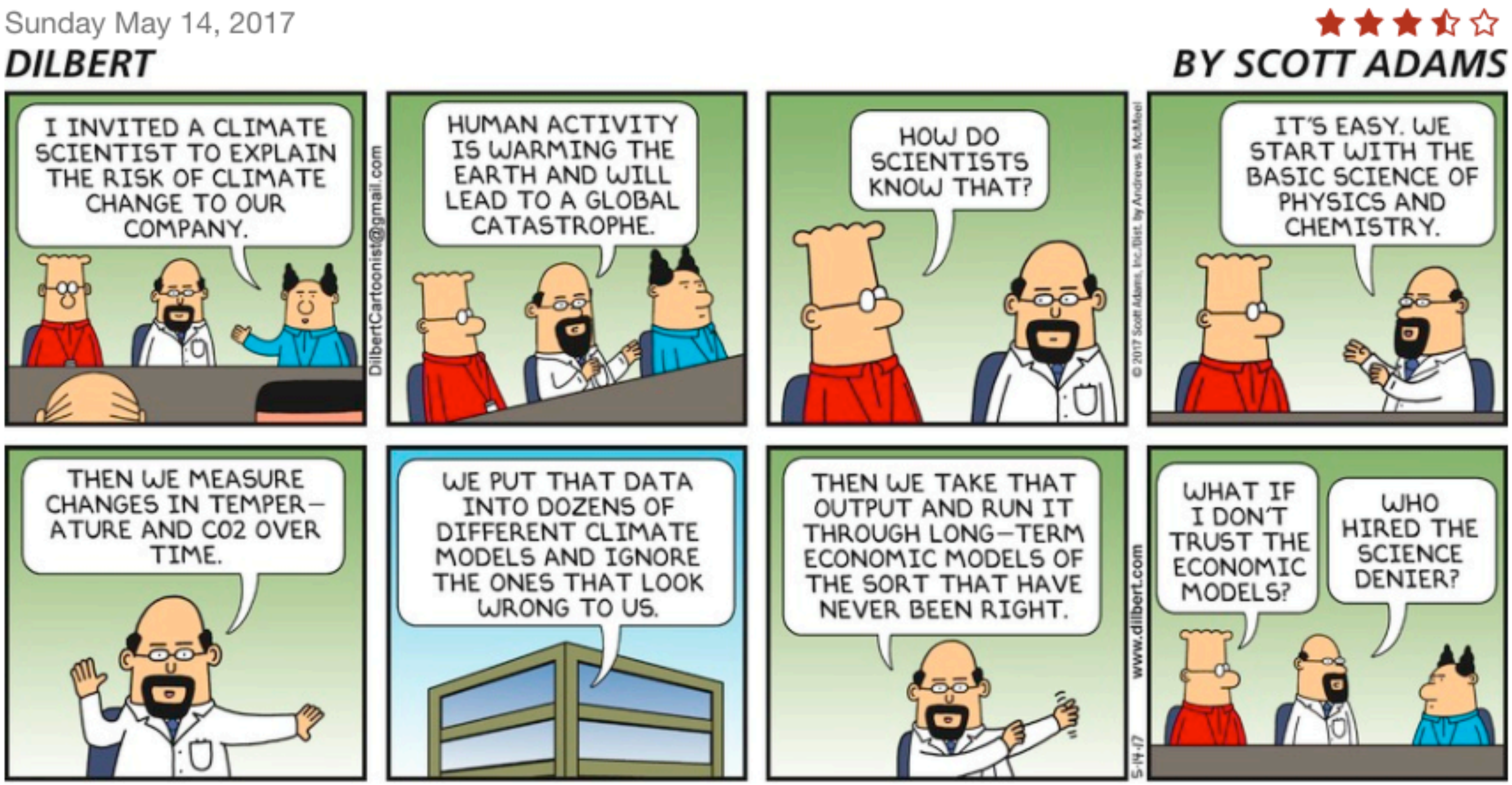
ECE 158A: Lecture 10

Fall 2017

Good models



Sunday May 14, 2017
DILBERT
BY SCOTT ADAMS
I INVITED A CLIMATE SCIENTIST TO EXPLAIN THE RISK OF CLIMATE CHANGE TO OUR COMPANY.
DilbertCartoonist@gmail.com
HUMAN ACTIVITY IS WARMING THE EARTH AND WILL LEAD TO A GLOBAL CATASTROPHE.
HOW DO SCIENTISTS KNOW THAT?
© 2017 Scott Adams, Inc./Dist. by Andrews McMeel
IT'S EASY. WE START WITH THE BASIC SCIENCE OF PHYSICS AND CHEMISTRY.
THEN WE MEASURE CHANGES IN TEMPER-ATURE AND CO2 OVER TIME.
WE PUT THAT DATA INTO DOZENS OF DIFFERENT CLIMATE MODELS AND IGNORE THE ONES THAT LOOK WRONG TO US.
THEN WE TAKE THAT OUTPUT AND RUN IT THROUGH LONG-TERM ECONOMIC MODELS OF THE SORT THAT HAVE NEVER BEEN RIGHT.
www.dilbert.com
5-14-17
WHAT IF I DON'T TRUST THE ECONOMIC MODELS?
WHO HIRED THE SCIENCE DENIER?

