A Quick Introduction to Network Science for Security Guys

2nd Division, 2nd R&D Institute,
Agency for Defense Development
Donghwan Lee
nomalee@gmail.com

Mafia Boy & Project Rivolta

- February 2000, large commercial websites including Yahoo, FIFA, Amazon, Dell, E*TRADE, eBay, and CNN were shutdown

- Suspect revealed as a high-school student, attacked listed sites with basic knowledges & techniques, caused 1.2 billion $ in global economic damage

Science

- “Science should make the wonderful and complex understandable and simple but not less wonderful” - Herbert A. Simon

- “Science is built of facts the way a house is built of bricks; but an accumulation of facts is no more science than a pile of bricks is a house” - Henri Poincaré

Pareto law

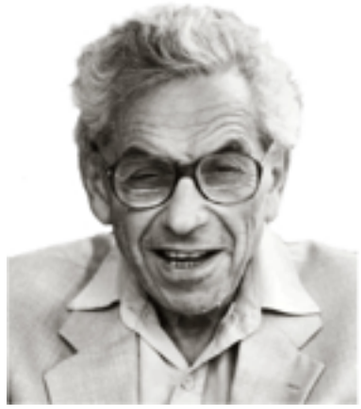
- 80% of results come from 20% sources

Small-World Experiment

- Frigyes Karinthy, "*Láncszemek*," in *Minden másképpen van*, 1929.
- "We could name any person among earth's one and a half billion inhabitants and through *at most five acquaintances*, one of which he knew personally, he could link to the chosen one"

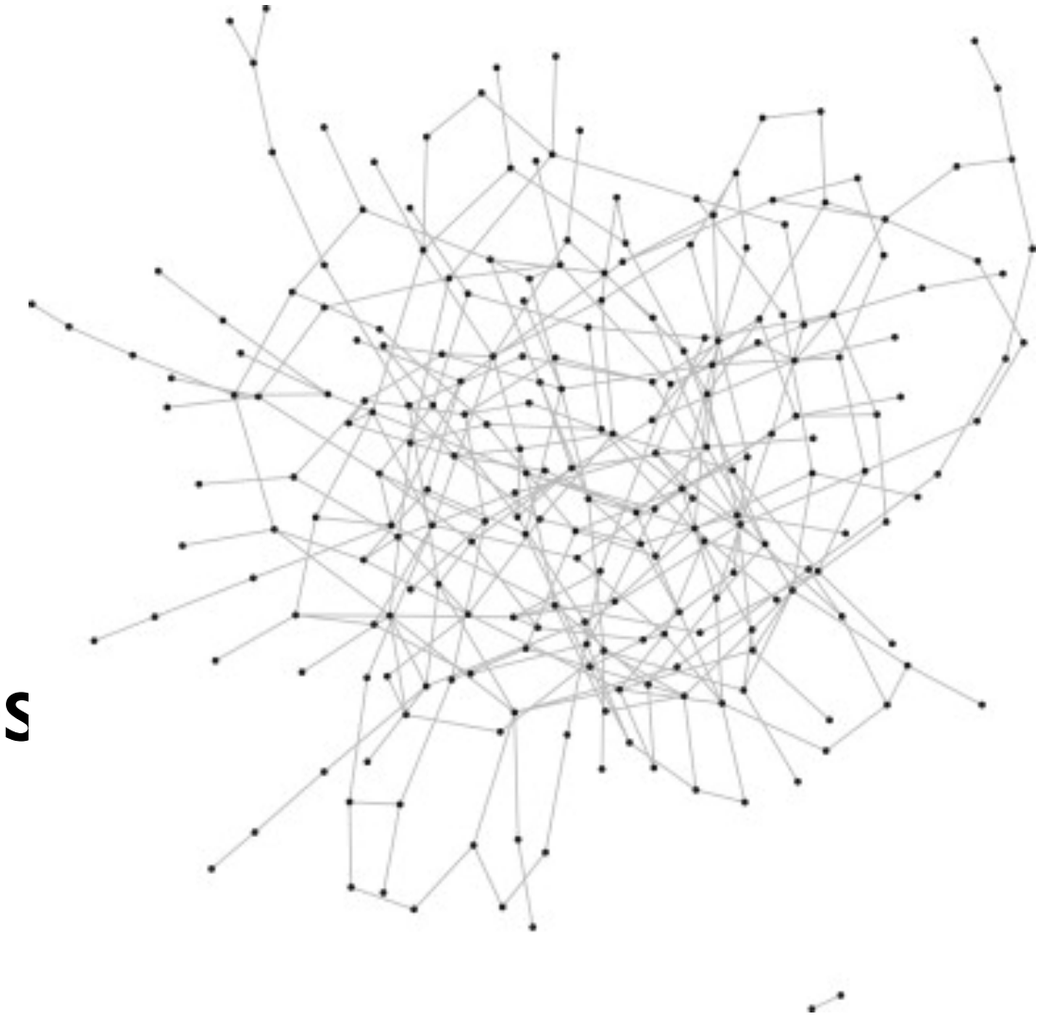
6 Degrees of Seperation

- Milgram's Experiment: counting the number of ties between any two people
- Result: people in the United States are separated by about six people on average
- Jeffrey Travers & Stanley Milgram. 1969. "An Experimental Study of the Small World Problem." Sociometry, Vol. 32, No. 4, pp. 425-443.



Random Tie

- Erdős–Rényi Model: random graph (1960)
- As the network gets larger, the number of required random edges dramatically decrease

