Analytical Background

John C.S. Lui

Notations

Probability	
$P[\ldots]$	Probability of an event
X, Y, etc.	Random variables
$p_X(x)$	PDF or distribution function of X
$p_{X,Y}(x,y)$	Joint PDF or distribution function of X and Y
$F_X(x)$	Cumulative distribution of X, $P[X \leq x]$
$E[X]$	Expected value of X
μ_X	Mean of X
$\mathrm{Var}(X)$, σ_X^2	Variance of X
$\mathrm{Cov}(X,Y)$	Covariance of X and Y
$\hat{a}$	An estimate or an estimator of a
$\mathcal{N}(\mu,\sigma^2)$	The Normal (Gaussian) distribution, mean μ and variance σ^2

Random variable & CDF

- **Definition**: is the outcome of a random event or experiment that yields a numeric values.
- For a given x, there is a fixed possibility that the random variable will not exceed this value, written as $P[X<=x]$.
- The probability is a function of x, known as $F_X(x)$. $F_X(.)$ is the cumulative distribution function (CDF) of X.

PDF & PMF

- A continuous random variable has a probability density function (PDF) which is:

$$p_X(x) \equiv \frac{dF(x)}{dx}.$$

- The possibility of a range $(x_1,x_2]$ is

$$P[x_1 < X \leq x_2] = F(x_2) - F(x_1) = \int_{x_1}^{x_2} p(u)\,du.$$

- For a discrete random variable. We have a discrete distribution function, aka. possibility massive function.

$$p_X(x) \equiv P[X = x].$$

Moment

- The expected value of a continuous random variable X is defined as

$$E[X] \equiv \int_{-\infty}^{\infty} u\, p(u)\, du.$$

- Note: the value could be infinite (undefined). The mean of X is its expected value, denote as μ_X
- The n^{th} moment of X is:

$$E[X^n] \equiv \int_{-\infty}^{\infty} u^n\, p(u)\, du.$$

Variability of a random variable

- Mainly use variance to measure:

$$\mathrm{Var}(X) \equiv E[(X-\mu)^2] = \int_{-\infty}^{\infty} (u-\mu)^2\, p(u)\, du.$$

- The variance of X is also denote as: σ^2_X
- Variance is measured in units that are the square of the units of X; to obtain a quantity in the same units as X one takes the standard deviation:

$$\sigma_X \equiv \sqrt{\sigma_X^2}.$$

Joint probability

- The joint CDF of X and Y is:

$$F_{X,Y}(x,y) \equiv P[X \leq x, Y \leq y]$$

- The covariance of X and Y is defined as:

$$\mathrm{Cov}(X,Y) \equiv E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,y\,p_{X,Y}(x,y)\,dx\,dy.$$

- Covariance is also denoted: $\sigma^2_{X,Y}$
- Two random variable X and Y are *independent* if: $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y) \quad \forall x, y.$

Conditional Probability

- For events A and B the conditional probability defined as:

$$P[A \mid B] \equiv \frac{P[A, B]}{P[B]}.$$

- The conditional distribution of X given an event denoted as:

$$p(x \mid \text{event}).$$

- It is the distribution of X given that we know that the event has occurred.

Conditional Probability (cont.)

- The conditional distribution of a discrete random variable X and Y

$$p(x \mid Y = y)$$

- Denote the distribution function of X given that we happen to know Y has taken on the value y.
- Defined as:

$$p(x \mid Y = y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}.$$

Conditional Probability (cont.)

- The conditional expectation of X given an event:

$$E[X \mid \text{event}] \equiv \int_{-\infty}^{\infty} u\, p(u \mid \text{event})\, du.$$

Central Limit Theorem

- Consider a set of independent random variable $X_1, X_2, \ldots X_N$, each having an arbitrary probability distribution such that each distribution has mean μ and variance σ^2
- When $N \to \infty$

$$\frac{1}{N} \sum_{i=1,\ldots,N} X_i \xrightarrow{d} \mathcal{N}(\mu, \sigma^2/N).$$

With parameter μ and variance σ^2/N

Commonly Encountered Distributions

- Some are specified in terms of PDF, others in terms of CDF. In many cases only of these has a closed-form expression

Distribution	Definition	Domain
Exponential	$p(x) = \lambda e^{-\lambda x}$	$x > 0$
Normal	$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$	$-\infty < x < \infty$
Gamma	$p(x) = \frac{(x-\gamma)^{\alpha-1} \exp[-(x-\gamma)/\beta]}{\beta^\alpha \Gamma(\alpha)}$	$x > \gamma$
Extreme	$F(x) = \exp\left[-\exp\left(-\frac{(x-\alpha)}{\beta}\right)\right]$	$-\infty < x < \infty$
Lognormal	$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2\right]$	$x > 0$
Pareto	$p(x) = \alpha k^\alpha x^{-\alpha-1}$	$x > k$
Weibull	$p(x) = \frac{bx^{b-1}}{a^b} \exp\left[-\left(\frac{x}{a}\right)^b\right]$	$x > 0$

Stochastic Processes

- Stochastic process: a sequence of random variables, such a sequence is called a stochastic process.
- In Internet measurement, we may encounter a situation in which measurements are presented in some order; typically such measurements arrived.

Stochastic Processes

- A stochastic process is a collection of random variables indexed on a set; usually the index denote time.
- Continuous-time stochastic process:

$$\{X_t,\ t \geq 0\}$$

- Discrete-time stochastic process:

$$\{X_n,\ n = 1, 2, ...\}.$$

Stochastic Processes

- Simplest case is all random variables are independent.
- However, for sequential Internet measurement, the current one may depend on previous ones.
- One useful measure of dependence is the auto-covariance, which is a second-order property:

$$\sigma_{X_i,X_j} = \mathrm{Cov}(X_i, X_j) \equiv E[(X_i - \mu_{X_i})(X_j - \mu_{X_j})]$$

Stochastic Processes

- First order to n-order distribution can characterize the stochastic process.
  - First order: $p_{X_1}(\cdot),\ p_{X_2}(\cdot),\ \ldots$
  - Second order:

$$p_{X_1,X_2}(\cdot),\ p_{X_1,X_3}(\cdot),\ \ldots,\ p_{X_2,X_3}(\cdot),\ p_{X_2,X_4}(\cdot),\ \ldots$$

- Stationary
  - Strict stationary

$$p_{X_n,X_{n+1},\ldots,X_{n+N-1}}(\cdot) = p_{X_{n+k},X_{n+k+1},\ldots,X_{n+k+N-1}}(\cdot)$$

For all n,k and N

Stochastic Processes

- Stationary
  - Wide-sense Stationary (weak stationary)
    - If just its mean and autocovariance are invariant with time.

$$\begin{aligned} E[X_n] &= E[X_1] \\ \mathrm{Cov}(X_n, X_{n+k}) &= \mathrm{Cov}(X_1, X_{k+1}) \end{aligned}$$

Stochastic Processes

- Measures of dependence of stationary process
  - Autocorrelation: normalized autocovarience

$$r(k) \equiv \gamma(k)/\gamma(0) = \gamma(k)/\sigma_X^2.$$

  - Entropy rate
    - Define entropy:

$$H(X) \equiv \sum_{x \in \mathcal{H}} p(x) \log 1/p(x)$$

    - Joint entropy:

$$H(X_1, X_2) \equiv \sum_{x_1 \in \mathcal{H}} \sum_{x_2 \in \mathcal{H}} p(x_1, x_2) \log 1/p(x_1, x_2).$$

Stochastic Processes

- Measures of dependence of stationary process
  - Entropy rate
    - The entropy per symbol in a sequence of n symbols

$$H_n(\mathcal{H}) = \frac{1}{n} H(X_1, X_2, ..., X_n).$$

    - The entropy rate

$$H(\mathcal{H}) = \lim_{n \to \infty} H_n(\mathcal{H}).$$

Special Issues in the Internet

- Relevant Stochastic Processes
  - Arrivals: events occurring at specific points of time
  - Arrival process: a stochastic process in which successive random variables correspond to time instants of arrivals:

$$\{A_n,\ n = 0, 1, ...\}$$

    - Property: non-decreasing & not stationary
  - Interarrival process (may or may not stationary)

$$\{I_n,\ n = 1, 2, ...\} \text{ where } I_n \equiv A_n - A_{n-1})$$

Special Issues in the Internet

- Relevant Stochastic Processes
  - Timeseries of counts
    - Fixed-size time intervals and counts how many arrivals occur in each time interval. For a fixed time interval T, the yields $\{C_n, \ n = 0, 1, ...\}$ where:

      $$C_n \equiv \#\{A_m \,|\, nT < A_m \leq (n+1)T\}$$
    - T called timescale
    - Can use an approximation to the arrival process by making additional assumption (such as assuming Poisson)
    - A more compact description of data

Short tails and Long tails

"In the case of network measurement large values can dominate system performance, so a precise understanding of the probability of large values is often a prime concern"

- As a result we care about the upper tails of a distribution
- Consider the shape of

$$1 - F(x) = P[X > x] \text{ for large } x$$

Short tails and Long tails

- Declines exponentially if exists >0, such that:

  $$1 - F(x) \sim e^{-\lambda x}$$

  – AKA. Short-tailed or light-tailed

  – Decline as fast as exponential or faster.

