Simplifying complexity

—

Large communities and random interactions

Matthieu Barbier

(Plant Health Institute Montpellier, CIRAD
& Institut Natura e Teoria en Pirenèus)

20/03/2025

Before everything – background for why I am giving you this lecture, in 2 questions and 1 graph!

The questions: when I analyze data, do I believe

The questions: when I analyze data, do I believe

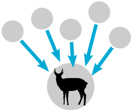
- I can neatly separate what I measure into causes and consequences (independent & dependent variables), or not?

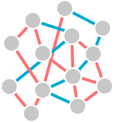
The questions: when I analyze data, do I believe

- I can neatly separate what I measure into causes and consequences (independent & dependent variables), or not?
- I can only focus on a few things that matter, or not?

The questions: when I analyze data, do I believe

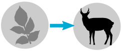
- I can neatly separate what I measure into causes and consequences (independent & dependent variables), or not?
- I can only focus on a few things that matter, or not?



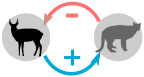


Many things matter

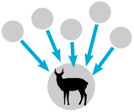
Few things matter

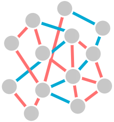


One-way causality



Feedbacks





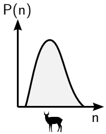
forget who does what,
count what happens

Many things matter

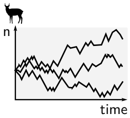
PCA,
network
metrics,
...

Few things matter

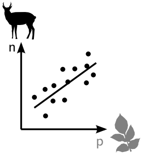
identify who does what



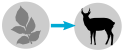
histograms



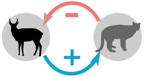
random walks



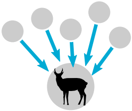
regressions

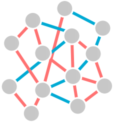


One-way causality



Feedbacks





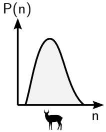
forget who does what,
count what happens

Many things matter

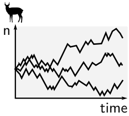
PCA,
network
metrics,
...

Few things matter

identify who does what

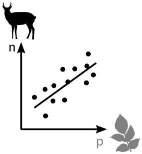


histograms



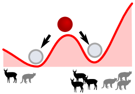
random walks

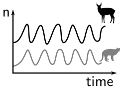
???



regressions

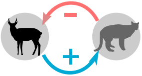
joint states or dynamics
e.g. alternate states, cycles





One-way causality

more stats ⟷ more dyn



Feedbacks

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

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

Note: Hereafter I will be talking about community ecology & species interactions, but many of the ideas apply for complex systems of genes, behaviors...

Preamble: an illustration in simulated data

Model

Assume we can use a model roughly like Lotka-Volterra with S species

$$\frac{dN_i}{dt} = r_i N_i (1 - a_{ii} N_i - \sum_{j \neq i}^{S} a_{ij} N_j) \quad (1)$$

Model

Assume we can use a model roughly like Lotka-Volterra with S species

$$\frac{dN_i}{dt} = r_i N_i (1 - a_{ii} N_i - \sum_{j \neq i}^{S} a_{ij} N_j) \tag{1}$$

we need many parameters:

Model

Assume we can use a model roughly like Lotka-Volterra with S species

$$\frac{dN_i}{dt} = r_i N_i (1 - a_{ii} N_i - \sum_{j \neq i}^{S} a_{ij} N_j) \tag{1}$$

we need many parameters:

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

$$r = (?, ?, ?....) \tag{2}$$

Model

Assume we can use a model roughly like Lotka-Volterra with S species

$$\frac{dN_i}{dt} = r_i N_i (1 - a_{ii} N_i - \sum_{j \neq i}^{S} a_{ij} N_j) \tag{1}$$

we need many parameters:

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

$$r = (?, ?, ?....) \tag{2}$$

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

$$a = \begin{pmatrix} ? & ? & ... \\ ? & & \\ ... & & \end{pmatrix} \tag{3}$$

Model

Practical modeller's strategy: don't know r_i and a_{ij} in your system? take them at random!

$$a = \begin{pmatrix} 0.29 & 0.54 & 0.53 & 0.02 & ... \\ 0.57 & 0.86 & 0.90 & 0.81 & ... \\ ... \end{pmatrix} \tag{4}$$

Model

Practical modeller's strategy: don't know r_i and a_{ij} in your system? take them at random!

