

ALIEN

Prédire les interactions entre espèces avec R

Par Steve Vissault (M.Sc), et Kévin Cazelles (Ph.D)

Chaire de recherche en Écologie Intégrative



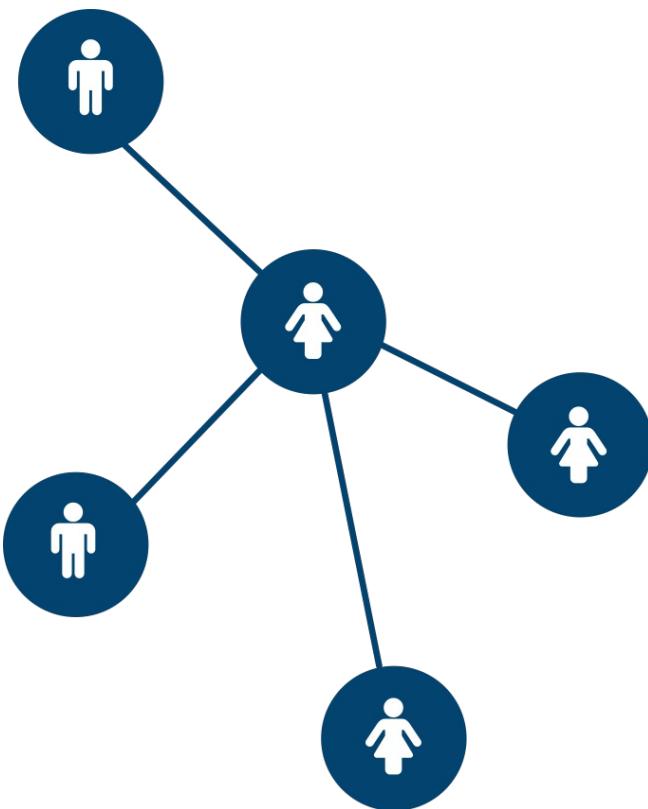
MISE EN CONTEXTE

Les réseaux: une façon d'abstraire le monde

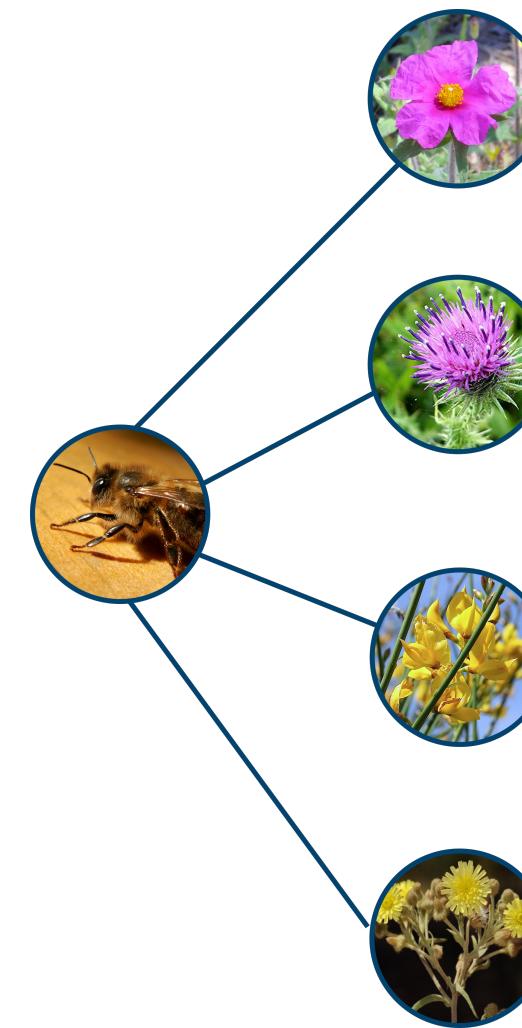
Réseaux de transports



