Dynamics of Information Spreading in a Homogenous Population

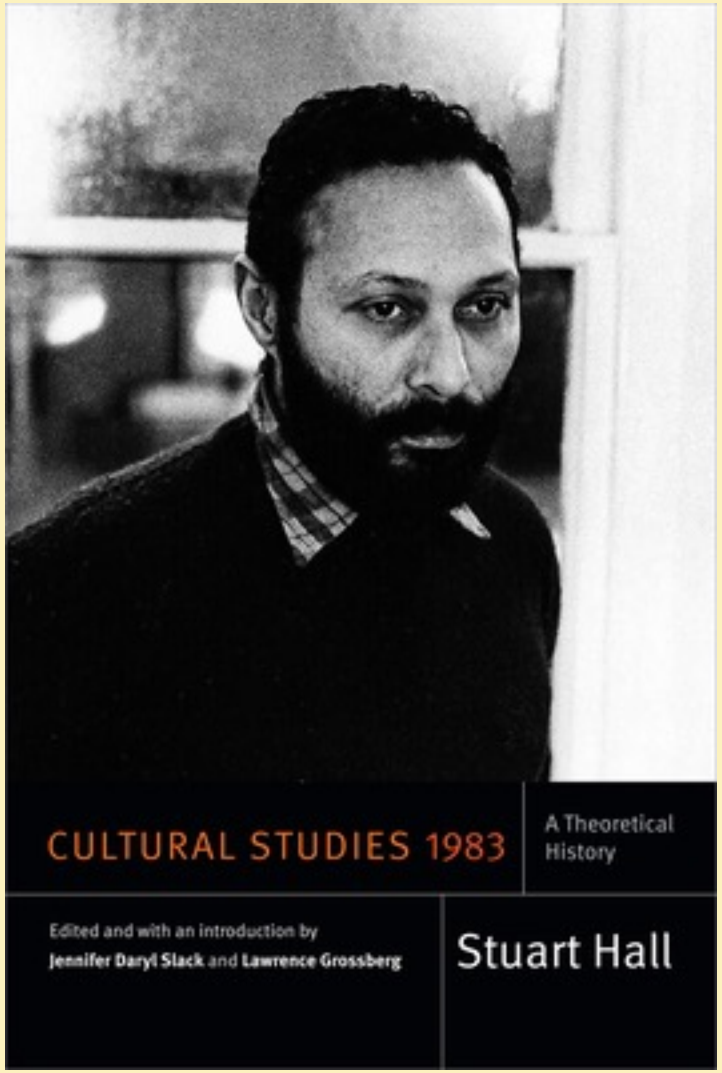
By Sagar Kumar

Overview

- Motivation

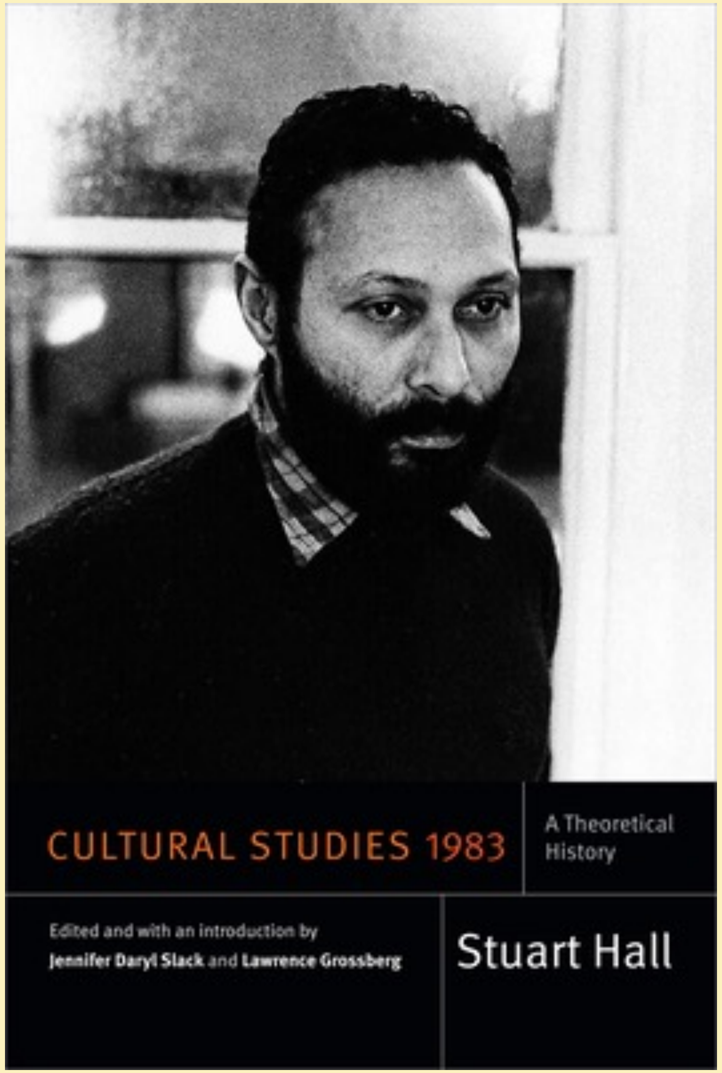
Overview

- Motivation
    - Problem



Overview

- Motivation
  - Problem
  - Question



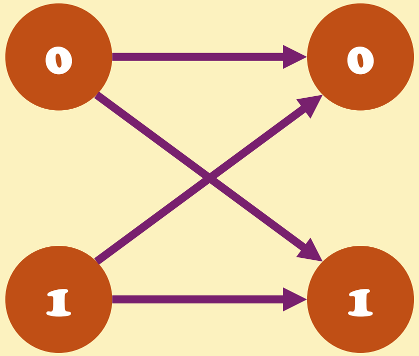
Overview

- Motivation
  - Problem
  - Question
- Background

Overview

- Motivation
  - Problem
  - Question
- Background
  - Shannon Entropy & Mutual Information

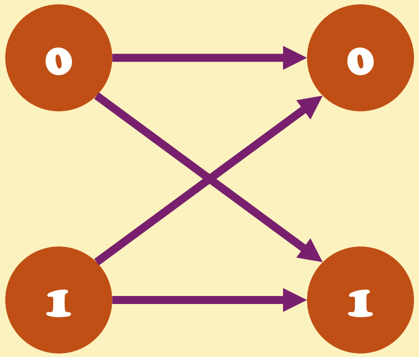
Overview

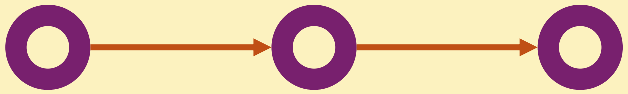


- Motivation
  - Problem
  - Question
- Background
  - Shannon Entropy & Mutual Information
  - Discrete Noisy Channels

Overview

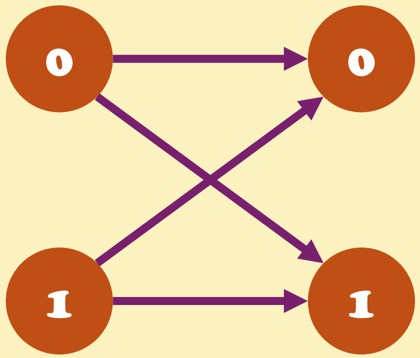
- Motivation
  - Problem
  - Question
- Background
  - Shannon Entropy & Mutual Information
  - Discrete Noisy Channels
  - Discrete Memoryless Relay Channels

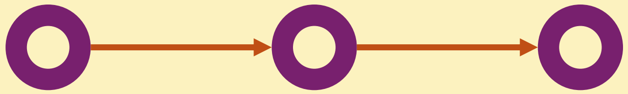


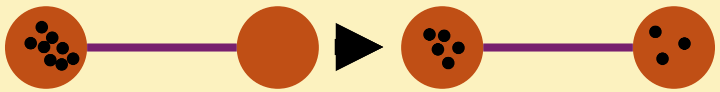


Overview

- Motivation
  - Problem
  - Question
- Background
  - Shannon Entropy & Mutual Information
  - Discrete Noisy Channels
  - Discrete Memoryless Relay Channels
  - Random Walks

