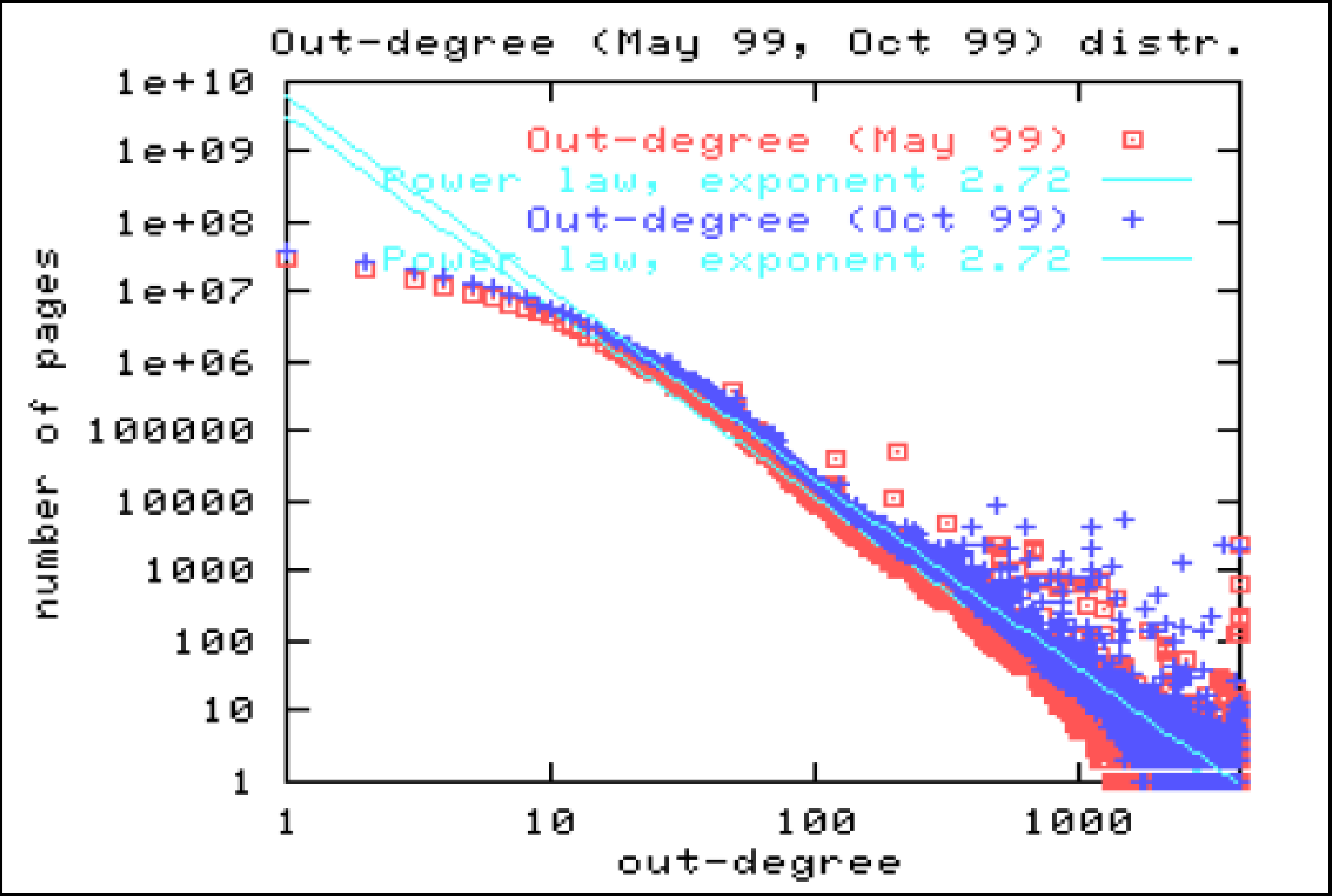
1) Connectivity? There is a giant connected component
2) Diameter? Small diameter
3) Degree?
4) Clustering?



0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Frequency
0
500
1000
1500
2000
2500
3000
3500
4000
In degree



In-degree (May 99, Oct 99) distr.
In-degree (May 99)
Power law, exponent 2.09
In-degree (Oct 99)
Power law, exponent 2.09
number of pages
1e+10
1e+09
1e+08
1e+07
1e+06
100000
10000
1000
100
10
1
1
10
100
100000
in-degree


Out-degree (May 99, Oct 99) distr.
Out-degree (May 99)
Power law, exponent 2.72
Out-degree (Oct 99)
Power law, exponent 2.72
number of pages
out-degree
1e+10
1e+09
1e+08
1e+07
1e+06
100000
10000
1000
100
10
1
1
10
100
1000

Our goal: What "structure" does the web graph have?

1) Connectivity? There is a giant connected component
2) Diameter? Small diameter
3) Degree? Heavy-tailed degree distribution
4) Clustering?

Our goal: What "structure" does the web graph have?

1) Connectivity? There is a giant connected component
2) Diameter? Small diameter
3) Degree? Heavy-tailed degree distribution
4) Clustering? Highly clustered

