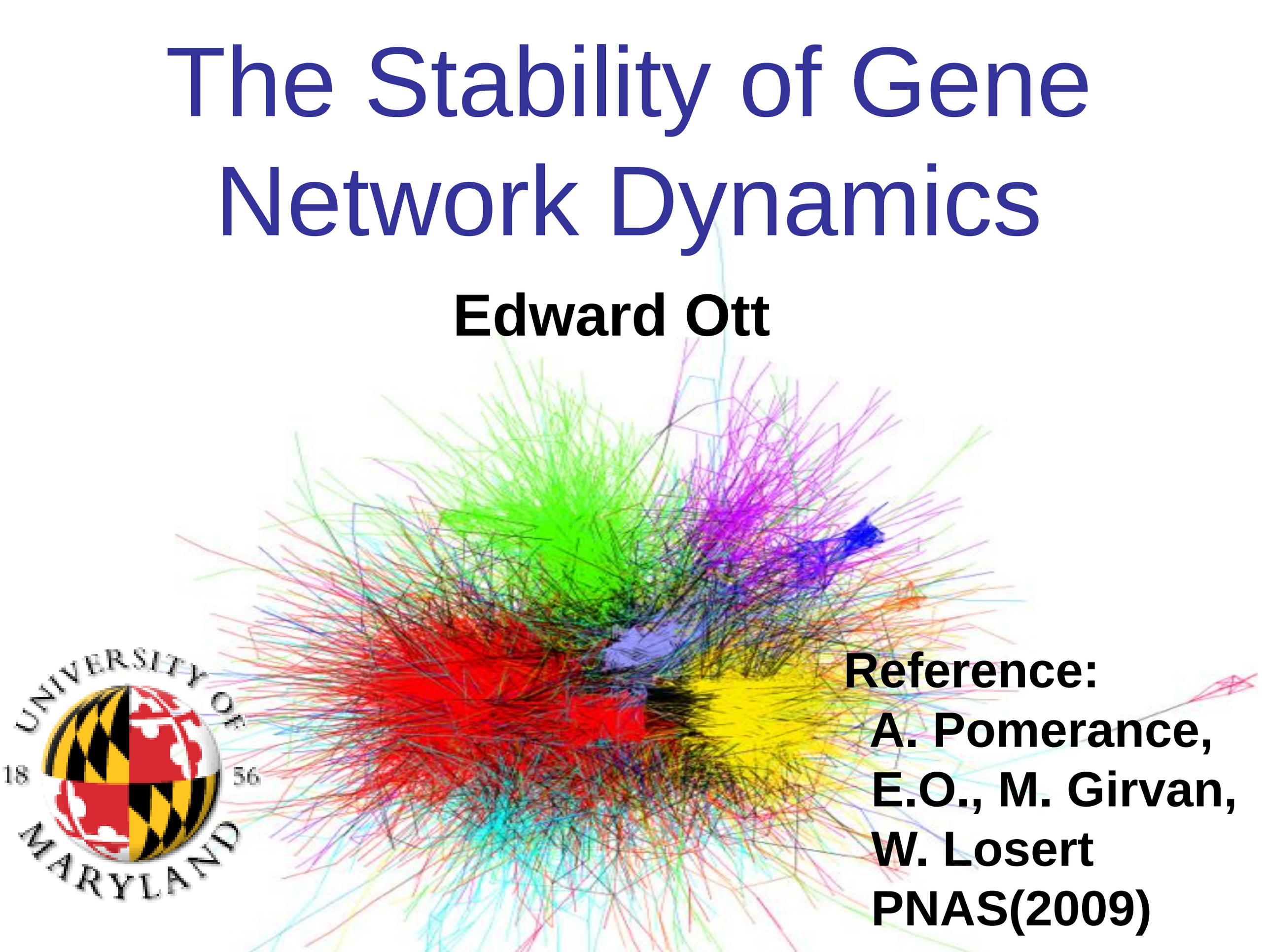
The Stability of Gene Network Dynamics

Edward Ott



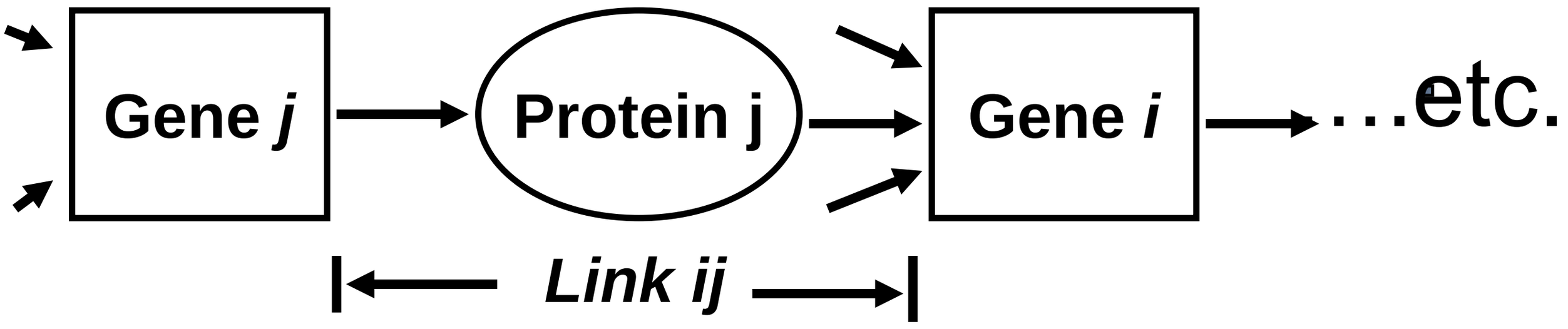
Reference:
A. Pomerance, E.O., M. Girvan, W. Losert
PNAS(2009)

Overview

- **Our main result is a very general, simple and easily applied criterion for determining the stability of discrete state dynamical models of gene regulation.**
- **This result makes possible treatment of network topology and of many other real factors previously inaccessible to study. Importantly, the result can be applied to networks derived from experimental data.**
- **We hypothesize that a dynamical instability in the gene network may be a causal mechanism contributing to the occurrence of cancer.**

Gene Expression and Regulation

- **Genes:** Functional segments of DNA.
- **Gene expression:** The function of a given gene may be to produce ("express") a specific protein.
- **Regulation:** A gene's expression is regulated, in large part, by the binding of particular proteins to its promoter region.