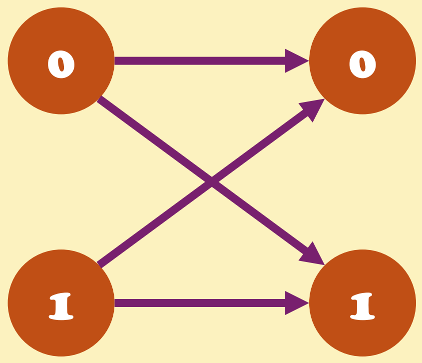


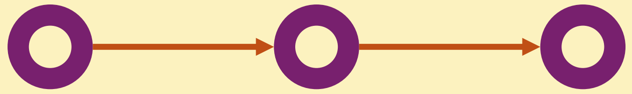


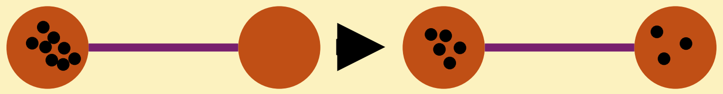


Overview

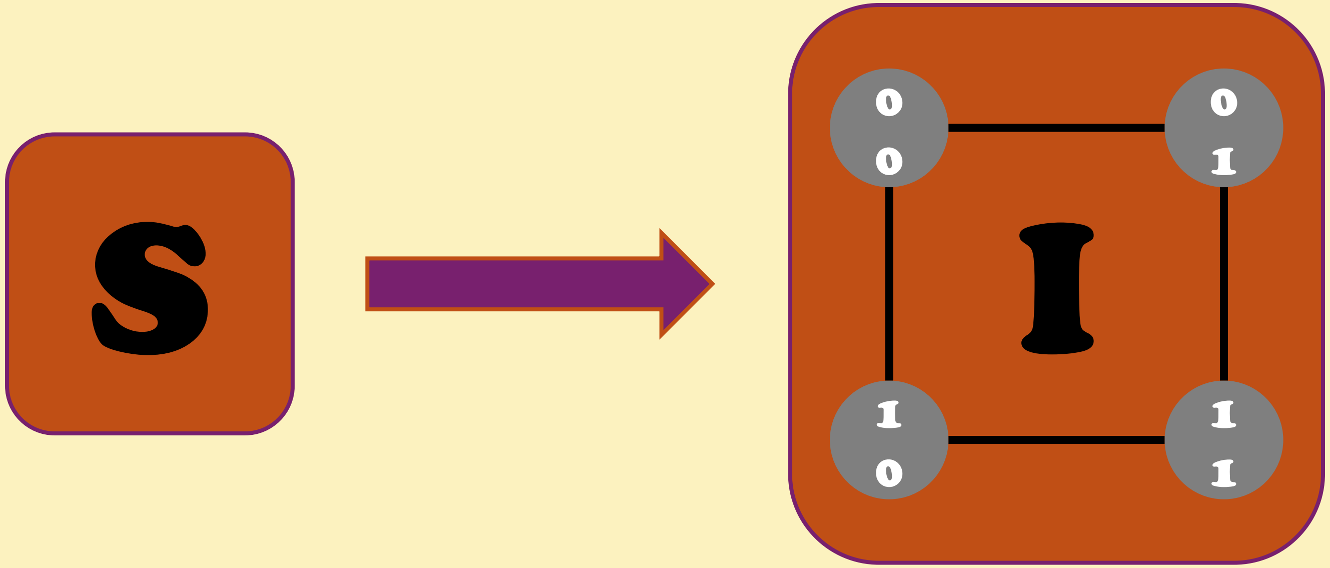
- Motivation
  - Problem
  - Question
- Background
  - Shannon Entropy & Mutual Information
  - Discrete Noisy Channels
  - Discrete Memoryless Relay Channels
  - Random Walks
  - SI(R) Model







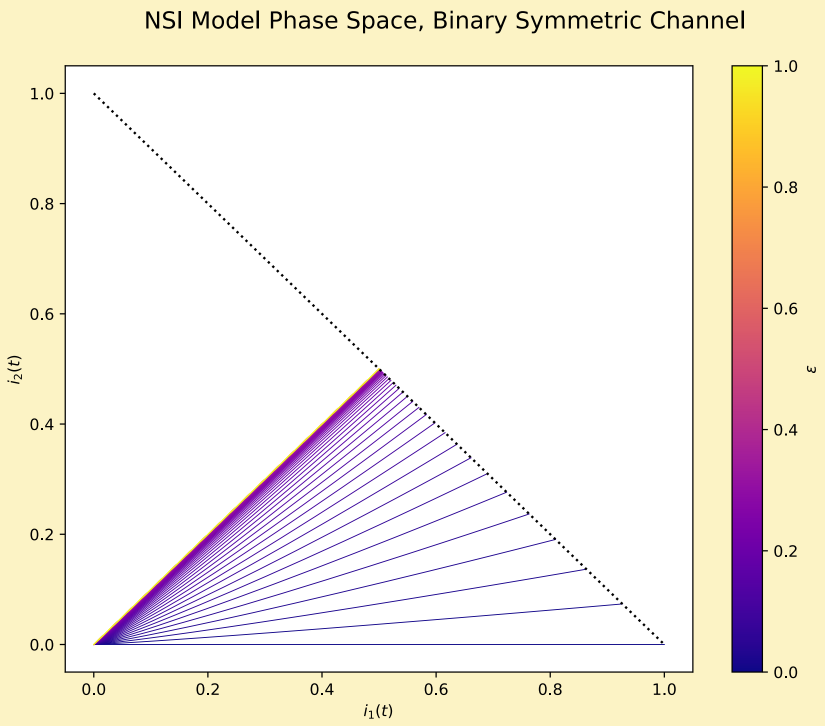
Overview



- Motivation
  - Problem
  - Question
- Background
  - Shannon Entropy & Mutual Information
  - Discrete Noisy Channels
  - Discrete Memoryless Relay Channels
  - Random Walks
  - SI Model
- Model

Overview

- Motivation
  - Problem
  - Question
- Background
  - Shannon Entropy & Mutual Information
  - Discrete Noisy Channels
  - Discrete Memoryless Relay Channels
  - Random Walks
  - SI Model
- Model
- Results



Overview

- Motivation
  - Problem
  - Question
- Background
  - Shannon Entropy & Mutual Information
  - Discrete Noisy Channels
  - Discrete Memoryless Relay Channels
  - Random Walks
  - SI Model
- Model
- Results
- Questions!

Problem

- Information does not “spread” in the real world, it is *communicated*
- Inherent to this process is the introduction of ***noise*** through the encoding and decoding processes.
- See Shannon, Krippendorff, Hall, etc.

Question

Suppose I hear a piece of news from a friend-of-a-friend-of-a-friend-of-a-... how much of that news can I expect to be true?

Motivation

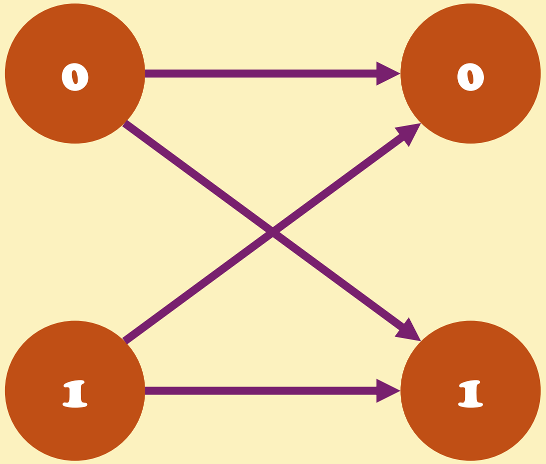
- Important to understand the emergence and nature of misinformation and propaganda
- Collective Intelligence & Group Decision-Making
- Useful in the study of strain dynamics in disease spreading

Background

- We will be relying on these tools heavily:
  - <u>Shannon Entropy</u> – A measure of the amount of information in a probability distribution. Consider this as the theoretical limit of what can be expected to be known with certainty given a distribution.
  - <u>Mutual Information</u> – A measure of how much information can be transmitted across as “channel”