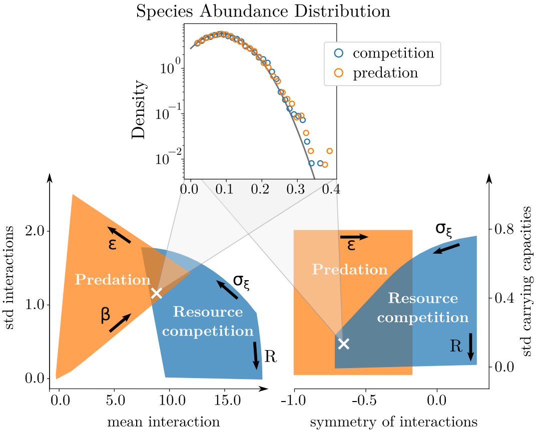
$$a = \begin{pmatrix} 0.29 & 0.54 & 0.53 & 0.02 & \dots \\ 0.57 & 0.86 & 0.90 & 0.81 & \dots \\ \dots & & & & \end{pmatrix} \tag{4}$$

Problem: results could depend on many details of the matrix, e.g. each value, or how we drew the random numbers (normal, uniform, etc.)

Predictions

In fact, if you simulate again and again, many results only depend on a few global statistics of the community:

- mean of interactions $\langle a_{ij} \rangle$ ($= \mathrm{E}[a_{ij}]$ if you prefer)
- standard deviation $\mathrm{std}(a_{ij})$
- and symmetry $\mathrm{corr}(a_{ij}, a_{ji})$

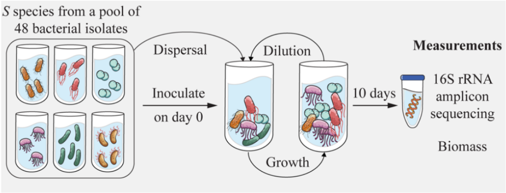


In particular, nature of interactions (competitive, trophic, parasitism...) is *irrelevant*!

e.g. two models, one with predation, one with competition, give same results (abundance distribution, etc.) if they correspond to the same mean, variance and symmetry of a_{ij}

Empirical test

Experimental setup: soil bacteria competition **(Hu et al, Science 2022)**

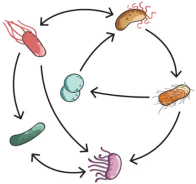


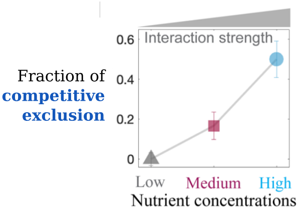

Jiliang Hu


Jeff Gore

Very well controlled environment & replicable results

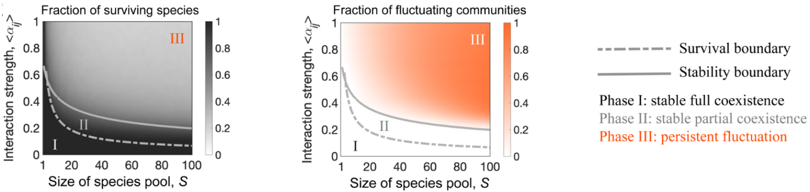
Unique feature: ability to **control how strongly species interact**



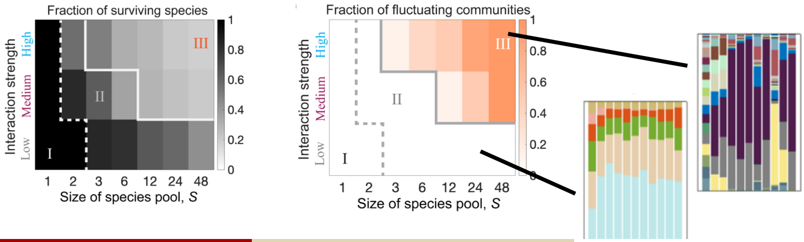


white = unique equilibrium, orange = fluctuations (in theory, chaos)

Random Lotka-Volterra Theory



Microbial experiments



Now that I gave this illustration, two questions:

- how is that possible?
- how does this help us understand ecology?

I. Introduction: Complexity and simplicity

Where to put complexity

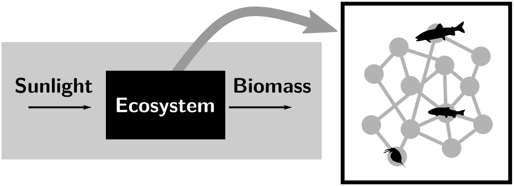
Basic question of modelling: which details are important to include?

Where to put complexity

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

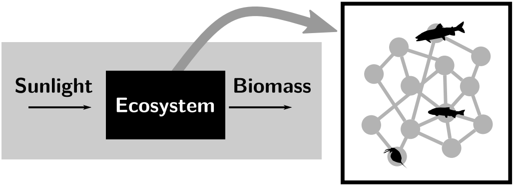
- whenever we write a simple model in biology, we are hiding complexity



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

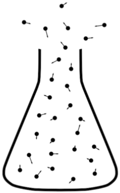


- is there a *principled* way of understanding when this is a valid choice?