The resulting directed network of interactions is given by the network's 'adjacency matrix' A

$A_{ij} = 1$ if there is a link or 0 otherwise.

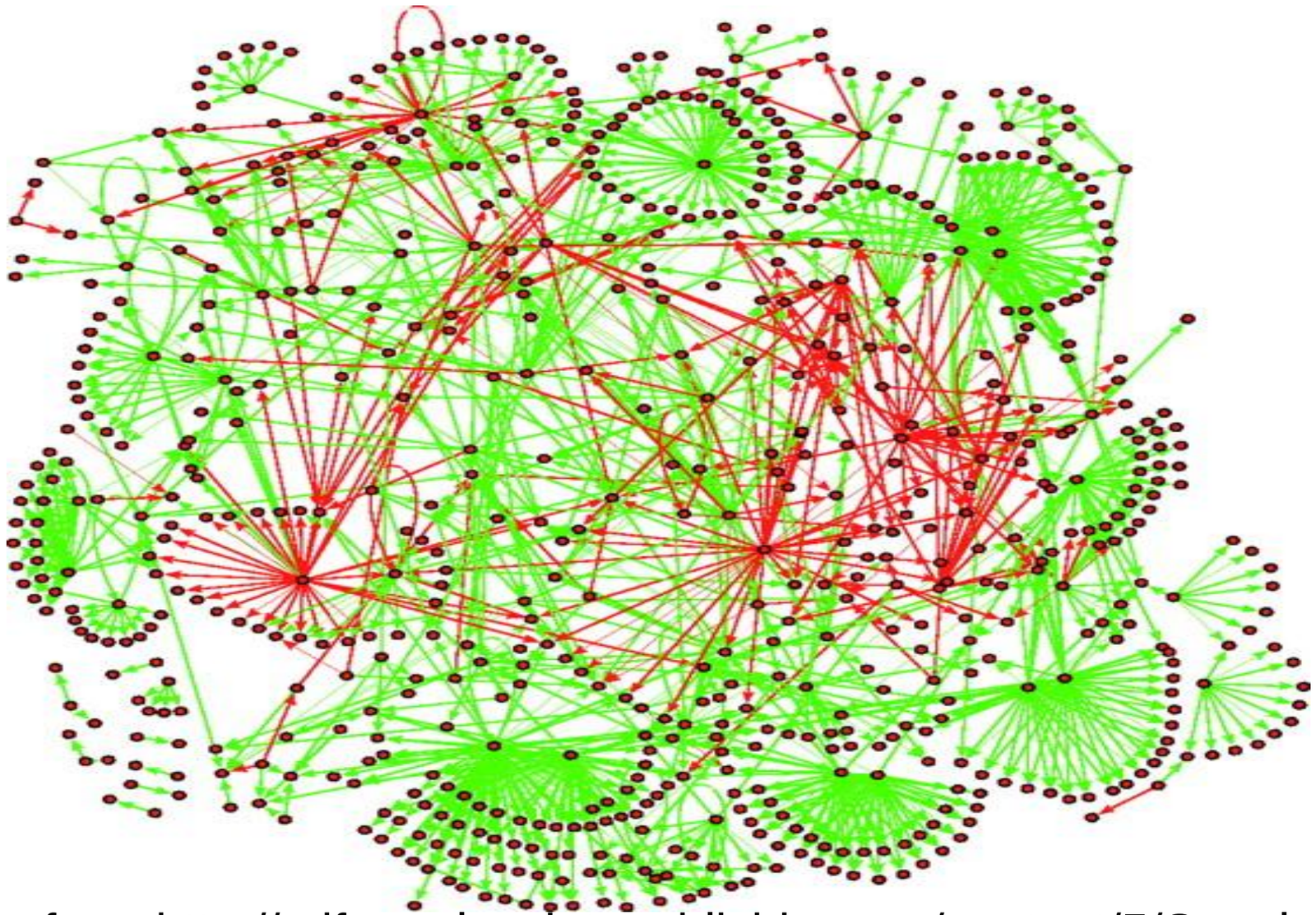
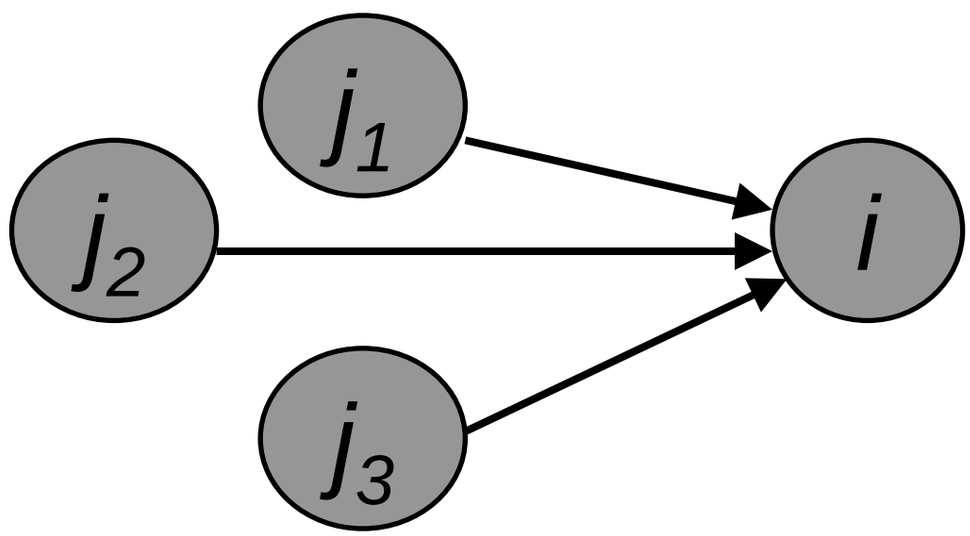


Figure taken from http://rsif.royalsocietypublishing.org/content/5/Suppl_1/S85.full

Naïve Boolean Model

- The state of each gene (network node) is either on (symbol 1) or off (symbol 0).
- Time is discretized: t = 0,1,2,3,….
- At time t, the states of each node i, denoted σ_i^t (= 0 or 1), are simultaneously updated using a deterministic Boolean function (truth table) of the time t-1 states, σ_j^{t-1}, of the K_i^{in} nodes, j, that input to node i. (K_i^{in} = the in-degree of node i.)

