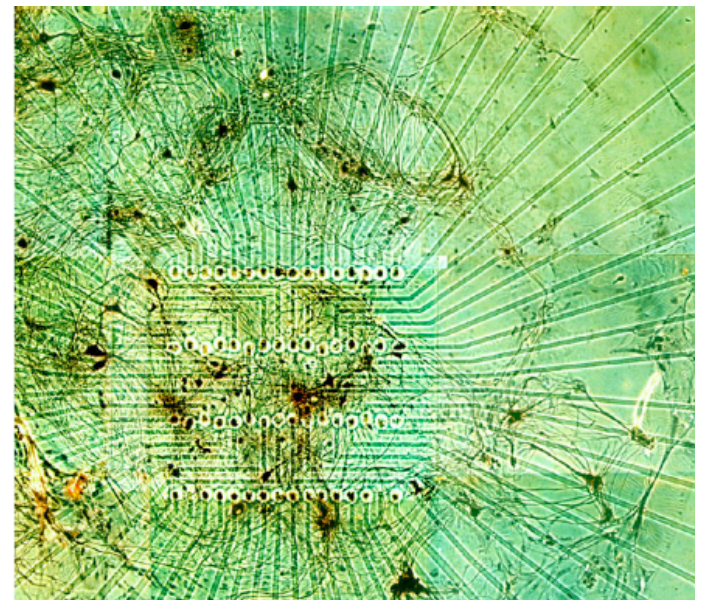
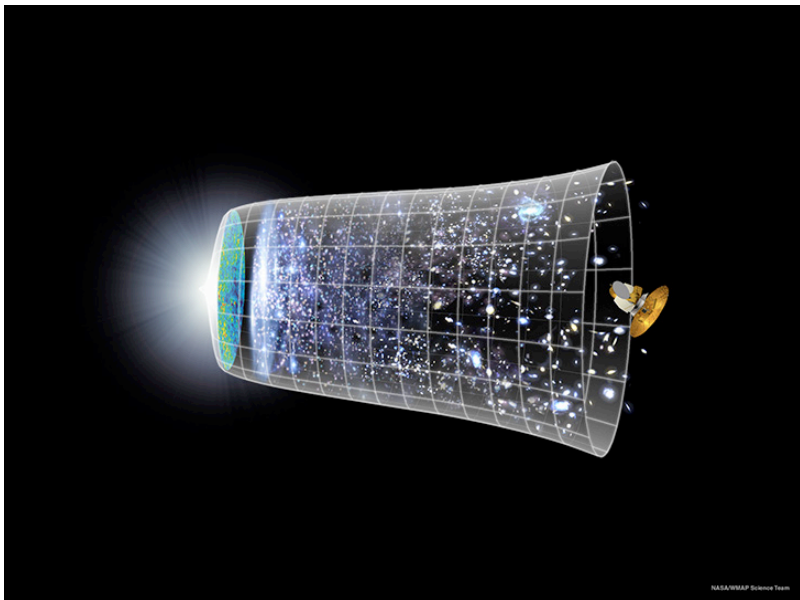


# Uncertainty quantification in complex systems: emerging diseases, predicting science and the functional structure of the brain



Luís M. A. Bettencourt

Los Alamos National Laboratory & Santa Fe Institute

<http://math.lanl.gov/~lmbett>

# Uncertainty quantification in complex systems

## **Examples:**

climate, markets, cities, emerging diseases, science and innovation ecosystems and biodiversity, brain and cognition

## **Properties:**

many components, heterogeneity, uncertainty in initial conditions and parameters, exposed to externalities, learn and adapt

## **Modeling and theory in complex systems:**

No models exist that give long time predictability

Few models have been proposed that give short term reliable predictions

Causality is not well understood

**Uncertainty** in models and predictions **must be quantified** for falsification

Lots of data are coming in !

# Synopsis

**Predicting the epidemic potential of emerging infectious diseases**

with Ruy Ribeiro

**Accelerating Science and Technology**

with David Kaiser, OSTI DOE

**How do complex networks process information?**

the functional information structure of living neural networks

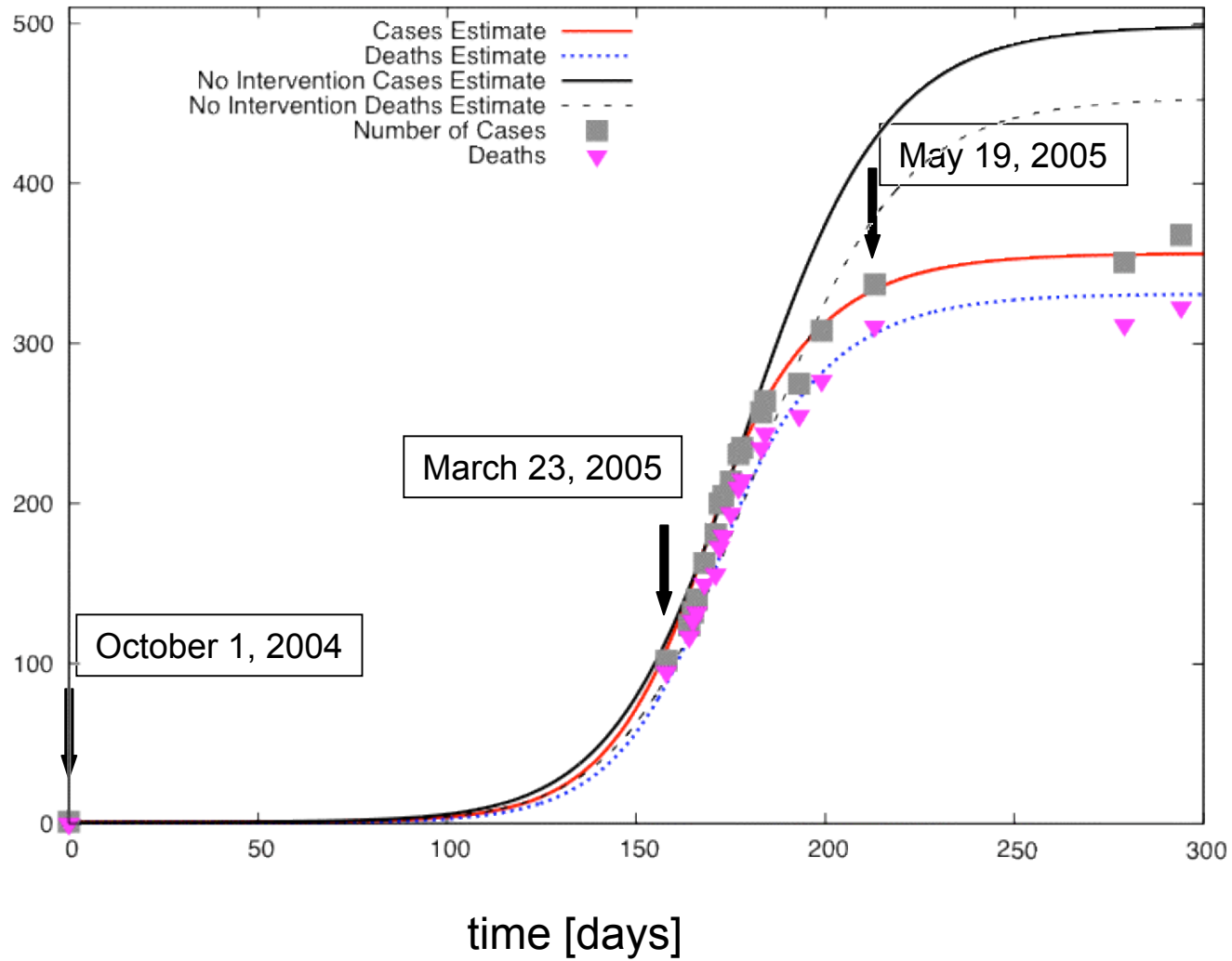
with Vadas Gintautas and Michael Ham

# Estimating the epidemic potential of emerging infectious diseases

“What’s the risk of a H5N1 (bird) flu pandemic in the next 3 years?”

anonymous DHS program manager

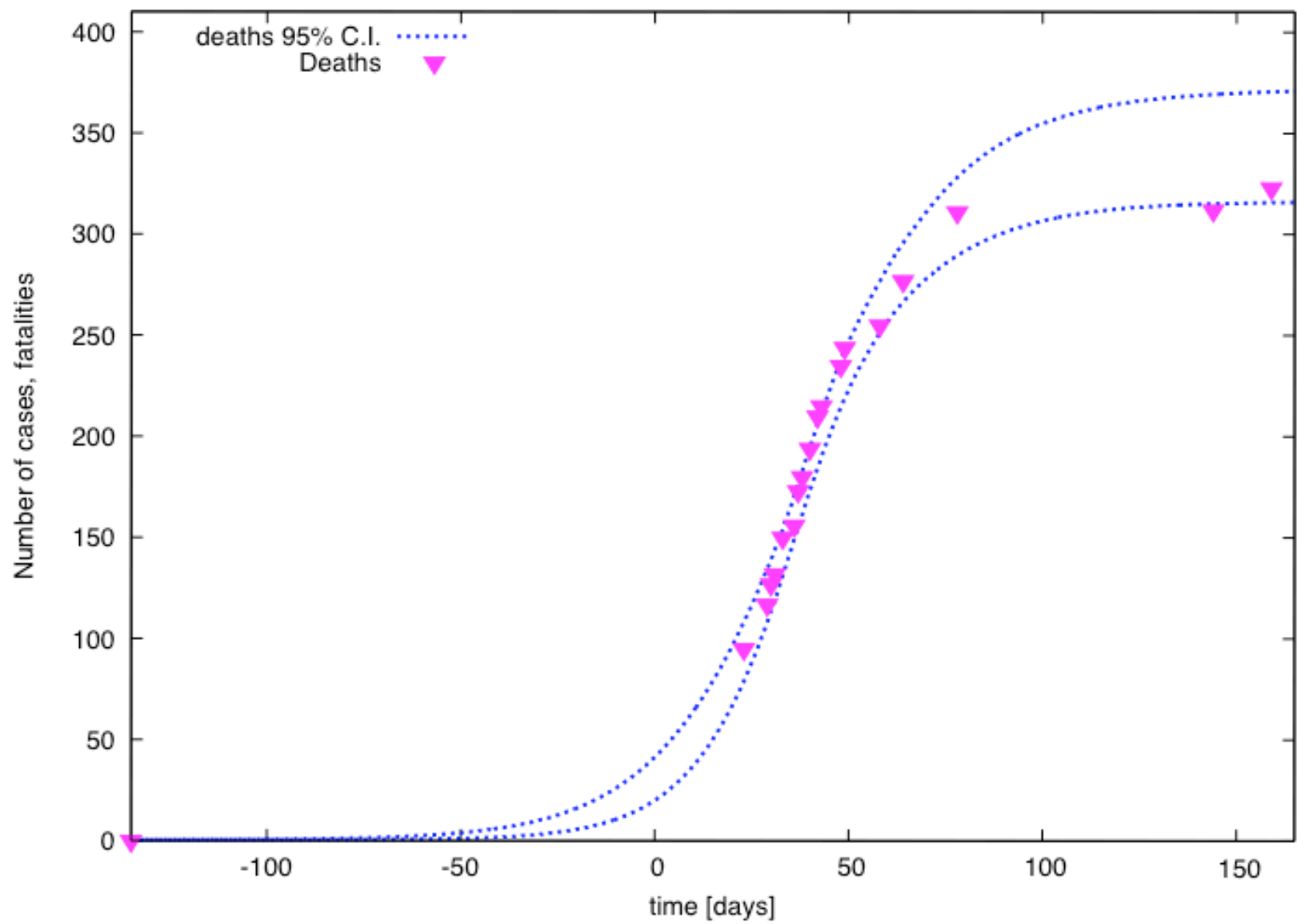
# Outbreak of Marburg fever: Angola, 2005





# Conditions in Uige:







“What’s the risk of a H5N1 (bird) flu pandemic in the next 3 years?”

anonymous DHS program manager

# H5N1 avian influenza

Influenza A virus, very contagious among birds:

pandemic: Asia, Europe, Africa 51 countries

Caused over 381 human cases, with 240 deaths 63% case mortality

Presently very low transmissibility among humans:  $0 < R_0 \ll 1$

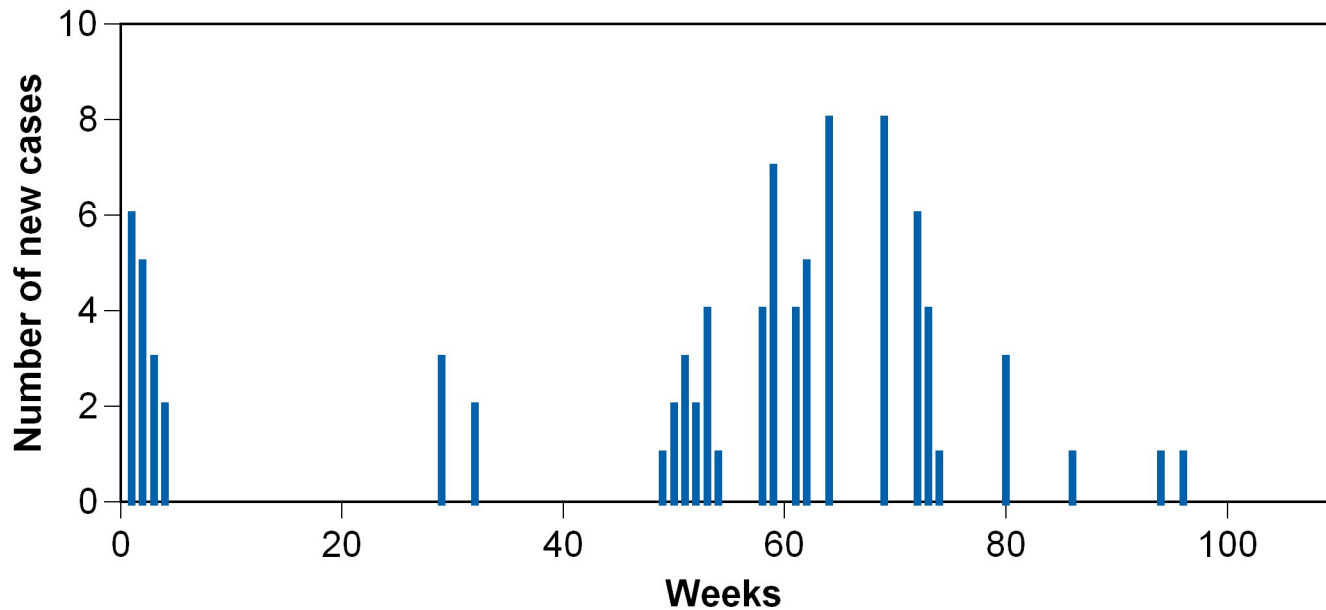
## How will H5N1 influenza evolve?

What will be the signs of sustained human transmission:

Number of new cases induced by an infectious individual:  $R_0 < 1$ ,  $R_0 \rightarrow 1$ ?