Power law degree distribution

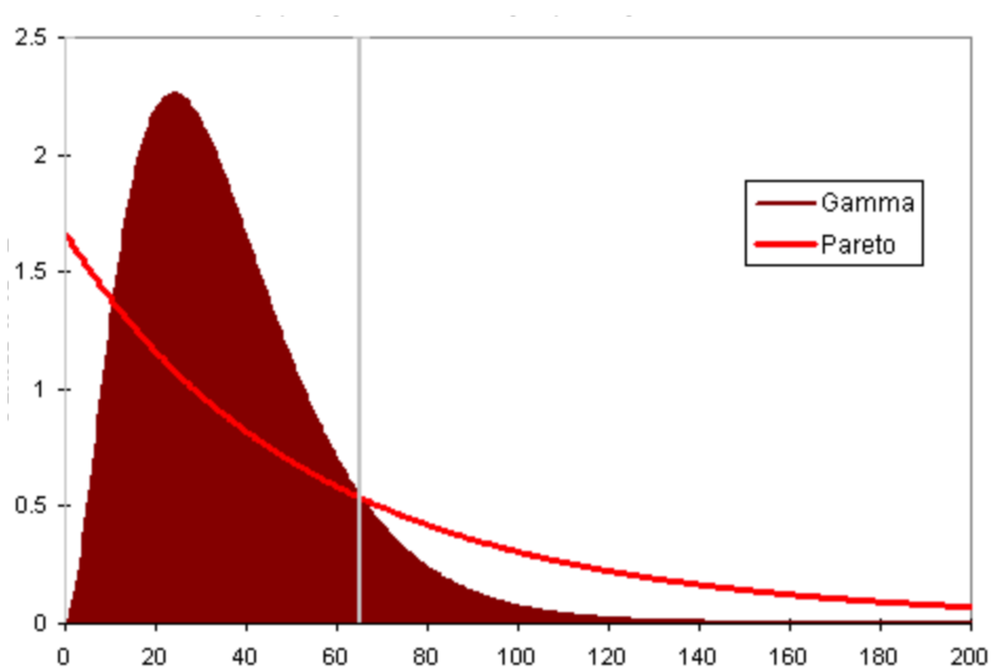
In 1999 it was observed that the Internet graph both at the router level (IP) and at the AS level (BGP) has a power law degree distribution

$$P(k_i = k) = Ak^{-\alpha}, \quad 2 < \alpha < 3$$

Power law degree distribution

In 1999 it was observed that the Internet graph both at the router level (IP) and at the AS level (BGP) has a power law degree distribution

$$P(k_i = k) = Ak^{-\alpha}, \quad 2 < \alpha < 3$$



Power law degree distribution

This indicates large variability in node degrees, as the average node degree is essentially uninformative

This is incompatible with random graph models that lead to a distribution with an exponentially decaying tail

A Preferential attachment model (PA) was then proposed as a "universal" model for complex networks exhibiting power laws

Preferential attachment

connected graph of m nodes

$PA \implies$ new node connects to i with probability

$$p_i = \frac{k_i}{\sum_j k_j}$$

$$\boxed{\lim_{n \to \infty} P(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \sim k^{-3}}$$

The model can also be made more general to obtain a power law degree distribution of any power in (2,infinity)

Preferential attachment

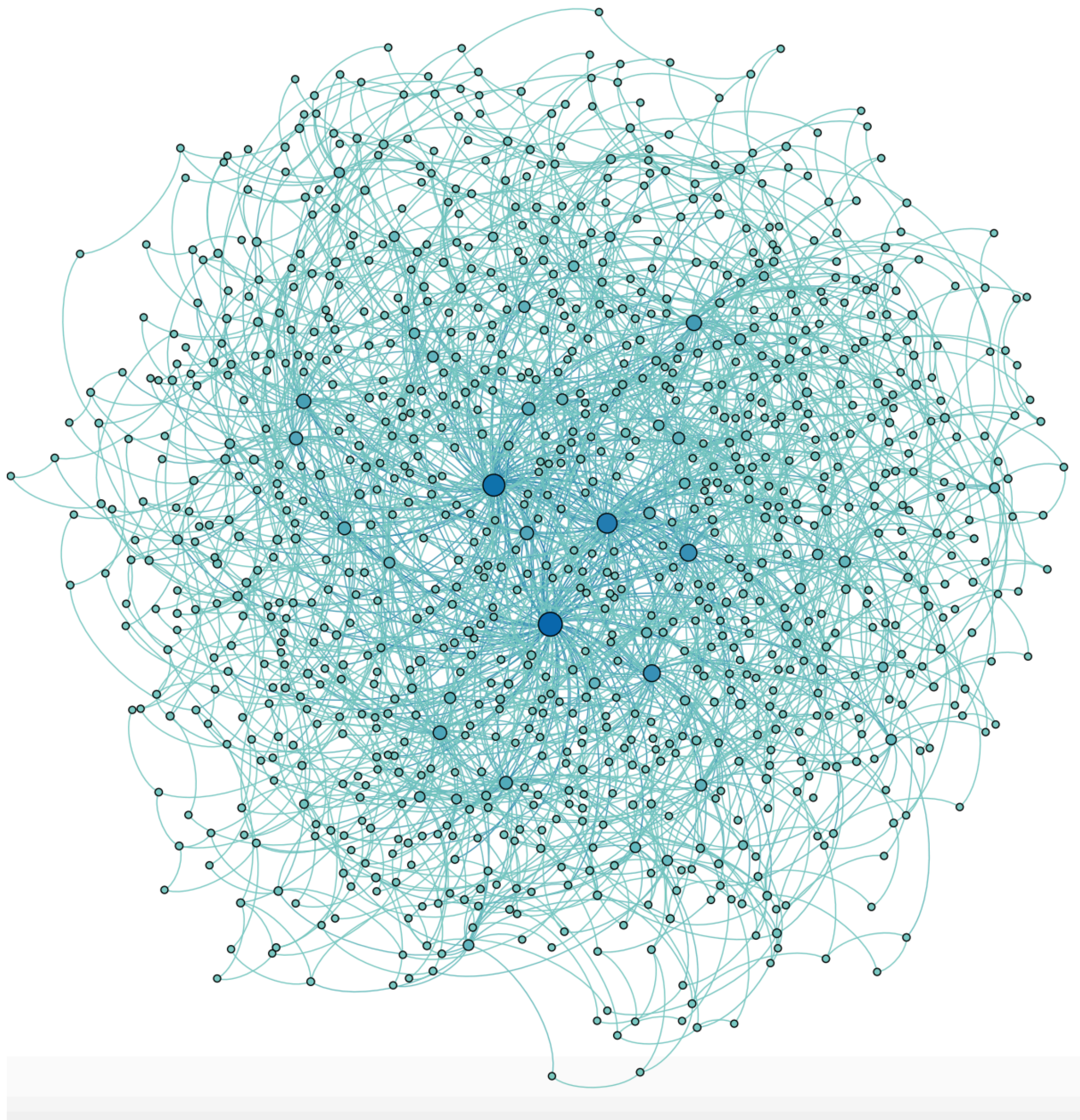
Properties of preferential attachment:

Power law degree distribution

Small diameter $\sim \dfrac{\log n}{\log \log n}$

Emergence of "hubs": these highly connected nodes appear to be at the core of the network

Preferential attachment “hubs”



Preferential attachment “hubs”

Preferential attachment “hubs”

Removing 95% of the links makes “little damage”

Removing 2% of the hubs breaks the network in a multitude of small components

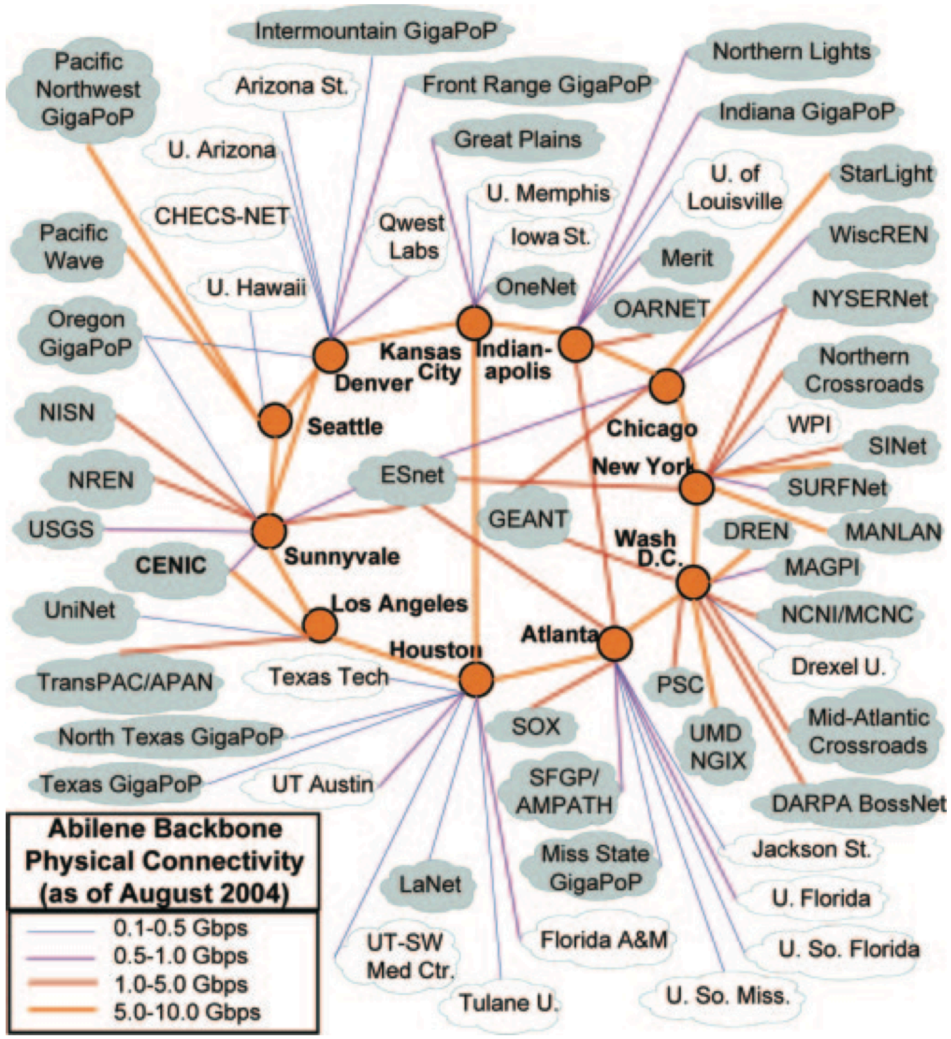
Preferential attachment

Does preferential attachment really occurs?

It has been proposed to explain power laws in WWW, Internet, collaboration networks, sexual partner networks, protein networks…

In reality they have very little in common. If you carefully look at the data, there are fundamental differences that cannot be explained using a single model.