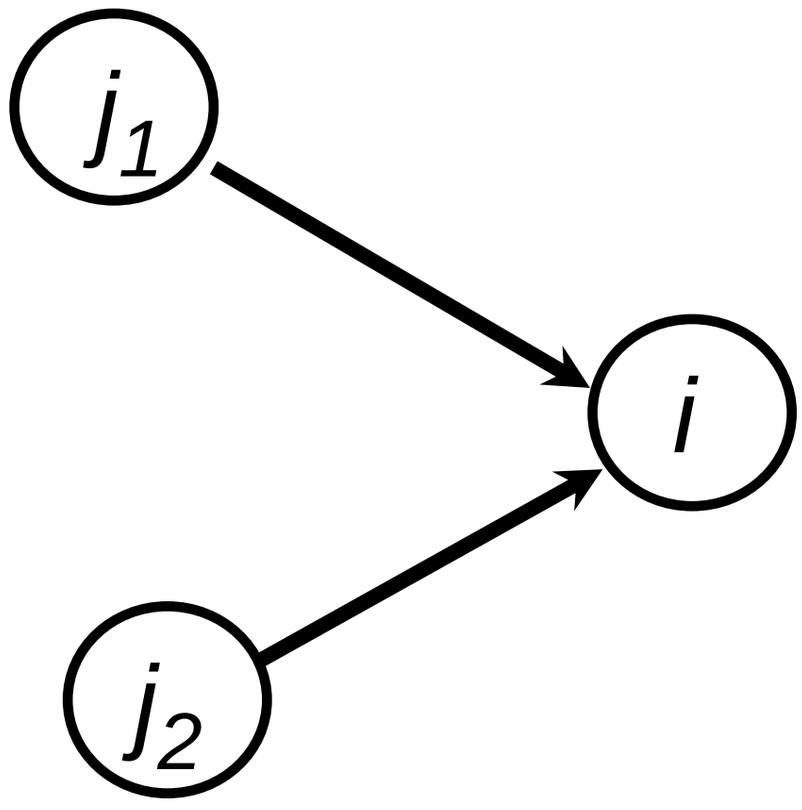
K_i^{in}=3:



Update Truth Table

current state at time *t* (input to gene *i*)		State of gene *i* at time *t+1* (output)
Gene j_1	Gene j_2	
0	0	0
0	1	0
1	0	1
1	1	0

$K_i^{in} = 2$

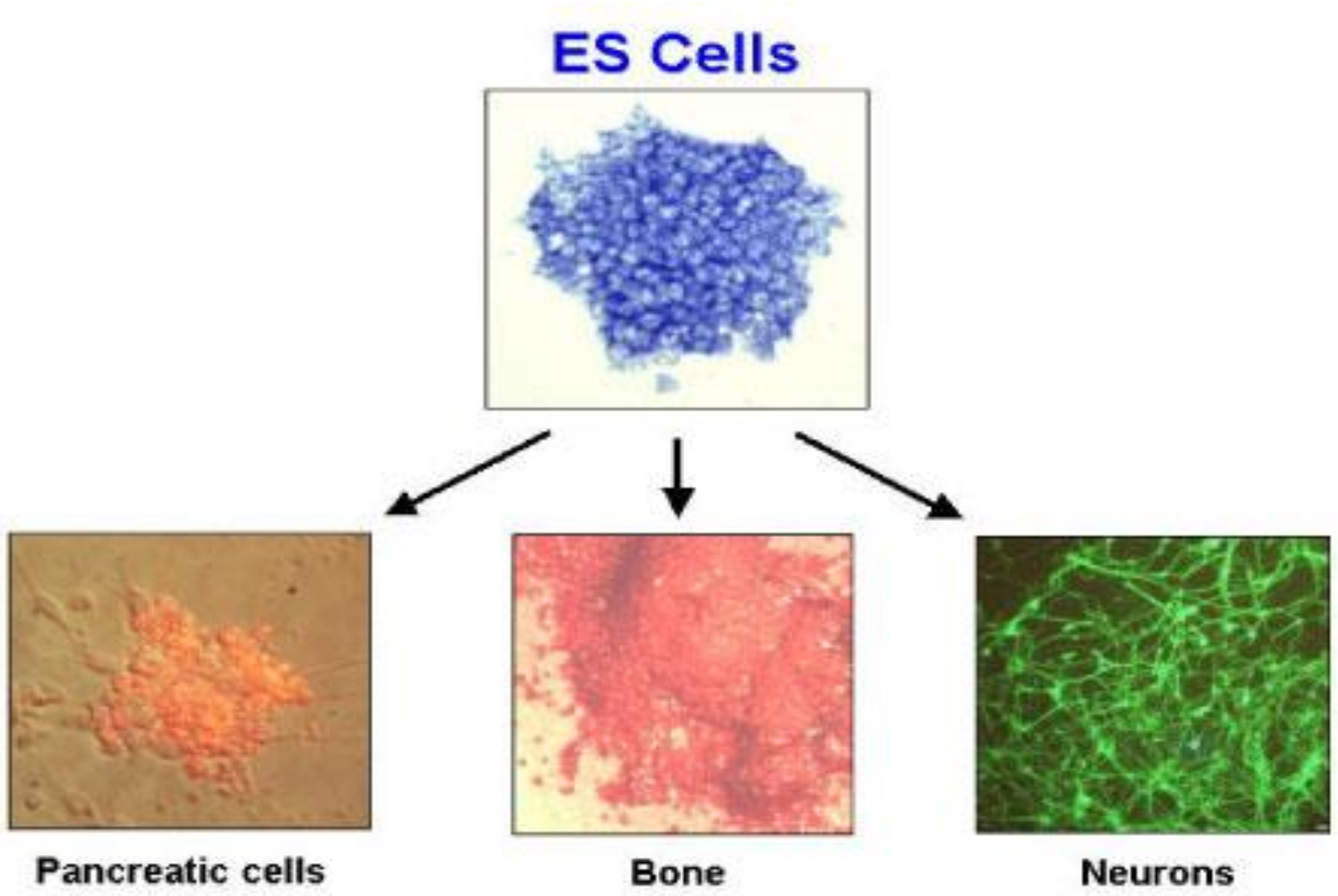


Properties of the Boolean Model

- **Finiteness: Eventually the system must return to a previously visited state.**
- **Determinism: Upon this return, the subsequent dynamics will be the same as for the previous visit.**
- **Attractors: Every initial condition produces a trajectory that eventually goes to a periodic orbit, called the "attractor" of that initial condition, and different initial conditions can go to different periodic orbit attractors.**

Assumed significance of attractors

- **Attractor may be thought of as representing a specific patterns of protein expression that defines the character of cells [Kauffman ('69)].**



Kauffman's *N-K* Net Model

S. Kauffman, J.Theor.Biol. 22,437 (1969).

K_i^{in} = *K* independent of *i*. *N*= # of nodes.