# Typical EID time series

## human cases of H5N1 influenza in Vietnam



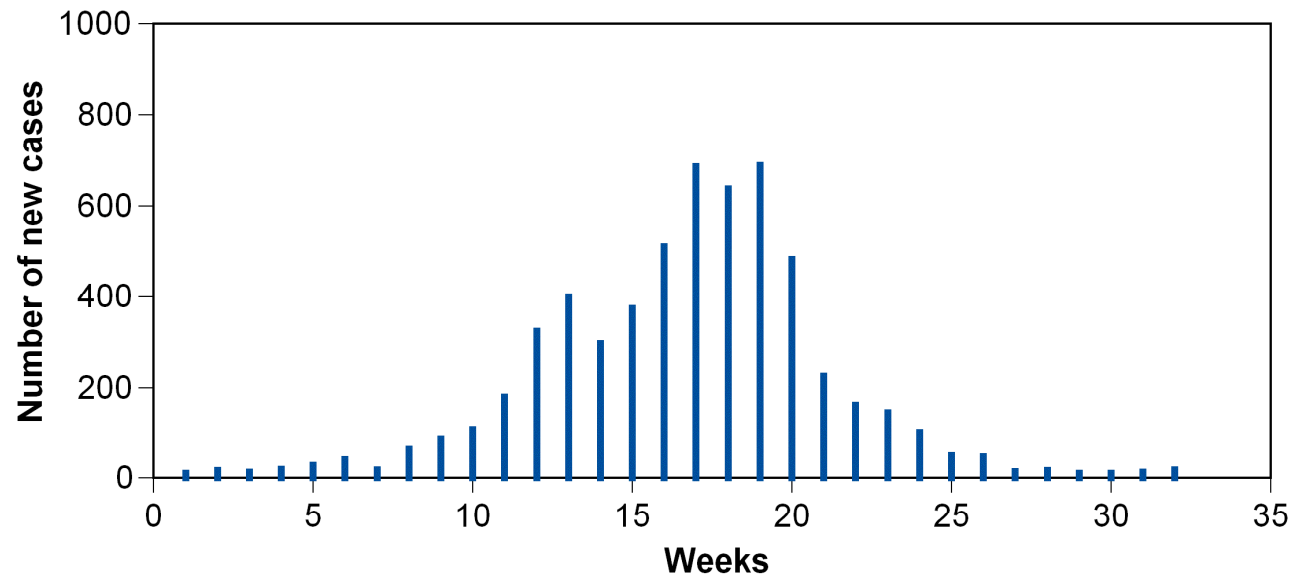
From “Situation Updates: Avian Influenza”

World Health Organization (WHO)

Laboratory confirmed H5N1 cases (<http://www.who.int/csr/don/en>).

# vs. Seasonal flu epidemic time series

## H3N2 USA isolates 2004-05



H3N2 seasonal influenza isolates  
from the Center for Disease Control (CDC)  
*Surveillance Weekly Reports* in the United States  
(<http://www.cdc.gov/flu/weekly/fluactivity.htm>)

# Requirements for model of EID

A successful model for EID should:

- i) incorporate **reservoir sources**, associated with **multiple introductions**
- ii) formulate the model in **discrete probabilistic form**
- iii) **quantify uncertainty** in epidemiological parameters and use new data to reduce it
- iv) cast **state variables in terms of observable quantities**,  
reported **from field surveillance**
- v) estimation procedure should **not** depend on  
**future unknown data**, such as final case cluster size
- vi) **supply surveillance with real-time probabilistic expectations**,  
which when violated may indicate that:
  - there are **errors** in the new data
  - the **pathogen is evolving**
  - the host **population is changing**

# Epidemic *Mean* Dynamics in terms of 'observables'

# SIR model without Sources

Consider a standard SIR model:

$$R_0 = \beta/\gamma$$

$$\dot{s} = -\beta \frac{S}{N} I, \quad \dot{i} = \left[ \beta \frac{S}{N} - \gamma \right] I \quad \text{Total Cases: } \dot{T} = \beta \frac{S}{N} I$$

The solution is:

$$I(t + \tau) = I(t) \exp \left[ \gamma \int_t^{t+\tau} \left( R_0 \frac{S(t')}{N} - 1 \right) dt' \right] \cong I(t) \exp \left[ \gamma \tau \left( R_0 \frac{S}{N} - 1 \right) \right]$$

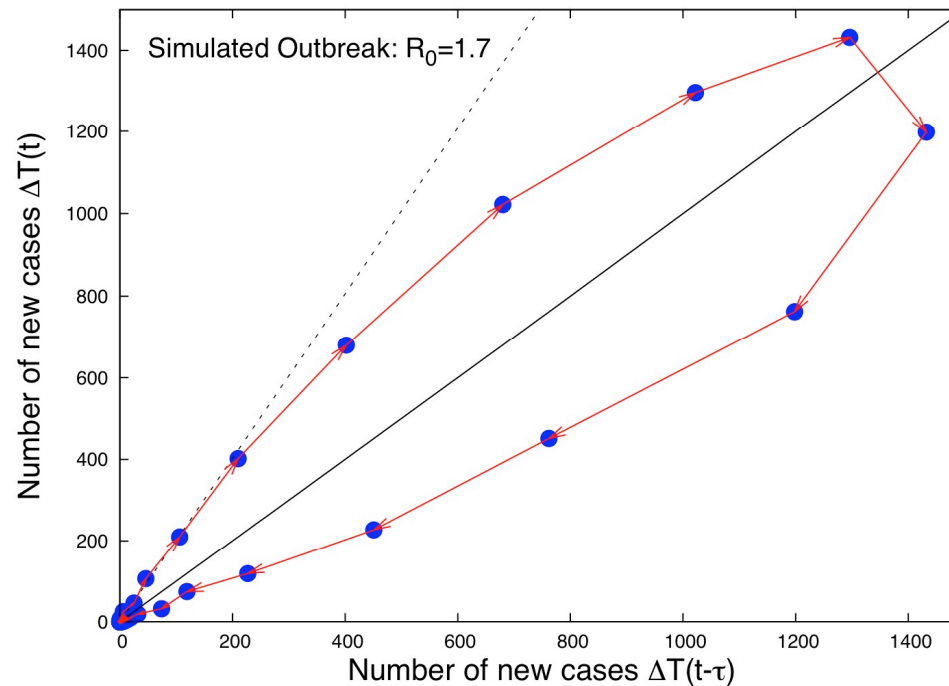
Evolution of the **expectation value** for **New Cases**

$$\langle \Delta T(t + \tau) \rangle = b(R) \Delta T(t) \quad b(R_0) = \exp \left[ \gamma \tau \left( R_0 \frac{S}{N} - 1 \right) \right]$$

$b(R)$  is the branching parameter:  $R_{ML} = 1 + \frac{1}{\gamma \tau} \ln \frac{\Delta T(t + \tau)}{\Delta T(t)}$

# Epidemic time delay diagrams: $R > 1$

$$\langle \Delta T(t + \tau) \rangle = b(R) \Delta T(t)$$



$b(R)$  is the  
**slope** of the  
tangent at  
the origin

$R_0$  can be determined *geometrically*,  
- without complex parameter estimation.



# Model with Sources

(multiple introductions)

Model with two classes of human infected:

$$\dot{I}^h = \left[ \beta \frac{S(t)}{N} - \gamma \right] I^h(t) + \beta_{bh} \frac{S(t)}{N} f_c K(t), \quad \dot{I}^b = \beta_{bh} \frac{S(t)}{N} (1 - f_c) K(t) - \gamma I^b$$

Take:  $\dot{B} = \beta_{bh} \frac{S(t)}{N} K(t)$  to give:

$$\dot{I} = \left[ \beta \frac{S(t)}{N} - \gamma \right] I^h(t) + \dot{B} - \gamma I^b, \quad \dot{T} = \beta \frac{S(t)}{N} I^h(t) + \dot{B}.$$

Leading to the solution:

$$I^h(t + \tau) = b(R_0) \left[ I^h(t) + \int_t^{t+\tau} e^{-\gamma(R_0 S/N - 1)} f_c \dot{B}(t') dt' \right] \equiv b(R_0) [I^h(t) + f_c \psi(t, \tau, B)]$$

# Model with Sources (cont.)

(multiple introductions)

The progression of the **expectation value** for **New Cases** obeys:

$$\langle \Delta T(t + \tau) \rangle = \Delta B(t + \tau) + b(R_0) \left[ \Delta T(t) - \Delta B(t) + \tau \gamma R_0 \frac{S}{N} f_c \Delta B(t) \right]$$

new cases from sources (birds)

time evolution

old cases from human transmission

evolution of introduced infectious cases

New cases are treated as a stochastic variable with this average

$$\Delta T(t + \tau) \sim P[\Delta T(t + \tau) \leftarrow \Delta T(t) | \Gamma]$$

The functional form of **P** is constrained by the mean

# Probabilistic Epidemic Models

Real Time] Bayesian Parameter Estimation

# Estimating the probability distribution of $R$ , $\gamma$ , etc from surveillance time series

Usual perspective:

“Initial value problem”

Previous cases + Model ( $\Gamma$ )  $\Rightarrow$  probability dist. of New Cases

**is equivalent to:**

Alternative perspective:

“Boundary value problem”

Previous Cases + New Cases  $\Rightarrow$  probability dist. of model ( $\Gamma$ )

**surveillance time series**

**$P(\Gamma)$**

# This results from Bayes' Theorem

$$P[\Gamma \mid \Delta T(t + \tau) \leftarrow \Delta T(t)] = \frac{P[\Delta T(t + \tau) \leftarrow \Delta T(t) \mid \Gamma]P[\Gamma]}{P[\Delta T(t + \tau) \leftarrow \Delta T(t)]}$$

$\Gamma$  are the model parameters

- $P[\Gamma]$  is the ‘prior’ [ it expresses the expected distribution of  $\Gamma$  ]
- $P[\Delta T(t + \tau) \leftarrow \Delta T(t)]$  is a normalization factor:

$$P[\Delta T(t + \tau) \leftarrow \Delta T(t)] = \int d\Gamma P[\Delta T(t + \tau) \leftarrow \Delta T(t) \mid \Gamma]P[\Gamma]$$

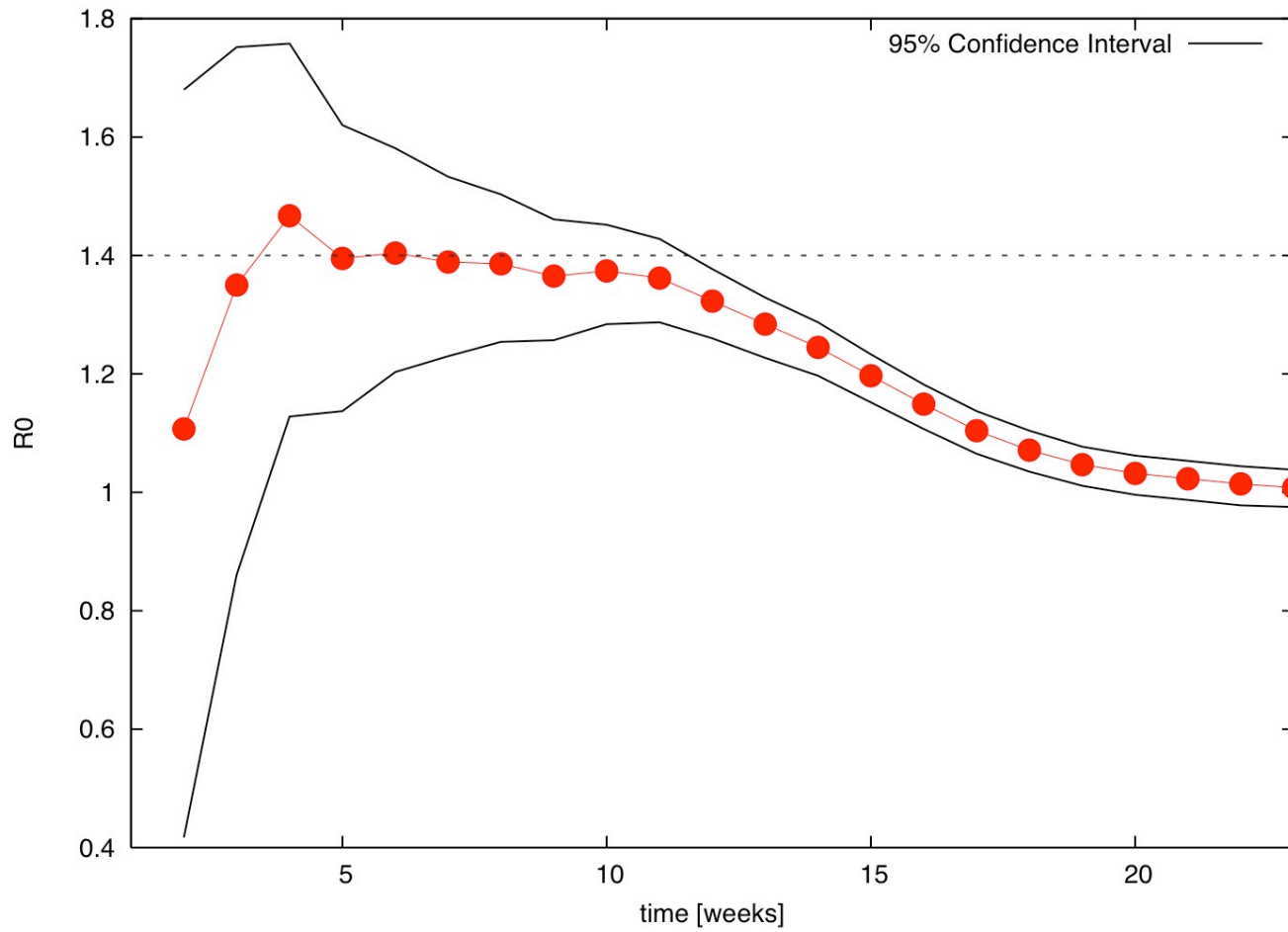
# Iterative estimation and uncertainty reduction

At the next time

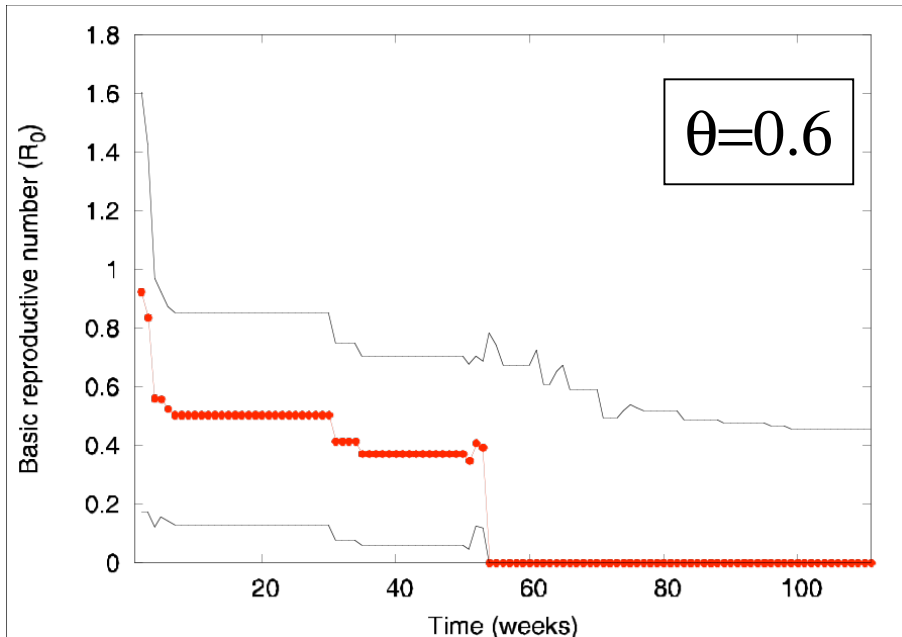
$$P[\Gamma \mid \Delta T(t + \tau) \leftarrow \Delta T(t)] = \frac{P[\Delta T(t + \tau) \leftarrow \Delta T(t) \mid \Gamma]P[\Gamma]}{P[\Delta T(t + \tau) \leftarrow \Delta T(t)]}$$