Réseaux sociaux

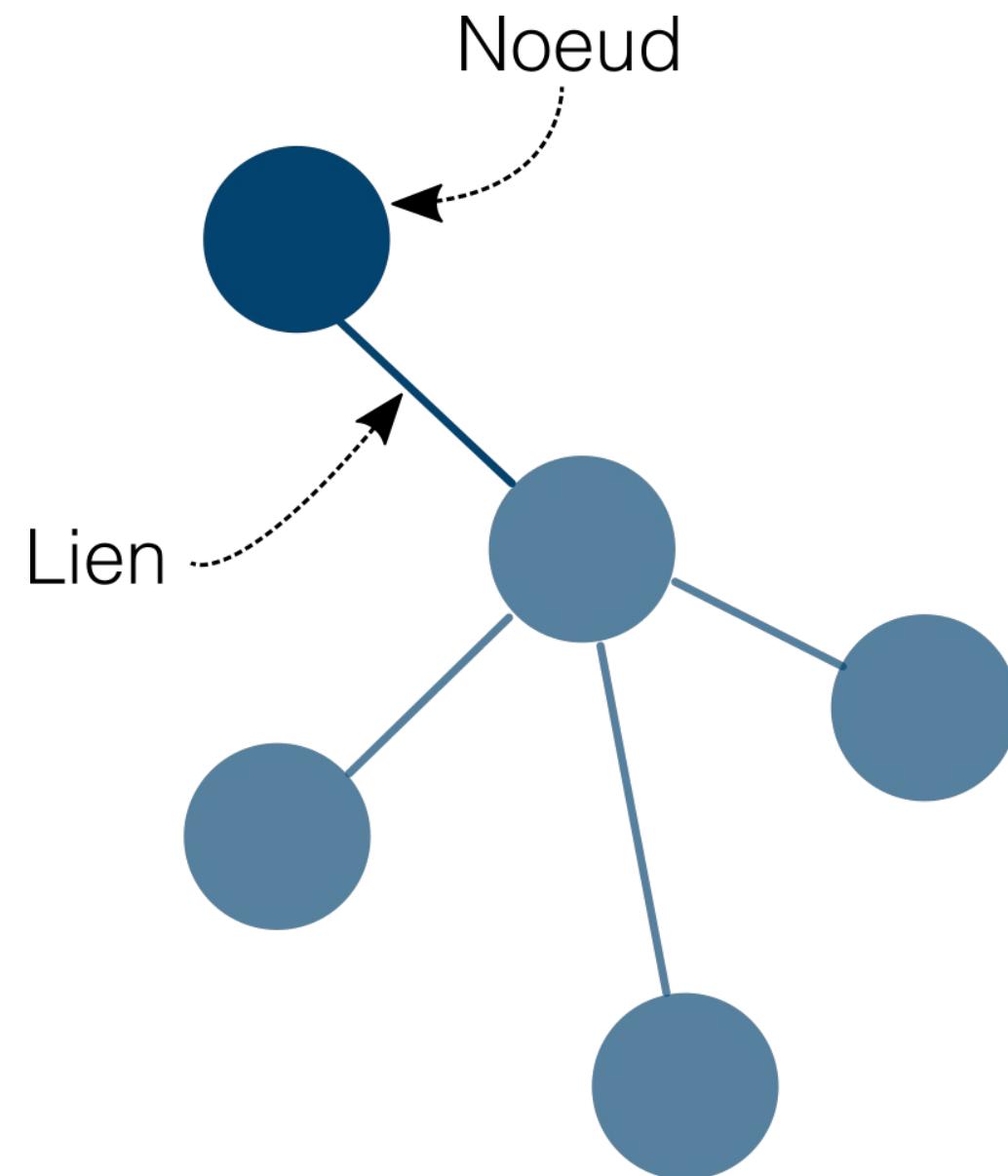


Réseaux écologiques



DÉFINITION D'un réseau

Une sémantique conjointe

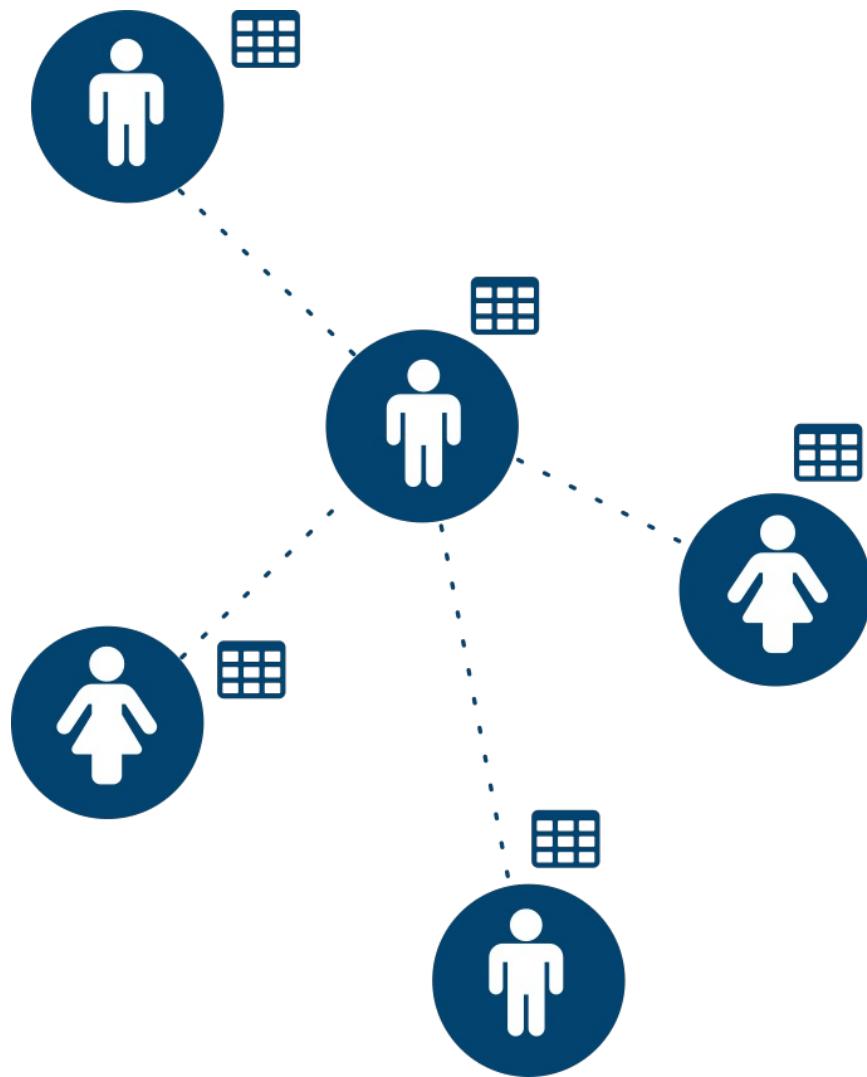


- Théorie des graphs
- Connus dans les disciplines de l'informatique et des mathématiques
- Un graph est un assemblage de **noeuds** reliés par des **liens**