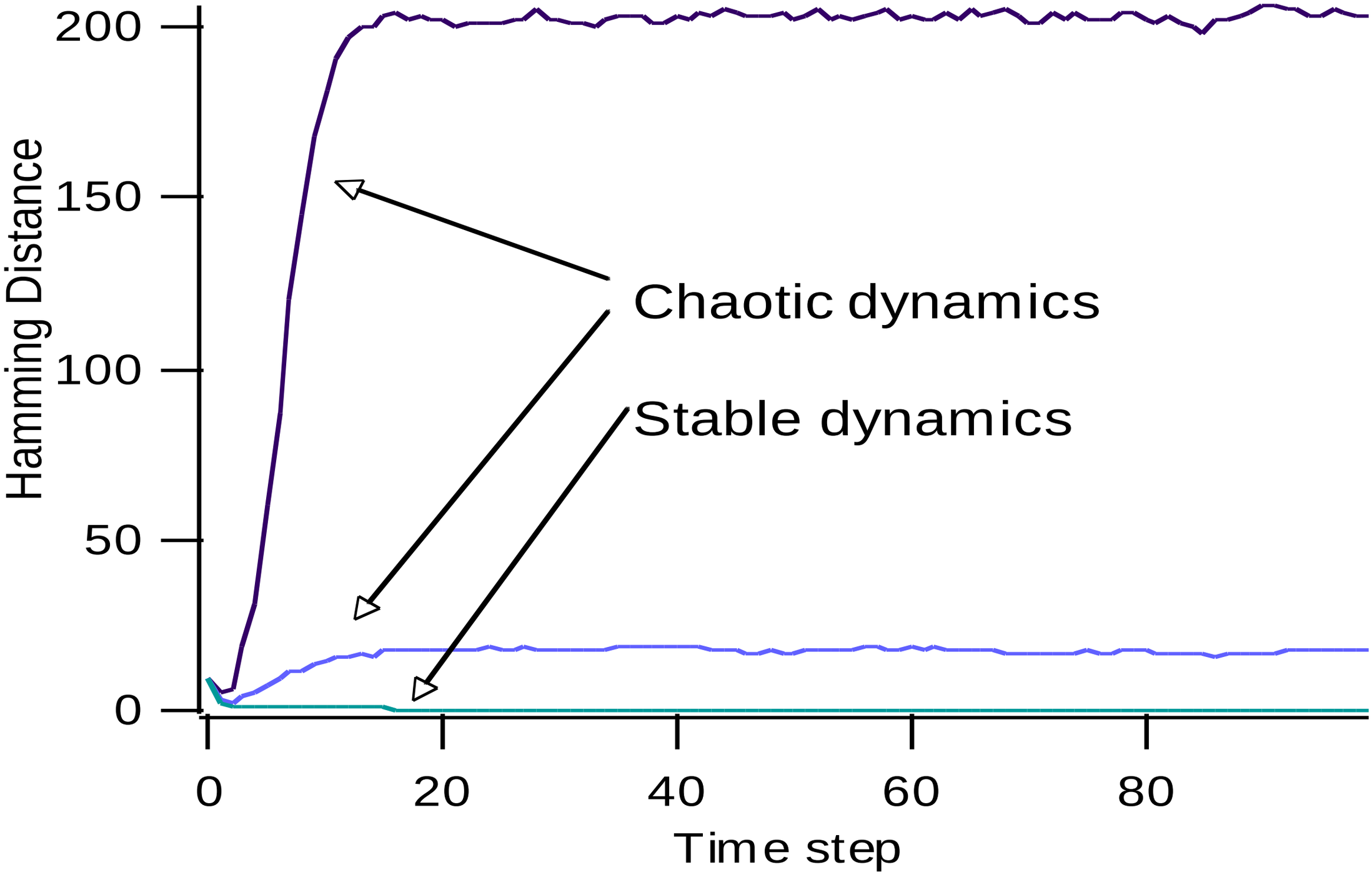
The *K* inputs to each node *i* are randomly chosen from amongst all the *N-1* other nodes.

The output entries for each node's truth table are randomly chosen with probability of 0 being ½ and probability of 1 being ½.

Stability

- **(State) = $\underline{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)$**
- **Distance between two states, $\underline{\sigma}$ and $\underline{\tilde{\sigma}}$: $H(\underline{\sigma}, \underline{\tilde{\sigma}}) = \sum_i |\sigma_i - \tilde{\sigma}_i|$ (the Hamming distance)**
- **Instability (stability): If at $t=0$, $H << N$, then H initially grows (shrinks) with increasing time t.**

Chaotic and stable dynamics for different networks



Past Work on Stability

Derrida & Pomeau, Europhys. Lett. 1, 45 ('86).

***p* = probability of a *0* in the truth table output.**

Network and truth table are "annealed".

Unstable if *2p(1-p) > 1/K*.

Others consider in-degree distribution $P(K^{in})$ with input nodes still chosen uniformly randomly [e.g., Aldana and Cluzel (PNAS (2003)): $P(K^{in})$ scale-free].

Motivation for Our Work

Real networks are far from the model networks previously analyzed.

We would like to be able to analyze any fixed network, and we are interested in such effects as assortativity, community structure, a different p_i at each node, asynchronous update, nonsynchronous updates with heterogeneous link time delays.

We can do these [Pomerance et al. PNAS (‘09)]!