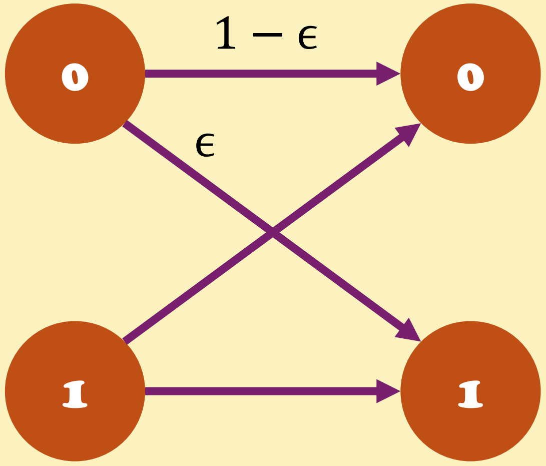
Background

Discrete Noisy Channels



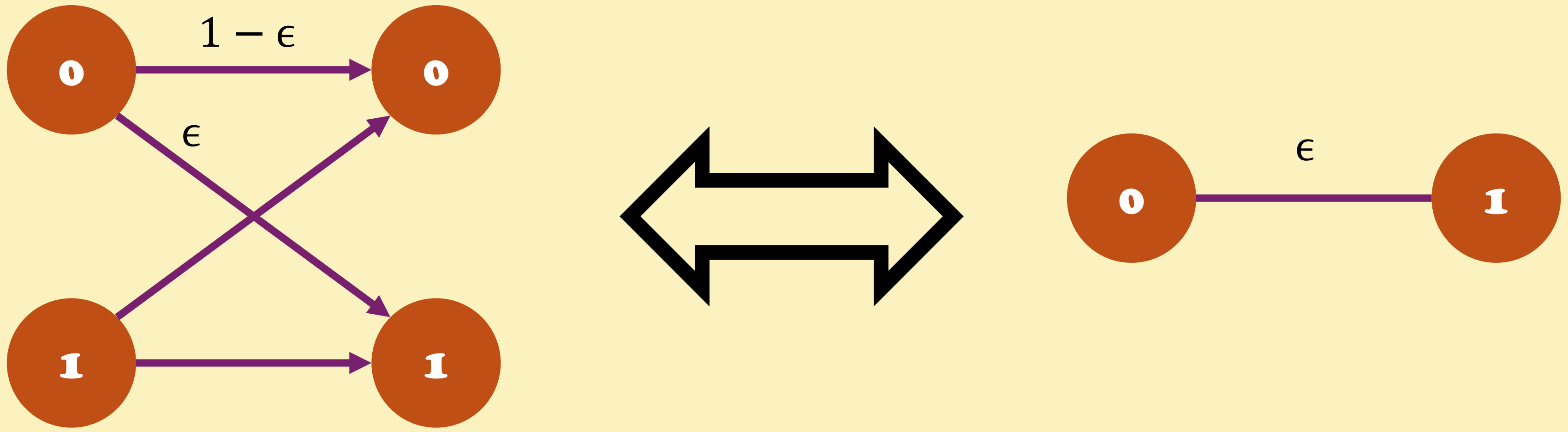
Background

Discrete Noisy Channels



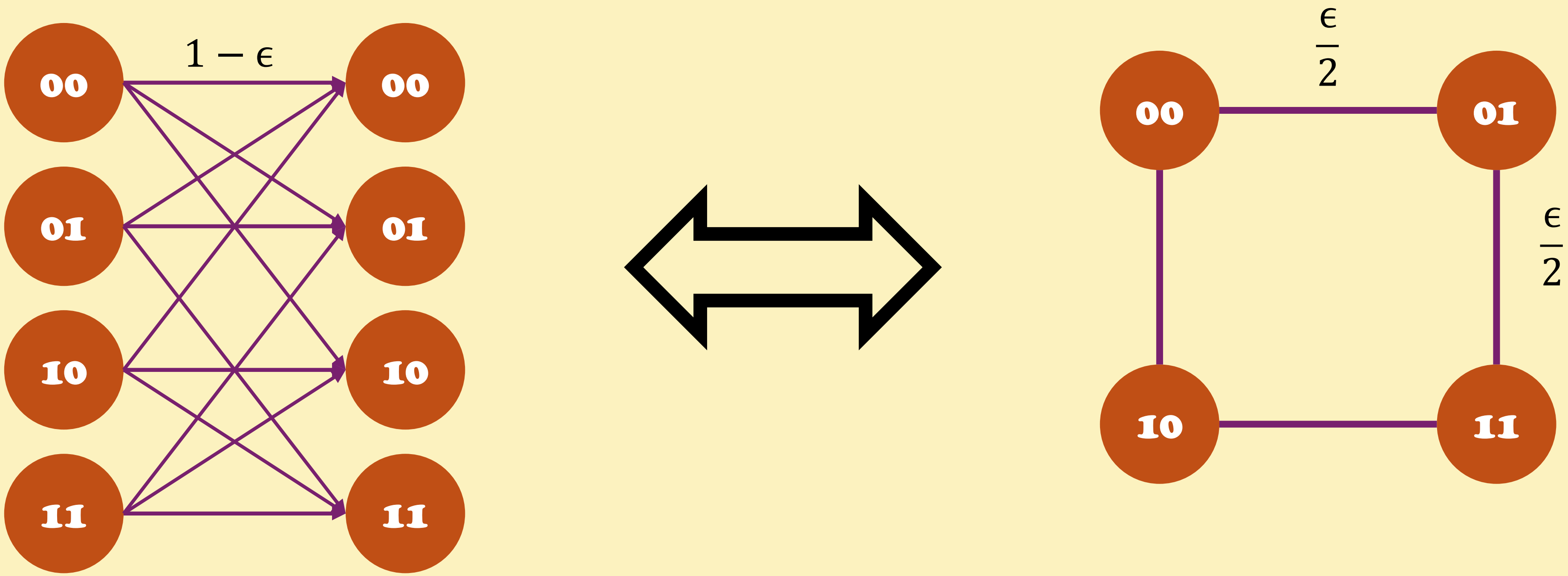
Background

Discrete Noisy Channels



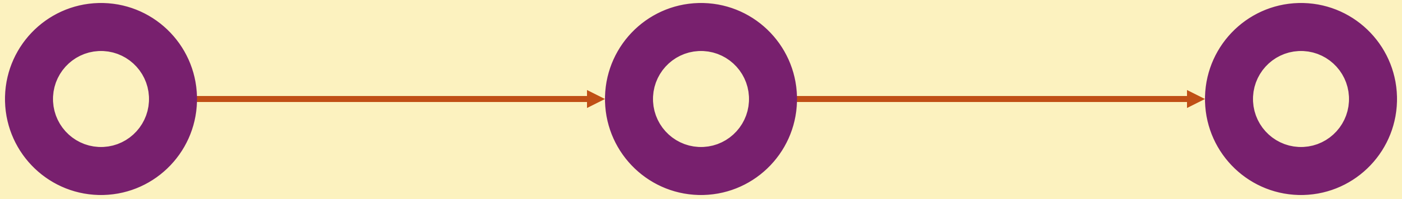
Background

Discrete Noisy Channels



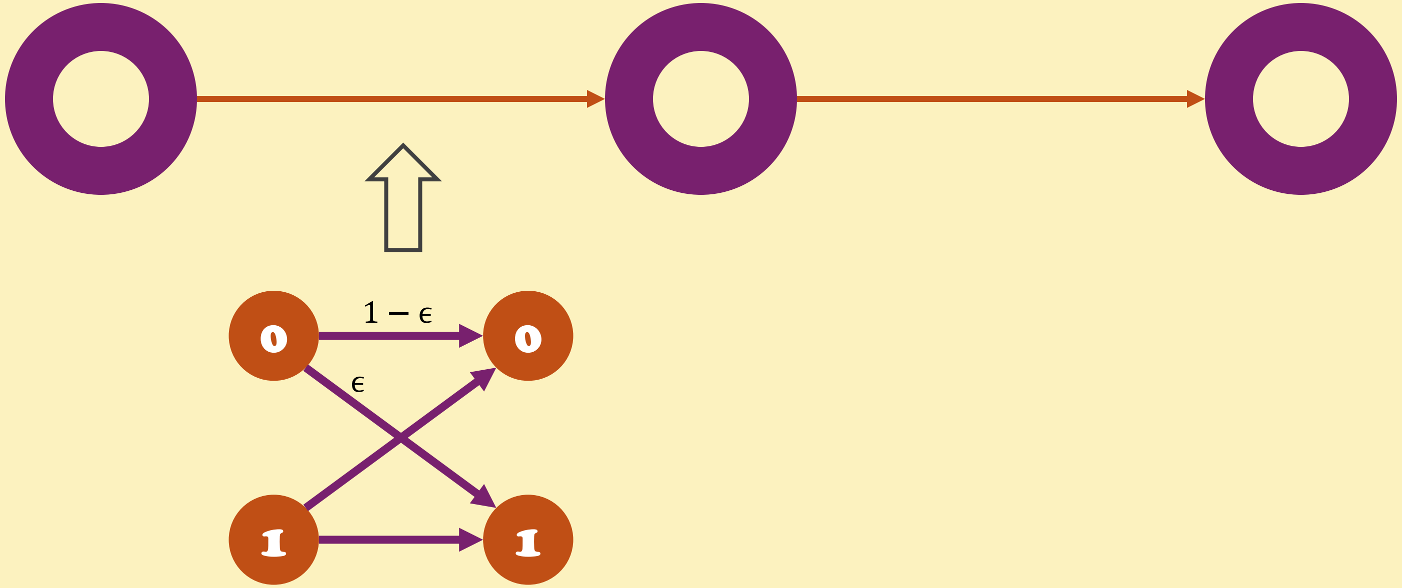
Background

Discrete (Memoryless) Relay Channels