- Subexponential distribution

  $$(1 - F(x))e^{\lambda x} \rightarrow \infty \text{ as } x \rightarrow \infty \text{ for all } \lambda > 0$$

  – A long tail

  – The practical result is that the samples from such distributions show extremely large observations with non-negligible frequency

Short tails and Long tails

- Heavy-tailed distribution:
  - a special case of the subexponential distributions
  - Asymptotically approach a hyperbolic (power-law) shape
  - Formally:

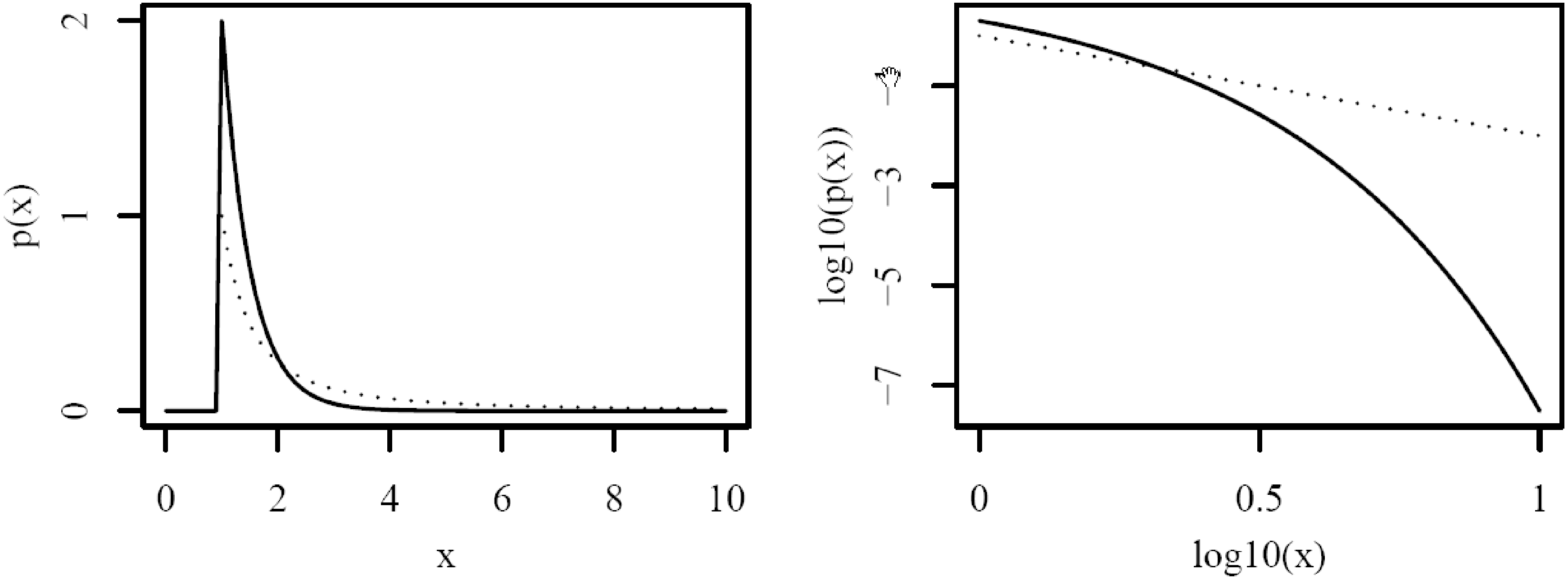
$$1 - F(x) \sim x^{-\alpha} \qquad 0 < \alpha \leq 2$$

  - Such a distribution will have a PDF also follow a power law:

$$p(x) \sim x^{-\alpha-1}$$

Short tails and Long tails

- A comparison of a short-tailed and a long-tailed distribution



(a) (b)

Figure 3.1 Comparison of Short- and Long-tailed Distributions: (a) Linear Scale (b) Log-log Scale. Dotted line: Heavy-tailed Pareto distribution; Solid line: light-tailed Exponential distribution.

Statistics

Terms of measured data

- Terms used in describing data
  - For example: “mean of a dataset”
  - An objectively measurable quantity which is the average of a set of known values
- Terms used in describing probability models
  - For example: “mean of a random variable”
  - A property of an abstract mathematical construct
- To emphasize the distinction, we add the adjective “empirical” to describe data
  - Empirical mean vs. mean
- Classification of measured data
  - Numerical: *i.e. numbers*
  - Categorical: *i.e. symbols, names, tokens, etc.*

Central tendency

- Definition
  - Given a dataset $\{x_i, i=1,\ldots N\}$, it tells where on the number line the values tend to be located.
- ***Empirical mean (average)***

$$\bar{x} = \frac{1}{N} \sum_{i=1,\ldots,N} x_i$$

- ***Mode***
  - most common value
- ***Median***
  - value which divides the sorted dataset into two equal parts

Dispersion

- Measured methods
  - ***Empirical variance***: squared units

$$s^2 = \frac{1}{N} \sum_{i=1,\ldots,N} (x_i - \overline{x})$$

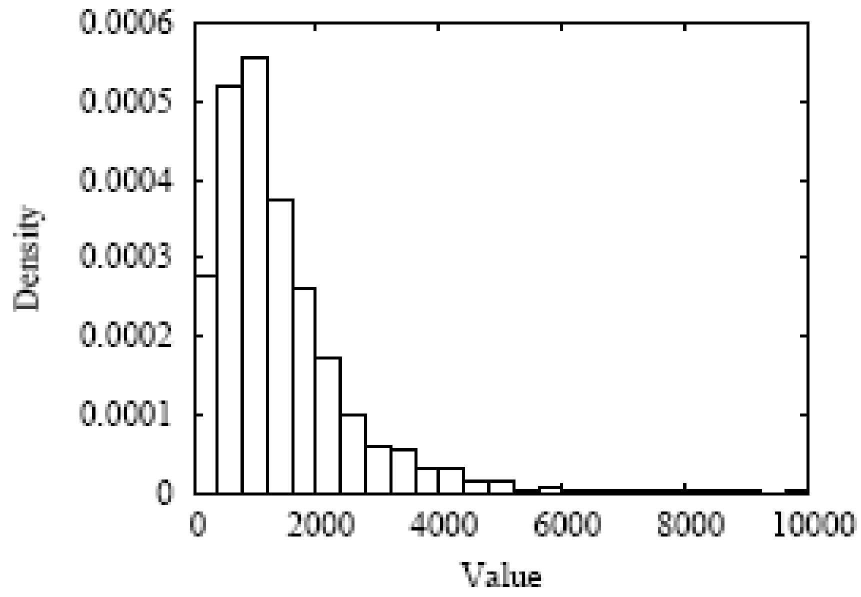
  - ***Standard deviation***: the square root of variance
  - ***Coefficient of variation***: $s / \overline{x}$

More detailed descriptions

- ***Quantiles***
  - The p^{th} quantile is the value below which the fraction p of the values lies.
  - Median is the 0.5-quantile
- ***Percentile***
  - The 90^{th} percentile is the value that is larger than 90% of the data

Histogram

- Defined in terms of bins which are a particular of the observed values
- Counts how many values fall in each bin
- A natural empirical analog of a random variable's probability density function (PDF) or distribution function
- Practical problem:
  - How to determine the bin boundaries



Entropy

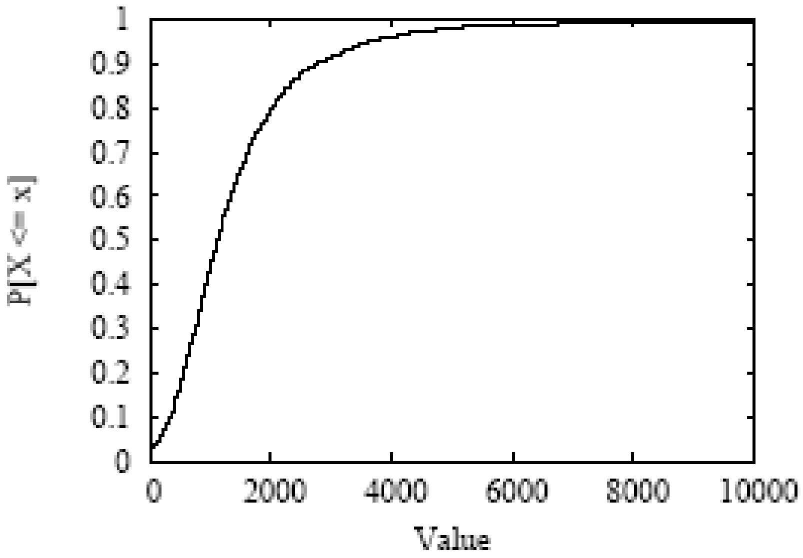
- Definition

  Let *P* be a probability mass function on the symbol set *A*, the entropy of P is $H(P) = -\sum_{x \in A} P(x) \log P(x)$