A look at the real Internet



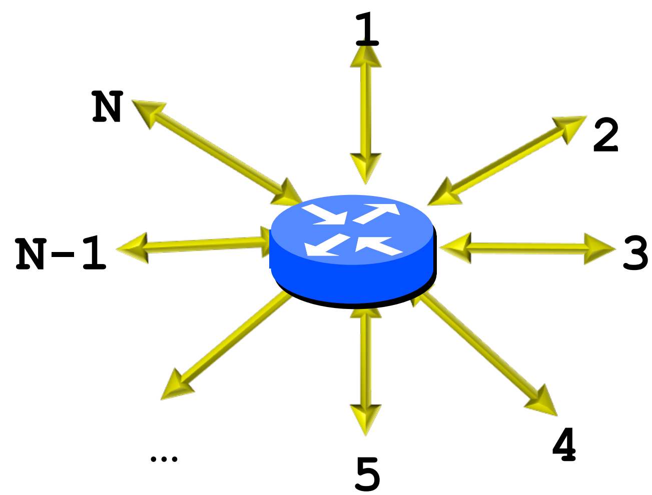
- Example of ISP router level map from ISP

A look at the real Internet

- Complete absence of “hubs”;
- High-degree vertices can exist but are found only within the local networks at the far periphery of the network and would not appear anywhere close to the backbone.
- This shows absence of “Achille’s heel”

Routers capacity

- N = number of external router "ports"
- R = speed ("line rate") of a port
- Router capacity = N x R



Routers capacity

Juniper T4000

- R= 10/40 Gbps
- NR = 4 Tbps

Cisco CRS

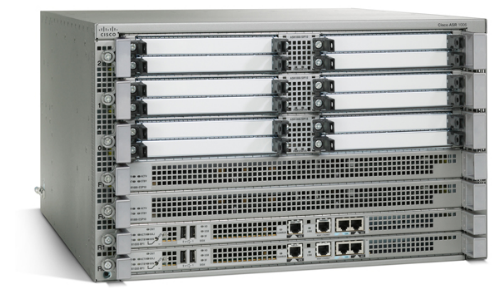
- R=10/40/100 Gbps
- NR = 322 Tbps

72 racks, 1MW

Routers capacity

Cisco ASR 1006

- R=1/10 Gbps
- NR = 40 Gbps



Juniper M120

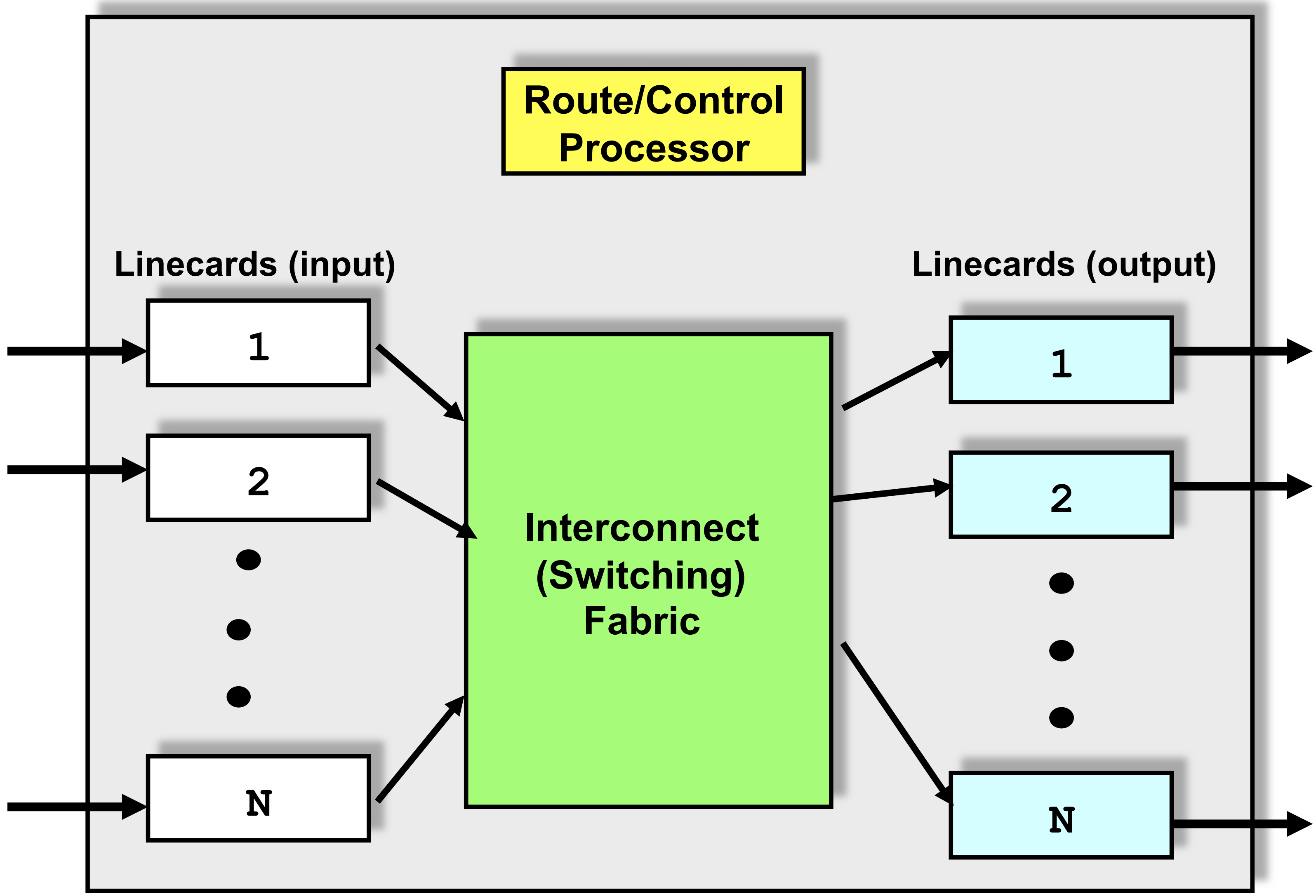
- R= 2.5/10 Gbps
- NR = 120 Gbps

Routers capacity

Cisco 3945E

- R = 10/100/1000 Mbps
- NR < 10 Gbps

Routers capacity



Alternative models

Alternative models must capture more details of the real system

In the case of the internet is clear that the backbone cannot be composed of highly connected hubs

