The power of a good idea

Predictive population dynamics of scientific discovery and information diffusion

Luís M. A. Bettencourt
Theoretical Division
Los Alamos National Laboratory

David I Kaiser (MIT)
Jasleen Kaur (Indiana)
Carlos Castillo-Chavez (ASU)

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

1943 - 2003 Los Alamos NATIONAL LABORATORY
Ideas That Change the World

T Theoretical Division

a predictive science of science?

Can the course of science be forecast?

-when is a field opening or closing?

-what are the signatures of new scientific discoveries?

-"Paradigm shifts" vs. "normal science"

can it be measured from streaming data?

Prediction enables interventions:

How should agencies and institutions allocate resources:
Students? Meetings? Individual PIs?

How can scientific discovery be accelerated?

Predictive Models

Agent based models:

Detailed but difficult for data assimilation

(too) many parameters?

Statistical (and network models)

Issue of non-stationarity

Population Models:

better suited for estimation

motivation of ideas as epidemics

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

Current Comments

The Epidemiology of Knowledge and the Spread of Scientific Information

Number 35   September 1, 1980

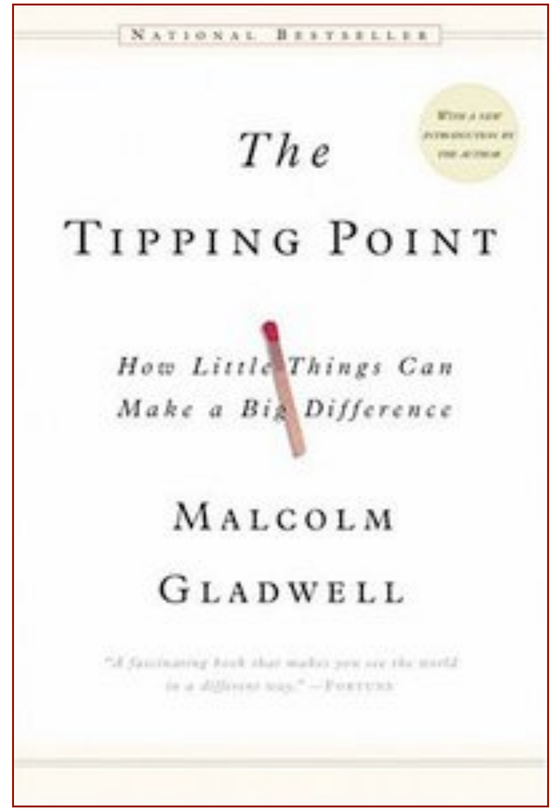
article

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



Los Alamos NATIONAL LABORATORY 1943 - 2003
Ideas That Change the World

T Theoretical Division

Parallels between social dynamics and epidemiology

Individual

Social (population)

Host/pathogen dynamics

Susceptible

Exposed

Infectious

Recovered

subpopulation in classes
contact rate
incubation time
infectious period

no intentionality in standard disease contagion

1943 - 2003 Los Alamos NATIONAL LABORATORY
Ideas That Change the World

T Theoretical Division

Population States

The concepts of:

- Susceptible: can acquire the idea
- Exposed [but not infectious]: knows the idea, may traning to acquire it but cannot yet transit
- Infectious: has acquired the idea and can transmit it to Susceptibles or Exposed.

Generalize naturally

from epidemics to the social transmission of ideas

Population States II

- Immune Recovery does not occur for ideas
- Competing strands or antagonists may reduce the spread of an idea:

Competition/distraction ~ Immunization

Disease	**Ideas**
Susceptibles → Recovered	Susceptibles → Recovered/Other
Vaccination	Competition/Distraction

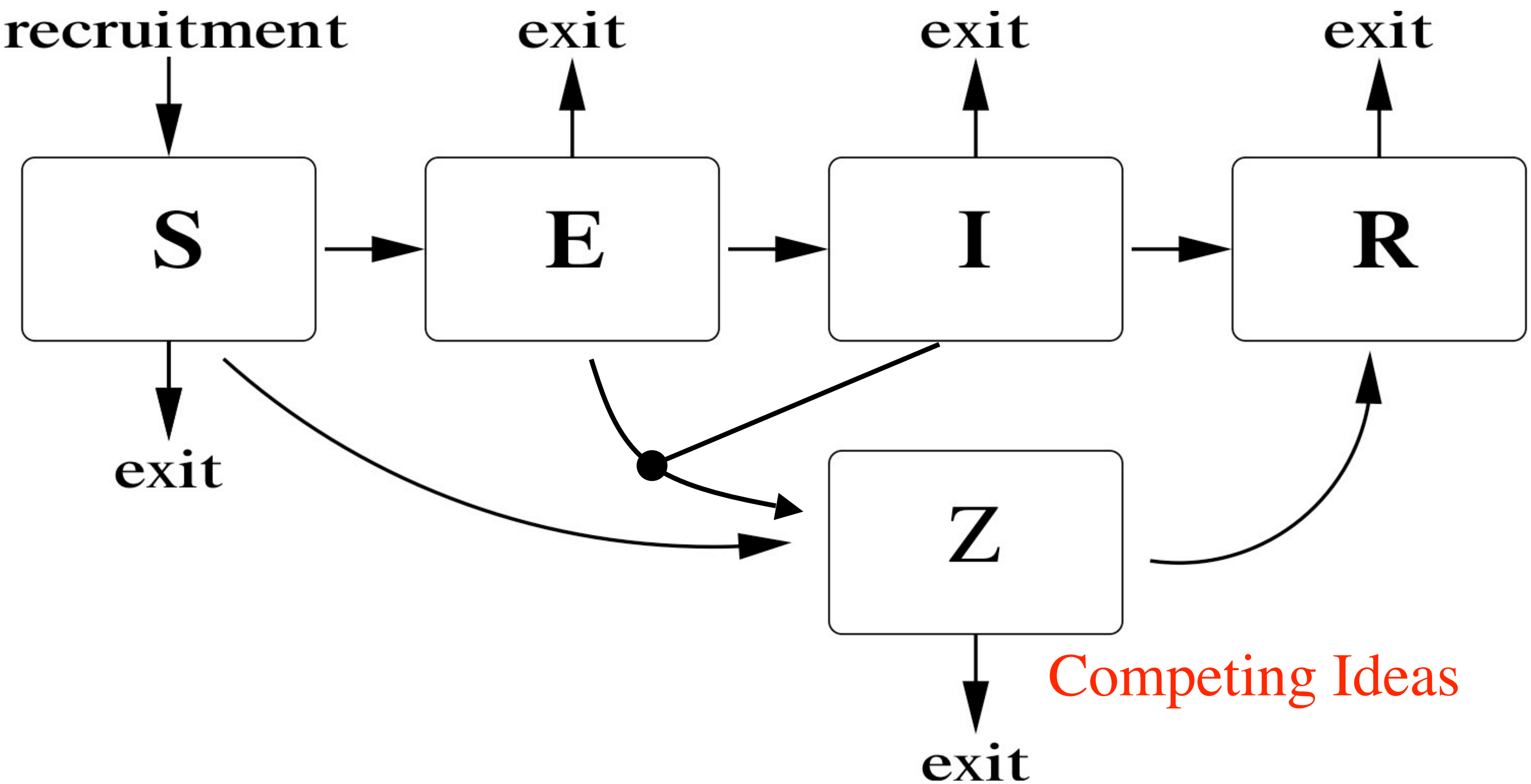
Intentionality is essential in social processes

E.g. the spread of (scientific) ideas:

Ideas are **desirable** to acquire	intentionally extended training (PhD, postdocs)
They require effort and training	intense repeated contacts
They may never be forgotten	structures to prolong memory reservoirs: papers, libraries

Ideas appear as hard to catch diseases characterized by small contact rates and long infectious periods i.e typically large R, but slow dynamics

Population compartment models



Competing Ideas

The advent and spread of Feynman Diagrams
(1948-54)

- Quantum Electrodynamics is being developed independently - Feynman (Cornell), Schwinger (Harvard), and Tomonaga (Tokyo)
- Early 1947: these formulations seem independent, particular and possibly incomplete [renormalization]
- Feynman introduces diagrams in the Pocono conference 1947. Reaction: “completely baffling”
- 1947: Dyson spends the year at Cornell with Bethe, Feynman
- Early Summer 1948: Feynman and Dyson drive together from Cleveland to Albuquerque
- Summer 1948: Dyson attends lectures by the “great Schwinger” at Ann Arbor Conference
- August 1948: Dyson finishes his unifying paper and gives precise meaning to diagrams
- Fall 1948: Dyson comes to IAS as a postdoc
- Winter 1948 : Diagrams are eventually accepted by Oppenheimer and spread from the IAS to the rest of the community in the US and abroad

TABLE 8–2

The correspondence between diagrams and S-matrix elements in momentum space

Component of Diagram	Factor in S-Matrix Element	
Internal photon line	$g_{\nu\lambda}\dfrac{1}{k^2 - i\mu}$	photon propagation function
Internal electron line	$\dfrac{i\not{p} - m}{p^2 + m^2 - i\mu}$	electron propagation function
Corner	$\gamma^\nu\delta(p - p' - k)$	
External photon lines	$\dfrac{1}{\sqrt{2\omega}}e_\mu(\mathbf{k}),\ \dfrac{1}{\sqrt{2\omega}}e_\mu(\mathbf{k})$	ingoing and outgoing photons
External negaton lines	$\sqrt{\dfrac{m}{\epsilon}}\,u(\mathbf{p}),\ \sqrt{\dfrac{m}{\epsilon}}\,\bar{u}(\mathbf{p})$	ingoing and outgoing negatons
External positon lines	$\sqrt{\dfrac{m}{\epsilon}}\,\bar{v}(\mathbf{p}),\ \sqrt{\dfrac{m}{\epsilon}}\,v(\mathbf{p})$	ingoing and outgoing positons

The “Feynman Rules” in the momentum-space representation, following Dyson’s prescriptions. Reproduced from J.M. Jauch and F. Rohrlich, *The Theory of photons and Electrons* (Cambridge: Addison-Wesley, 1955), 154.

