# Graph Mining

Selected and adapted slides from:
Jiawei Han and Micheline Kamber
Department of Computer Science
University of Illinois at Urbana-Champaign
www.cs.uiuc.edu/~hanj

Acknowledgements: Based on the slides by Sangkyum Kim and Chen Chen



# Natural Network Analysis

- Natural Networks: Examples
- Primitives for Natural Network Analysis
- Natural Network Characteristics
- Natural Network Generation
- Natural Network Mining

## Social Networks

- Social network: A social structure made of nodes (individuals or organizations) that are related to each other by various interdependencies: friendship, kinship, like, ...
- Graphical representation
  - Nodes = members
  - Edges = relationships

## Social Networks

**Nodes**: individuals

**Links**: social relationship (family/work/friendship/etc.)


S. Milgram (1967) **Six Degrees of Separation (hops)**

The Small-World Problem. Psychology Today, vol 1. no. 1, May 1967, pp66-67

John Guare (play-writer)

Social networks: Many individuals with diverse social interactions between them

For example: What is Facebook's degree of Separation?

# Social Networks

Mark Zuckerberg
3.17 degrees of separation

Sheryl Sandberg
2.92 degrees of separation

From: https://research.fb.com/three-and-a-half-degrees-of-separation/

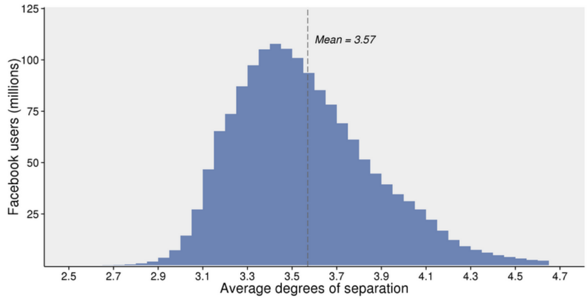



*Figure 1. Estimated average degrees of separation between all people on Facebook. The average person is connected to every other person by an average of 3.57 steps. The majority of people have an average between 3 and 4 steps.*

Facebook Study

2008: on average 4.28 intermediate 'friends' between any two users, i.e., 5.28 hops.

11-2011: on average 3.74 Facebook users in between (4.74 hops)

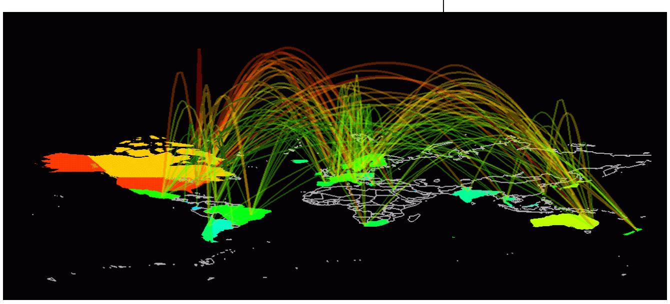
2-2016: on average 3.57 (4.57 hops) (Note there are 1.6 x $10^9$ Facebook users )

5

# Communication Networks

The Earth developed an electronic nervous system, a network with diverse **nodes** and **links** are

| nodes | links |
| --- | --- |
| -computers | -network cables |
| -routers | -TV cables |
| -satellites | -EM waves |

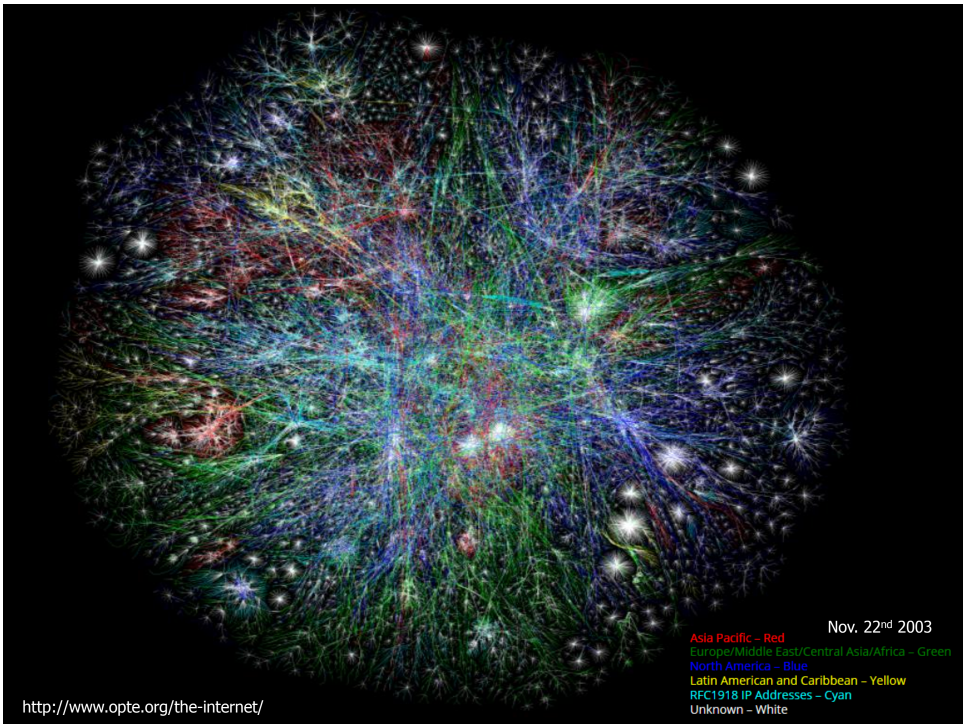


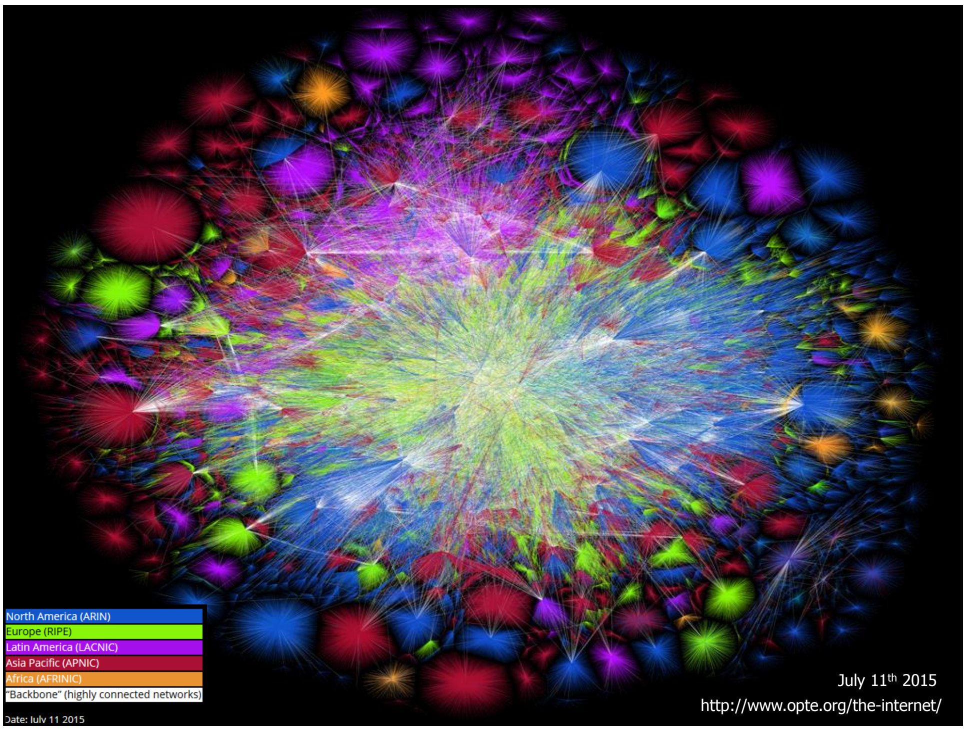
Communication networks: Many non-identical components with diverse connections between them