- Entropy measures the unevenness of a distribution
- The maximum entropy is log|A|, when all symbols are equally likely, $P(x) = 1/|A|$ for every x $\in$A

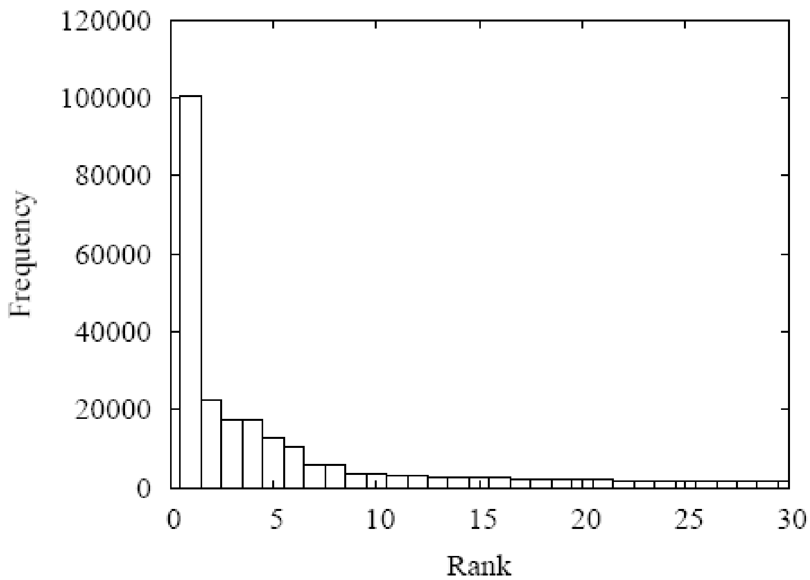
Empirical cumulative distribution function (CDF)

- CDF involves no binning or averaging of data values
- CDF potentially provides more information about the dataset than does the histogram.
- For each unique value in the data set, the fraction of data items that are smaller than that value (quantile).
- CDF involves no binning or averaging of data values
- CCDF: *complementary cumulative distribution function*



Categorical data description

- Probability distribution
  - Measure the empirical probability of each symbol in the dataset
  - Use histogram in decreasing order



Describing memory and stability

- ***Timeseries*** data
  - Question: Do successive measurements tend to have any relation to each other?
- ***Memory***
  - When the value of a measurement tends to give some information about the likely values of future measurements
  - Empirical autocorrelation function (ACF)

$$\hat{r}(k) = \frac{\frac{1}{N-k}\sum_{i=1,\ldots,N-k}(x_i - \overline{x})(x_{i+k} - \overline{x})}{s^2}$$

- Stability
  - If its empirical statistics do not seem to be changing over time.
  - Subjective
  - Objective measures
    - Break the dataset into *windows*

Special issues

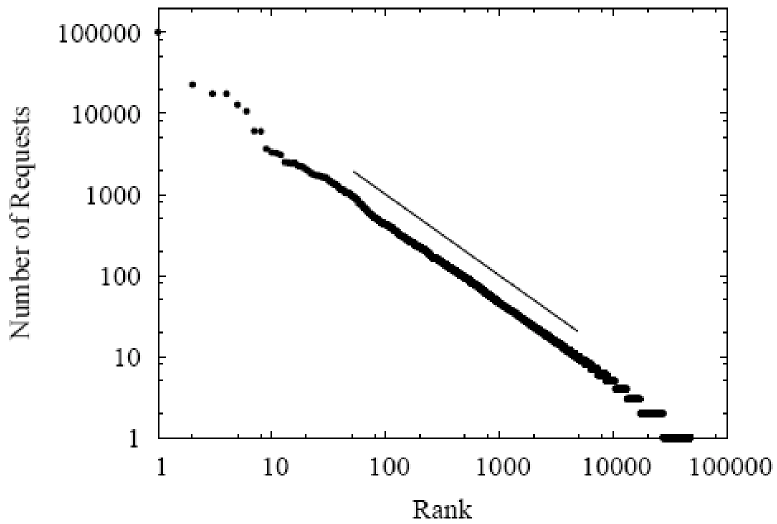
- High variability (Numeric data distribution)
  - Traditional statistical methods focuses on low or moderate variability of the data, *e.g. Normal distribution*
  - Internet data shows high variability
    - It consists of many small values mixed with a small number of large value
    - A significant fraction of the data may fall many standard deviations from the mean
    - Empirical distribution is highly skewed, and empirical mean and variance are strongly affected by the rare, large observations
    - It may be modeled with a subexponential or heavy tailed distribution
    - Mean and variance are not good metrics for high variability data, while quantiles and the empirical distribution are better, e.g. empirical CCDF on log-log axes for long-tailed distribution

Special issues

- Zipf's law (a categorical distribution)
  - A model for the shape of a categorical distribution when data values are ordered by decreasing empirical probability, *e.g. URLs of Web pages*

$$R = cn^{-\beta} \quad (\beta \approx 1 \; or \; \beta < 1)$$

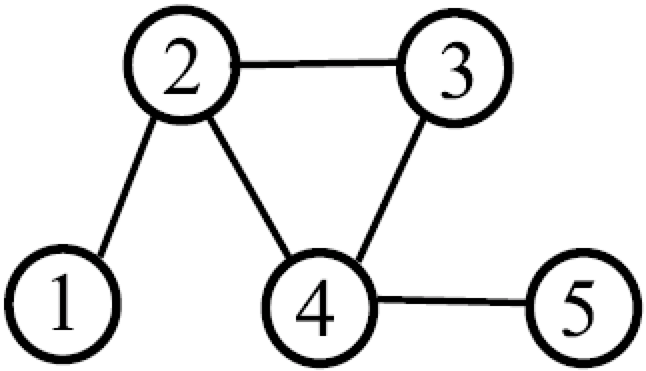
  - Categorical data distributions



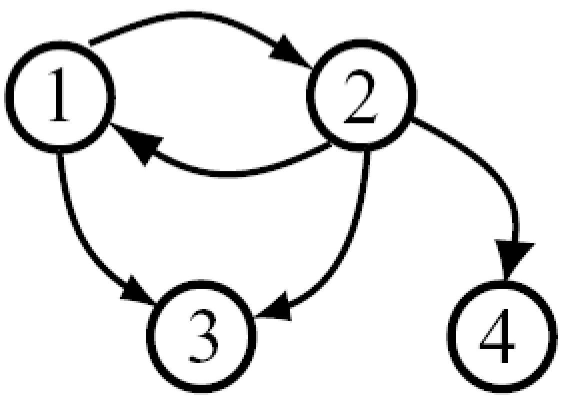
Graphs and Topology

Background of Graph

- A graph is a pair $\mathcal{G}=(V,E)$
  - Undirected graph and directed graph
  - Weighted graph and unweighted graph



(a)



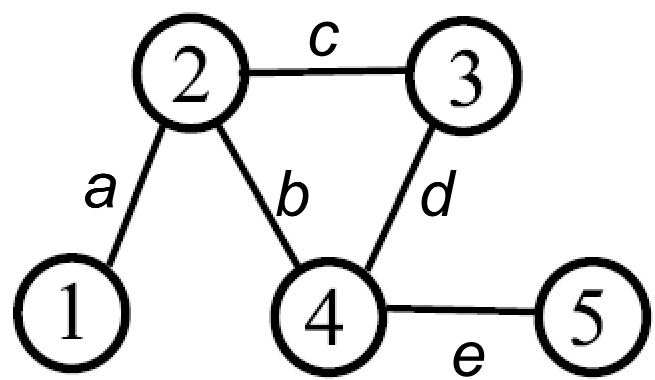
(b)

Subgraph

- Subgraph $\mathcal{G}'$=(V',E') of $\mathcal{G}$=(V,E)
  - $V' \subset V$
  - $(v_1, v_2) \in E'$ if and only if $(v_1, v_2) \in E$ $v_1, v_2 \in V'$
- Clique
  - Complete subgraph

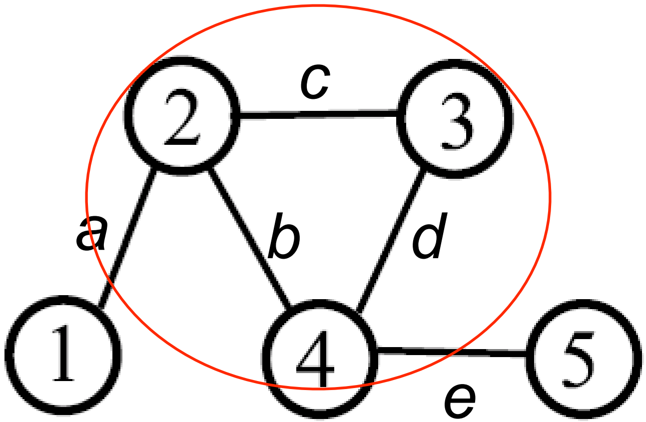
Subgraph

- Subgraph $\mathcal{G}'=(V',E')$ of $\mathcal{G}=(V,E)$
  - $V' \subset V$
  - $(v_1, v_2) \in E'$ if and only if $(v_1, v_2) \in E$ $v_1, v_2 \in V'$
- Clique
  - Complete subgraph