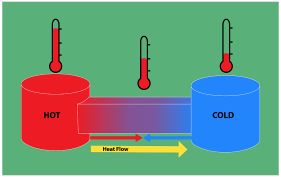
Idea originating from physics

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

10^{23} variables:
position of every molecule

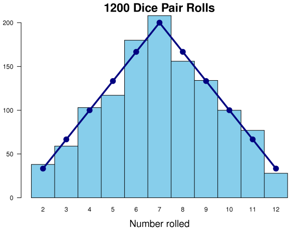


2 variables: temperature & pressure



uncountable factors, chaotic motion

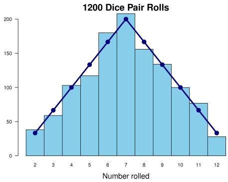
1 probability distribution



Meaning of randomness

uncountable factors, chaotic motion

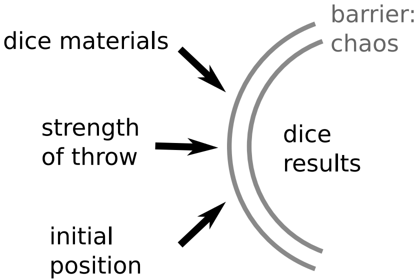
1 probability distribution



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

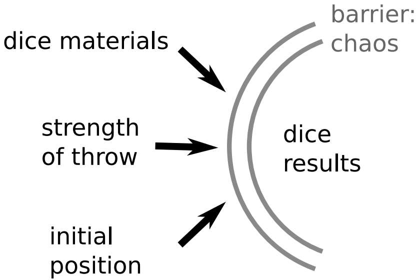
(vs e.g. basketball throws, where you need to know who is throwing, how far...)

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

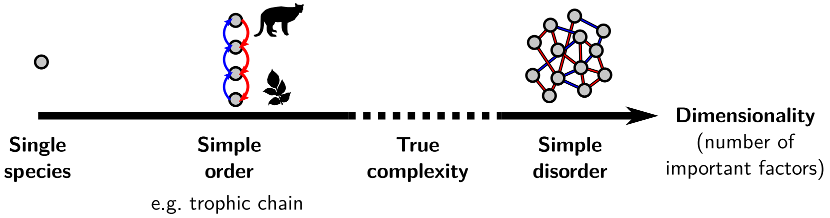
“Random” means “too many factors”, so complex mechanistically that it becomes simple statistically

Small and large systems

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...

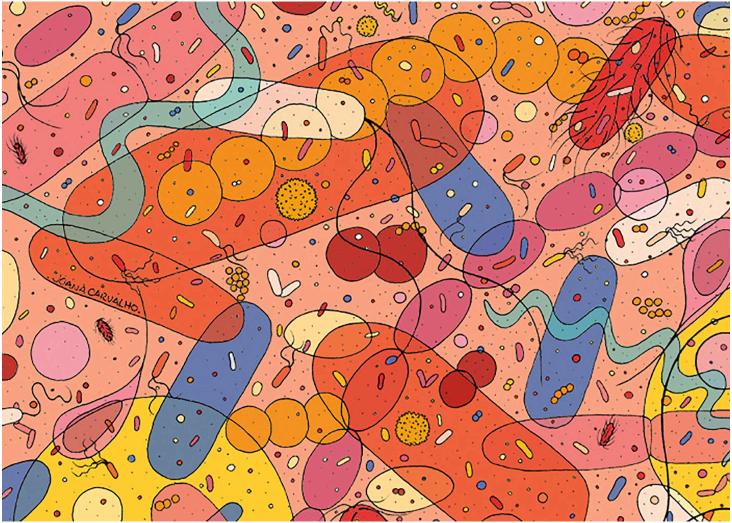
  but when & how are they simple?

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 strains, phenotypes, etc.

Observables in large communities

What is interesting in large communities:

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

Observables in large communities

What is interesting in large communities:

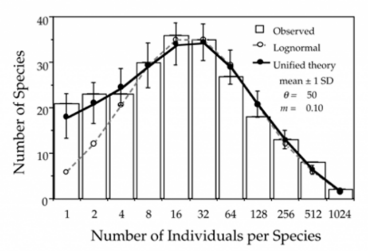
- we lose focus on individual species – they are often unpredictable, maybe impacted by dozens or hundreds of others
- we gain aggregate properties
  - static properties
  - dynamical properties

Static properties

Measurable from a single/few snapshots:

Measurable from a single/few snapshots:

- Distributions (= histograms, frequencies)

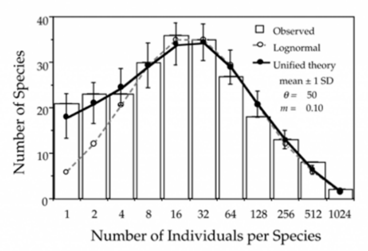


abundances, number of offsprings/production, variation in space or time...