Earliest “Feynman Graphs” from Dyson’s 1949 Papers

F.J. Dyson PR **75** 486 (1949)

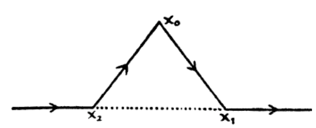


FIG. 1.

F.J. Dyson PR **75** 1736 (1949)

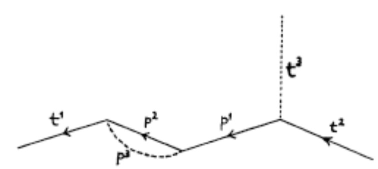
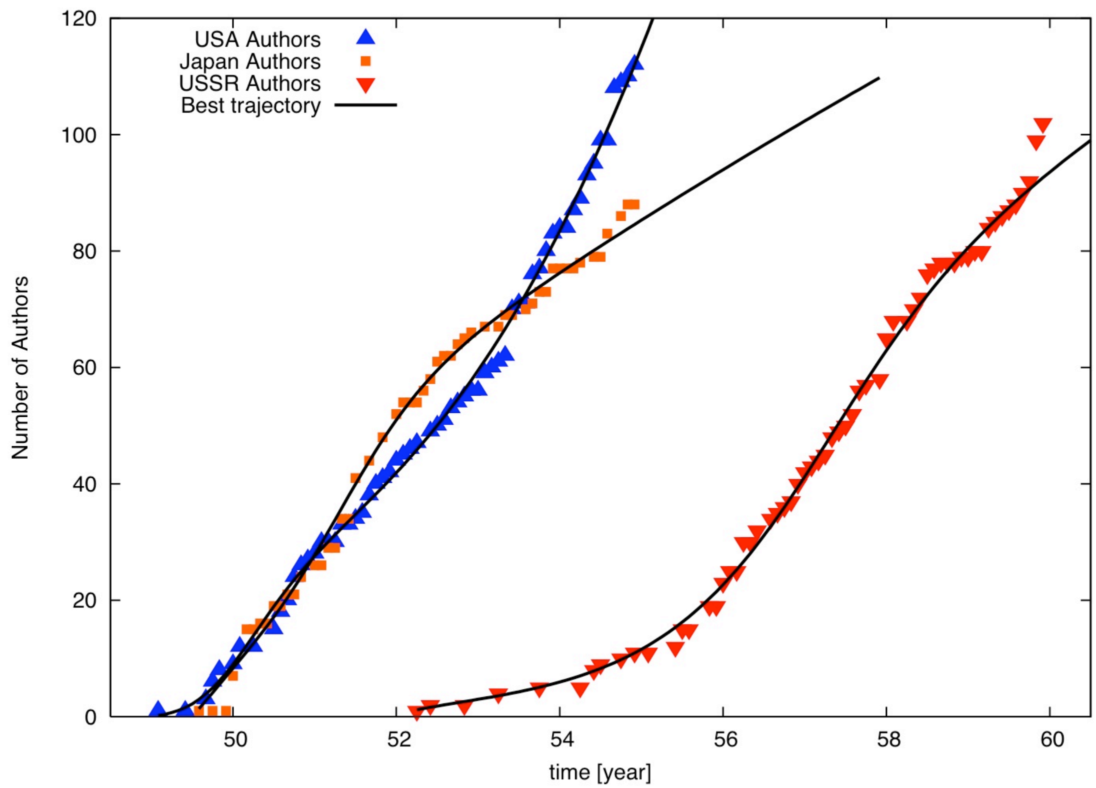


FIG. 3.

1943 - 2003
Los Alamos
NATIONAL LABORATORY
Ideas That Change the World

The spread of Feynman Diagrams

USA, Japan, USSR



Los Alamos NATIONAL LABORATORY 1943 - 2003 Ideas That Change the World

T Theoretical Division

Parameter Estimates

Parameter	Best-fit	Mean	Std
USA			
$S(t_0)$	478.515	398.691	61.990
$E(t_0)$	60.989	44.686	4.728
$I(t_0)$	0.020	0.160	0.135
ε	0.257	0.391	0.055
β	1.041	0.951	0.086
μ	0.025	0.040	0.012
Λ	45.385	40.052	6.467
R_0	37.711	23.172	5.798
Japan			
$S(t_0)$	30.248	31.037	2.190
$E(t_0)$	11.569	12.022	1.400
$I(t_0)$	0.153	0.165	0.129
ε	2.361	2.009	0.279
β	5.956	4.417	0.787
μ	0.039	0.044	0.013
Λ	12.067	12.578	1.093
R_0	150.136	105.372	35.223
USSR			
$S(t_0)$	3.074	0.810	0.722
$E(t_0)$	3.344	3.462	0.647
$I(t_0)$	0.682	0.738	0.266
ε	1.713	1.613	0.476
β	3.715	3.589	0.753
μ	0.067	0.075	0.035
Λ	17.819	19.372	3.668
R_0	53.257	55.892	28.788

Modeling the dynamics of scientific discovery

Rationale:

- Identify the birth and development of scientific fields
- Extract their temporal dynamics [papers, authors,…]
- Extract the characteristics of their social networks
  [recruitment, densification,
  components of collaboration]

Is there something special - dynamically and structurally - to the emergence of scientific fields?

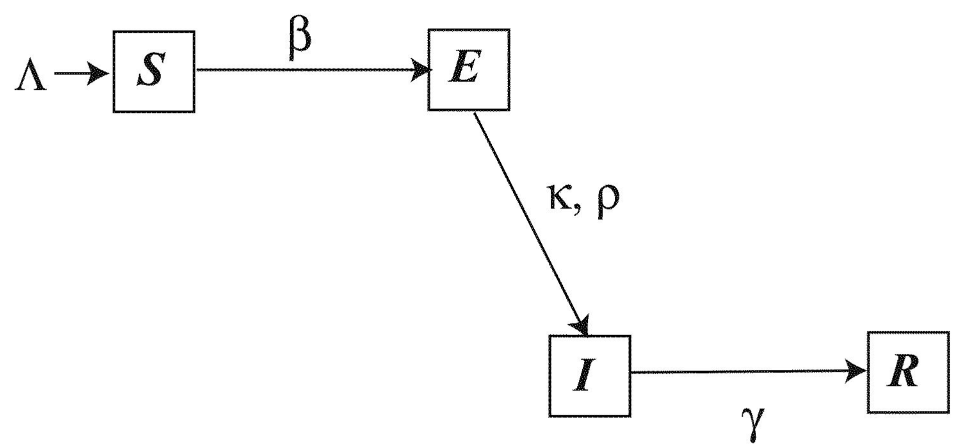
Dynamical Model

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

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

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

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



$R_0 = \beta/\gamma$ 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 solution from which we construct:

Joint probability distribution for model parameters conditional on observed data:

$$P[\Gamma|_{I^o}]$$

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

"I remember my friend Johnny von Neumann used to say, 'with four parameters I can fit an elephant and with five I can make him wiggle his trunk.'"
A meeting with Enrico Fermi, Nature **427**, 297; 2004.

