Simplifying complexity

—

Large ecosystems and random interactions

16/05/2022

Points of this lecture:

- We always think that simple system = small system (e.g. only study one species in isolation, or a pair)

Points of this lecture:

- We always think that simple system = small system (e.g. only study one species in isolation, or a pair)
- In fact, it is possible to be large (many-species...) and still simple

Points of this lecture:

- We always think that simple system = small system (e.g. only study one species in isolation, or a pair)
- In fact, it is possible to be large (many-species...) and still simple
- Complexity with simple consequences is (or can be modelled by) "randomness"

Points of this lecture:

- We always think that simple system = small system (e.g. only study one species in isolation, or a pair)
- In fact, it is possible to be large (many-species...) and still simple
- Complexity with simple consequences is (or can be modelled by) "randomness"
- Small and large simplicity are both wrong, but both are valid starting points, and they can be combined to model reality

I. Introduction: Complexity and simplicity

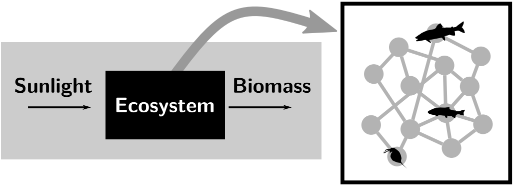
Where to put complexity

Basic question of modelling: which details are important to include?

Where to put complexity

Basic question of modelling: which details are important to include?

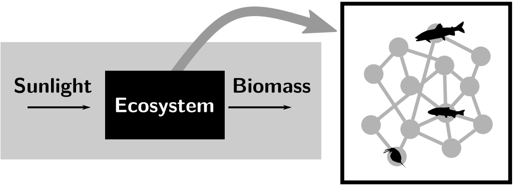
- whenever we write a simple model in biology, we are hiding complexity



Where to put complexity

Basic question of modelling: which details are important to include?

- whenever we write a simple model in biology, we are hiding complexity



- is there a *principled* way of understanding when this is a valid choice?

Imagine if we only tracked colors? (grouping all lifeforms of same color)

$$\frac{d\,\mathrm{Red}}{dt} = a\,\mathrm{Red} + b\,\mathrm{Blue} \tag{1}$$

Seems absurd (except maybe green for photosynthesis) but why exactly?

Lotka-Volterra model

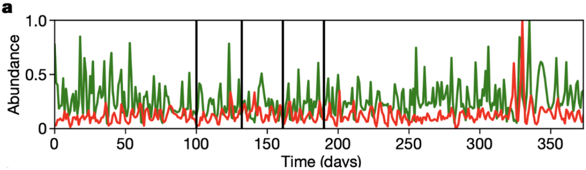
For instance, take our favorite dynamical model:

$$\frac{dN_i}{dt} = r_i N_i \left(1 - \sum_j a_{ij} N_j \right) \tag{2}$$

- why species abundances and interactions, rather than
  - individual movement, size, social and sexual behavior
  - genes, proteins
  - nutrient fluxes, biochemical processes (redox, denitrification...)
  - ...

$$\frac{dN_i}{dt} = r_i N_i (1 - \sum_j a_{ij} N_j) \qquad (3)$$

- choice guided by what we can measure e.g. abundance time series (more available than social behavior time series)



but not only: colors are probably easier to observe than species abundances

Lotka-Volterra model

$$\frac{dN_i}{dt} = r_i N_i (1 - \sum_j a_{ij} N_j) \tag{4}$$

$\Rightarrow$ assumptions about which processes are important & independent

- species growth & interactions are important forces
  (N_i is not fixed by some other force like human experimenter)

$$\frac{dN_i}{dt} = r_i N_i (1 - \sum_j a_{ij} N_j) \quad (4)$$

$\Rightarrow$ assumptions about which processes are important & independent

- species growth & interactions are important forces
  (N_i is not fixed by some other force like human experimenter)
- other processes (e.g. evolution, individual movements) can be ignored because on different scales, e.g. much slower or much faster

Lotka-Volterra model

$$\frac{dN_i}{dt} = r_i N_i (1 - \sum_j a_{ij} N_j) \quad (4)$$

$\Rightarrow$ assumptions about which processes are important & independent

- species growth & interactions are important forces

  (N_i is not fixed by some other force like human experimenter)

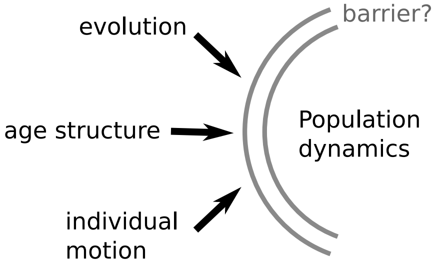
- other processes (e.g. evolution, individual movements) can be ignored because on different scales, e.g. much slower or much faster

- other processes on same scale (e.g. population genetics, age structure) can be ignored because they *do not interfere* somehow

  same abundance dynamics could exist in systems *without* age, genes... e.g. computer viruses

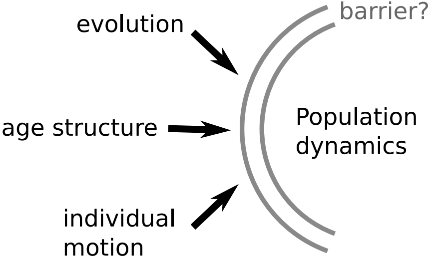
Toward large systems

- **Idea that will keep coming back:** not all details matter for everything; sometimes, there are "barriers" that details don't cross