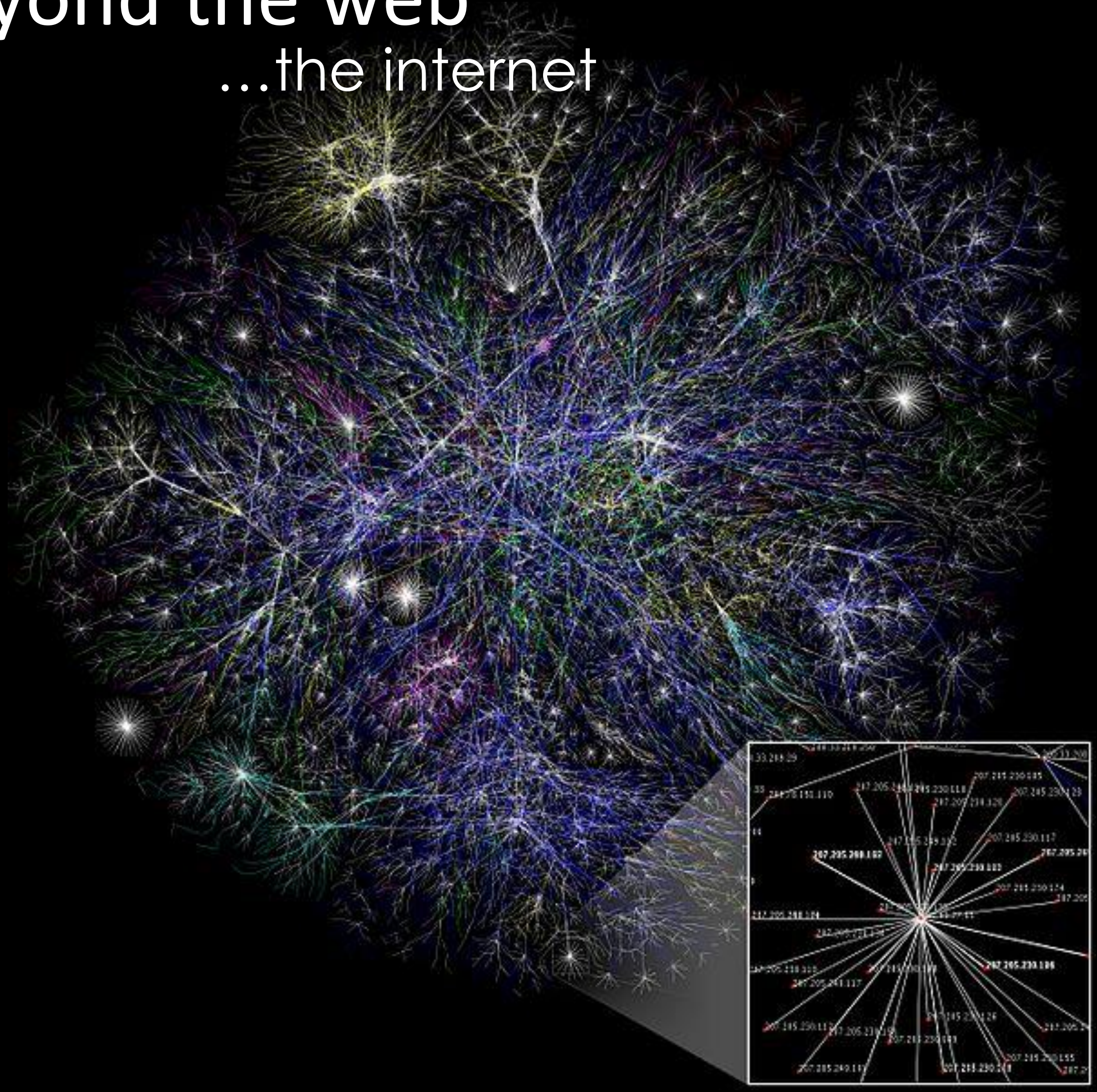
It turns out that these are "universal"

Beyond the web

…there are lots of other networks people study

Beyond the web

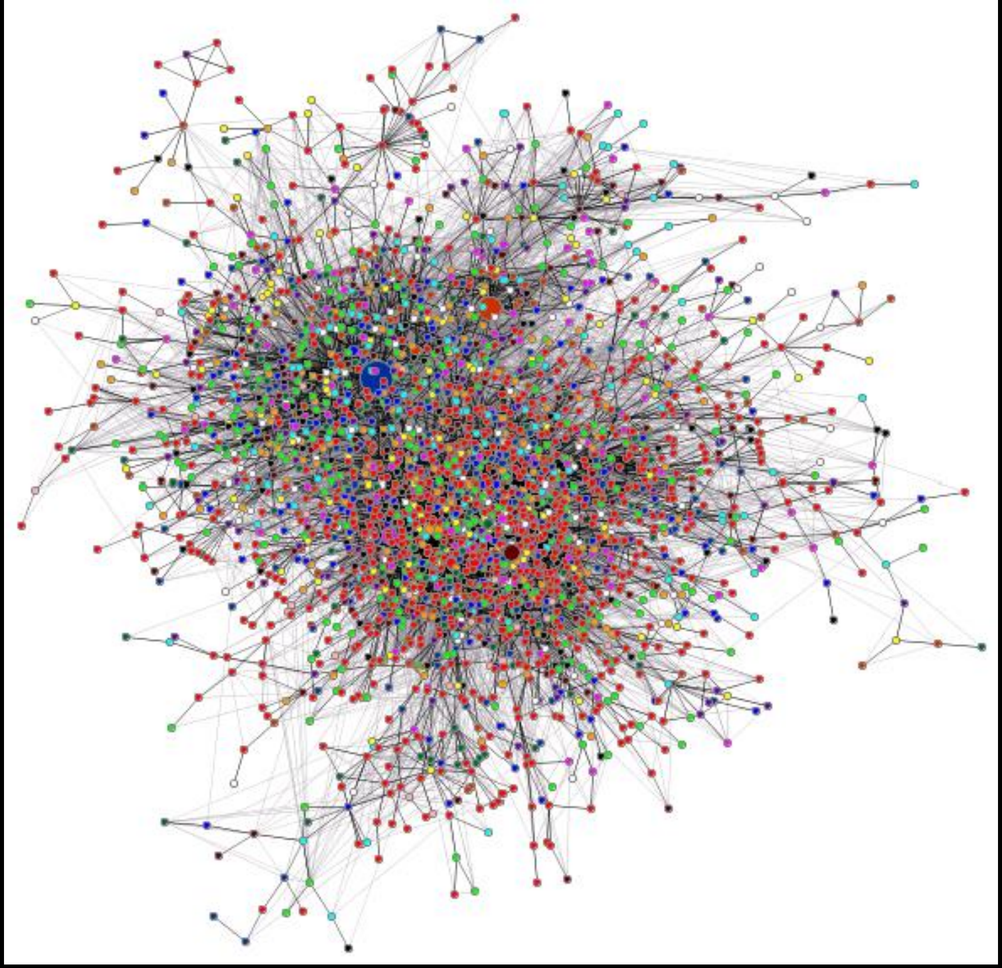
...the internet



Beyond the web

…online social networks
(Facebook, Myspace, blogs, flickr, IMs, etc)