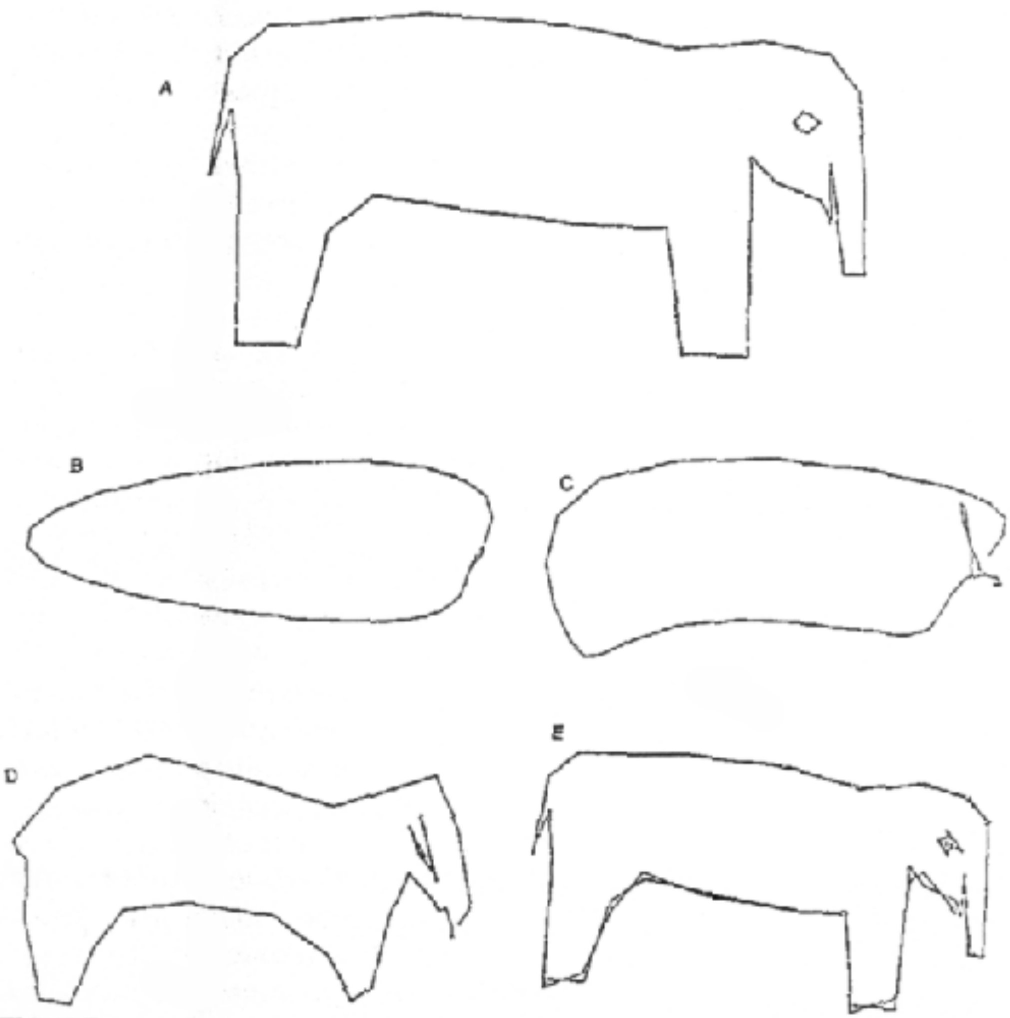


FIGURE 1.2. "How many parameters does does it take to fit an elephant?" was answered by Wel (1975). He started with an idealized drawing (A) defined by 36 points and used least squares Fourier sine series fits of the form $x(t) = \alpha_0 + \sum \alpha_i \sin(it\pi/36)$ and $y(t) = \beta_0 + \sum \beta_i \sin(it\pi/36)$ for $i = 1, \ldots, N$. He examined fits for $K = 5, 10, 20$, and 30 (shown in B–E) and stopped with the fit of a 30 term model. He concluded that the 30-term model "may not satisfy the third-grade art teacher, but would carry most chemical engineers into preliminary design."

1943 - 2003 Los Alamos NATIONAL LABORATORY
Ideas That Change the World

T Theoretical Division

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

Deviation (action):

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

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

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

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

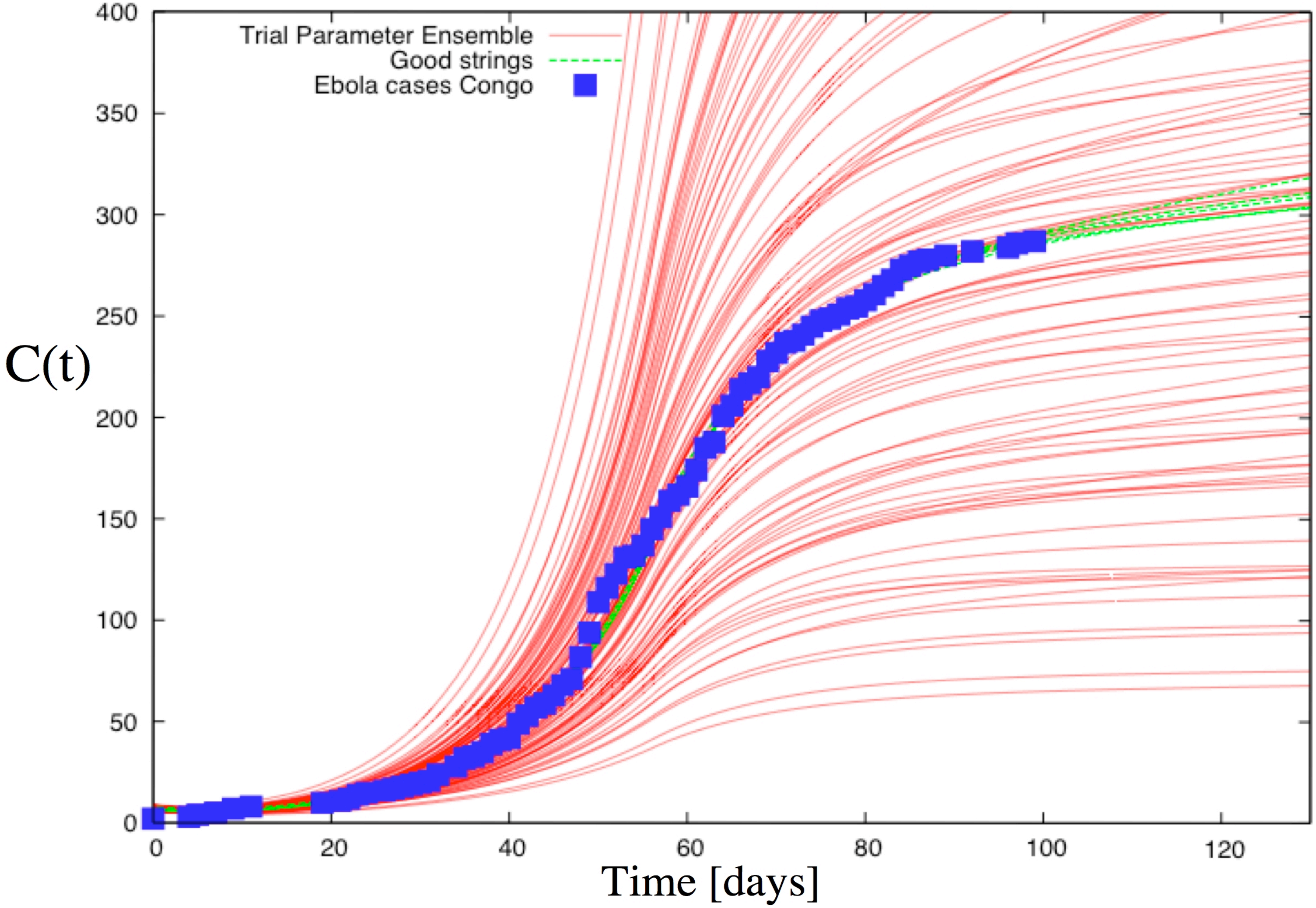
Ensemble Estimation in practice:

The joint probability distribution is estimated from
an ensemble of trajectories:

$$P[\Gamma \,|\, I^O] = \sum_{i=1}^{\infty} \delta(\Gamma - \Gamma_i) w_{\Gamma_i}$$

