# Darwin's daemon and a mechanistic approach to evolution

**Stefan Thurner**

www.complex-systems.meduniwien.ac.at www.santafe.edu

with R. Hanel and P. Klimek

Physical Review E 72, 036117 (2005)
Physical Review E 76, 030101 (2007)
Physical Review E 76, 046101 (2007)
European Physical Journal, B 62 (2008)
Europhysics Letters 82, 28008 (2008)
Scientific American (May 2008)
J. Theoretical Biology 256, 142146 (2009)
New Journal of Physics (2010) in print
Editorial of Nature Physics, (Nov 2009)

# What is evolution?

- Universal mechanism how diversity of a system changes endogenously

biological, ecological, chemical, social, industrial, economical, financial, historical

# What is an evolutionary process ?

Evolution is a three step process

- Step 1: new thing comes into being into a given environment

- Step 2: new thing has the chance to interact with environment. Result of this interaction: get selected or destroyed

- Step 3: if new thing gets selected (survives) in this environment it becomes part of this environment – it becomes part of the definition of the environment for all future new and arriving elements

Evolution is not physics

# Evolution is not physics

How does physics work?

Since Newton physics follows a recipe

(1) write down differential equations: $\frac{d^2x}{dt^2} = mg$
(2) fix boundary conditions: $x(0) = x_0$ , $v(0) = v_0$
(3) solve equations: $x(t) = \frac{1}{2}mgt^2 + v_0 t + x_0$
(4) test the solution in nature: drop stone, measure $x(t)$, compare

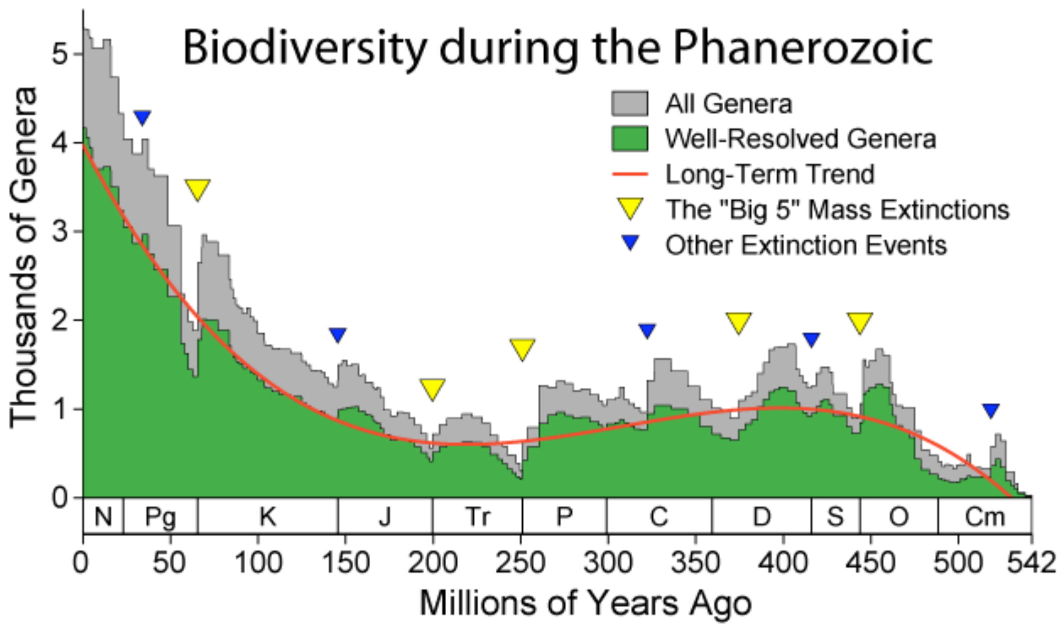
Two problems

(1) can not fix boundary conditions

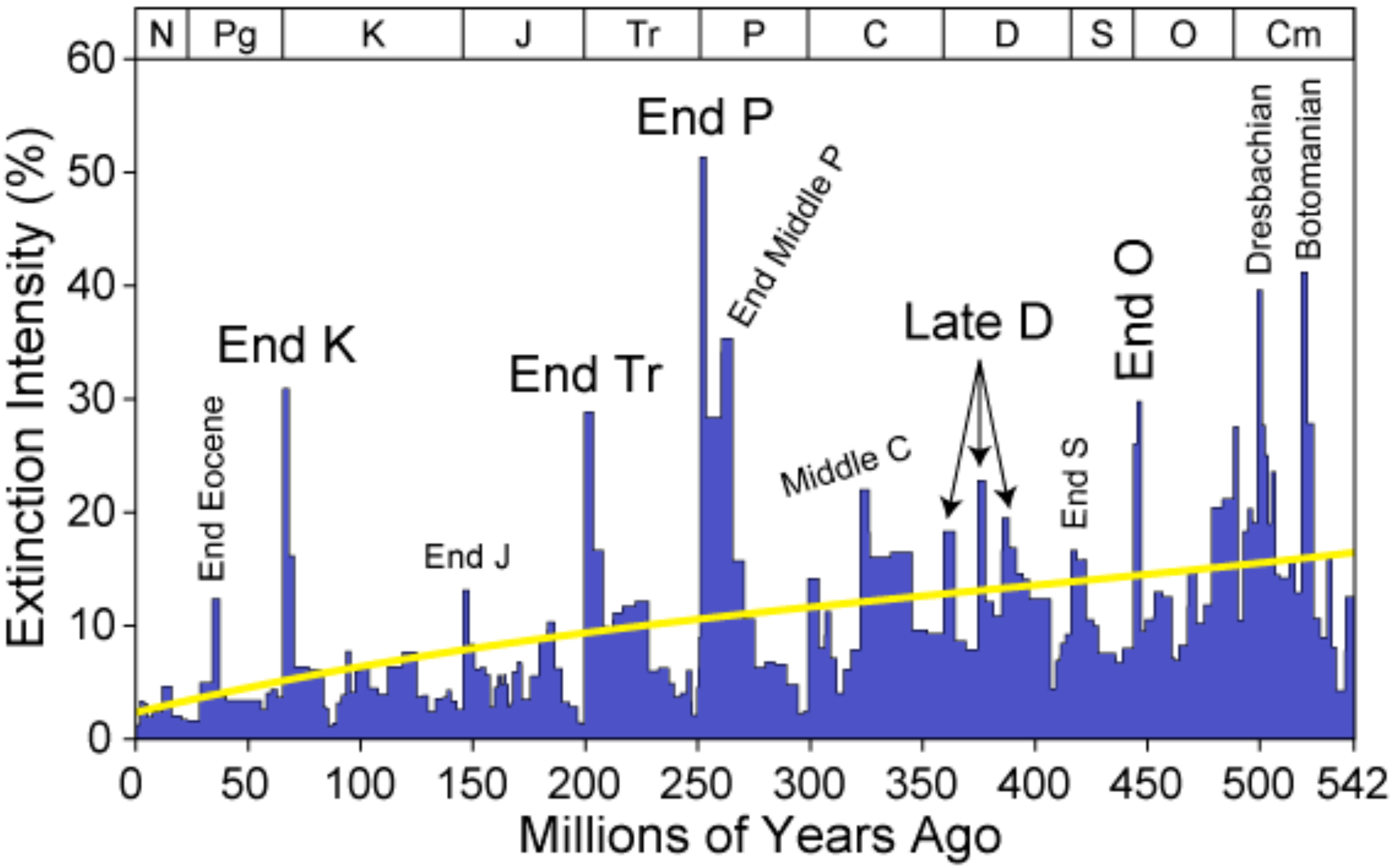
(2) can not prestate potential outcomes

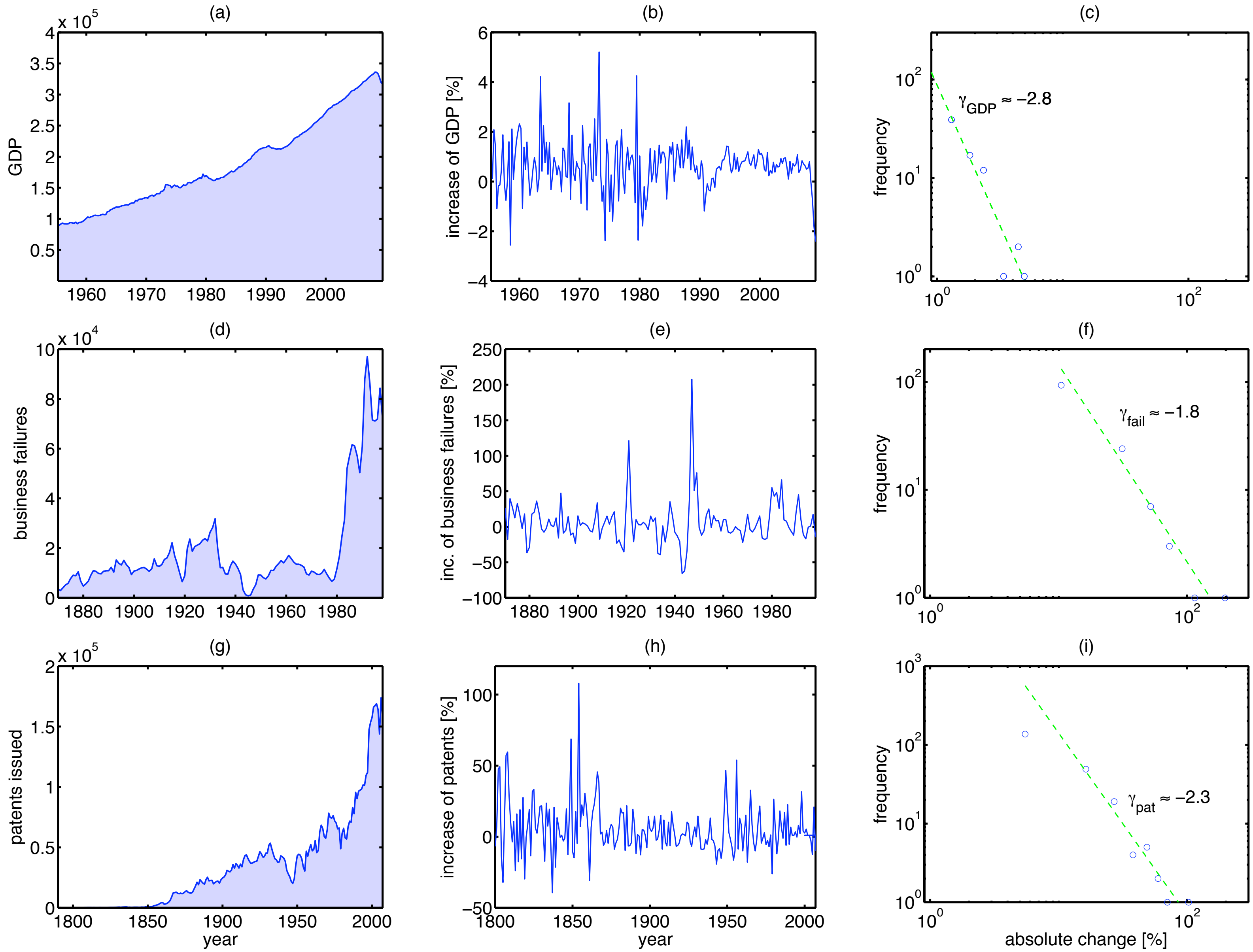
# Facts on evolution dynamics?

- Evolution involves fast and dramatic changes of diversity
- Bursts of massive creation and extinction happen with little / no precursor
- Transitional species are practically missing



# Marine Genus Biodiversity: Extinction Intensity




(a)
x 10^5
GDP
(b)
increase of GDP [%]
(c)
frequency
γGDP ≈ −2.8
(d)
x 10^4
business failures
(e)
inc. of business failures [%]
(f)
frequency
γfail ≈ −1.8
(g)
x 10^5
patents issued
year
(h)
increase of patents [%]
year
(i)
frequency
γpat ≈ −2.3
absolute change [%]

# Traditional view on evolutionary dynamics

Species $i$ produces species $k$ under the influence of species $j$ – can be expressed by a network equation of the type

$$\dot{x}_k = \sum_{i,j} \alpha_{kij} x_i x_j - x_k \Phi \quad , \quad \Phi = \sum_{i,j,k} \alpha_{kij} x_i x_j$$

• entities: $x_i$ • rules: $\alpha_{kij}$ (density of entries $r$)

| | | | | |
|---|---|---|---|---|
| replication | $x_i$ | $\rightarrow$ | $2x_i$ | $f_i$ |
| competition | $x_i + x_j$ | $\rightarrow$ | $x_j$ | $p_{ij}$ |
| mutation | $x_i$ | $\rightarrow$ | $x_j$ | $q_{ij}$ |
| recombination | $x_j + x_l$ | $\rightarrow$ | $x_i$ | $\alpha_{ijl}$ |
| development | $x_j + x_l$ | $\rightarrow$ | $p_i$ | $d_{ijl}$ |

# Problems

What can you measure?

- $x$: YES
- $\alpha$: NO

Even if one could $\rightarrow$ fundamental problem of predictive value of traditional of evolutionary dynamics

## Darwin's daemon – limitations of the fitness concept

Fitness-based population dynamics can not make quantitative, falsifiable predictions on long-term behavior of evolutionary systems

Thought experiment: *Darwin's Demon*

Suppose a demon has exact knowledge on the abundance and fitness of each species in the universe. (fitness $\equiv$ functional dependence of its proliferation rate on all other species). Daemon can measure fitness landscape exactly

What can the demon *predict*? Surprisingly little

New species is created $\rightarrow$ part of environment $\rightarrow$ environment changes

Demon has information on environment existing *before* arrival of new species

Thus the demon might have an exact description of the *current* biosphere, but with the advent of each new species this description loses accuracy.