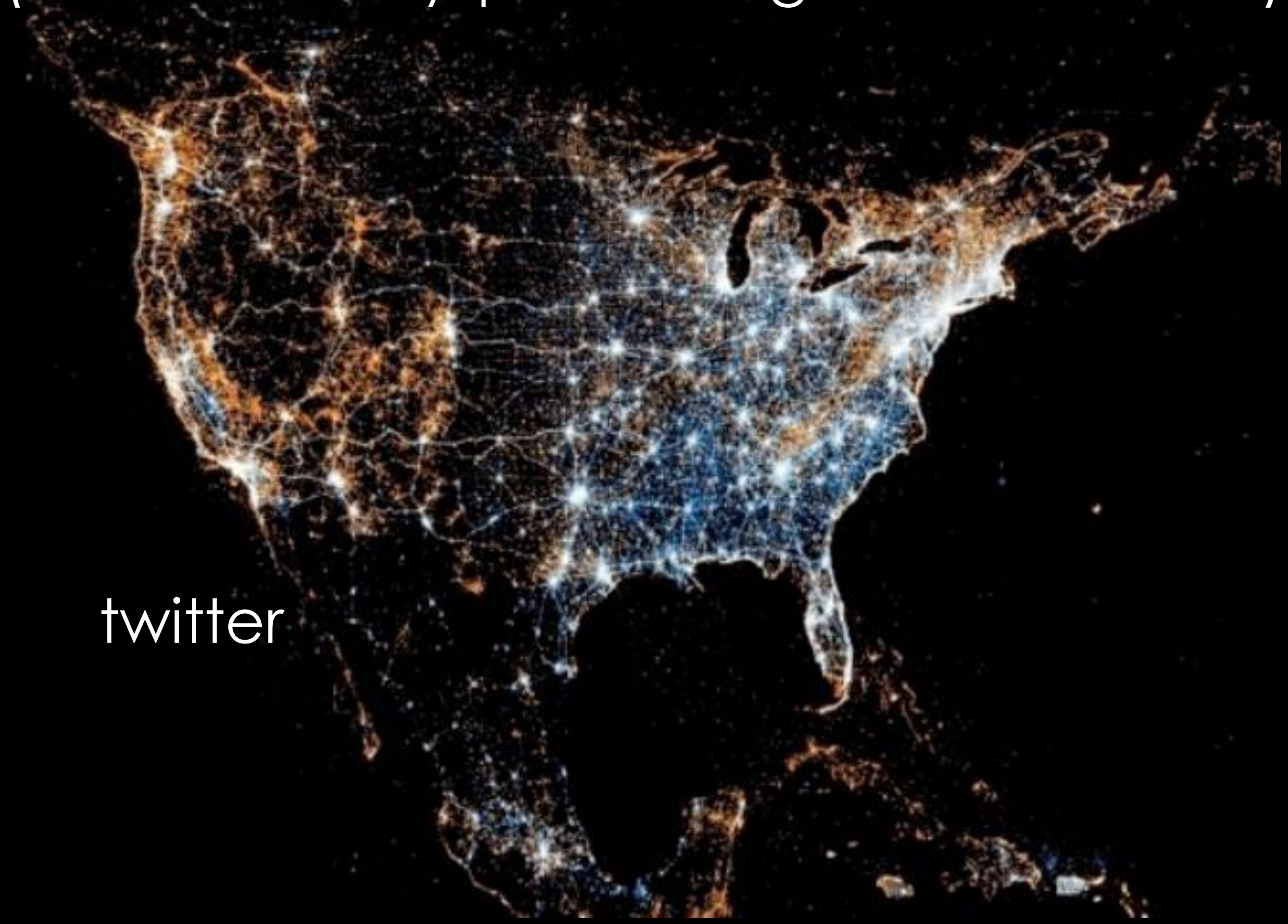
Beyond the web
…online social networks
(Facebook, Myspace, blogs, flickr, IMs, etc)



Blogosphere links

Beyond the web

...online social networks
(Facebook, Myspace, blogs, flickr, IMs, etc)