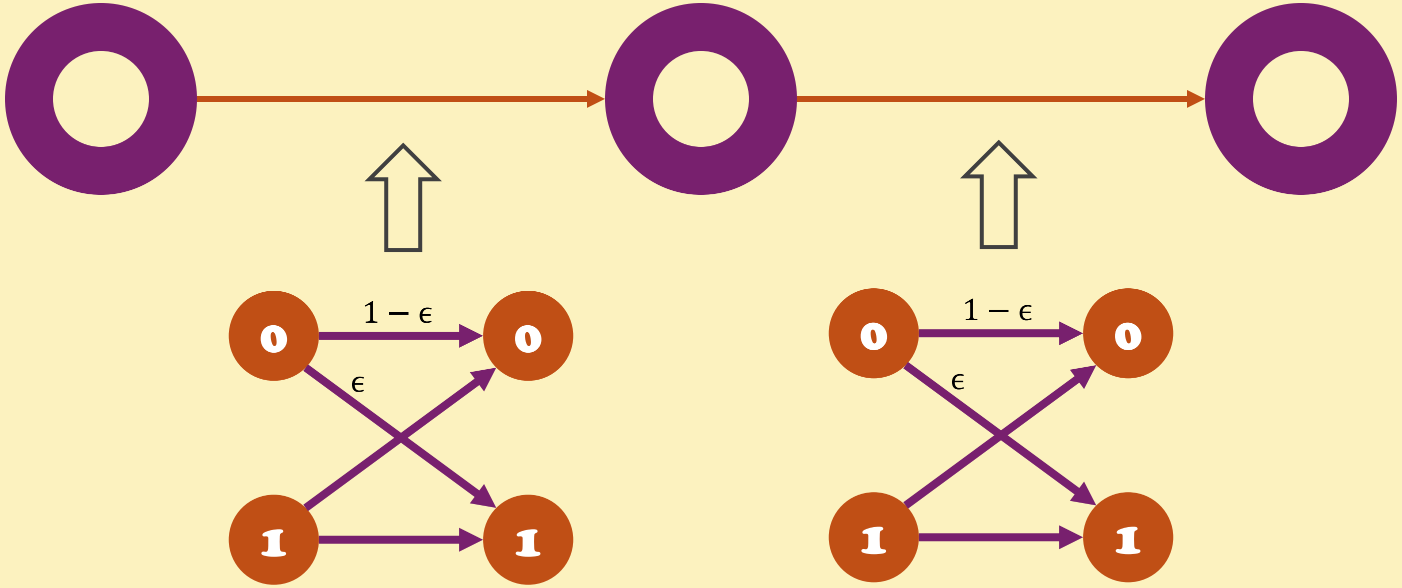
Background

Discrete Relay Channels



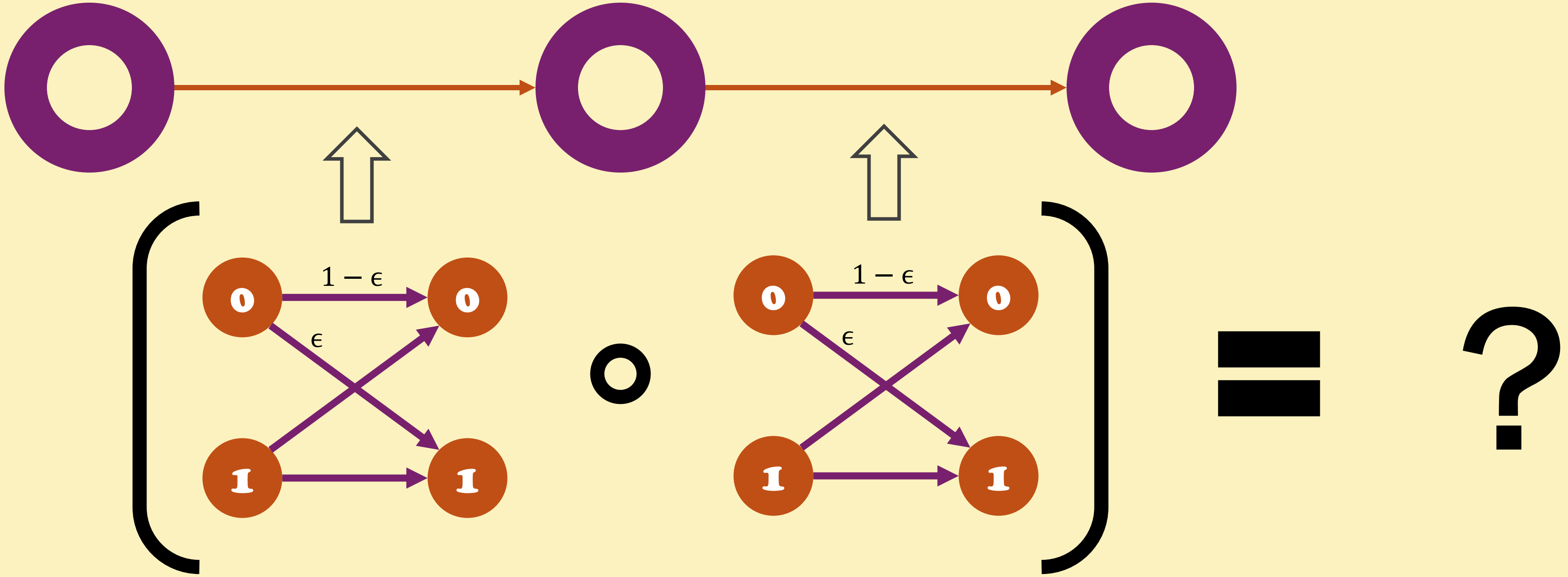
Background

Discrete Relay Channels



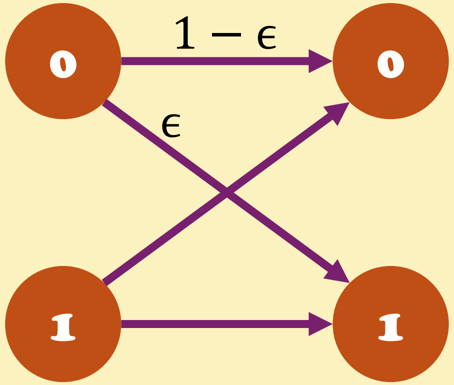
Background

Discrete Relay Channels

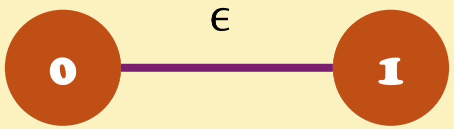


Background

Discrete Noisy Channels



$\Longleftrightarrow$



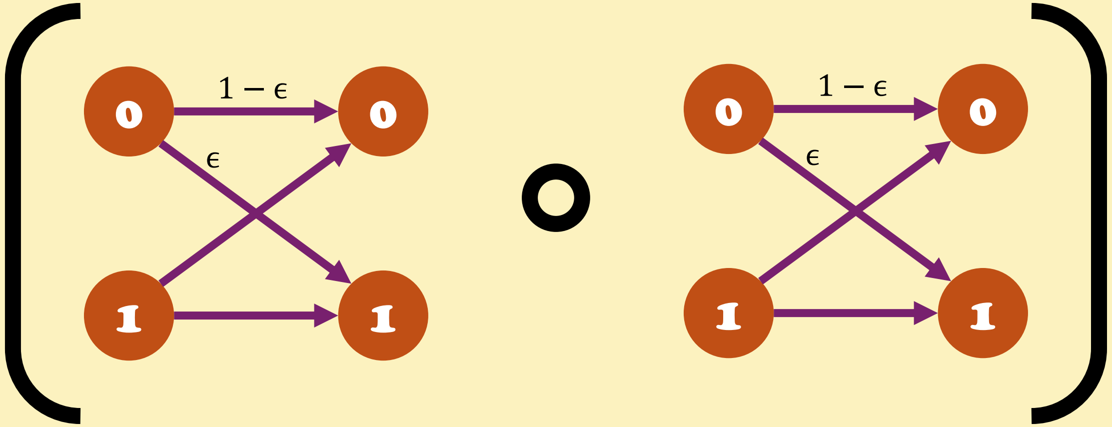
$\Longleftrightarrow$