Fitness always encodes *a posteriori* info $\rightarrow$ useless for falsifiable predictions

# Needed: co-evolution of fitness and species

Not fruitful to predict future fitness from present fitness

Instead one should understand how species and their fitness landscapes co-construct each other – *co-evolve*

We propose a variational principle in a spin-model-like setup

We derive a functional which is minimized under the most general evolutionary formulation of a dynamical system, i.e. evolutionary trajectories causally emerge as a minimization of a functional

# Three components of a model

System is characterized by

- States: open system $\longrightarrow$ infinite state vector. For simplicity binary
- Interactions: mutual influences of things $\longrightarrow$ production rate tensor. binary
- Stochastic component ($\beta$)

# Ingredients: Species

Species exist or not

$\sigma_i(t) = 1$ species $i$ exists at $t$

$\sigma_k(t) = 0$ species $k$ does not exist at $t$: because not produced yet, or got eliminated from the system

$N$-dimensional vector $\vec{\sigma}(t)$ contains all thinkable species

Species has to be produced by its components at every timestep

Production of species $i$ can happen only if all necessary components (parts) are simultaneously available

*Diversity* of system: $D(t) = \frac{1}{N}\sum_{i=1}^{N}\sigma_i(t)$

# How do new species (goods, ideas, ...) appear?

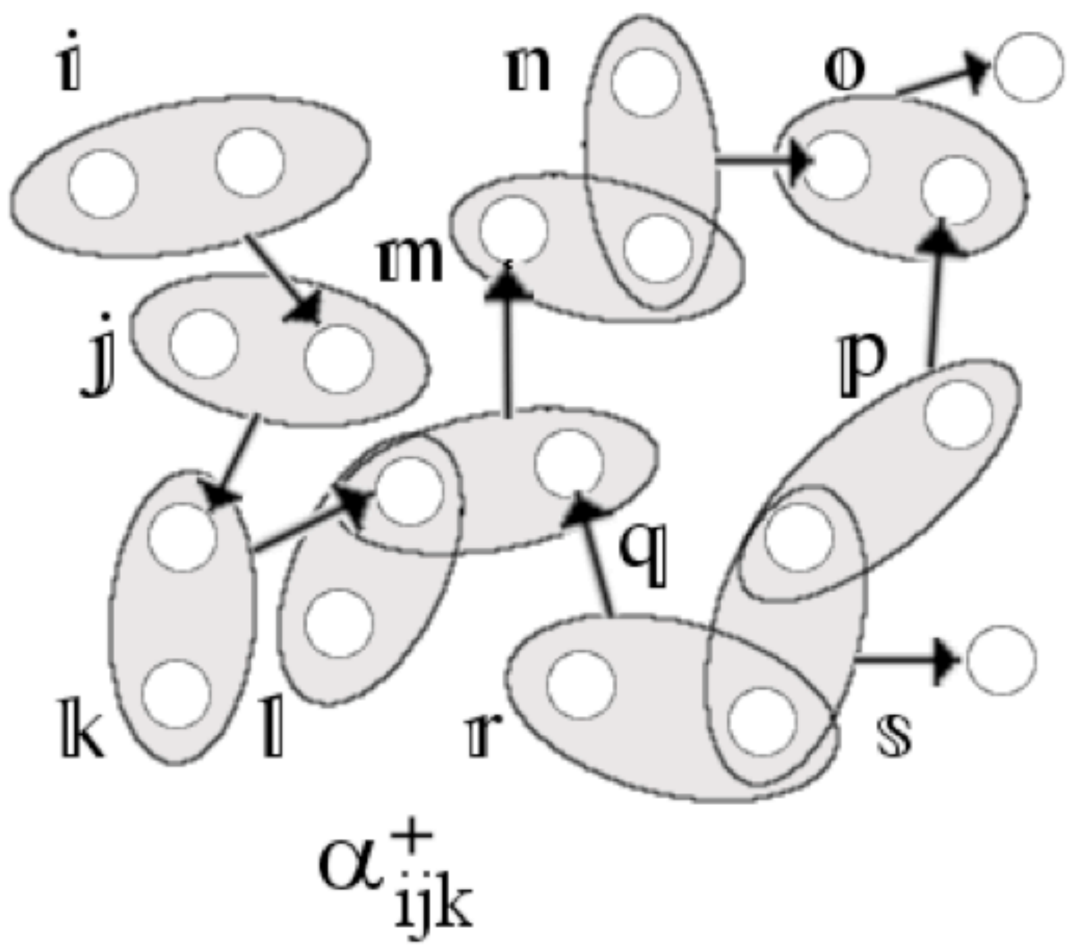
New things emerge through combination and substitution of existing things

Examples

- $\exists$ rule: a blacksmith can make a knife from a block of steel + a hammer
- $\exists$ rule: hydrogen with oxygen can form water
- $\exists$ rule: a cock and a hen can be the parents of a chicken
- $\neg\exists$ rule: a fish and a dog can be the parents of a chicken
- $\neg\exists$ rule: two blocks of U 235 can be welded into one big block of uranium

Whether a thing $i$ can be produced from components $j$ and $k$ is encoded in the production map, $\alpha^{+}_{ijk} = 0$ or $1$

# Ingredients: Production map

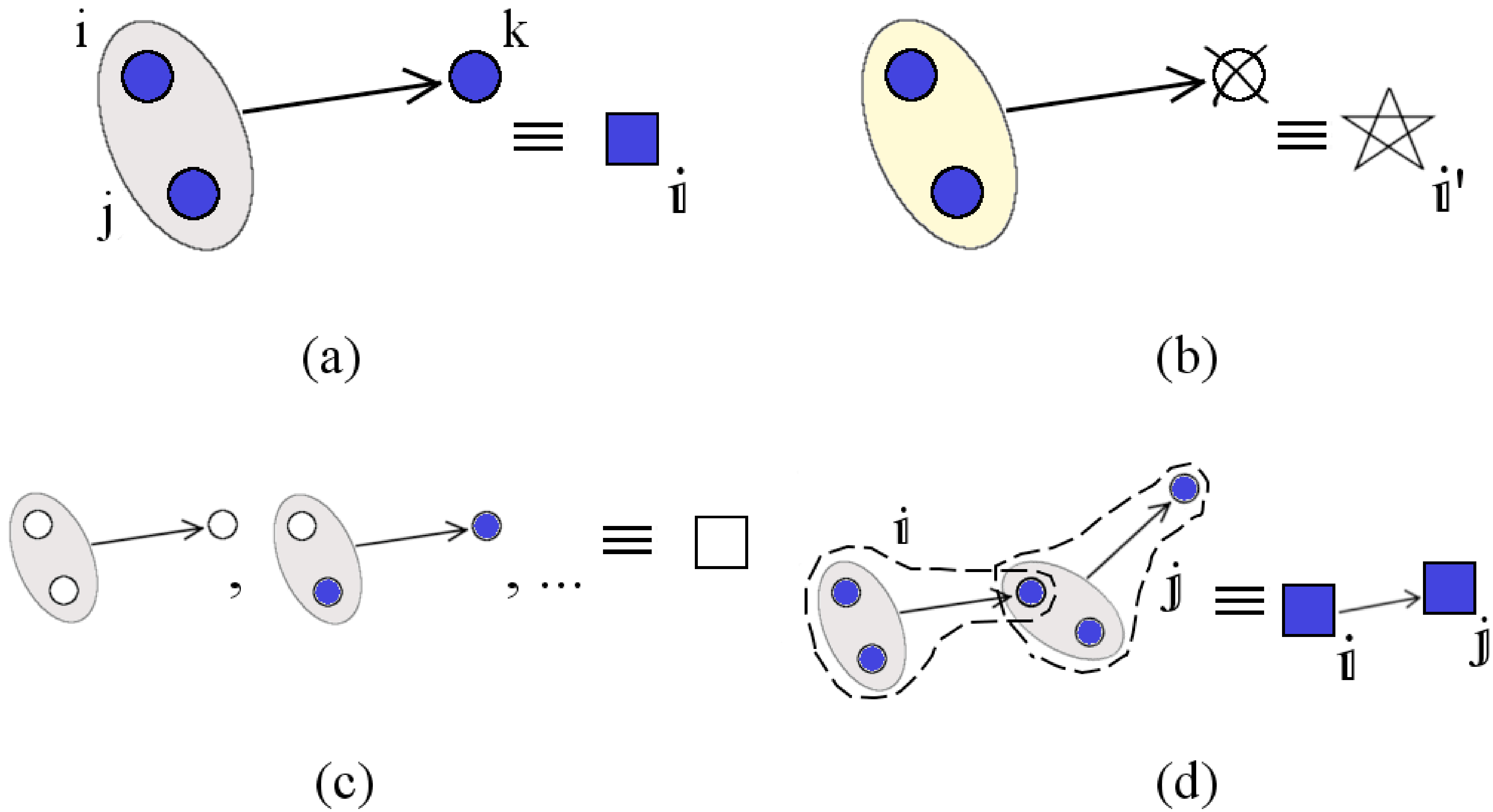


# Ingredients: Production

If possible to produce thing $i$ from $j$ and $k$, call it a production ( $\alpha^{+}_{ijk} = 1$)

If production actually produces $i$, call it active production
$(\sigma_i(t) = \sigma_j(t) = \sigma_k(t) = \alpha^{+}_{ijk} = 1)$

Production process: $\sigma_i(t+1) = \alpha^{+}_{kij}\sigma_j(t)\sigma_k(t)$


i
k
j
i
i'
(a)
(b)
i
j
i
j
(c)
(d)

# Ingredients: Selection/Competition/Destruction

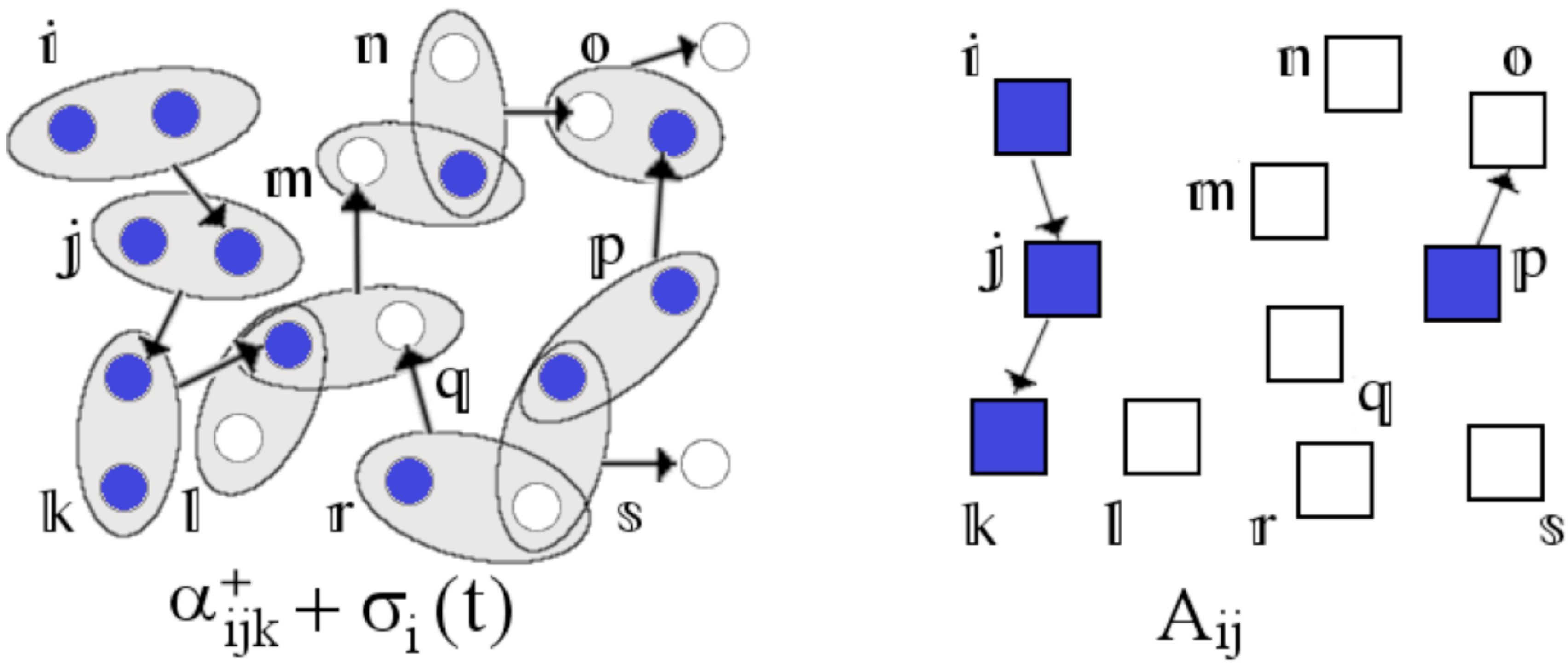
New species, good,... can cause competition with existing one: think of good which can be produced cheaper or that is more robust etc., will drive the other from the market, biosphere,...

Mechanism: combination of things $i$ and $j$ produces a thing $l$ which then drives product $k$ out.

*active destruction*: if $\sigma_i(t) = \sigma_j(t) = \sigma_k(t) = \alpha^-_{ijk} = 1$

Destruction process (if $\sigma_i(t) = 1$): $\sigma_i(t+1) = 1 - \alpha^-_{jki}\sigma_j(t)\sigma_k(t)$

# The active production network




Distinguish production rules ($\alpha^{+}$) and the active production networks $A(t)$

# Spontaneous inventions/mutations

• From time to time spontaneous ideas or inventions happen

• Introduce probability $p$ with which a non-existing good gets spontaneously invented or an existing good is spontaneously annihilated

# Summary of the model

- Phase space: $\Gamma \equiv \{0,1\}^N = \{\boldsymbol{\sigma} \,|\, \sigma_i \in \{0,1\}, 1 \le i \le N\}$, $N \gg 1$
- Define quadratic forms: $\Delta_i(\boldsymbol{\sigma}) \equiv \sum_{j,k}^{N} \alpha_{ijk}\sigma_j\sigma_k$ with $\alpha_{ijk} \in \{-1,0,1\}$
- Select a positive value $p \le 1$
- Dynamical update is the map $F : \Gamma \to \Gamma$ via the difference equation

$$\boldsymbol{\sigma}(t+1) = F(\boldsymbol{\sigma}(t)) \qquad , \qquad F = \boldsymbol{\Psi} \circ \boldsymbol{\Phi}$$

- Deterministic part $\boldsymbol{\Phi}(\boldsymbol{\sigma}) \equiv \mathbf{x}$ where $x_i = 1(0)$ when $\Delta_i(\boldsymbol{\sigma}) > 0(< 0)$, and $x_i = \sigma_i$, when $\Delta_i(\boldsymbol{\sigma}) = 0$.

- Stochastic part $\boldsymbol{\Psi} \equiv \mathbf{x}$, where $x_i = 1 - \sigma_i$ with probability $p$, and $x_i = \sigma_i$ else

# An evolutionary algorithm

Update from time $t$ to $t+1$

- pick a thing $i$ at random (random sequential update)
- count all active productive and destructive influences on $i$

For example: If $N_i^{\mathrm{prod}} > N_i^{\mathrm{destr}} \longrightarrow$ produce $i$

- with probability $p$ switch the state of $\sigma_i(t+1)$
- continue with next thing

Initial condition ($t=0$): fraction of randomly chosen initial things which exist

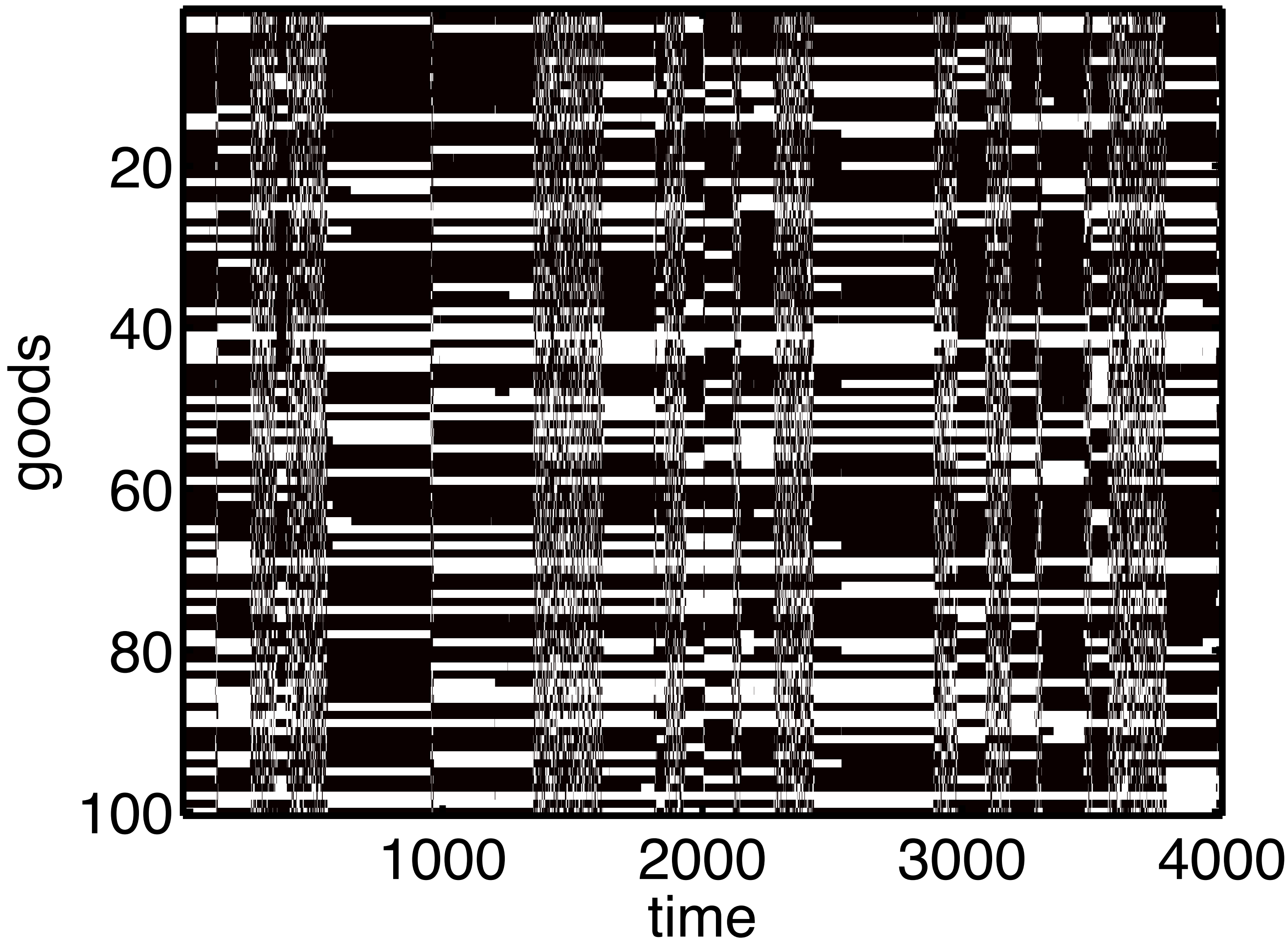
For simplicity: $\alpha_{ijk}$ binary random matrices, characterized by $r^{\pm}$