Network Instability and Cancer

- **Tumor dissections show that nearby cells in cancer tumors are very heterogeneous and have widely different gene expression patterns. (This contrasts with normal tissue, where, e.g., cells in muscle are similar.)**
- **We conjecture that this observed variability of nearby cells in the same tumor might be due to mutation-induced breakdown of the dynamical stability of the gene network.**
- **Ongoing, rapid experimental advances in gene network reconstruction may soon make tests and applications of this conjecture feasible.**

Review of Derrida-Pomeau Method

(Basis of all previous work on stability.)

K^{in} and p are the same at each node.

Annealing: At each time-step

(1) Truth table outputs are randomly chosen.

(2) At each node the K^{in} other nodes that input to that node are randomly chosen.

But we are really interested in the *frozen* case.

However, the annealed situation can be analyzed, but not the frozen case.

Assumption: For N>>1,
(annealed results) ~ (frozen results).

Our Technique: 'Semi-Annealing'

Pomerance, Ott, Girvan, Losert, PNAS (2009)

The network is held fixed.

Each node *i* has its own p_i (bias to be on or off).

Semi-annealed: At each time step, only the truth table is randomly chosen, and this is done at each node *i* according to the fixed bias probability p_i for that node.

For *N*>>1, our numerical simulations show:

(semi-annealed) ~ (frozen)

Semi-Annealed Analysis

- **Consider two node states, $\sigma(t)$ and $\tilde{\sigma}(t)$, where σ and $\tilde{\sigma}$ evolve from slightly different initial conditions.**

- **Let $y_i(t)$ = (the probability that $\sigma_i(t)$ and $\tilde{\sigma}_i(t)$ differ).**

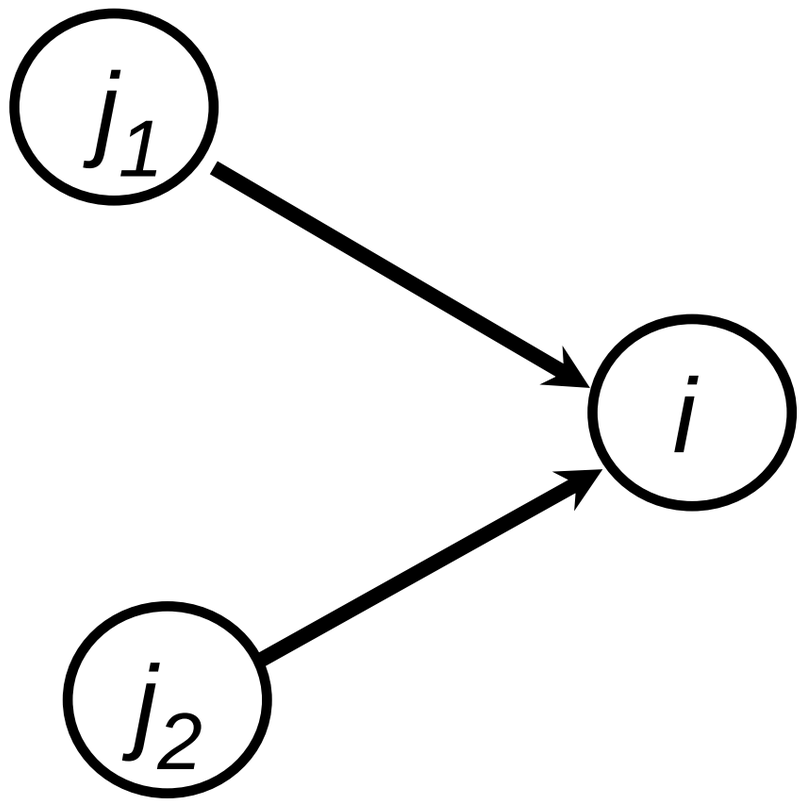
- **Let q_i = [the probability that $\sigma_i(t)$ and $\tilde{\sigma}_i(t)$ differ,** (Re-show slide #6)
  given that the $\sigma(t\text{-}1)$ and $\tilde{\sigma}(t\text{-}1)$ inputs to i differ]

$$q_i = 1 - [p_i^2 + (1 - p_i)^2] = 2p_i(1 - p_i)$$

Update Truth Table

current state at time *t* (input to gene *i*)		State of gene *i* at time *t+1* (output)
Gene j_1	Gene j_2	
0	0	0
0	1	0
1	0	1
1	1	0

$K_i^{in} = 2$



Update of Nodal Probability $y_i(t)$

prob. that the inputs at *t-1* to *i* are not all the same

$$y_i(t) = q_i \left\{ 1 - \prod_{j, A_{ij}=1} \left[1 - y_j(t-1) \right] \right\}$$

prob. that inputs are all the same

Perturbe around $\underline{\sigma} = \underline{\tilde{\sigma}}$ (i.e., $y_i << 1$); linearization gives

$$y_i(t) \cong q_i \sum_{j=1}^{N} A_{ij}\, y_j(t-1) = \sum_{j=1}^{N} Q_{ij}\, y_j(t-1) \text{ , or } \underline{y}(t) = \underline{\underline{Q}}\, \underline{y}(t\text{-}1)$$

where $Q_{ij} = q_i A_{ij}$ ="modified adjacency matrix".

Stability

$$y_i(t) = \sum_j Q_{ij}\, y_j(t-1), \quad Q_{ij} = q_i A_{ij}$$

λ_Q = **largest eigenvalue of** $\underline{\underline{Q}}$ **, which, according to the Perron - Frobenius theorem, is real and positive** ($Q_{ij} \geq 0$).