Toward large systems

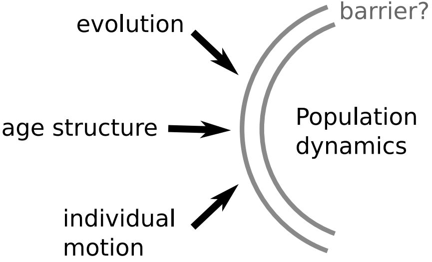
- **Idea that will keep coming back:** not all details matter for everything; sometimes, there are "barriers" that details don't cross



- if this wasn't the case, science would be impossible

Toward large systems

- **Idea that will keep coming back:** not all details matter for everything; sometimes, there are "barriers" that details don't cross



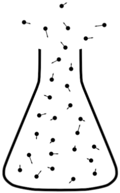
- if this wasn't the case, science would be impossible

- One such source of simplicity: "largeness" (high-dimensionality)

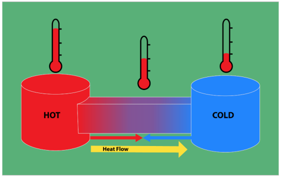
Idea originating from physics

When a system has *many* variables, a much simpler description is often possible

10^{23} variables:
position of every molecule

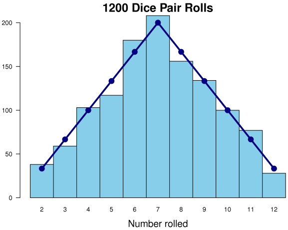


2 variables: temperature & pressure



uncountable factors, chaotic motion

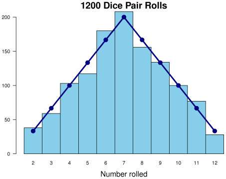
1 probability distribution



Meaning of randomness

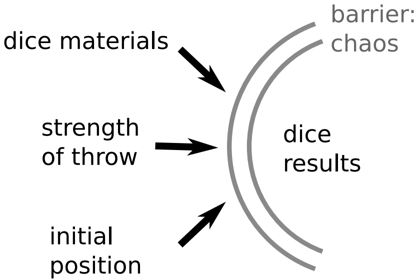
uncountable factors, chaotic motion

1 probability distribution



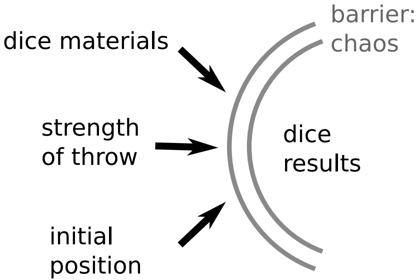
dice are simple *because* they are extremely sensitive to many details, making their movement chaotic

Meaning of randomness



- "barrier" against details = chaos, motion unpredictable even if you know almost all details

Meaning of randomness



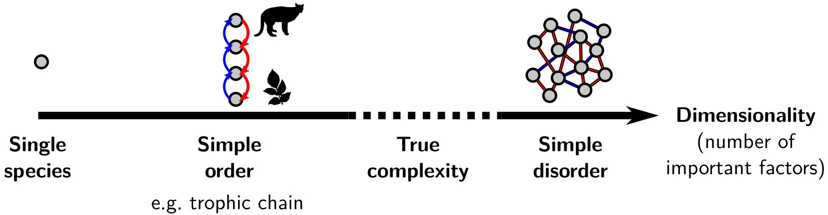
- "barrier" against details = chaos, motion unpredictable even if you know almost all details
- result = randomness, unpredictability becomes simplicity

"Random" means "too many factors", so complex mechanistically that it becomes simple statistically

All that to ask: is there simplicity from apparent complexity in ecology?

Small and large systems

All that to ask: is there simplicity from apparent complexity in ecology?



Modelling an ecological community can start

- from "small simplicity" (e.g. a 3-species trophic chain)
- or from "large simplicity" = many-species networks...

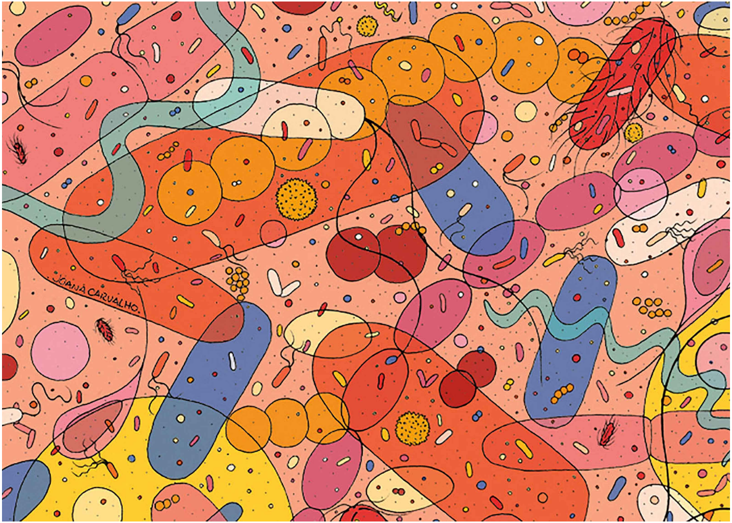
  but when & how are they simple?