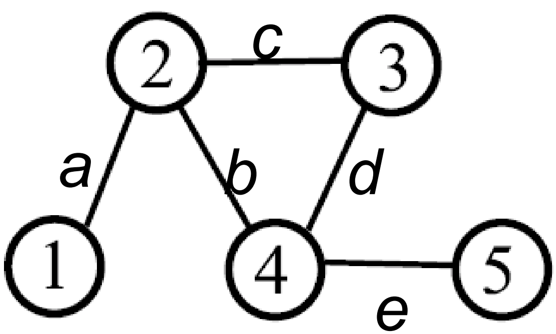
Subgraph

- Subgraph $\mathcal{G}'=(V',E')$ of $\mathcal{G}=(V,E)$
  - $V' \subset V$
  - $(v_1, v_2) \in E'$ if and only if $(v_1, v_2) \in E$ $v_1, v_2 \in V'$
- Clique
  - Complete subgraph



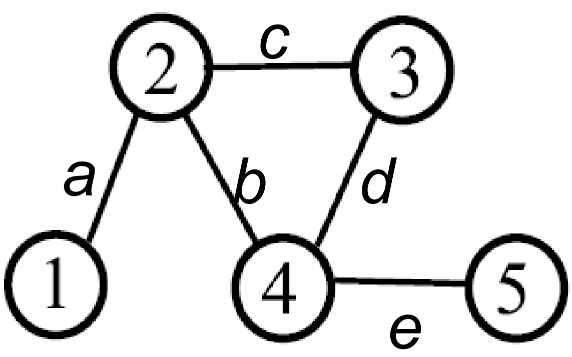
Connected Graph

- Path
  - A sequence of vertices v_1, v_2,…,v_n that there is an edge from each vertex to the next vertex in the sequence
  - If the graph is directed, then each edge must in the direction from the current vertex to the next vertex in sequence
- Connected vertices
  - Two vertices v_i and v_j are connected if there is a path that starts with v_i and ends with v_j.
- Connected graph:
  - Undirected graph with a path between each pair of vertices
- Strong connected graph
  - Directed graph with a path between each pair of vertices



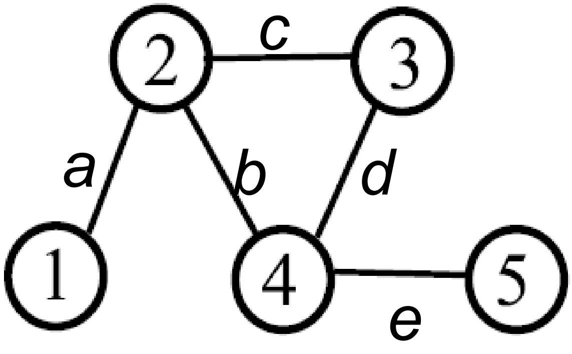
Characterizing Graph Structure (1)

- Degree
  - The degree of a vertex is the number of edges incident to it
    - Indegree and outdegree of directed graph
- Shortest path
  - The shortest path length between two vertices *i* and *j* is the number of edges comprising the shortest path (or a shortest path) between *i* and *j*.
  - Diameter of a connected graph
- Characteristic path length
  - Average length of all the shortest paths between any pair of vertices



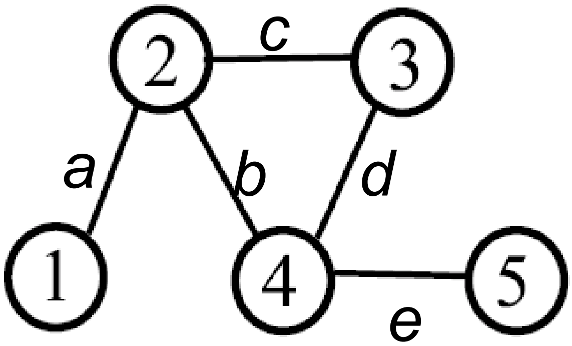
Characterizing Graph Structure (2)

- Clustering
  - The tendency for neighbors of a node to themselves be neighbors
  - Clustering efficient
- Betweenness
  - The centrality of a vertex
  - Give the set of shortest paths between all pairs of vertices in a graph, the betweenness b_i of a vertex *i* is the total number of those paths that pass through that vertex.



Associated Matrices

- Incidence matrix of $\mathcal{G}=(V,E)$
  - $n \times n$ matrix with ($n=|V|$)



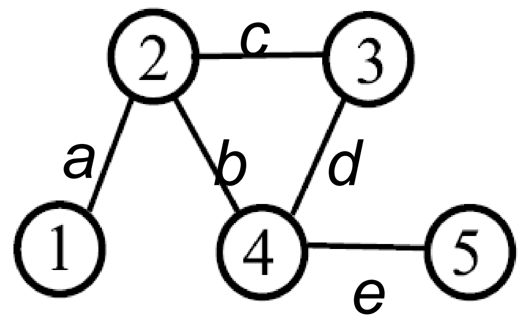
$$M = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

- Routing matrix
  - All-pairs paths

$$A = \begin{matrix} a \\ b \\ c \\ d \\ e \end{matrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

Applications of Routing Matrix (1)

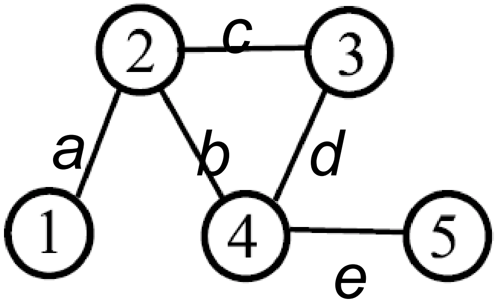
- Origin-destination (OD) flow



$$\vec{y} = A\vec{x} \Rightarrow \begin{matrix} a \\ b \\ c \\ d \\ e \end{matrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 2 \\ 2 \\ 4 \end{bmatrix}$$

Applications of Routing Matrix (2)

- $G \equiv A^T$
- Delay of paths



$$\vec{z} = G\vec{w} \Rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 3 \\ 1 \\ 1 \\ 2 \\ 1 \\ 2 \\ 1 \end{bmatrix}$$

Commonly Encountered Graph Models (1)

- *Erdös-Rényi* random graph
  - $G_{N,p}$
    - N vertices, $(N^2-N)/2$ possible edges
    - Each edge is present with probability of *p* independently
    - Expected number of edges:

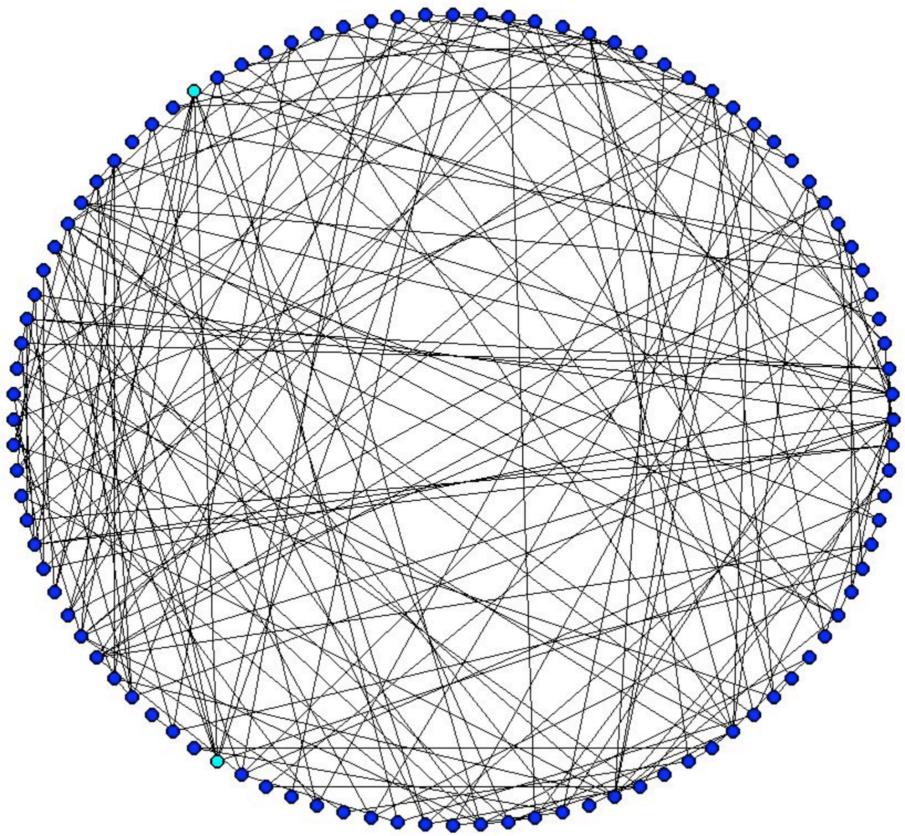
$$\mu_M = \frac{(N^2 - N)p}{2}$$

    - Expected vertex degree (Poisson distribution)

$$\mu_D = \frac{2\,\mu_M}{N} = Np - p$$

Commonly Encountered Graph Models (1)