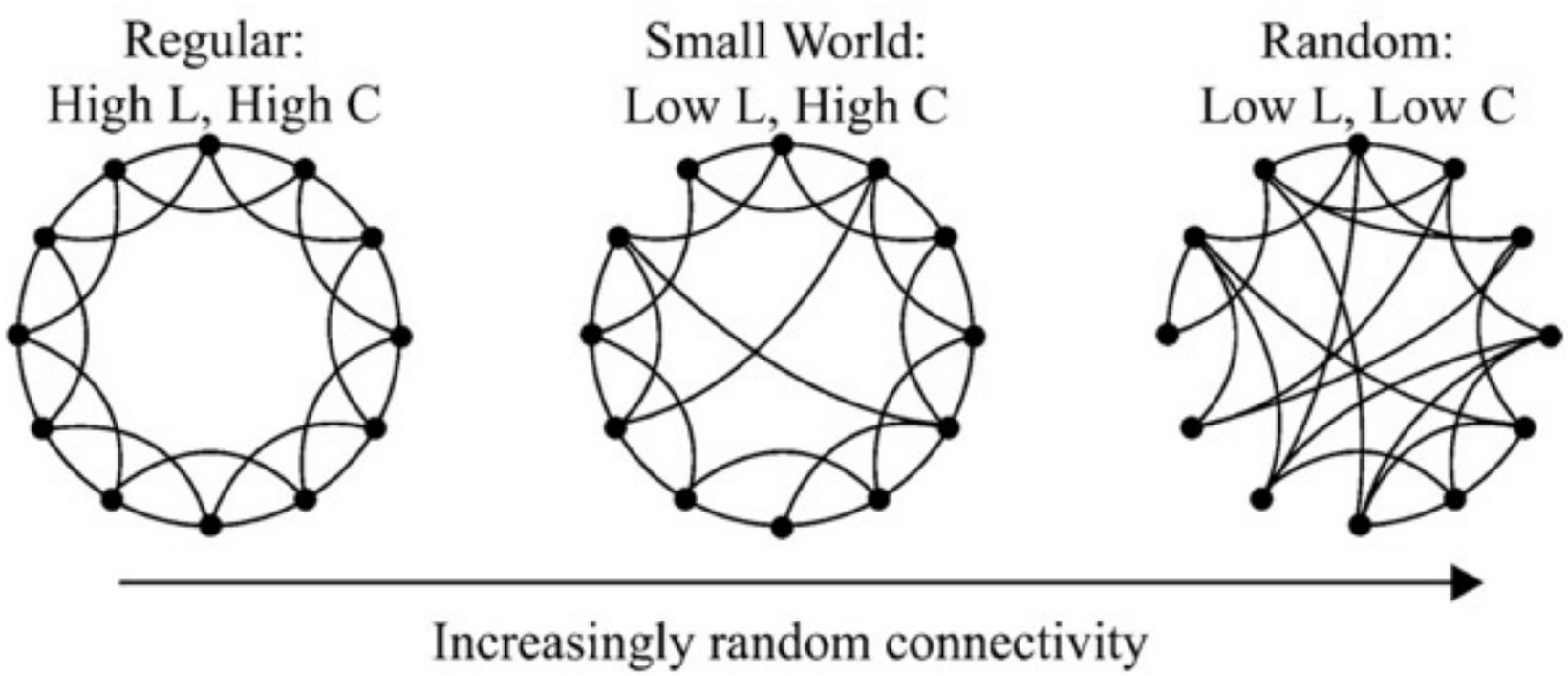


Weak Tie

- From whom do people often get jobs?
- Mark Granovetter, "The Strength of Weak Ties," Amercican Journal of Sociology, Vol. 78, Issue 6, 1360-1380, May 1973.
- Capability-based Security

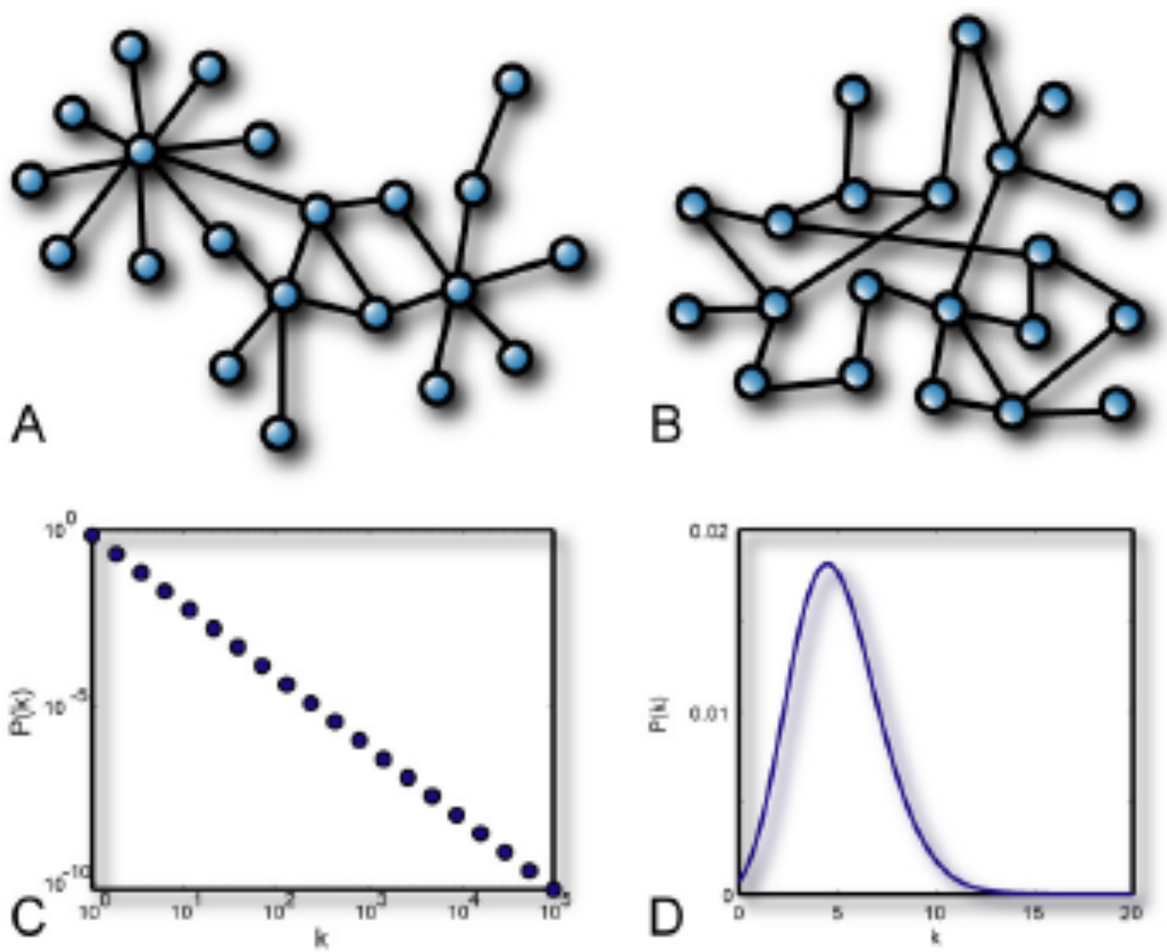
Small World: Watts-Strogatz

- D. J. Watts & S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature* vol. 393, no. 6684, pp. 409–410, 1988.



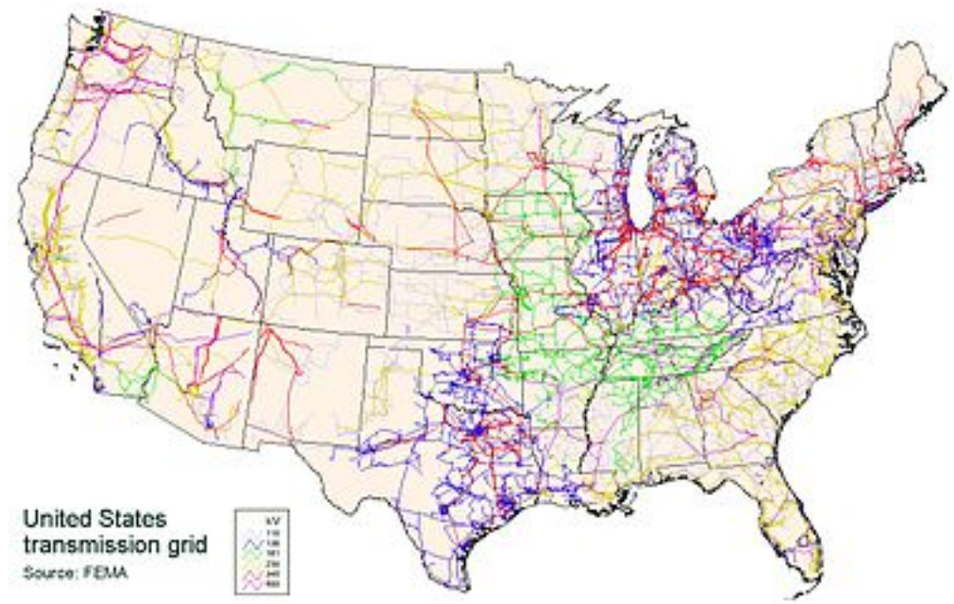
Scale-free Network

- A scale-free network is a network whose degree distribution follows the power law