Static properties

Measurable from a single/few snapshots:

- Distributions (= histograms, frequencies)

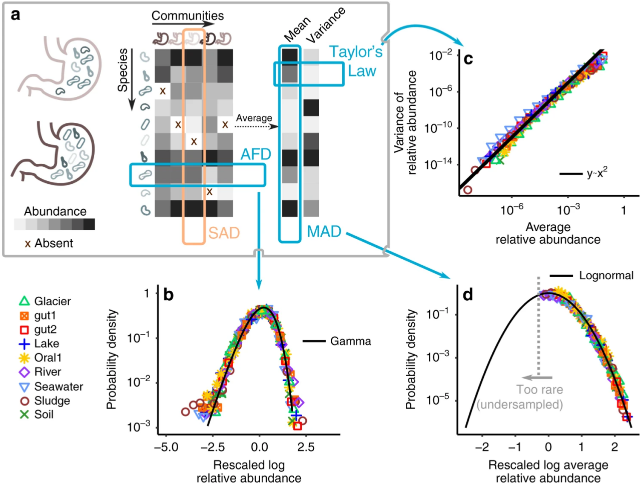


  abundances, number of offsprings/production, variation in space or time...

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

Static properties

Many common patterns are different ways of aggregating same basic data



Grilli (2020) Nature Communications

Dynamical properties

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

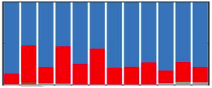
- Is an ecosystem in a stable equilibrium or other dynamical regime?
- How does it respond when you disturb it?

Dynamical properties

What is the usual regime of a given ecosystem?

- equilibrium (with small fluctuations)

  example: near-constant populations of microbial functional groups

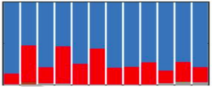


Dynamical properties

What is the usual regime of a given ecosystem?

- equilibrium (with small fluctuations)

  example: near-constant populations of microbial functional groups

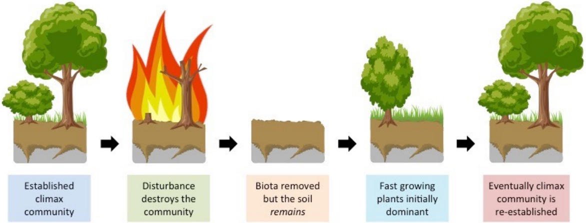


- directional trajectory

  example: microbial succession during 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

  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

  example: gut microbiome disturbed by sickness then re-colonized

- "malleable", remains modified, does not go back

  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

  example: gut microbiome disturbed by sickness then re-colonized

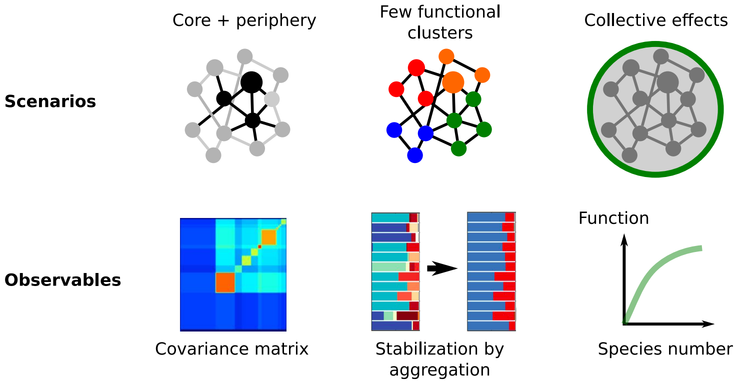
- "malleable", remains modified, does not go back

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

- "chaotic", becomes more and more different

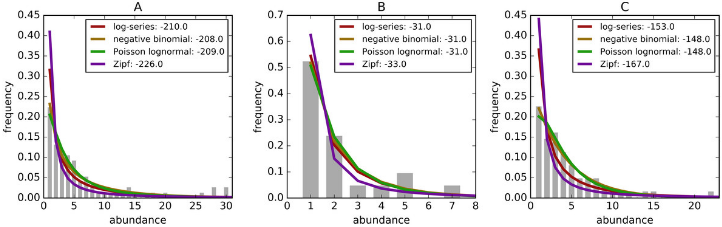
  example: a single invasive species causes a cascade of extinctions, then invasions, then extinctions...

Why do we care? Often, these patterns are used as “fingerprints” to reveal some ecological scenarios...



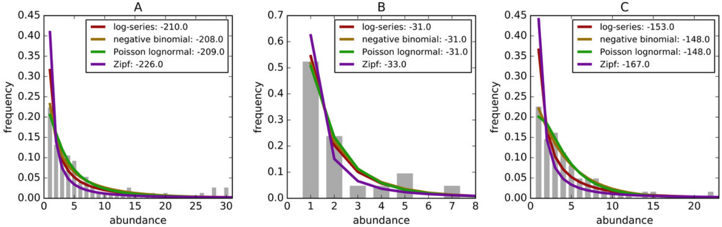
Fingerprints of ecological scenarios

... But usually no single pattern is a "smoking gun" (e.g. many different explanations for same abundance histograms)



Fingerprints of ecological scenarios

... But usually no single pattern is a "smoking gun" (e.g. many different explanations for same abundance histograms)



