Combinatorial Species

A tool for the perplexed mathematical biologist

Egor Lappo

October 16, 2023

Follow along at https://github.com/EgorLappo/tanglegram-species

motivation

(perfectly plausible) problems

- **Tanglegrams.** How many tanglegrams are there for your set of species?

(perfectly plausible) problems

- **Tanglegrams.** How many tanglegrams are there for your set of species? What if trifurcations are allowed?

(perfectly plausible) problems

- **Tanglegrams.** How many tanglegrams are there for your set of species? What if trifurcations are allowed? k-furcations?

(perfectly plausible) problems

- **Tanglegrams.** How many tanglegrams are there for your set of species? What if trifurcations are allowed? k-furcations?
- **Allele identity configurations.** For n individuals with k alleles at a locus, how many identity states are there?

(perfectly plausible) problems

- **Tanglegrams.** How many tanglegrams are there for your set of species? What if trifurcations are allowed? k-furcations?
- **Allele identity configurations.** For n individuals with k alleles at a locus, how many identity states are there? What if individuals are distinguishable?

(perfectly plausible) problems

- **Tanglegrams.** How many tanglegrams are there for your set of species? What if trifurcations are allowed? k-furcations?
- **Allele identity configurations.** For n individuals with k alleles at a locus, how many identity states are there? What if individuals are distinguishable?
- **Monkeys.** Each of the n monkeys gives a fruit to another monkey. How many exchange configurations need to be considered?

(perfectly plausible) problems

- **Tanglegrams.** How many tanglegrams are there for your set of species? What if trifurcations are allowed? k-furcations?
- **Allele identity configurations.** For n individuals with k alleles at a locus, how many identity states are there? What if individuals are distinguishable?
- **Monkeys.** Each of the n monkeys gives a fruit to another monkey. How many exchange configurations need to be considered? What if monkeys can't give fruits to themselves?

(perfectly plausible) problems

- **Tanglegrams.** How many tanglegrams are there for your set of species? What if trifurcations are allowed? k-furcations?
- **Allele identity configurations.** For n individuals with k alleles at a locus, how many identity states are there? What if individuals are distinguishable?
- **Monkeys.** Each of the n monkeys gives a fruit to another monkey. How many exchange configurations need to be considered? What if monkeys can't give fruits to themselves? What if there are blue monkeys and red monkeys?

what is the *first* thing you do?

what is the *first* thing you do?

- Compute.

what is the *first* thing you do?

- Compute.
- Guess.

what is the *first* thing you do?

- Compute.
- Guess.
- Write proofs.

what is the *first* thing you do?

- Compute.
- Guess.
- Write proofs.
- Try out the combinatorial species approach!

definition

Definition
A **species of structures** is a rule F that

- for each finite set U gives a finite set $F[U]$,
- for each bijection $U \to V$ gives a function $F[U] \to F[V]$,
- respects composition.

Definition
A **species of structures** is a rule F that

- for each finite set U gives a finite set $F[U]$,
- for each bijection $U \to V$ gives a function $F[U] \to F[V]$,
- respects composition.

Example
Species $\mathcal{A}$ of rooted trees, $\mathcal{G}$ of simple graphs, $\mathcal{S}$ of permutations, Par of partitions, etc.

Definition

A **species of structures** is a rule F that

- for each finite set U gives a finite set $F[U]$,
- for each bijection $U \to V$ gives a function $F[U] \to F[V]$,
- respects composition.

Example

Species $\mathcal{A}$ of rooted trees, $\mathcal{G}$ of simple graphs, $\mathcal{S}$ of permutations, Par of partitions, etc.

The juice of the theory is in the *combinators* of species and *associated series*.

U, V are finite sets, σ is a bijection.

$$\begin{tikzcd}
U \arrow[r] \arrow[d, "\sigma"'] & F[U] \arrow[d, "F[\sigma]"] \\
V \arrow[r] & F[V]
\end{tikzcd}$$

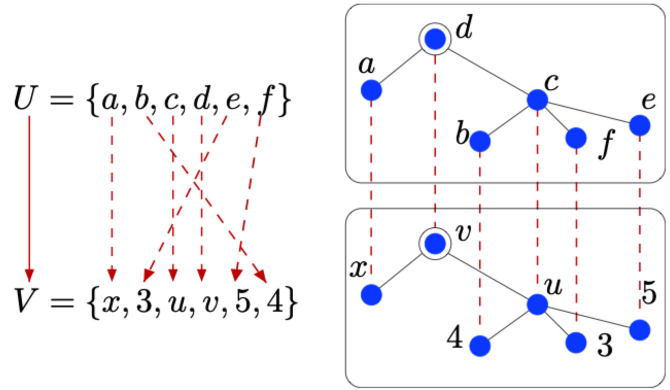
definition

U, V are finite sets, σ is a bijection.