Beyond the web

…traditional social networks

e.g. friends

hierarchy in business

family trees

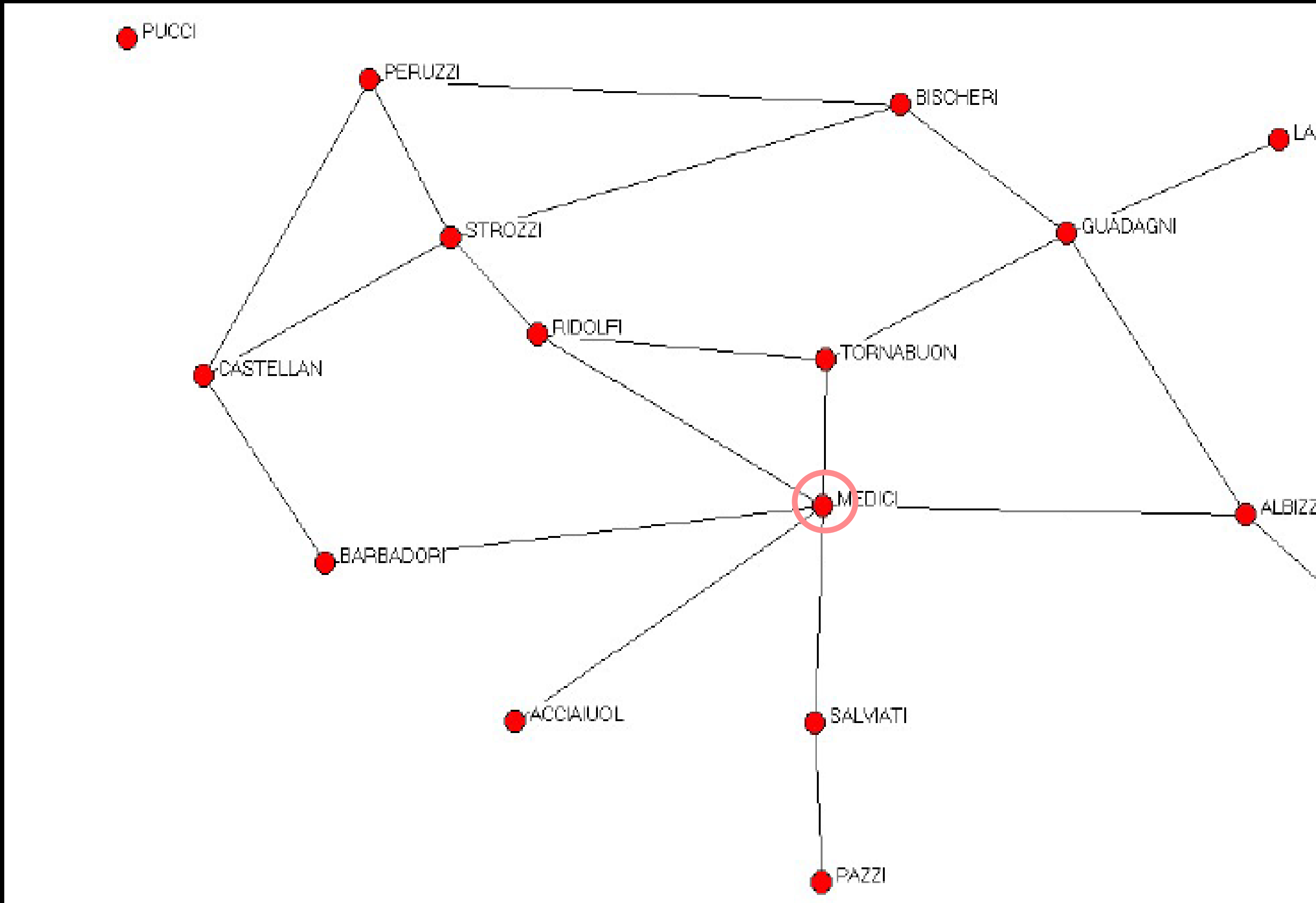
marriages

sexual relations

…many others

Beyond the web

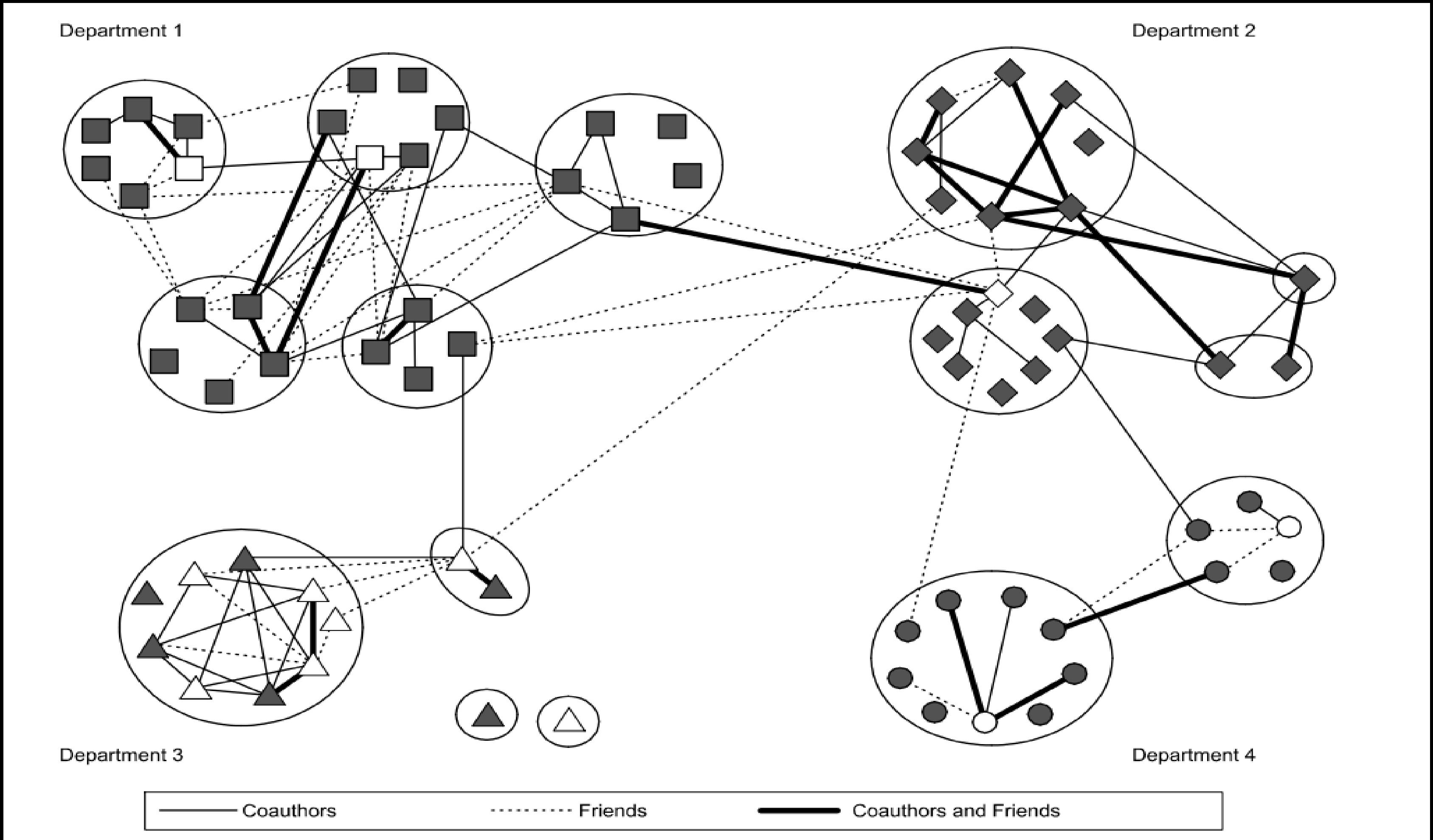