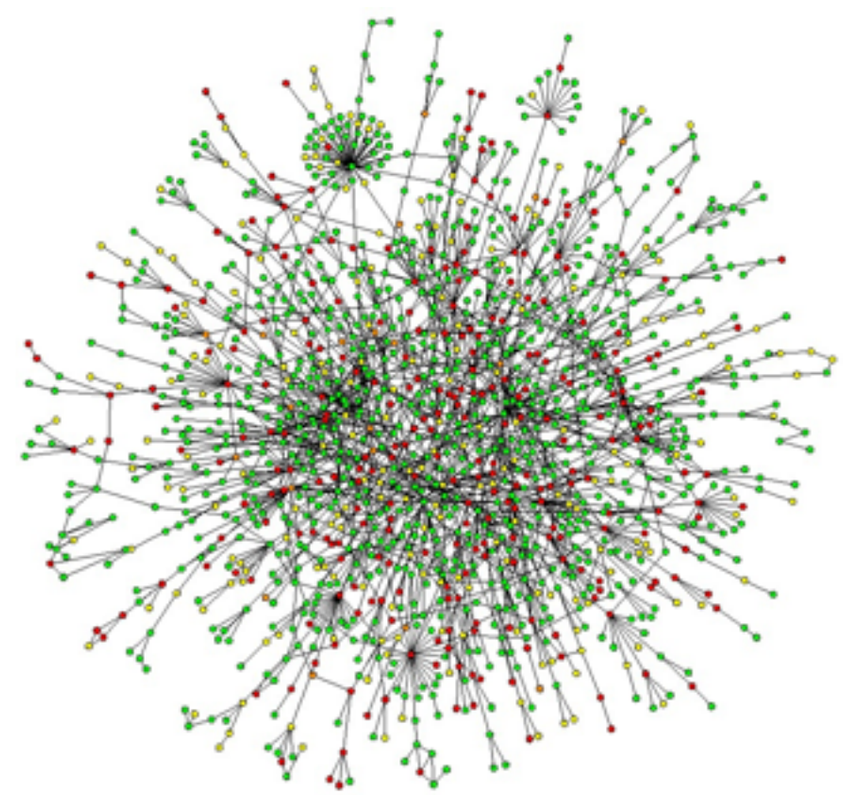


Scale-free Examples

- Power-grid
- Transporta-tion networks

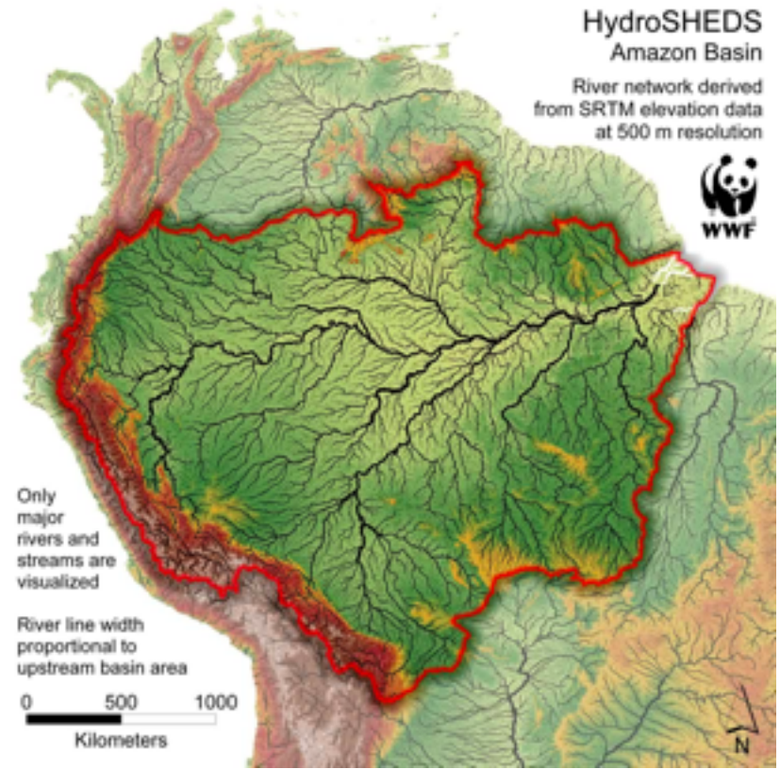


- Metabolic networks
- Food circle
- Worm epidemic



Scale-free Examples

- River networks



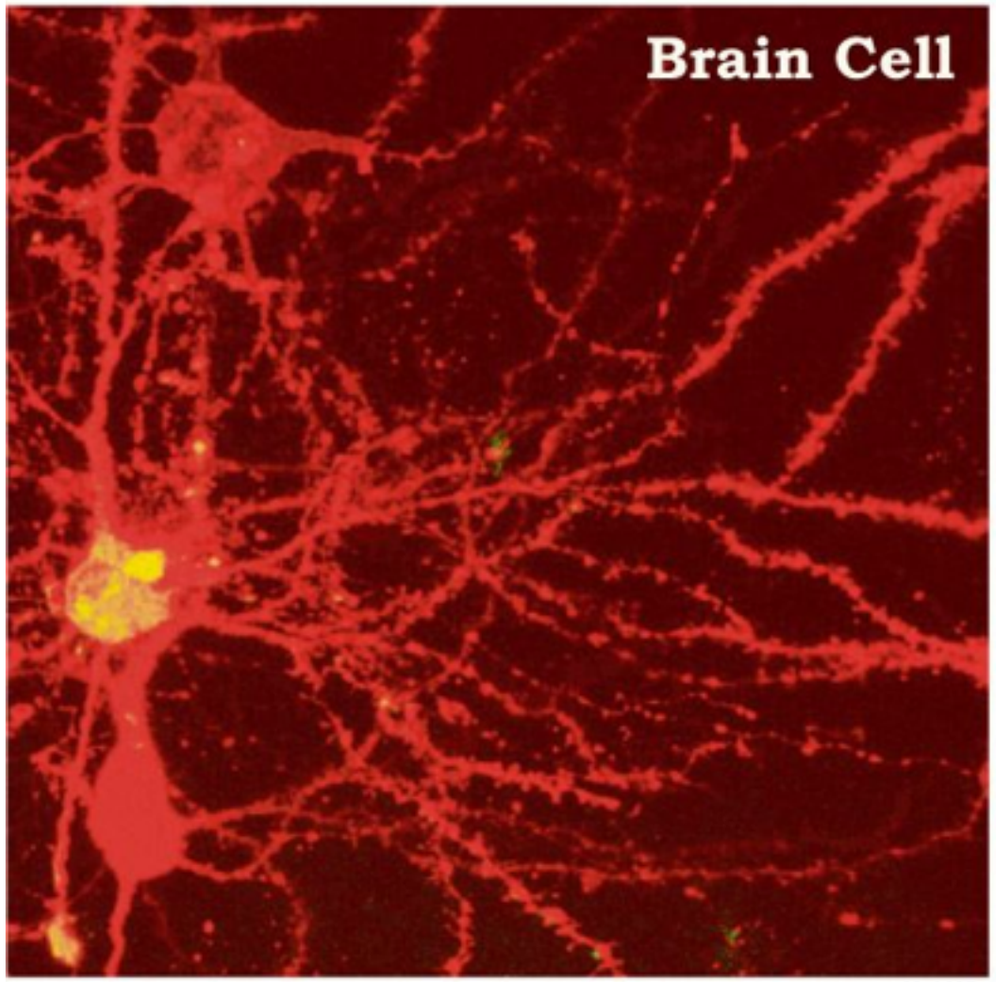
- Synchronization networks of fireflies or hand claps

Scale-free Examples

- Social Networks!