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

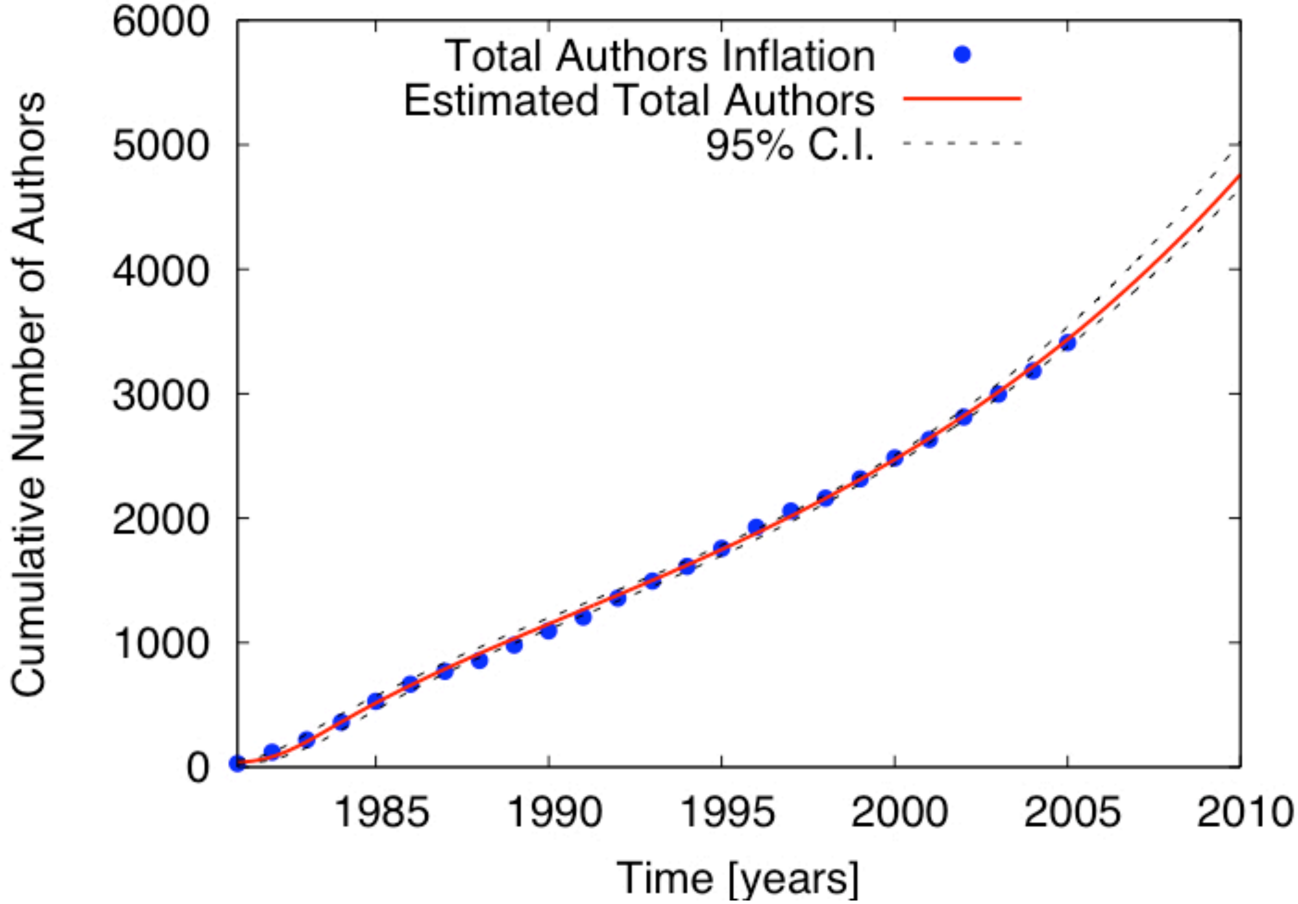
In practice this expression can be used as an *estimator* for a finite sized ensemble of N_S realizations.


Trial Parameter Ensemble
Good strings
Ebola cases Congo
C(t)
Time [days]
0
50
100
150
200
250
300
350
400
0
20
40
60
80
100
120

Six examples of scientific discovery

Cosmological Inflation Cosmic Strings	Theoretical Physics
Prions H5N1 Influenza	BioMedical
Quantum Computing & Computation Carbon Nanotubes	Applied Physics Material Science Engineering