- *Erdös-Rényi* random graph



Commonly Encountered Graph Models (2)

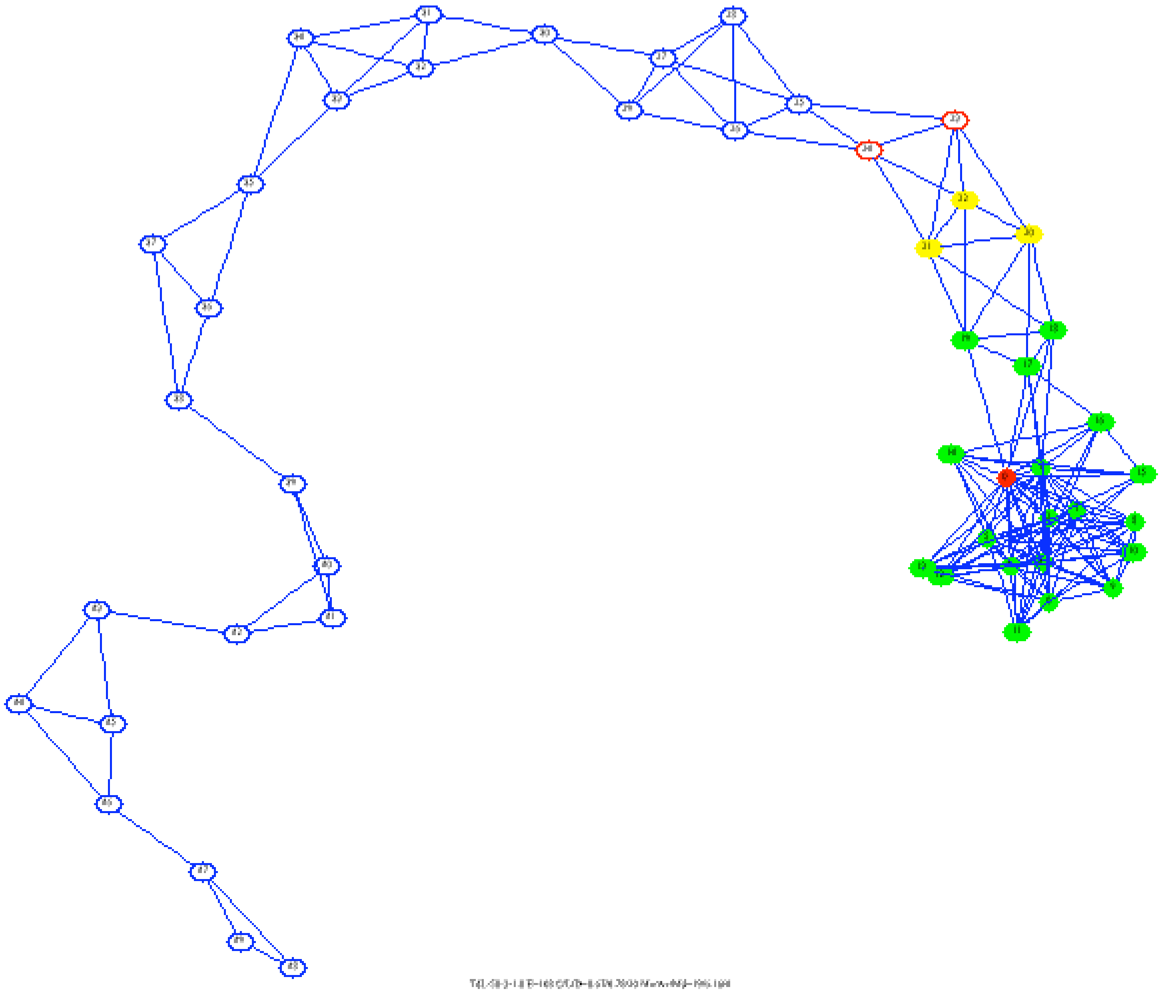
- Generalized random graph
  - Given fixed degree sequence $\{d_i, i=1,\ldots,N\}$
  - Each degree d_i is assigned to one of the N vertices
  - Edge are constructed randomly to satisfy the degree of each vertex
  - Self-loop and duplicate links may occur

Commonly Encountered Graph Models (3)

- Preferential attachment model
  - Ideas:
    - Growing network model
    - The addition of edges to the graph is influenced by the degree distribution at the time the edge is added
  - Implementation
    - The graph starts with a small set of m_0 connected vertices
    - For each added edge, the choice of which vertex to connect is made randomly with probability proportional to d_i
    - E.g. power law distribution of the degree

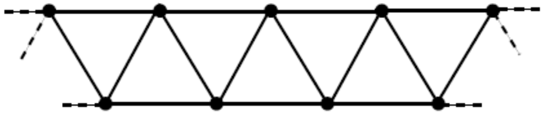
$$p_D(d) \sim d^{-3}$$

Commonly Encountered Graph Models (3)

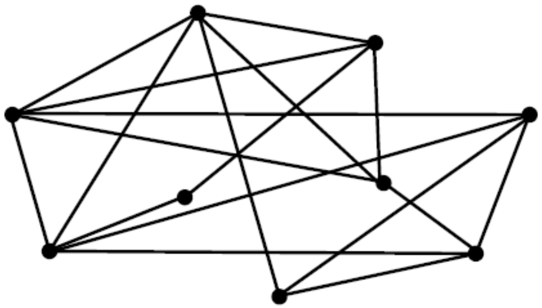


Regular Graph *vs* Random Graph

- Regular graph
  - Long characteristic path length
  - High degree of clustering
- Random Graph
  - Short paths
  - Low degree of clustering
- Small world graph
  - Short characteristic path length
  - High degree of clustering



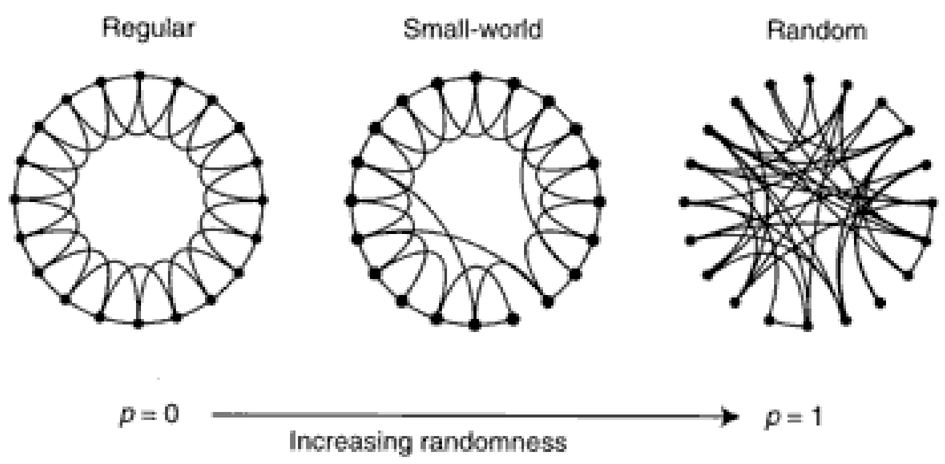
(a)