II. Many-species communities

–

Part 1: What observations are we trying to explain

Forget about randomness for now, just study communities with many populations



Hereafter “species”, but could be intraspecific phenotypes, etc.

Observables in large communities

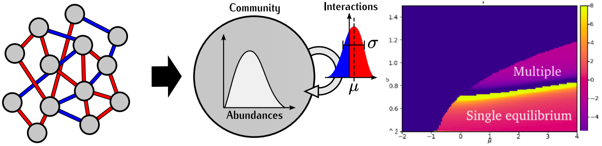
What is interesting in large communities:

- we lose focus on individual species – they are usually unpredictable, maybe impacted by dozens or hundreds of others

Observables in large communities

What is interesting in large communities:

- we lose focus on individual species – they are usually unpredictable, maybe impacted by dozens or hundreds of others
- we gain aggregate properties



  - static properties
  - dynamical properties

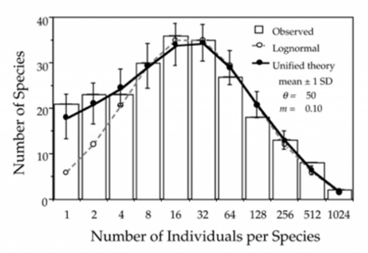
Static properties

Measurable from a single/few snapshots:

Static properties

Measurable from a single/few snapshots:

- Distributions (= histograms, frequencies)

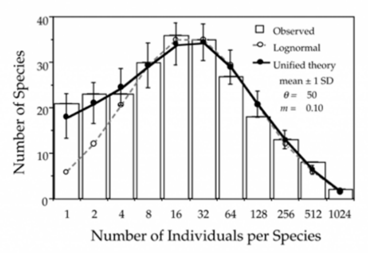


abundances, number of offsprings/production, variation in space, correlations between species' fluctuations

Static properties

Measurable from a single/few snapshots:

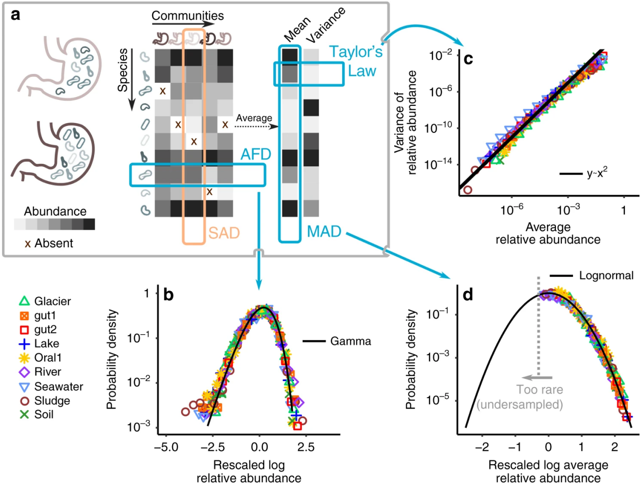
- Distributions (= histograms, frequencies)



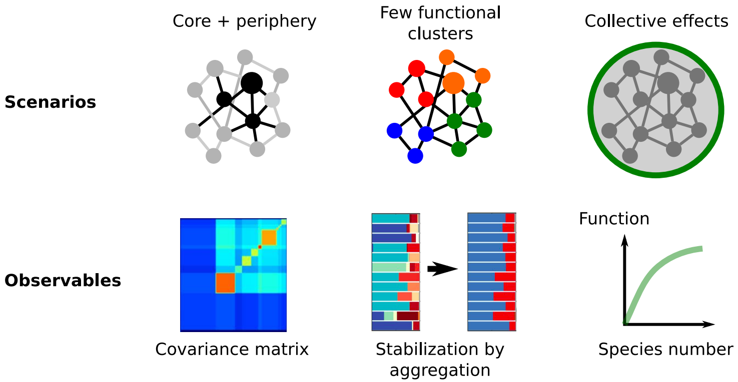
abundances, number of offsprings/production, variation in space, correlations between species' fluctuations

- Statistics on these distributions:
  - diversity (number of coexisting species)
  - total abundance $\sum_i N_i$, total production $\sum_i r_i N_i$

Many common patterns are different ways of aggregating same basic data



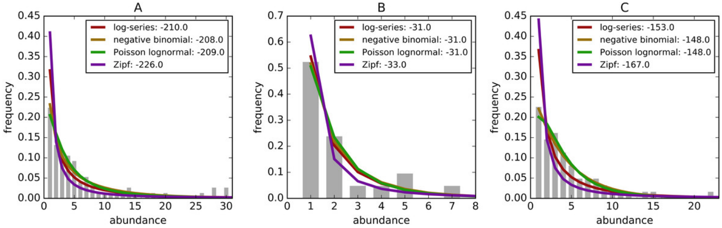
Various patterns used as "fingerprints" to test some ecological scenarios...



Fingerprints of ecological scenarios

... But I will insist that usually no “smoking gun”:

- single pattern almost never enough to know underlying ecology and processes
- e.g. many different models can fit empirical abundance histograms



Dynamical properties

Properties that can only be observed by tracking species over time, e.g.