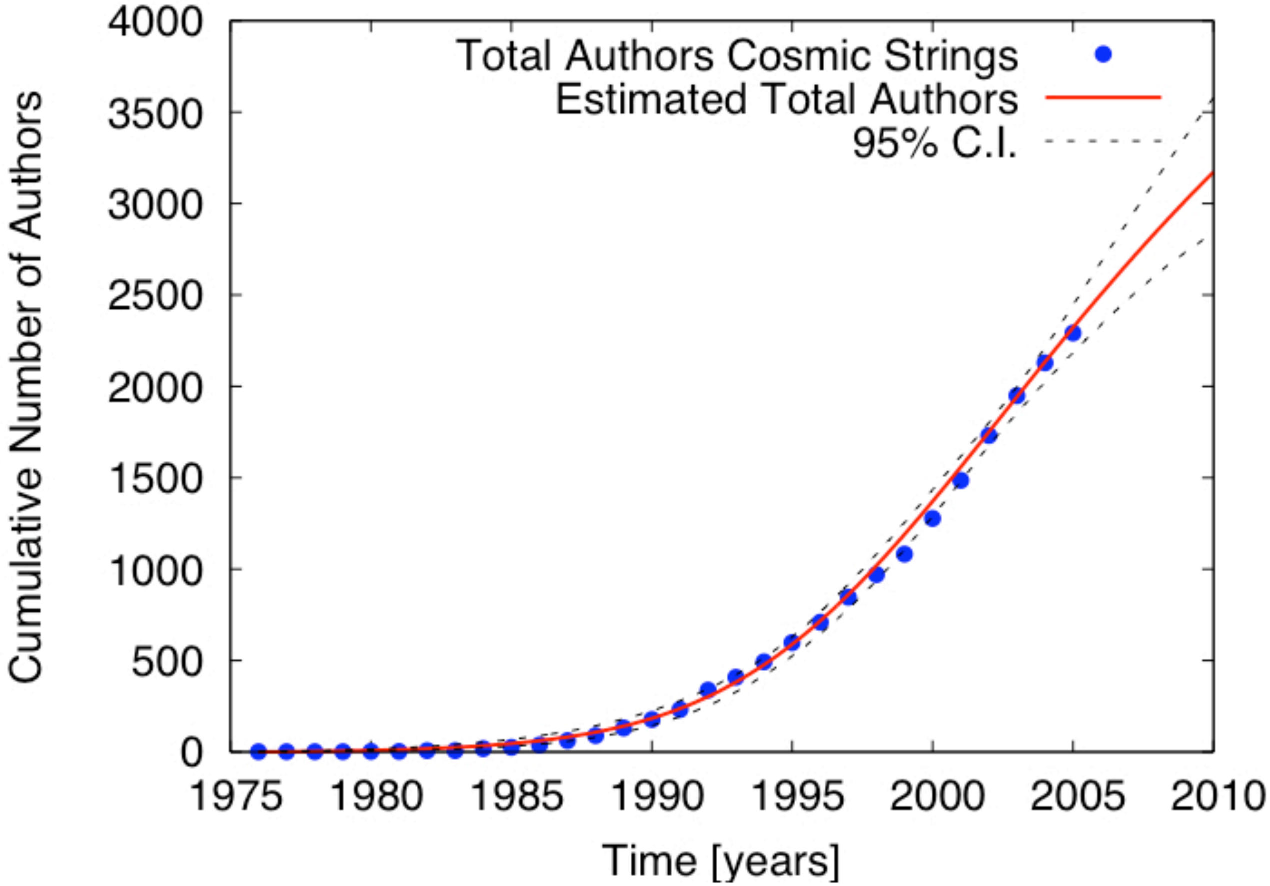
Cosmological Inflation



Alan Guth 1981
Andrei Linde 1982

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

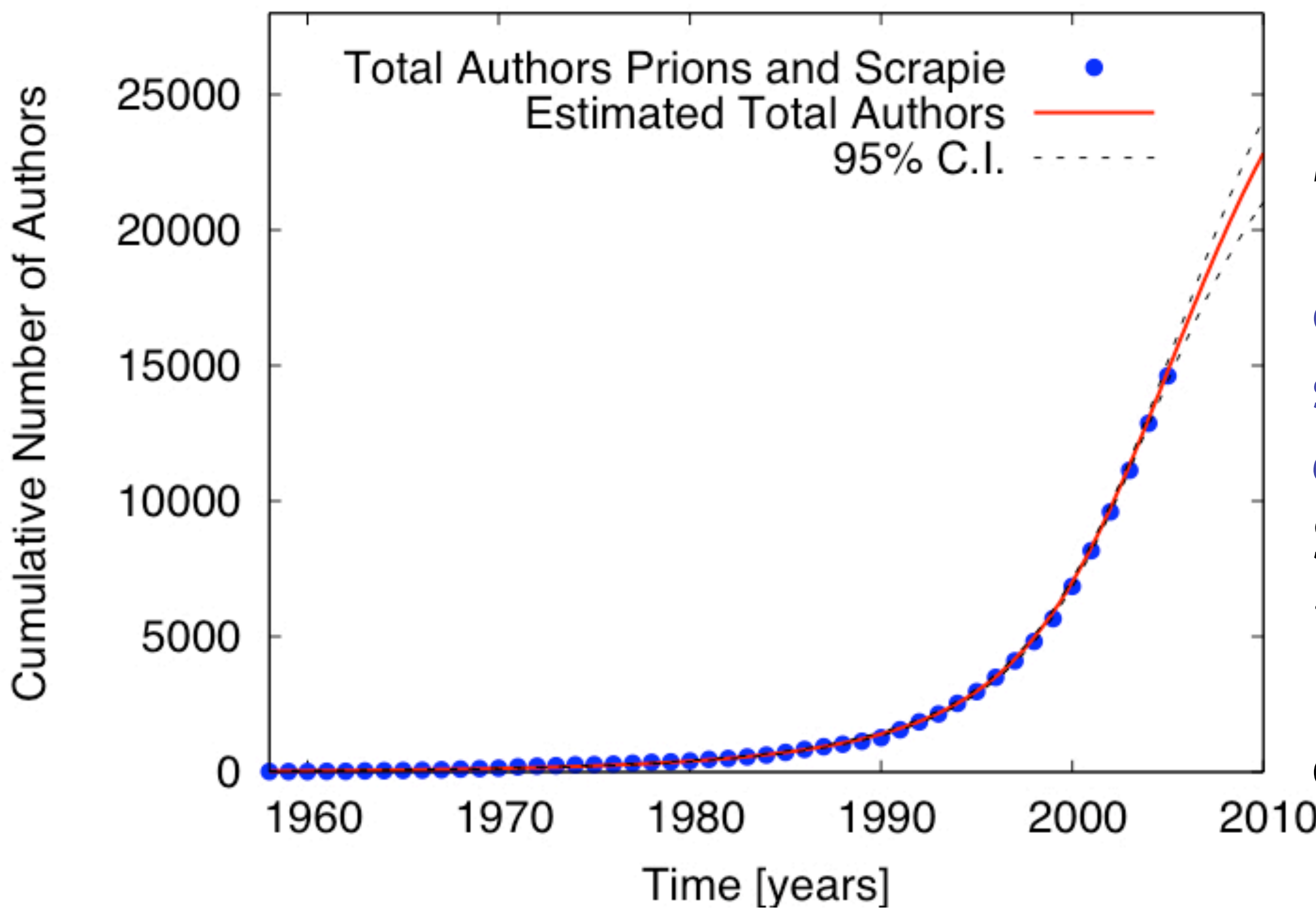
Cosmic Strings and Topological Defects



TWB Kibble 1976
Y Zeldovich 1980

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

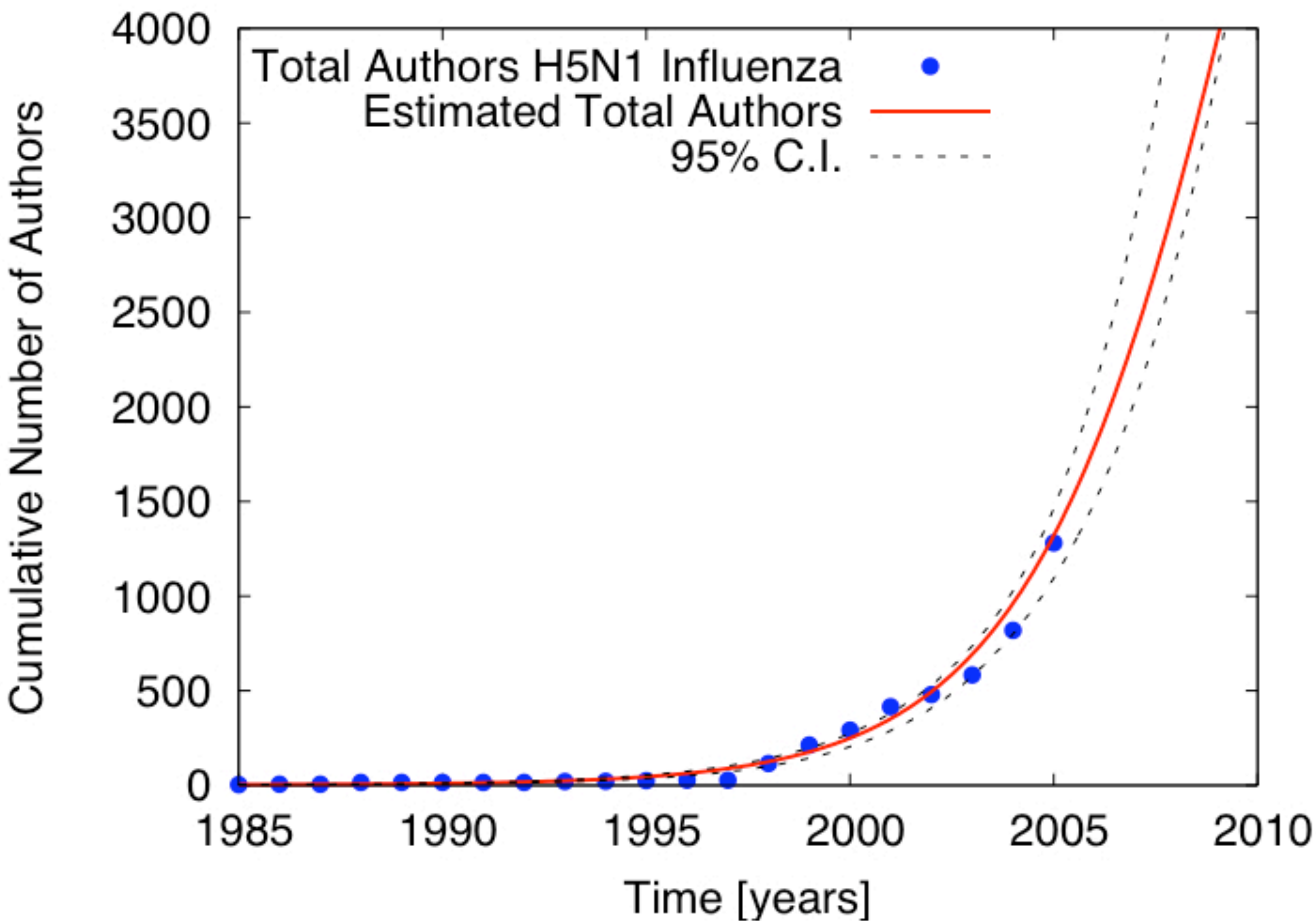
Prions



Prusiner 1982
Nobel Prize 1997

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

H5N1 Influenza (bird flu)