At this time

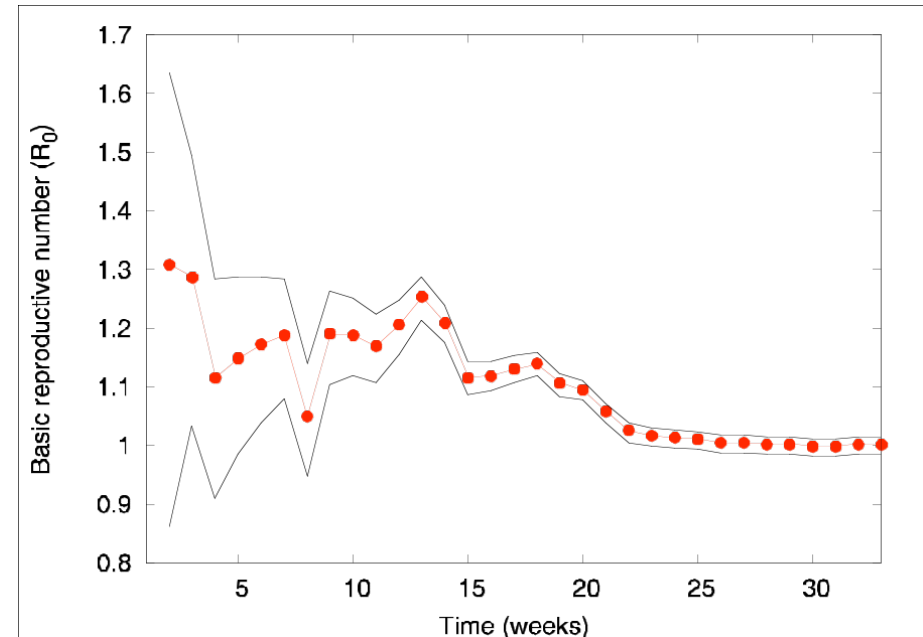
# Simulated Outbreaks



# Real time evolution of maximum likelihood $R_0$ and 95% confidence interval



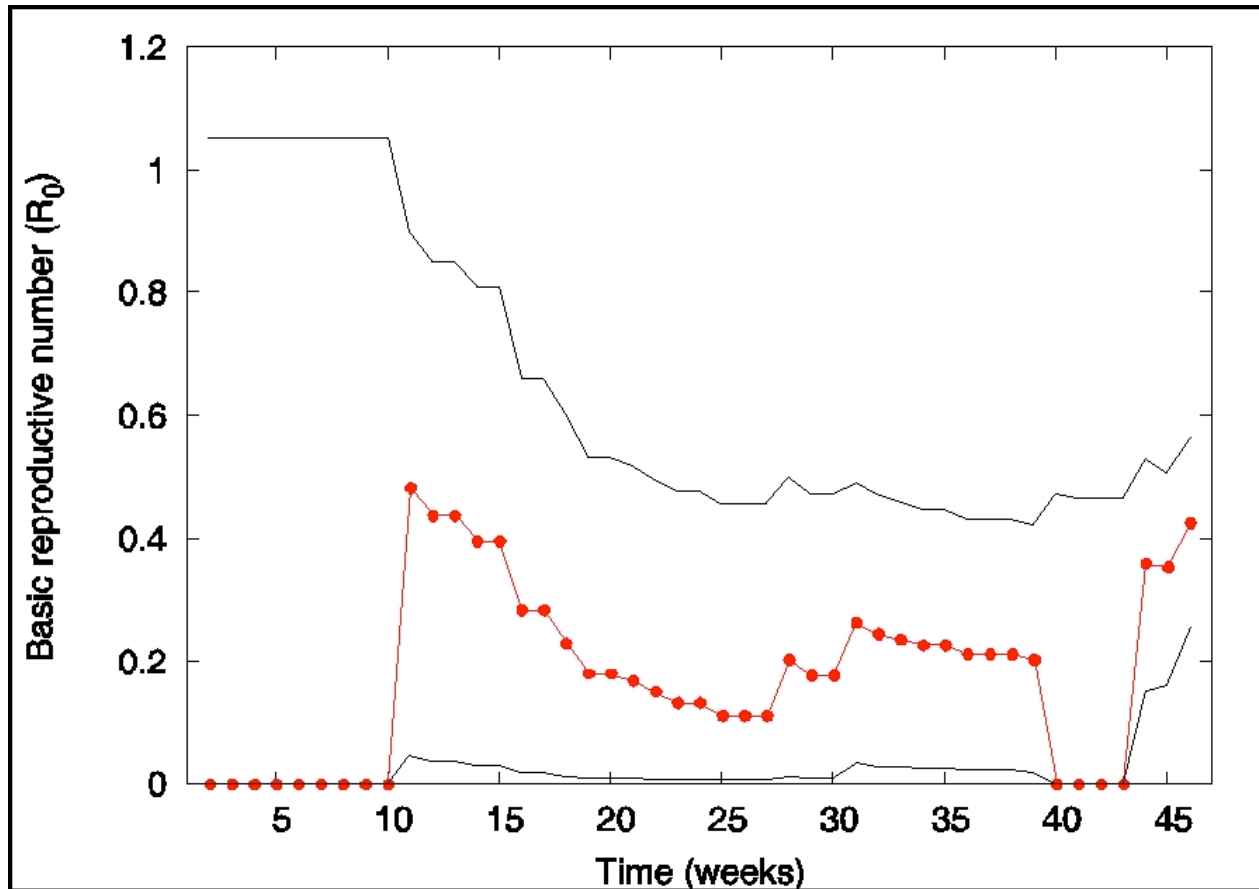
**H5N1 avian influenza: Vietnam**



**H3N2 seasonal influenza: USA**



# Indonesia



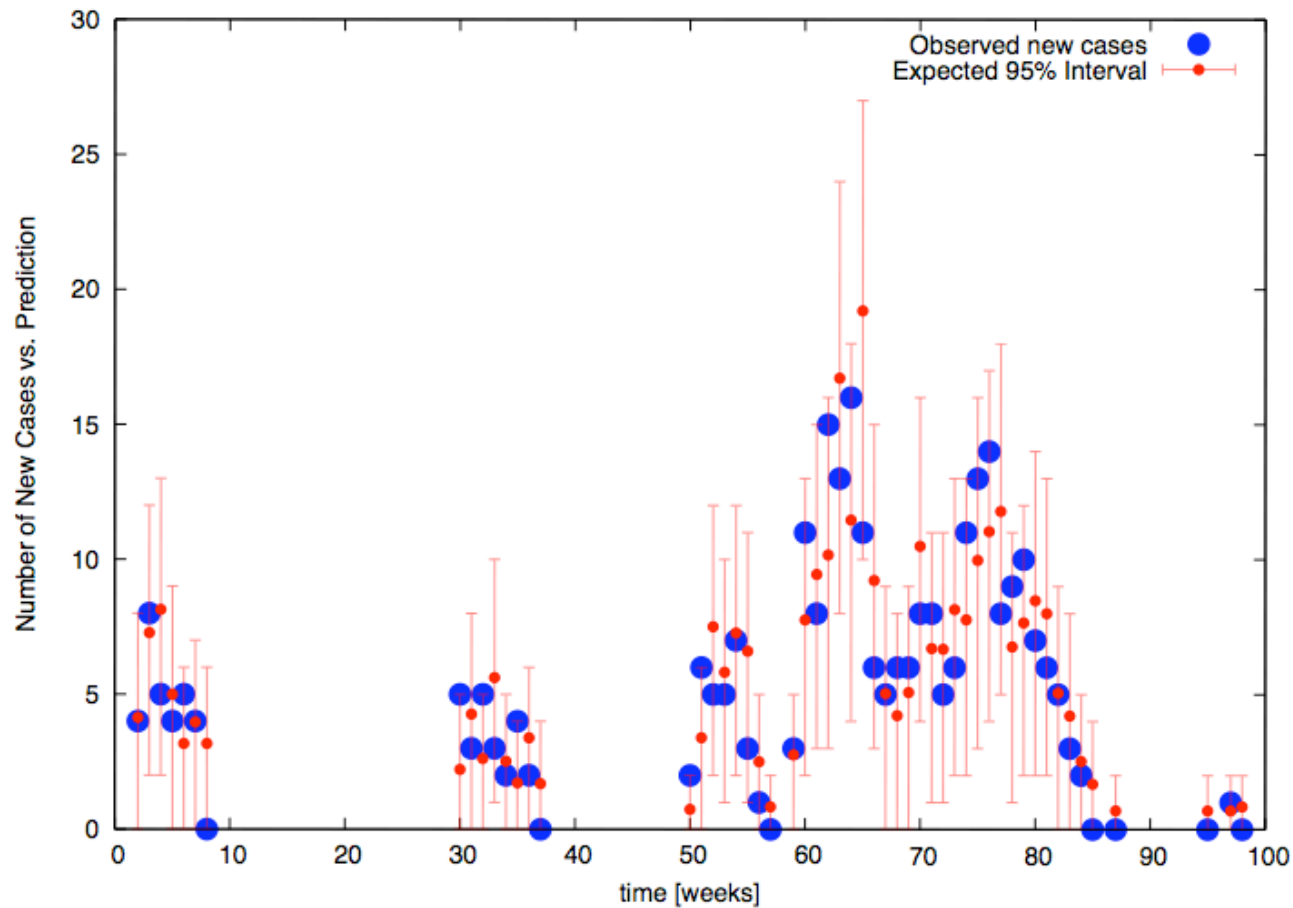
# The $R_0$ of H5N1 influenza in humans

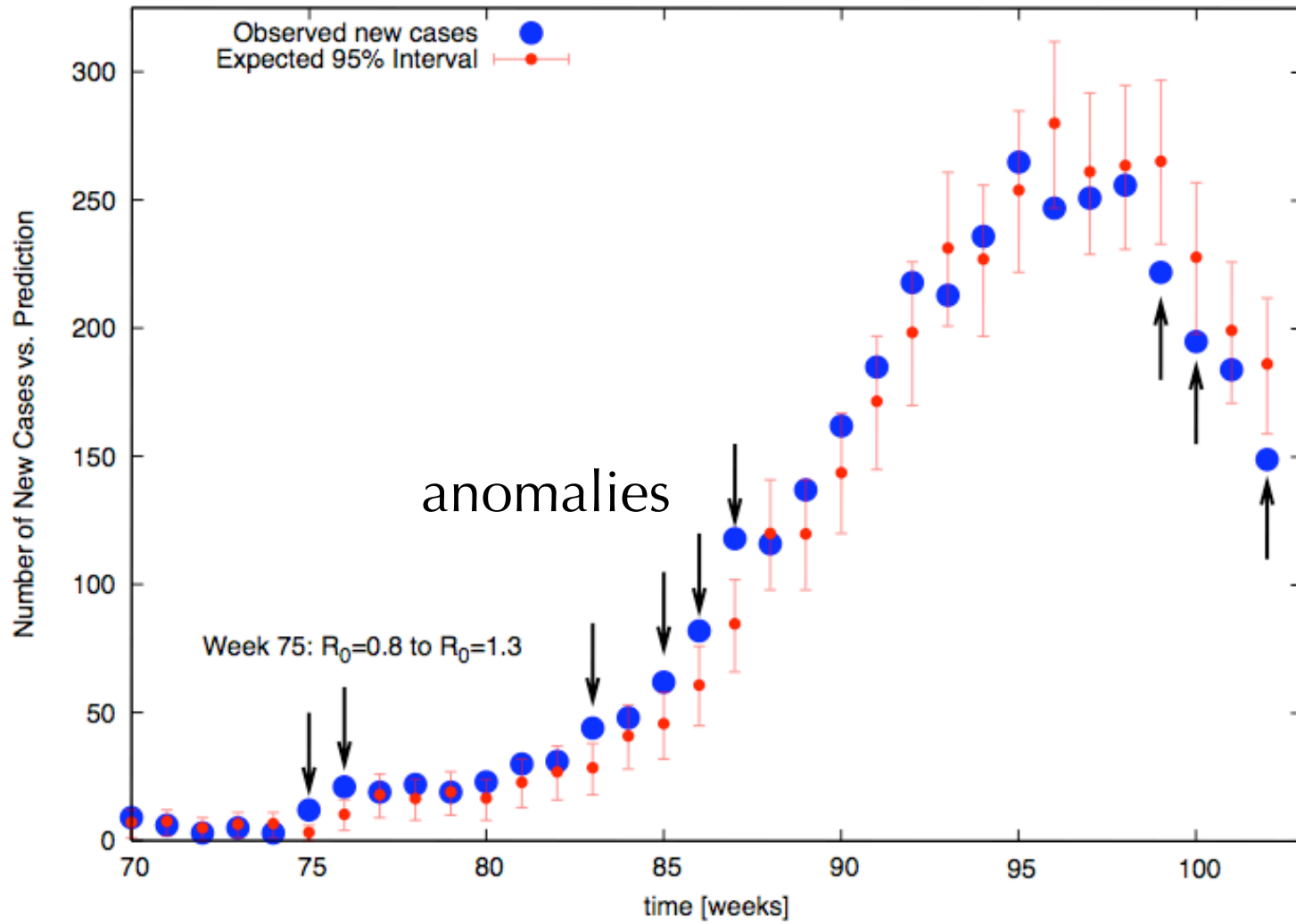
	VIETNAM				INDONESIA			
	Average fraction of cases attributable to human contagion ( $\theta$ )							
	1.0	0.8	0.4	0.2	1.0	0.8	0.4	0.2
$R_0$ min	0.26	0.23	0	0	0.68	0.56	0.26	0
ML $R_0$	0.53	0.46	0	0	0.84	0.71	0.43	0
Mean $R_0$	0.52	0.46	0.08	0	0.83	0.70	0.42	0
$R_0$ max	0.77	0.68	0.46	0	0.97	0.83	0.56	0

even in worst case scenario:  $R_0 < 1$

# Active surveillance through real time prediction and anomaly detection

$$\Gamma, \Delta T(t) \quad \rightarrow \quad \Delta T(t + \tau) \sim P[\Delta T(t + \tau) \leftarrow \Delta T(t) | \Gamma]$$





Standard **p-value test** at 95% significance

# **Accelerating Science and technology**

a predictive “science of science”?

NSF, DOE OSTI

Can science be forecast?

-when is a field opening or closing?

-what are the signatures of new scientific discoveries?

-“Paradigm shifts” vs. “normal science”

can they be distinguished from analysis of literature?

Prediction enables “interventions”:

How should agencies and institutions allocate resources:

Students? Meetings? Individual PIs?

How can scientific discovery be accelerated?

# The structure of scientific revolutions

T. Kuhn 1961



Are there quantitative signatures  
of where a field is ?

**“paradigm shift”**

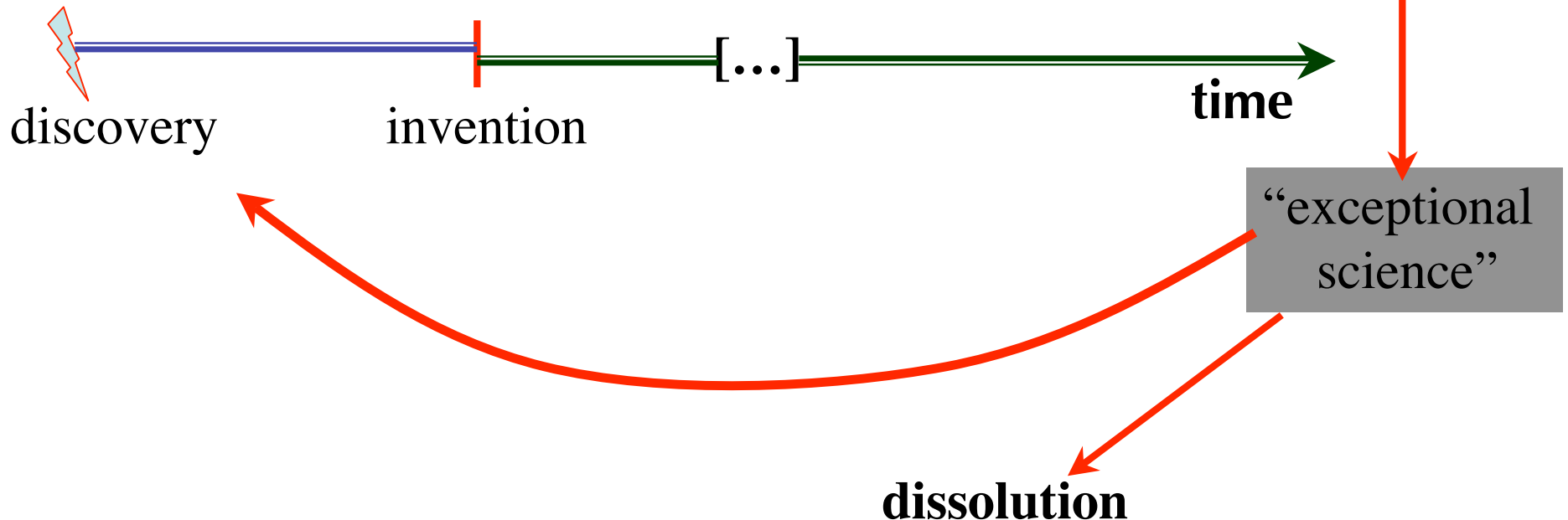
**“normal science”**

**crisis**

**[discovery & invention]**

**[puzzle solving]**

**inconsistencies**



## 6+2 examples of scientific discovery

Cosmological Inflation

Cosmic Strings

String Theory



Theoretical Physics

Prions

H5N1 Influenza



BioMedical

Quantum Computing & Computation

Carbon Nanotubes



Applied Physics  
Material Science  
Engineering

Cold Fusion

“Pathological” Science



# Data sources and retrieval

*SearchPlus* developed by the LANL's Research Library  
Library Without Walls (<http://library.lanl.gov/lww/>)