…traditional social networks



Florentine marriages in 1400

Beyond the web

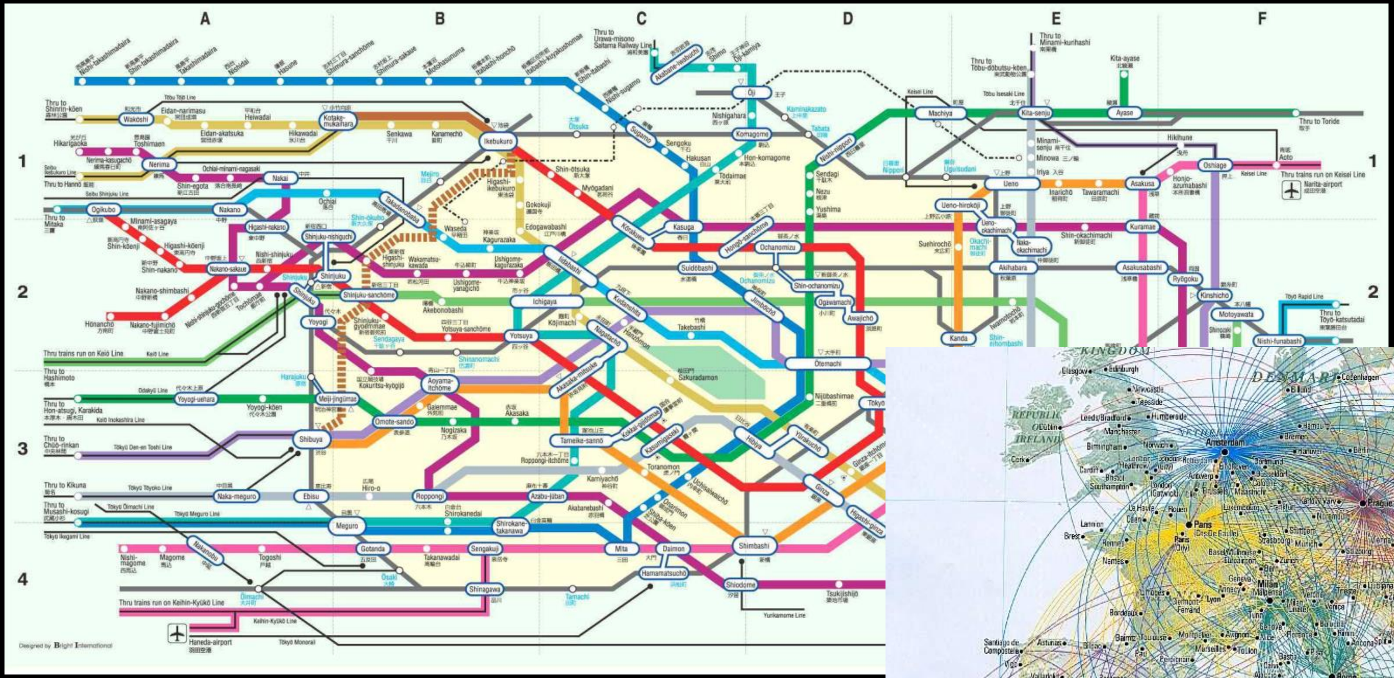
…traditional social networks



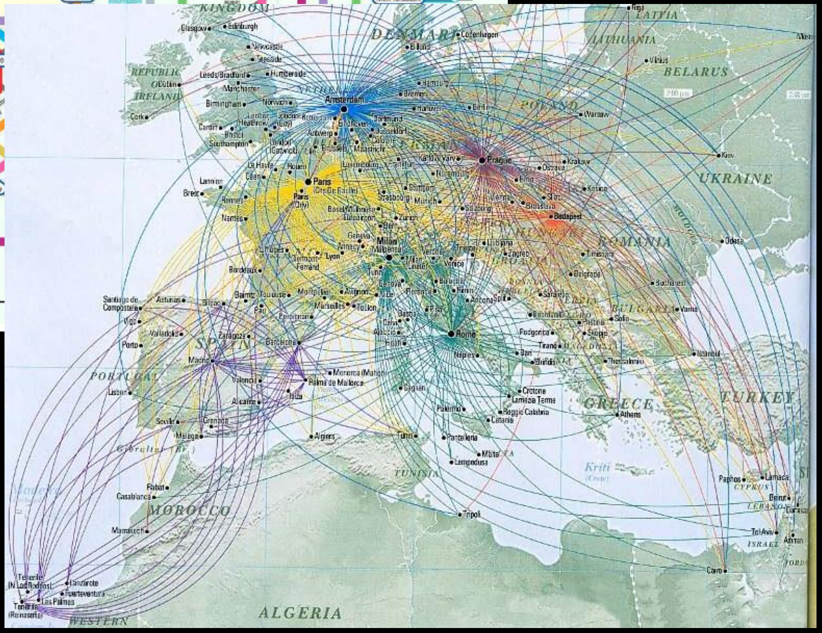
(from Leeat Yariv)