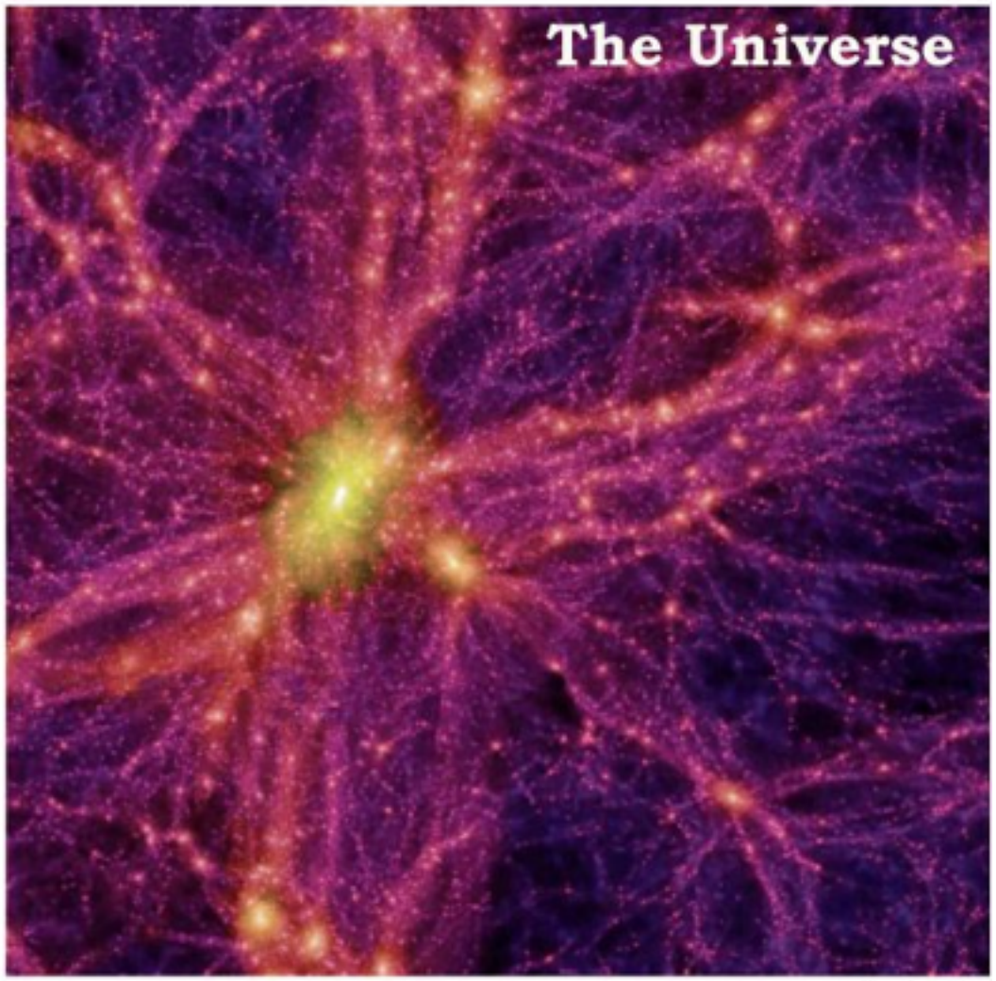
Brain vs. The Universe

One is only micrometers wide. The other is billions of light-years across. One shows neurons in a mouse brain. The other is a simulated image of the universe. Together they suggest the surprisingly similar patterns found in vastly different natural phenomena. *DAVID CONSTANTINE*



Mark Miller

Mark Miller, a doctoral student at Brandeis University, is researching how particular types of neurons in the brain are connected to one another. By staining thin slices of a mouse's brain, he can identify the connections visually. The image above shows three neuron cells on the left (two red and one yellow) and their connections.



Virgo Consortium

An international group of astrophysicists used a computer simulation last year to recreate how the universe grew and evolved. The simulation image above is a snapshot of the present universe that features a large cluster of galaxies (bright yellow) surrounded by thousands of stars, galaxies and dark matter (web).

Source: Mark Miller, Brandeis University; Virgo Consortium for Cosmological Supercomputer Simulations; www.visualcomplexity.com