# Variables and parameters

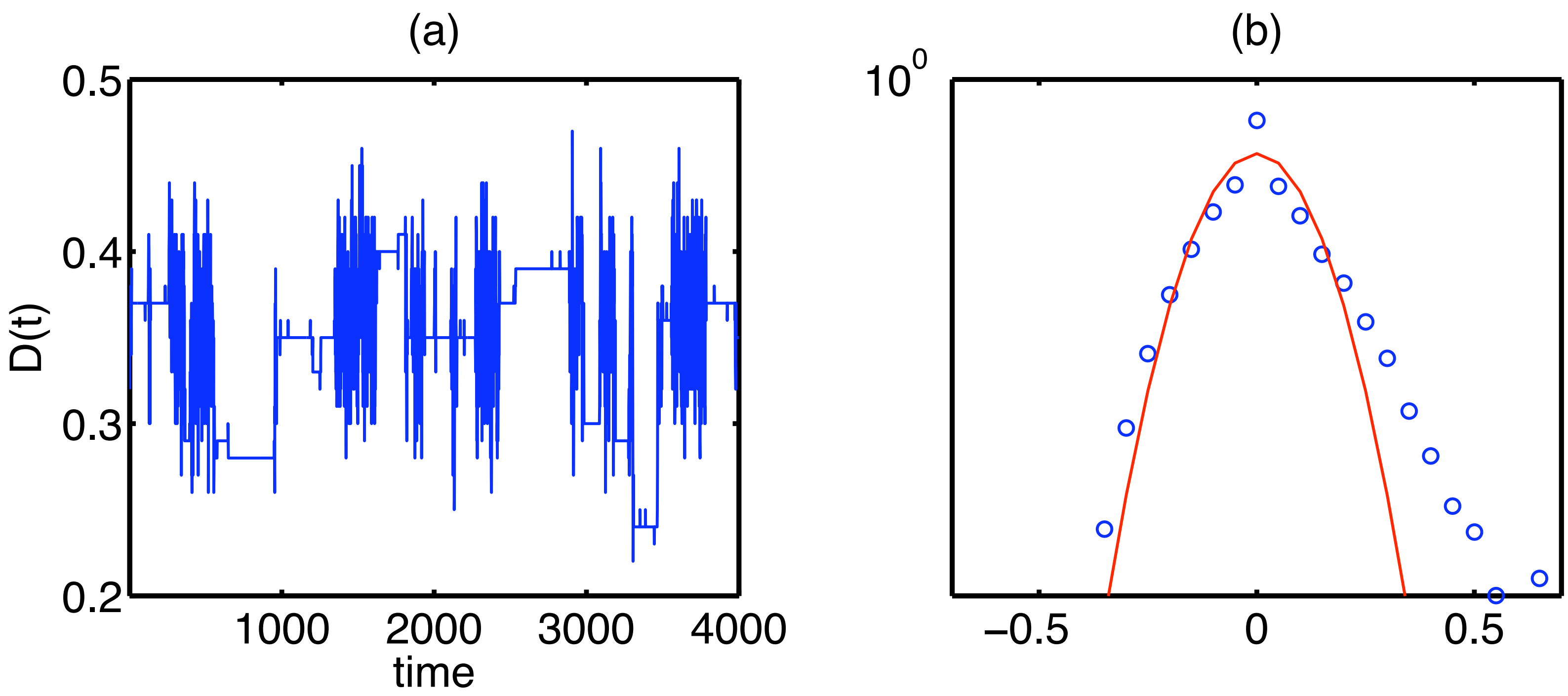
| Variable | Description | |
|---|---|---|
| $\sigma_i(t)$ | state of species $i$. exists / does not exist | dynamic |
| $D(t)$ | diversity at time $t$ | dynamic |
| $A(t)$ | active production network | dynamic |

| Parameter | | |
|---|---|---|
| $\alpha^{\pm}$ | productive/destructive interaction topology | fixed |
| $r^{\pm}$ | rule densities | fixed |
| $p$ | mutation rate / innovation parameter | fixed |
| $N$ | number of nodes | fixed |

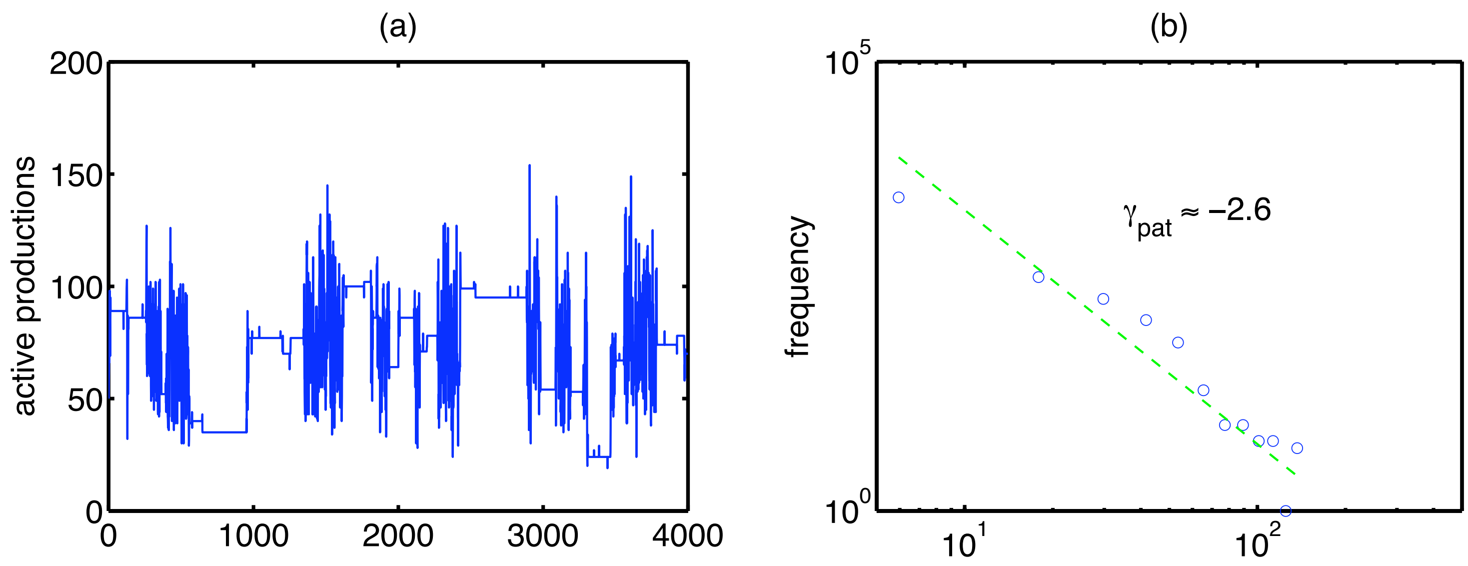
# Dynamics



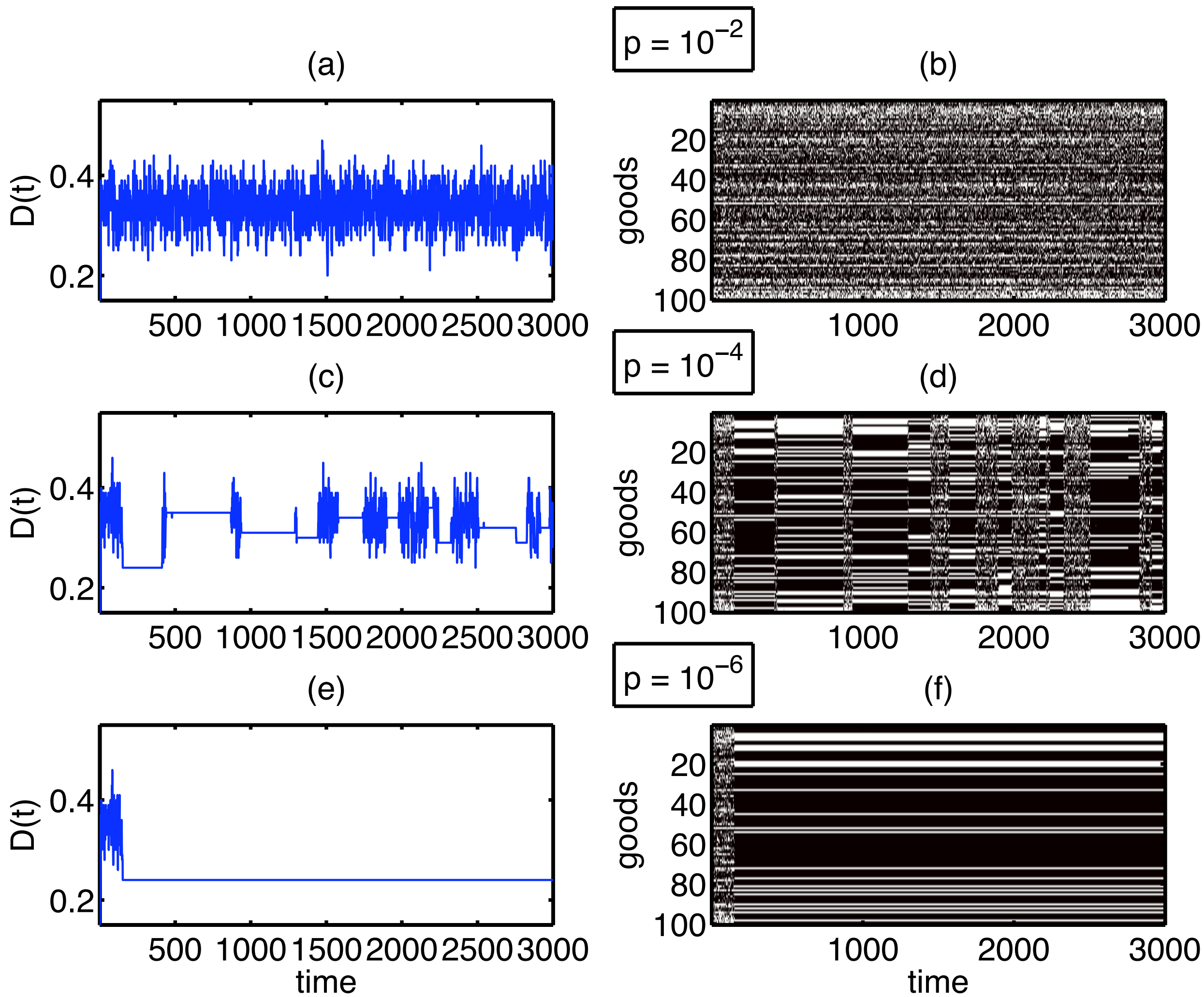
# Diversity dynamics



# Production dynamics – 'GDP'



# Dependence on invention rates



# Model variants

- More realistic competition

hierarchical suppression as a more realistic mechanism of competition

- Topology of production and destruction networks

*scale-free* production/destruction networks
Presence of hubs stabilizes system: increases lifetime of plateaus

- Asymmetry in production and destruction

easier to destroy than to 'create': one destructive influence stops production

- Modular structure of production/destruction networks

use modules $m$ (up to ten) with different random topologies $r_m^{\pm}$ linked by a variable number of connecting links

# Model variants II

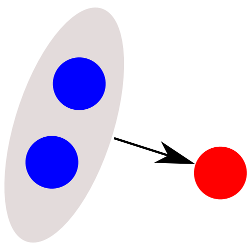
- Finite lifetime of things

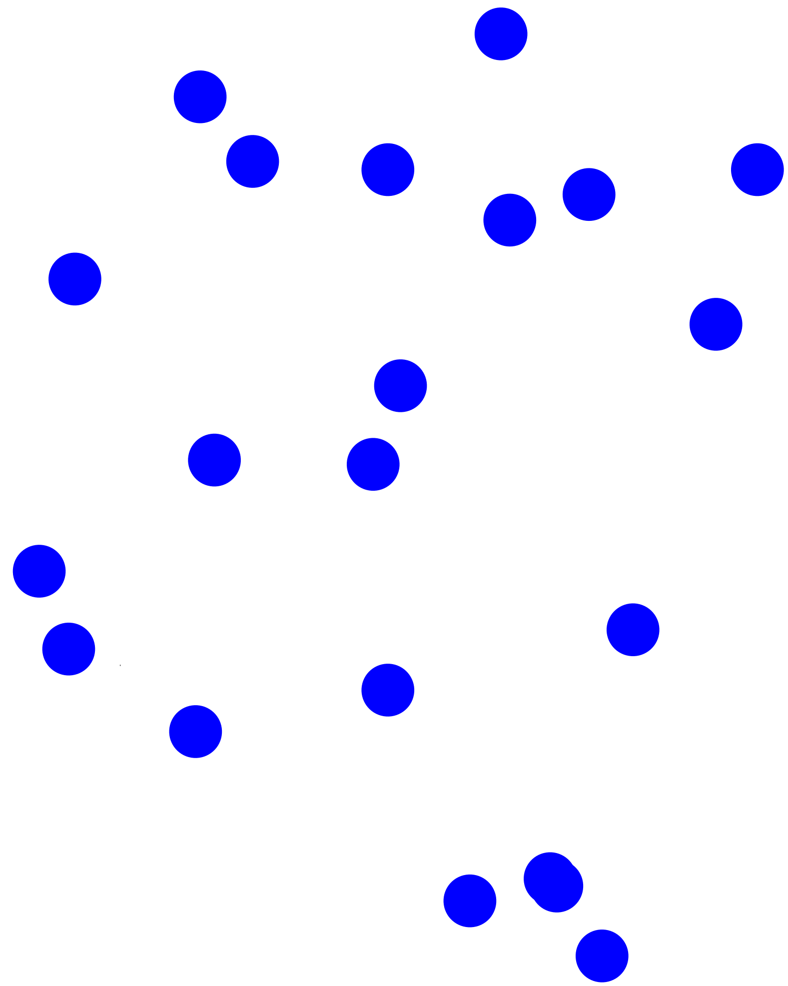
things decay with probability $\lambda$ (can serve as a stochastic driving force, for $p = 0$)