Beyond the web

…transportation networks

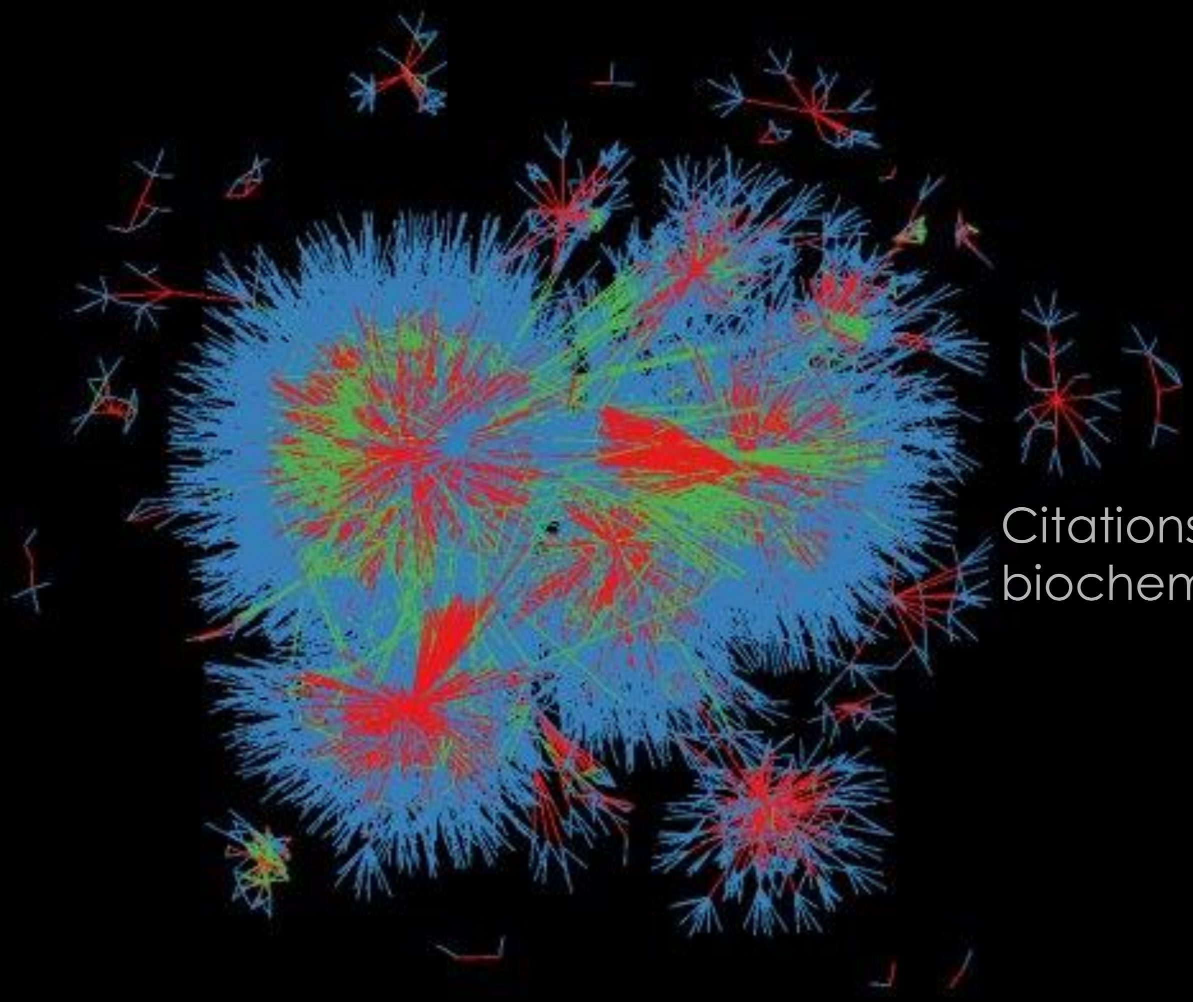


Tokyo metro



European airlines

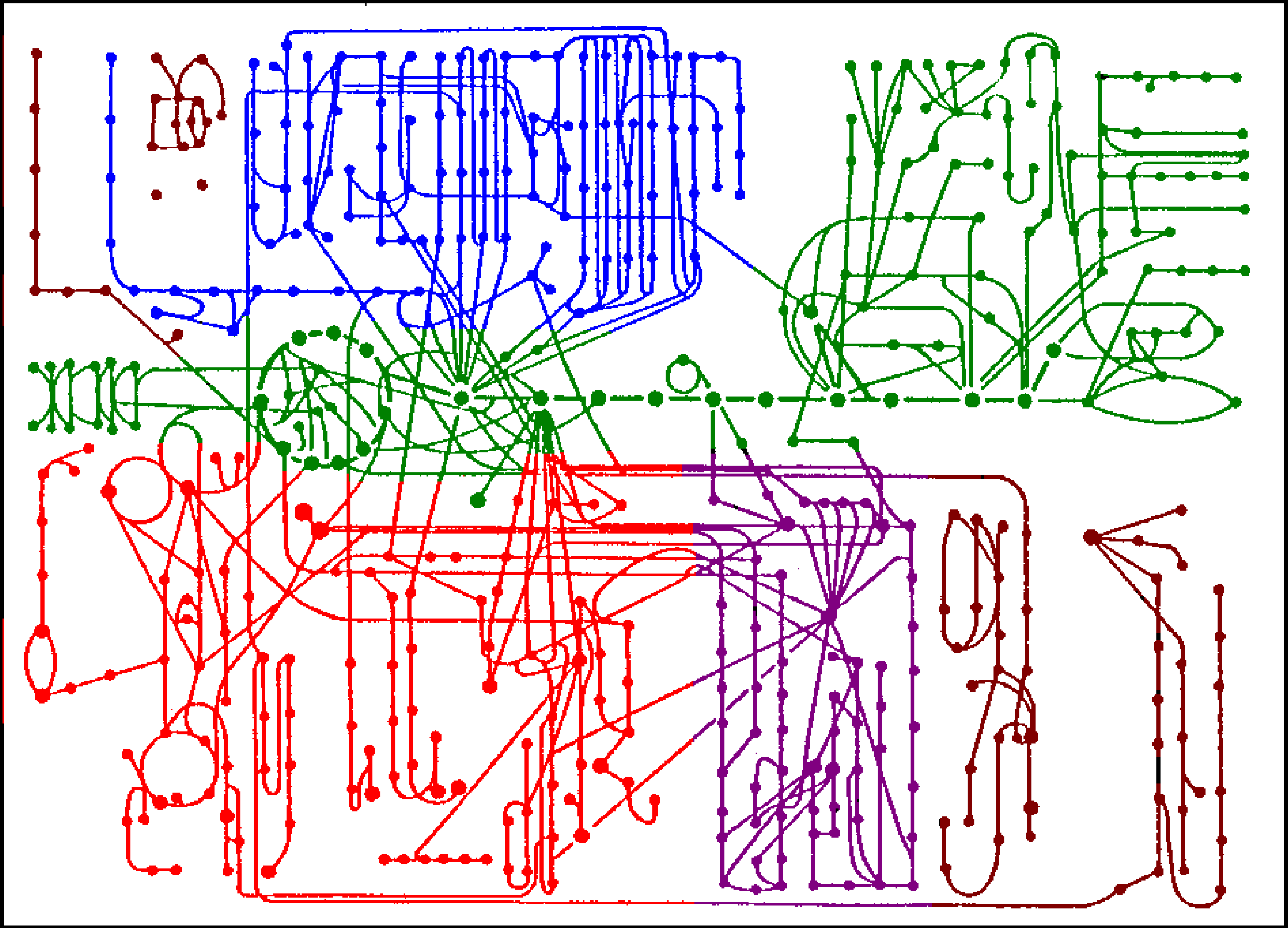
Beyond the web
…coauthor & citation networks



Citations among biochem patents

Beyond the web

…biological networks



Biochemical pathways of malaria

Beyond the web
…biological networks

The Economist

Log in | Register | My account

Home | World | Business & finance | Science & technology | Markets

Railways and slime moulds

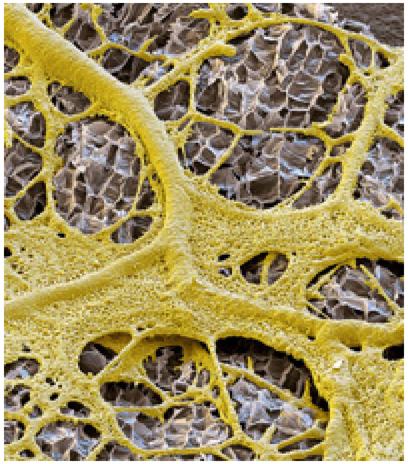
A life of slime

Network-engineering problems can be solved by surprisingly simple creatures

Jan 21st 2010

From *The Economist* print edition

FROM adhesives that mimic the feet of geckos to swimsuits modelled on shark skin, biologically inspired design has taken off in recent times. Copying nature's ideas allows people to harness the power of evolution to come up with clever products. Now a group of researchers has taken this idea a step further by using an entire living organism—a slime mould—to solve a complex problem. In this case, the challenge was to design an efficient rail network for the city of Tokyo and its outlying towns.