⇒ **Goal: How do we construct a simple model that explains as many community patterns as possible?** (to allow a real test)

II. Many-species communities

–

Part 2: How do we explain observations

Model

Back to our Lotka-Volterra

$$\frac{dN_i}{dt} = r_i N_i (1 - a_{ii} N_i - \sum_{j \neq i}^{S} a_{ij} N_j) \tag{5}$$

Model

Back to our Lotka-Volterra

$$\frac{dN_i}{dt} = r_i N_i (1 - a_{ii} N_i - \sum_{j \neq i}^{S} a_{ij} N_j) \tag{5}$$

- maybe the snapshot properties (e.g. abundance distribution) are given by some *equilibrium* of the equation?

$$\frac{dN_i}{dt} = 0 \Rightarrow \begin{cases} N_i = 0 & \text{extinct species} \\ N_i = (1 - \sum_{j \neq i}^{S} a_{ij} N_j)/a_{ii} & \text{surviving species} \end{cases} \tag{6}$$

Model

Back to our Lotka-Volterra

$$\frac{dN_i}{dt} = r_i N_i (1 - a_{ii} N_i - \sum_{j \neq i}^{S} a_{ij} N_j) \tag{5}$$

- maybe the snapshot properties (e.g. abundance distribution) are given by some *equilibrium* of the equation?

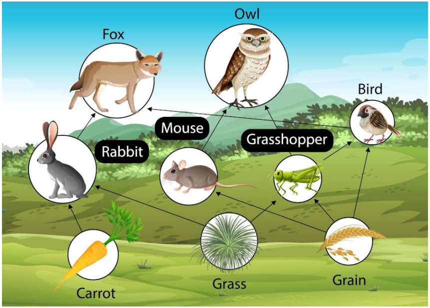
$$\frac{dN_i}{dt} = 0 \Rightarrow \begin{cases} N_i = 0 & \text{extinct species} \\ N_i = (1 - \sum_{j \neq i}^{S} a_{ij} N_j)/a_{ii} & \text{surviving species} \end{cases} \tag{6}$$

- maybe dynamical properties can be understood, e.g. how does $N_i(t)$ respond to perturbation, does it go to equilibrium or cycle...

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

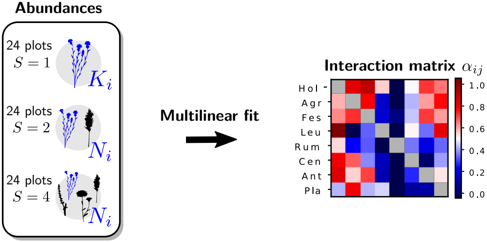


How do we obtain the matrix of interactions a_{ij}?

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

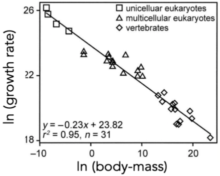
Grassland plants (Wageningen)

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



Species interaction networks

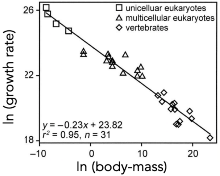
- Most of the time, theoretical assumptions are needed to put numbers into the model:
  - Metabolic scaling, r_i and a_{ij} given by body sizes of species i and j



  - 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



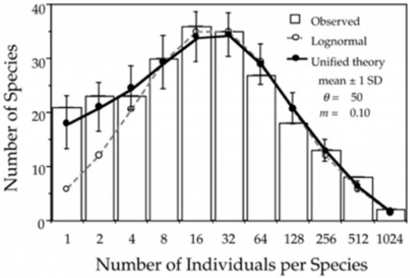
    - Ecopath model (see with Claire this afternoon)
    - ...
- What do we do if we cannot or do not want to assume anything?

Neutrality

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

Neutrality

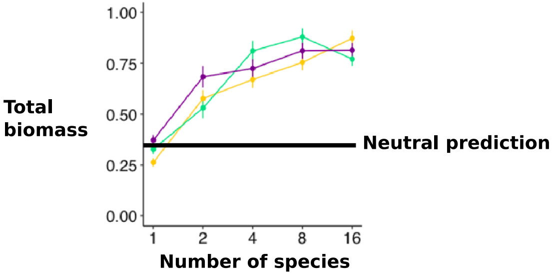
- Extreme simplification: neutrality, all species identical, e.g. $r_i = a_{ij} = 1$
- Different outcomes for different species only due to chance: random events of birth, death and migration
- Why use it? Because it can suffice to predict some patterns, e.g. abundance distributions