The New York Times

Internet: revisited

- Paul Baran: father of ARPANET, designed survivable architecture of Internet

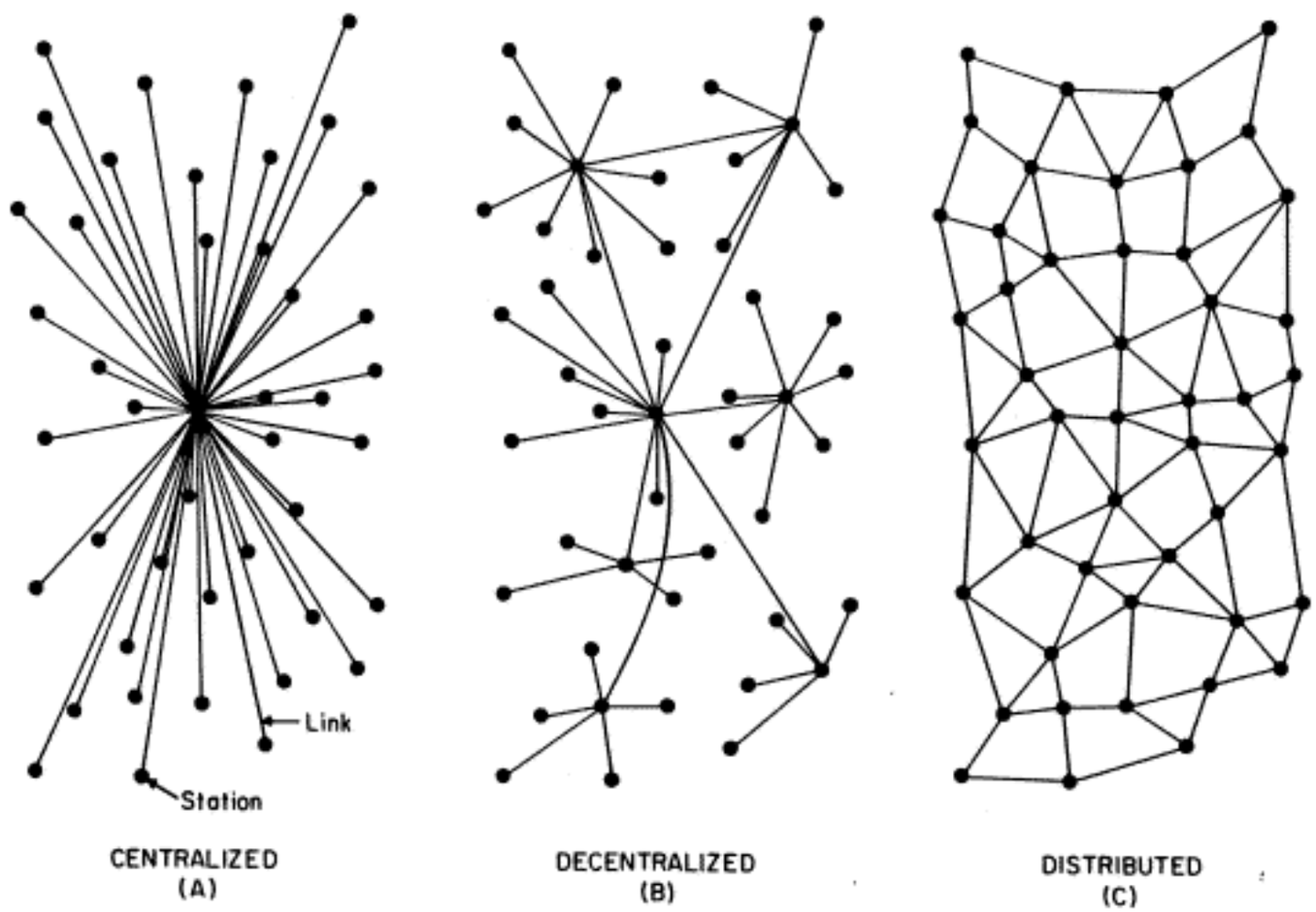
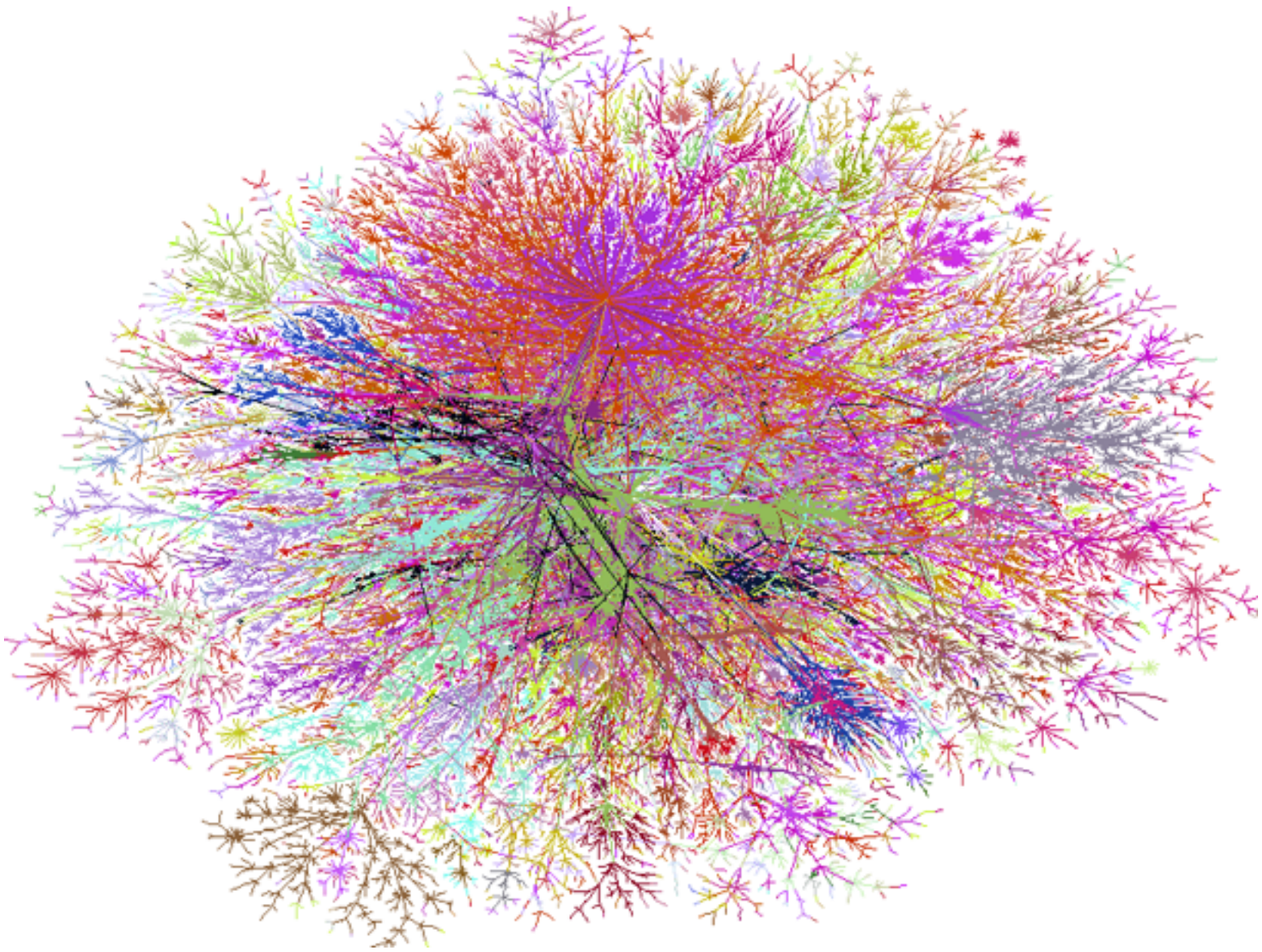