- Is an ecosystem in a stable equilibrium or not?
- How does it respond when you disturb it?

Dynamical properties

What is the usual state of a given ecosystem?

- equilibrium
  example: constant populations of bacteria feeding in different niches

Dynamical properties

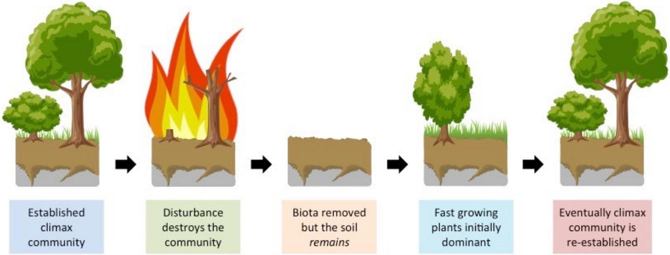
What is the usual state of a given ecosystem?

- equilibrium
  example: constant populations of bacteria feeding in different niches
- directional trajectory
  example: microbial succession during organic decomposition

Dynamical properties

What is the usual state of a given ecosystem?

- equilibrium
  example: constant populations of bacteria feeding in different niches
- directional trajectory
  example: microbial succession during organic decomposition
- stationary nonequilbrium
  example: cycles, chaos, constant flux of species invading and dying



Dynamical properties

How does an ecosystem respond when you disturb it?

- "elastic": goes back to its state or trajectory (unique attractor)
  example: gut microbiome disturbed by sickness then re-colonized

Dynamical properties

How does an ecosystem respond when you disturb it?

- "elastic": goes back to its state or trajectory (unique attractor)
  example: gut microbiome disturbed by sickness then re-colonized
- "plastic": remains modified, does not go back (multiple attractors)
  example: humans plant trees outside their original range, they remain in the new biome

Dynamical properties

How does an ecosystem respond when you disturb it?

- “elastic”: goes back to its state or trajectory (unique attractor)

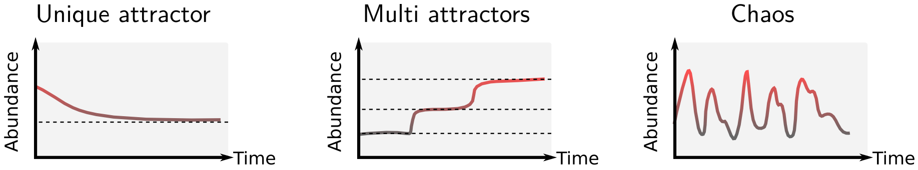
  example: gut microbiome disturbed by sickness then re-colonized

- “plastic”: remains modified, does not go back (multiple attractors)

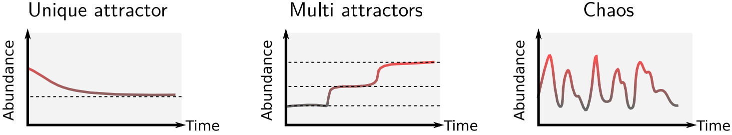
  example: humans plant trees outside their original range, they remain in the new biome

- “chaotic”: becomes more and more different

  example: a single invasive species causes a cascade of extinctions and other invasions



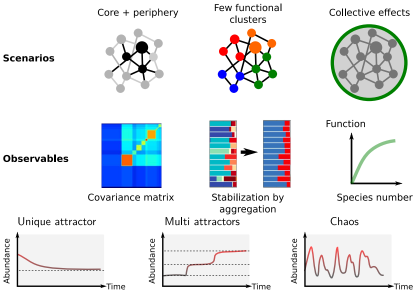
Dynamical properties



Challenge: How to predict any of these dynamics for many species?

Quick recap

Brief summary:



- Various aggregate patterns & dynamics to explain
- Many possible ecological scenarios & explanations, each with specific assumptions

$\Rightarrow$ How do we construct a simple "generic" model that explains as many patterns as possible?

II. Many-species communities

–

Part 2: How do we explain observations

Parameter explosion

If we use a model like Lotka-Volterra with S species

$$\frac{dN_i}{dt} = r_i N_i (1 - \sum_j^S a_{ij} N_j) \tag{5}$$

we need many parameters:

Parameter explosion

If we use a model like Lotka-Volterra with S species

$$\frac{dN_i}{dt} = r_i N_i (1 - \sum_j^S a_{ij} N_j) \tag{5}$$

we need many parameters:

- growth rates r_i (S numbers, 1 per species)

$$r = (?, ?, ?....) \tag{6}$$

Parameter explosion

If we use a model like Lotka-Volterra with S species

$$\frac{dN_i}{dt} = r_i N_i (1 - \sum_j^S a_{ij} N_j) \tag{5}$$

we need many parameters:

- growth rates r_i (S numbers, 1 per species)

$$r = (?, ?, ?....) \tag{6}$$

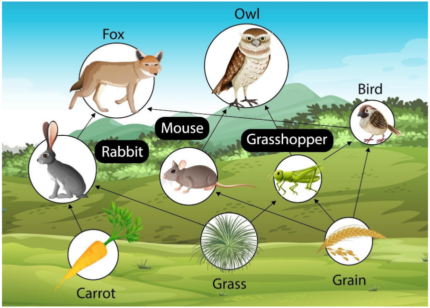
- interactions a_{ij} (S^2 numbers, S per species)

$$a = \begin{pmatrix} ? & ? & ... \\ ? & & \\ ... & & \end{pmatrix} \tag{7}$$

Species interaction networks

How do we obtain the matrix of interactions a_{ij}?