Neutrality

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

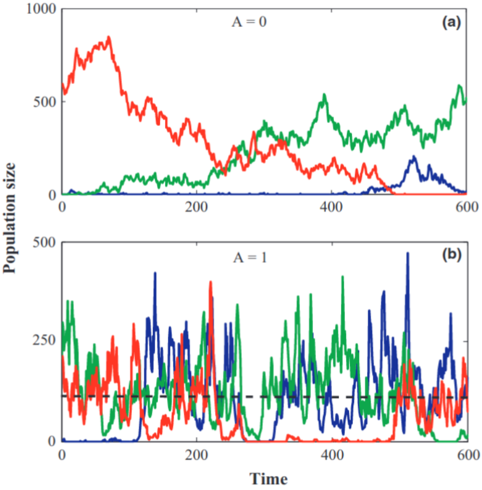
- More biomass when more species (neutral theory = zero-sum game, total biomass is fixed)



Neutrality

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)

Next simplest thing:

- neutrality = identical interactions

$$a = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{7}$$

Next simplest thing:

- neutrality = identical interactions

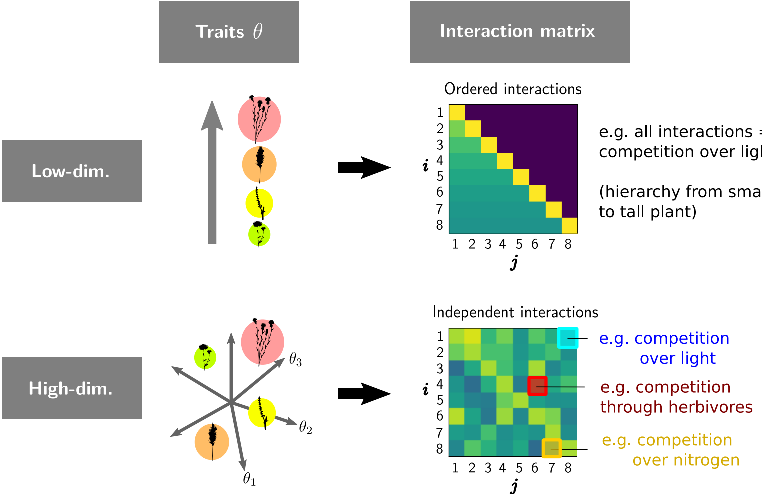
$$a = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{7}$$

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

$$a = \begin{pmatrix} 0.29 & 0.54 & 0.53 & 0.02 & 0.40 \\ 0.57 & 0.86 & 0.90 & 0.81 & 0.76 \\ 0.53 & 0.11 & 0.42 & 0.44 & 0.09 \\ 0.15 & 0.72 & 0.84 & 0.27 & 0.94 \\ 0.87 & 0.85 & 0.61 & 0.36 & 0.63 \end{pmatrix} \tag{8}$$

Random interactions

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



Predictions

As I said at the start, under broad conditions, results only depend on 3 parameters

- mean of interactions $\langle a_{ij} \rangle$ ($= \mathrm{E}[a_{ij}]$ if you prefer)
- standard deviation $\mathrm{std}(a_{ij})$
- and symmetry $\mathrm{corr}(a_{ij}, a_{ji})$

How?

From a system of S equilibrium equations on all the abundances
(here $a_{ii} = 1$ for simplicity)

$$N_1 = 1 - \sum_{j \neq 1}^{S} a_{1j} N_j \quad (9)$$

$$N_2 = 1 - \sum_{j \neq 2}^{S} a_{2j} N_j \quad (10)$$

$$\ldots \quad (11)$$

How?

From a system of S equilibrium equations on all the abundances
(here $a_{ii} = 1$ for simplicity)

$$N_1 = 1 - \sum_{j \neq 1}^{S} a_{1j} N_j \tag{9}$$

$$N_2 = 1 - \sum_{j \neq 2}^{S} a_{2j} N_j \tag{10}$$

$$\dots \tag{11}$$

We can go to a few equations on statistics of abundances, almost like this (a bit more complicated)

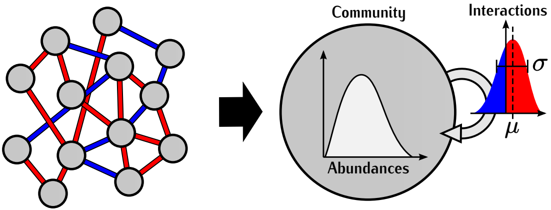
$$\langle N \rangle = 1 - S \, \langle a \rangle \langle N \rangle \tag{12}$$

$$\mathrm{var}(N) = S \, \mathrm{var}(a) \langle N^2 \rangle \tag{13}$$

where the **only parameters** are number of species S, and statistics of interactions $\langle a \rangle$, $\mathrm{var}(a)$ (or $\mathrm{std}(a)$)

Why?

Why does it work to just write equations on statistics (e.g. mean and variance) of abundance?