$\lambda_Q < 1$, $\underline{y} = 0$ **is stable,**

$\lambda_Q > 1$, $\underline{y} = 0$ **is unstable,**

$\lambda_Q = 1$, $\underline{y} = 0$ **is " edge of chaos ".**

Numerical Tests

We show tests of predictions of

(i) λ_Q stability criterion,

(ii) Predicted saturated normalized

Hamming distance between $\underline{\sigma}$ and $\underline{\tilde{\sigma}}$:

$$\overline{y} = \lim_{t\to\infty} \frac{1}{N} \sum_i y_i(t).$$

(re - show slide # 11)

We compare results of the semi-annealed theory with frozen simulations. Issues we study include:

Nodal in/out degree correlation

Assortativity / disassortativity

q_i correlation with nodal degrees

Differing time delays along different links

Community structure

Motifs

Finite size effects

Experimentally determined network for yeast

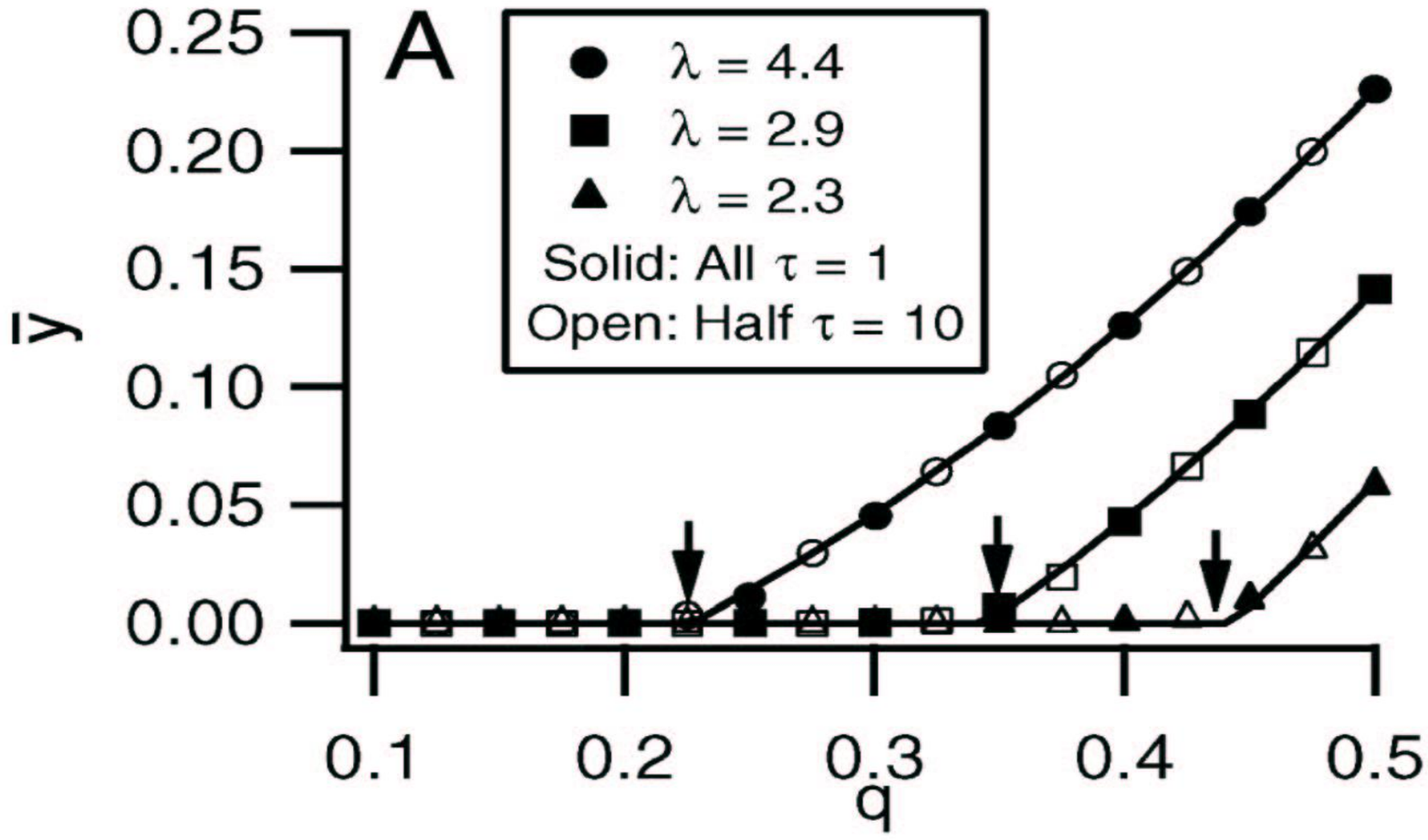
An Example

$N = 10{,}000$

$P(K) \propto K^{-2.1}$

for $K < 15$

$q_i = q \Rightarrow \lambda_Q = q\lambda_A$



$\left\langle K^{in} K^{out} \right\rangle$ tuned by swapping of out degrees

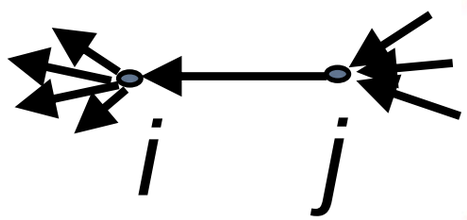
$\lambda_A \cong \left\langle K^{in} K^{out} \right\rangle_{node} / \left\langle K \right\rangle_{node}$

- Solid line and arrows = semi-annealed theory.
- Symbols = numerical results for frozen case.
- Open symbols: different delay times.

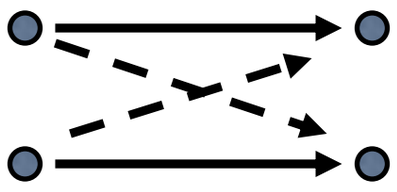
Another Example

- Same as in previous slide but for three networks with different assortativity / disassortativity.
- Starting with a randomly constructed non-assortative network, the assortativity coefficient,

$$\rho = \frac{\left\langle K_i^{out} K_j^{in} \right\rangle_{\text{edges } j \to i}}{\left\langle K^{in} K^{out} \right\rangle_{\text{node}}}$$