$$\boldsymbol{Q}$$

$$\begin{pmatrix} \epsilon & 1-\epsilon \\ 1-\epsilon & \epsilon \end{pmatrix}$$

Background

Discrete Relay Channels

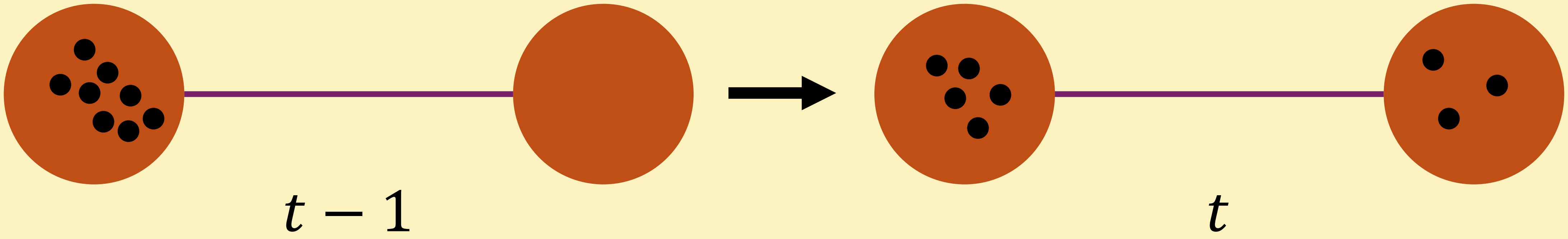


$$= \; Q[Q\vec{x}]$$

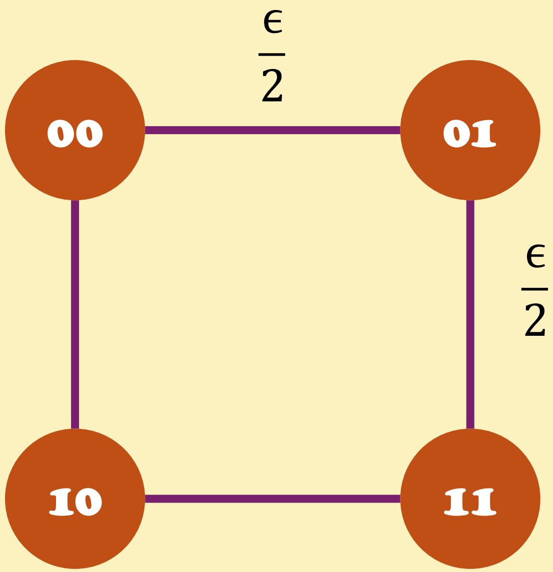
Background

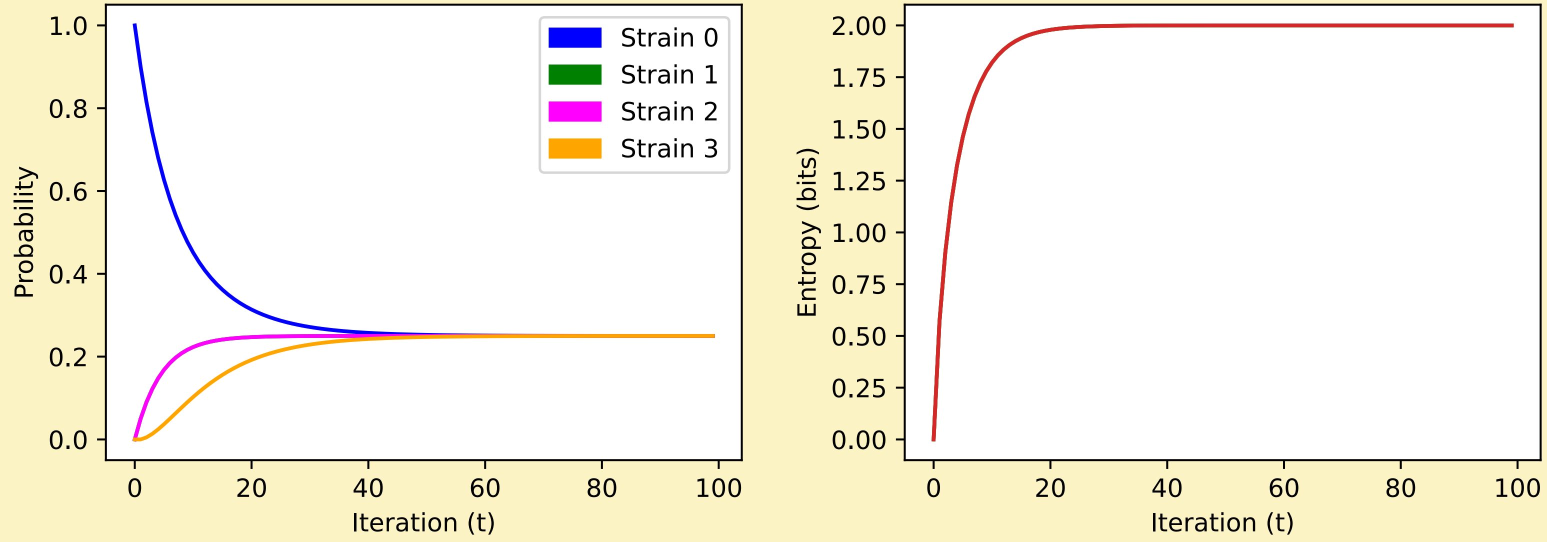
Random Walk / Dynamical System

$$\vec{x}_t = \boldsymbol{Q}\vec{x}_{t-1}$$



Background





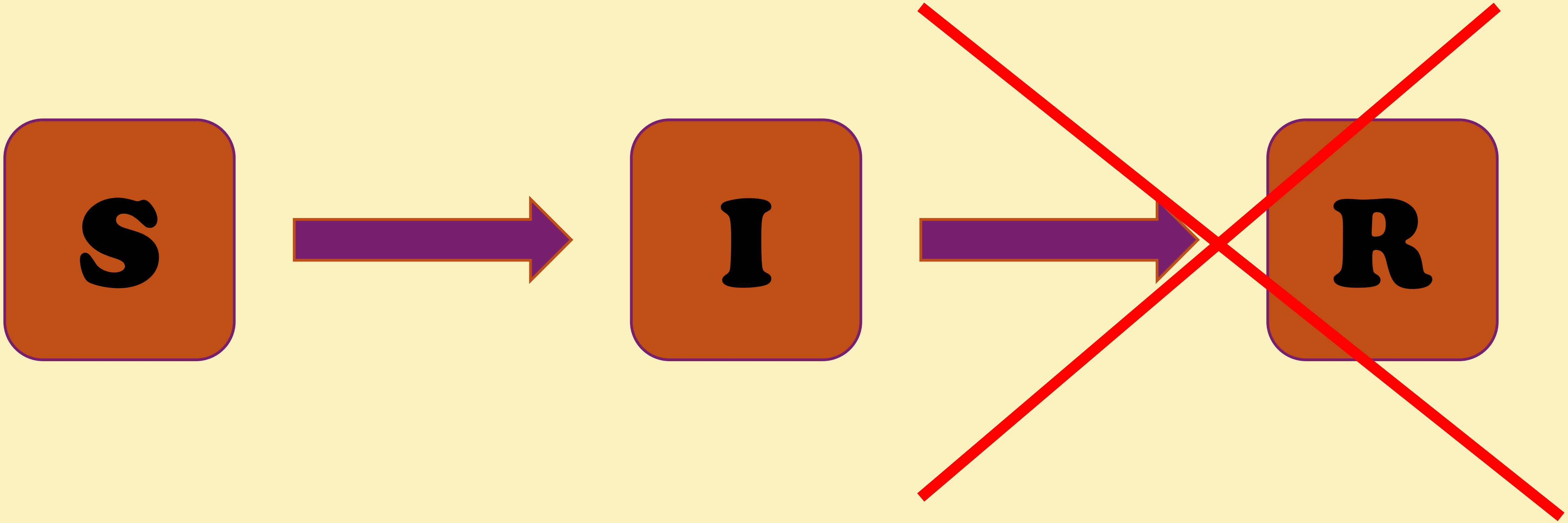