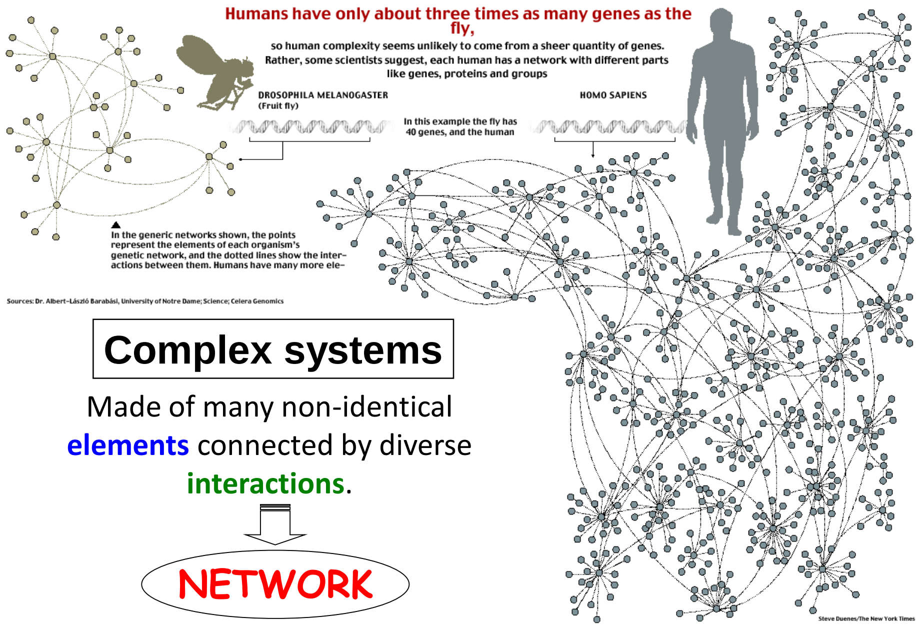
The Internet of Things

Note: the number of IPv4 addresses is $2^{32}$, IPv6 has an address space of size $2^{128}$.

6


Nov. 22nd 2003
Asia Pacific – Red
Europe/Middle East/Central Asia/Africa – Green
North America – Blue
Latin American and Caribbean – Yellow
RFC1918 IP Addresses – Cyan
Unknown – White
http://www.opte.org/the-internet/


North America (ARIN)
Europe (RIPE)
Latin America (LACNIC)
Asia Pacific (APNIC)
Africa (AFRINIC)
"Backbone" (highly connected networks)
Date: July 11 2015
July 11th 2015
http://www.opte.org/the-internet/




**Complex systems**

Made of many non-identical **elements** connected by diverse **interactions**.

NETWORK

Sequencing@home: The Internet of Sequences

9

# "Natural" Networks and Universality

- Consider many kinds of networks:
  - social, technological, business, economic, content,…
- These networks tend to share certain *informal* properties:
  - large scale; continual growth
  - distributed, **organic growth:** vertices "decide" who to link to
  - interaction restricted to links
  - mixture of local and long-distance connections
  - abstract notions of distance: geographical, content, social,…

10

## "Natural" Networks and Universality

- Consider many kinds of networks:
  - social, technological, business, economic, content,…
- *Social network theory and link analysis*
  - Do natural networks share more *quantitative* universals?
  - What would these "universals" be?
  - How can we make them precise and measure them?
  - How can we explain their universality?

11

## Natural Network Analysis

- Natural Networks: Examples
- Primitives for Natural Network Analysis
- Natural Network Characteristics
- Natural Network Generation
- Natural Network Mining

12/4/2018

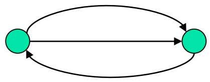
12

## Networks and Their Representations

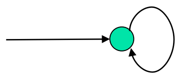
- A network (or a graph):

  G = (V, E), where V: vertices (or nodes), and E: edges (or links)

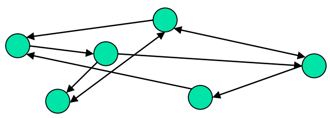
  - Multi-edge: if more than one edge between the same pair of vertices



  - Self-edge (self-loop): if an edge connects vertex to itself

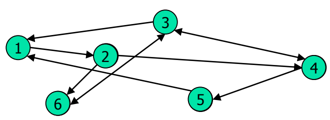


- Simple network/graph if a network has neither self-edges nor multi-edges



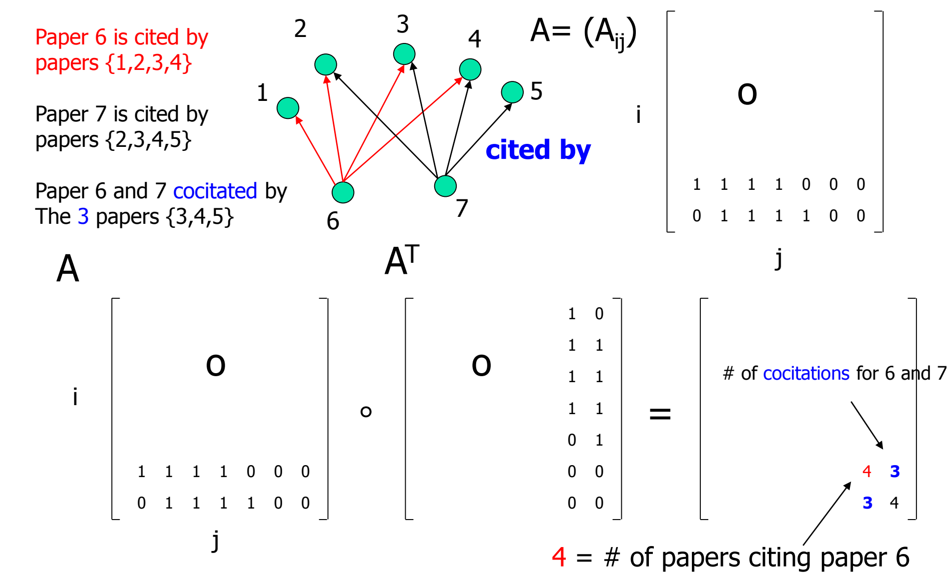
## Networks and Their Representations

- A network (or a graph): G = (V, E), where V: vertices (or nodes), and E: edges (or links)