Prédire à Partir des Propriétés

Les propriétés des noeuds

Les réseaux sociaux

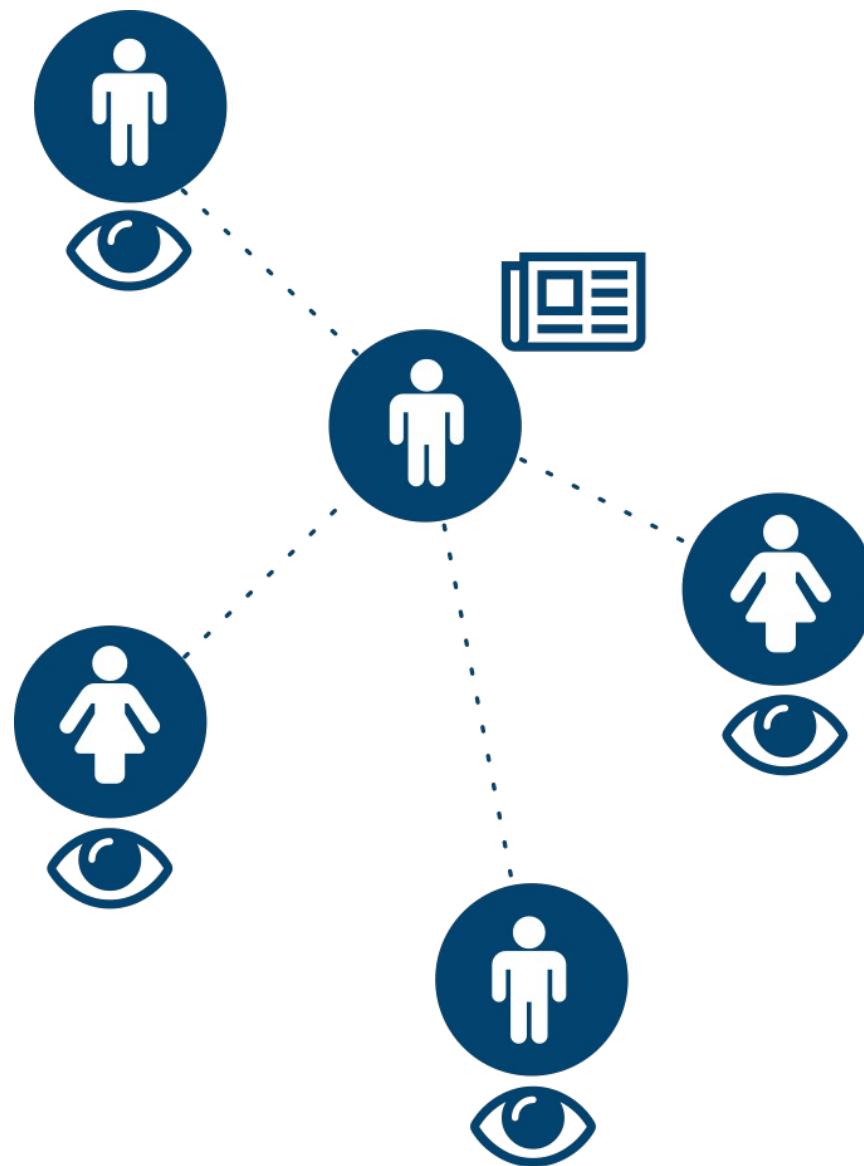


Chaque membre d'un réseau dispose de propriétés qui le définissent:

- Âge
- Sexe
- Lieu de naissance
- Profession
- Études etc.

Les propriétés des noeuds

Comment prédire les liens à partir des propriétés des noeuds?

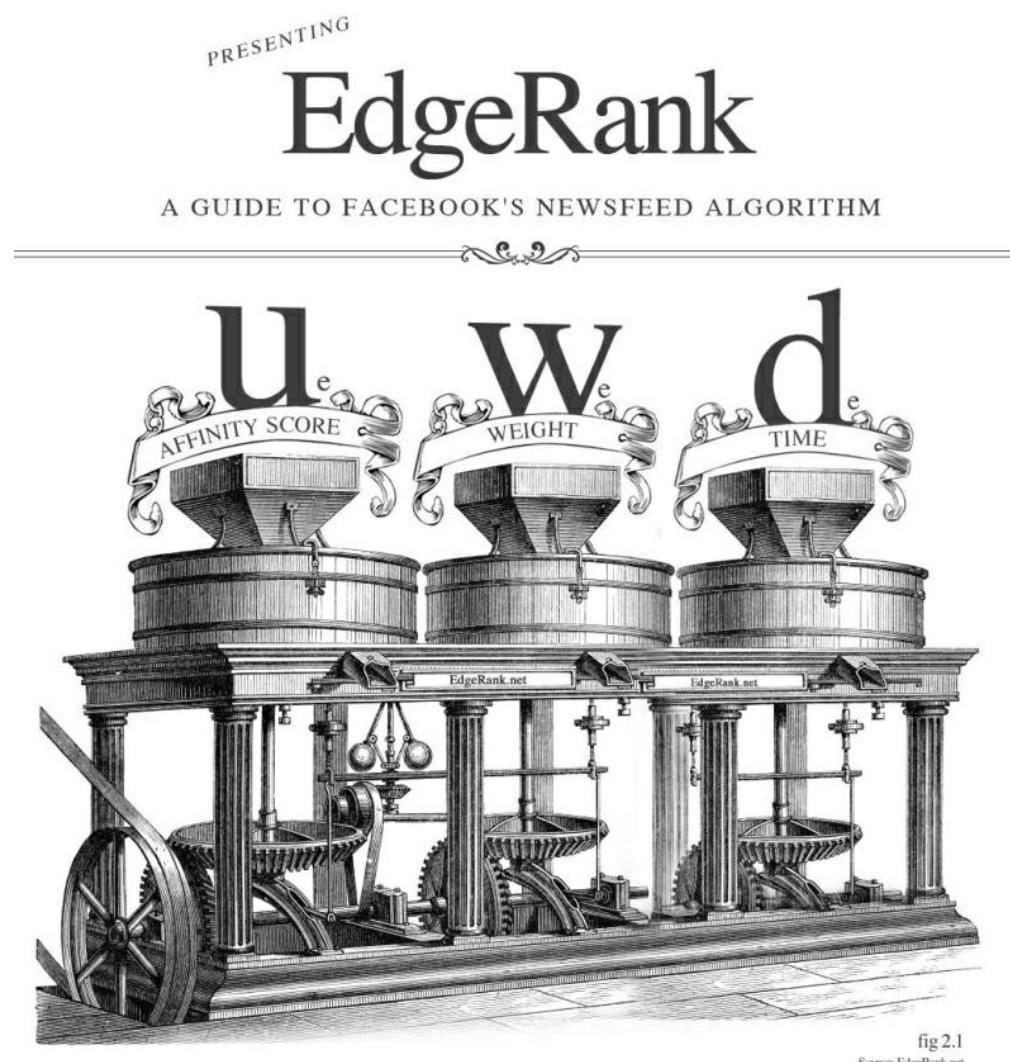


Si je diffuse une nouvelle...