$$I(X_0; Y_t) \to 0$$

Wait…
We’re Forgetting Something

Background

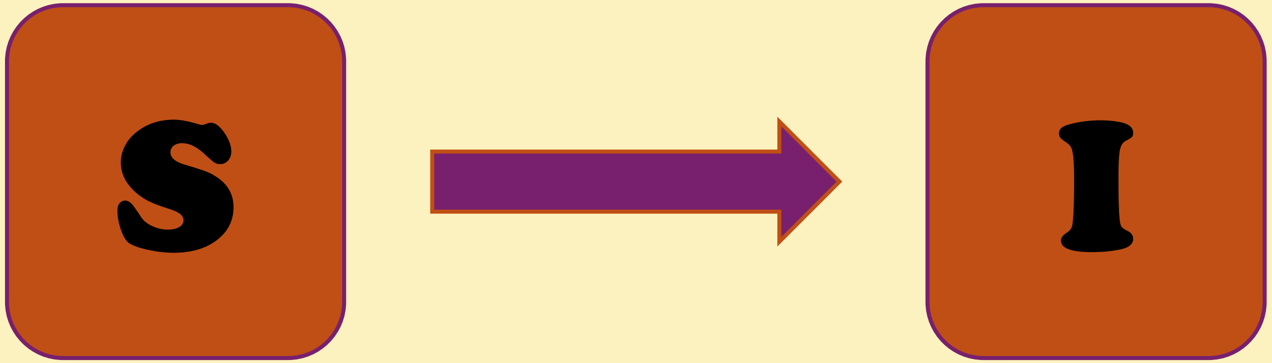
Background

Model



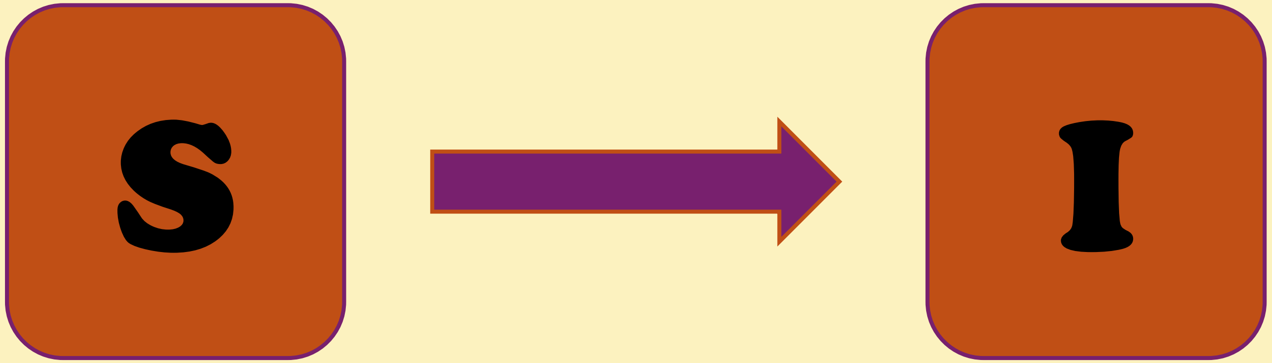
Model

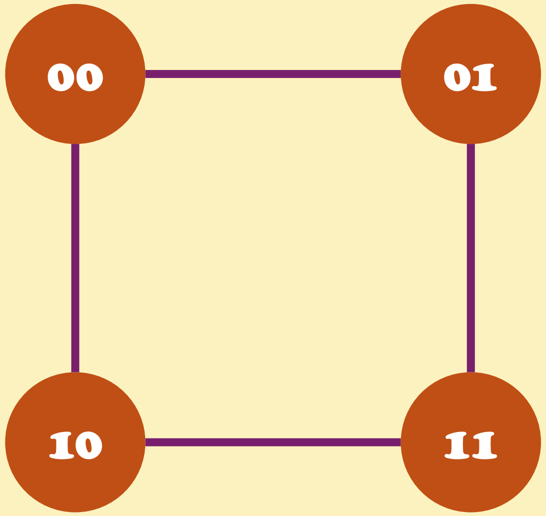
Noisy SI Model



Model

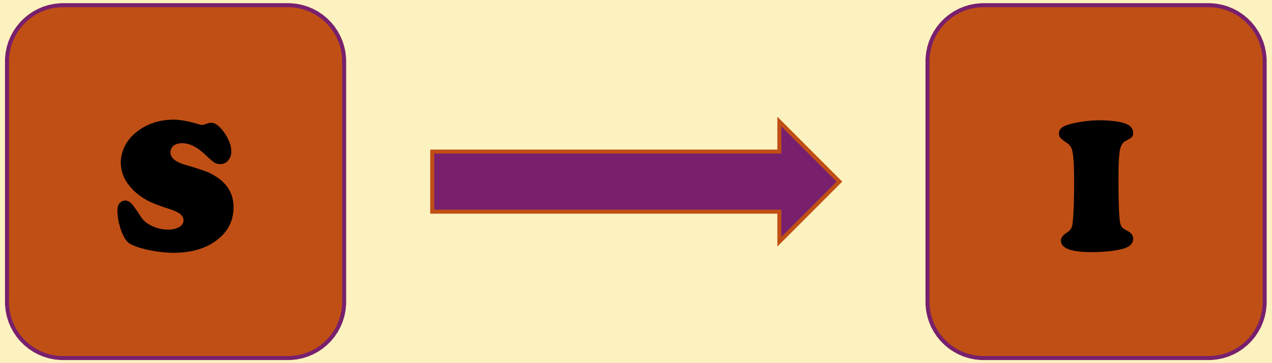
Noisy SI Model

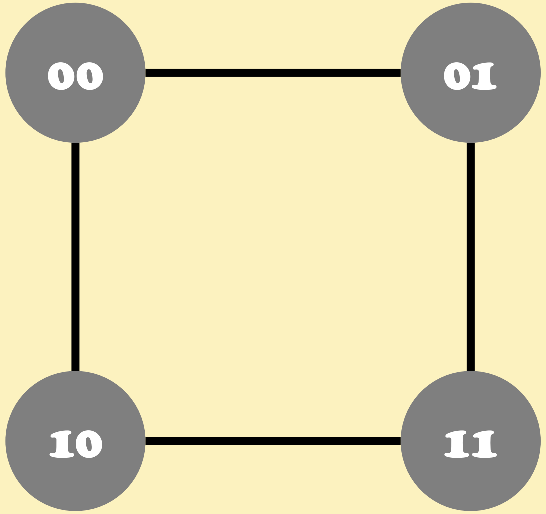




Model

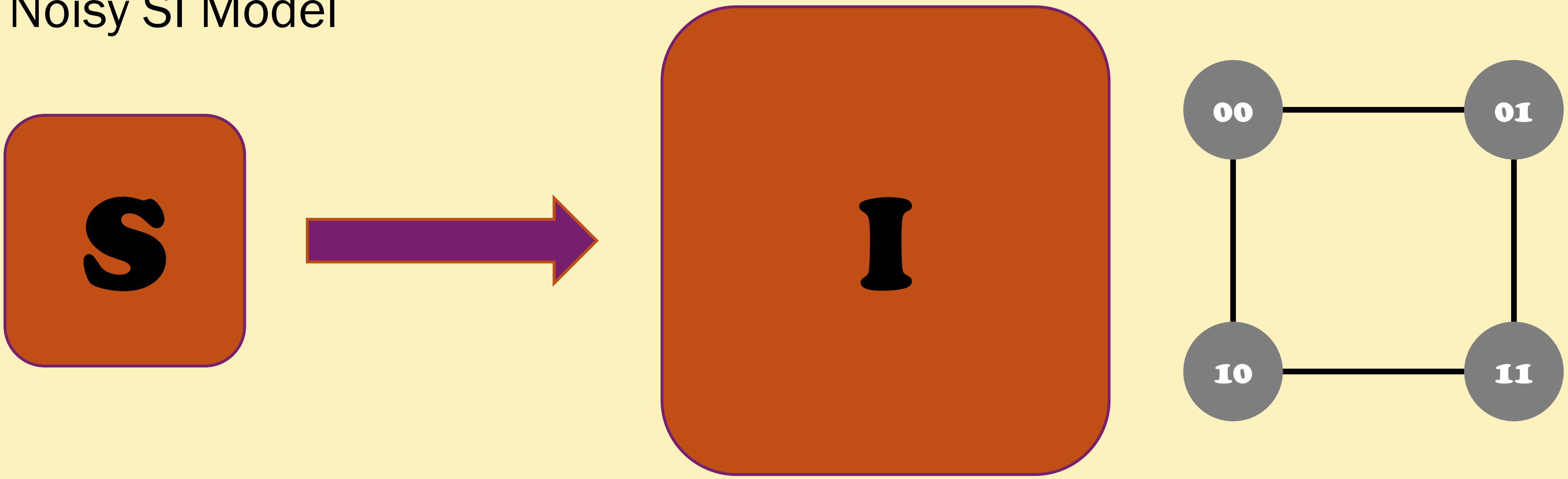