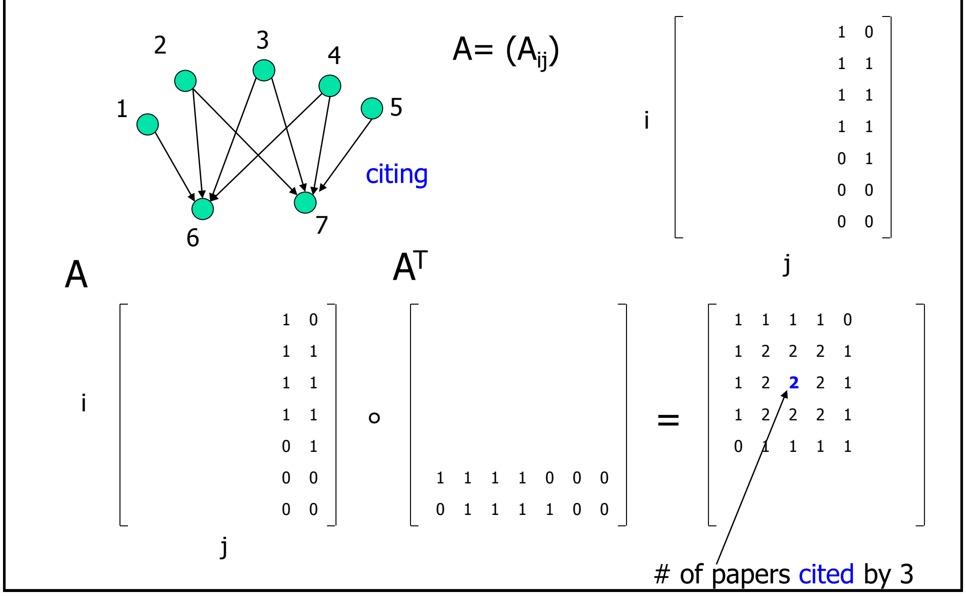



- Adjacency matrix:
  - $A_{ij}$ = 1 if there is an edge between vertices i and j; 0 otherwise
- Weighted networks:
  - Edges having weight (strength), usually a real number
- Directed network (directed graph): if each edge has a direction
  - $A_{ij}$ = 1 if there is an edge from i to j; 0 otherwise

# Cocitation (cited by same papers)




# Bibliographic Coupling



## Cocitation & Bibliographic Coupling: Comparison

- For strong cocitation: must have **a lot of outgoing edges**
  - Must be well-cited => **(influential) papers, surveys, or books**
  - Takes time to accumulate citations
- Strong bib-coupling if two papers have **similar citations**
  - A more uniform indicator of **similarity between papers**
  - Can be computed as soon as a paper is published
  - No change over time
- Analysis algorithms
  - HITS (Hyperlink Induced Topic Search) (J. Kleinberg, 1998) explores both cocitation and bibliographic coupling
  - Current methods use additional full text analysis.

## Degree and Network Density

- Degree of a vertex *i*: $k_i = \sum_{j=1}^{n} A_{ij}$

  where n is the number of vertices

For an undirected graph:

- # of edges m = 1/2 of the sum of degrees of all the vertices:

$$m = \frac{1}{2}\sum_i^n k_i = \frac{1}{2}\sum_{ij} A_{ij}$$

- The mean degree c of a vertex in an undirected graph:

$$c = \frac{1}{n}\sum_i^n k_i = \frac{2m}{n}$$

What is a dense graph?

18

## Degree and Network Density

- Degree of a vertex *i*: $k_i = \sum_{j=1}^{n} A_{ij}$
- # of edges m = 1/2 of sum of degrees of all the vertices:
$$m = \frac{1}{2}\sum_{i}^{n} k_i = \frac{1}{2}\sum_{ij} A_{ij}$$
- The mean degree c of a vertex in an undirected graph:
$$c = \frac{1}{n}\sum_{i}^{n} k_i = \frac{2m}{n}$$
- Density ρ of a graph: $\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)} = \frac{c}{n-1}$
- A network is **dense** if density ρ tends to be a constant as n → ∞
- A network is **sparse** if density ρ → 0 as n → ∞. The fraction of nonzero element in the adjacency matrix tends to zero
- Internet, WWW and friendship networks are usually regarded as sparse



## Natural Network Analysis

- Natural Networks: Examples
- Primitives for Natural Network Analysis
- Natural Network Characteristics
- Natural Network Generation
- Natural Network Mining

## Some Interesting Network Quantities

- *Connected components:*
  - how many, and how large?
- *Network diameter:*
  - maximum (worst-case) or average?
  - exclude infinite distances? (disconnected components)
  - the small-world phenomenon
- *Clustering:*
  - to what extent links tend to cluster "locally"?
  - what is the balance between local and long-distance connections?
  - what roles do the two types of links play?
- *Degree distribution:*
  - what is the typical degree in the network?
  - what is the overall distribution?

21

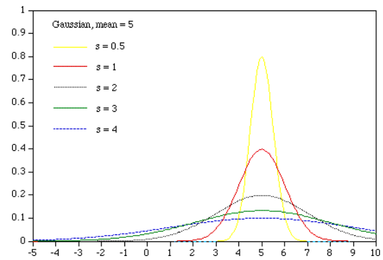
## Natural Network Characteristics

- *Few* connected components:
  - often only 1 or a small number, indep. of network size
- *Small* diameter:
  - often a constant independent of network size (like 6, 3.57)
  - or perhaps growing only logarithmically with network size or even shrink? 4.28 -> 3.74 -> 3.57 (4 years in between)
  - typically exclude infinite distances
- A *high* degree of clustering:
  - considerably more so than for a random network
  - in tension with small diameter
- A *heavy-tailed* degree distribution:
  - a small but reliable number of high-degree vertices
  - often of *power law* form (random variable X assuming integer values (> 0) probability of value $x \sim 1/x^a$ (typically $0 < a < 2$) ( … See next slides)

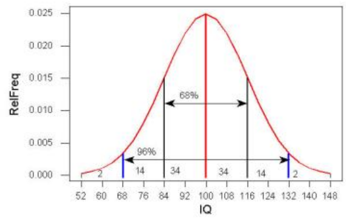
22

# The Normal Distribution

- The *normal* or *Gaussian* density:
- characterized by mean m and standard deviation s
- *density* at x is defined as
  - $(1/(s\ \mathrm{sqrt}(2pi)))\ \exp(-(x-m)^2/2s^2)$
  - special case m = 0, s = 1: a $\exp(-x^2/b)$ for some constants a,b > 0
- peaks at x = m, then dies off *exponentially* rapidly