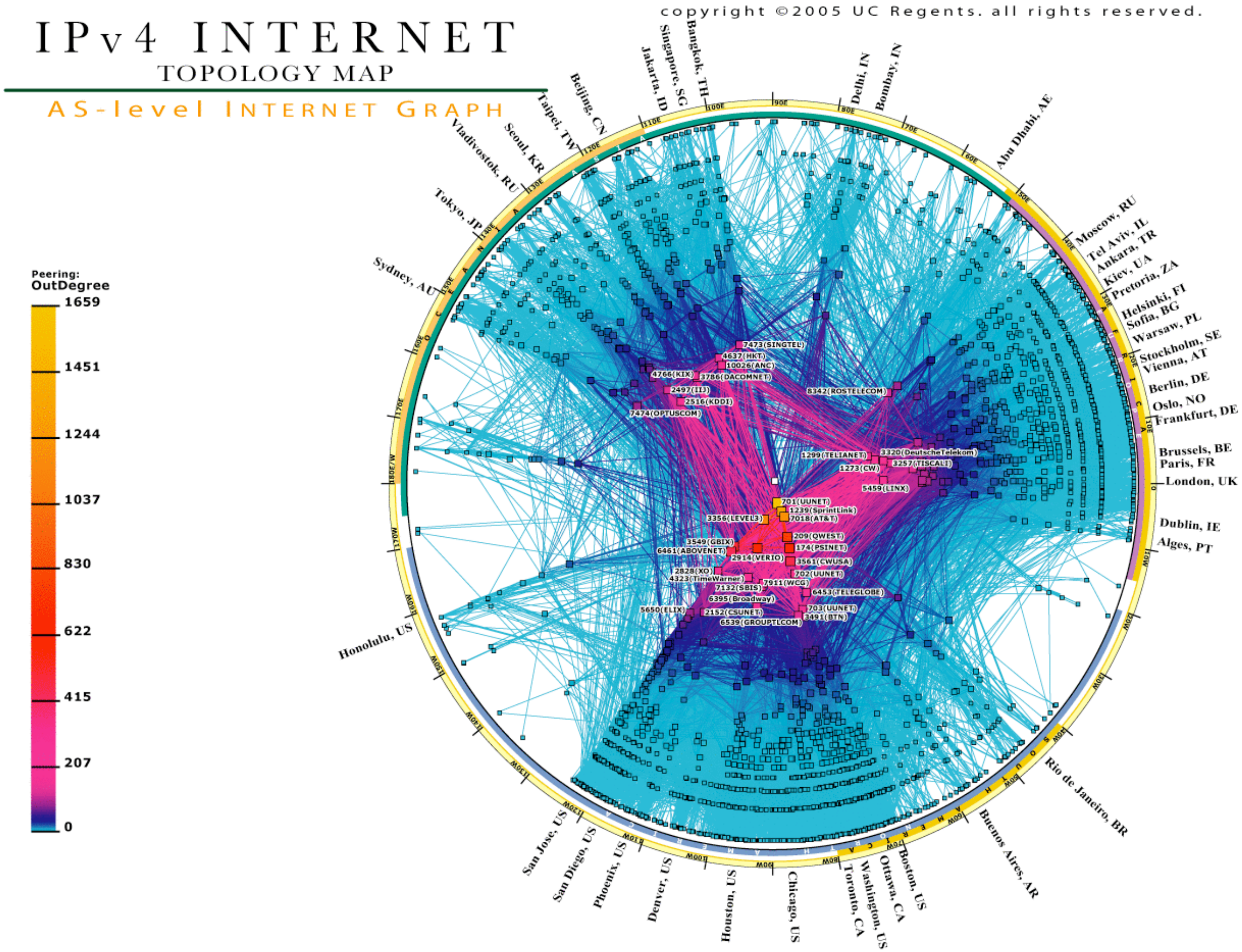
(b)

Regular Graph *vs* Random Graph

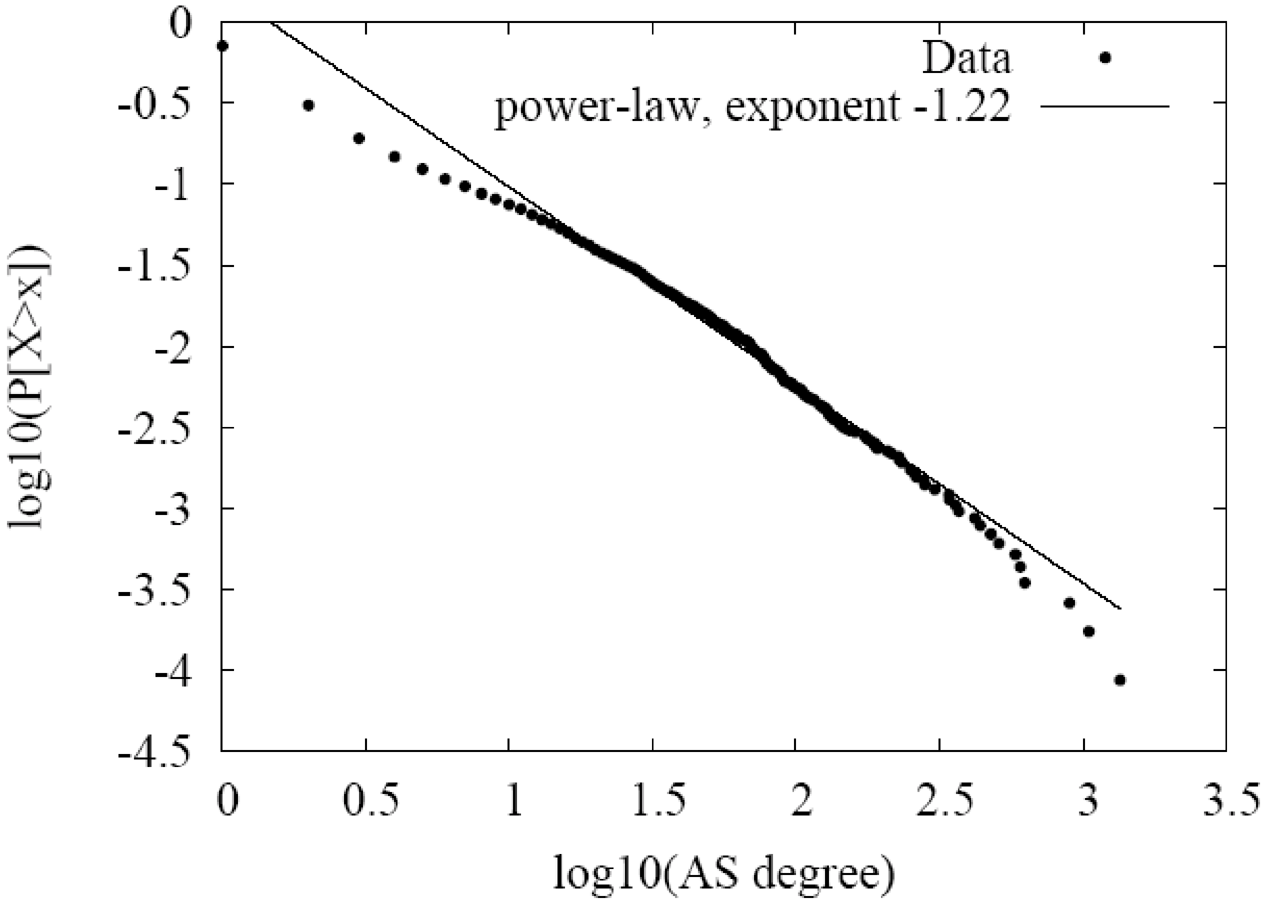
- Regular graph
  - Long characteristic path length
  - High degree of clustering
- Random Graph
  - Short paths
  - Low degree of clustering
- Small world graph
  - Short characteristic path length
  - High degree of clustering



AS-level Topology



AS-level Topology



AS-level Topology

- High variability in degree distribution
  - Some ASes are very highly connected
    - Different ASes have dramatically different roles in the network
    - Node degree seems to be highly correlated with AS size
  - Generative models of AS graph
    - “Rich get richer” model
    - Newly added nodes connect to existing nodes in a way that tends to simultaneously minimize the physical length of the new connection, as well as the average number of hops to other nodes
    - New ASes appear at an exponentially increasing rate, and each AS grows exponentially as well