FIG. 1 — Centralized, Decentralized and Distributed Networks

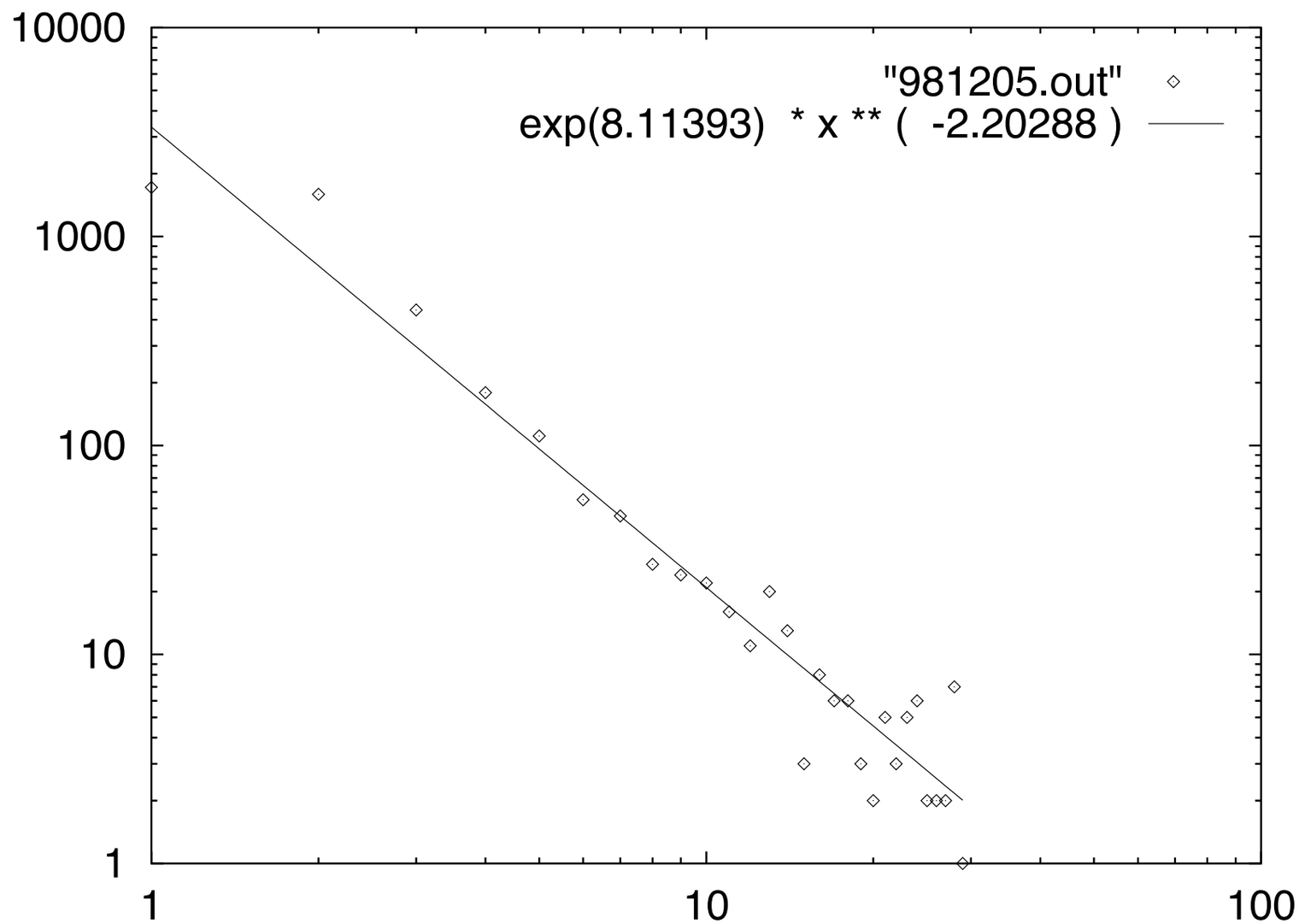
Internet: revisited

- But the result shows..

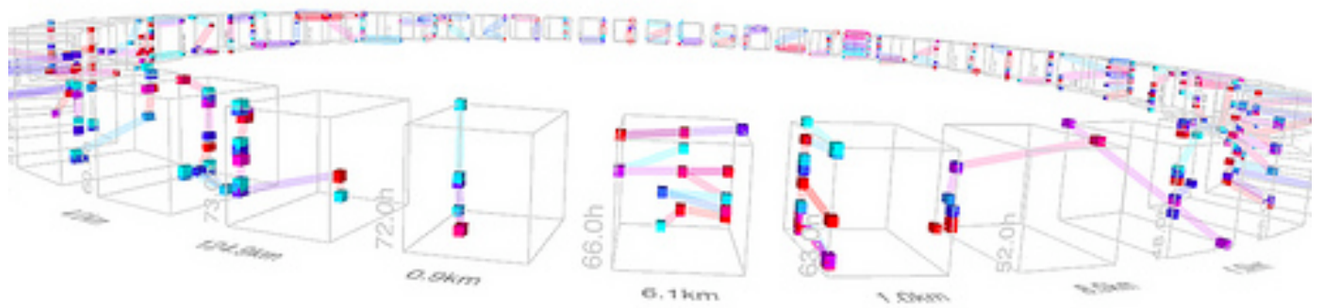


Power-Law-Topology of Internet

- No. of Router vs. Number of Nodes (Links)



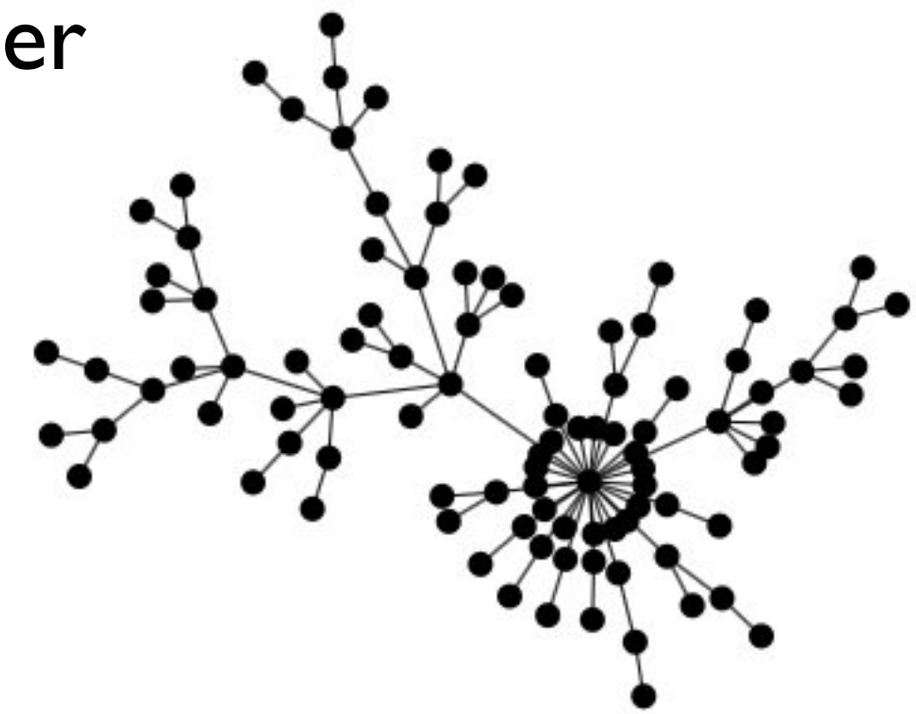
Diameter of WWW



- WWW is a scale-free network
- Diameter of WWW = 19
  (avg. 19 clicks between web pages)
- R. Albert, H. Jeong, and A.-L. Barabási, "The diameter of the WWW," *Nature*, vol. 401, no. 6749, pp. 130-131, 1999.

And How the Web Grows...

- Rich-get-richer



- A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," Science, vol. 286, pp. 509-512, October 15, 1999.