Est-ce que cette nouvelle sera visible
sur le fil d'actualité de mes amis
Facebook (création d'un lien)?

Prédire un lien

Définir un lien à partir des propriétés d'un noeud



Simple algorithme:

$$\sum_{lien=e} u_e w_e d_e$$

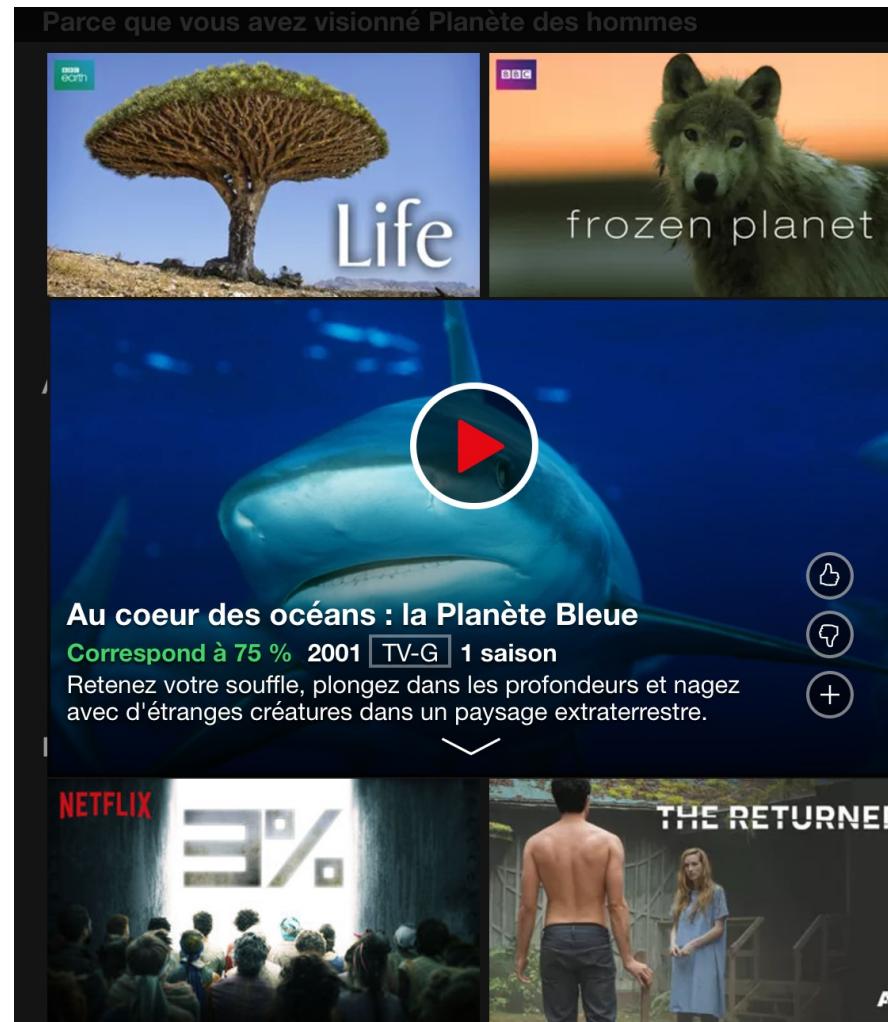
Où:

- u_e est le lien entre mon ami et moi (ayant posté la nouvelle)
- w_e le poids attribué à ce lien
- d_e le temps depuis la création de la nouvelle

Prédire un lien

Ce type d'algorithms de rangs sont répandus à travers de nombreuses plateformes.

NETFLIX



amazon

Frequently bought together



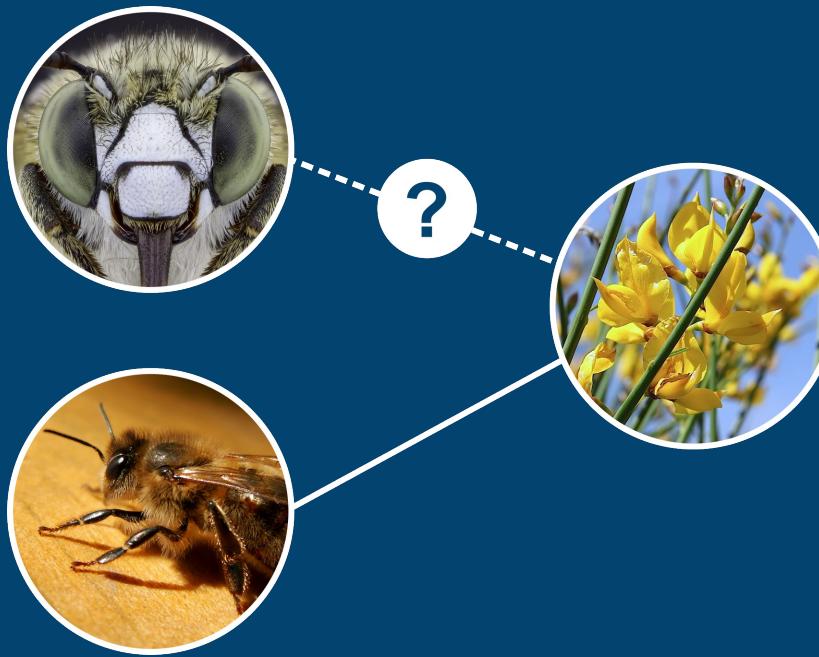
- This item: Eloquent JavaScript: A Modern Introduction to Programming by Marijn Haverbeke Paperback CDN\$ 25.57
- JavaScript: The Good Parts by Douglas Crockford Paperback CDN\$ 25.57
- Web Design with HTML, CSS, JavaScript and jQuery Set by Jon Duckett Paperback CDN\$ 55.34