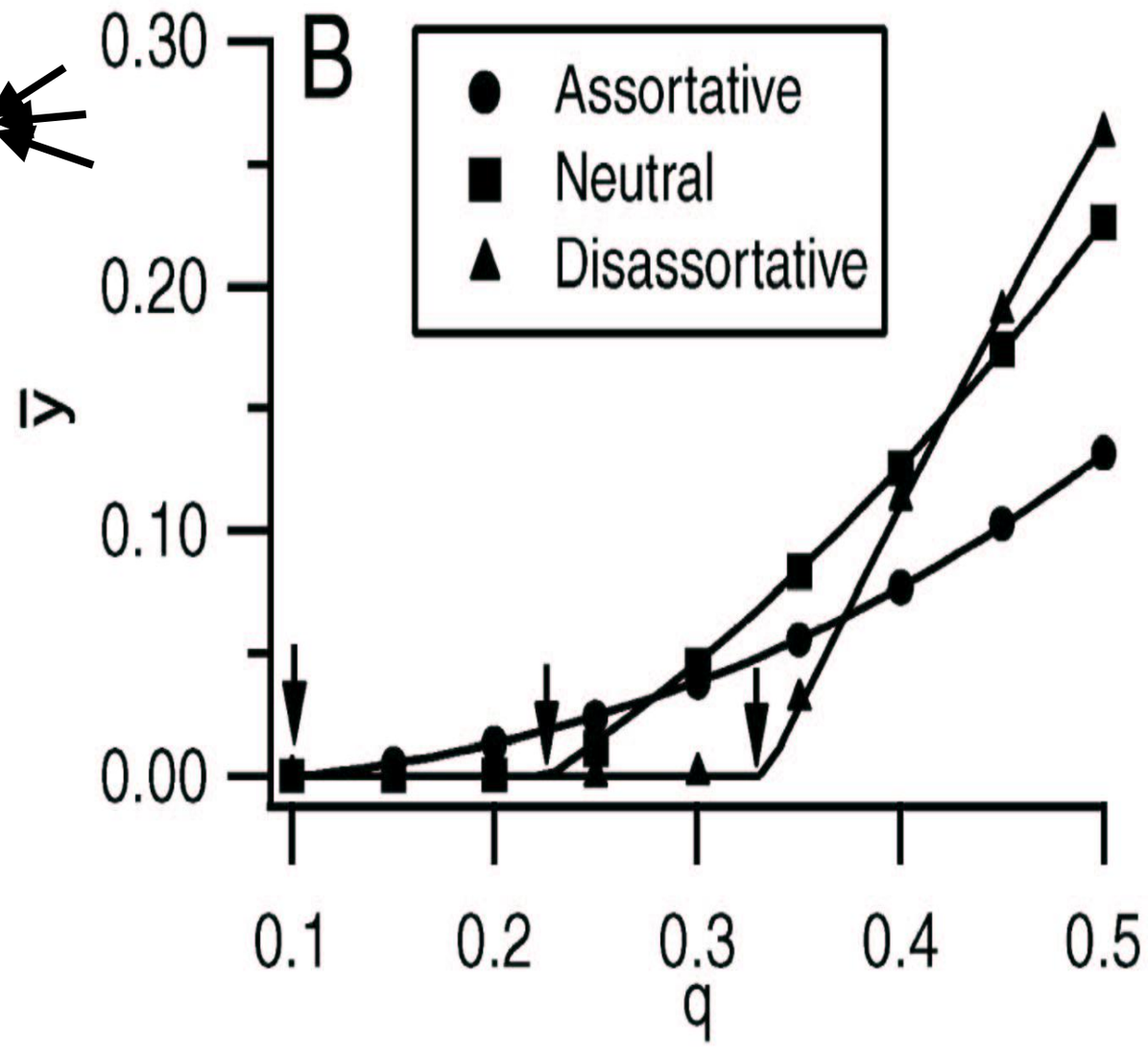


is tuned by swapping edges :



- $\lambda_Q = q\lambda_A$:

$$\lambda_A \cong \frac{\left\langle K^{in} K^{out} \right\rangle_{\text{node}}}{\left\langle K \right\rangle_{\text{node}}} \rho$$



Restrepo, Ott, Hunt, PhysRevE '07

A Possible Strategy for Cancer Treatment

- **Our cancer/network-stability hypothesis combined with our stability analysis suggests a possible cancer therapy strategy: Namely, design drugs that target those genes or links whose disabling would most reduce λ_Q.**

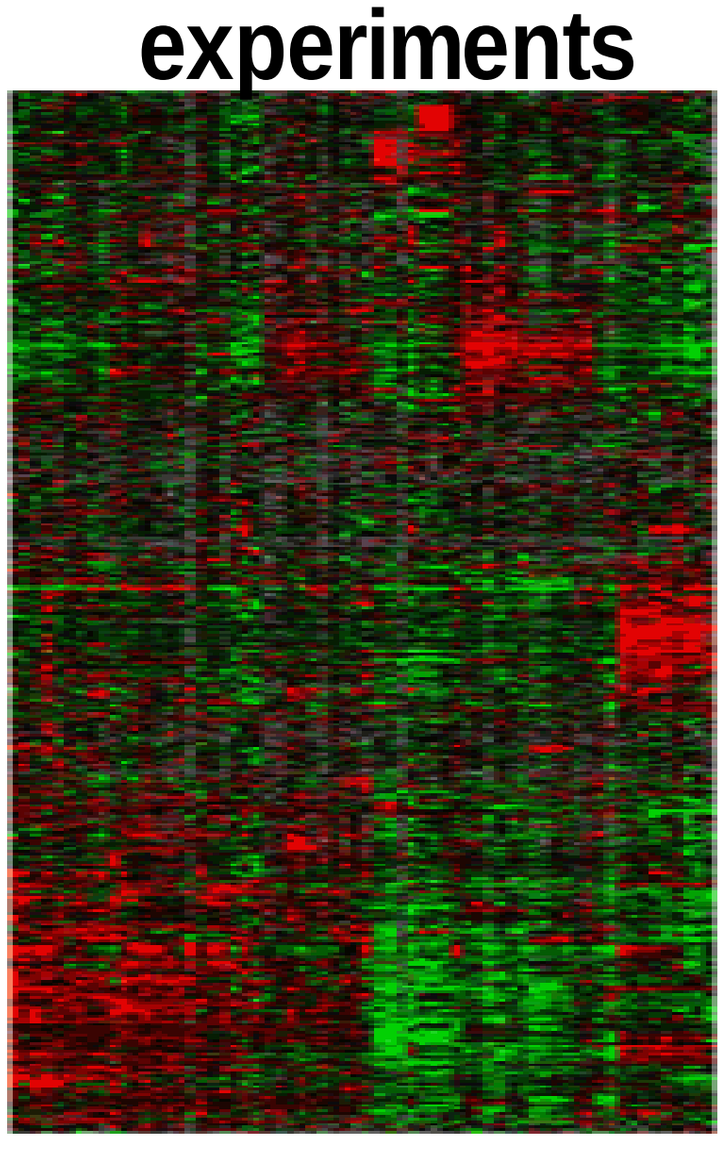
Summary

- **We introduce a ‘semi-annealing’ technique for studying the stability of discrete state gene network models.**
- **Our technique greatly expands the range of factors that can be treated, allowing treatment of arbitrary network topology, node-specific bias p_i, etc.**
- **We conjecture that gene network instability may be an important causal factor in some cancers.**
- ***Reference:*** **A. Pomerance, E. Ott, M. Girvan, and W. Losert, “The effect of network topology on the stability of discrete state models of genetic control”, *PNAS* 106, 8209 (2009).**

Estimating the network and the node bias probabilities p_i from data

- **Network:** Undirected links can be inferred from data by looking at co-expression patterns across a range of perturbation experiments, and other techniques can determine directed links.