Locating the Source of Diffusion in Large-Scale Network

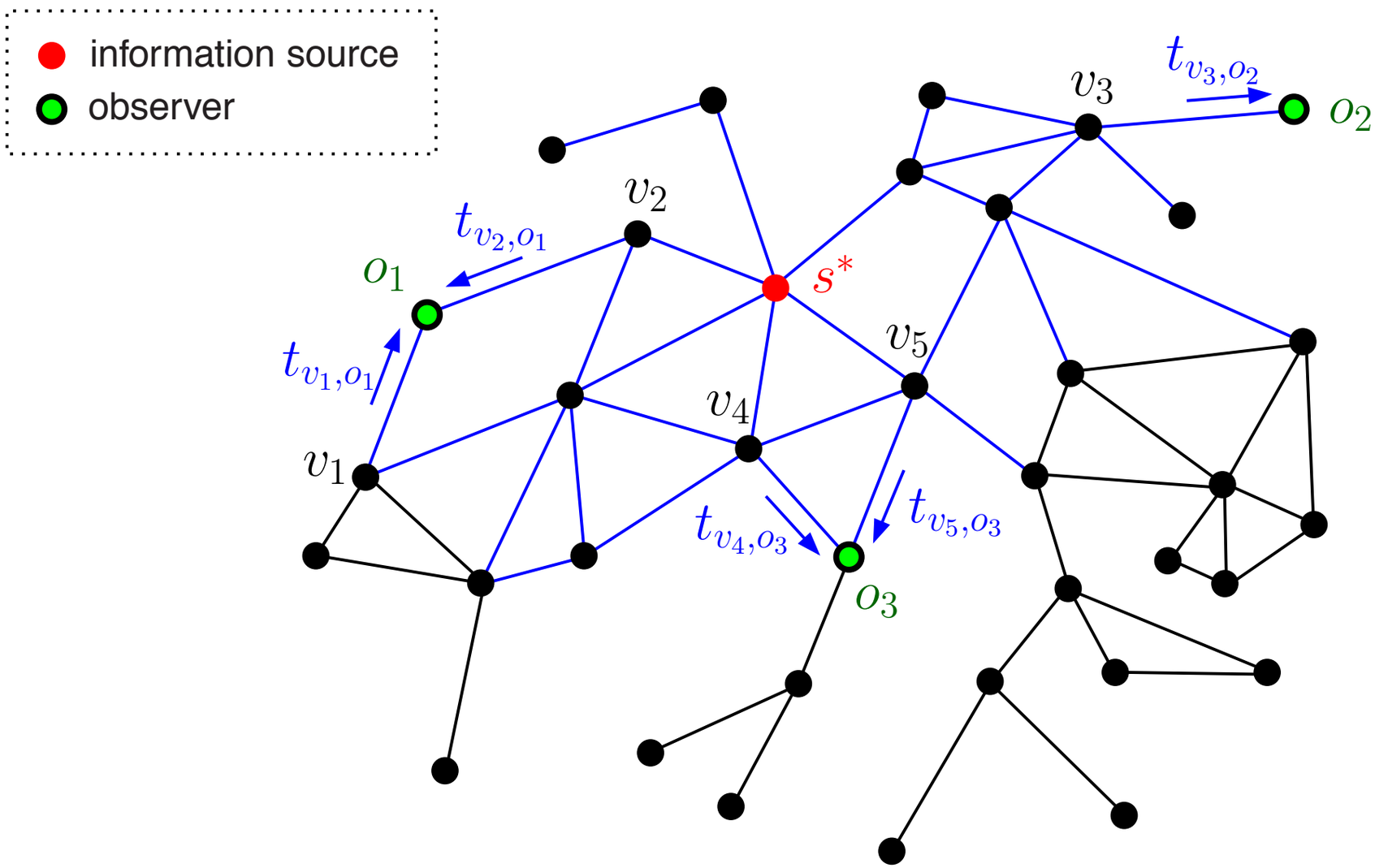
2nd Division, 2nd R&D Institute,
Agency for Defense Development
Donghwan Lee
nomalee@gmail.com

Locating Source of Diffusion

- *forward problem*: on the diffusion process and its dependence on the rates of infection and cure
- but, now we have; *inverse problem*: inferring the original source of diffusion

Network Model

- For a finite, undirected graph $\mathcal{G} = \{V, E\}$



Observers & Observation Set

- if $t_{v,o}$ denotes the absolute time at which observer o receives the information from its neighbour v, then the observation set is composed of tuples of direction and time measurements, i.e., $\mathcal{O} \triangleq \{(o,\ v,\ t_{v,o})\}$, for all $o \in O$ and $v \in V(o)$

The Estimator

- for all source $s \in \mathcal{G}$, the maximum likelihood estimator is,

$$\hat{s}(\mathcal{O}) = \underset{s \in \mathcal{G}}{\operatorname{argmax}}\, \mathbf{P}(\mathcal{O}|s^* = s)$$

$$= \underset{s \in \mathcal{G}}{\operatorname{argmax}} \sum_{\Pi_s} \mathbf{P}(\Pi_s|s^* = s) \times$$

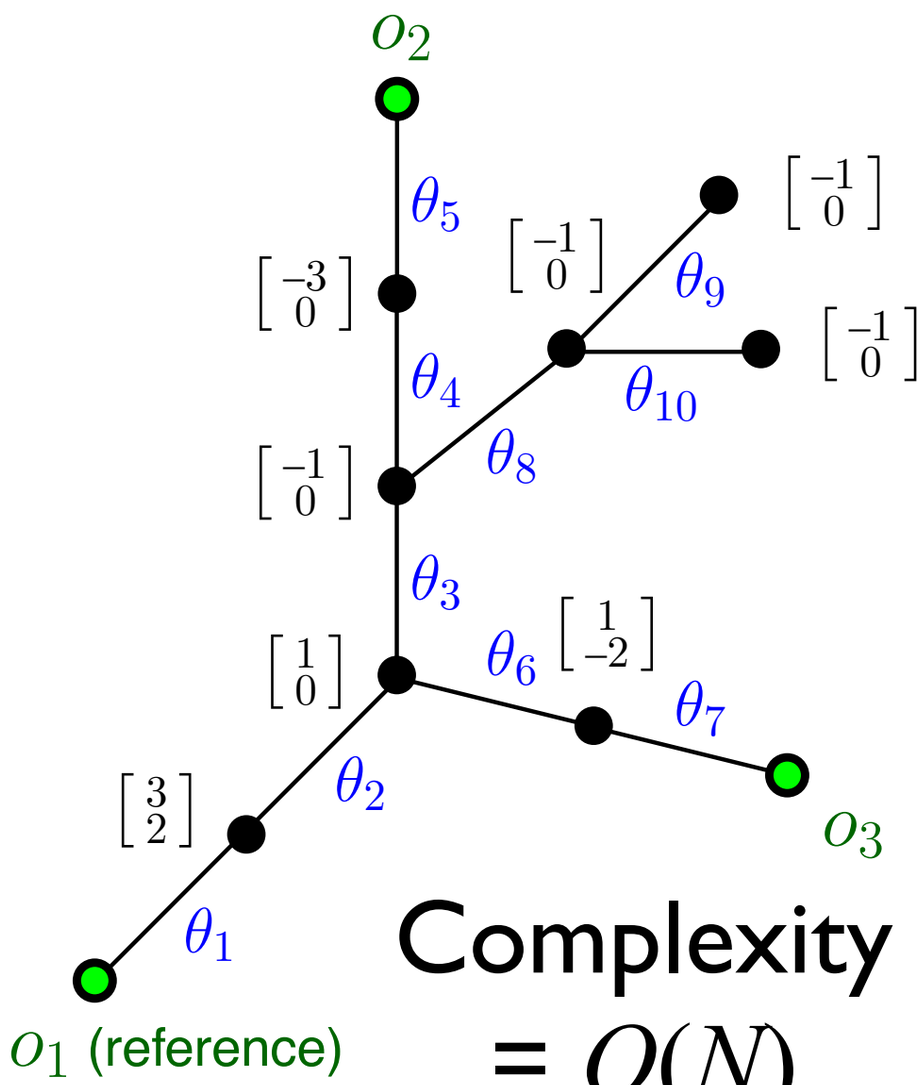
$$\int \cdots \int g(\theta_1, \cdots, \theta_L, \mathcal{O}, \Pi_s, s) d\theta_1 \cdots d\theta_L$$

where Π_s is the set of all possible paths $\{\mathcal{P}_{s,o_k}\}_{k=1}^{K}$ between the source s and the observers in the graph $\mathcal{G}$