Searches the standard set of largest scientific databases:

BIOSIS

Engineering Index Proceedings

Inspec

**ISI databases (Thomson Scientific):**

ISI Proceedings

ISI SciSearch

ISI Social SciSearch

ISI Arts & Humanities



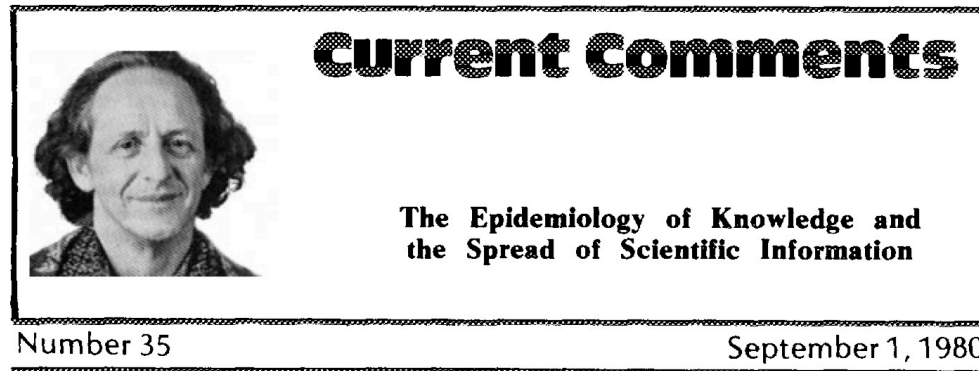
Each field is built from a  
combination of searches and  
analyzed by a domain expert

**Retrieved data (HTML) -> Parsed -> Relational Databases**

authors, title, date, journal reference

# Ideas as ‘epidemics of knowledge’

Essays of an Information Scientist, Vol:4, p.586-591, 1979-80 Current Contents, #35, p.5-10, September 1, 1980



## article

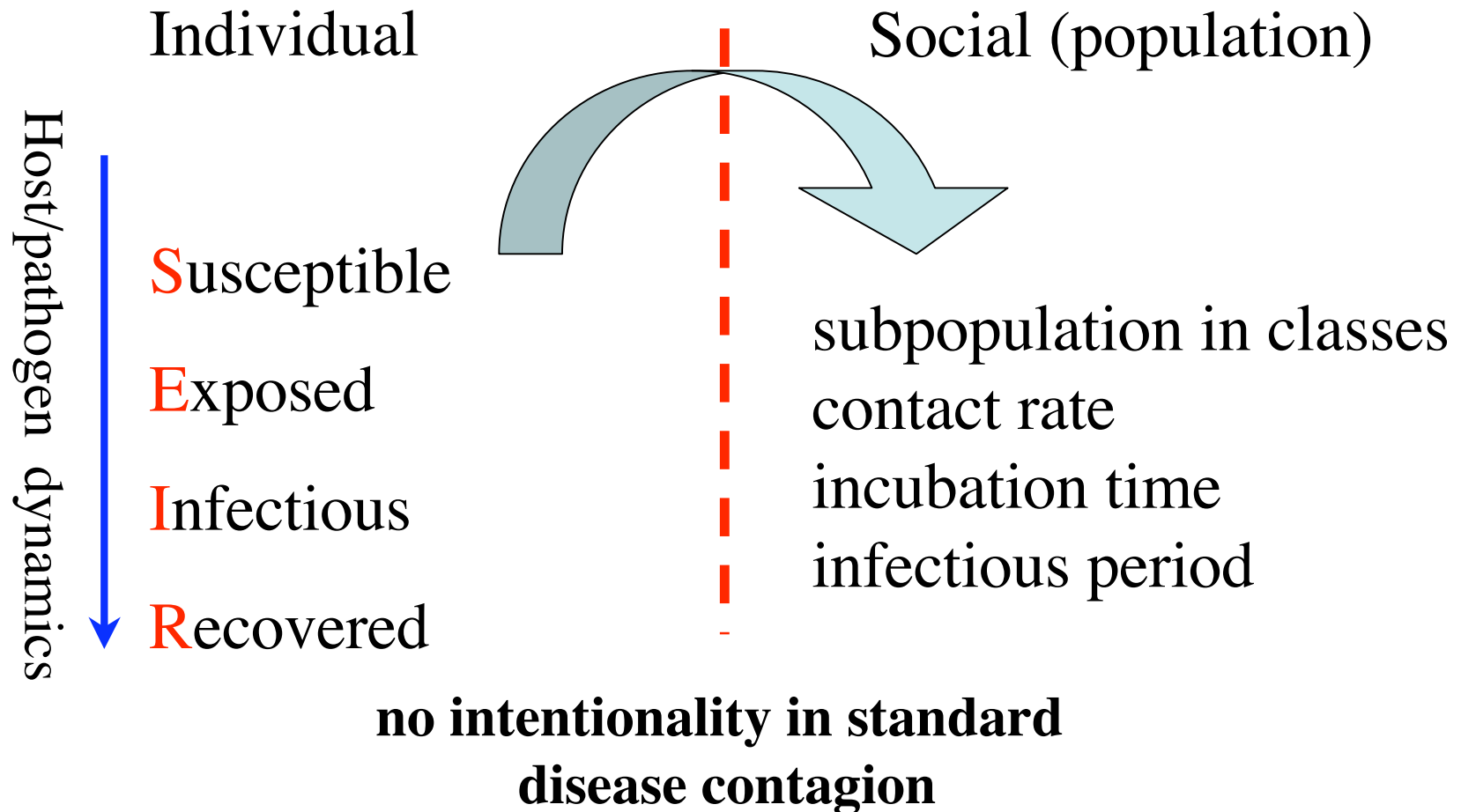
*Nature* 204, 225 - 228 (17 October 1964); doi:10.1038/204225a0

## Generalization of Epidemic Theory: An Application to the Transmission of Ideas

WILLIAM GOFFMAN & VAUN A. NEWILL

Center of Documentation and Communication Research, School of Library Science, Western Reserve University  
School of Medicine, Western Reserve University, Cleveland, Ohio

# Parallels between social dynamics and epidemiology



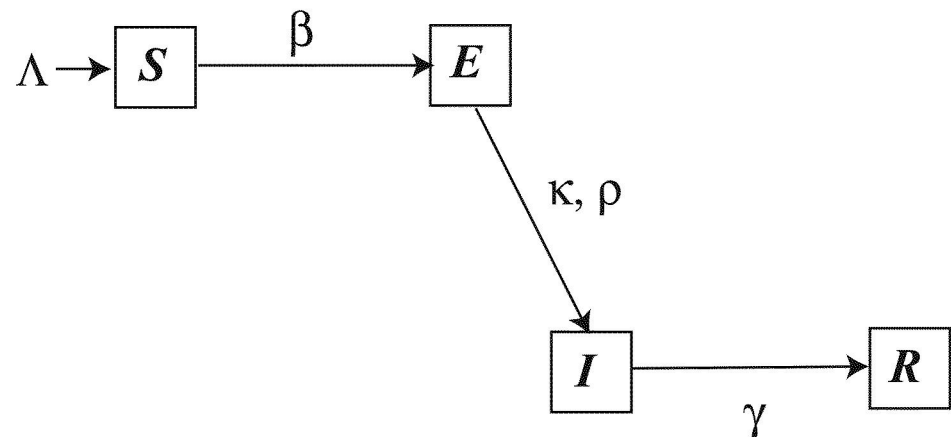
# Dynamical Model

$$\frac{dS}{dt} = \Lambda - \beta S \frac{I}{N},$$

$$\frac{dE}{dt} = \beta S \frac{I}{N} - \rho S \frac{I}{N} - \kappa E,$$

$$\frac{dI}{dt} = \rho S \frac{I}{N} + \kappa E - \gamma I,$$

$$\frac{dR}{dt} = \gamma I$$



$R_0 = \beta/\gamma$  is a measure of transmissibility  
Basic reproduction number

# Parameter Search and Optimization

## Strategy:

- Search for the best parameters is an optimization problem: minimizing the deviation of the model relative to the data
- **Optimization** within a fixed tolerance leads to many good solutions from which we construct:

Joint probability distribution for model parameters conditional on observed data:


$$P[\Gamma | I^o]$$

$$\Gamma = (\underbrace{S(t_0), I(t_0), E(t_0), R(t_0)}_{\text{Initial State}}, \underbrace{\beta, \Lambda, \kappa, \rho, \gamma}_{\text{Dynamical Parameters}})$$

Indirect estimation of  $P[\Gamma|I^o]$  from trajectories:

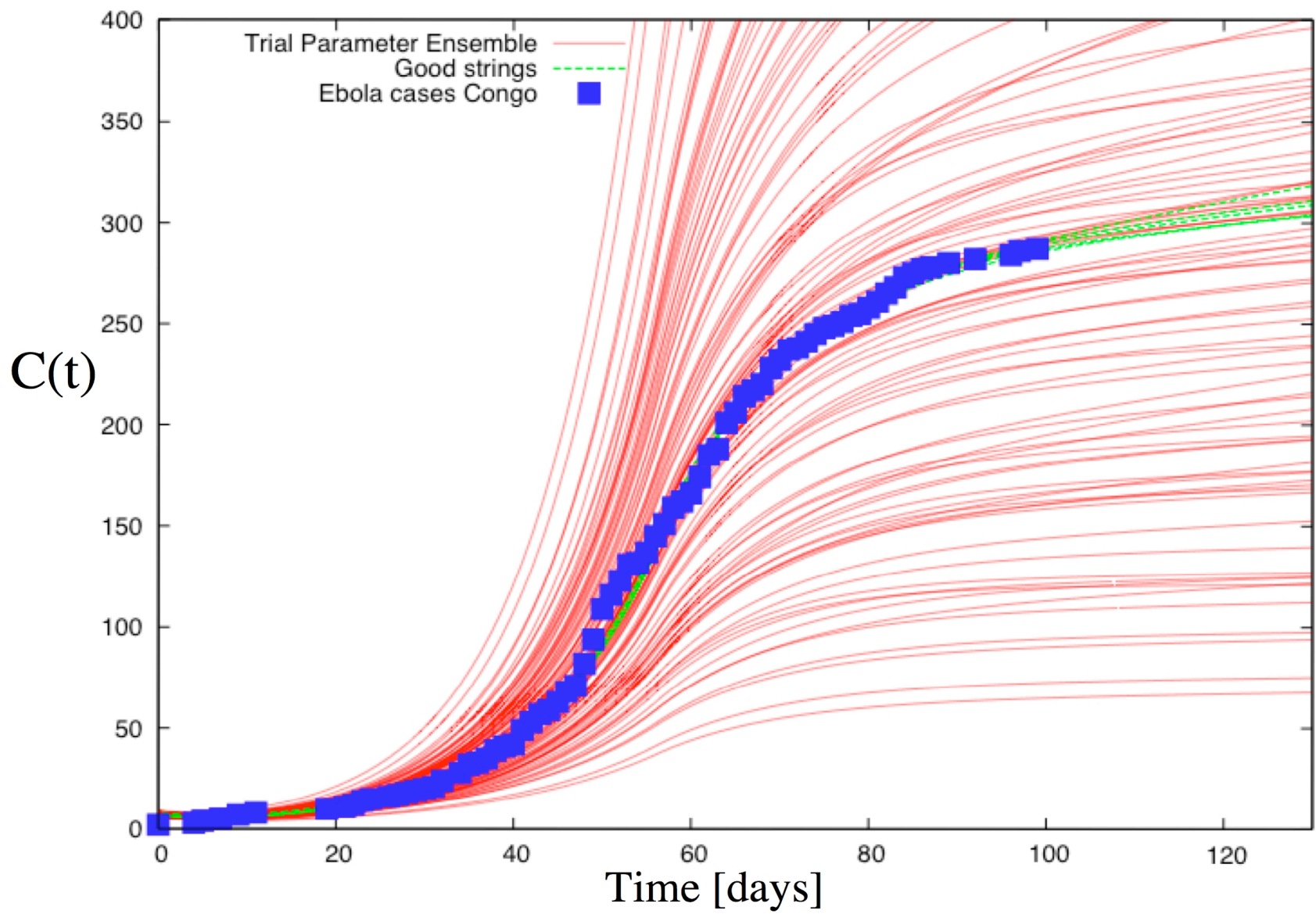
Deviation (action):

$$A(\Gamma) = \frac{1}{N} \sum_{i=1}^N \frac{(I^\Gamma(t_i) - I^o(t_i))^2}{2\sigma_{t_i}}$$

where  $I^\Gamma(t_i)$  is the state given by solving the model with initial conditions and dynamical parameters given by  $\Gamma$ , evaluated at the data points  Inverse Problem

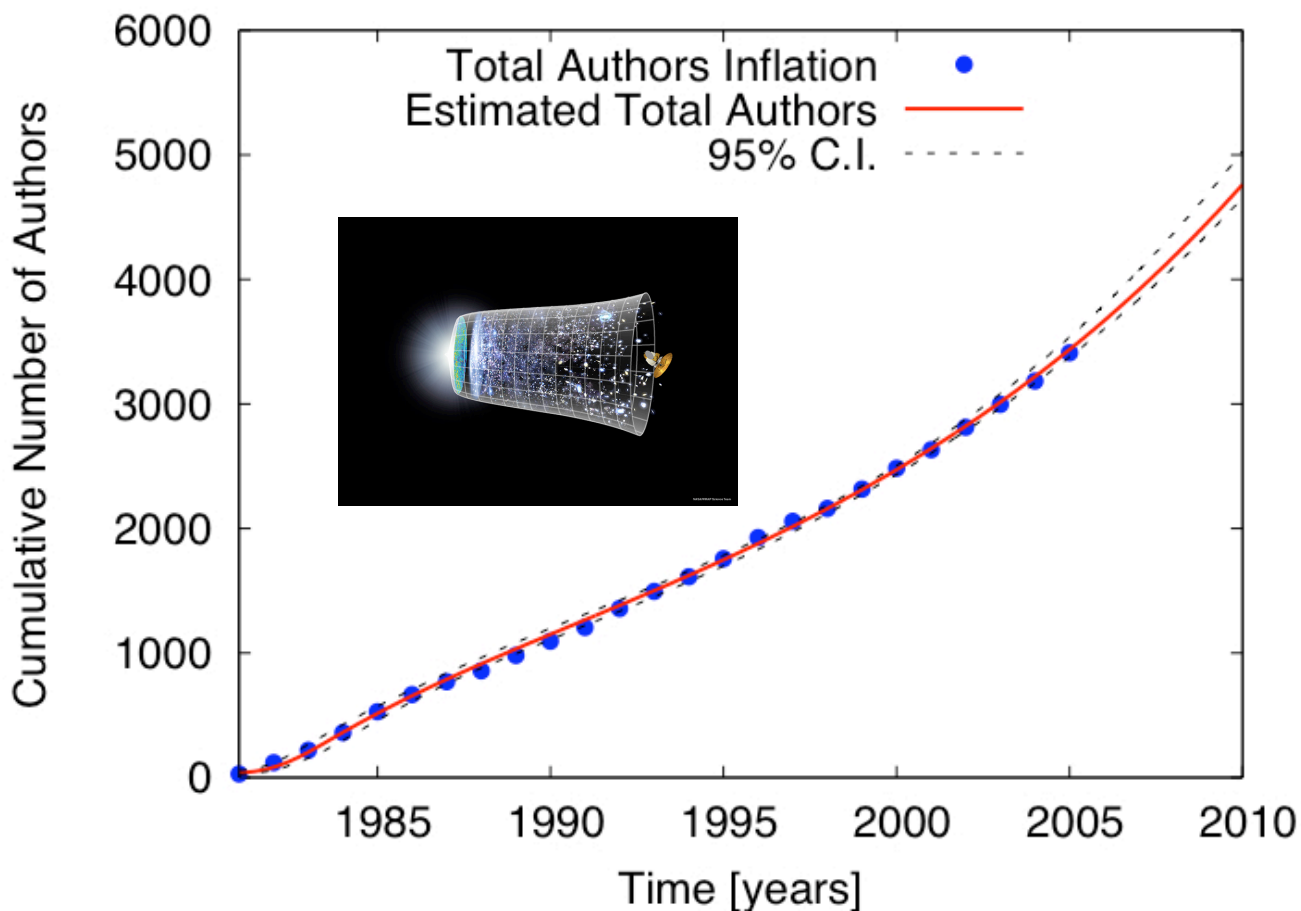
Thus we can associate a (goodness of fit) probability for the trajectory  $I^\Gamma(t)$  as

$$w_\Gamma = \frac{1}{N_w} e^{-A_\Gamma}, \quad N_w = \text{Tr}[w_\Gamma]$$



# Cosmological Inflation

[2005: 3410 authors, 5135 papers]