Show me the way to Shinjuku

SPL

Slime moulds are unusual critters—neither animal, nor plant nor fungus. If they resemble anything, it is a colonial amoeba. *Physarum polycephalum*, the species in question, consists of a membrane-bound bag of

All these (and many others) tend to have the same "universal" properties as the web

1) A "giant" connected component
2) Small diameter
3) Heavy-tailed degree distribution
4) High clustering coefficient

We’ll look at 6 examples

Social network → Slashdot network in 2009

Citation network → US Patents

Web Graph → Google 2002

Internet Graph → 2005

Product co-purchasing → Amazon 2003

Road networks → California 2008

(all data from http://snap.stanford.edu/data/index.html)

1) A “giant” connected component
2) Small diameter
3) Heavy-tailed degree distribution
4) High clustering coefficient

Social network → ??

Citation network → ??

Web Graph → ??

Internet Graph → ??

Product co-purchasing → ??

Road networks → ??

1) A “giant” connected component
2) Small diameter
3) Heavy-tailed degree distribution
4) High clustering coefficient

Social network → 86% nodes in SCC, 100% in WCC

Citation network → 0% in SCC, 99% in WCC

Web Graph → 50% in SCC, 98% in WCC

Internet Graph → 99% in SCC, 99% in WCC

Product co-purchasing → 98% in SCC, 100% in WCC

Road networks → 99% in WCC

1) A “giant” connected component
2) Small diameter
3) Heavy-tailed degree distribution
4) High clustering coefficient

Social network → ??
Citation network → ??
Web Graph → ??
Internet Graph → ??
Product co-purchasing → ??
Road networks → ??

1) A “giant” connected component
2) Small diameter
3) Heavy-tailed degree distribution
4) High clustering coefficient

Social network → 12, 90%tile effective diameter is 4.7

Citation network → 22, 90%tile effective diameter is 9.4

Web Graph → 22, 90%tile effective diameter is 8.1

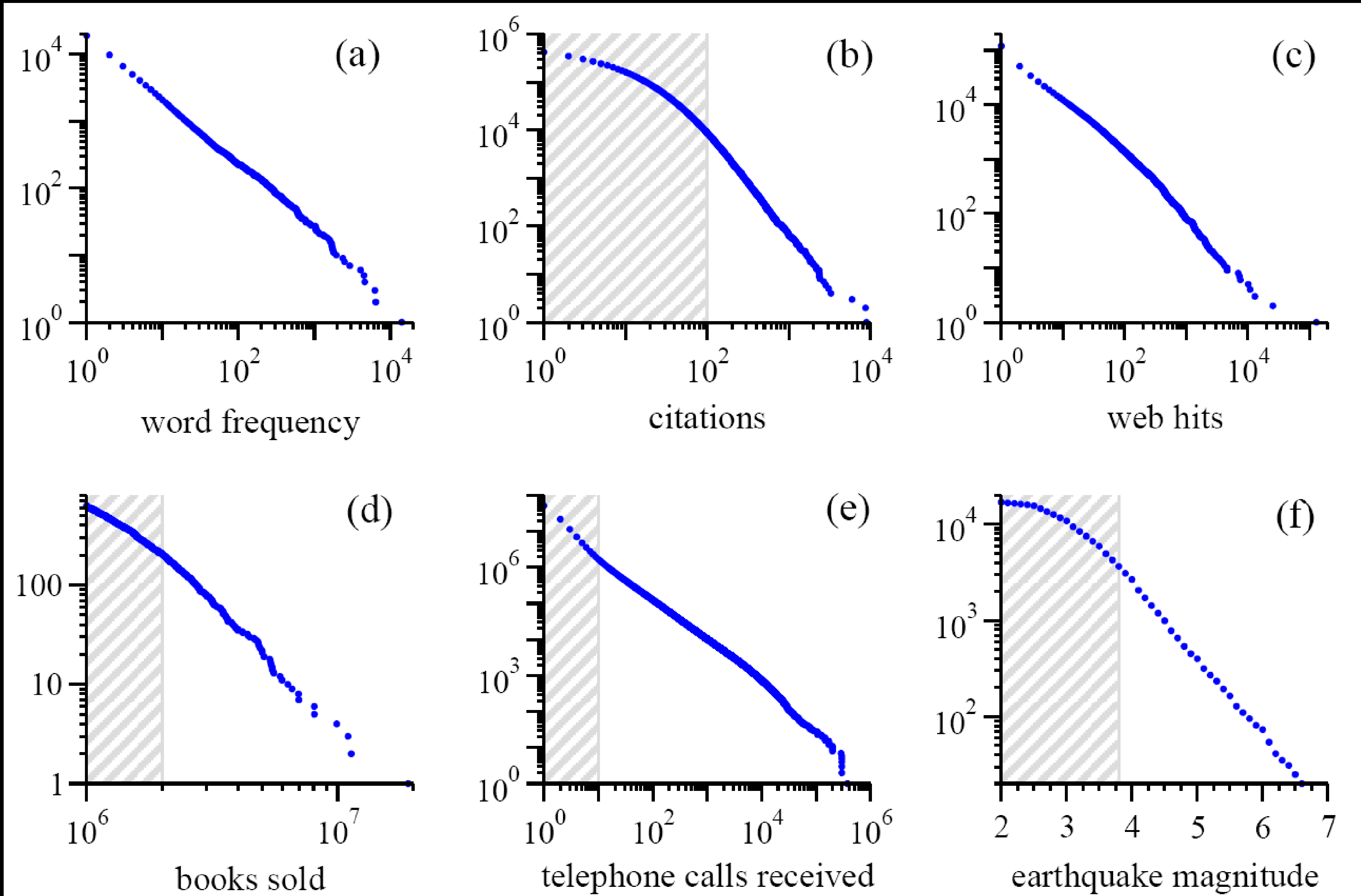
Internet Graph → 25, 90%tile effective diameter is 5.9