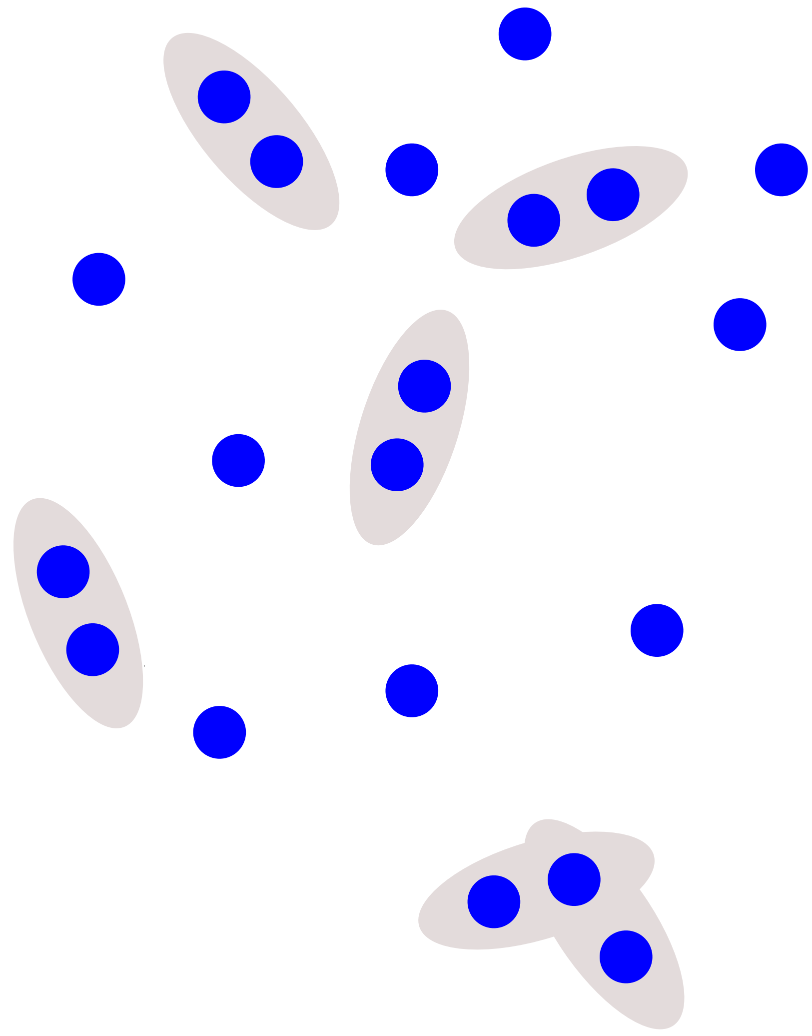
- Bounded rationality

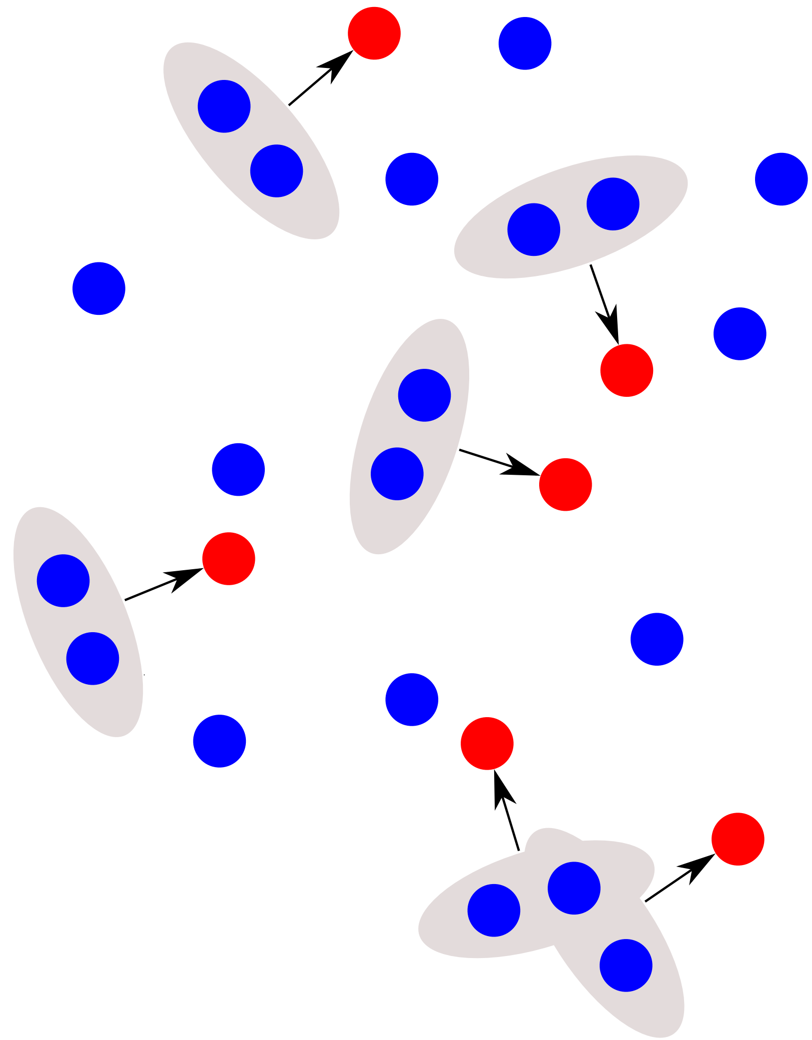
if a product *can* get produced it does not mean that it actually *will* get produced. To incorporate this possibility we say that if a good can get produced, it will actually get produced with a probability $q$

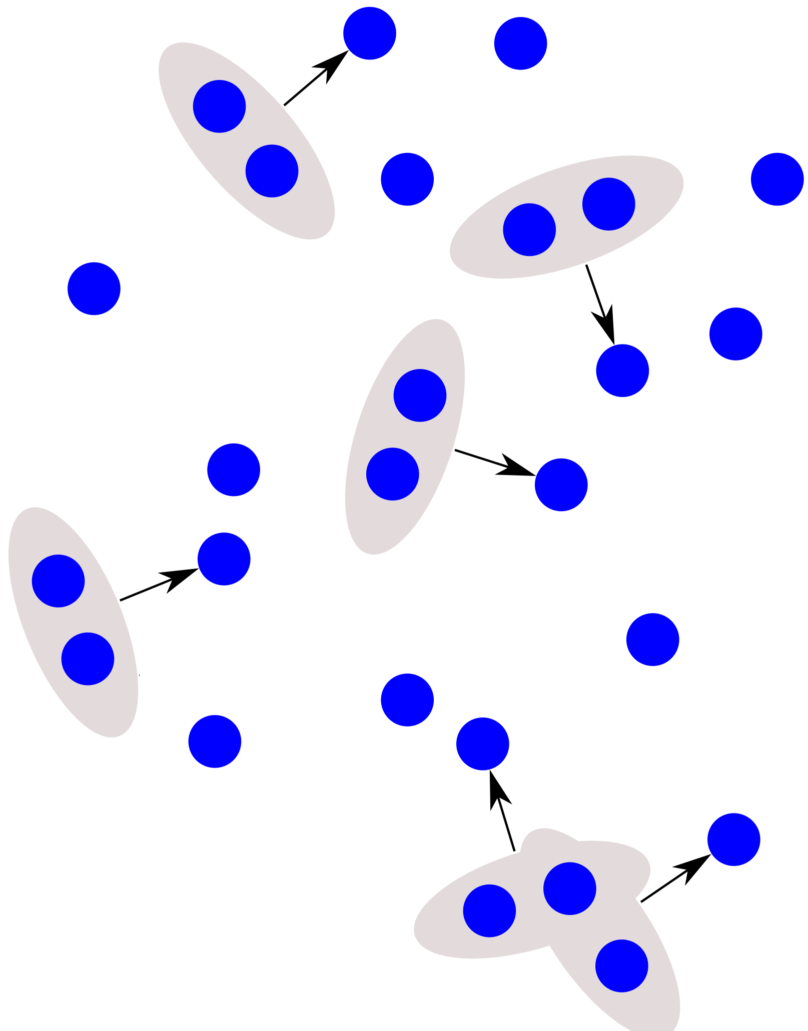
- Variations in the update: robust for parallel update

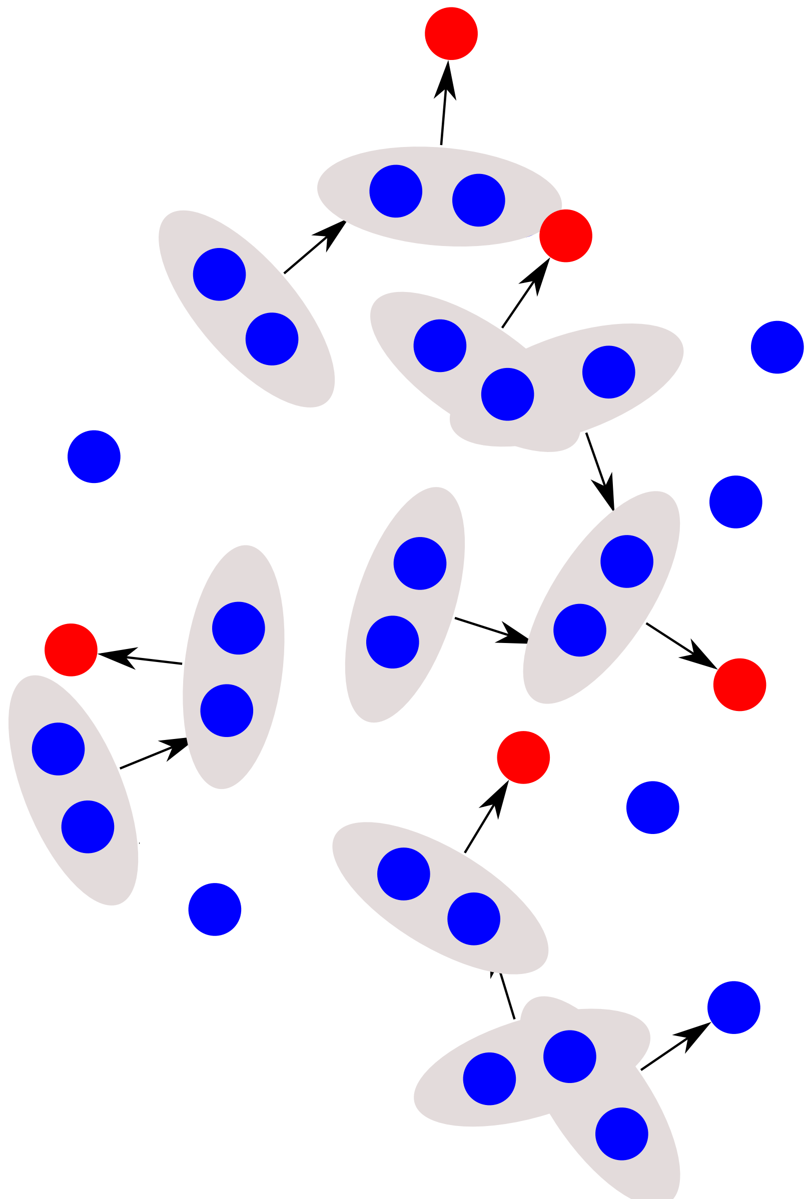










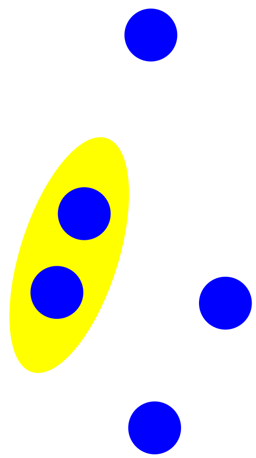


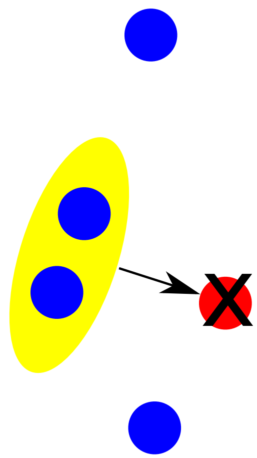
# Competition ?

- How to get competition without introducing new ingredients

- → destructive pairs

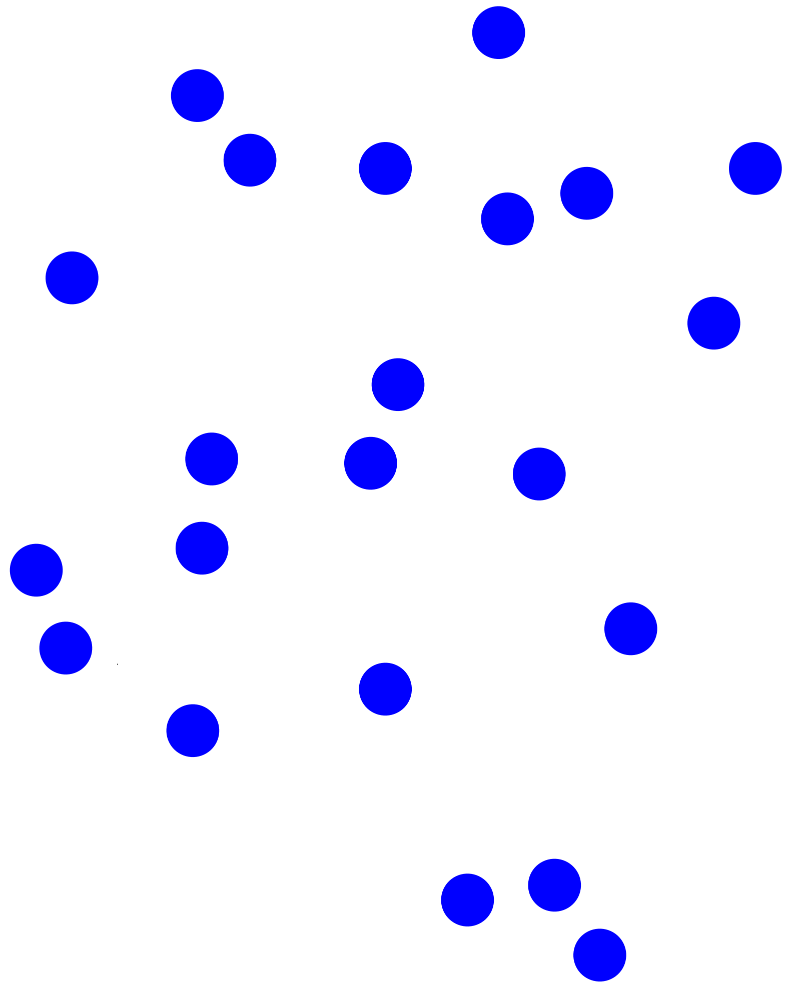
Taboo: utility, fitness, niches, selection pressure, ...

