# NBA 600
# Networks and the Internet
**Class 3, Mon 10/22**

**Prof. Dan Huttenlocher**



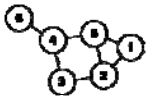
## Today's Class

- What are networks
- Properties of networks
  - Attributes of individuals in a network
    - Very different from those of individuals alone
  - Value of a network
    - Metcalfe's "law", network effect
- Structure of the Internet
  - Collection of networks
  - Internet Protocol (IP)
  - Services: email, Web, file transfer, …



## What's a Network?

- Abstractly: a system of inter-connected components
  - Represent as graph with nodes and edges
- Physical networks
  - Railroad, highway, telephone, cable
- Electronic networks
  - Television, radio, Internet
- Information networks
  - Web, payment



Pages, links



## Kinds of Networks

- Centrally engineered networks
  - Action of a single controlling agent
    - E.g., Cornell's campus network, Sprint's network, global corporate data network
- Distributed or self-organizing networks
  - Actions of many "independent" agents
    - E.g., Internet, trade networks
- Natural networks
  - Biological and ecological systems
    - E.g., nervous system, waterways



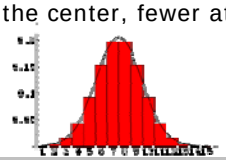
## Properties of Networks

- The connectivity in networks makes nodes behave differently than isolated individuals
  - Particularly for self organizing networks
- Generally people more familiar with properties of individuals than networks
  - E.g., human heights
  - Distribution of these properties
    - Normal or bell-shaped
    - Central tendency (mean)
    - Degree of variation (variance or stand. dev.)



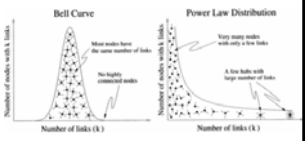
## Normal Distribution

- Normal or bell-shaped curve characterizes many properties of individuals
- X-axis property, e.g. height
- Y-axis number (or proportion) of individuals with that property
- More near the center, fewer at "tails"



## Power Law Distribution

- Quantitative properties of nodes in network often power law distributed
  - Very different from normal distribution
- For instance
  - Links to Web page
  - Visitors to Web site
- Mean doesn't make much sense
  - No central or average case
- Long tail, 80/20 rule





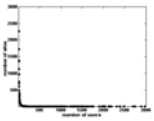

## Consider Following

- Individuals each with a certain score on some test or quantity of some good
- Bell-shaped
  - 25 with 1
  - 60 with 2
  - 150 with 3
  - 60 with 4
  - 25 with 5
- Power law
  - 300 with 1
  - 30 with 10
  - 3 with 100
- What is average (mean) in each case?
  - How representative?

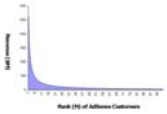


## Examples of Power Laws

- Number of AOL users to visit given Web sites in particular day
  - X, visitors to a site (property)
  - Y, sites with this many visitors (count)



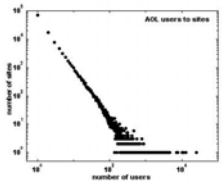
- Revenues of AdSense advertisers
  - X, rank of advertiser (rank of count)
  - Y, revenue at that rank (property)





## Log-Log Plot

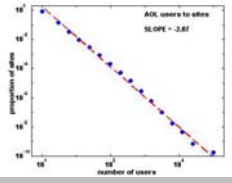
- Hard to see what is going on in plot of power law distribution
  - Very large number of low values dominates
- Instead plot logarithms of both quantities






## Power Law Distribution

- Characteristic of power law is a good linear fit to log-log plot
  - Exponentially many items have low values compared to those with high values
  - High values are exponentially larger than low values






## Why this is Important

- What value does "the majority" have in a power law vs. a normal distribution?
  - Near the bottom vs. near the middle
  - Winner-take-most phenomena
- In a networked world vast differences between top, nearly top, and low rank
  - Web site traffic, daily page views per million
    - #1 Ranked Yahoo: 80,000
    - #9 Ranked Youtube: 8,000
    - #98 Ranked cnet: 450

(source: alexa.com)

## Aside: Distribution of Wealth

- Is wealth normally distributed, power law distributed, something else?
  - Appears to be power law
- Network property rather than individual
  - Arises from independent actions of many other individuals
  - For information networks often very few limits on degree of network-based properties of individuals
    - E.g., physical constraint of having shaken hands with Bill Gates vs. contributed to his wealth



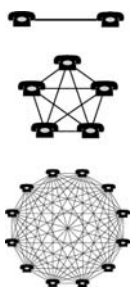
## Value of a Network

- We have been focusing on properties of individual "nodes" in a network
- What about the network itself
  - Particularly centrally engineered networks such as traditional telephone networks
- Simple hypothesis is that value grows proportional to number of "nodes"
  - Each new user increases value by one unit
  - Linear model
  - E.g., pay per subscriber to acquire a network



## Metcalfe's "Law"

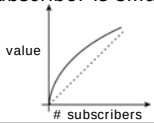
- Bob Metcalfe, inventor of the Ethernet, postulated the value of a network increases with the square of the number of users
  - In graph, edges between pairs, count edges rather than nodes
  - Pairwise communication such as telephone, IM, fax, email
  - Value for each user goes up when add one new user
    - Increase of n rather than 1





## Valuing Networks in Practice

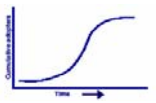
- Is a network with 100M subscribers worth 10x or 100x a network with 10M subscribers?
  - Linear vs. quadratic, n vs. $n^2$
- Studies have suggested lies in between
  - As number of subscribers gets large additional value of new subscriber is smaller