Customers who bought this item also bought



ÉCOLOGIE

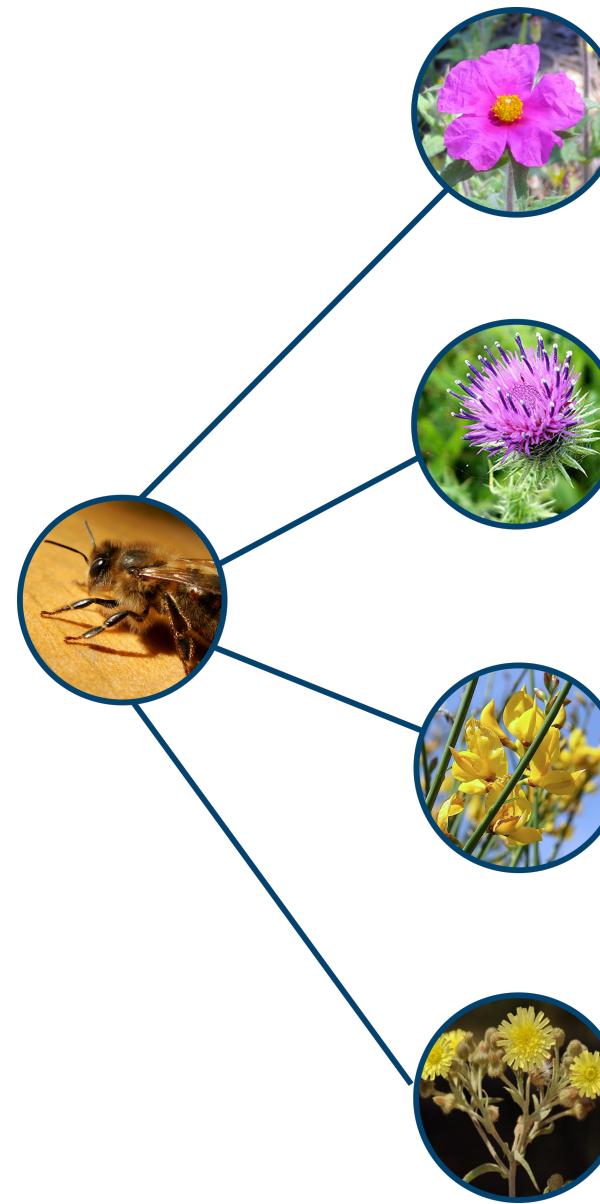
Peux-t-on faire la même chose
en écologie?



Les propriétés des noeuds

Les réseaux écologiques

Réseau plantes-pollinisateur

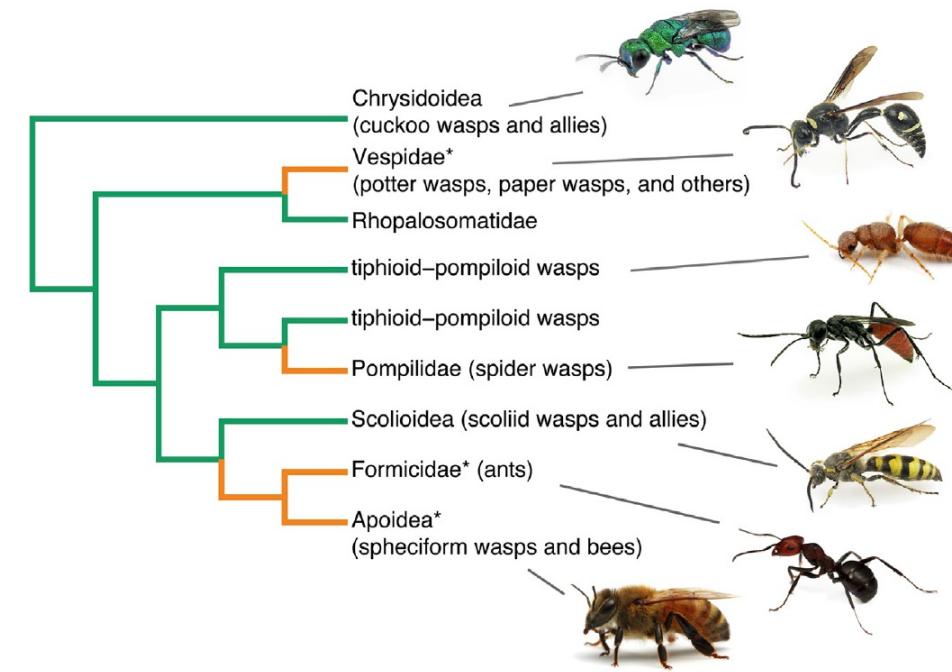


Bartomeus *et al.*, 2017

Les propriétés des organismes

DNA Phylogénie

Indique le degré de parenté d'un organisme avec les autres organismes du réseau (distance génétique)