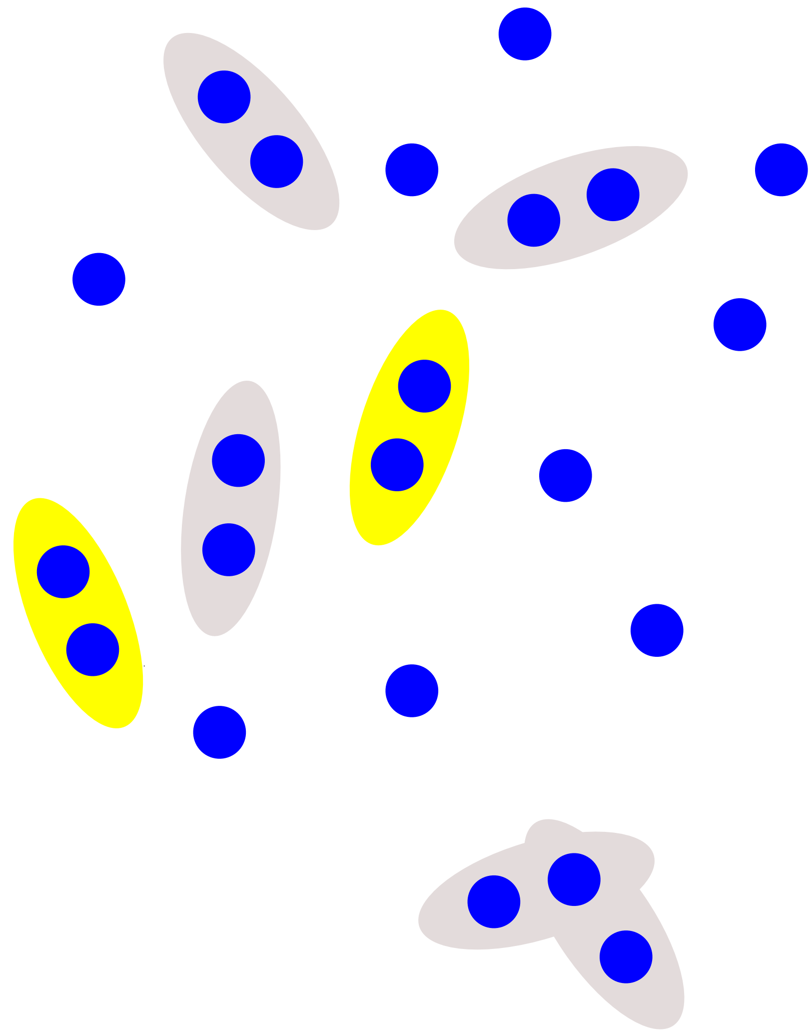


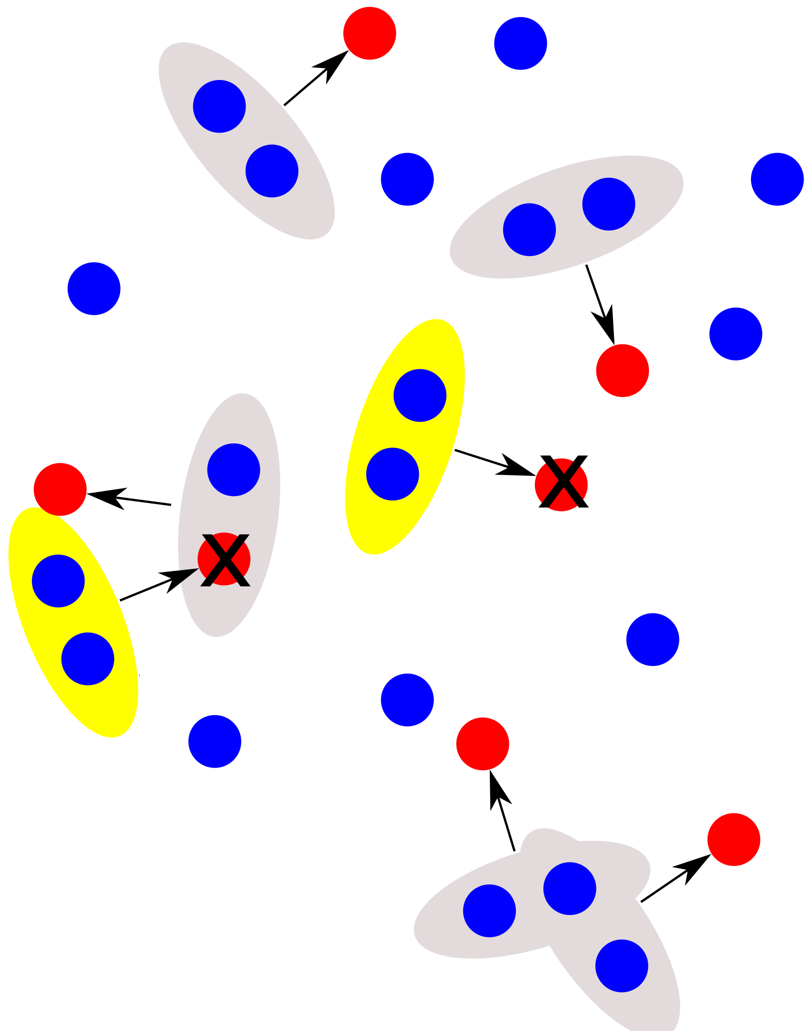

X

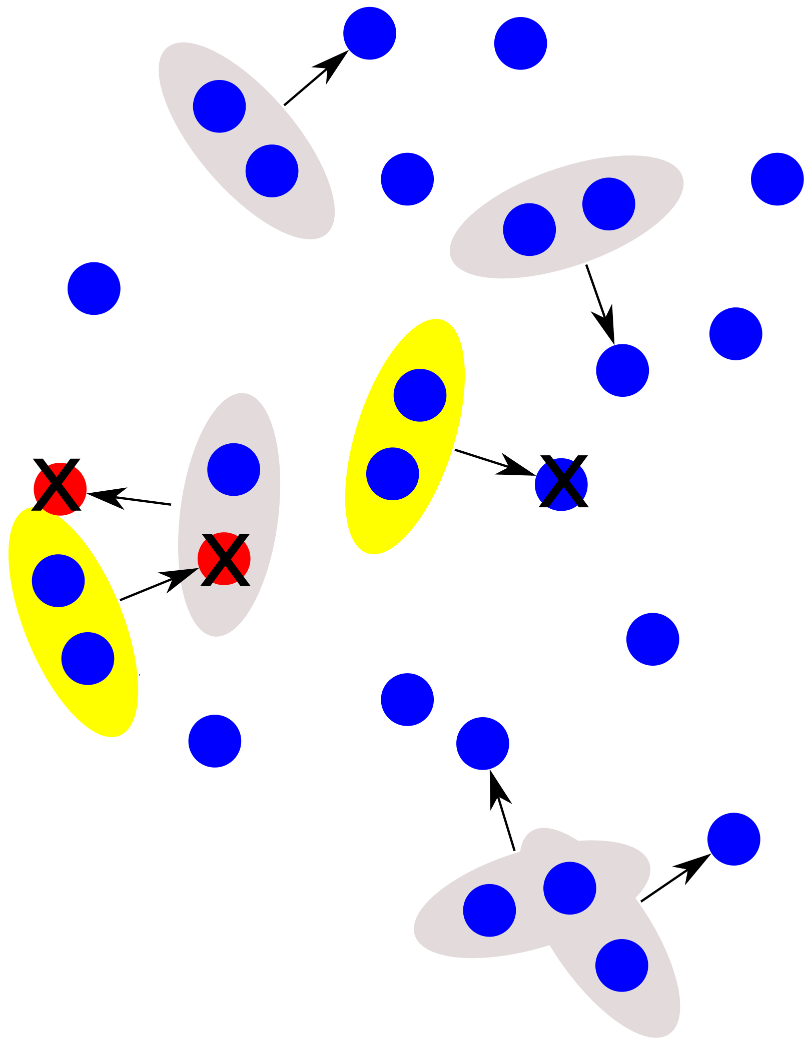
Combine $\alpha^{+}_{ijk}$ (constructive) and $\alpha^{-}_{ijk}$ (destructive) pairs