- the classic "bell-shaped curve"
  - exam scores, human body temperature
- remarks:
  - can control mean and standard deviation independently
  - can make as "broad" as we like, but always have *finite variance*

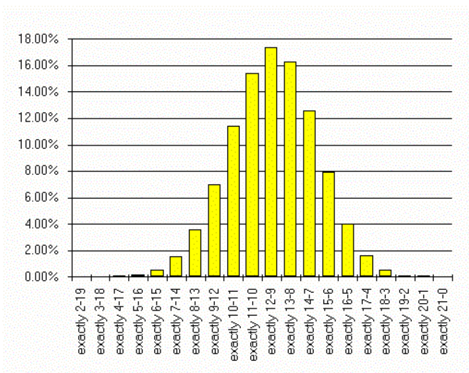





# The Binomial Distribution

- Coin with Pr[heads] = p, flip n times, probability of getting exactly k heads:
  - choose (n, k) = $p^k(1-p)^{n-k}$
- For large n and p *fixed:*
  - approximated well by a normal with

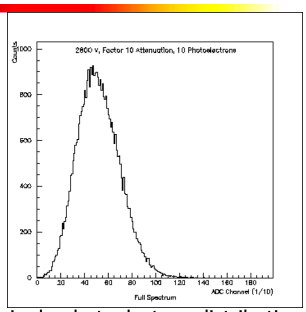
    m = np, s = sqrt(np(1-p))
  - s/m → 0 as n grows
  - leads to strong large deviation bounds



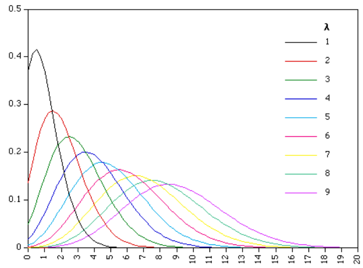

[www.professionalgambler.com/binomial.html](www.professionalgambler.com/binomial.html)

# The Poisson Distribution

- Like binomial, applies to variables taken on integer values > 0
- Often used to model *counts* of events
  - number of phone calls placed in a given time period
  - number of times a neuron fires in a given time period
- Single free parameter l, probability of exactly x events:
  - exp(-l) $l^x$/x!
  - mean and variance are both l
- Binomial distribution with n large, p = l/n (l fixed)
  - converges to Poisson with mean l

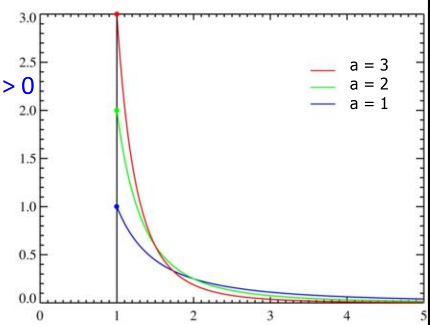



single photoelectron distribution




# Power Law (or Pareto) Distributions

- Heavy-tailed, pareto, or *power law* distributions:
  - For variables assuming integer values > 0
  - probability of value x ~ $1/x^a$
  - Typically 0 < a < 2; smaller a gives heavier tail
  - sometimes also referred to as being *scale-free*
- Note: for binomial, normal, and Poisson distributions the tail probabilities approach 0 *exponentially* fast
- **What kind of phenomena does this distribution model?**
- **What kind of process would *generate* it?**




26

# Natural Network Analysis

- Natural Networks: Examples
- Primitives for Natural Network Analysis
- Natural Network Characteristics
- Natural Network Generation
- Natural Network Mining



# Probabilistic Models of Networks

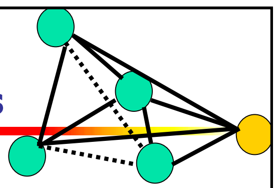
- All of the network generation models we will study are *probabilistic* or *statistical* in nature
- They can generate networks of any size
- They often have various *parameters* that can be set:
  - size of network generated
  - average degree of a vertex
  - fraction of long-distance connections
- The models generate a *distribution* over networks
- Statements are always *statistical* in nature:
  - *with high probability*, diameter is small
  - on average, degree distribution has heavy tail

## Some Models of Network Generation

- *Random graphs (Erdös-Rényi models, 1959):*
  - gives few components and small diameter
  - does not give high clustering and heavy-tailed degree distributions
  - is the mathematically most well-studied and understood model
- *Watts-Strogatz models (1998):*
  - give few components, small diameter and high clustering
  - does not give heavy-tailed degree distributions
- *Scale-free Networks (1965, 1976, … 1999):*
  - gives few components, small diameter and heavy-tailed distribution
  - does not give high clustering
- *Hierarchical networks:*
  - few components, small diameter, high clustering, heavy-tailed
- *Affiliation networks:*
  - models group-actor formation

29

## Degrees and Clustering Coefficients



- Let a network G = (V, E), degree of a vertex
- Undirected network: $d(v_i)$: $d(v_i) = |v_j| \; s.t. \; e_{ij} \in E \wedge e_{ij} = e_{ji}$
- Directed network
  - In-degree of a vertex $d_{in}(v_i)$: $d_{in}(v_i) = |v_j| \; s.t. \; e_{ji} \in E$
  - Out-degree of a vertex $d_{out}(v_i)$: $d_{out}(v_i) = |v_j| \; s.t. \; e_{ij} \in E$
- Clustering coefficients
  - Let $N_v$ be the set of adjacent vertices of v, $k_v$ be the number of adjacent vertices to node v
  - Local clustering coefficient for directed network

$$C_v = \frac{|e_{ij}|}{k_v(k_v-1)} \; s.t. \; v_i, v_j \in N_v, e_{ij} \in E$$

  - Local clustering coefficient for undirected network

$$C_v = \frac{2|e_{ij}|}{k_v(k_v-1)} \; s.t. \; v_i, v_j \in N_v, e_{ij} \in E$$

  - For the whole network: Averaging the local clustering coefficient of all the vertices (Watts & Strogatz): $C = \frac{1}{|V|} \Sigma_{v \in V} C_v$

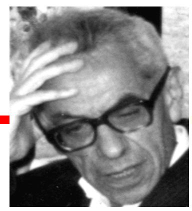
30

# The Erdös-Rényi (ER) Model: A Random Graph Model

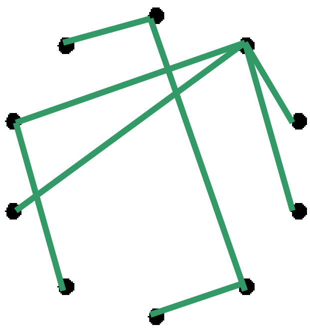
- A random graph is obtained by starting with a set of *N* vertices and adding edges between them at random
- Different *random graph models* produce different *probability distributions* on graphs
- Most commonly studied is the *Erdős–Rényi model*, denoted *G(N,p)*, in which **every possible edge occurs independently with probability *p***
- Random graphs were first defined by Paul Erdős and Alfréd Rényi in their 1959 paper "On Random Graphs"
- The usual *regime of interest* is when *p ~ 1/N, N is large*
  - e.g., p = 1/2N, p = 1/N, p = 2/N, p=10/N, p = log(N)/N, etc.
  - in expectation, each vertex will have a "small" number of neighbors
  - will then examine what happens when N → infinity
  - can thus study properties of *large networks* with *bounded degree*
  - Sharply concentrated; *not* heavy-tailed