Alan Guth 1981  
Andrei Linde 1982

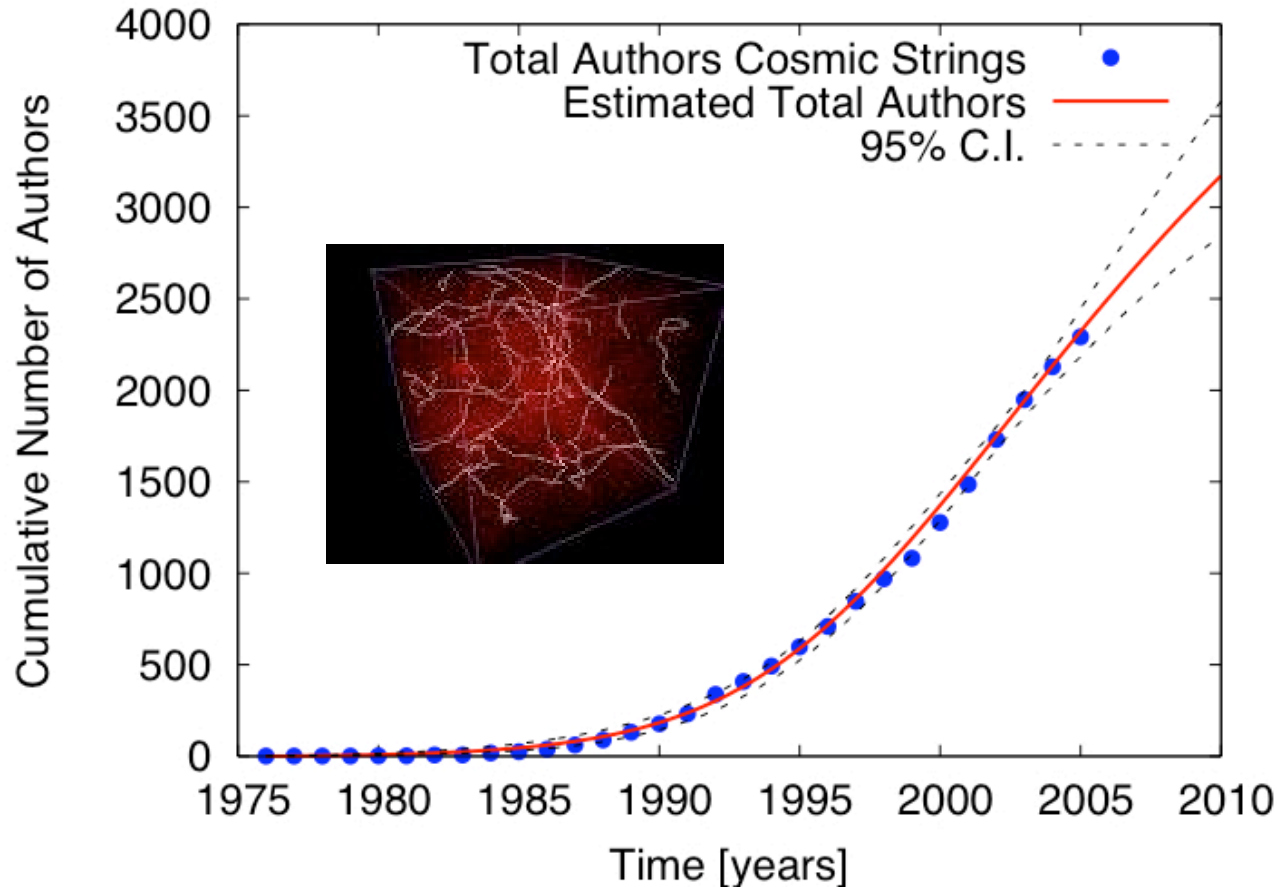


Proposes  
Explanations  
for many  
cosmological  
problems:  
Boosted by recent  
Cosmic Microwave  
Background  
Measurements



# Cosmic Strings and Topological Defects

[2005: 2292 authors; 2443 authors]



TWB Kibble 1976

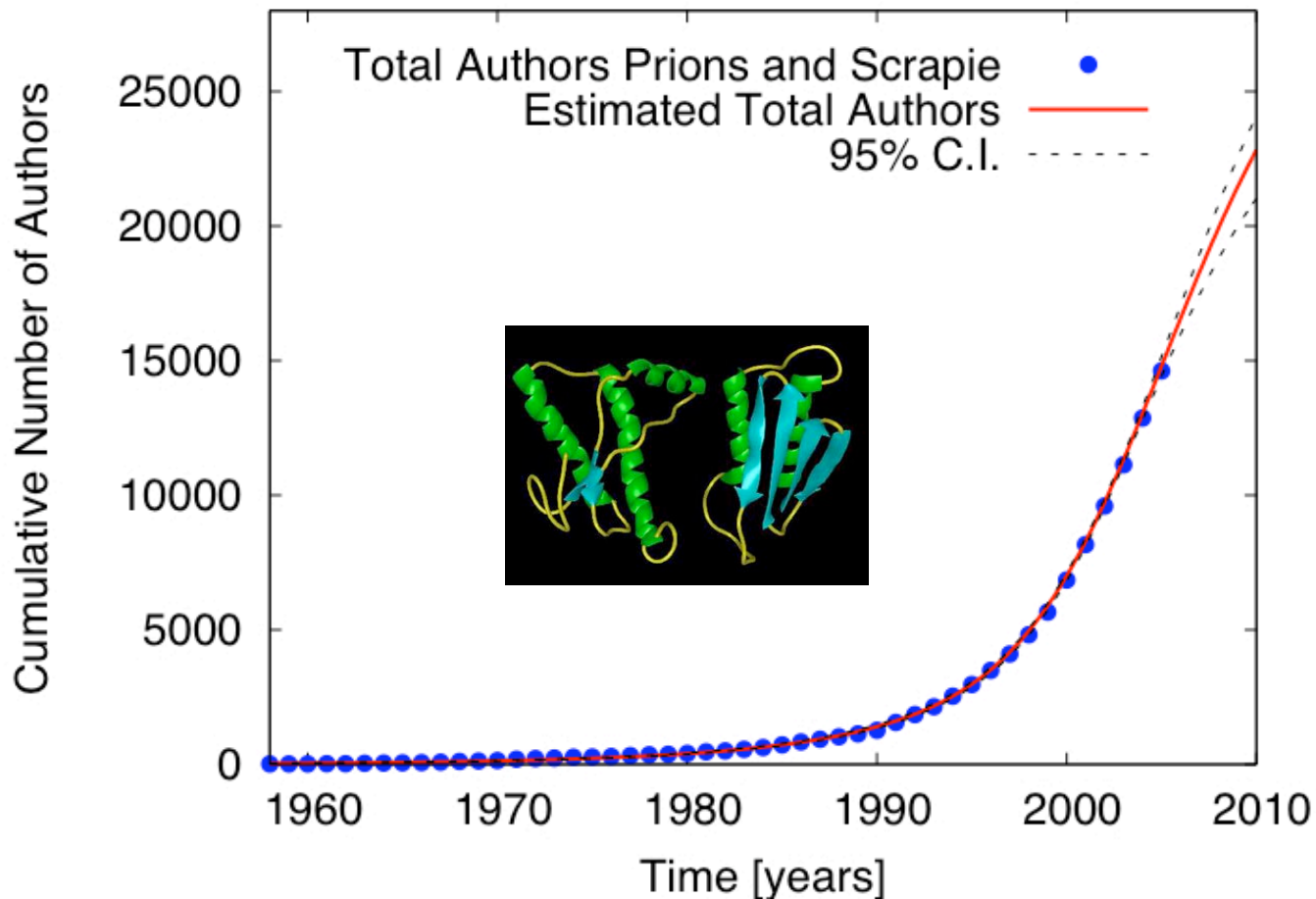
Y Zeldovich 1980



Unavoidable features of the Early Universe:  
Could they have seeded structure?  
Disfavored by Current CMB measurements

# Scrapie and Prions

[2005:14620 authors, 11074 papers]



Prussiner 1982

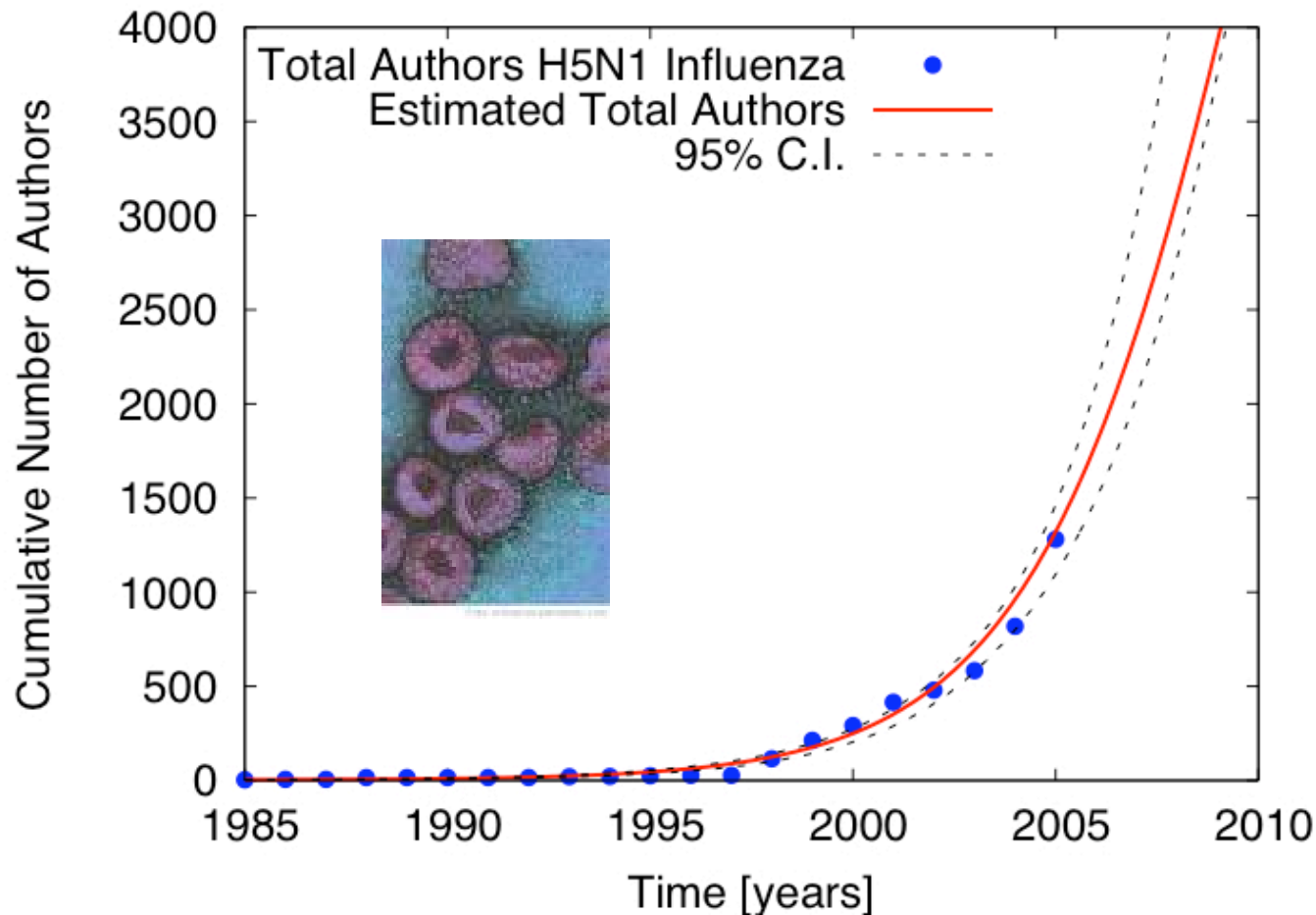
Nobel Prize 1997



Misfolding  
Proteins that  
cause transmissible  
spongiform  
encephalopathies:  
Scrapie,  
“mad cow disease”  
Kreuzberg-Jacob  
disease in humans

# H5N1 Influenza (bird flu)

[2005:1281 authors, 604 papers]



Disease of birds

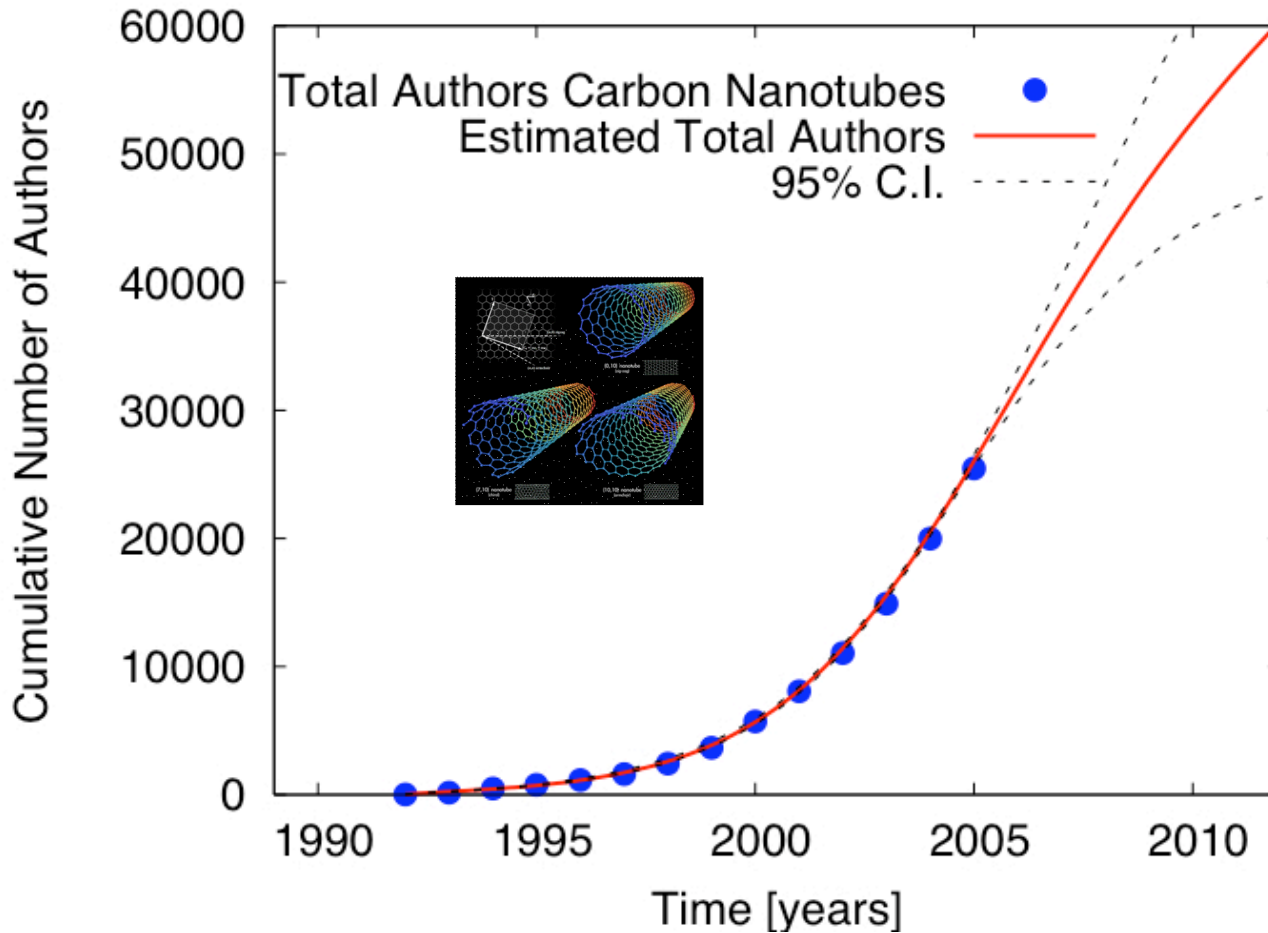
First infected humans in 1997 in Hong Kong