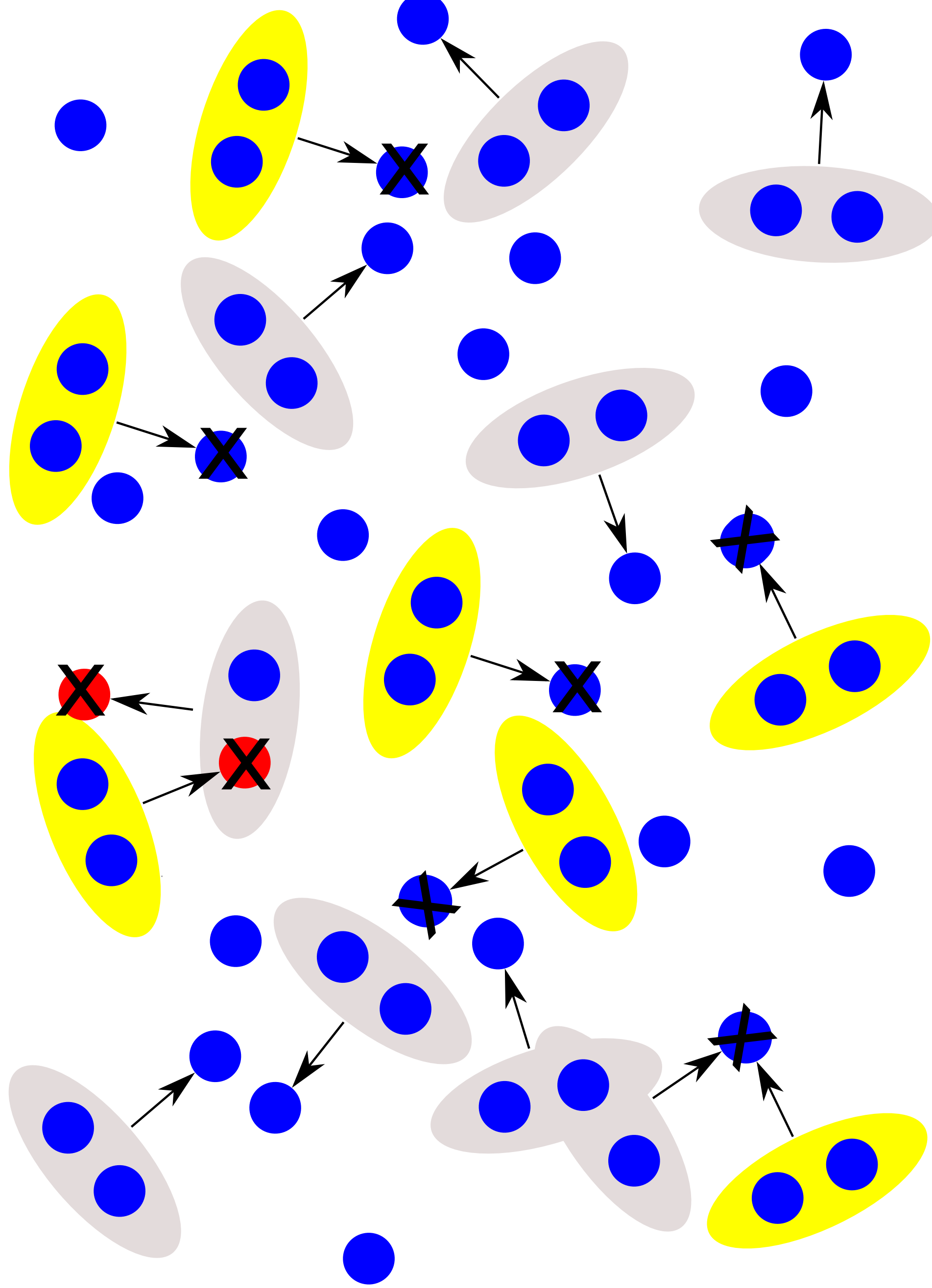
for implicit competition mechanism

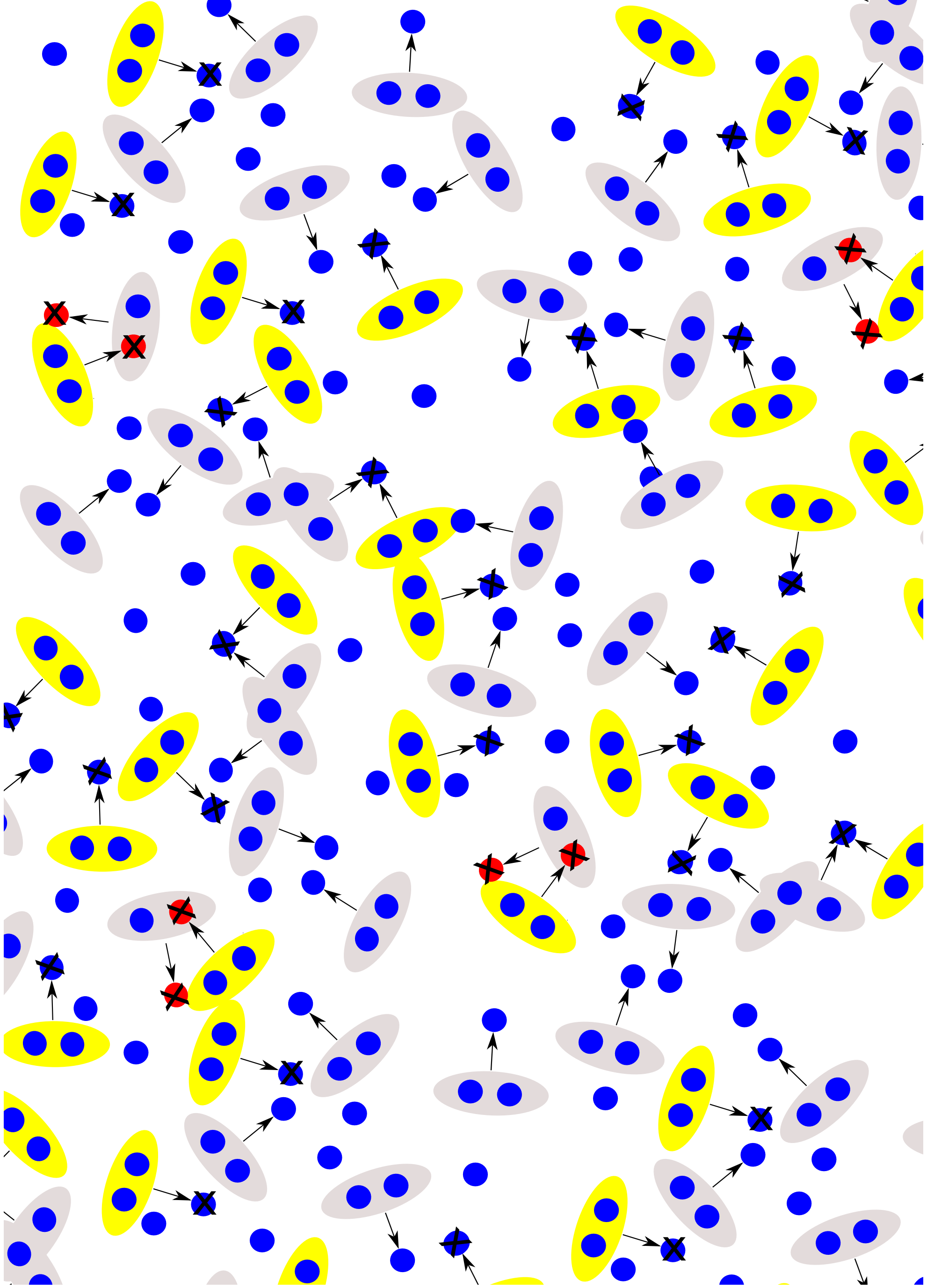


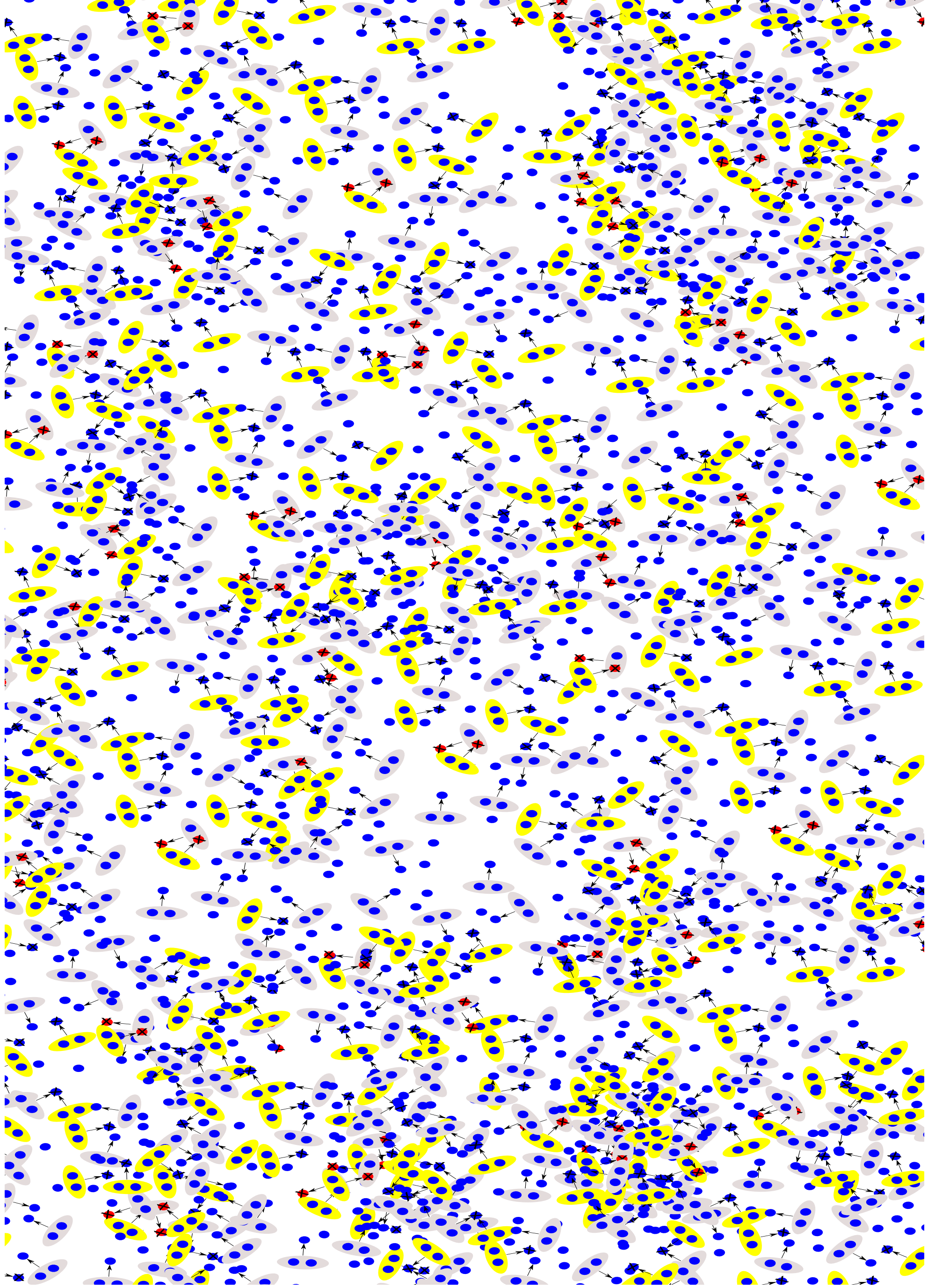


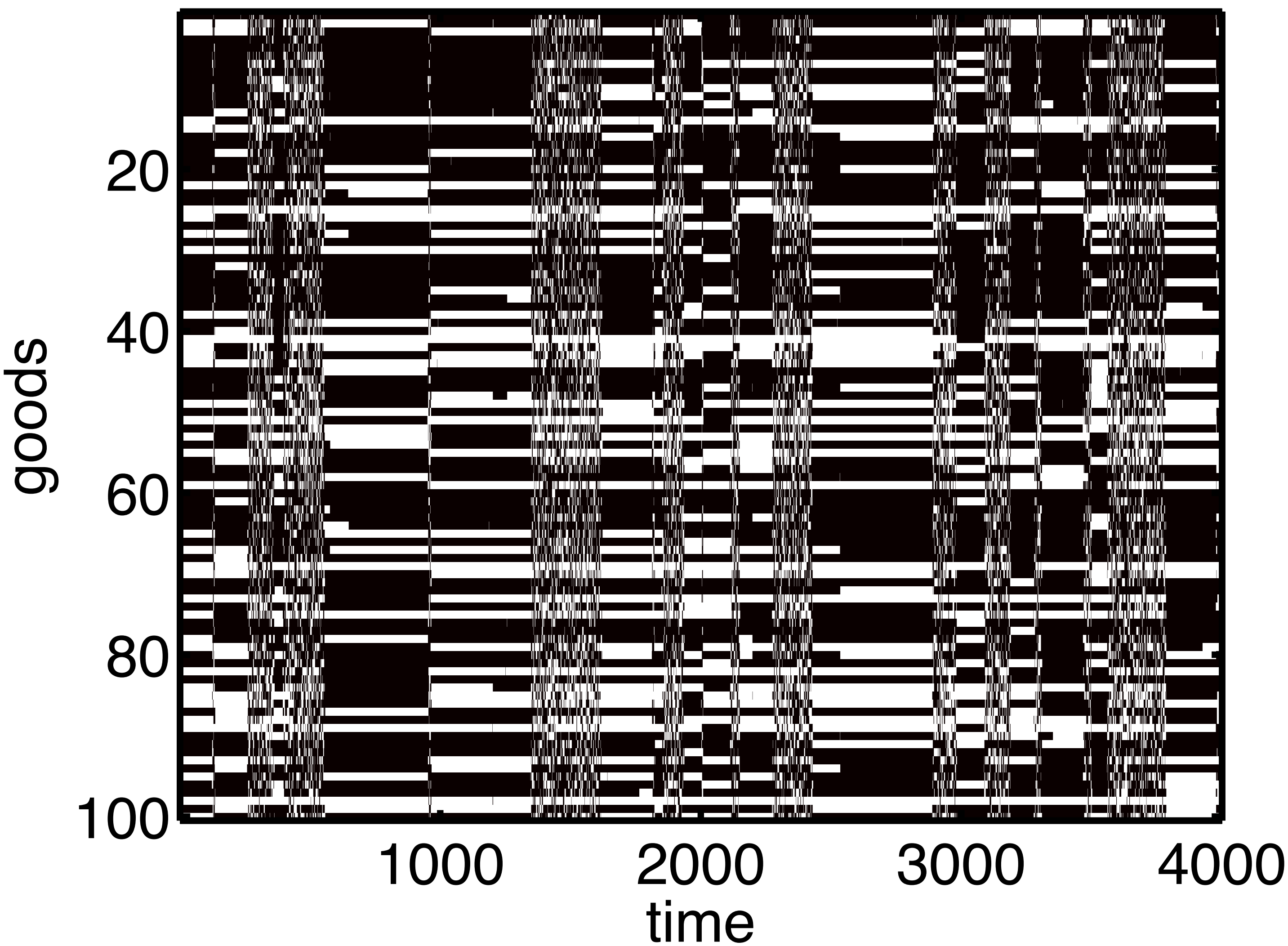



X
X
X

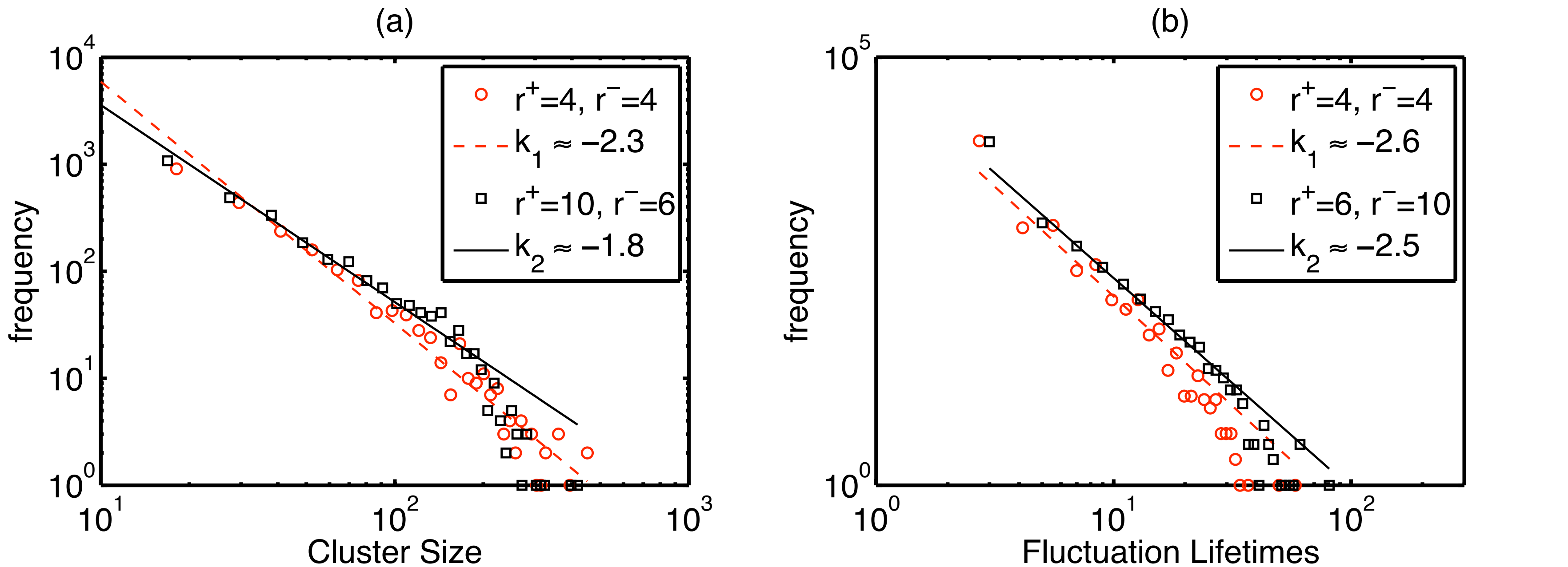






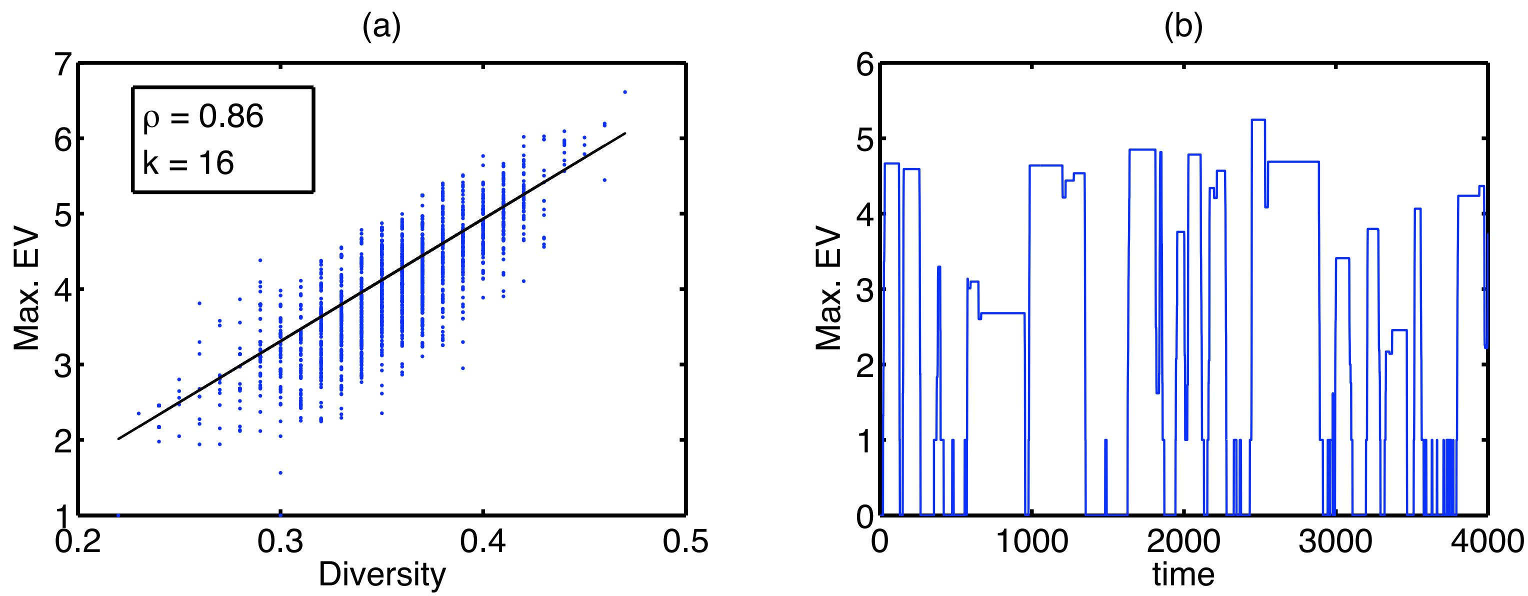

goods
20
40
60
80
100
1000
2000
3000
4000
time

# Understanding the dynamics I: it's a sandpile




Set $p = 0$ – let system relax – flip one state – measure avalanche of activity

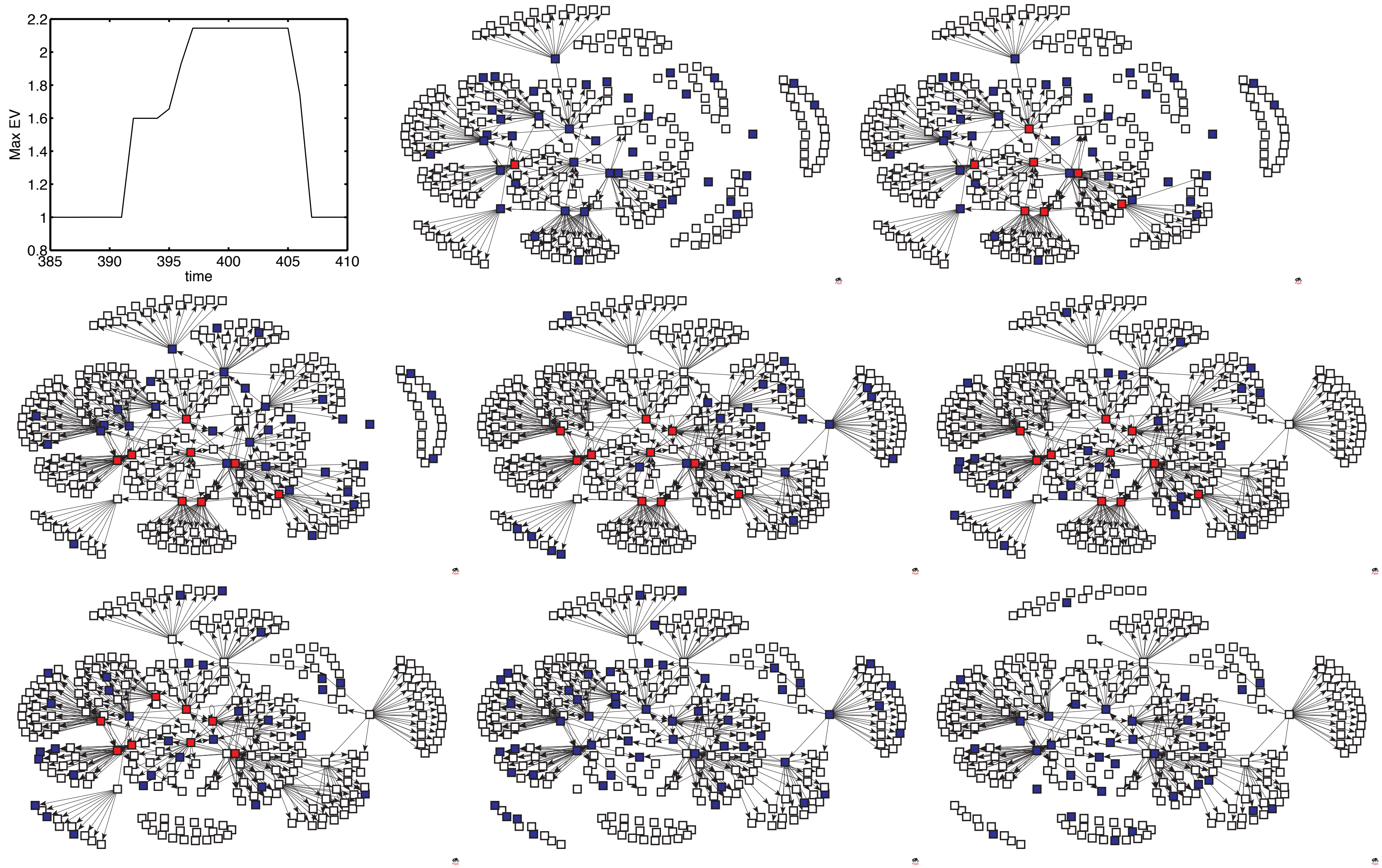
# Understanding the dynamics II: Eigenvalues & Keystone species