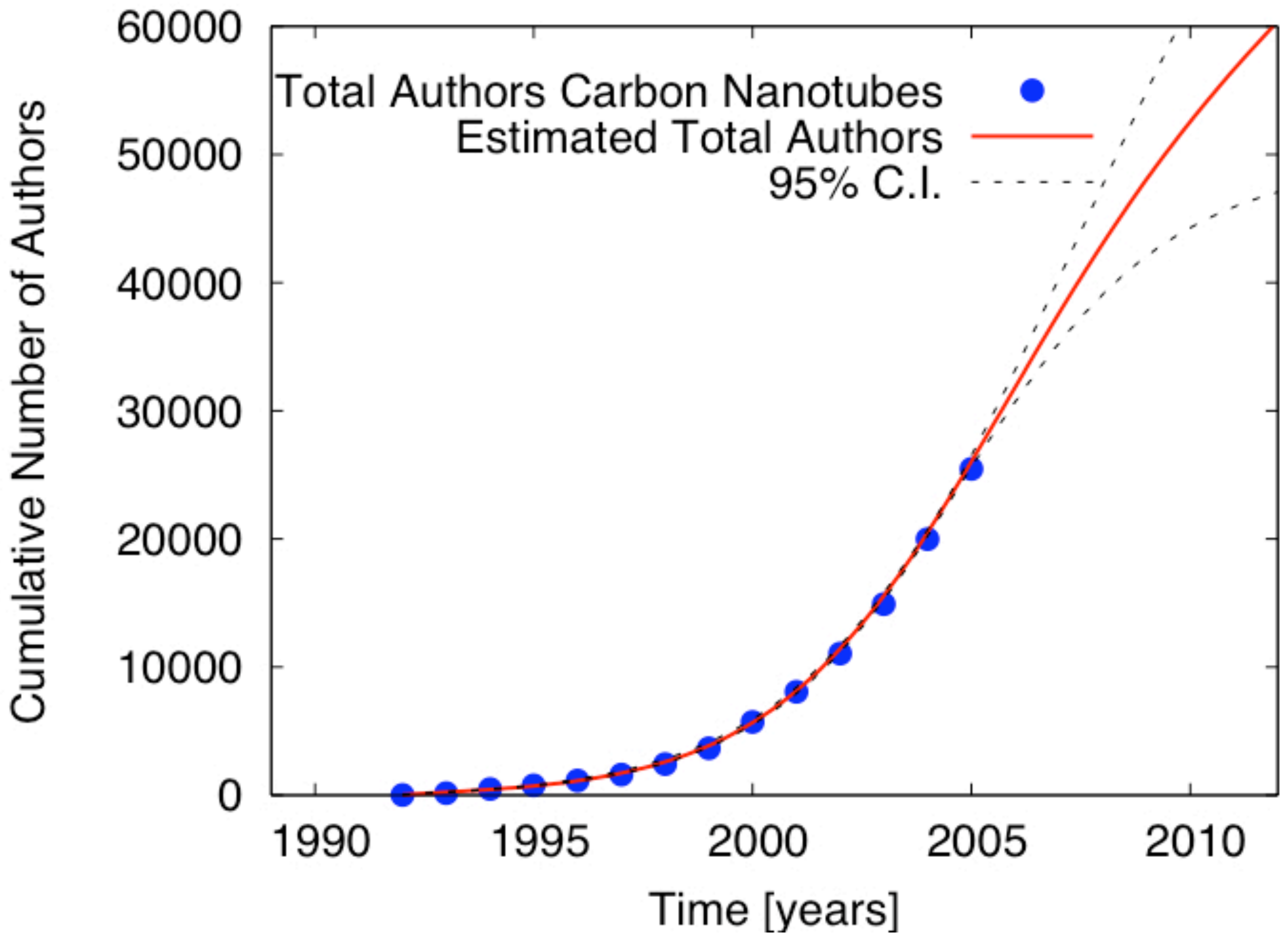
Disease of birds

First infected humans in 1997 in Hong Kong

280 humans infected

~60% case mortality

Carbon Nanotubes



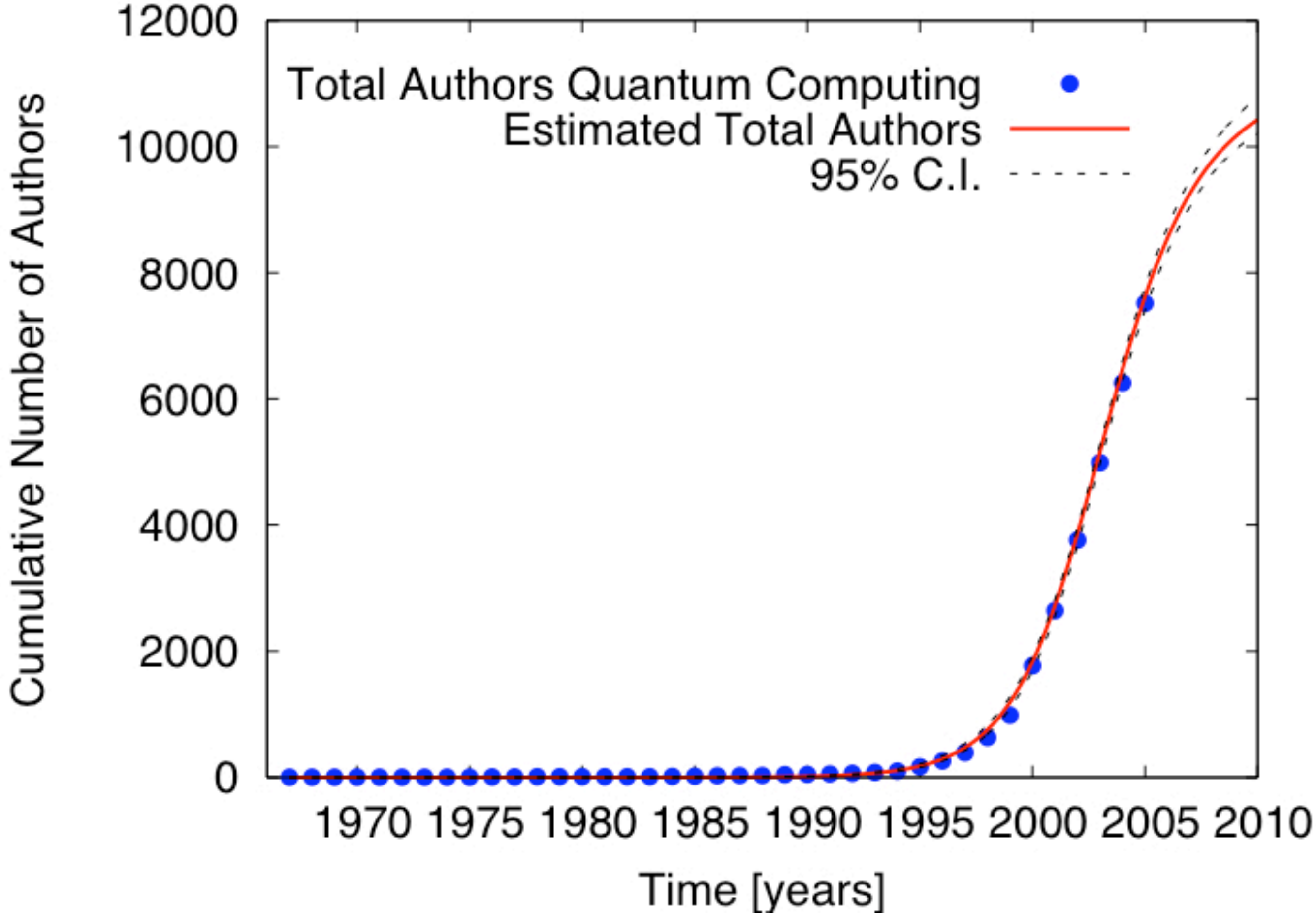
S. Ijima 1992

Important subfield of nanotech

Allotrope of Carbon

Promises to revolutionize Nano-engineering

Quantum Computers and Computation



First references 1960s-70s

Feynman 1982

Deutsch 1985

Algorithms:

Shor, Grover ~1995

NMR Experiments ~1996

Revolution in Computing & Cryptography?

Estimated parameters

Parameters	$S(t_D)$	$E(t_D)$	$I(t_D)$	$R(t_D)$	β	Λ	κ	ρ	γ	R_D
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 Λ, not ΛN in the equations for S.
** Susceptible population growth starts in 1990.

Measures of Scientific Productivity

Marginal Returns [from Economics]

Output →

Input →

$$\frac{\Delta Y(t')}{\Delta X(t)} = f[\Delta X(t)] \sim [\Delta X(t)]^{\beta}, \qquad t' \geq t$$

scaling relation (?)

"**Returns to Scale**" in ΔY=Papers *versus* ΔX=Authors:

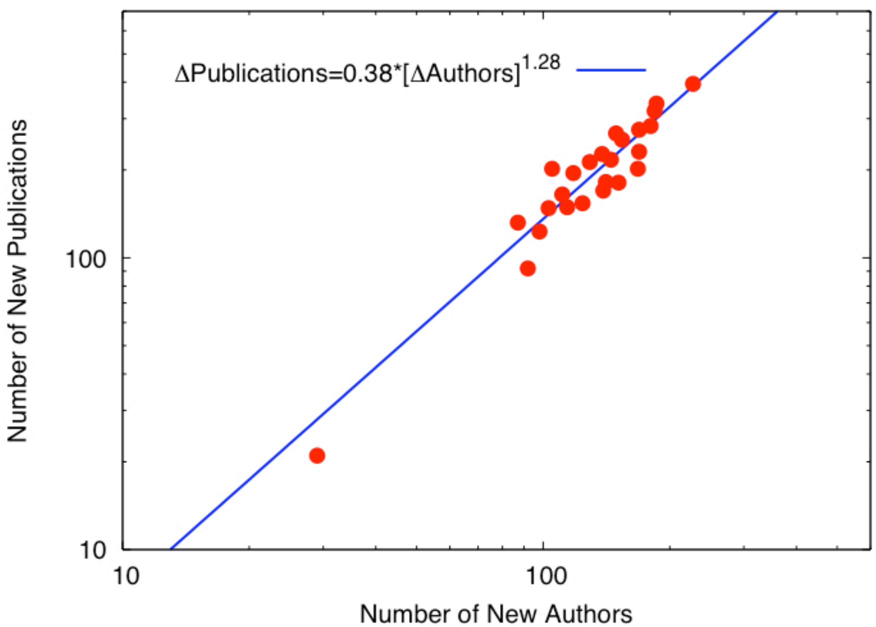
citations, patents funding, reputation

β=1 : each unit of input produces one unit of output

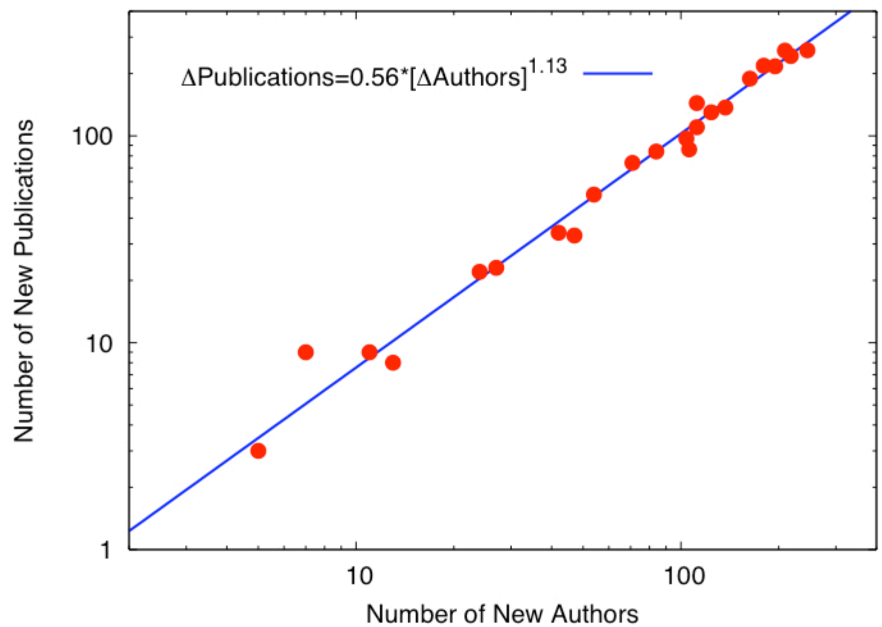
*$\beta < 1$: **diminishing returns**: each new author -> less papers/author*