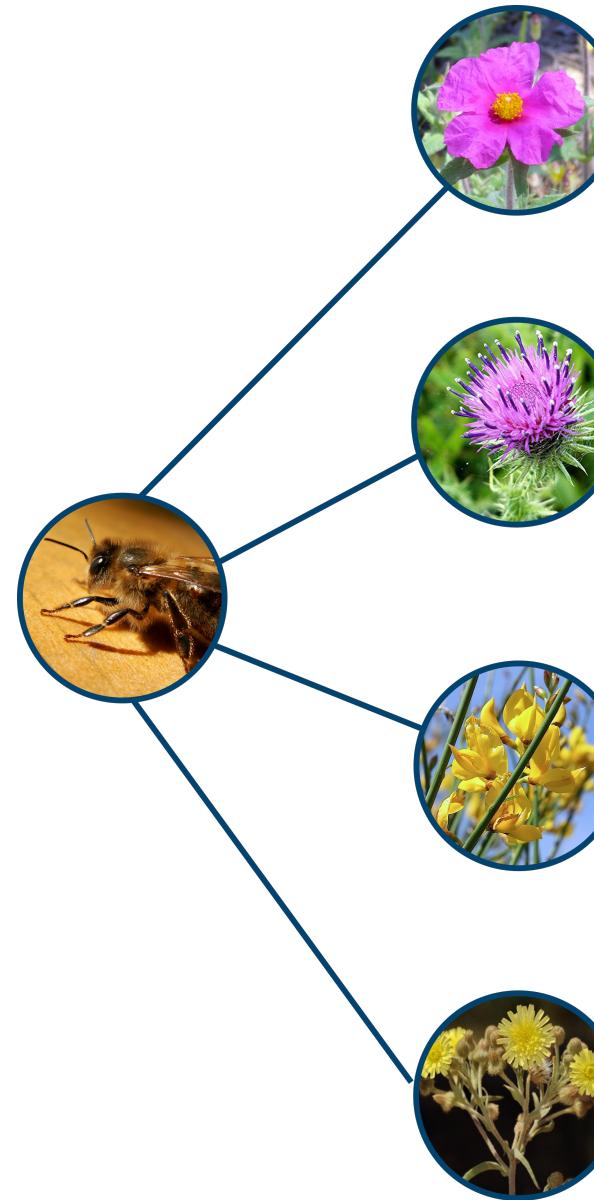


A. Wild & K. Schaefer

Les propriétés des noeuds

Les réseaux écologiques

Réseau plantes-pollinisateur

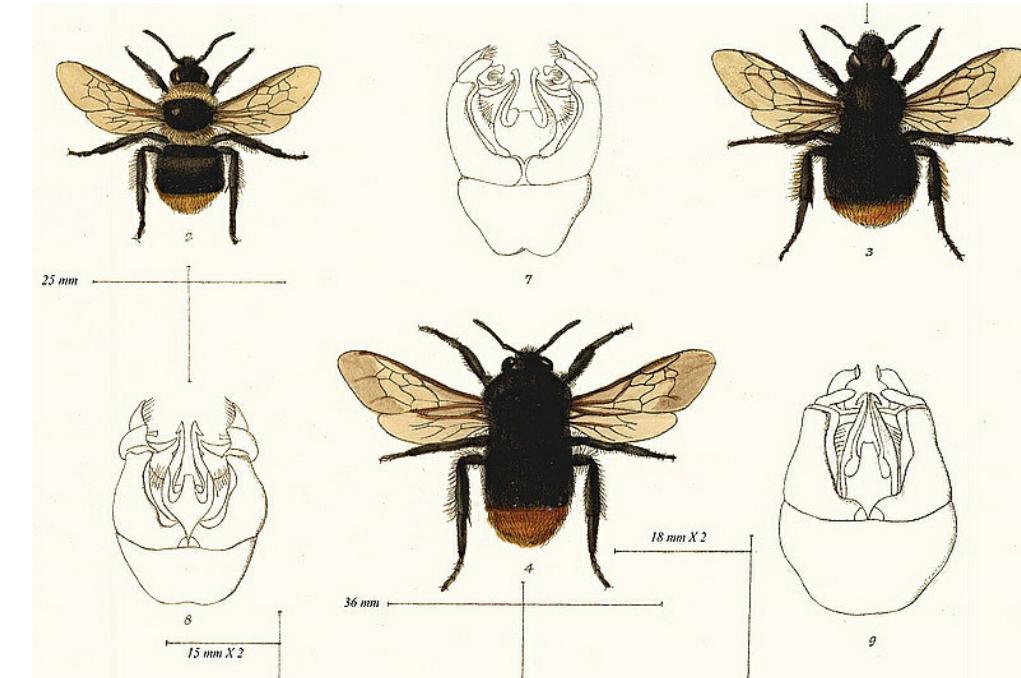


Bartomeus *et al.*, 2017

Les propriétés des organismes



Mesure qualitative ou quantitative prise sur un organisme (i.e. Taille des ailes, Type d'appareil bucal etc.)

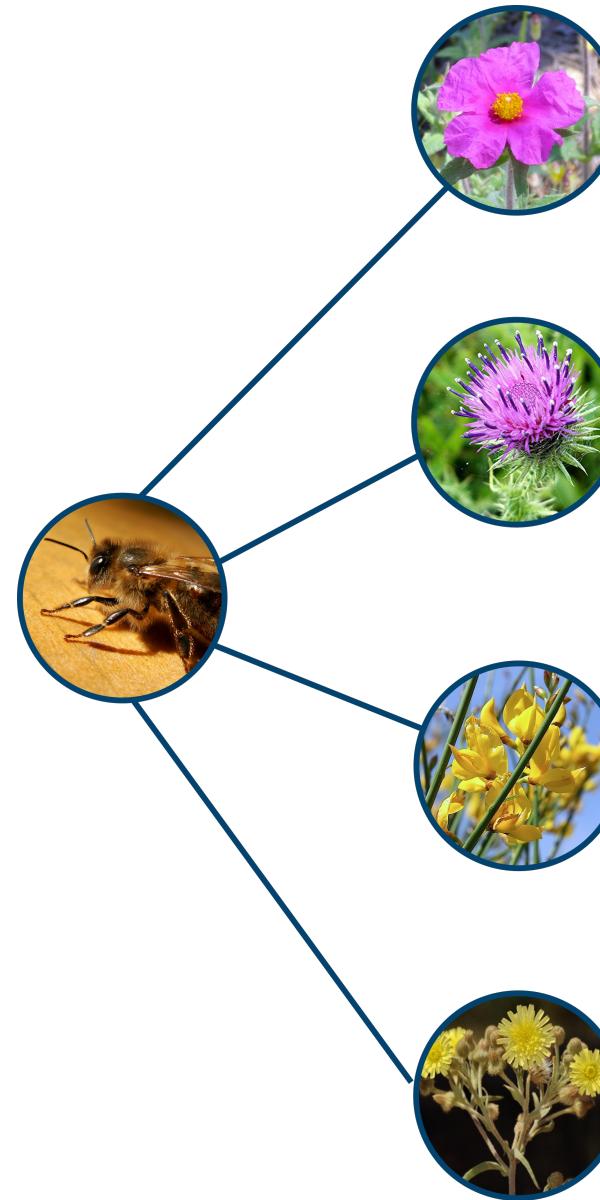


L. Watson & M. J. Dallwitz

Les propriétés des noeuds

Les réseaux écologiques

Réseau plantes-pollinisateur



Bartomeus *et al.*, 2017

Les propriétés des organismes



Co-occurrence et Co-abondance

Les deux organismes doivent être présent sur le site au moment de l'échantillonnage.