## Growth of Networks

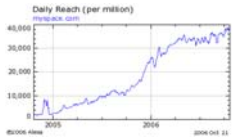
- Tends to follow s-shaped curve over time
  - Slow growth, the increasing rate, then decreasing
- Similar form seen in diffusion of innovation, crossing the chasm
  - Early, middle and late adopters normally distributed yield similar growth curve



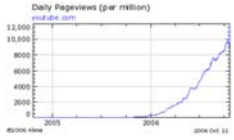


## Examples of Growth

- Myspace




- Youtube



## Network Effects Arise Broadly

- Not always an obvious network
  - Transmission of information on underlying networks of social interaction
- Format effects
  - Videotape, CD, DVD formats
  - Electronic document formats, Word, pdf
- Expertise/training effects
  - Software packages
- Brand or lifestyle effects
  - Making choices for what others think



## The Internet

- Differs from other large manmade networks, highly distributed
  - E.g., electricity, telephone, transportation, …
  - Internet not planned or administrated by one or small number of authorities/providers
    - Grows and evolves based on actions of many companies
- Internet is collection of many networks
  - Each network consists of computers and routers (two types of components)
    - Use common "Internet Protocol" (IP)



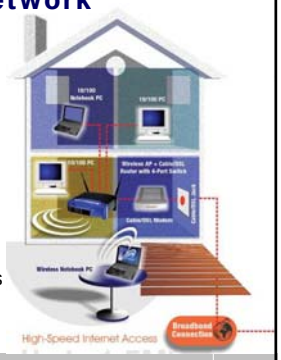
## The Internet

- Different network layers
- Physical layer – connections between places
  - Wires, fiber, microwave, wireless, …
- Electronic layer – getting information from one place to another
  - Internet Protocol (IP), BGP, …
- Information layer – content that people are interested in
  - Web, IM, email, …



## Home Network

- All computers communicate via router
- Broadband connection is link to "public Internet" via service provider (ISP)
  - Note: router can protect your computers from being accessed





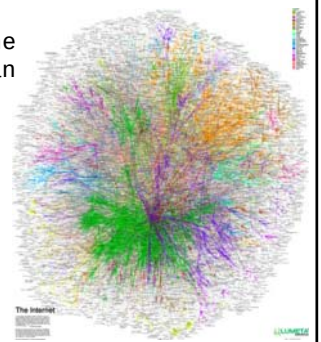

## Make-Up of Internet

- Your home network is not by itself one of the networks making up the Internet
  - Your ISP gives you one or more IP addresses
    - An IP address is somewhat like a phone number, it specifies a particular computer on the Internet
  - Generally on a home network the IP address is for your router, and your computers are not directly accessible
  - Your ISP often uses "dynamic IP" where your IP address is not fixed
    - Better for outgoing than incoming connections



## Make-Up of the Internet

- Each network in the Internet is called an autonomous system (or AS)
  - Your ISP is (generally) an AS
  - Cornell is an AS
- There are tens of thousands of AS's
  - Each a "provider"



## How Does This Work?

- Separate inter-connected networks around the globe
  - Some are large ISP's with millions of customers on a single network
- No central planning or authority
  - Not like the few big telco's in the US coordinating, which is hard enough
- Key is in the design of the underlying communication protocols of the Internet
  - IP routing: getting data between computers



## IP Routing Between Networks

- Each router connects two or more networks together
  - Home router connects home network to ISP's network
  - Routers can be within or between AS's
- Certain AS's are connected by routers
  - Have "peering agreements" about what data they will send over these connections
    - Terms, financial and otherwise, up to each pair of AS's that decide to connect directly
    - Generally large global providers have multiple connections



## IP Routing Analogy

- Like driving without a map
  - At each intersection either take a road that says it is heading towards your destination or one that says "all other destinations"
- Data broken into packets
  - Like having people on same itinerary travel separately, each making own choices at intersections
  - Signs at intersections may change based on congestion so routes differ
- Generally good for avoiding congestion



## Internet Structure

- A particular structure has evolved
  - Tier 1 providers (ATT, Level3, Sprint, …)
    - Generally operate own multinational or global networks
    - All have peering arrangements with each other, where do not pay for traffic
  - Regional providers (commercial), Tier 2
  - Local commercial providers
  - Residential providers
    - Local phone companies
    - Cable companies



## Internet Payment Models

- Would be substantial overhead to charging for each packet of data
  - Lots of bookkeeping
  - Potentially networks from many intermediate providers traversed
- Model that has evolved is based on end user payment for bandwidth
  - Peak amount of data that can be sent at once
  - Total amount of data sent over some time period
- Peering agreements between providers



## End-User Internet Payment

- Peak bandwidth (millions of bits/sec)
  - $15/mo 56kb (consumer dialup)
  - $30/mo 128kb upstream - 768kb downstream broadband (consumer "broadband")
  - $300/mo 1.6mb T1 (commercial)
  - $5-20k/mo 155mb OC3 (commercial)
- Commercial users tend to use their full bandwidth 24/7, consumers not
  - Asymmetric broadband disproportionately cheaper as a result
- Each of these categories generates billions of annual revenue to ISP's

## Internet Structure and Payment

- Nothing cast in stone about these payment models or structures
  - Whatever customers are willing to pay for
  - Anyone can set up a business that charges for access in other ways
    - But needs to cover costs of infrastructure and of paying others for access
  - Recently hot area is virtual system operators or providers of network services
    - Has been for wireless phone networks more than Internet



## Internet Services

- End users of the Internet rarely see anything about the IP-level
  - Except maybe configuring an IP address at some point
- Services on top of basic network
  - DNS maps names of sites to IP addresses
  - HTTP web protocols deliver content
  - SMTP email protocols deliver mail
  - Various (incompatible) IM protocols



## Next Time

- Discuss strategy and the Internet
  - Read Porter, Hammonds and Tapscott articles