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

Theoretical Physics



Cosmological Inflation β=1.28

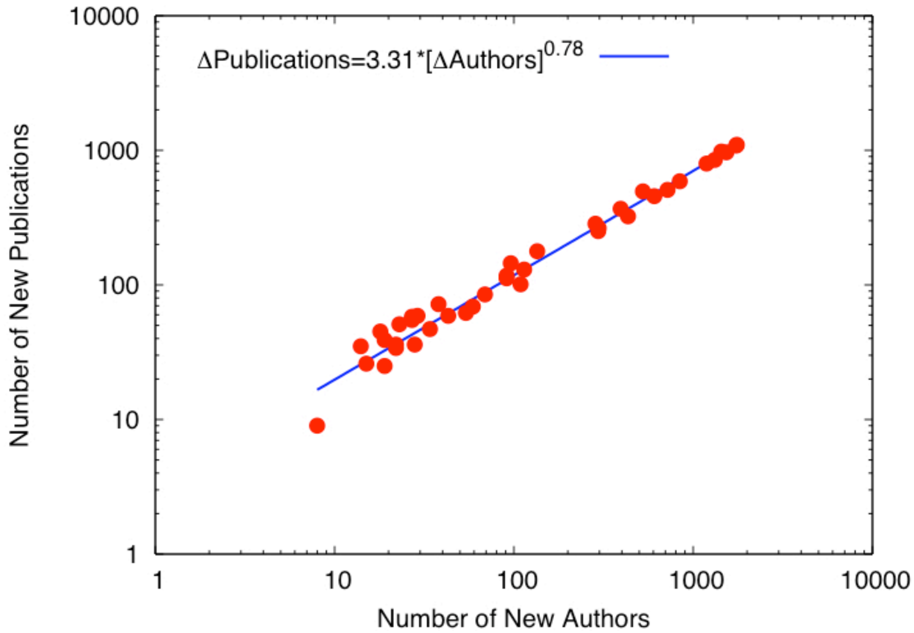


Cosmic Strings β=1.13

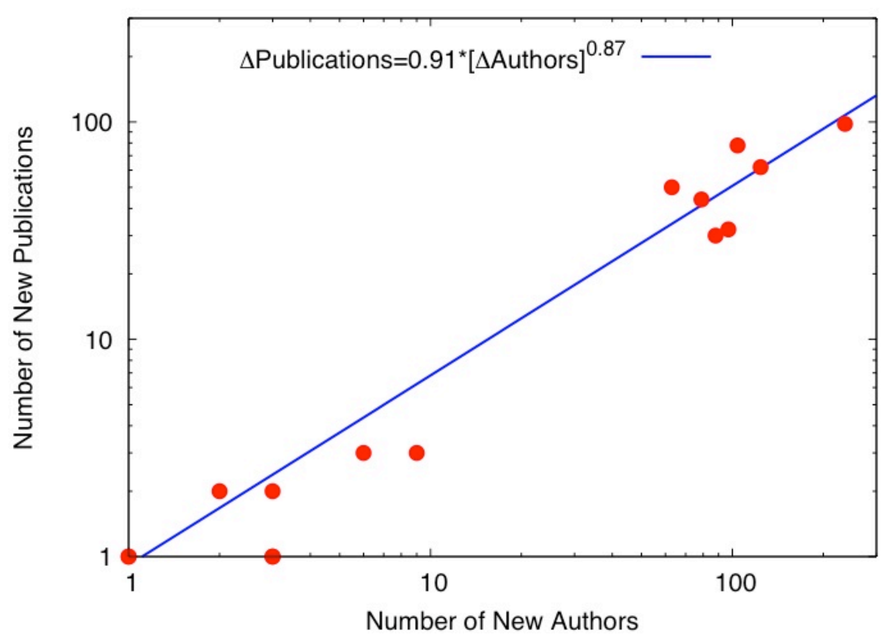
1943 - 2003 Los Alamos NATIONAL LABORATORY *Ideas That Change the World*

T Theoretical Division

BioMedical Fields



Prions β=0.78

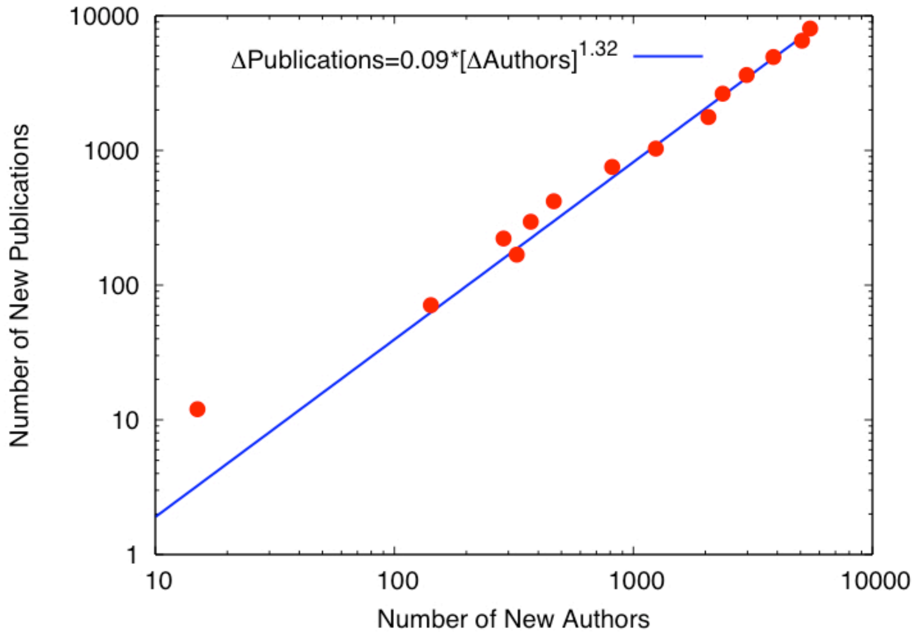


H5N1 Influenza β=0.87

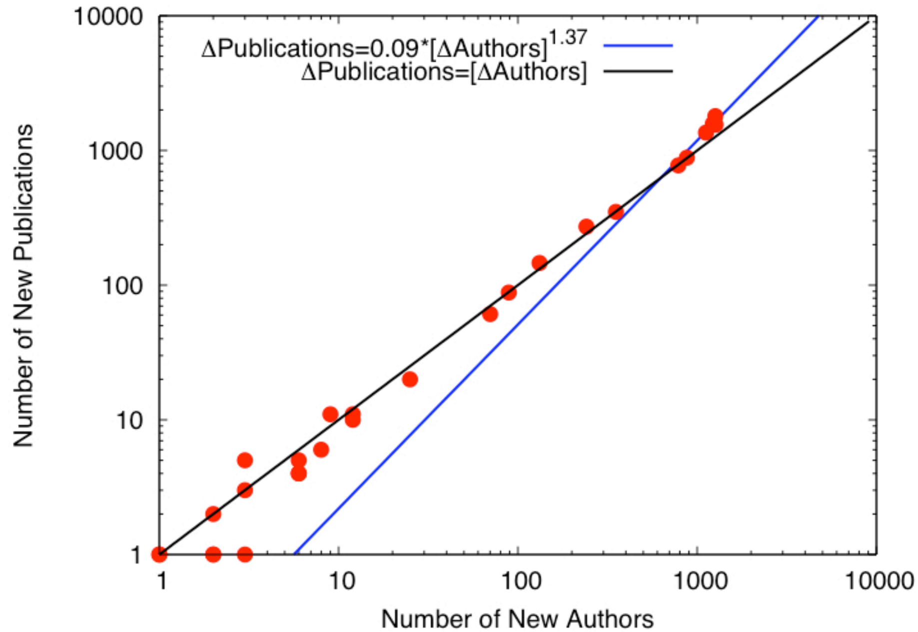
Los Alamos National Laboratory 1943 - 2003 Ideas That Change the World

T Theoretical Division

Technological Fields



Carbon NanoTubes β=1.32



Quantum Computation β=1 vs. 1.37

Scientific Intervention and Model Validation
dynamical sensitivity measures

-how much do I need to change a Γ_i to produce a change in output?
-what parameter leads to the greater changes [most sensitive]?