- The nodal bias probabilities p_i can be estimated from clinical expression data.



The Perron-Frobenius Eigenvalue

- **Recall that, since A_{ij} and Q_{ij} are non-negative, their stability-determining eigenvalue of maximum magnitude is real and positive (Perron-Frobenius Theorem).**
- **"Markov" theory for λ_A, the Perron-Frobenius eigenvalue (Note: $\lambda_Q = q\lambda_A$ if $q_i = q$ is the same on all nodes.)**

  Restrepo, Ott, Hunt, Phys. Rev. E 76, 056119(2007).

  E.g., $$\lambda_A \cong \frac{\left\langle K^{in} K^{out} \right\rangle_{nodes}}{\left\langle K \right\rangle_{nodes}} \rho$$

Perron-Frobenius Eigenvalue (cont'd)

- **Markov theory for λ_Q with $Q_{ij} = q_i A_{ij}$:**

  Ott and Pomerance, Phys. Rev. E 79, 056111 (2009).

  E.g., if q_i is correlated with $K_i^{in} K_i^{out}$,

$$\lambda_Q \cong \frac{\left\langle qK^{in}K^{out} \right\rangle_{nodes}}{\left\langle K \right\rangle_{nodes}}$$

Finite Size Effects

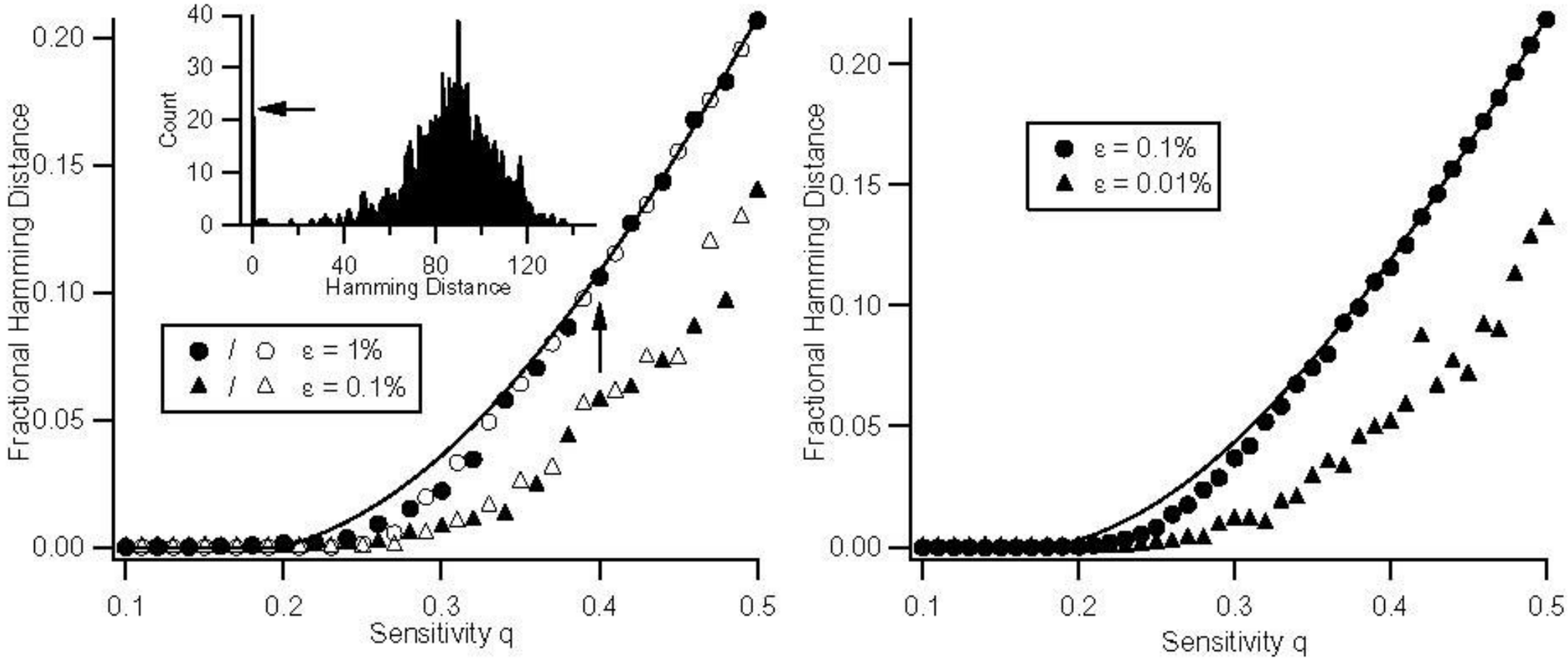


FIg. S2. The steady-state fractional Hamming distance h/N for (a) $N = 10^3$ and (b) $N = 10^4$ as a function of the sensitivity q for various values of ε, both in the frozen case (filled symbols) and the annealed case (open symbols). The largest eigenvalue of this network's adjacency matrix is $\lambda \approx 5$. While the theory does not depend on the value of ε, finite-size effects cause a dependence on the number of flipped bits. The inset to (a) shows a histogram of measured Hamming distances at $q = 0.4$ and $\varepsilon = 0.01$ (up arrow).

ε = fraction of initially flipped bits.

no. of initially flipped bits = εN

Apparently we need $\varepsilon N \geq 5$.

Gene expression and regulation

Transcriptional regulation: Proteins called transcription factors bind to specific sequences of the DNA to help determine whether or not an individual gene produces its mRNA and the subsequent protein which, in turn, may then regulate another gene by binding to its associated DNA sequence.