We cannot build fast switches serving a large number of ports

Constrained optimization models

ISP exploit traffic aggregation

Many links with small bandwidth

A few links with large bandwidth

The real architecture arises as the solution to a constrained optimization problem

We will examine generative graph models based on constrained optimization next

Constrained optimization models

Start with a low-degree low-diameter “backbone”

Attach “tree-like” regional points of access

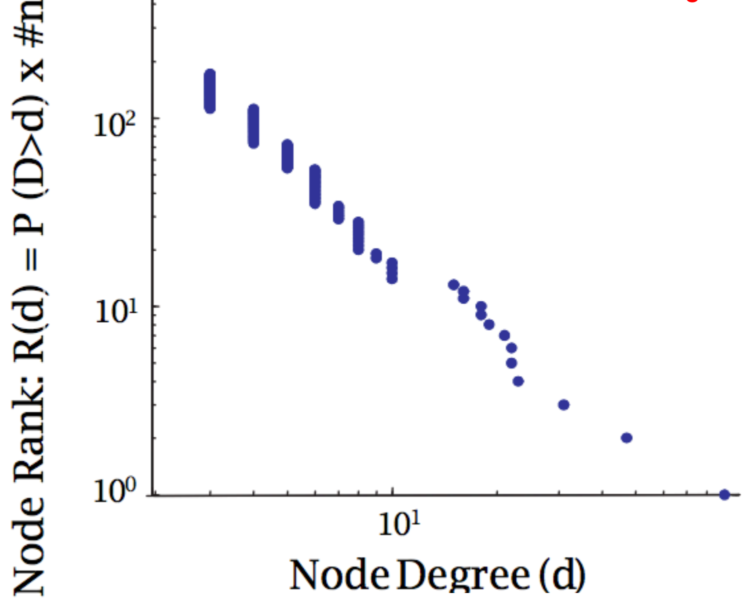
Determine the routing matrix through shortest-path algorithm

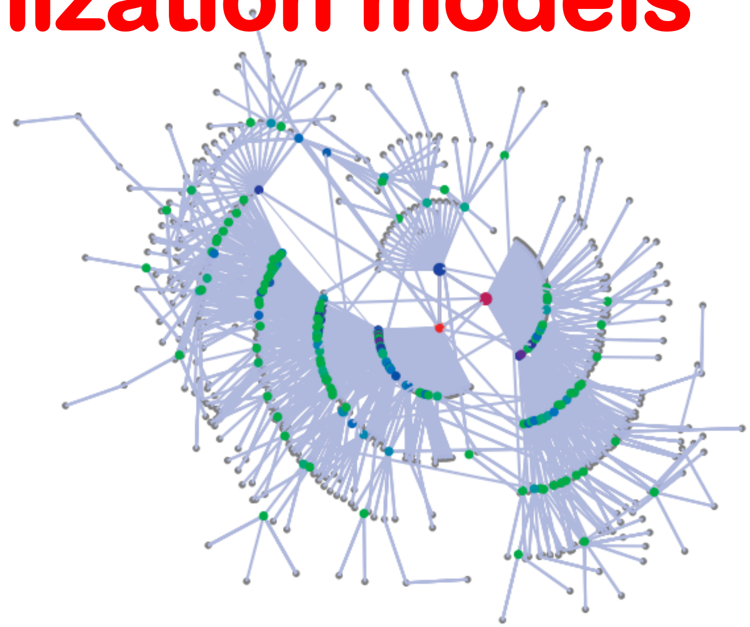
Solve constrained flow optimization problem

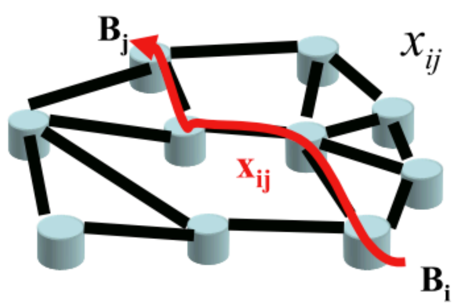
The value for the obtained flows is higher than the one obtained solving the same problem using a PA generative model

The reason is because the network model reflects real engineering insights

Constrained optimization models



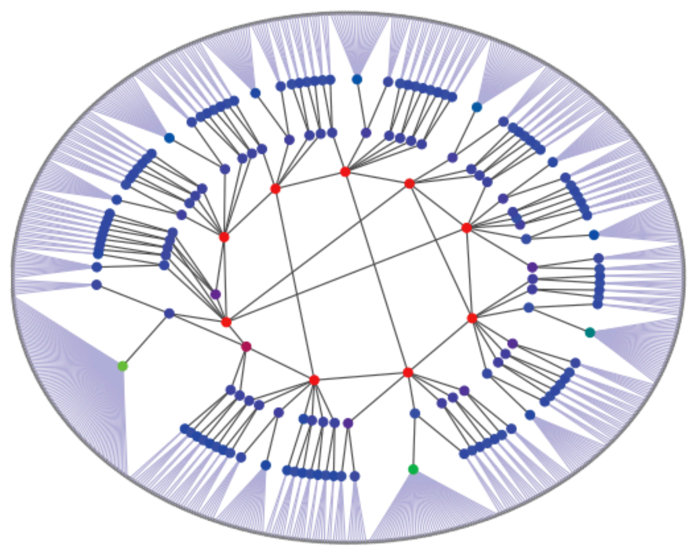




$$x_{ij} \propto B_i B_j$$

$$\max_{\alpha} \sum_{i,j} x_{ij} = \max \sum_{i,j} \alpha B_i B_j$$

$$s.t. \sum_{i,j:k \in r_{ij}} x_{ij} \le B_k, \forall k$$



(a) (b)

Figure 5. Generating networks using constrained optimization. (a) Engineers view netw