Noisy SI Model





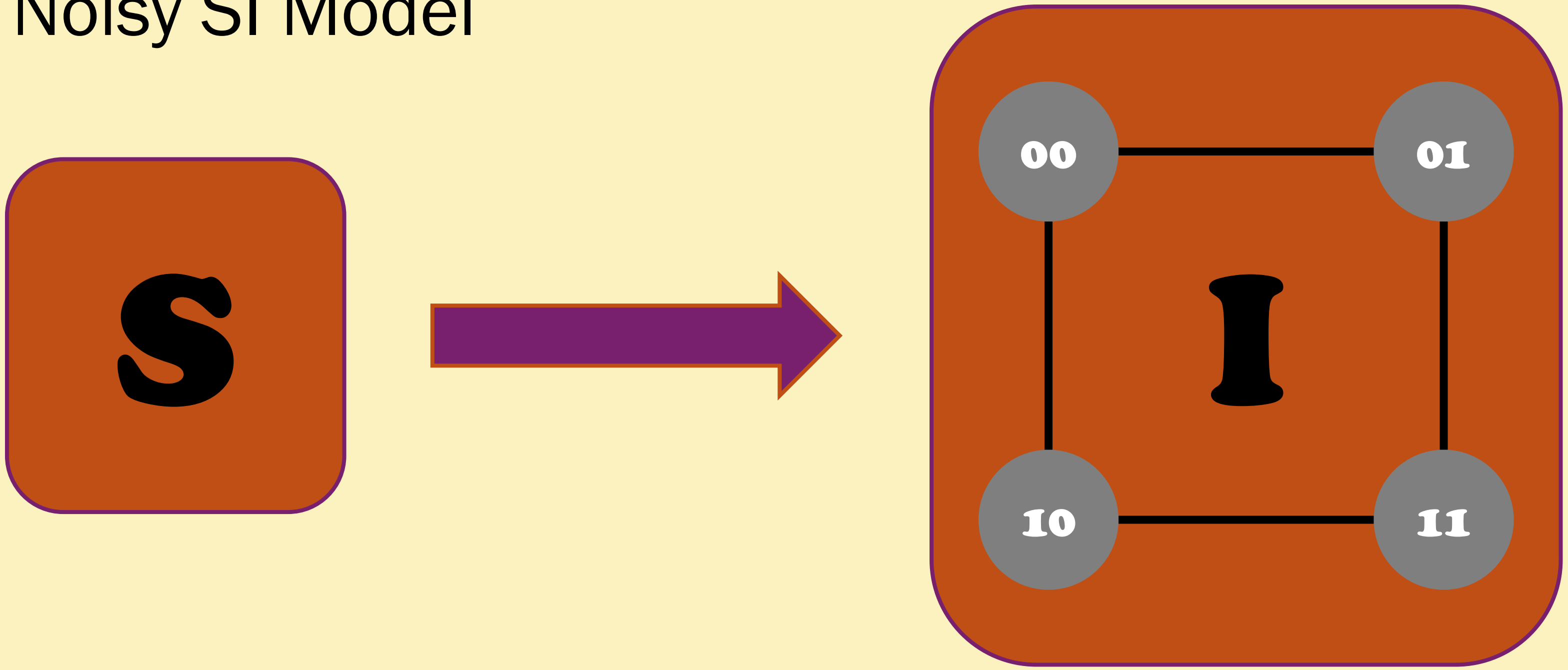
Model

Noisy SI Model



Model

Noisy SI Model



SI

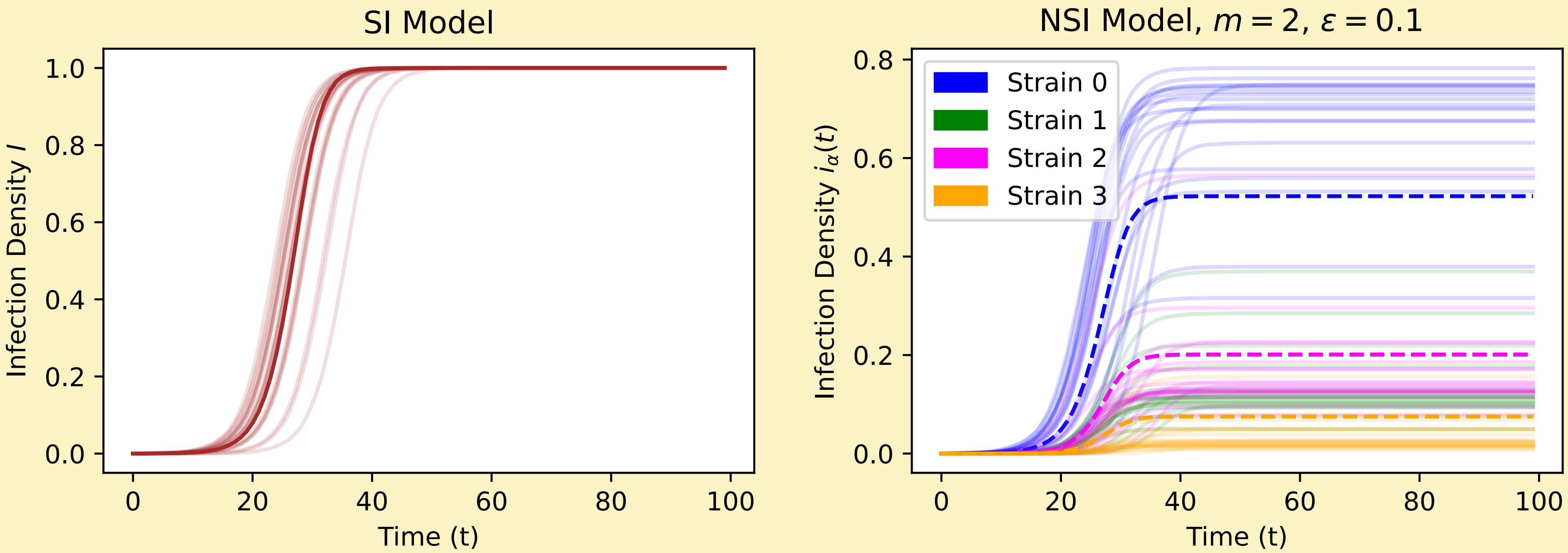
$$\frac{di(t)}{dt} = \beta\langle k\rangle\big(1 - i(t)\big)i(t)$$

$$\Downarrow$$

NSI

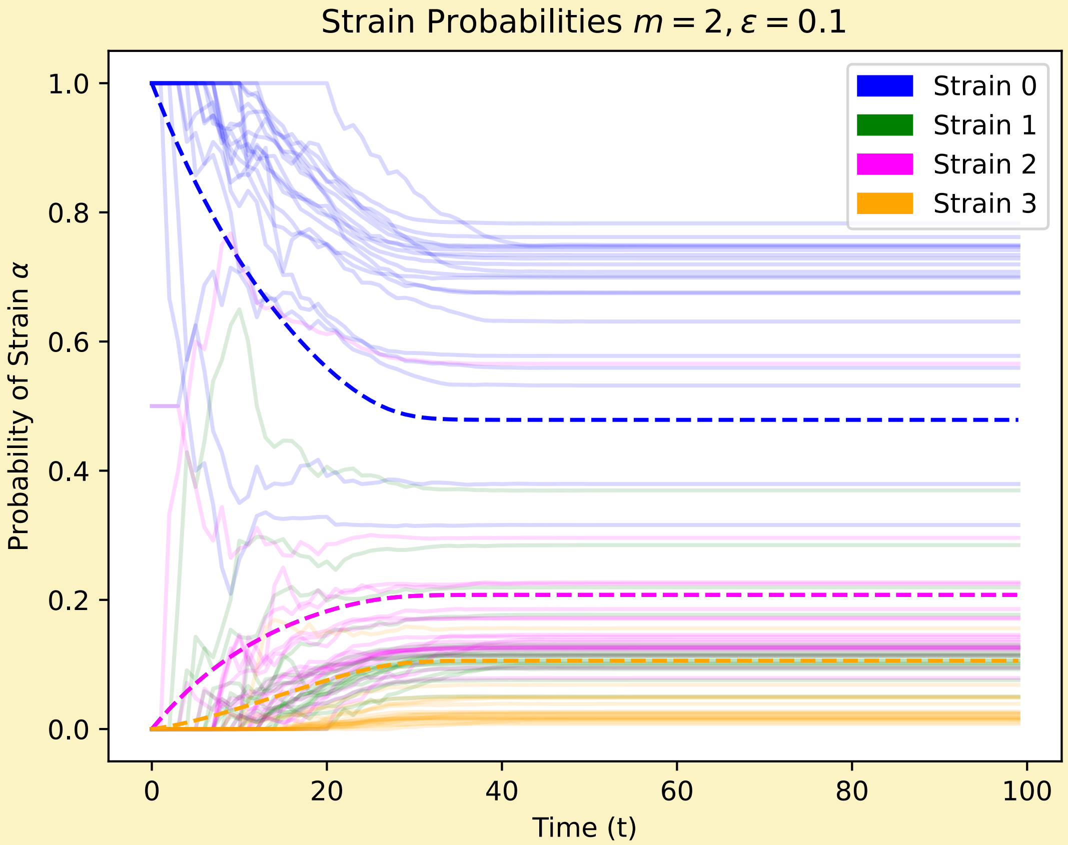
$$\frac{d\vec{i}(t)}{dt} = \beta\langle k\rangle\left(1 - \sum_{\alpha}\vec{i}_{\alpha}(t)\right)\boldsymbol{Q}\vec{i}(t)$$