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

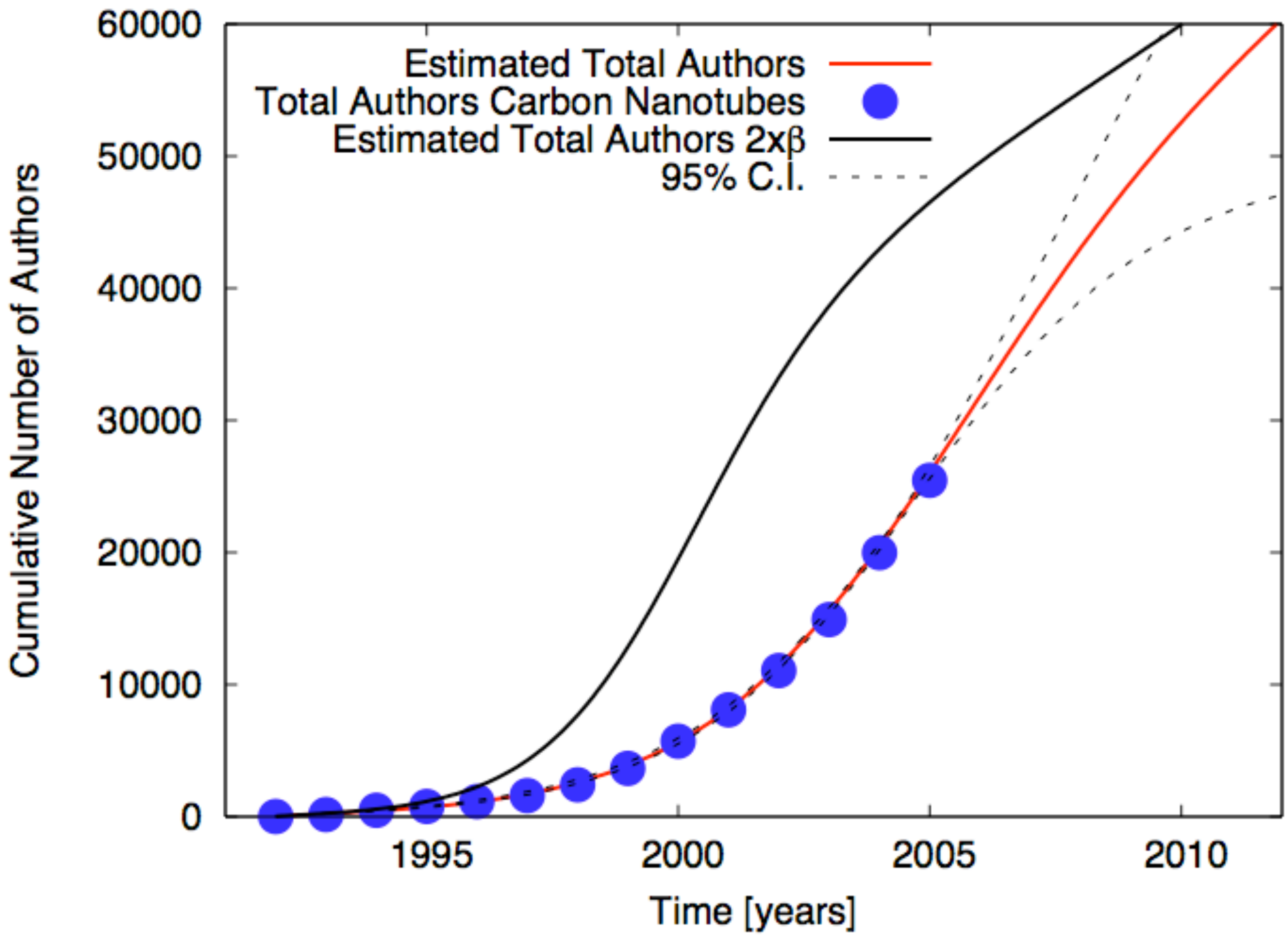
$$\frac{dE}{dt} = \beta S \frac{I}{N} - \rho S \frac{I}{N} - \kappa E,$$

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

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

$$\frac{d}{dt}\frac{dI}{d\Gamma_i} = \frac{dG(Y,\Gamma)}{d\Gamma_i} + \frac{dG_I}{dY_k}\frac{dY_k}{d\Gamma_i}$$

Susceptibility Equations



Los Alamos NATIONAL LABORATORY 1943 - 2003
Ideas That Change the World

T Theoretical Division

Models for the spread of scientific ideas

insights for a science of science and forecasting

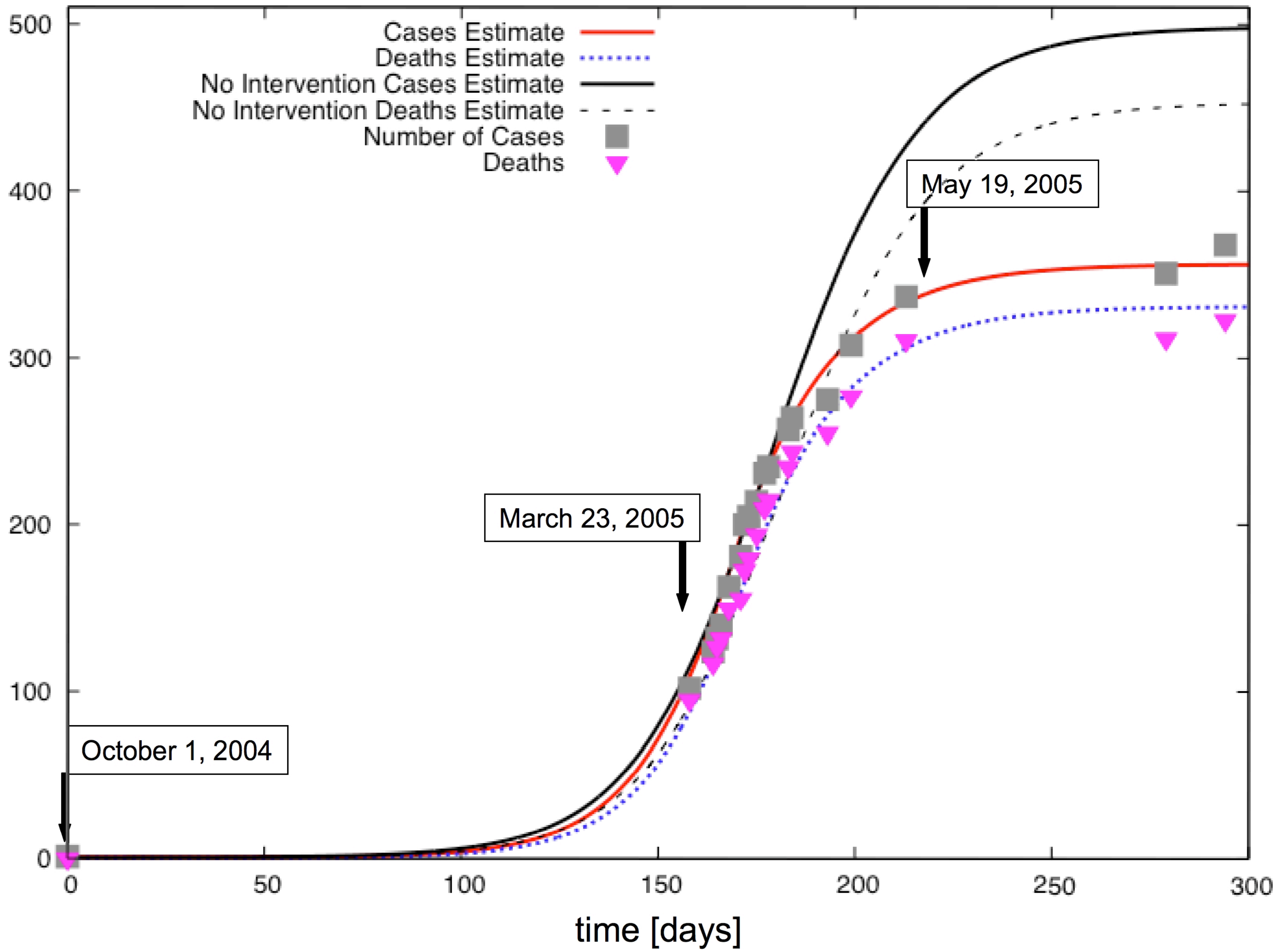
Some aspects of social dynamics are essential parts of the phenomenology and must be accounted in population models:

- Intentionality to learn
- Repeated contacts
- Importance of recruitment
- Absence of true recovery
- Idea Competition as a form of removal of susceptibles

What are the relevant ingredients for dynamical theories of science?

Population models work very well and can be used for **forecasting**

Measures of **Scientific Productivity** can be obtained from correlations



Los Alamos NATIONAL LABORATORY 1943-2003 Ideas That Change the World

T Theoretical Division