Comparing models

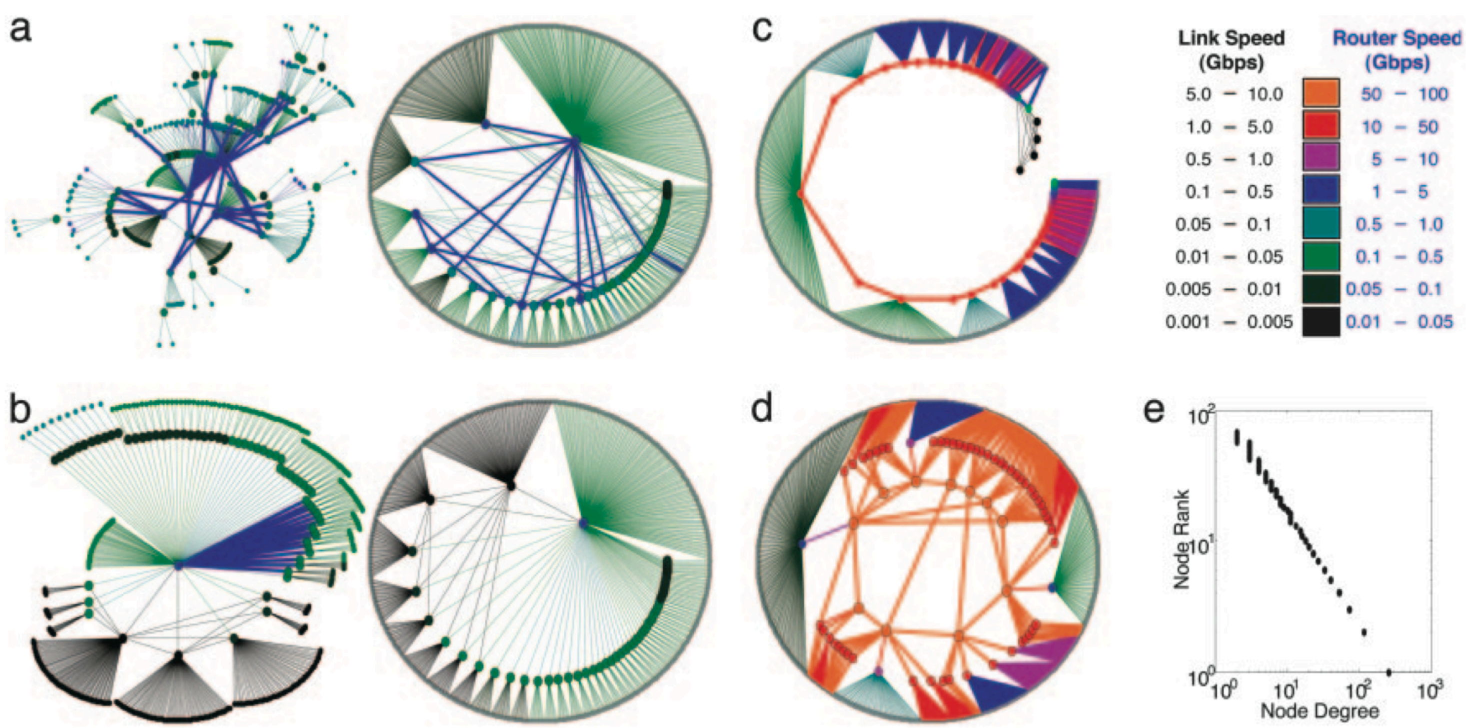


Fig. 1. Diversity among graphs having the same degree sequence D. (*a*) RNDnet: a network consistent with construction by PA. The two networks represent the same graph, but the figure on the right is redrawn to emphasize the role that high-degree hubs play in overall network connectivity. (*b*) SFnet: a graph having the most preferential connectivity, again drawn both as an incremental growth type of network and in a form that emphasizes the importance of high-degree nodes. (*c*) BADNet: a poorly designed network with overall connectivity constructed from a chain of vertices. (*d*) HOTnet: a graph constructed to be a simplified version of the Abilene network shown in Fig. 2. (*e*) Power-law degree sequence D for networks shown in *a*–*d*. Only $d_i > 1$ is shown.

Comparing models

Feature	PA net	HOTnet	Real Internet
High-degree vertices	Core	Periphery	Periphery
Degree distributions	Power law	Power law	Highly variable
Generated by	Random	Design	Design
Core vertices	High degree	Low degree	Low degree
Throughput	Low	High	High
Attack tolerance	Fragile	Robust	Robust
Fragility	High-degree/ hubs	Low-degree/core	Hijack network

In Summary

- In PA high degree nodes are essential for connectivity
- In real Internet removal of high degree nodes has only local effects
- PA model leads to poor performance in terms of maximum throughput
- Real Internet is the result of an optimization process with many constraints
  - Tech and economic constraints restrict feasible topologies
  - Maximize throughput with router flow constraints
- Power law degree distributions naturally arise from constrained optimization problems

In Summary

- A small world ring with random shortcuts (Backbone)
- With attached local multi-level tree structures (Point of access that aggregate traffic)
- Might be a better toy model for Internet Graph
- It exhibits:
  - Small diameter
  - Power-law degree distribution
  - Clustering
  - High aggregate throughput
  - Resilient to both random failures and targeted failures
  - It is the result of an optimization process

Heuristically optimized trade-offs

Consider a random tree driven by a uniform distribution of points in the unit square

$$i \leftrightarrow j : \min_{j<i} \quad \alpha d_{i,j} + h_j$$

Every newly added node minimizes the weighted sum of two objectives

"Last Mile" connection cost (Euclidean distance)

"Centrality" (Hop-distance to other nodes)

$$h_j = \mathbb{E}(\text{hops to others})$$

$$h_j = \max(\text{hops to others})$$

$$h_j = \text{hops to central node}$$

Heuristically optimized trade-offs

Fabrikant, Koutsoupias, Papadimitriou (2002)

$$\alpha < 1/\sqrt{2} \implies \text{T is a star}$$

$$\alpha > c_1\sqrt{n} \implies \mathbb{E}(|\{i : \deg_i \geq k\}|) < n^2 \exp(-c_2 k)$$

$$\alpha > 4,\ \alpha = o(\sqrt{n}) \implies \mathbb{E}(|\{i : \deg_i \geq k\}|) > c(k/n)^{-\alpha}$$

Heuristically optimized trade-offs

This suggests that power laws can be the manifestation of trade-offs, complicated optimization problems with multiple and conflicting objectives.