31

# Erdös-Rényi Model (1959)



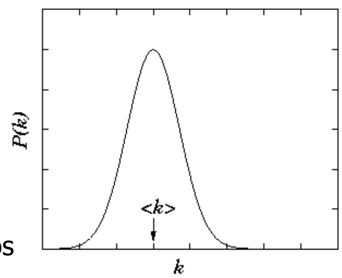
Pál Erdős
(1913-1996)



**Connect with probability p**

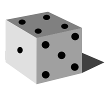
p=1/6
N=10
⟨k⟩~1.5

Poisson distribution




**- Democratic**

**- Random**



Btw.: My Erdös Number = 3
E.M. Bakker – J. Van Leeuwen – S. Zaks – P. Erdös

## Erdös-Rényi Model (1959)

## The First, Well-Studied

## Random Graph Model

## BUT

## Not a Good Model for Natural Networks!



# The Watts and Strogatz Model

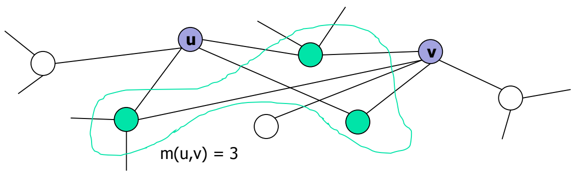
- Proposed by Duncan J. Watts, Steven Strogatz in their **1998 Nature paper**
- A random graph generation model that produces graphs with ***small-world properties***, including **short average path lengths** and **high clustering**
- Known as the (Watts) *beta* model after Watts used β to formulate it in his popular science book *Six Degrees*

- The Erdos-Renyi graphs fail to explain two important properties observed in real-world networks:
  - Do not account for local clustering, i.e., having a low *clustering coefficient* as a result of assuming a constant and independent probability of two nodes being connected
  - Do not account for the formation of hubs. Formally, the degree distribution of Erdos-Renyi graphs converges to a Poisson distribution, rather than a *power law* observed in most real-world, *scale-free networks*

# The $\alpha$-model: Propensity

- For any vertices u and v:
  - m(u,v) equal to the number of common neighbors (so far)
- Key quantity: the *propensity* R(u,v) of u to connect to v

```
if m(u,v) >= k then      // parameter k average degree
    R(u,v) = 1           // u,v share too many friends not to be connected
if m(u,v) = 0 then
    R(u,v) = p           // no mutual friends → no bias to connect)
else
    R(u,v) = p + (m(u,v)/k)^α (1-p)      // probability has m(u,v) as bias
```




35

# The $\alpha$-model

- The $\alpha$-model has the following parameters:
  - N: *size* of the network to be generated
  - k: the *average degree* of a vertex in the network to be generated
  - p: the *default probability* that two vertices are connected
  - $\alpha$: adjustable parameter dictating bias towards local connections
- For any vertices u and v:
  - define m(u,v) to be the number of common neighbors (so far)
- Key quantity: the *propensity* R(u,v) of u to connect to v
  - if m(u,v) >= k, R(u,v) = 1 (u,v share too many friends *not* to be connected)
  - if m(u,v) = 0, R(u,v) = p (no mutual friends → no bias to connect)
  - else, $R(u,v) = p + (m(u,v)/k)^{\alpha} (1-p)$
- Generate new edges incrementally
  - using R(u,v) as the edge probability; details omitted
- Note: $\alpha$ = infinity is “like” Erdos-Renyi (but not exactly)

36

## The $\alpha$-model
## Small Worlds and Occam's Razor

- For small $\alpha$, should generate large clustering coefficients
  - we "programmed" the model to do so
  - Watts claims that proving precise statements is hard…
- But we do *not* want a new model for every little property
  - Erdos-Renyi → small diameter
  - $\alpha$-model → high clustering coefficient
- In the interests of *Occam's Razor*, we would like to find
  - a *single, simple* model of network generation…
  - … that *simultaneously* captures *many* properties

=> Watt's $\beta$-Model, small world: small diameter *and* high clustering

37

## Watts $\beta$-Model Discovered by Examining the Real World…

- Watts examines three real networks as case studies:
  - the **Kevin Bacon** graph
  - the Western states power grid
  - the C. elegans nervous system
- For each of these networks, he:
  - computes its size, diameter, and clustering coefficient
  - compares diameter and clustering to *best* Erdos-Renyi approx.
  - shows that the *best* $\alpha$-model approximation is better
  - important to be "fair" to each model by finding best fit
- Overall:
  - if we care only about diameter and clustering:
    $\alpha$ is better than p

38

# Case 1: Kevin Bacon Graph

- Vertices: actors and actresses
- Edge between u and v if they appeared in a film together

| Kevin Bacon |
|---|
| No. of movies : 46<br>No. of actors : 1811<br>Average separation: 2.79 |

*Is Kevin Bacon the most connected actor?*

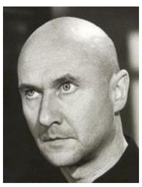
***NO!***

| Rank | Name | Average distance | # of movies | # of links |
|---|---|---|---|---|
| 1 | Rod Steiger | 2.537527 | 112 | 2562 |
| 2 | Donald Pleasence | 2.542376 | 180 | 2874 |
| 3 | Martin Sheen | 2.551210 | 136 | 3501 |
| 4 | Christopher Lee | 2.552497 | 201 | 2993 |
| 5 | Robert Mitchum | 2.557181 | 136 | 2905 |
| 6 | Charlton Heston | 2.566284 | 104 | 2552 |
| 7 | Eddie Albert | 2.567036 | 112 | 3333 |
| 8 | Robert Vaughn | 2.570193 | 126 | 2761 |
| 9 | Donald Sutherland | 2.577880 | 107 | 2865 |
| 10 | John Gielgud | 2.578980 | 122 | 2942 |
| 11 | Anthony Quinn | 2.579750 | 146 | 2978 |
| 12 | James Earl Jones | 2.584440 | 112 | 3787 |
| … | | | | |
| **876** | **Kevin Bacon** | **2.786981** | **46** | **1811** |
| … | | | | |