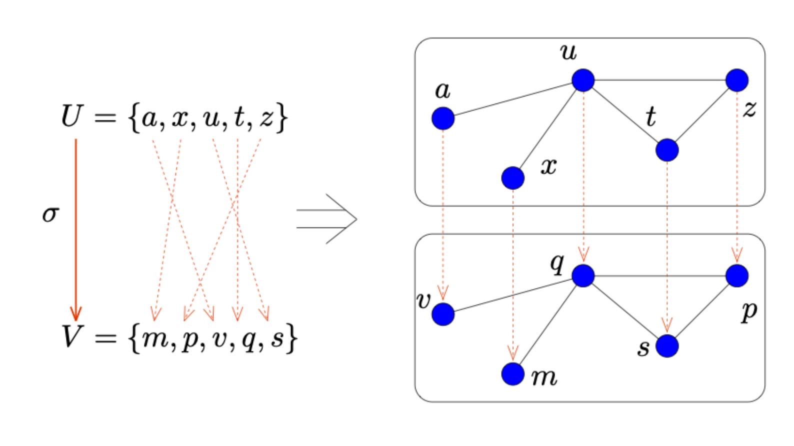
$$
\begin{tikzcd}
U \arrow[r] \arrow[d, "\sigma"'] & F[U] \arrow[d, "F[\sigma]"] \\
V \arrow[r] & F[V]
\end{tikzcd}
$$

$F[\sigma]$ is *not* necessarily a bijection.

species of rooted trees



species of simple graphs



key aspects

- Several basic constructions — building blocks
- Combinators — generalizing generating functions
- Associated series — enumeration
- Functional equations, Lagrange inversion, virtual species, …

Species of sets E: $E[U] = \{U\}$.

$$\begin{tikzcd} U \arrow[r] \arrow[d, "\sigma"'] & E[U] = \{U\} \arrow[d, "E[\sigma]"] \\ V \arrow[r] & E[V] = \{\sigma(U)\} \end{tikzcd}$$

Unique choice for each U.

basic species: sets

Species of elements ε: $\varepsilon[U] = U$.

$$\begin{tikzcd}
U \arrow[r] \arrow[d, "\sigma"'] & \varepsilon[U] = U \arrow[d, "\varepsilon[\sigma]=\sigma"] \\
V \arrow[r] & \varepsilon[V] = V
\end{tikzcd}$$

Kind of identity.

basic species: specific cardinality

- Species 1, characteristic of the empty set, with

$$1[U] = \begin{cases} \{U\}, & \text{if } U = \varnothing \\ \varnothing, & \text{otherwise.} \end{cases}$$

basic species: specific cardinality

- Species 1, characteristic of the empty set, with

$$1[U] = \begin{cases} \{U\}, & \text{if } U = \varnothing \\ \varnothing, & \text{otherwise.} \end{cases}$$

- Species X, characteristic of singletons, with

$$X[U] = \begin{cases} \{U\}, & \text{if } |U| = 1 \\ \varnothing, & \text{otherwise.} \end{cases}$$

basic species: specific cardinality

- Species 1, characteristic of the empty set, with

$$1[U] = \begin{cases} \{U\}, & \text{if } U = \varnothing \\ \varnothing, & \text{otherwise.} \end{cases}$$

- Species X, characteristic of singletons, with

$$X[U] = \begin{cases} \{U\}, & \text{if } |U| = 1 \\ \varnothing, & \text{otherwise.} \end{cases}$$

- Species E_2 of pairs, etc.

the part where is all pays off

Definition
Let F be a species. The **cycle index series** of F is the formal power series

$$Z_F(x_1, x_2, x_3, \ldots) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\sum_{\sigma \in \mathcal{S}_n} \operatorname{fix} F[\sigma]\, x_1^{\sigma_1} x_2^{\sigma_2} x_3^{\sigma_3} \cdots \right),$$

where σ are all permutations of $[n]$, $\operatorname{fix} F[\sigma]$ is the number of fixed points of $F[\sigma]$, and σ_i is the number of cycles of σ of length i.