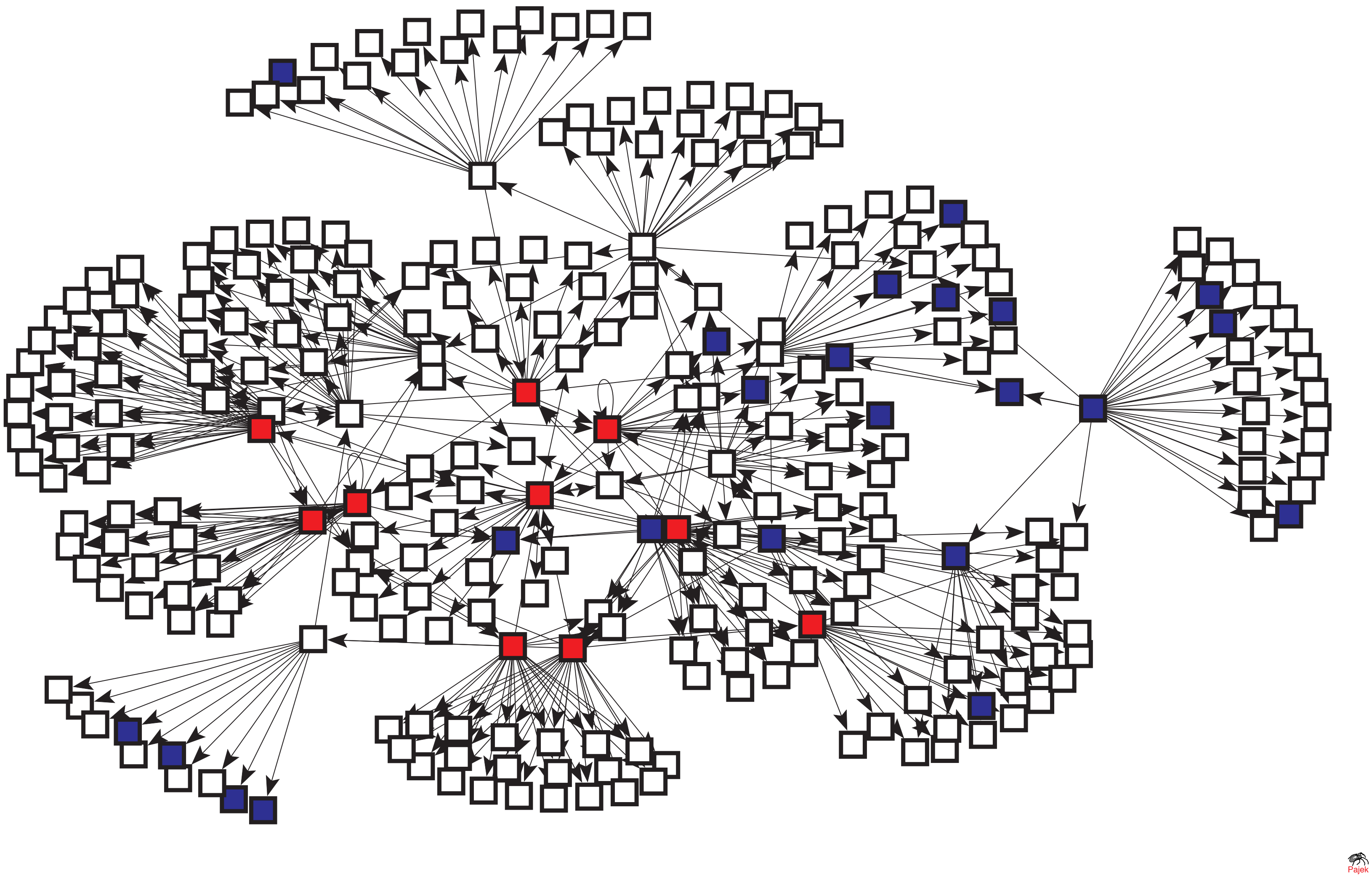


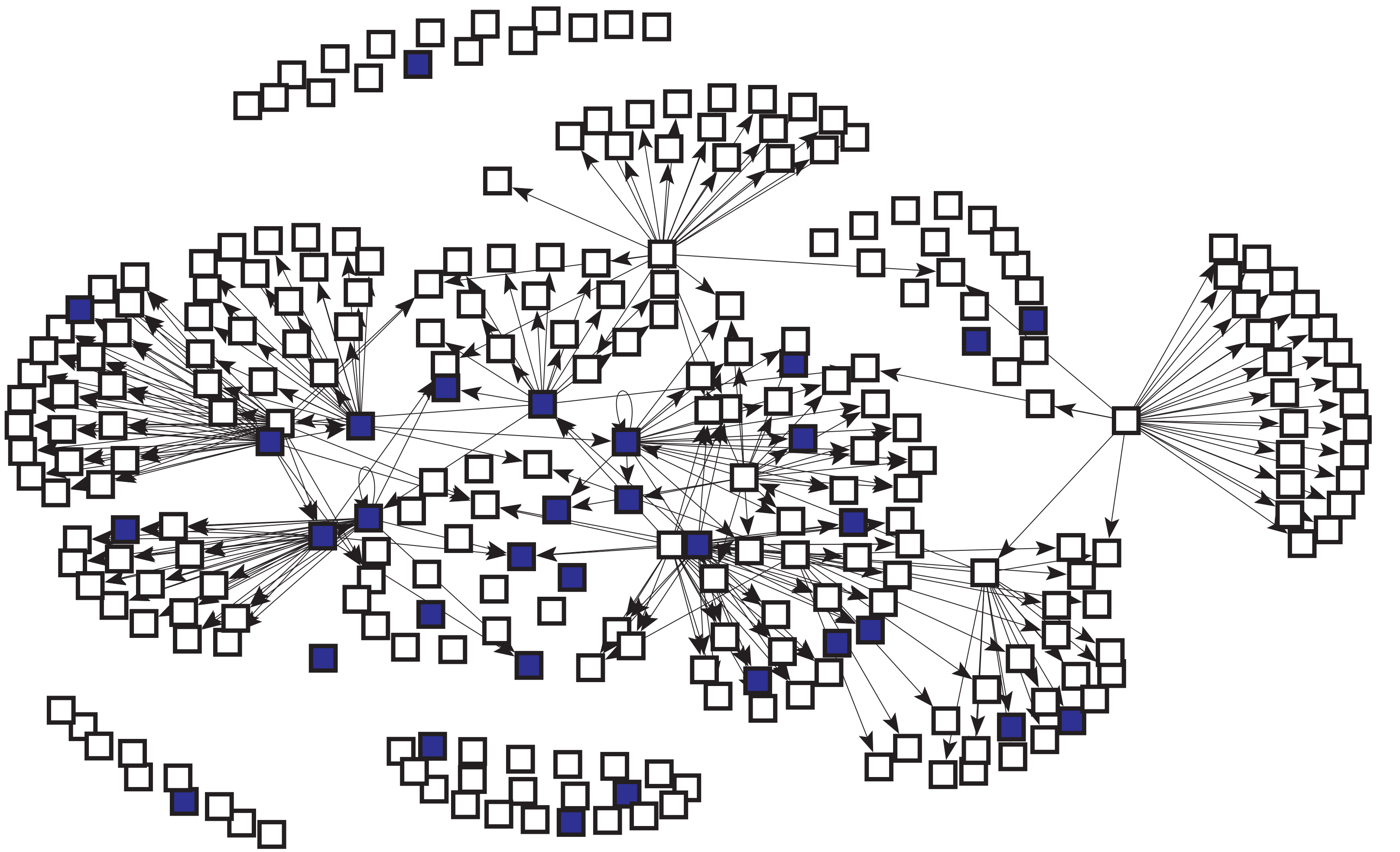

Relate topology of active production network with diversity

Similar to what was found in Jain&Krishna model of biological evolution

Kill keystone species (in autocatalytic core) → diversity collapses


Max EV
time
2.2
2
1.8
1.6
1.4
1.2
1
0.8
385
390
395
400
405
410


Pajek



# Toward a unified mathematical framework

At each time each element $i$ experiences one of three cases:

(i) creation $\sigma_i(t) = 0 \rightarrow \sigma_i(t+1) = 1$
(ii) annihilation $\sigma_i(t) = 1 \rightarrow \sigma_i(t+1) = 0$
(iii) nothing $\sigma_i(t) = \sigma_i(t+1)$

Imagine a function, $f_i(\vec{\sigma}(t)) : \{0,1\}^N \rightarrow \mathbb{R}$ to indicate

$$\begin{array}{lll} \text{(i)} & f_i(\vec{\sigma}(t)) > 0 & \Rightarrow \sigma_i(t+1) = 1 \\ \text{(ii)} & f_i(\vec{\sigma}(t)) < 0 & \Rightarrow \sigma_i(t+1) = 0 \\ \text{(iii)} & f_i(\vec{\sigma}(t)) = 0 & \Rightarrow \sigma_i(t+1) = \sigma_i(t) \end{array}$$

Use ramp function $R(x) \equiv \max(0, x)$

$$\begin{aligned} \sigma_i(t+1) &= \sigma_i(t) + \Delta\sigma_i(t) \quad , \\ \Delta\sigma_i(t) &= \mathrm{sgn}\left[(1 - \sigma_i(t))R(f_i(\vec{\sigma}(t))) - \sigma_i(t)R(-f_i(\vec{\sigma}(t)))\right] \end{aligned}$$

## A variational principle for diversity dynamics

Consider displacement: $\sigma_i'(t) = \sigma_i(t) + \delta\sigma_i(t)$

- Quadratic distance: $K_i(\sigma_i'(t), \sigma_i(t)) \equiv \frac{\mu}{2}\left[\sigma_i'(t) - \sigma_i(t)\right]^2$
- Potential: $V_i(\sigma_i'(t), \vec{\sigma}(t)) \equiv |(1 - \sigma_i'(t))R(f_i(\vec{\sigma}(t))) - \sigma_i'(t)R(-f_i(\vec{\sigma}(t)))|$
- Balance function: $B_i \equiv K_i + V_i$ contains all full dynamical information
- Variational principle (Theorem):

Given $\vec{\sigma}(t)$ the solution $\sigma_i(t+1)$ is identical to the value of $\sigma_i'(t)$ for which $B_i$ assumes its minimum, i.e.

$$\sigma_i(t+1) = \underset{\sigma_i'(t)}{\operatorname{argmin}} \left[ B_i \Big( \sigma_i'(t), \vec{\sigma}(t) \Big) \right]$$

with $\underset{x}{\operatorname{argmin}}\,[f(x)]$ the $x$ for which $f(x)$ takes its minimum

# Stochastic formulation

Probability to find a configuration $\sigma_i(t)$

$$p(\sigma_i(t)) \propto \mathrm{e}^{-\beta B_i(\vec{\sigma}(t))} \qquad \beta \equiv 1/T$$

Demand detailed balance

$$\frac{p(\sigma_i(t) \to \hat{\sigma}_i(t))}{p(\hat{\sigma}_i(t) \to \sigma_i(t))} = \frac{p(\hat{\sigma}_i(t))}{p(\sigma_i(t))} = \mathrm{e}^{-\beta(\hat{B}_i - B_i)}$$

with $\hat{B}_i \equiv B_i(\hat{\sigma}_i(t), \sigma(t)_{j \neq i})$

Now use e.g. Metropolis transition probabilities

# Mean-field approximation

Expectation value $q_i(t) = \langle \sigma_i(t) \rangle$

Assume probability factorizes: $p(\vec{\sigma}(t)) = \prod_i p_i(\sigma_i(t))$

Boltzmann-Gibbs entropy $s$ for element $i$ is

$$s(\sigma_i(t)) = -\langle \ln p_i(\sigma_i(t)) \rangle \equiv s(q_i(t))$$

Free energy

$$\phi(q_i(t)) = \langle B_i \rangle_{p(\vec{\sigma}(t))} - \frac{s(q_i(t))}{\beta}$$

Asymptotic state of species $i$, $q_i(t \to \infty) \equiv q_i$ at minimum in $\phi$

Necessary condition $\partial\phi(q_i)/\partial q_i = 0$, is $\frac{\partial \langle B_i \rangle}{\partial q_i} + \frac{1}{\beta} \ln\left(\frac{q_i}{1-q_i}\right) = 0$

$$q_i = \frac{1}{2}\left\{\tanh\left[-\frac{\beta}{2}\frac{\partial \langle B_i \rangle}{\partial q_i}\right] + 1\right\}$$

The self-consistent solution yields the asymptotic configuration

# Mean-field approximation II

Calculate $\langle B_i \rangle$ for random interaction topologies

rule densities: $r^{\pm}$

constructive / desructive set size: $n^{\pm}$

$$\langle K_i \rangle_{p(\sigma)} = \frac{1}{2}\Big( (1-q_i)p^{+} + q_i p^{-} \Big)^2$$

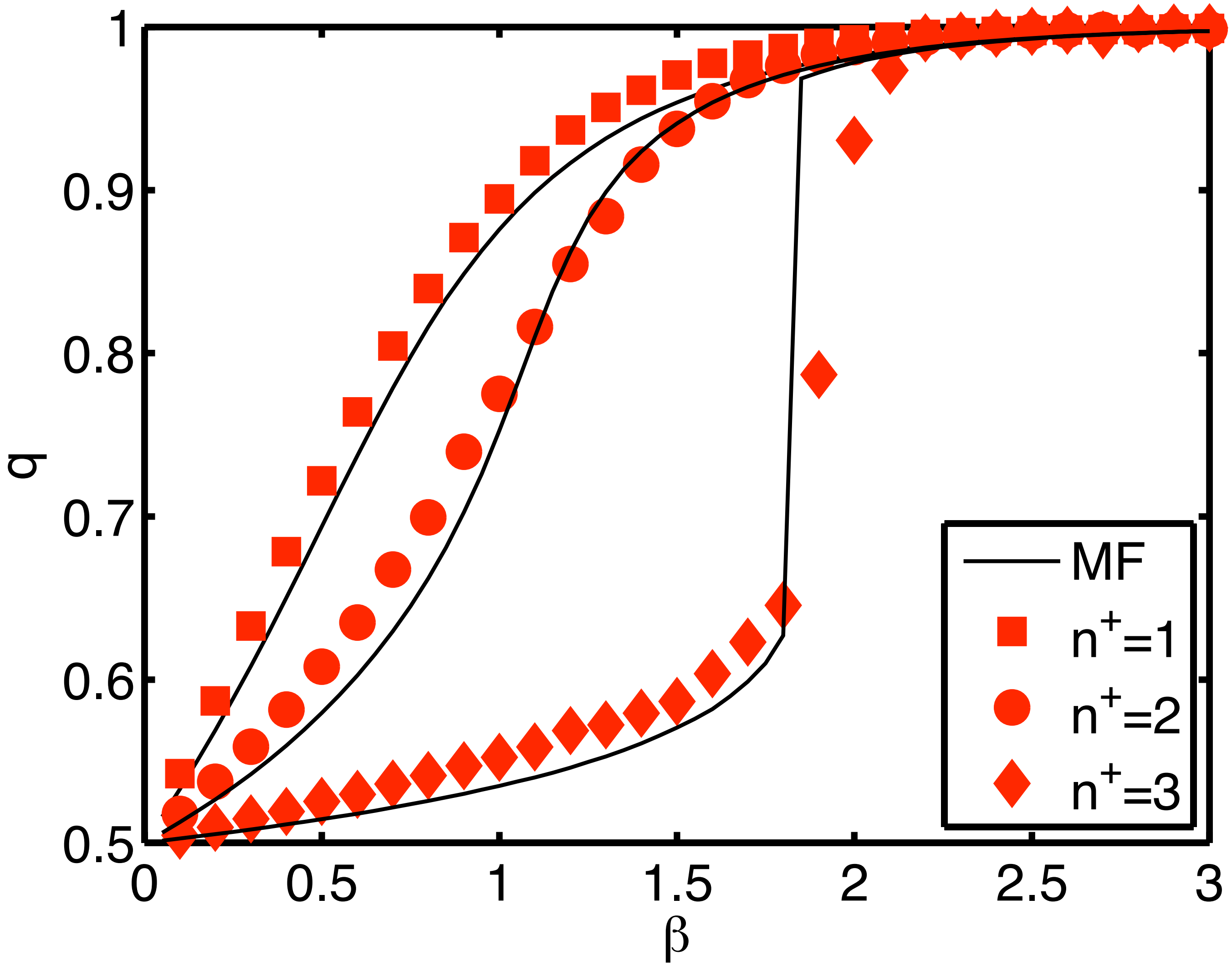
and

$$\langle V_i \rangle_{p(\sigma)} = \left| (1-q_i)\, R(f_i) - q_i R(-f_i) \right|$$