280 humans infected

~60% case mortality

# Carbon Nanotubes

[2005: 25464 authors, 30521 papers]



S. Ijima

1992



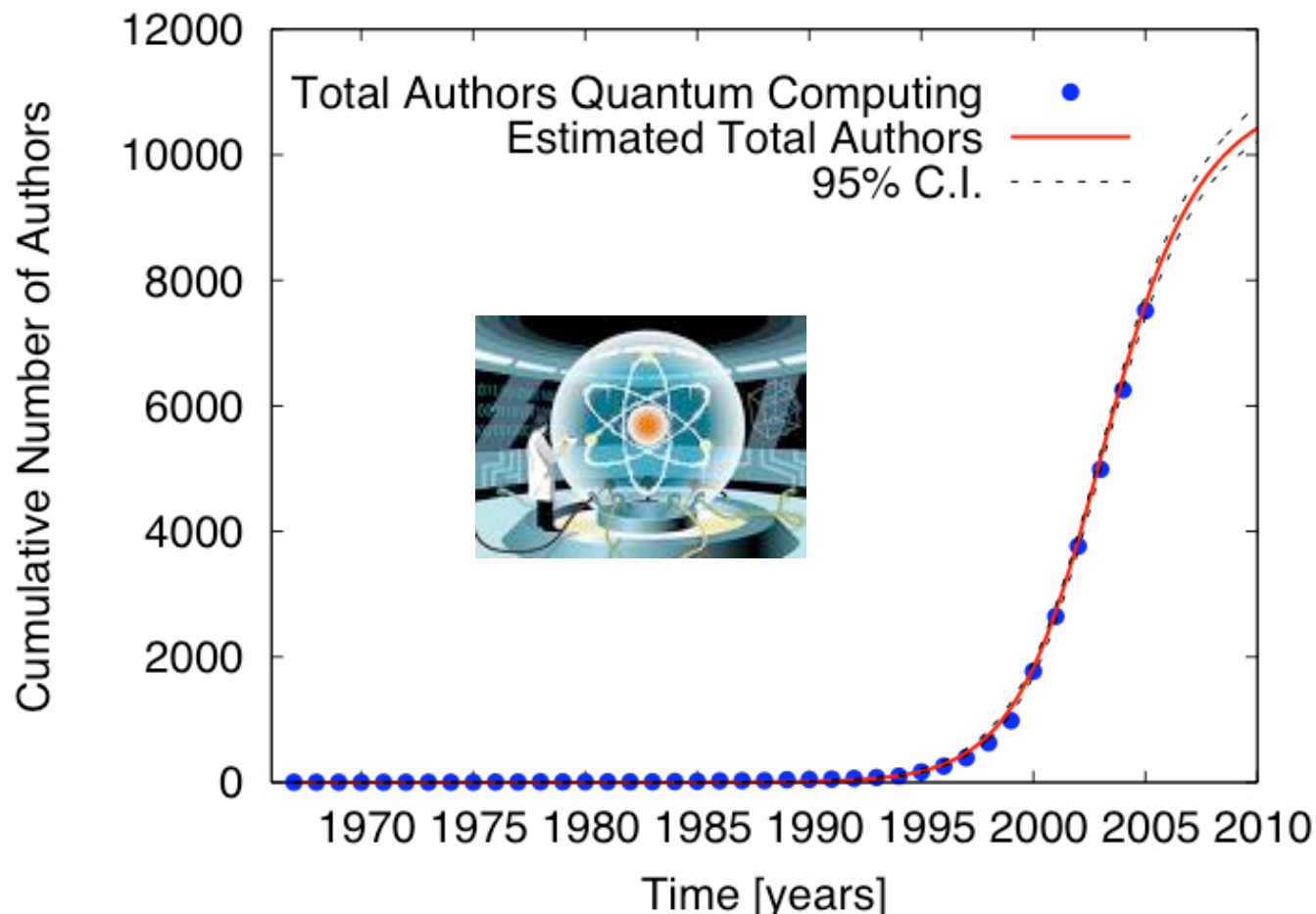
Important subfield  
of nanotech

Allotrope of Carbon

Promises to  
revolutionize  
Nano-engineering

# Quantum Computers and Computation

[2005: 7518 authors; 8946 papers]



First references  
1960s-70s

Feynman 1982  
Deutsch 1985

Algorithms:

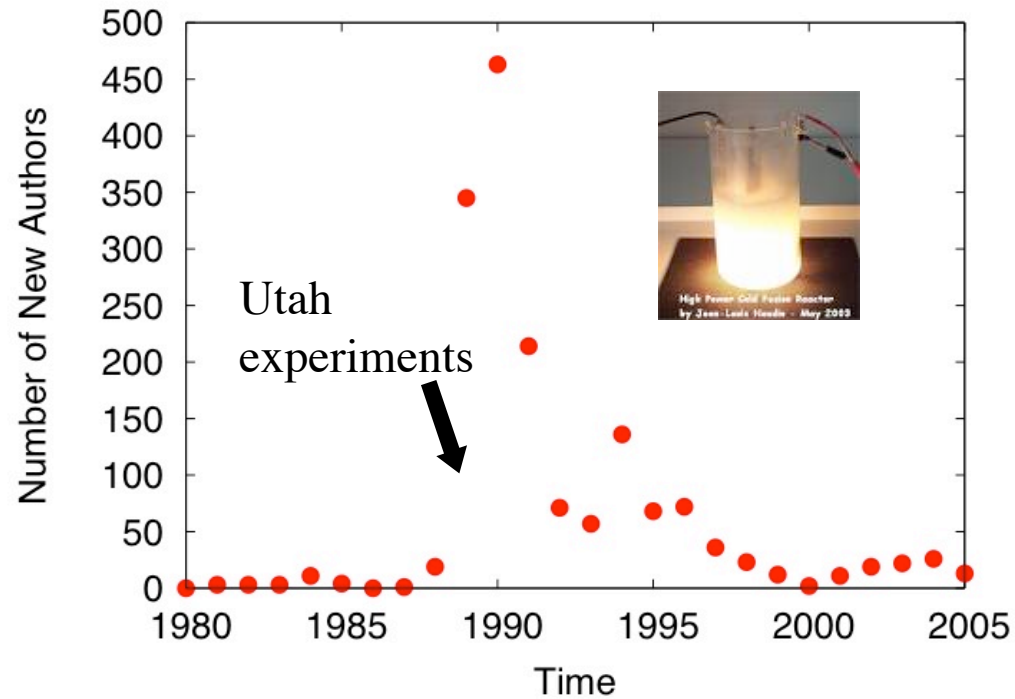
Shor, Grover  
~1995

NMR Experiments  
~1996

Revolution in  
Computing &  
Cryptography?

# Cold Fusion

## pathological science



[2005: 1637 authors; 871 papers]

# Estimated parameters

## infectiousness and recruitment pull of scientific ideas

Parameters	$S(t_0)$	$E(t_0)$	$I(t_0)$	$R(t_0)$	$\beta$	$\Lambda$	$\kappa$	$\rho$	$\gamma$	$R_0$
Cosmological Inflation	930±1	6	37±1	2	13.41±0.28	0.07	0.20	0	0.21	64.6±1.5
Cosmic Strings	14±9	5	0	0	4.45±0.42	159.1±2.7*	0.25±0.02	0	1.73±0.19	2.58±0.11
Prions & Scrapie	14262±1368	1	8±1	7±2	0.69±0.05	469±25*	0.22±0.01	18.4±1.24	0.37±0.03	1.87±0.03
H5N1 Influenza	9057±200	1	0	0	1.47±0.02	138±10*	0.71±0.01	0	0.6±0.01	2.44±0.03
Carbon Nanotubes	30464±5976	501±24	1	1	0.99±0.05	0.04±0.01	0.50±0.03	0.03±0.06	0.10±0.05	9.72±1.71
Quantum Computing	11627±91	0	0	0	3.78±0.09	1.03±0.02**	0.41±0.02	0.77±0.03	1.18±0.02	3.20±0.11

\* Indicates a linear growth term  $\Lambda$ , not  $\Lambda N$  in the equations for  $S$ .

\*\* Susceptible population growth starts in 1990.

# Measures of Scientific Productivity

## Marginal Returns

$$\begin{array}{l} \text{Output} \longrightarrow \\ \text{Input} \longrightarrow \end{array} \frac{\Delta Y(t')}{\Delta X(t)} = f[\Delta X(t)] \sim [\Delta X(t)]^\beta, \quad t' \geq t$$

**scaling relation (?)**

**“Returns to Scale”** in  $\Delta Y$ =Papers *versus*  $\Delta X$ =Authors:

**citations, patents**

**funding, reputation**

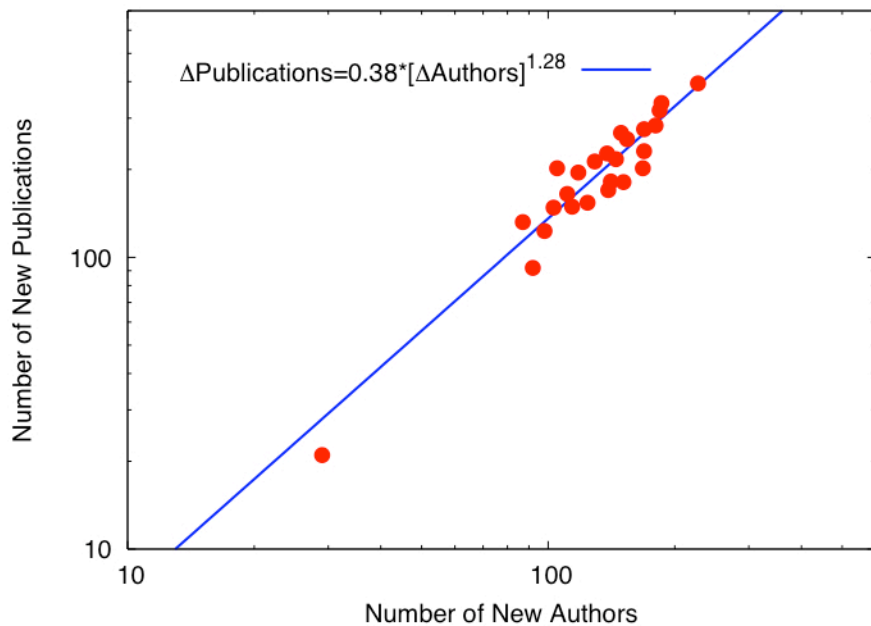
$\beta=1$  : *each unit of input produces one unit of output*

$\beta < 1$  : **diminishing returns**: *each new author -> less papers/author*

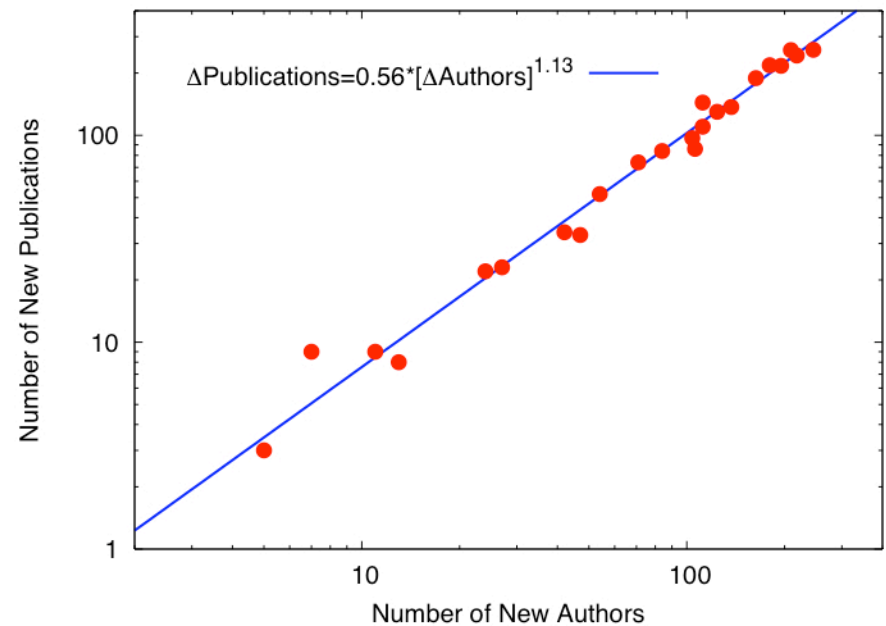
$\beta > 1$  : **increasing returns**: *each new author -> more papers/author*