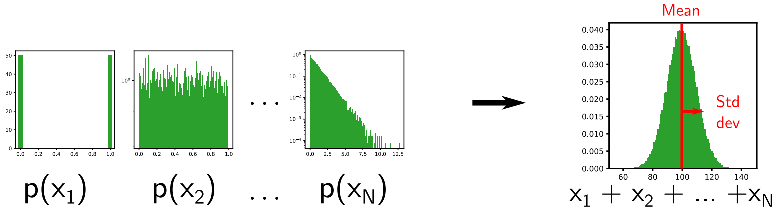


In a disordered network, all species are different but statistically equivalent

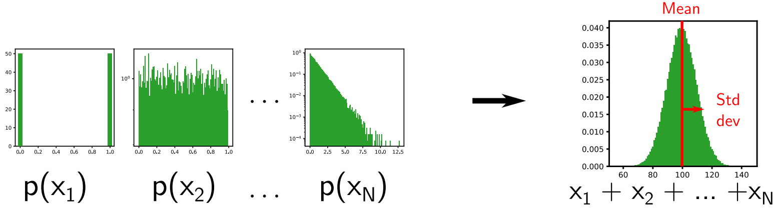
(no special role/position, any one at random is a fair sample from whole distribution)

Why?

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



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

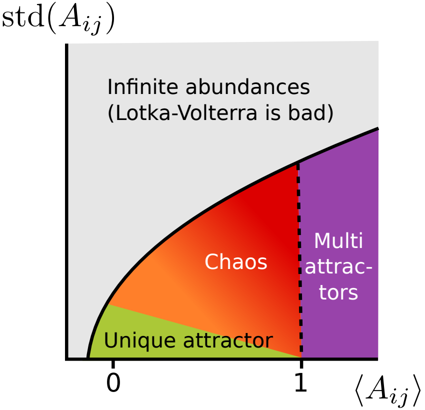


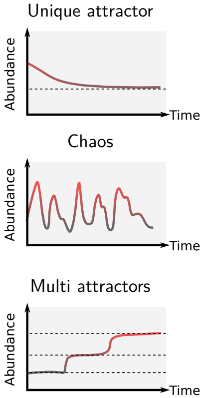
- Same is true with networks: many independent interactions together create a simple statistical result with only 3 parameters

How do we prove that result? Mathematical methods from physics

Predictions: dynamics

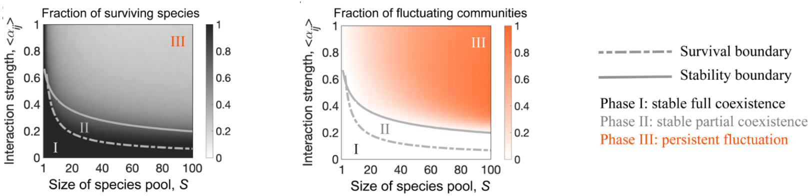
- Same reasoning works for dynamics, "$d\langle N\rangle/dt$, $d\mathrm{var}(N)/dt$..."
- Hence, with few parameters, we can answer the question of how a many-species network behaves dynamically "by default" (when random)



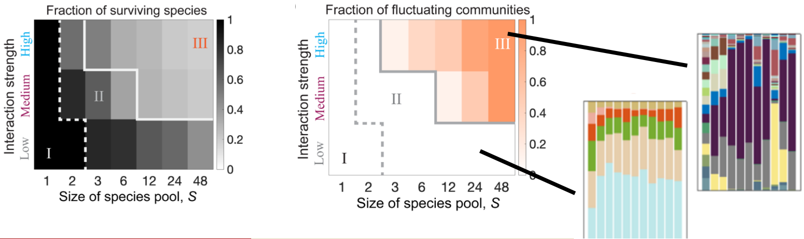


NB: Chaotic phase shows "realistic" fluctuations white = unique equilibrium, orange = fluctuations (in theory, chaos)

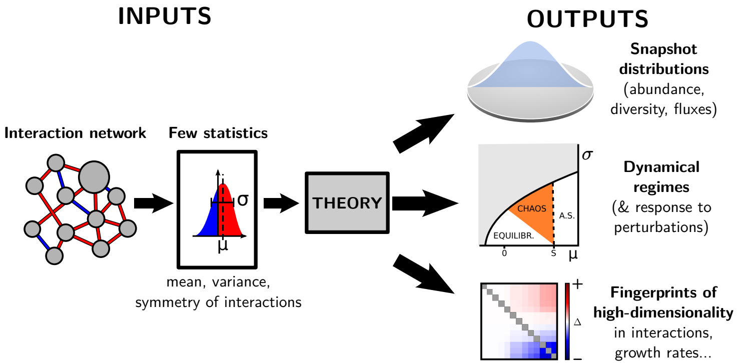
Random Lotka-Volterra Theory



Microbial experiments



Random interactions = a few input parameters, many testable outputs



Random interactions

40 years of empirical successes for random interaction models:

- In physics: glass, large atoms
- In chemistry: large reaction networks
- In neuroscience: neural networks
- In computer science: complex satisfaction problems (NP-hardness)

Random interactions

40 years of empirical successes for random interaction models:

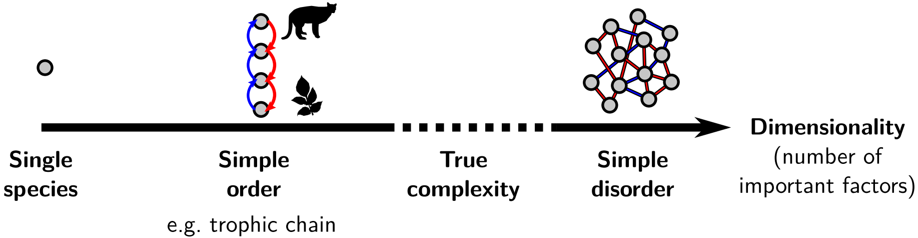
- In physics: glass, large atoms
- In chemistry: large reaction networks
- In neuroscience: neural networks
- In computer science: complex satisfaction problems (NP-hardness)

... but do we really believe that ecological systems can be modelled as completely random?

III. Order and disorder

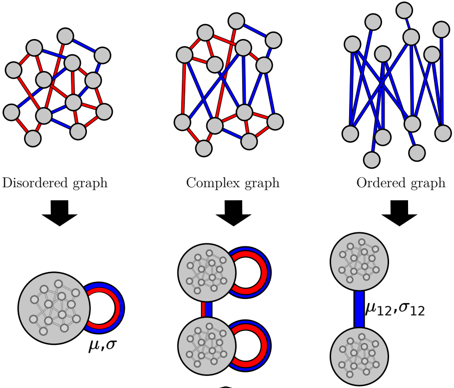
Combining order and disorder

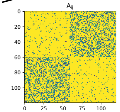
"There is a fundamental dichotomy between structure and randomness, which in turn leads to a decomposition of any object into a structured (low-complexity) component and a random (discorrelated) component."
– Terence Tao



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

Example 1: competitors and mutualists

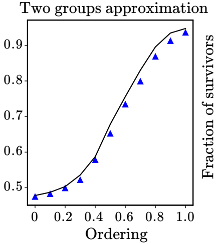




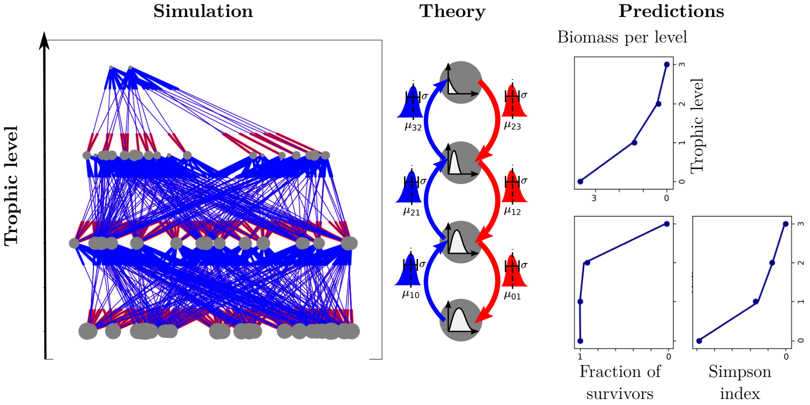
Matrix μ $\begin{pmatrix} 0.4 & 2 \\ 2 & 0.4 \end{pmatrix}$

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

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



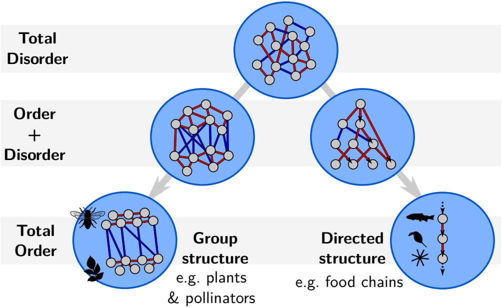
Example 2: food webs



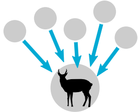
(but also size hierarchy, nestedness, trade-offs...)

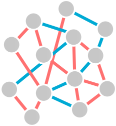
Two simplicities

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)





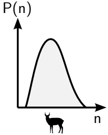
forget who does what,
count what happens

Many things matter

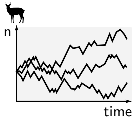
keystone spp/traits, PCA, network metrics, ...

Few things matter

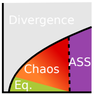
identify who does what



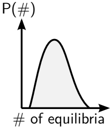
histograms



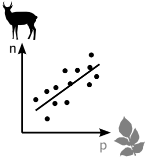
random walks



collective dynamical regimes

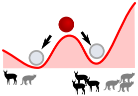


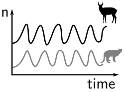
counting states



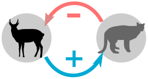
regressions

joint states or dynamics
e.g. alternate states, cycles





more stats ⟷ more dyn



One-way causality

Feedbacks

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
  - **drawing things at random does not mean you explore all possibilities!**
  - 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