- *Good news:* qualitative structure ($a_{ij} = 0$ or $\neq 0$) can be known for some interaction types, e.g. who eats who



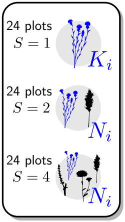
How do we obtain the matrix of interactions a_{ij}?

- *Bad news:* quantitative strength (a_{ij} values) is very rarely measured directly for every pair of species i, j (few experiments doing all that)

Grassland plants (Wageningen)

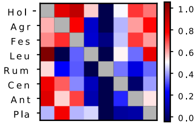
Agrostis, Anthox, Centaur, Festuca, Holcus, Leucant, Plantag, Rumex

Abundances



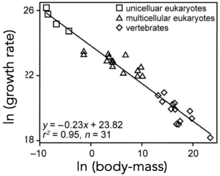
Multilinear fit

Interaction matrix α_{ij}



Species interaction networks

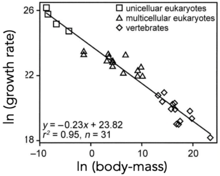
- Most of the time, theoretical assumptions are needed to put numbers into the model:
  - Metabolic scaling, r_i and a_{ij} given by body sizes of species i and j



  - Ecopath model (see with Claire this afternoon)
  - ...

Species interaction networks

- Most of the time, theoretical assumptions are needed to put numbers into the model:
  - Metabolic scaling, r_i and a_{ij} given by body sizes of species i and j



  - Ecopath model (see with Claire this afternoon)
  - ...
- What do we do if we cannot or do not want to assume anything?

Neutrality

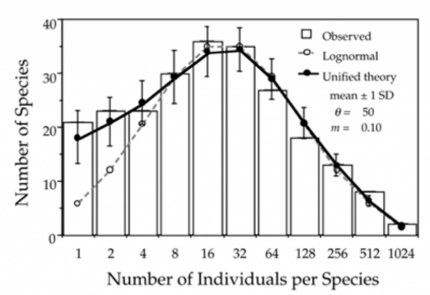
(Remember Day 1 lecture by Isabelle)

- Extreme simplification: neutrality, all species identical, $a_{ij} = 1$
- Different outcomes for different species only due to chance: random events of birth, death and migration

Neutrality

(Remember Day 1 lecture by Isabelle)

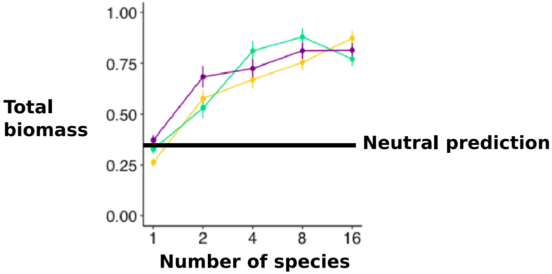
- Extreme simplification: neutrality, all species identical, $a_{ij} = 1$
- Different outcomes for different species only due to chance: random events of birth, death and migration
- Why use it? Because it can suffice to predict some patterns, e.g. abundance distributions



Neutrality

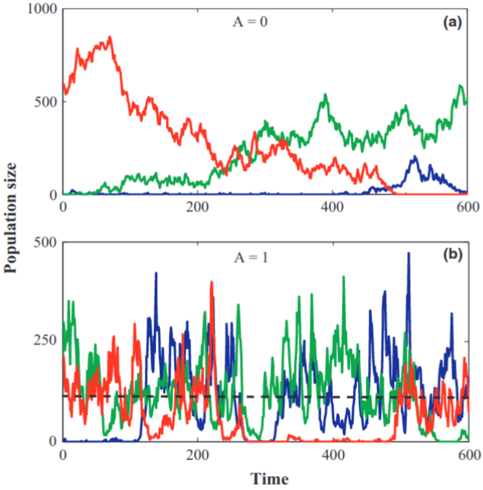
Why go beyond neutral? It fails for other patterns, e.g.

- More biomass when more species (neutral theory = zero-sum game, total biomass is fixed)



Why go beyond neutral? It fails for other patterns, e.g.

- Temporal fluctuations from original neutral theory are too slow



Neutral theory prediction
(fixed environment,
random birth-death)

More realistic timeseries
(fluctuating environment)

Random interactions

Next simplest thing:

- neutrality = identical interactions

$$a = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{8}$$

Random interactions

Next simplest thing:

- neutrality = identical interactions

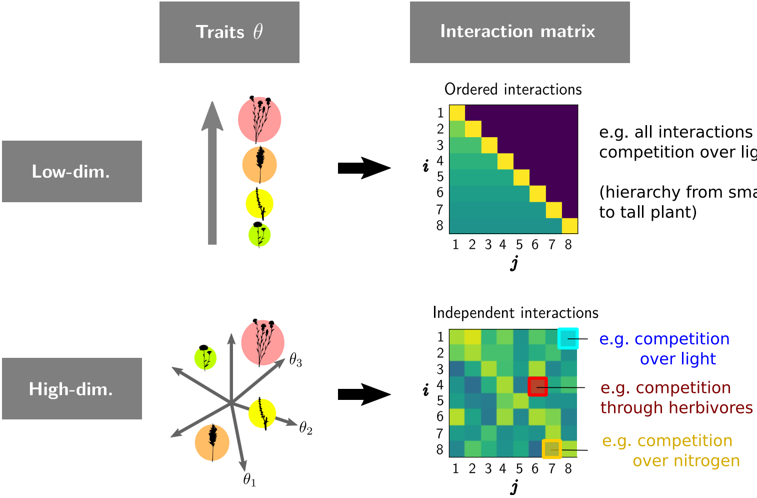
$$a = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{8}$$

- instead, take interactions a_{ij} that are different, but drawn at random

$$a = \begin{pmatrix} 0.29 & 0.54 & 0.53 & 0.02 & 0.40 \\ 0.57 & 0.86 & 0.90 & 0.81 & 0.76 \\ 0.53 & 0.11 & 0.42 & 0.44 & 0.09 \\ 0.15 & 0.72 & 0.84 & 0.27 & 0.94 \\ 0.87 & 0.85 & 0.61 & 0.36 & 0.63 \end{pmatrix} \tag{9}$$

what we do by default in a simulation when we don't know what numbers to put!

Justification: interactions not "really" uncertain, but caused by many independent ecological traits, mechanisms, etc.



Predictions

$$\frac{dN_i}{dt} = r_i N_i \left(1 - \sum_j^S a_{ij} N_j \right) \tag{10}$$

In principle, results could depend on every detail of the matrix, e.g. how we drew the random numbers (normal, uniform, etc.)

$$a = \begin{pmatrix} 0.29 & 0.54 & 0.53 & 0.02 & ... \\ 0.57 & 0.86 & 0.90 & 0.81 & ... \\ ... & & & & \end{pmatrix} \tag{11}$$

Predictions

$$\frac{dN_i}{dt} = r_i N_i \left(1 - \sum_j^S a_{ij} N_j\right) \tag{10}$$

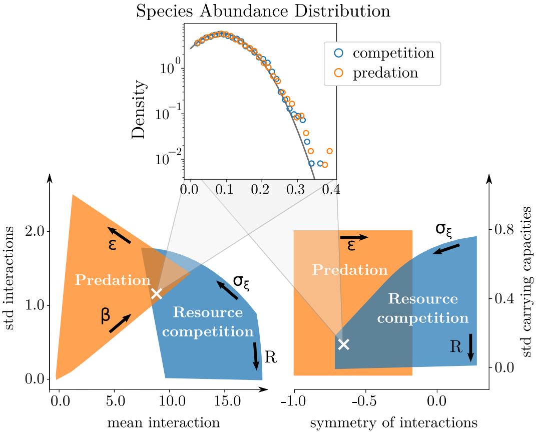
In principle, results could depend on every detail of the matrix, e.g. how we drew the random numbers (normal, uniform, etc.)

$$a = \begin{pmatrix} 0.29 & 0.54 & 0.53 & 0.02 & \dots \\ 0.57 & 0.86 & 0.90 & 0.81 & \dots \\ \dots & & & & \end{pmatrix} \tag{11}$$

In fact, under broad conditions, results only depend on 3 parameters

- mean of interactions $\langle a_{ij} \rangle$
- standard deviation $\mathrm{std}(a_{ij})$
- and symmetry $\mathrm{corr}(a_{ij}, a_{ji})$

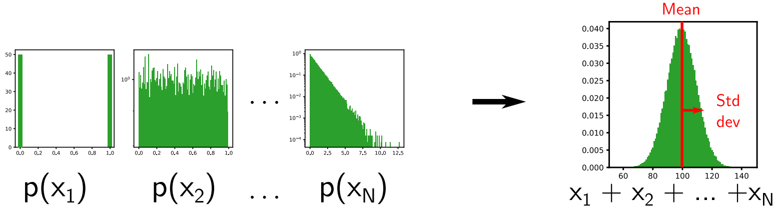
In particular, nature of interactions (competitive, trophic, parasitism...) is *irrelevant*, only statistics determine resulting patterns



e.g. two models, one with predation, one with competition, give same results (abundance distribution, etc.) if they produce the same statistics

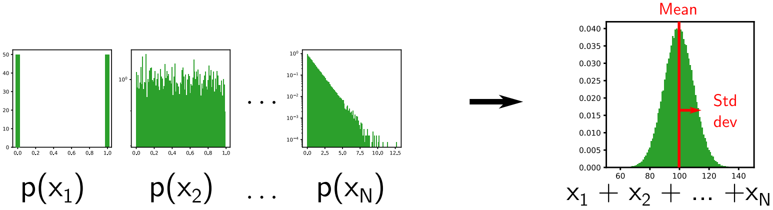
How is that possible?

- Like Central Limit Theorem: many independent variables together create a Gaussian, with only 2 parameters: mean and variance



How is that possible?