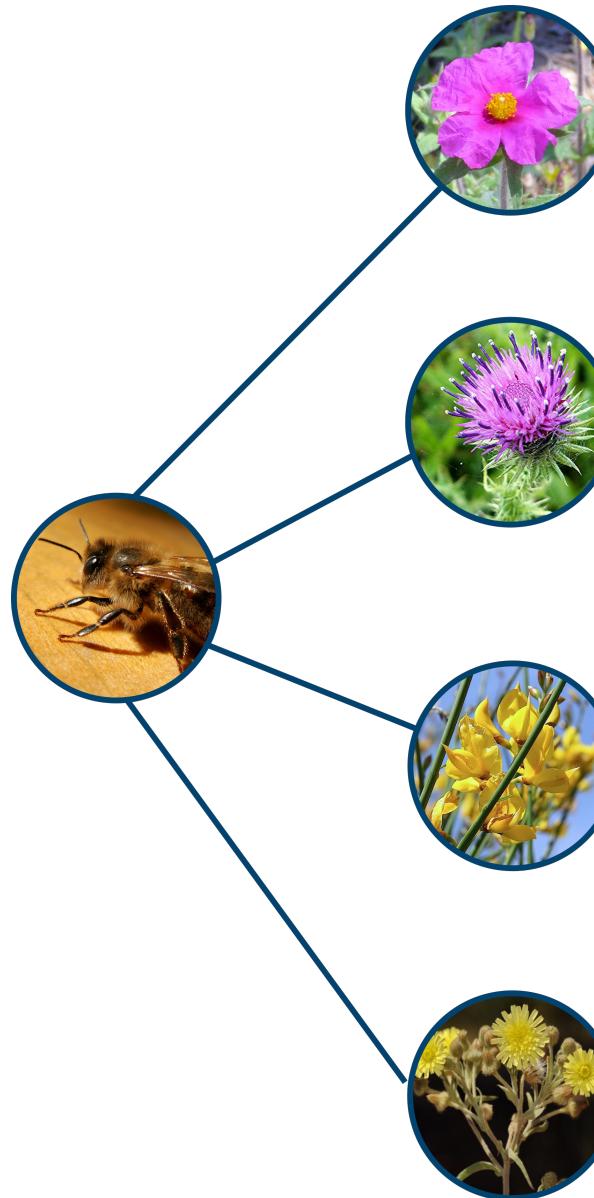


Ramon Nou

Les propriétés des noeuds

Les réseaux écologiques

Réseau plantes-pollinisateur



Les propriétés des organismes



Phylogénie



Traits



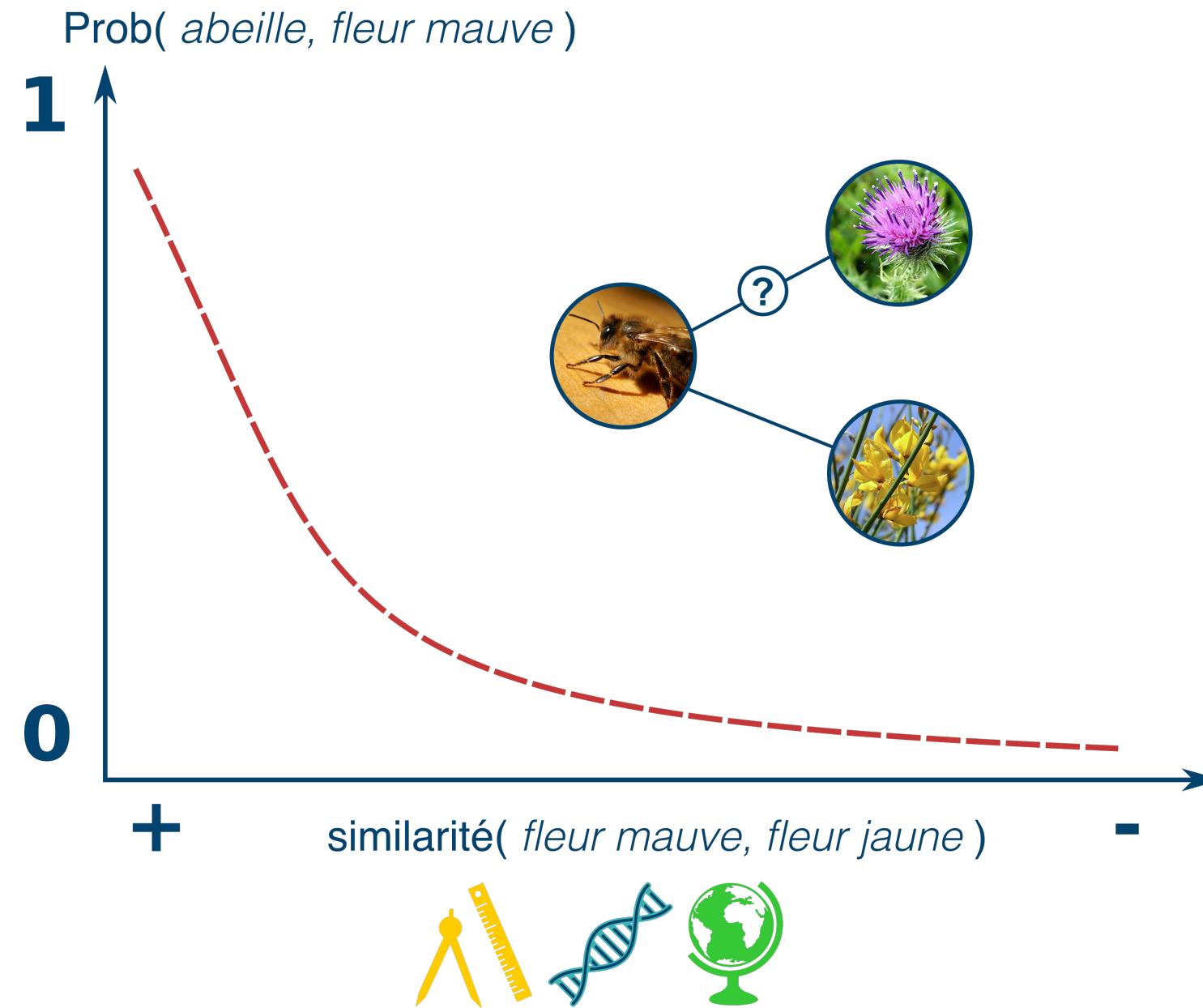
Co-occurrence et Co-abondance

Problématique 1

- Une grande diversité de structure et format de données
- L'information sur certaines espèces peut-être manquante

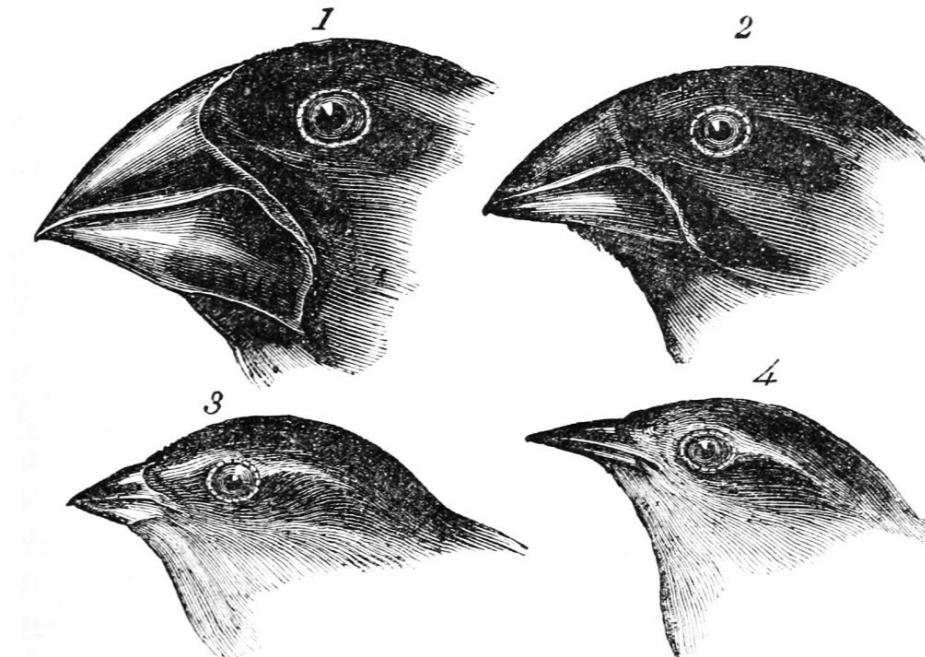
ÉCOLOGIE

Peux-t-on prédire les liens à partir de la similarité entre espèces?



ÉCOLOGIE

Peux-t-on prédire les liens à partir de la similarité entre espèces?