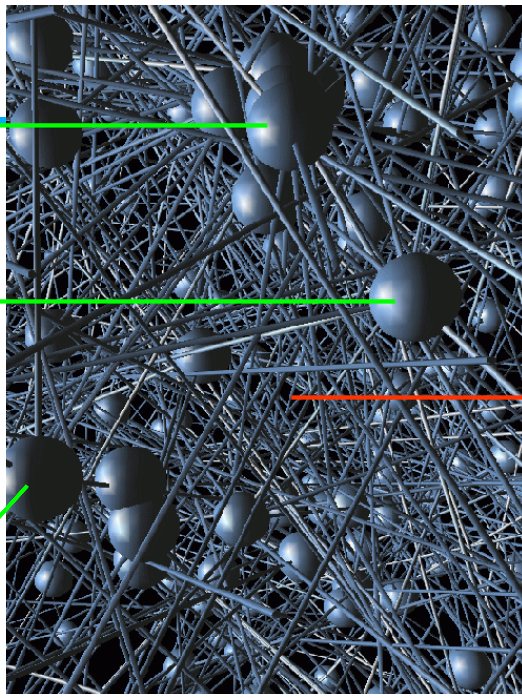
# Bacon -map

#1 Rod Steiger



#2 Donald Pleasence

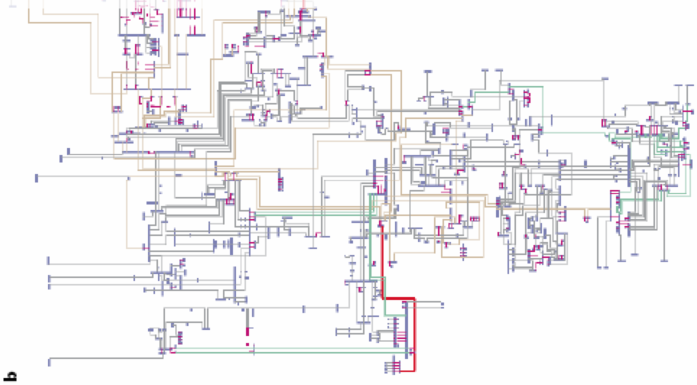
#3 Martin Sheen



#876 Kevin Bacon

# Case 2: New York State Power Grid

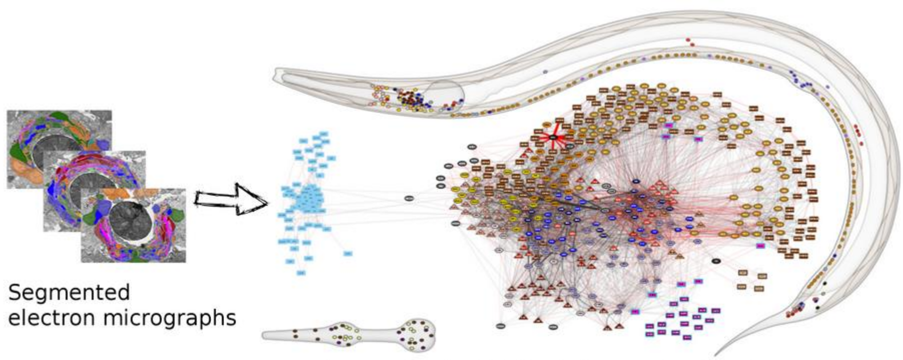
- Vertices: generators and substations
- Edges: high-voltage power transmission lines and transformers
- Line thickness and color indicate the voltage level
  - Red 765 kV, 500 kV; brown 345 kV; green 230 kV; grey 138 kV



41

# Case 3: C. Elegans Nervous System

- Vertices: neurons in the C. elegans worm
- Edges: axons/synapses between neurons




From: http://wormwiring.org/ Reconstructed biological neural network

42

# Two More Examples

- M. Newman on **scientific collaboration networks**
  - coauthorship networks in several communities
  - differences in degrees (papers per author)
  - empirical verification of
    - giant components
    - small diameter (mean distance)
    - high clustering coefficient
- Alberich et al. on the **Marvel Universe**
  - *purely fictional* social network
  - two characters linked if they appeared together in an issue
  - "empirical" verification of
    - heavy-tailed distribution of degrees (issues and characters)
    - giant component
    - rather *small* clustering coefficient

43

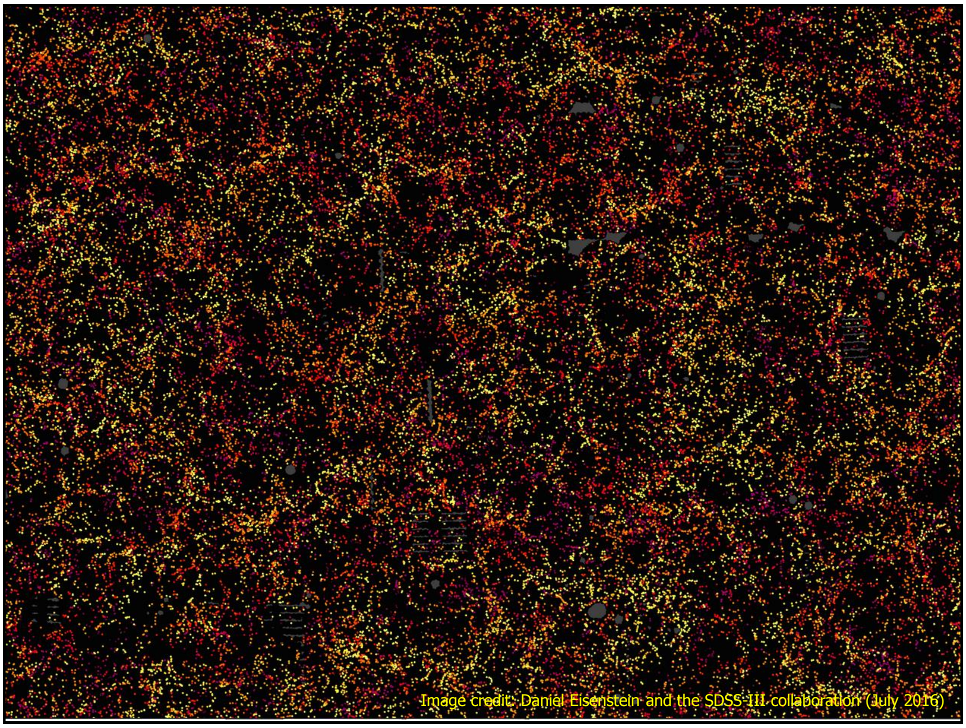


Image credit: Daniel Eisenstein and the SDSS-III collaboration (July 2016)

# Network Cosmology

**Network Cosmology**

Dmitri Krioukov, Maksim Kitsak, Robert S. Sinkovits, David Rideout, David Meyer3 and Marian Boguna

Prediction and control of the dynamics of complex networks is a central problem in network science. Structural and dynamical similarities of different real networks suggest that some universal laws might accurately describe the dynamics of these networks, albeit the nature and common origin of such laws remain elusive. Here we show that the causal network representing **the large scale structure of space-time in our accelerating universe** is a **power-law graph** with **strong clustering**, similar to many complex networks ***such as the Internet, social, or biological networks***. We prove that this structural similarity is a consequence of the asymptotic equivalence between the large scale growth dynamics of complex networks and causal networks. This equivalence suggests that unexpectedly similar laws govern the dynamics of complex networks and space-time in the universe, with implications to network science and cosmology.

arXiv: 1203.2109v2 (November 2012)

**Image on previous slide**: The image shows 48,741 galaxies, which is about 3% of the full survey dataset. It covers ~t 1/20th of the sky with a volume of 6 billion light-years (w) x 4.5 billion light-years (h) high x 500 million light-years (d). Color ranges from yellow to purple, where yellow is closest to earth (Sloan Digital Sky Survey III (SDSS-III), BOSS).