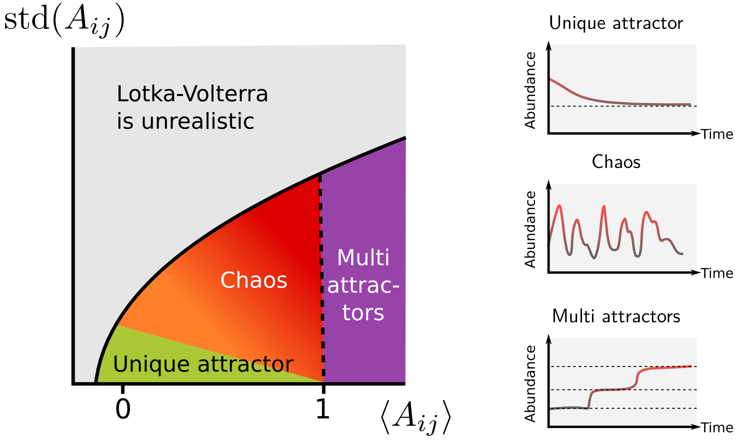
- Like Central Limit Theorem: many independent variables together create a Gaussian, with only 2 parameters: mean and variance



- Same is true with networks: many independent interactions together create a simple statistical result with only 3 parameters

How do we prove that result? Mathematical methods from physics

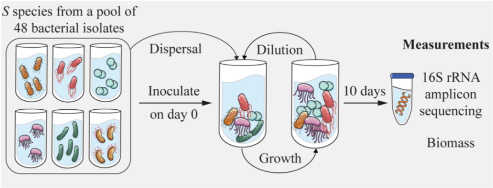
Predictions: dynamics



NB: Chaotic phase shows "realistic" fluctuations

Experimental setup: soil bacteria competition

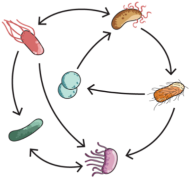
Hu et al. 2021 bioRxiv




Jiliang Hu

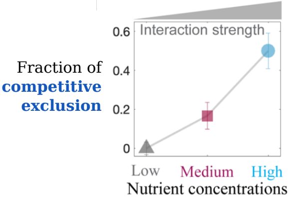

Jeff Gore

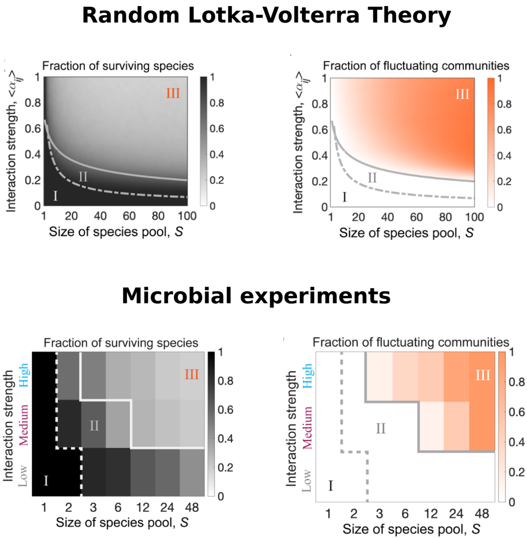
Unique feature: ability to control overall competition strength



Interaction matrix for bacteria pairs

1	0.69	<1	>1	0.08	0.11
0.31	1	0.41	0.99	0.26	0.25
>1	0.24	1	>1	0.20	0.20
<1	0.98	<1	1	>1	<1
0.21	0.36	0.18	<1	1	0.4
0.05	0.32	0.13	>1	0.87	1



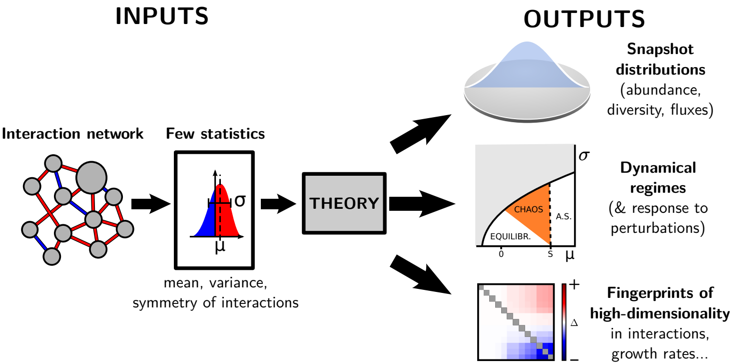


- - - - Survival boundary
—— Stability boundary

Phase I: stable full coexistence
Phase II: stable partial coexistence
Phase III: persistent fluctuation