Definition
Let F be a species. The **cycle index series** of F is the formal power series

$$Z_F(x_1, x_2, x_3, \ldots) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\sum_{\sigma \in \mathcal{S}_n} \mathrm{fix}\, F[\sigma]\, x_1^{\sigma_1} x_2^{\sigma_2} x_3^{\sigma_3} \cdots \right),$$

where σ are all permutations of $[n]$, $\mathrm{fix}\, F[\sigma]$ is the number of fixed points of $F[\sigma]$, and σ_i is the number of cycles of σ of length i.

This contains *all* information on the enumeration of F.

example

Let $F[n]$ be the species of rooted trees on n leaves.

$$F[3] = \big\{a = ((1,2),3), b = ((2,3),1), c = ((3,1),2)\big\}$$

Let $\sigma = \{1,2,3\} \to \{1,3,2\}$. We have $\sigma_1 = 1$, $\sigma_2 = 1$.

Then $F[3](a) = b$, $F[3](b) = a$, $F[3](c) = c$, so that $\operatorname{fix} F[\sigma] = 1$.

associated series II

Let $F(x)$ be the exponential generating function for F, counting labeled F-structures,

$$F(x) = \sum_{n=0}^{\infty} |F[n]| \frac{x^n}{n!}.$$

Let $\widetilde{F}(x)$ be the generating function enumerating **unlabeled** F-structures,

$$\widetilde{F(x)} = \sum_{n=0}^{\infty} \widetilde{f}_n x^n.$$

Theorem
For any species F, we have

$$F(x) = Z_F(x, 0, 0, \ldots),$$
$$\widetilde{F(x)} = Z_F(x, x^2, x^3, \ldots).$$

Theorem
For any species F, we have

$$F(x) = Z_F(x, 0, 0, \ldots),$$
$$\widetilde{F(x)} = Z_F(x, x^2, x^3, \ldots).$$

Combinations of species correspond to operations on Z_F.

what are the combinators?

A **sum species** $F + G$ is a disjoint union.

combinators: sum

A **sum species** $F + G$ is a disjoint union.

$$|(F+G)[n]| = |F[n]| + |G[n]|,$$

$$Z_{F+G} = Z_F + Z_G.$$

combinators: sum

A **sum species** $F + G$ is a disjoint union.

$$|(F + G)[n]| = |F[n]| + |G[n]|,$$

$$Z_{F+G} = Z_F + Z_G.$$

For example, $E = 1 + X + E_2 + E_3 + \cdots$.

A **product species** $F \cdot G$ is an F- and a G- structure on two complementary disjoint subsets.

For any finite set U,

$$(F \cdot G)[U] = \sum_{(U_1, U_2)} F[U_1] \times G[U_2].$$

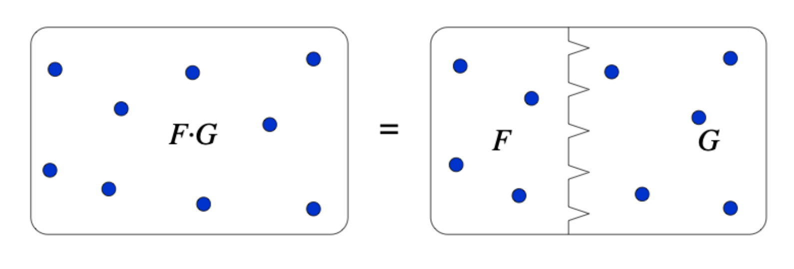
A **product species** $F \cdot G$ is an F- and a G- structure on two complementary disjoint subsets.

For any finite set U,

$$(F \cdot G)[U] = \sum_{(U_1, U_2)} F[U_1] \times G[U_2].$$

$$Z_{F \cdot G} = Z_F Z_G$$

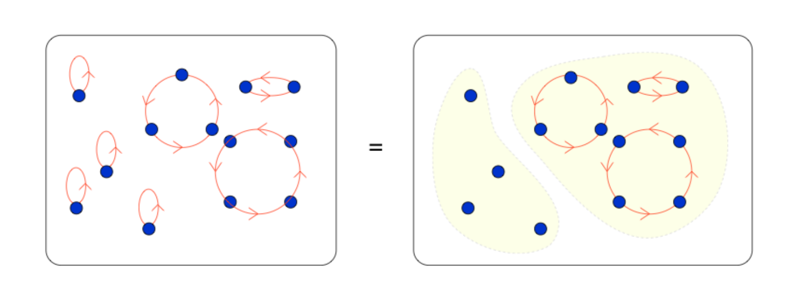
combinators: product



combinators: product

A permutation is a set of fixed points together with a derangement.