Product co-purchasing → 21, 90%tile effective diam is 7.6

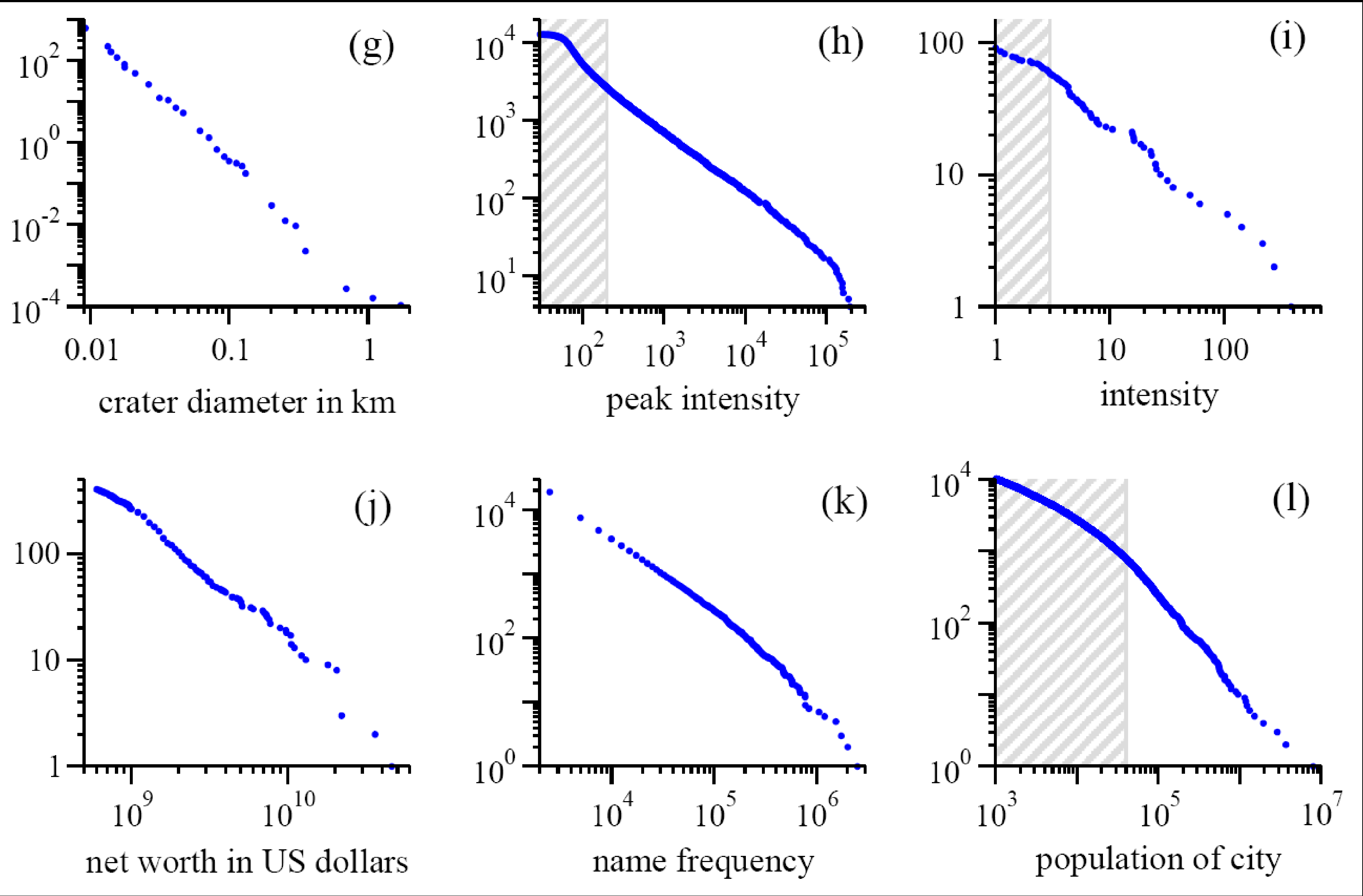
Road networks → 850, 90%tile effective diameter is 500

1) A "giant" connected component
2) Small diameter
3) Heavy-tailed degree distribution
4) High clustering coefficient

Here I'll start by showing you some other data sets…



(These are the ccdfs.)



(These are the ccdfs.)

The reason I showed you others is that the existence of heavy-tailed degree distributions is a bit controversial

…heavy-tails are "universal", but naïve statistics can lead you astray

On Power-Law Relationships of the Internet Topology

Michalis Faloutsos
U.C. Riverside
Dept. of Comp. Science
michalis@cs.ucr.edu

Petros Faloutsos
U. of Toronto
Dept. of Comp. Science
pfal@cs.toronto.edu

Christos Faloutsos *
Carnegie Mellon Univ.
Dept. of Comp. Science
christos@cs.cmu.edu

1999 Sigcomm paper – 4500+ citations!

BUT…

2005, STOC

On the Bias of Traceroute Sampling

or, Power-law Degree Distributions in Regular Graphs

Dimitris Achlioptas
Microsoft Research
Microsoft Corporation
Redmond, WA 98052
optas@microsoft.com

Aaron Clauset
Department of Computer Science
University of New Mexico
Albuquerque, NM 87131
aaron@cs.unm.edu

David Kempe
Department of Computer Science
University of Southern California
Los Angeles, CA 90089
dkempe@usc.edu

Cristopher Moore
Department of Computer Science
University of New Mexico
Albuquerque, NM 87131
moore@cs.unm.edu

IEEE/ACM TRANSACTIONS ON NETWORK… 1205

Understanding Internet Topology: Principles, Models, and Validation

David Alderson, *Member, IEEE*, Lun Li, *Student Member, IEEE*, Walter Willinger, *Fellow, IEEE*, and John C. Doyle, *Member, IEEE*

2005, ToN

There are similar stories in power nets, social nets, …

1) A “giant” connected component
2) Small diameter
3) Heavy-tailed degree distribution
4) High clustering coefficient

(without correlations clustering coefs would all be <0.001)

Social network → ??

Citation network → ??

Web Graph → ??

Internet Graph → ??

Product co-purchasing → ??

Road networks → ??

1) A “giant” connected component
2) Small diameter
3) Heavy-tailed degree distribution
4) High clustering coefficient

(without correlations clustering coefs would all be <0.001)

Social network → Avg clustering coef 0.06
Citation network → Avg clustering coef 0.09
Web Graph → Avg clustering coef 0.60
Internet Graph → Avg clustering coef 0.30
Product co-purchasing → Avg clustering coef 0.42
Road networks → Avg clustering coef 0.05

1) A “giant” connected component
2) Small diameter
3) Heavy-tailed degree distribution
4) High clustering coefficient

Hopefully you’re convinced that these are “universal”!

… the interesting thing to find these days are places where one or more doesn’t hold!

DISCLAIMER: We're moving on, but there are many other interesting properties of networks that we could discuss!

-- weak ties vs. strong ties

-- expansion

-- densifictation

-- shrinking diameter

-- power law eigenvalues

-- …

→ Identifying "universal" properties is still an active research area.

→ Any of these would make good projects

Two possible next steps for us:

Scientific question

What causes the emergence of these properties?

Engineering question

How can we exploit these properties for system design?

→ Since this is Caltech we'll be scientists first, and then engineers…

BUT

…to keep you motivated:

You should start to think of ways to exploit the properties we've discussed…