# Theoretical Physics

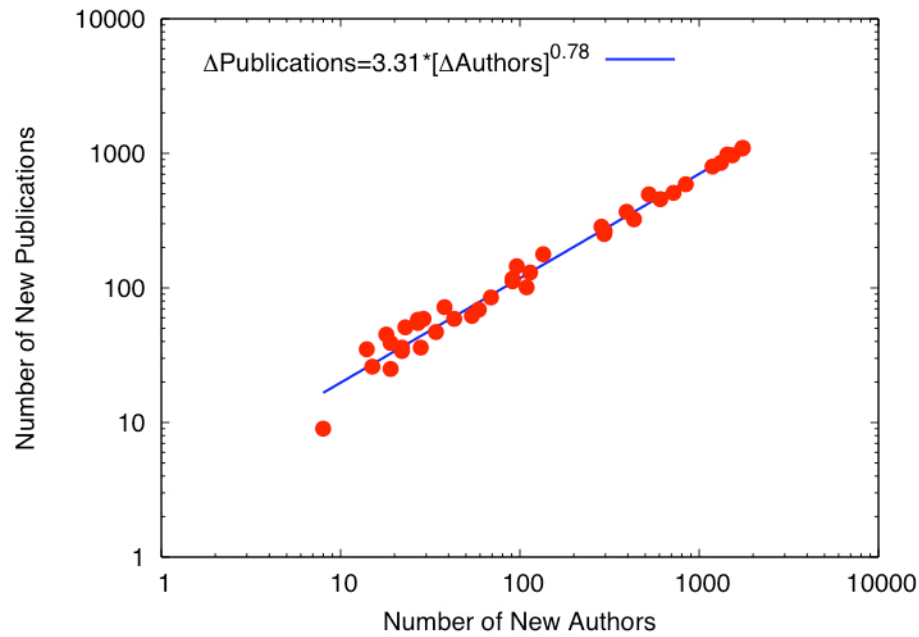


Cosmological Inflation  $\beta=1.28$

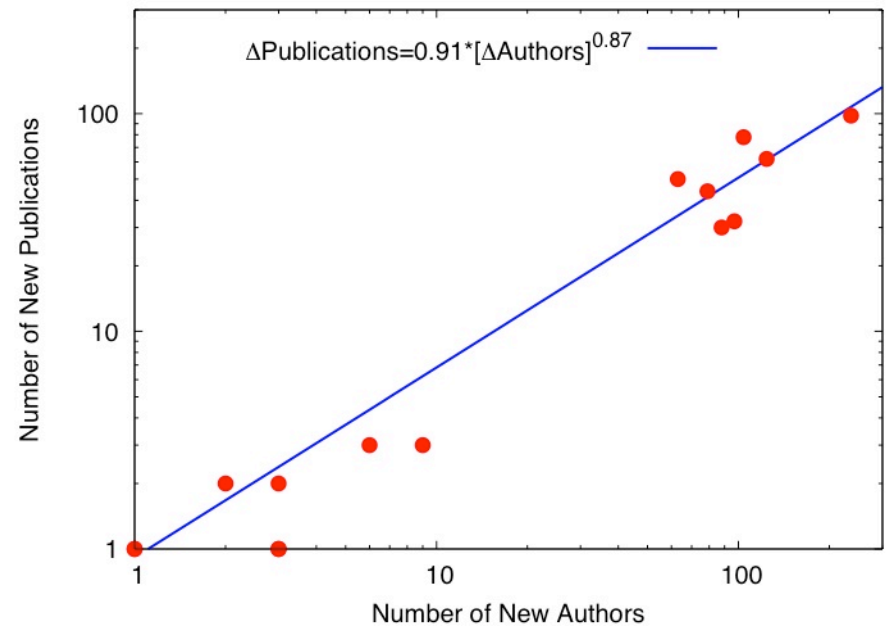


Cosmic Strings  $\beta=1.13$

# BioMedical Fields

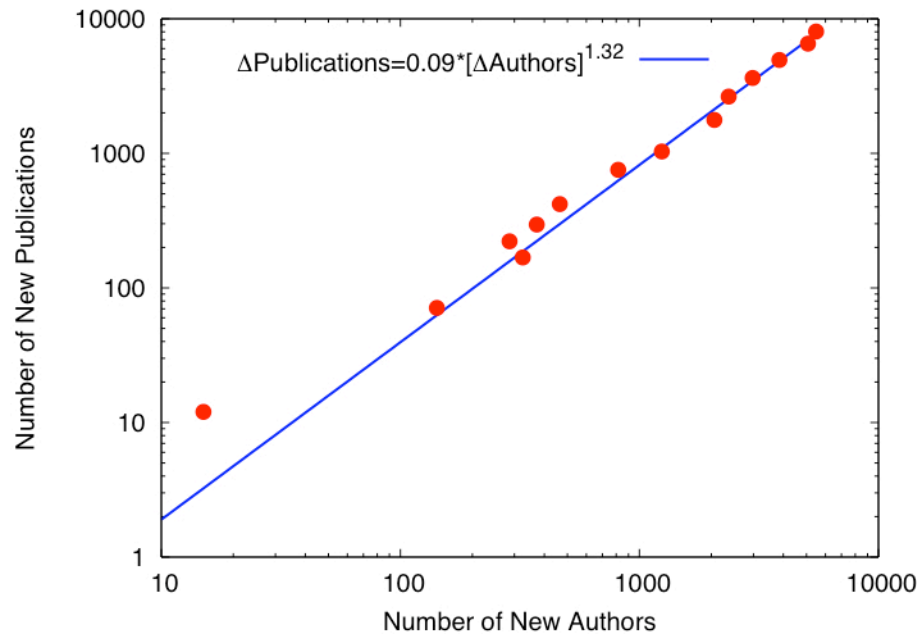


Prions  $\beta=0.78$

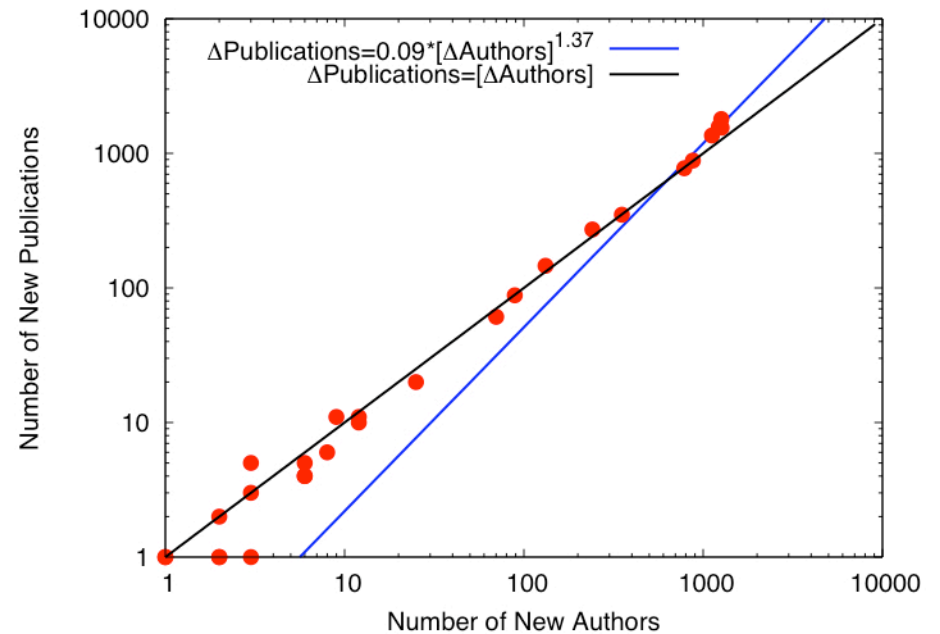


H5N1 Influenza  $\beta=0.87$

# Technological Fields



Carbon Nanotubes  $\beta=1.32$



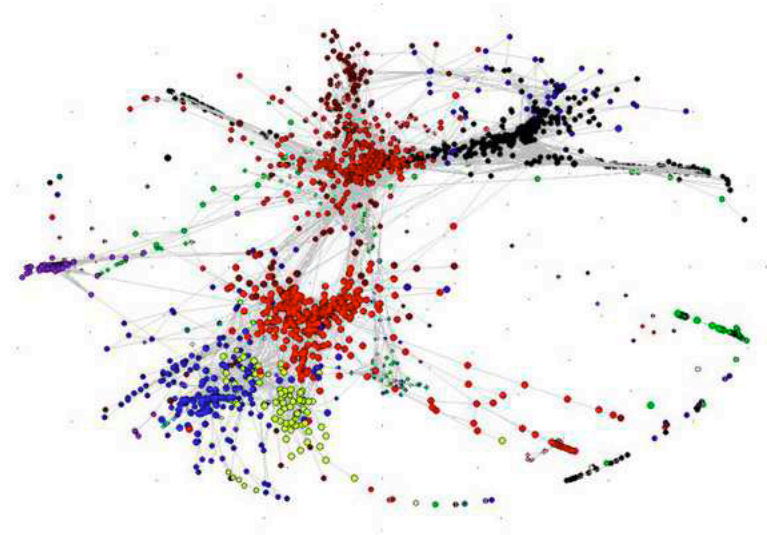
Quantum Computation  $\beta=1$  vs. 1.37

# **Science may be forecast**

uncertainty quantification in  
predictions is essential for model  
building and falsification

# Information processing in the nervous system

functional information modules in  
complex networks



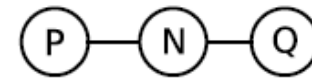
Mouse liver gene expression network from Jake Lusis Laboratory, UCLA

# Functional subgraphs

Types of functional units (building blocks)

- Redundant chains

(Bettencourt et al., Schneidman et al.)

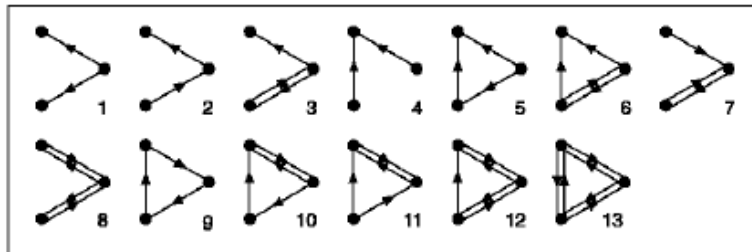


- Synergetic circuits

(Schneidman et al., Bettencourt et al., Gross et al.)



- Motifs



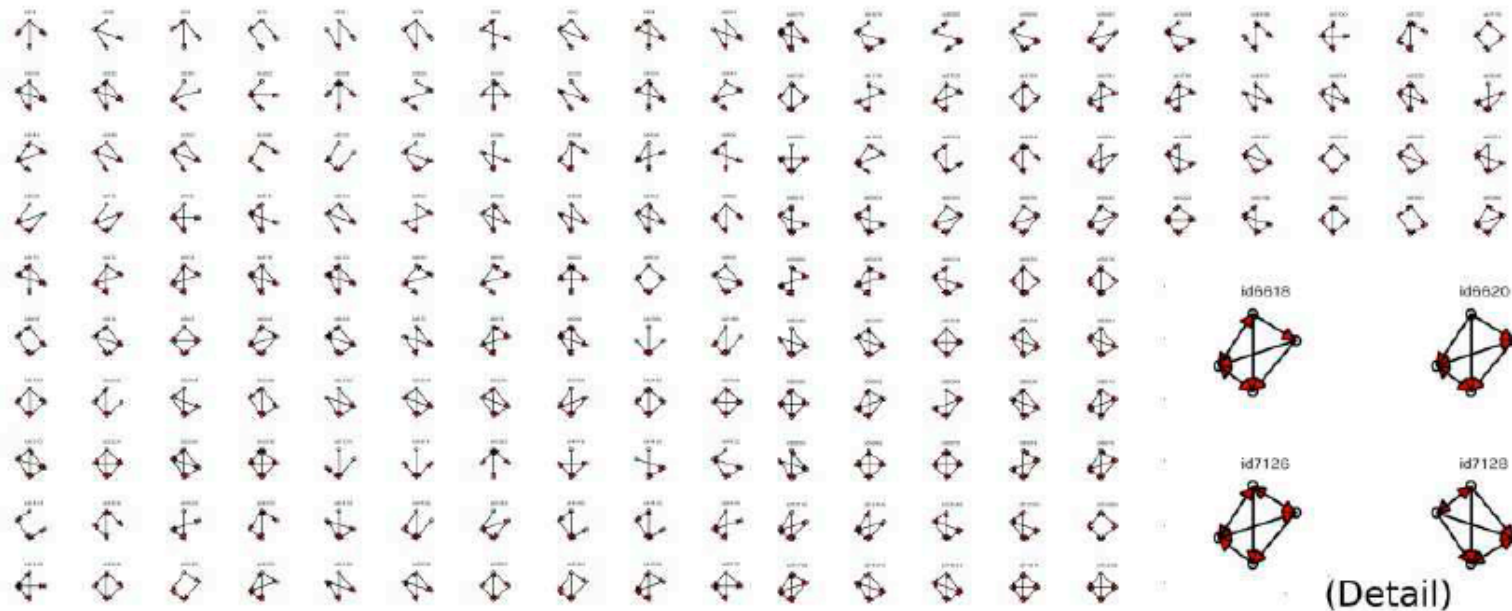
(Middendorf et al., Milo et al.)

(Image from Milo et al.)

# Curse of dimensionality motifs

There are too many!

(199 4-motifs, 9,364 5-motifs, 1,530,843 6-motifs, etc.)



(Image from Milo et al.)

# Entropy as uncertainty

## Mutual Information as uncertainty reduction

Shannon Entropy of X: 
$$S(X) = -\sum_x p(x) \log_2 p(x)$$

**measures number of states of X; stochasticity**

Shannon Entropy of {X;Y}: 
$$S(X;Y) = S(X) + S(Y) + I(X;Y)$$

Mutual Information {X;Y}: 
$$I(X;Y) = \sum_{x,y} p(x,y) \log_2 \left( \frac{p(x,y)}{p(x)p(y)} \right)$$

**measures correlation between states of X;Y**



# A [discrete] calculus in information measurement and information gain

Define the discrete calculus of the entropy under an additional measurement as

$$\frac{\Delta S(X)}{\Delta Y_i} \equiv S(X|Y_i) - S(X)$$

$$\begin{aligned} \frac{\Delta^2 S(X)}{\Delta Y_i \Delta Y_j} &\equiv \frac{\Delta}{\Delta Y_i} \left[ S(X|Y_i) - S(X) \right] \\ &= S(X|Y_i, Y_j) - S(X|Y_i) - S(X|Y_j) + S(X), \end{aligned}$$

...

# A cluster decomposition in terms of functional modules

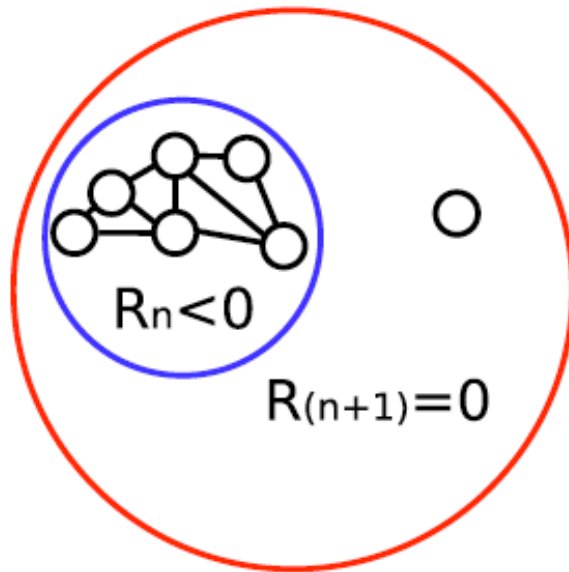
For a given set  $\{Y_1, \dots, Y_n\}$ ,

$$\begin{aligned} I(X, \{Y_1, Y_2, \dots, Y_n\}) &= S(X|Y_1, Y_2, \dots, Y_n) - S(X) \\ &= \sum_i \frac{\Delta S(X)}{\Delta Y_i} + \sum_{i>j} \frac{\Delta^2 S(X)}{\Delta Y_i \Delta Y_j} + \dots + \frac{\Delta^n S(X)}{\Delta Y_1 \dots \Delta Y_n} \end{aligned}$$

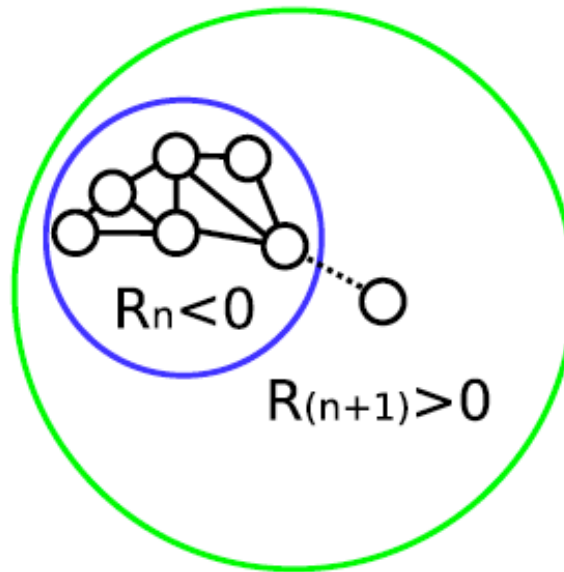
- This decomposition is analogous to a Taylor series.
- Each term isolates the pairs, triplets, etc.
- Define  $R_n(X, Y_{i_1}, \dots, Y_{i_n}) \equiv \frac{\Delta^n S(X)}{\Delta Y_{i_1} \dots \Delta Y_{i_n}} \neq R_n^S(X, Y_{i_1}, \dots, Y_{i_n})$

# $R_n$ gives redundancy or synergy exactly to $n^{\text{th}}$ order

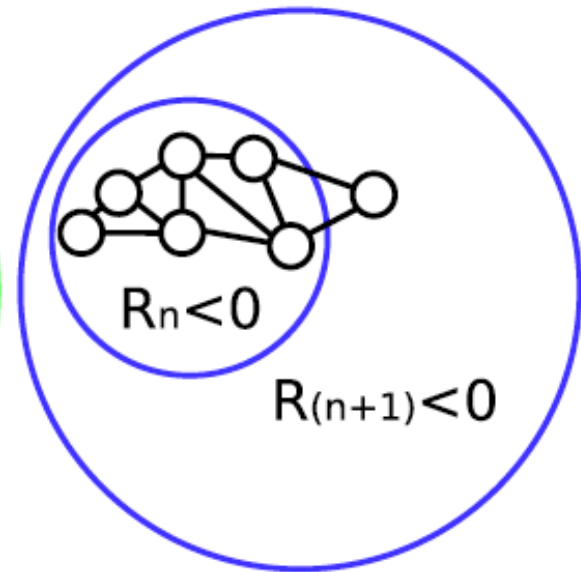
Independent



Redundant



Synergetic



Example:

- Start with set of  $n$  nodes with  $R_n < 0$
- Add one more node
- $R_{n+1}$  gives the relationship of that node to the previous set

# Architecture and information processing in the nervous system

Frontal cortex neurons  
from fetal mice  
[thousands/mm<sup>2</sup>]