$$\mathcal{S} = E \cdot \mathrm{Der}.$$



combinators: substitution

Finally, a composite species $F \circ G$ represents an F-assembly of disjoint G-structures.

combinators: substitution

Finally, a composite species $F \circ G$ represents an F-assembly of disjoint G-structures.

For any finite set U,

$$(F \circ G)[U] = \sum_{\pi \text{ partition of } U} F[\pi] \times \prod_{p \in \pi} G[p].$$

combinators: substitution

Finally, a composite species $F \circ G$ represents an F-assembly of disjoint G-structures.

For any finite set U,

$$(F \circ G)[U] = \sum_{\pi \text{ partition of } U} F[\pi] \times \prod_{p \in \pi} G[p].$$

For the associated series, we have

$$\begin{aligned} Z_{F \circ G}(x_1, x_2, x_3, \ldots) &= Z_F(Z_G(x_1, x_2, x_3, \ldots), Z_G(x_2, x_4, \ldots), \ldots, \\ \widetilde{(F \circ G)}(x) &= Z_F(\widetilde{G(x)}, \widetilde{G(x^2)}, \widetilde{G(x^3)}, \ldots) \\ (F \circ G)(x) &= F(G(x)) \end{aligned}$$

combinators: substitution

Finally, a composite species $F \circ G$ represents an F-assembly of disjoint G-structures.

For any finite set U,

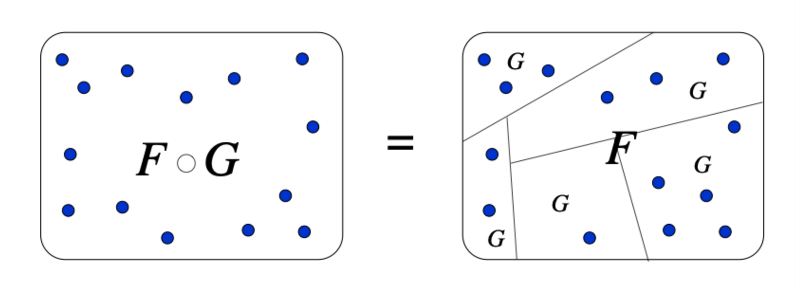
$$(F \circ G)[U] = \sum_{\pi \text{ partition of } U} F[\pi] \times \prod_{p \in \pi} G[p].$$

For the associated series, we have

$$\begin{aligned}
Z_{F \circ G}(x_1, x_2, x_3, \ldots) &= Z_F(Z_G(x_1, x_2, x_3, \ldots), Z_G(x_2, x_4, \ldots), \ldots, \\
\widetilde{(F \circ G)}(x) &= Z_F(\widetilde{G}(x), \widetilde{G}(x^2), \widetilde{G}(x^3), \ldots) \\
(F \circ G)(x) &= F(G(x))
\end{aligned}$$

Note that we can't do this unlabeled enumeration without the cycle index series.

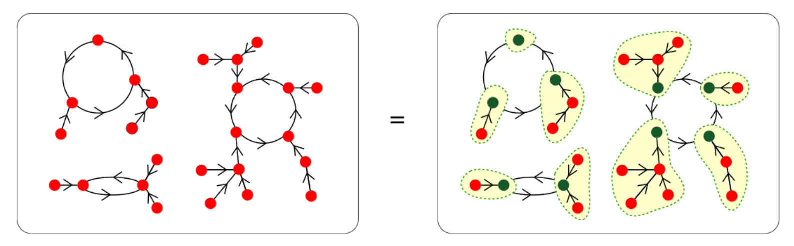
combinators: substitution



combinators: substitution

An endofunction is a permutation of trees,

$$\mathrm{End} = \mathcal{S} \circ \mathcal{A} = \mathcal{S}(\mathcal{A}).$$



Combinatorial species is a *language*.

Combinatorial species is a *language*.

What else is there?

- Species derivatives
- Weighted species, multisort species
- Virtual species
- Lagrange inversion to solve $Y = H(X, Y)$
- **Computer code**