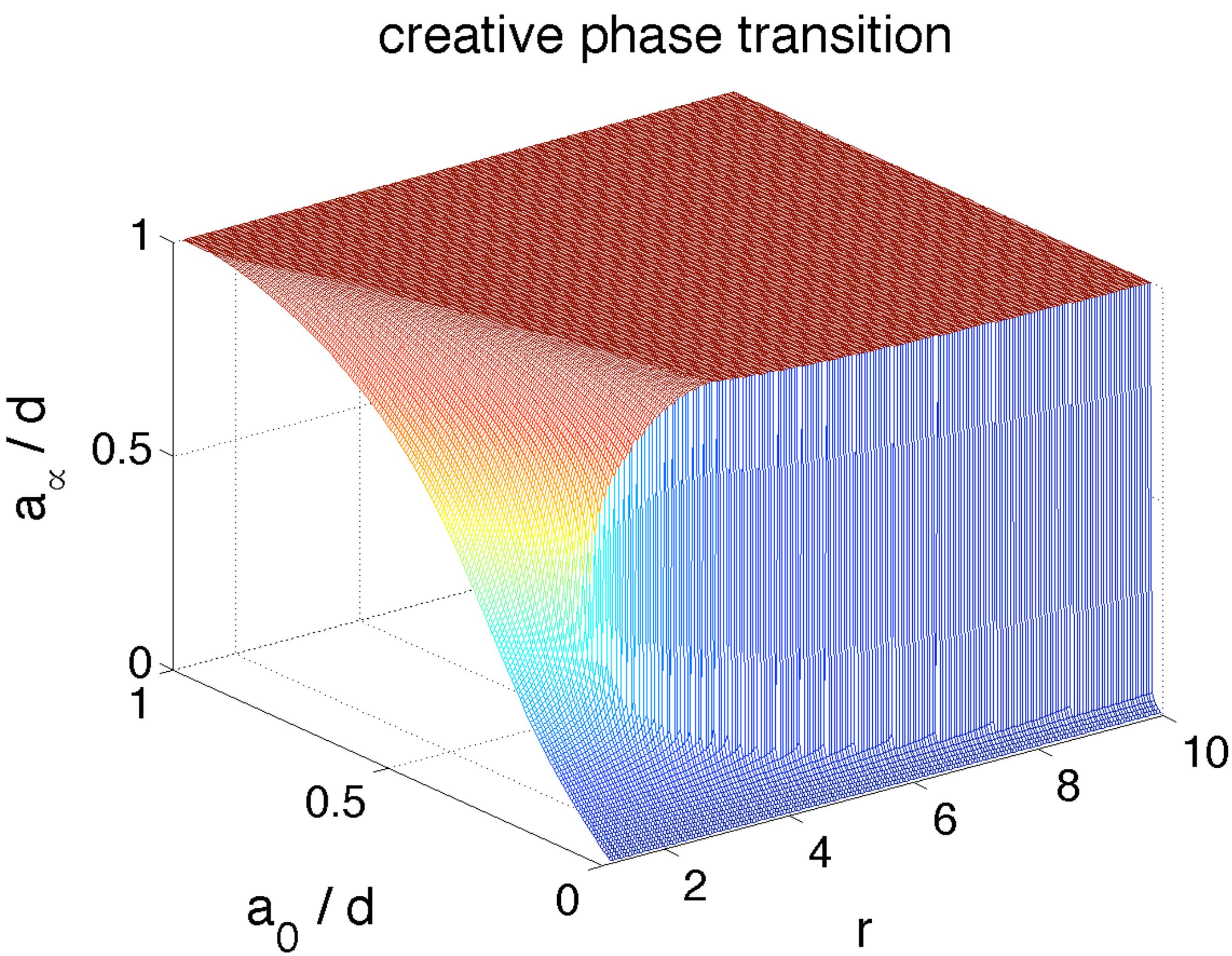
Take derivative w.r.t. $q_i$. MF result is the self-consistent solution for $q$

$$q = \frac{1}{2}\left\{ \tanh\left[ \frac{\beta}{2}\Big( r^{+} q^{n^{+}} - r^{-} q^{n^{-}} + \left[(1-q)p^{+} + qp^{-}\right](p^{+} - p^{-}) \Big) \right] + 1 \right\}$$

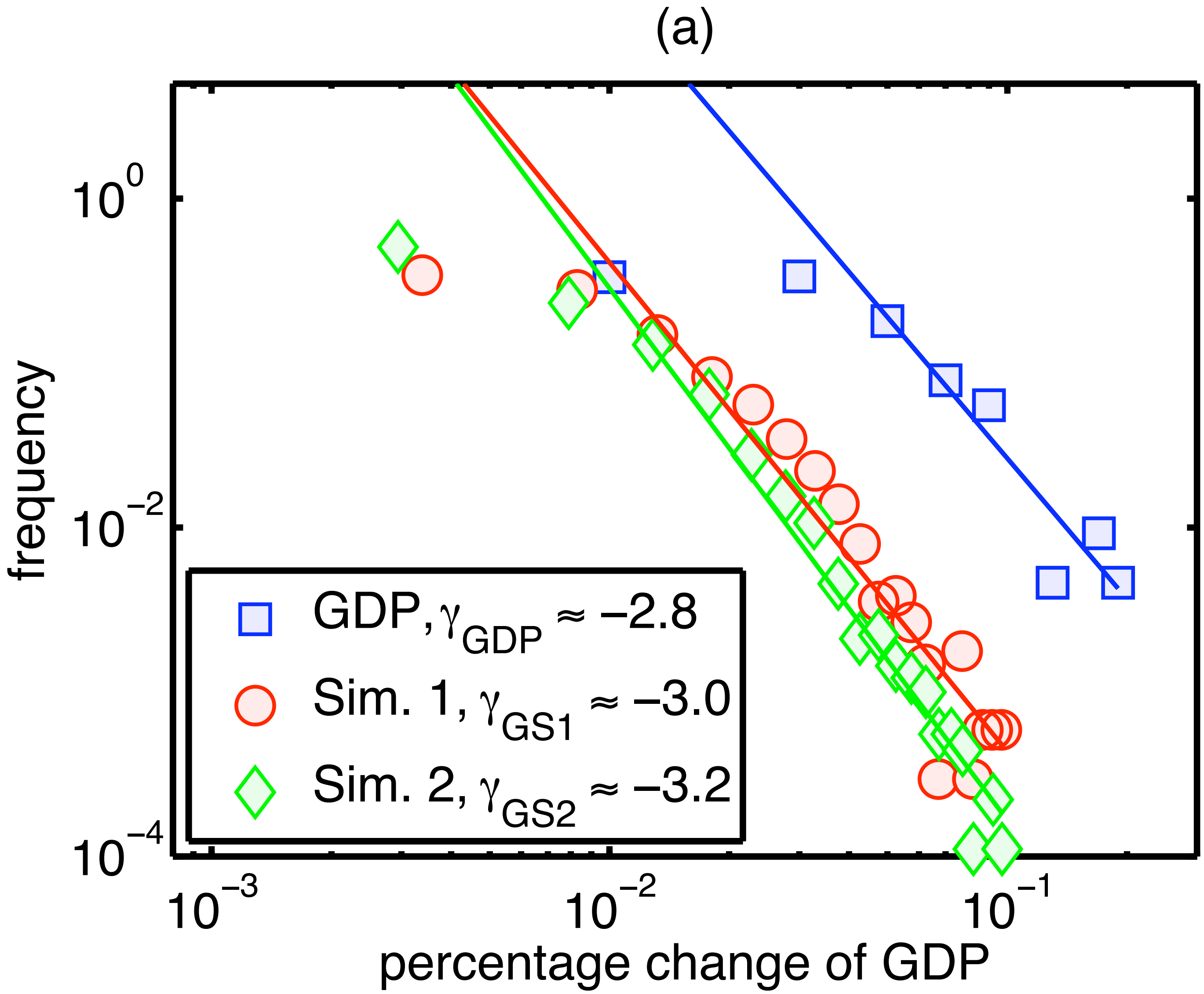
# Mean-field approximation III



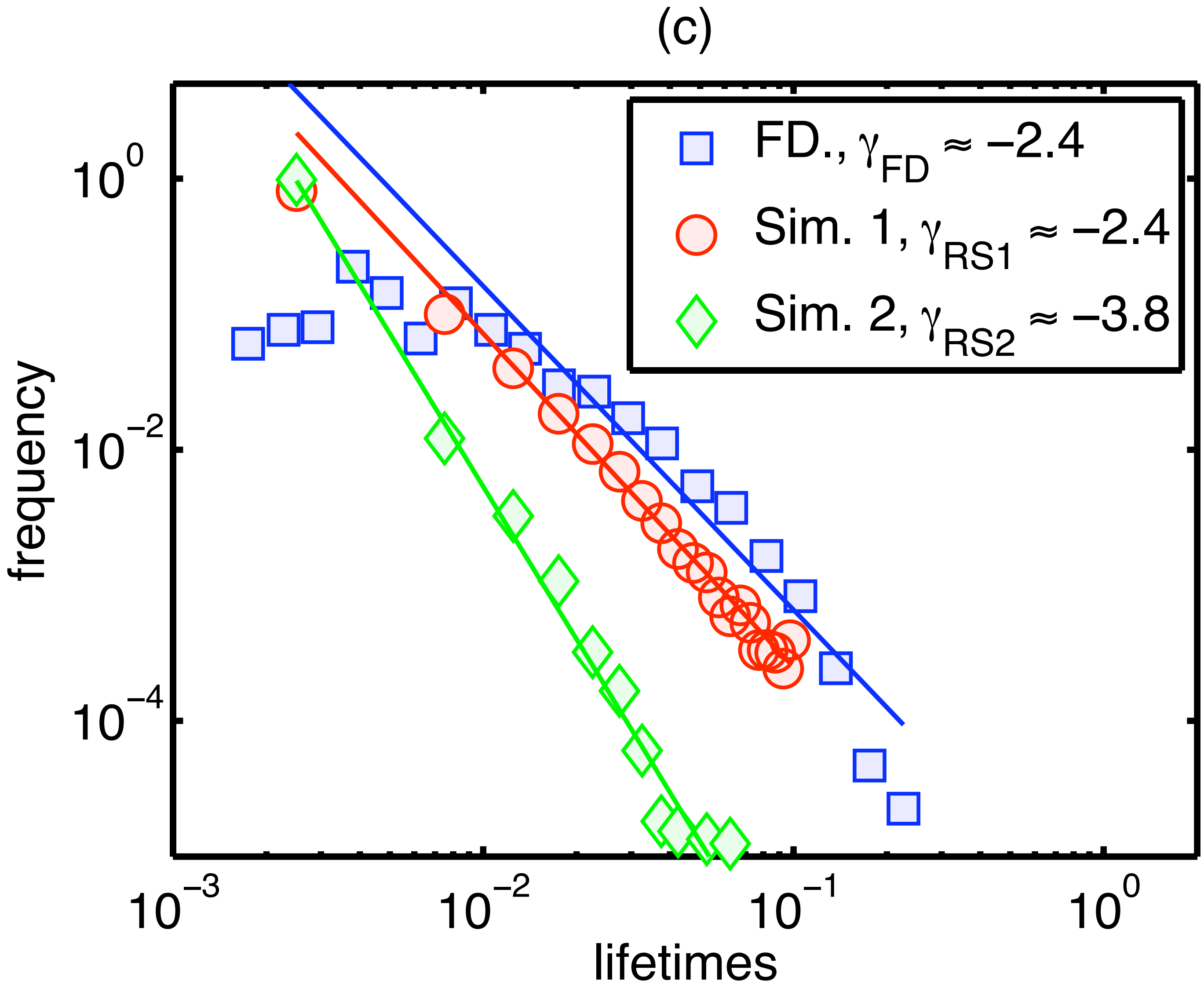
# Phase diagrams of evolution systems



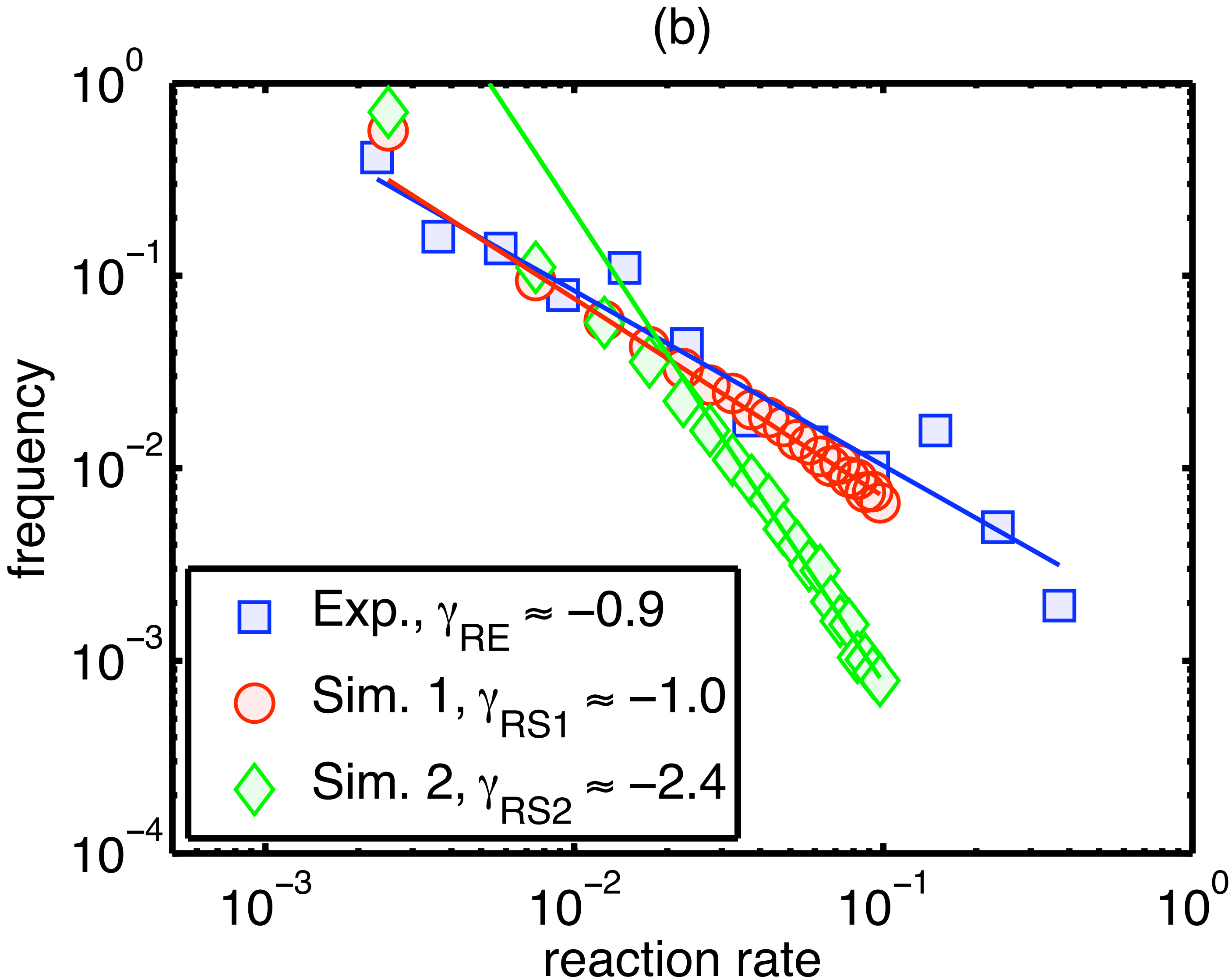
# Exaples I: GDP



# Exaples II: Life time of species



# Exaples III: Chemical reactions e. coli



# Conclusions

- identified problem with traditional evolutionary thinking
- proposed model: overlay production map with states
- formulate as a spin system
- compute its phase diagrams and compare to simulations
- show that actual timeseries are within predictions of the model