Finding the correct trade-offs requires an understanding of these complex processes that drive the network construction mechanism

Power law degree distribution

In 1999 it was observed that the Internet graph both at the router level (IP) and at the AS level (BGP) has a power law degree distribution

$$P(k_i = k) = Ak^{-\alpha}, \quad 2 < \alpha < 3$$

But is it really a power law?

Unreliable measurements

IP Alias resolution problem

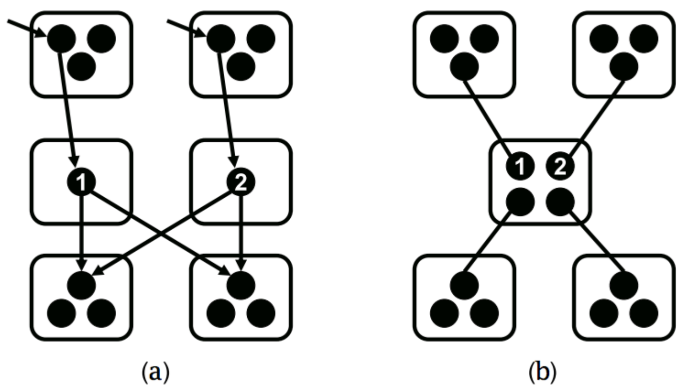
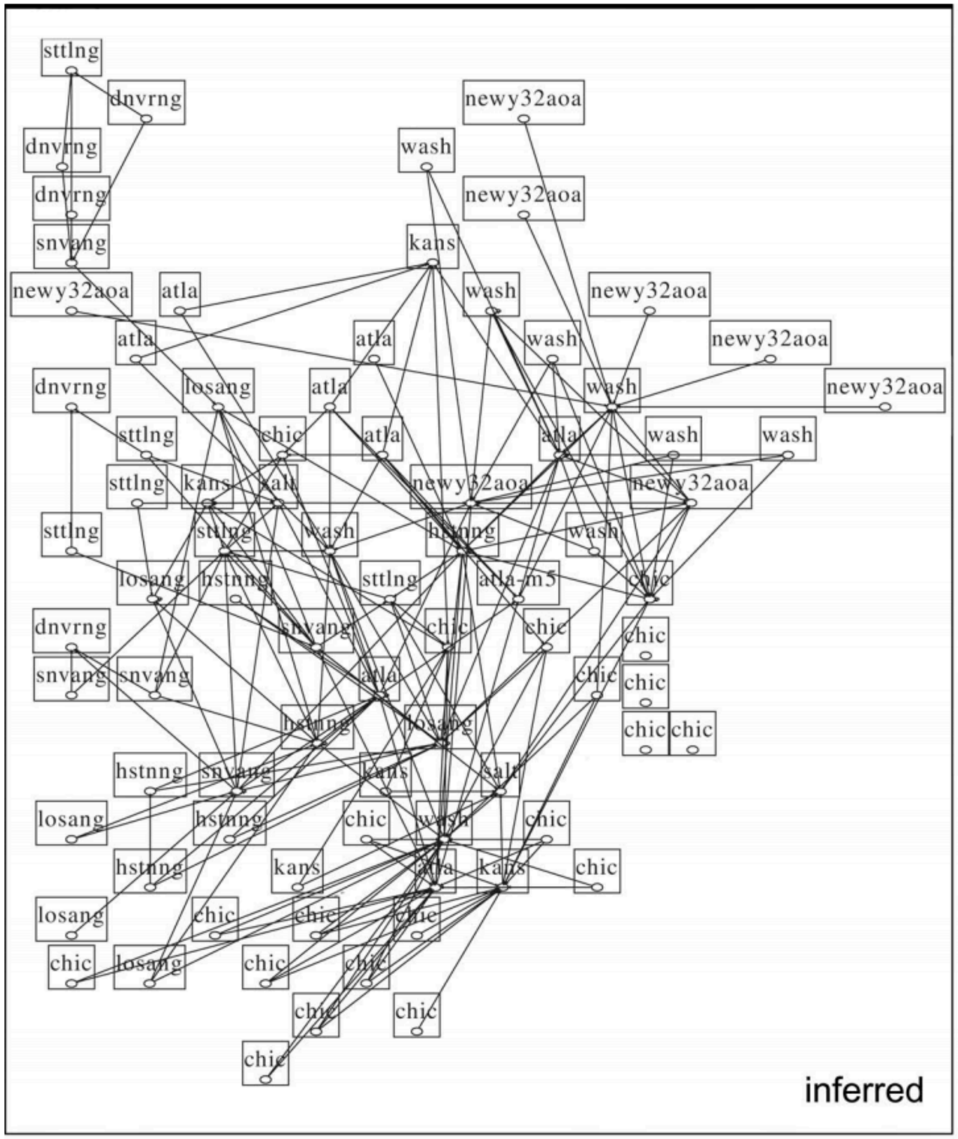
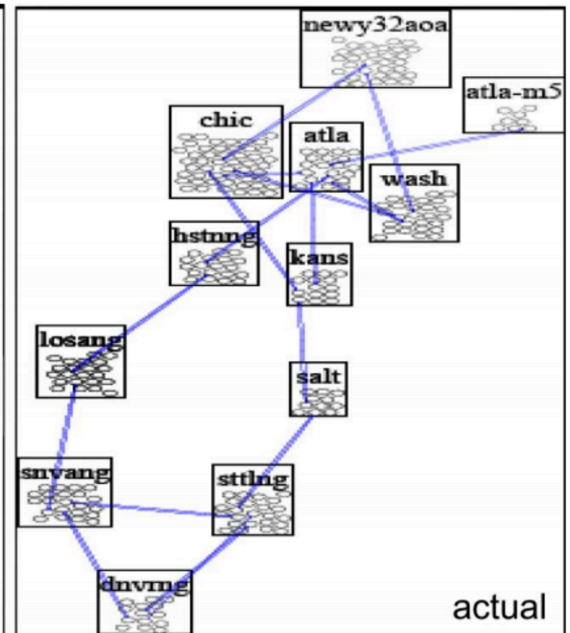
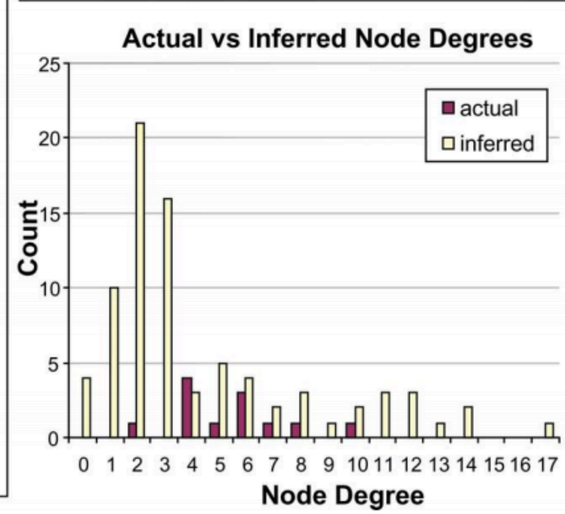