Grown in vitro over a  
1mm<sup>2</sup> microelectrode array

Disassociated Culture  
spontaneously form network

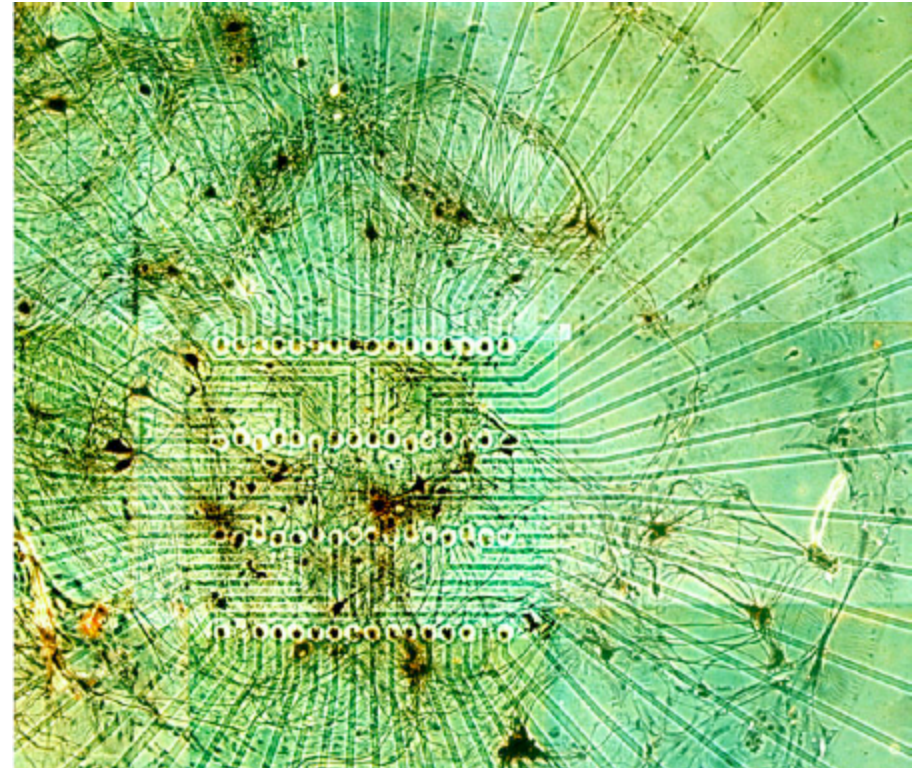
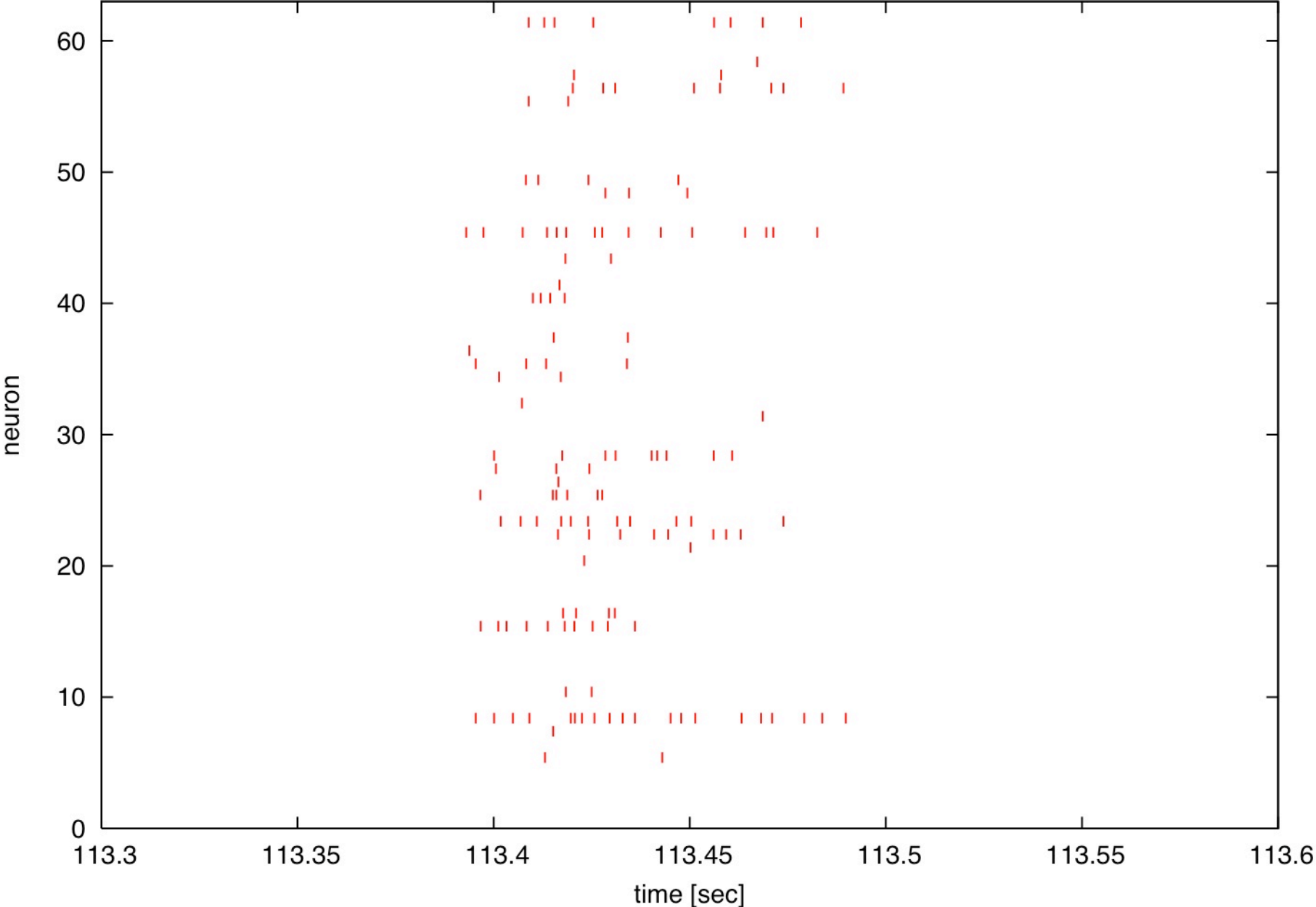


Image courtesy M. Ham and G. Gross

# Cortical neural network electrophysiological activity

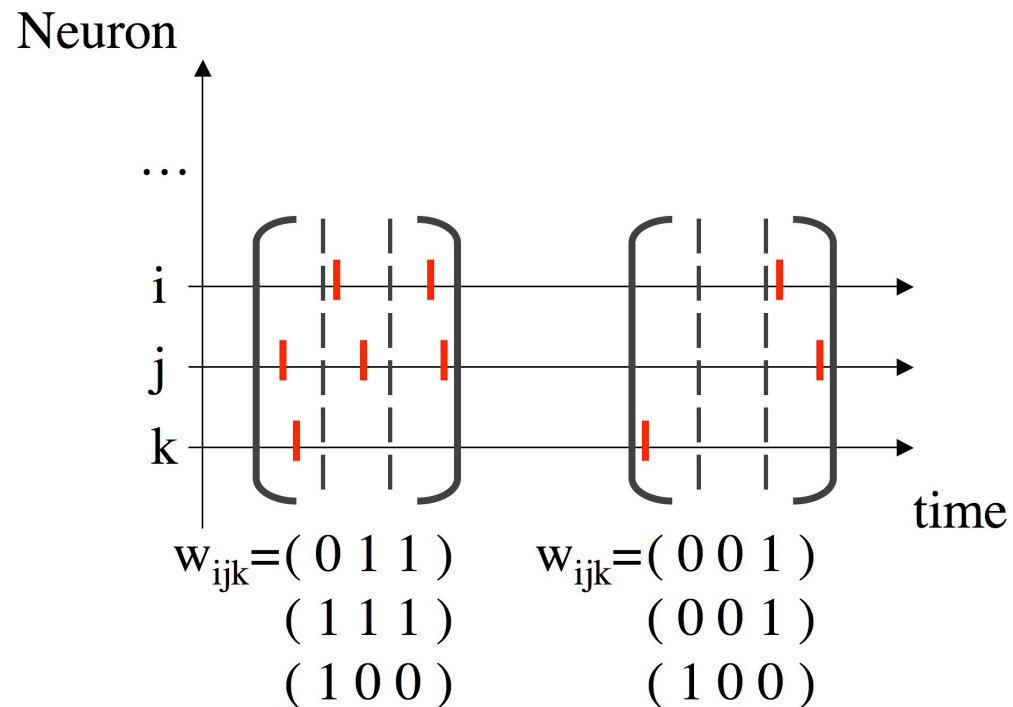


# Estimation in practice

## Binary 'words' from spike time series

"Spikes"

Rieke, Warland, de Ruyter van Steveninck, Bialek



And count word frequencies over time  $p_w = \frac{n_w}{N}$

# Motif search and identification optimization in uncertainty reduction

Maximize the unique information gain  $R_n$  one order at a time.  
Conveniently, the expansion

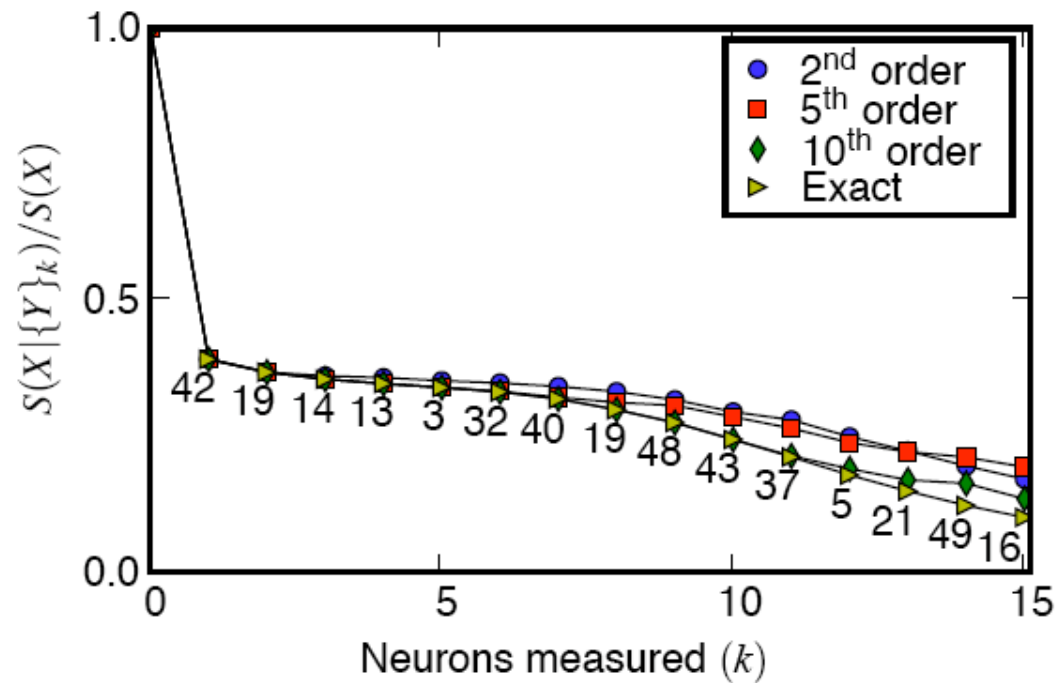
$$S(X|Y_1, Y_2, \dots, Y_n) - S(X) = \sum_i \frac{\Delta S(X)}{\Delta Y_i} \\ + \sum_{i>j} \frac{\Delta^2 S(X)}{\Delta Y_i \Delta Y_j} + \dots + \frac{\Delta^n S(X)}{\Delta Y_1 \dots \Delta Y_n}$$

allows us to do so order by order!

This is analogous to other optimization strategies such as the method of steepest descent.

# Approximate searches

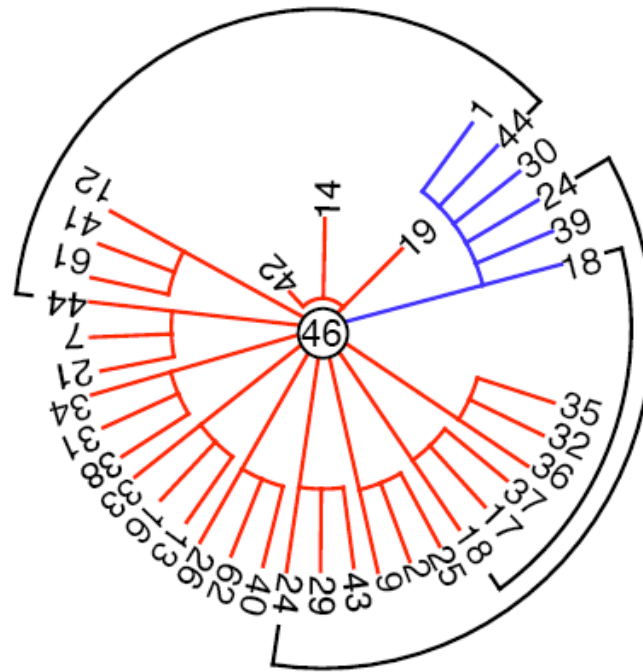
- Want to maximize  $I(X; \{Y\}_n)$  by choosing best set  $\{Y\}_n$
- Computationally expensive for  $n > 10$
- Instead truncate expansion to  $k < n$  and maximize
- How does the set  $\{Y\}_n$  found using the approximation compare to the set found by using the exact expression?





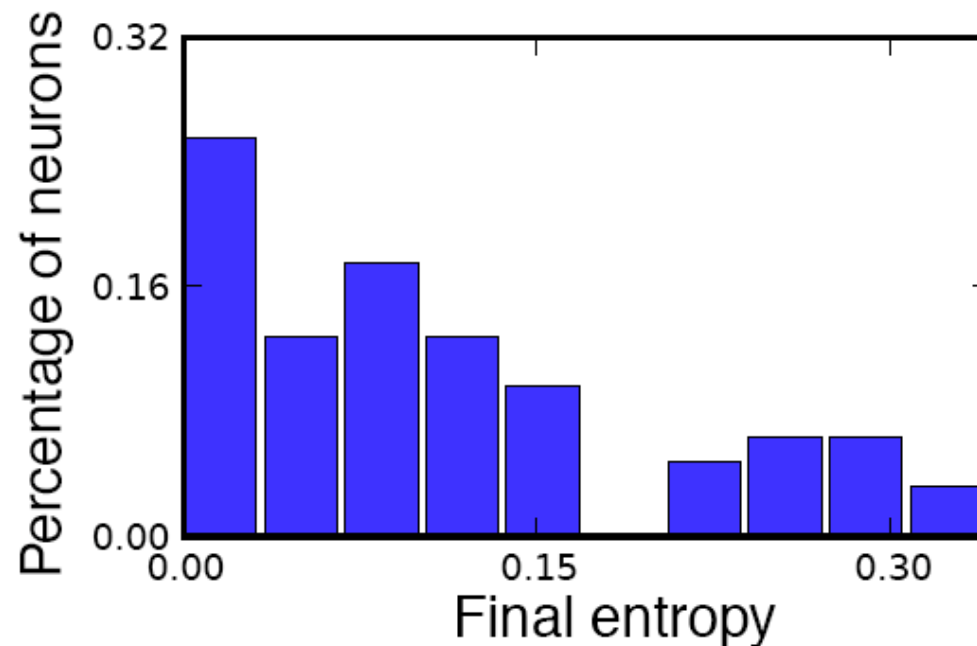
# Reverse engineering network circuits

Seek purely redundant and synergetic “cores” – a set of neurons and all possible subsets that share the same functional character



# Randomness or Structure?

Individual uncertainty is accounted for by other nodes



After all available neurons are measured, very little (0 – 30%) of each neuron's initial entropy remains!

# Uncertainty in models of complex systems

## Uncertainty quantification and management

is essential in complex systems

no [exact] predictive models exist  
many uncertainties in initial conditions and parameters  
exogenous shocks

## Uncertainty reduction via optimization reveals

the **functional network structure** of complex systems  
as information processing systems

generates **robust adaptive control protocols**, active learning and recovery