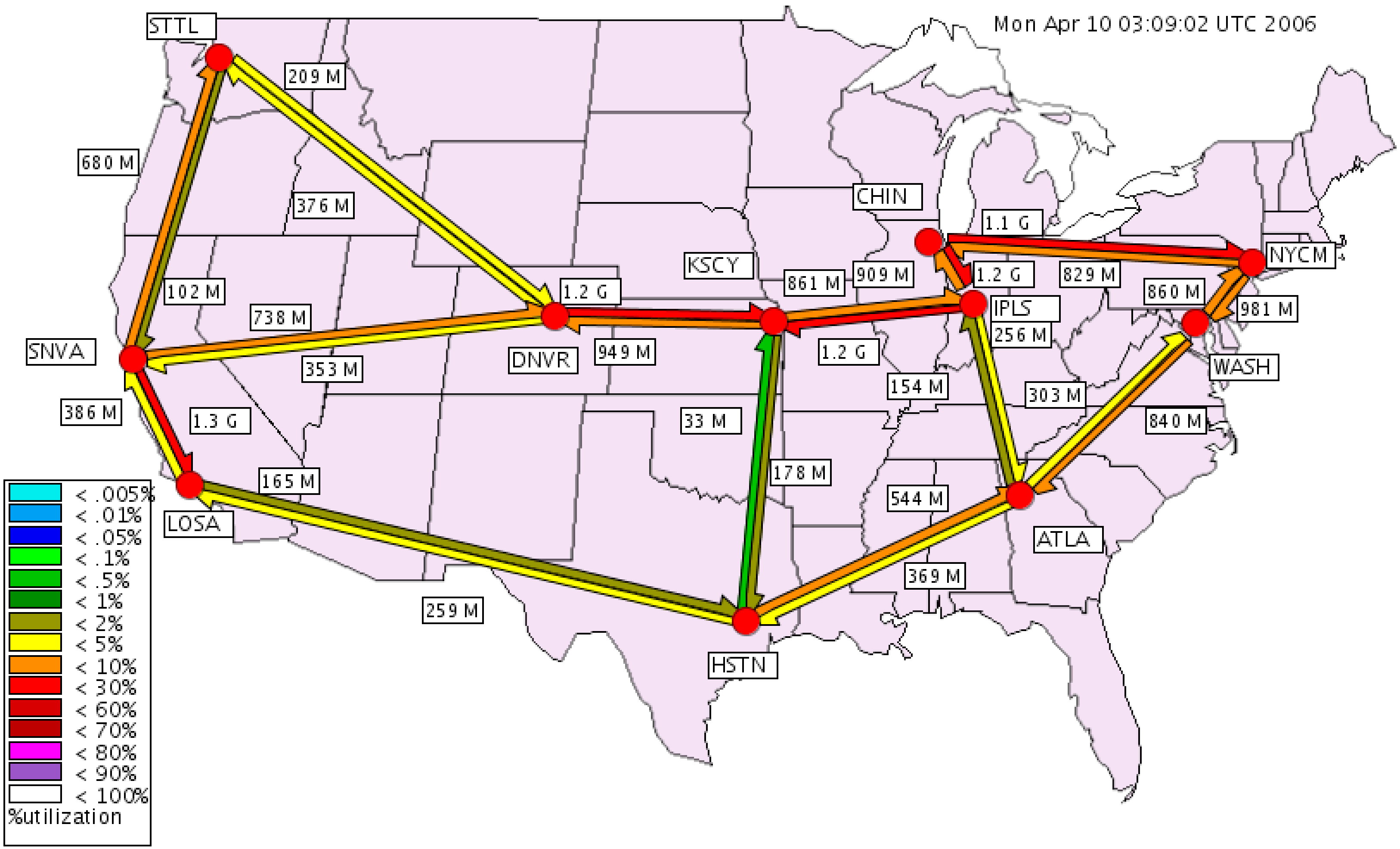
AS Graph Is Small World

- AS graph taken in Jan 2002 containing 12,709 ASes and 27,384 edges
  - Average path length is 3.6
  - Clustering coefficient is 0.46 (0.0014 in random graph)
  - It appears that individual clusters can contain ASes with similar geographic location or business interests

AS Relationships

- Four relationships
  - Customer-provider
  - Peering
    - Exchange only non-transit traffic
  - Mutual transit
    - typically between two administrative domains such as small ISPs who are located close to each other
  - Mutual backup
- Hierarchical structure?

Router-level Topology

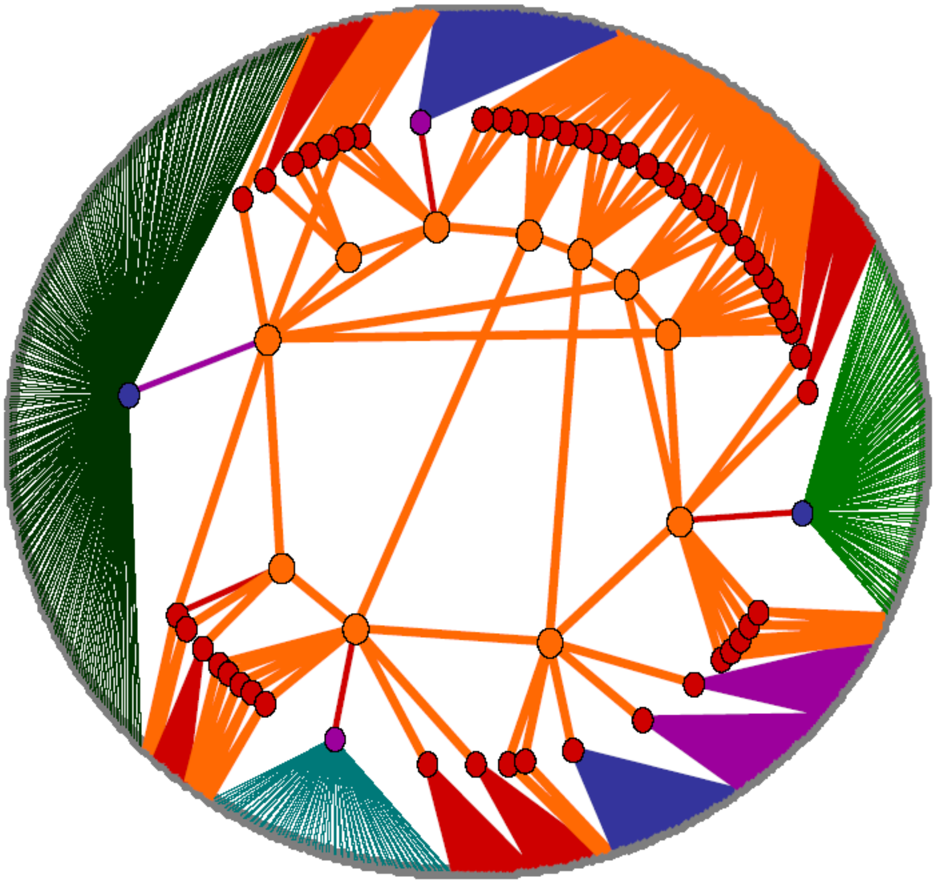


Router-level Topology

- High variability in degree distribution
  - Impossible to obtain a complete Internet topology
  - Most nodes have degree less than 5 but some can have degrees greater than 100
    - High degree nodes tend to appear close to the network edge
    - Network cores are more likely to be meshes
  - Sampling bias (Mercator and Rocketful)
    - Proactive measurement (Passive measurement for AS graph)
    - Nodes and links closest to the sources are explored much more thoroughly
    - Artificially increase the proportion of low-degree nodes in the sampled graph
- Path properties
  - Average length around 16, rare paths longer than 30 hops
  - Path inflation