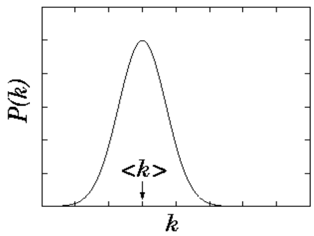


# Towards Improvements on Watts-Strogatz Model: Scale-Free Networks

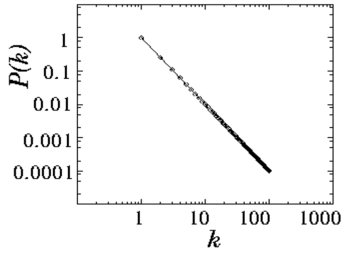
- The **Watts-Strogatz model** thus far:
  - Gives few components, small diameter and high clustering
  - It produces graphs that are homogeneous in degree, hence still do not exactly follow the heavy-tailed degree distributions
  - The Watts-Strogatz model also implies a fixed number of nodes and thus cannot be used to model network growth
- Proposal new model: **Scale-Free Networks**:
  - Real networks are often *scale-free networks* inhomogeneous in degree, having hubs and a scale-free degree distribution.
    - Such networks are better described by the *preferential attachment family of models*, such as the *Barabási–Albert (BA) model*
  - Degree distribution follows a ***power law***, at least asymptotically.

## What Does that Mean?

Poisson distribution

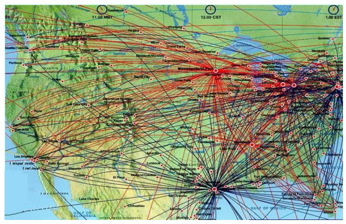



Power-law distribution





**Exponential Network**



**Scale-free Network**



## Scale-Free Networks

- The number of nodes (N) is not fixed
  - Networks continuously expand by additional new nodes
    - WWW: addition of new nodes
    - Citation: publication of new papers
- The attachment is not uniform
  - A node is linked with higher probability to a node that already has a large number of links
    - WWW: new documents link to well known sites (CNN, Yahoo, Google)
    - Citation: Well cited papers are more likely to be cited again

## Scale-Free Networks

- Start with (say) two vertices connected by an edge
- For i = 3 to N:
  - for each $1 <= j < i$
    d(j) is degree of vertex j so far
  - let $Z = \Sigma_j d(j)$ (sum of all degrees so far)
  - add new vertex i with k edges back to {1, …, i-1}:
    - i is connected back to j with probability $d(j)/Z$
- Vertices j with high degree are likely to get more links! —"Rich get richer"

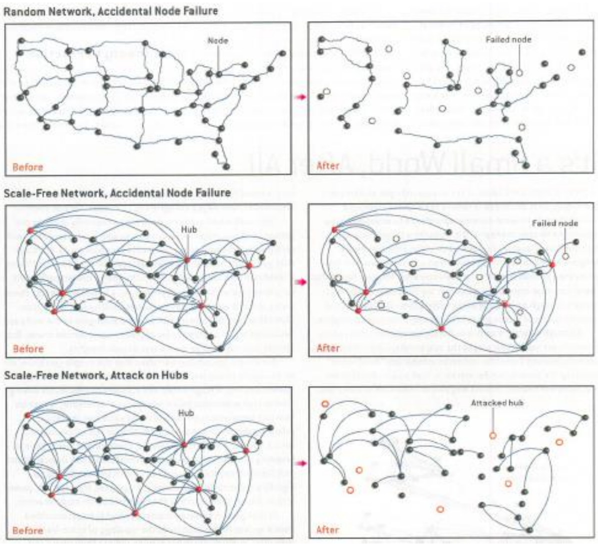


## Scale-Free Networks

- Start with (say) two vertices connected by an edge
- For i = 3 to N:
  - for each $1 <= j < i$, d(j) = degree of vertex j so far
  - let $Z = Sum_j d(j)$ (sum of all degrees so far)
  - add new vertex i with k edges back to {1, …, i-1}:
    - i is connected back to j with probability $d(j)/Z$
- Vertices j with high degree are likely to get more links! —"Rich get richer"

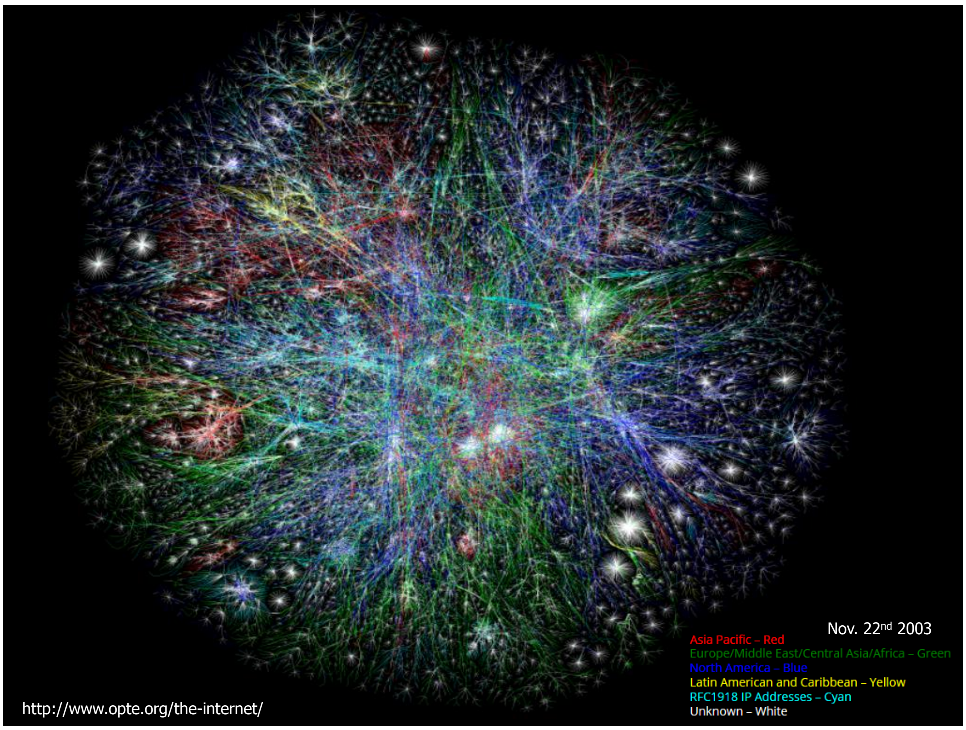
- Natural model for many processes:
  - hyperlinks on the web
  - new business and social contacts
  - transportation networks
- Generates a power law distribution of degrees
  - exponent depends on value of k
- Preferential attachment explains
  - heavy-tailed degree distributions
  - small diameter (~log(N), via "hubs")
- Will *not* generate high clustering coefficient
  - no bias towards local connectivity, but towards hubs