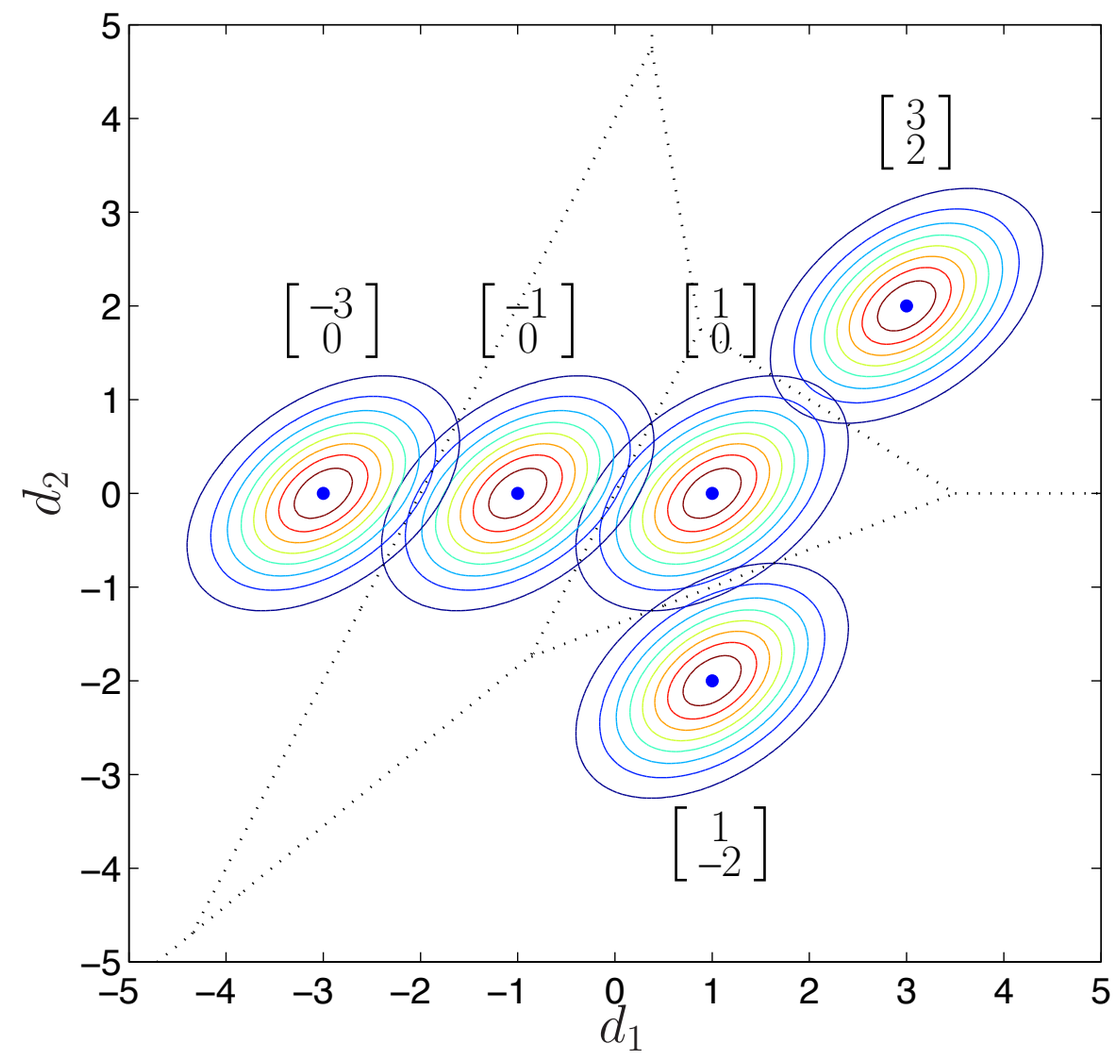
Complexity & Tractability

- Complexity increases exponentially with the number of nodes in $\mathcal{G}$
- Finding a solution in polynomial time
- Consider a tree
- Using direction & timing informations in sensors, build an *active tree* $\mathcal{T}_{\mathrm{a}} \subseteq \mathcal{T}$

Example



Complexity
$= O(N)$



Optimal Estimation in General Trees

Proposition 1 (Optimal Estimation in General Trees): For a general propagation tree $\mathcal{T}$, the optimal estimator is given by

$$\hat{s} = \underset{s \in \mathcal{T}_{\mathrm{a}}}{\operatorname{argmax}} \, \boldsymbol{\mu}_s^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} \left(\mathbf{d} - \frac{1}{2} \boldsymbol{\mu}_s \right) \tag{2}$$

where $\mathbf{d}$ is the *observed delay*, $\boldsymbol{\mu}_s$ is the *deterministic delay*, and $\boldsymbol{\Lambda}$ is the *delay covariance*, given by

$$[\mathbf{d}]_k = t_{k+1} - t_k, \tag{3}$$

$$[\boldsymbol{\mu}_s]_k = \mu \cdot \left(|\mathcal{P}(s, o_{k+1})| - |\mathcal{P}(s, o_1)| \right), \tag{4}$$

$$[\boldsymbol{\Lambda}]_{k,i} = \sigma^2 \cdot \begin{cases} |\mathcal{P}(o_1, o_{k+1})|, & k = i, \\ |\mathcal{P}(o_1, o_{k+1}) \cap \mathcal{P}(o_1, o_{i+1})|, & k \neq i, \end{cases} \tag{5}$$

for $k, i = 1, \ldots, K_{\mathrm{a}} - 1$, with $|\mathcal{P}(u, v)|$ denoting the number of edges (length) of the path connecting vertices u and v.

Optimal Estimation in General Graphs

- Relaxation : assuming that actual diffusion tree is breadth-first-search tree

$$\hat{s} = \operatorname*{argmax}_{s \in \mathcal{G}} \mathcal{S}(s, \mathbf{d}, \mathcal{T}_{\mathrm{bfs},s}), \tag{6}$$

where $\mathcal{S} = \boldsymbol{\mu}_s^{\mathrm{T}} \boldsymbol{\Lambda}_s^{-1} \left(\mathbf{d} - \frac{1}{2}\boldsymbol{\mu}_s\right)$, with parameters $\boldsymbol{\mu}_s$ and $\boldsymbol{\Lambda}_s$ computed with respect to the BFS tree $\mathcal{T}_{\mathrm{bfs},s}$ rooted at s.

Complexity
= $O(N^3)$

Effect of Multiple Cascades

- Time-resolution tradeoff

Proposition 2 (Effect of Multiple Cascades): If the source s^* transmits C independent cascades of information on a tree $\mathcal{T}$, then the probability of correct localization P_{loc} achieved by the optimal estimator is given by

$$P_{\text{loc}} = P_{\text{max}} - O\left(e^{-aC}\right),$$

where P_{max} is the maximum probability of localization achieved under *deterministic* propagation, and a is a constant.

Performance Results

- K/N rate necessary to achieve $P_{loc} = 90\%$

Network	Observer Placement	
	High-degree	Random
Apollonian	4%	25%
Barabási-Albert	18%	41%
Erdös-Rényi ($Np = 0.5$)	34%	49%
Erdös-Rényi ($Np = 2$)	32%	44%

Conclusion

- Now we can locate the origin of diffusion in networks,

  only within $O(N^3)$ complexity and with (comparably) quite a few sensors deployed