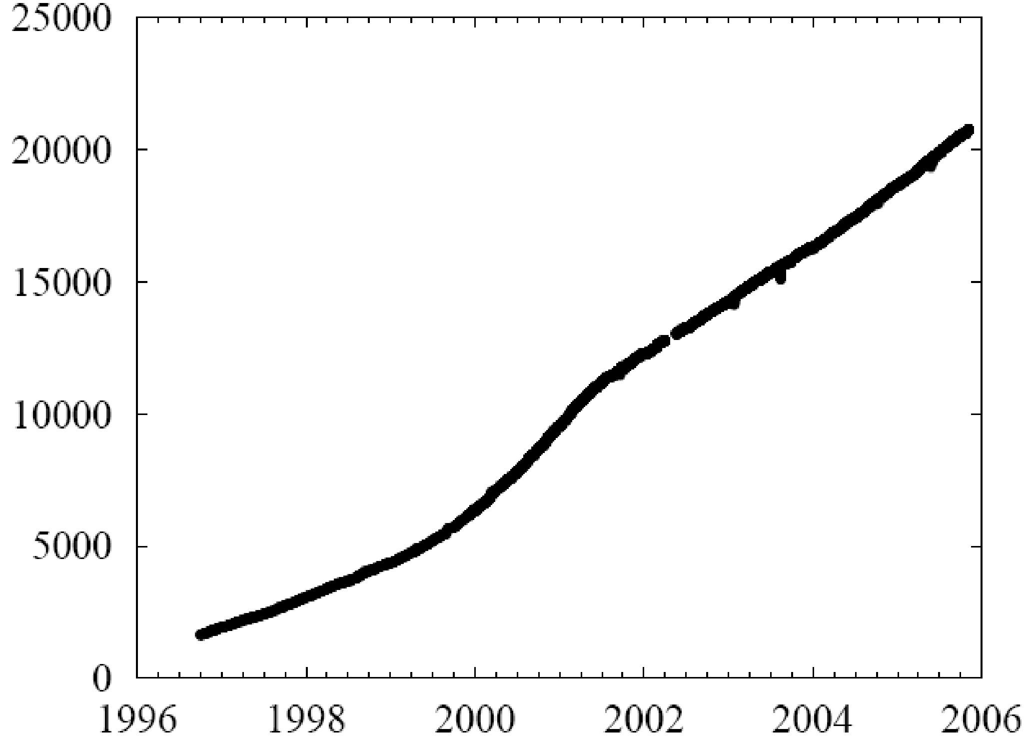
Generative Router-level Topology

- Based on network robustness & technology constrains



Dynamic Aspects of Topology

- Growing Internet



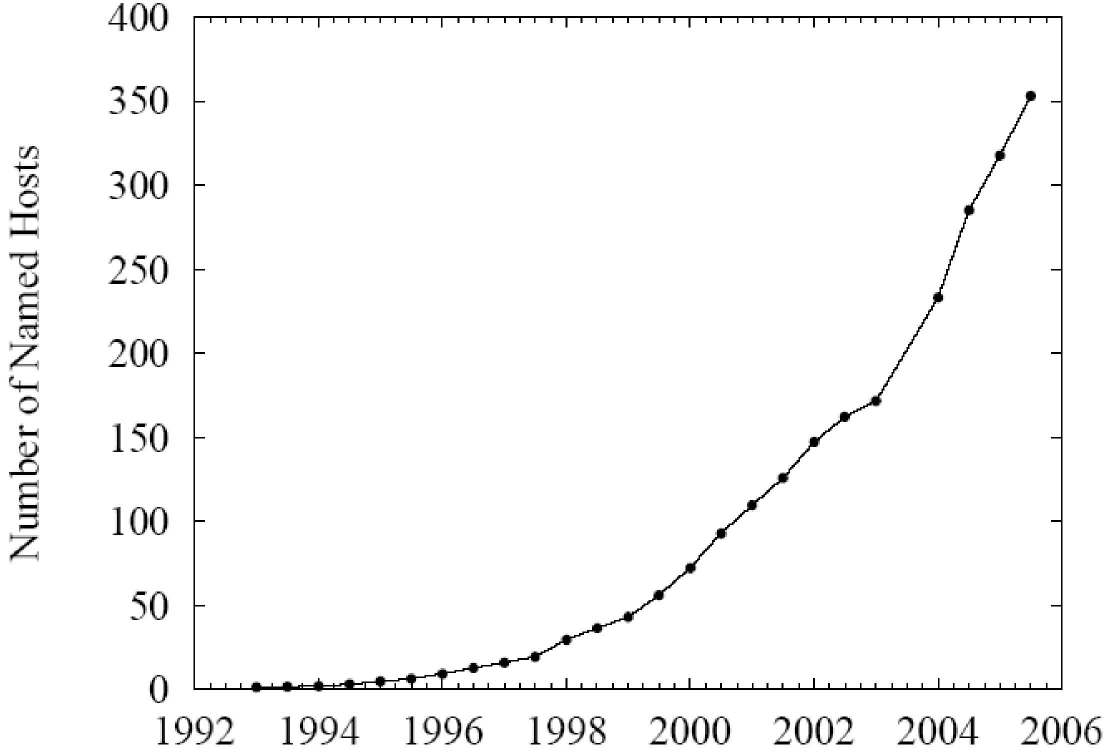
Number of unique AS numbers advertised within the BGP system

Dynamic Aspects of Topology

- Difficult to measure the number of routers
  - DNS is decentralized
  - Router-level graph changes rapidly
- Difficult measurement on endsystems
  - Intermittent connection
  - Network Address Translation (NAT)
  - No single centralized host registry

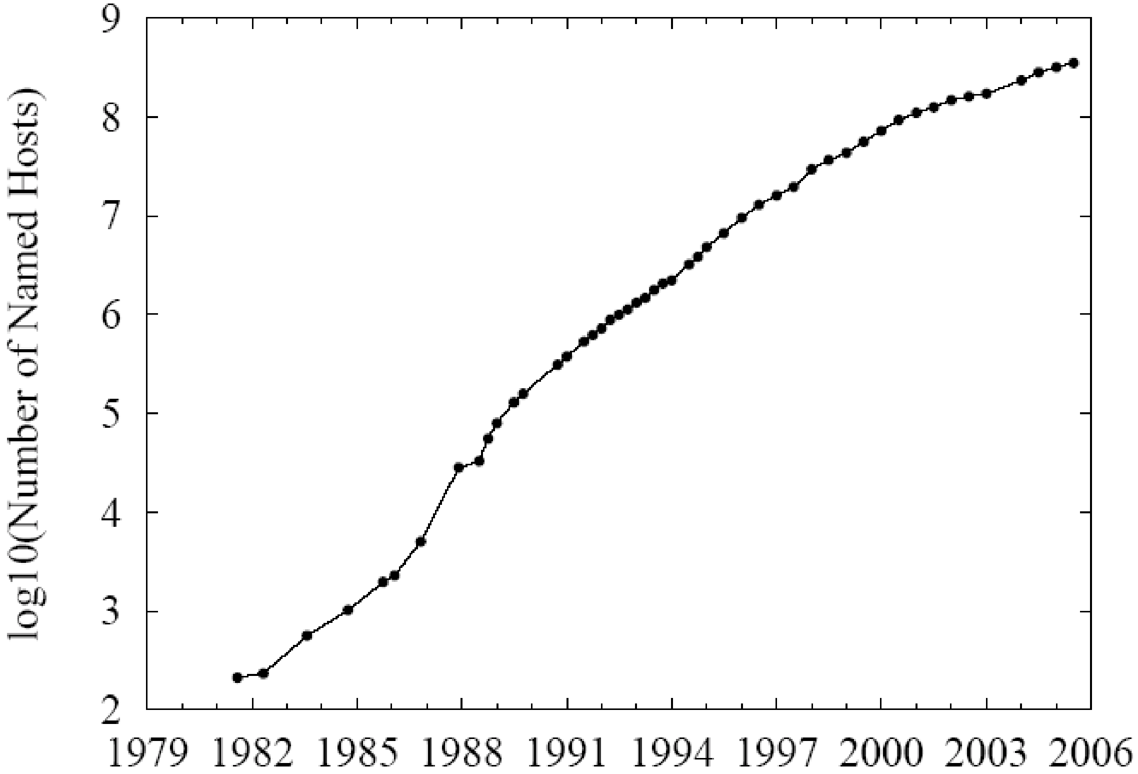
Registered Hosts in DNS

- During the 1990s Internet growing exponentially
- Slowed down somewhat today



Registered Hosts in DNS

- During the 1990s Internet growing exponentially
- Slowed down somewhat today



Stability of Internet

- BGP instability
  - Equipment failure, reconfiguration or misconfiguration, policy change
  - Long sequences of BGP updates
    - Repeated announcements and withdrawals of routers
    - Loop or increased delay and loss rates
  - Although most BGP routes are highly available, route stability in the interdoman system has declined over time
  - Instable BGP route affects a small fraction of Internet traffic
  - Unavailable duration can be highly variable

Stability of Internet

- Router level instability
  - Some routes do exhibit significant fluctuation
    - The trend may be increasing slightly
    - Consistent with the behavior AS-level paths
    - High variability of route stability
      - Majority of routes going days or weeks without change
    - High variability of route unavailable duration
  - Causes of instability of the router graph
    - Failure of links
      - Majority of link failures are concentrated on a small subset of the links
      - Marjority of link failures are short-lived (<10min)
    - Router failure

Geographic Location

- Interfaces -> IP addresses
- Population from CIESIN
- Online users from the repository of survey statistics by Nua, Inc

	Population (Millions)	Interfaces	People Per Interface	Online (Millions)	Online per Interface
Africa	837	8,379	100,011	4.15	495
South America	341	10,131	33,752	21.9	2,161
Mexico	154	4,361	35,534	3.42	784
W. Europe	366	95,993	3,817	143	1,489
Japan	136	37,649	3,631	47.1	1,250
Australia	18	18,277	975	10.1	552
USA	299	282,048	1,061	166	588
World	5,653	563,521	10,032	513	910

Table 5.1 Variation in People/Interface Density Across Regions [LBCM03].

Measurement and Modeling

Measurement and modeling

- Model
  - Simplified version of something else
  - Classification
    - A system model: simplified descriptions of computer systems
    - Data models: simplified descriptions of measurements
- Data models
  - *Descriptive* data models
  - *Constructive* data models

Descriptive data model

- Compact summary of a set of measurements
  - E.g. summarize the variation of traffic on a particular network as "a sinusoid with period 24 hours"
- An underlying idealized representation
- Contains *parameters* whose values are based on the measured data
- Drawback
  - Can not use all available information
  - Hard to answer "why is the data like this?" and "what will happen if the system changes?"

Constructive data model

- Succinct description of a process that gives rise to an output of interest
  - E.g. model network traffic as “the superposition of a set of flows arriving independently, each consisting of a random number of packets”
- The main purpose is to concisely characterize a dataset, instead of representing or simulating the real system
- Drawback
  - Model is hard to generalize --- such models may have many parameters
  - The nature of the output is not obvious without simulation or analysis
  - It is impossible to match the data in every aspect

Data model

- *"All models are wrong, but some models are useful"*
  - Model is approximate
  - Model omits details of the data by their very nature
  - Modeling introduces the tension between the simplicity and utility of a model
- *Under which model is the observed data more likely?*
  - *Models involves a random process or component*
- *Three key steps in building a data model:*
  - *Model selection*
    - *Parsimonious: prefer models with fewer parameters over those with a larger number of parameters*
  - *Parameters estimation*
  - *Validating the model*

Why build models

- Provides a compact summary of a set of measurements
- Exposes properties of measurements that are important for particular engineering problems, when parameters are interpretable
- Be a starting point to generate random but “realistic” data as input in simulation

Probability models

- Why use random models in the Internet?
  - Fundamentally, the processes involved are random
    - The value is an immense number of particular system properties that are far too tedious to specify
- Random models and real systems are very different things
  - It is important to distinguish between the properties of a probabilistic model and the properties of real data.

Probability models

Model Properties		
Autocorrelation	$E[(X_i - \mu_{X_i})(X_j - \mu_{X_j})] \neq 0, \quad i \neq j$	(Sec. 3.2.1)
Stationarity	$E[X_n] = E[X_1] \quad \forall n$, *etc.*	(Sec. 3.2.1)
Long Tail	$\bar{F}(x)\, e^{\lambda x} \to \infty$ as $x \to \infty$ for all $\lambda > 0$	(Sec. 3.2.2)

Data Properties	
System Memory	Tendency for nearby observations to be similar
Stability	Tendency for empirical statistic to not vary with time
High Variability	Highly skewed histogram which may be well modeled by a long tailed distribution