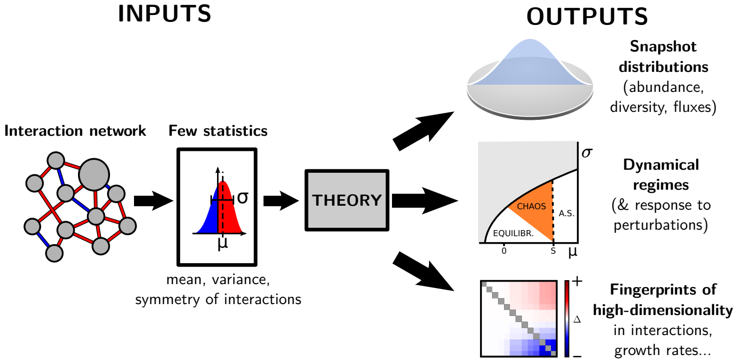
Random communities: a summary

Random interactions = a few input parameters, many testable outputs



Random communities: a summary

Random interactions = a few input parameters, many testable outputs



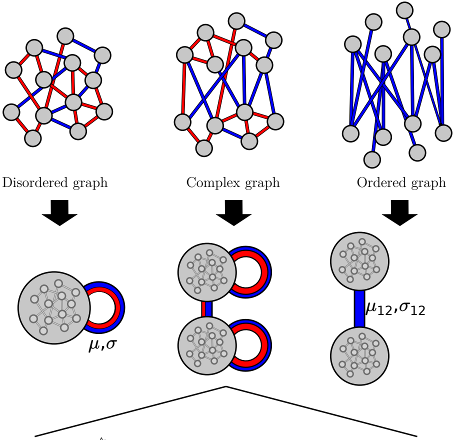
But do we really believe that systems are completely random?

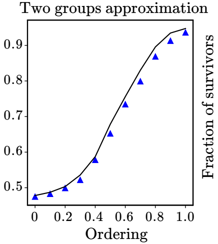
III. Order and disorder

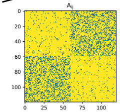
Combining order and disorder

"There is a fundamental dichotomy between structure and randomness, which in turn leads to a decomposition of any object into a structured (low-complexity) component and a random (discorrelated) component."
– Terence Tao

Claim: Often, apparently complex systems behave like interpolation between simple order & disorder





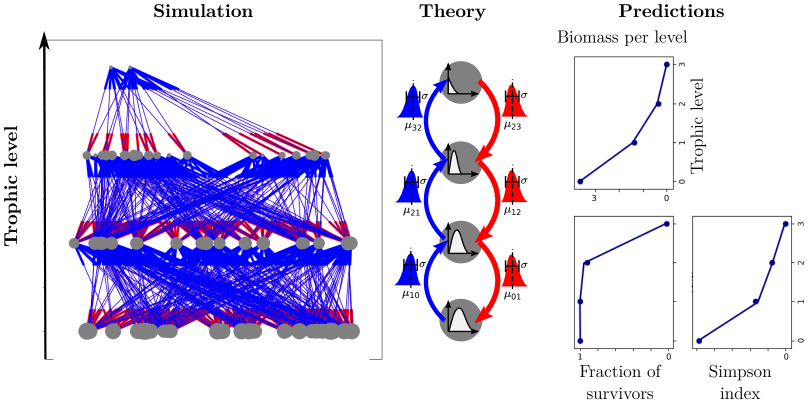


Matrix μ
$$\begin{pmatrix} 0.4 & 2 \\ 2 & 0.4 \end{pmatrix}$$

Matrix σ
$$\begin{pmatrix} 0.1 & 0.4 \\ 0.4 & 0.1 \end{pmatrix}$$

Matrix γ
$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

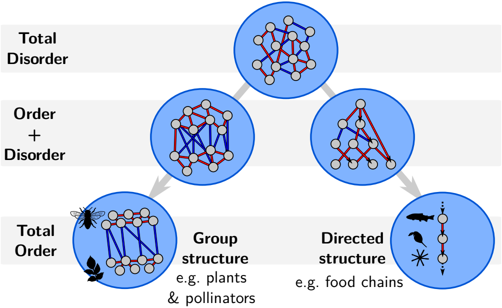
Example 2: food webs



(but also size hierarchy, nestedness, trade-offs...)

Two simplicities

In brief:



- Disorder = plausible null model for (single-functional group) communities with many factors causing interactions
- Order+disorder decomposition can reduce more complex systems to only few more parameters, but there are different types of simple order (most classically: blocks, nestedness, directedness)

Take aways

- Observables in large systems:
  - snapshot patterns: distribution and statistics
  - dynamics: number & nature of attractors, sensitivity to perturbations

Take aways

- Observables in large systems:
  - snapshot patterns: distribution and statistics
  - dynamics: number & nature of attractors, sensitivity to perturbations

- Can we really identify ecological mechanisms from observables, or can we explain observables without knowing much about mechanisms?
  - For many observables, not all details of mechanisms matter; the art of modelling involves understanding when and which details are lost

Take aways

- Observables in large systems:
  - snapshot patterns: distribution and statistics
  - dynamics: number & nature of attractors, sensitivity to perturbations
- Can we really identify ecological mechanisms from observables, or can we explain observables without knowing much about mechanisms?
  - For many observables, not all details of mechanisms matter; the art of modelling involves understanding when and which details are lost
- Randomness = particular case where we can prove that all details are lost except a few basic statistics
  - useful as null model; to know if network structure is important for a result, compare to result of random networks with similar statistics

Take aways

- Observables in large systems:
  - snapshot patterns: distribution and statistics
  - dynamics: number & nature of attractors, sensitivity to perturbations
- Can we really identify ecological mechanisms from observables, or can we explain observables without knowing much about mechanisms?
  - For many observables, not all details of mechanisms matter; the art of modelling involves understanding when and which details are lost
- Randomness = particular case where we can prove that all details are lost except a few basic statistics
  - useful as null model; to know if network structure is important for a result, compare to result of random networks with similar statistics
  - can be mixed with simple structure (e.g. functional groups, nestedness...) to model "complex" networks
    $\Rightarrow$ what seems complex may be largely random