Results



SI Model
Infection Density I
Time (t)
NSI Model, $m = 2$, $\varepsilon = 0.1$
Infection Density $i_\alpha(t)$
Strain 0
Strain 1
Strain 2
Strain 3
Time (t)

Results



$$\boxed{I(X_0; Y_t) \neq 0}$$

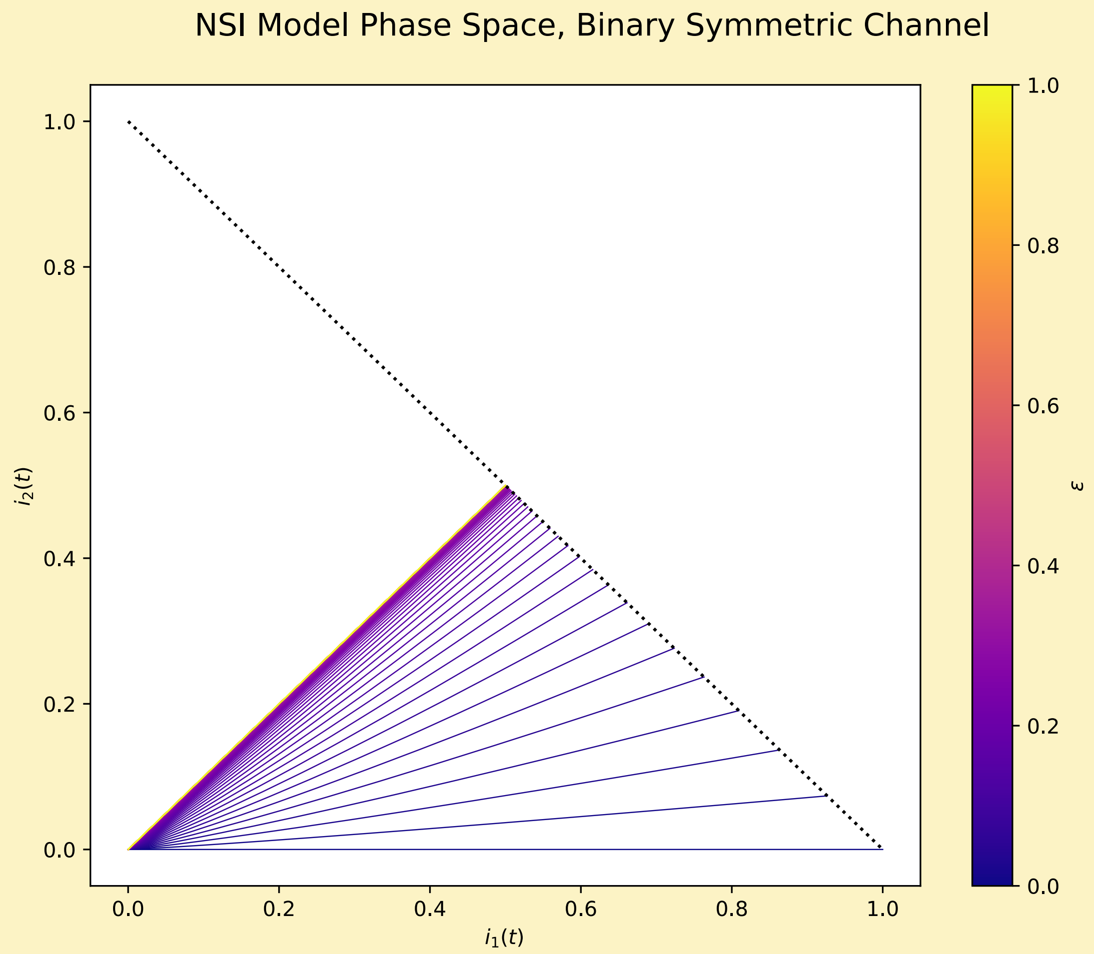
We can analytically calculate how much information we can expect to be "true".

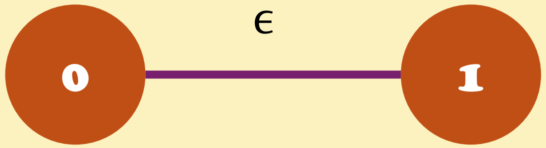
Results

Ok, but that is just for one model with fixed error.

Let’s start by varying the error

Results





$$H(Y) \neq 1$$

Results

SI

$$i(t) = \frac{i_0 e^{\beta \langle k \rangle t}}{1 + I_0 (e^{\beta \langle k \rangle t} - 1)}$$

NSI

$$\vec{i}(t) = \frac{\vec{i}_0 e^{\boldsymbol{Q} \beta \langle k \rangle t}}{1 + \vec{i}_0 (e^{\boldsymbol{Q} \beta \langle k \rangle t} - 1)}$$

Results

NSI

$$\vec{i}(t) = \frac{\vec{i}_0 e^{\boldsymbol{Q}\beta\langle k\rangle t}}{1 + \vec{i}_0 (e^{\boldsymbol{Q}\beta\langle k\rangle t} - 1)}$$

Results

NSI

$$\vec{i}(t) = \frac{\vec{i}_0 e^{\boldsymbol{Q}\beta\langle k\rangle t}}{1 + \vec{i}_0 (e^{\boldsymbol{Q}\beta\langle k\rangle t} - 1)}$$

$$\frac{\vec{i}_0 e^{\boldsymbol{Q}\beta\langle k\rangle t}}{1 + \vec{i}_0 (e^{\boldsymbol{Q}\beta\langle k\rangle t} - 1)} = \frac{\vec{i'}_0 e^{\boldsymbol{Q}\beta\langle k\rangle t}}{1 + \vec{i'}_0 (e^{\boldsymbol{Q}\beta\langle k\rangle t} - 1)}$$

Results

NSI

$$\vec{i}(t) = \frac{\vec{i}_0 e^{\boldsymbol{Q}\beta\langle k\rangle t}}{1 + \vec{i}_0 (e^{\boldsymbol{Q}\beta\langle k\rangle t} - 1)}$$

$$\frac{\vec{i}_0 e^{\boldsymbol{Q}\beta\langle k\rangle t}}{1 + \vec{i}_0 (e^{\boldsymbol{Q}\beta\langle k\rangle t} - 1)} = \frac{\vec{i'}_0 e^{\boldsymbol{Q}\beta\langle k\rangle t}}{1 + \vec{i'}_0 (e^{\boldsymbol{Q}\beta\langle k\rangle t} - 1)} \Longrightarrow \boxed{\vec{i_0} = \vec{i'_0}}$$

Results

Why is this important?

It shows us that regardless of the channel, the distribution of strains in the population is **uniquely** determined by the starting condition.

Thus, knowledge of the population statistics is fully informative of the initial state.

Results

Moreover (at least for symmetric channels—others have to be tested), the dominant strain in the population will *always* be the initial seed.

If you want to know the truth, ask your neighbors and accept the majority.

For large, homogenous populations, the # of neighbors required to ask can be arbitrarily small

Results

This mathematics holds in a number of important epidemiological settings as well:

1. Inferring initial strain
2. Inferring location of the initial case in a metapopulation model
    1. If mobility dynamics are uncertain but strain is strongly localized, this also works

Future Work

- Measuring and understanding the effects of structured populations on information dynamics in the system
  - Early evidence suggests that there is some self-correction or non-monotonous entropy increase, likely due to loops
- Empirical testing
- Implications for studies in communication and behavioral research
- Possibilities of reverse-engineering the channel from the final spread