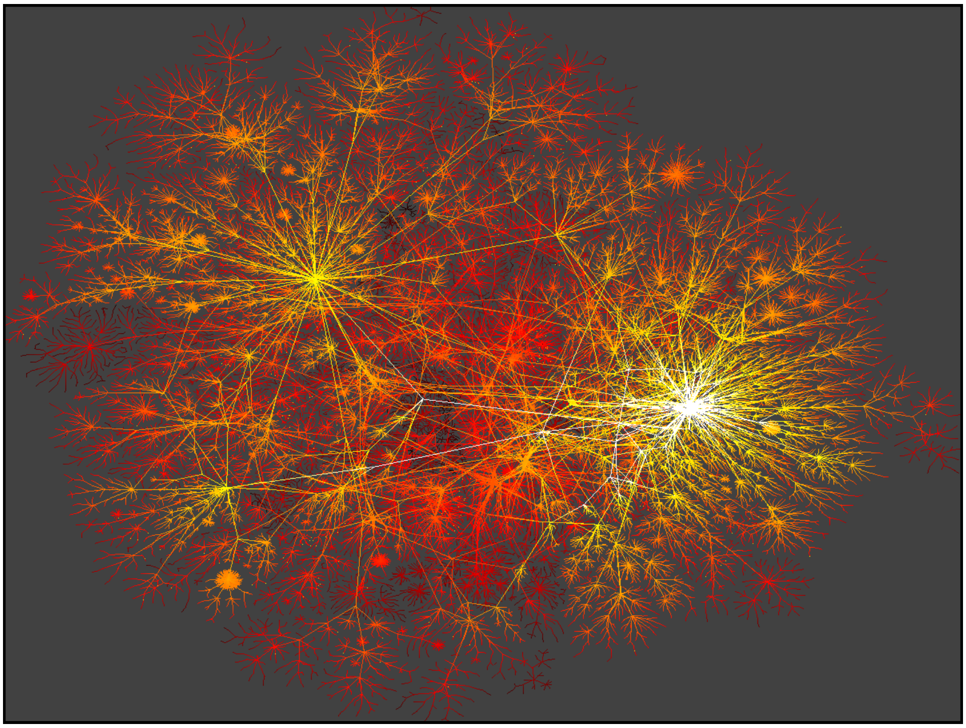
# Robustness of Random vs. Scale-Free Networks

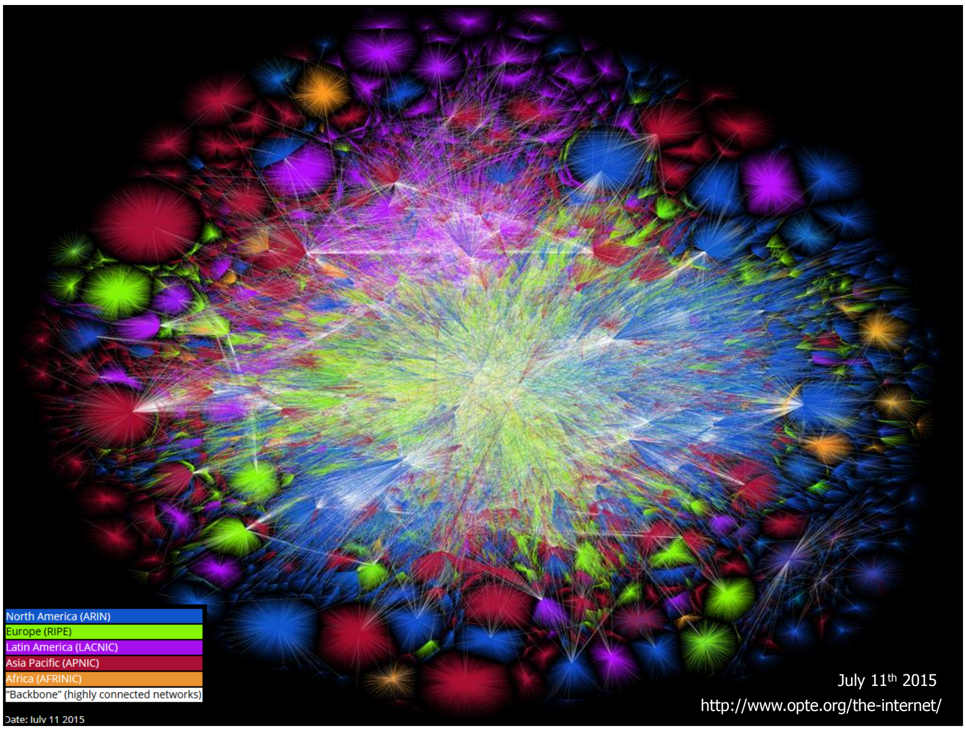



- The accidental failure of a number of nodes in a random network can fracture the system into non-communicating islands.
- Scale-free networks are more robust in the face of such failures
- Scale-free networks are highly vulnerable to a coordinated attack against their hubs








North America (ARIN)
Europe (RIPE)
Latin America (LACNIC)
Asia Pacific (APNIC)
Africa (AFRINIC)
"Backbone" (highly connected networks)
Date: July 11 2015
July 11th 2015
http://www.opte.org/the-internet/

# References


- D. Liben-Nowell and J. Kleinberg. The Link Prediction Problem for Social Networks. CIKM'03
- P. Domingos and M. Richardson, Mining the Network Value of Customers. KDD'01
- M. Richardson and P. Domingos, Mining Knowledge-Sharing Sites for Viral Marketing. KDD'02
- D. Kempe, J. Kleinberg, and E. Tardos, Maximizing the Spread of Influence through a Social Network. KDD'03.
- P. Domingos, Mining Social Networks for Viral Marketing. IEEE Intelligent Systems, 20(1), 80-82, 2005.
- S. Brin and L. Page, The anatomy of a large scale hypertextual Web search engine. WWW7.
- S. Chakrabarti, B. Dom, D. Gibson, J. Kleinberg, S.R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, Mining the link structure of the World Wide Web. IEEE Computer'99
- D. Cai, X. He, J. Wen, and W. Ma, Block-level Link Analysis. SIGIR'2004.
- Lecture notes from Lise Getoor's website: www.cs.umd.edu/~getoor/