Figure 2. The IP alias resolution problem.

Unreliable measurements

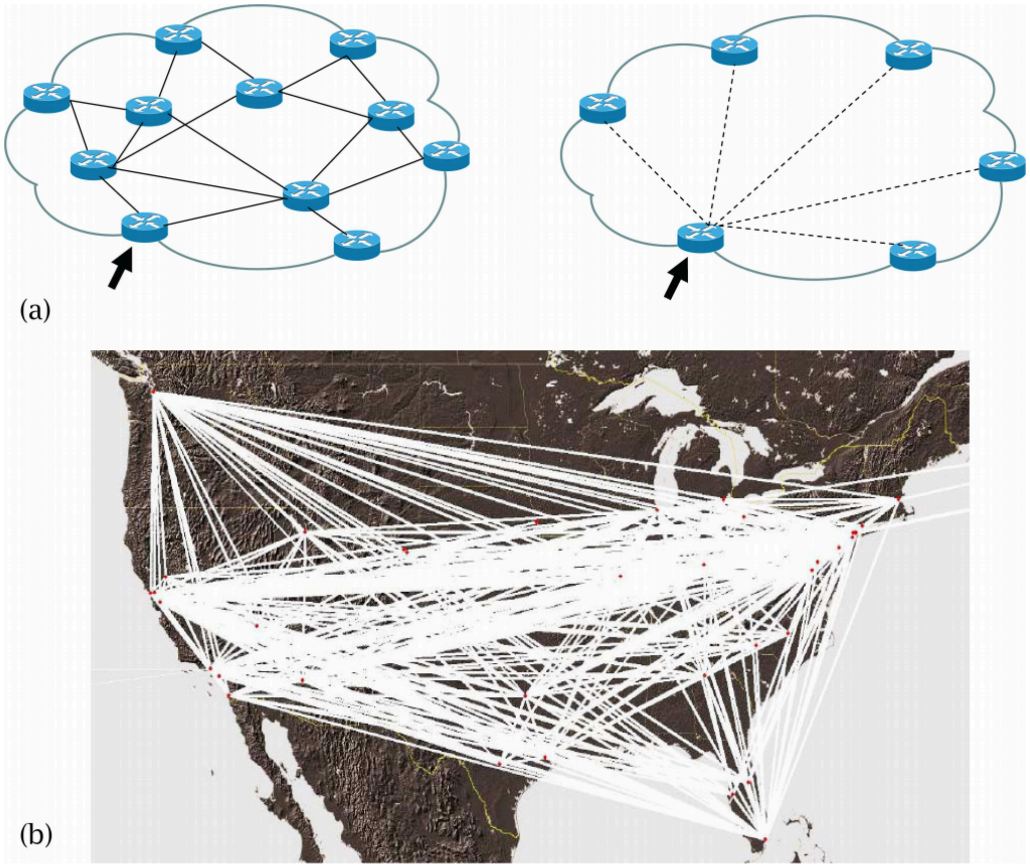






Unreliable measurements

Hidden layer network problem



Biased sampling

Even if we assume measurements are reliable and we sample a BSF tree

High degree nodes and nodes close to the root are more likely to be sampled

Sampling is biased with respect to the property to be sampled!

The bias introduced by BSF sampling can make power laws appear where they do not exist. Even a random (ER) graph or a regular random graph, where each vertex has the same degree is reported to have a power law degree distribution.

Biased sampling

How can we infer the true degree distribution from sampling?

This is a key problem in network science, beyond internet modeling, for example in social networks exploration. Methods usually involve some amount of bias that needs to be controlled.