1. *Geospiza magnirostris*.
3. *Geospiza parvula*.

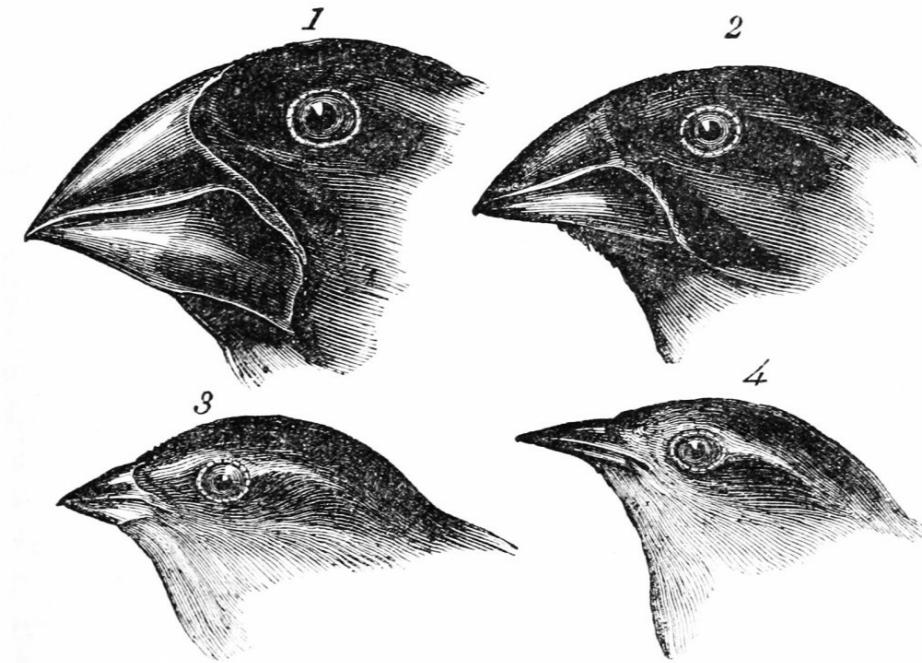
2. *Geospiza fortis*.
4. *Certhidea olivacea*.

Plusieurs méthodes existent en écologie:

- 2 méthodes de Machine learning (*approche informatique*)
- 3 méthodes orientées par les données (*approche bayésienne*)
- 6 méthodes basées sur la correspondance entre traits (*approche fréquentiste*)

ÉCOLOGIE

Peux-t-on prédire les liens à partir de la similarité entre espèces?



1. *Geospiza magnirostris*.
3. *Geospiza parvula*.

2. *Geospiza fortis*.
4. *Certhidea olivacea*.

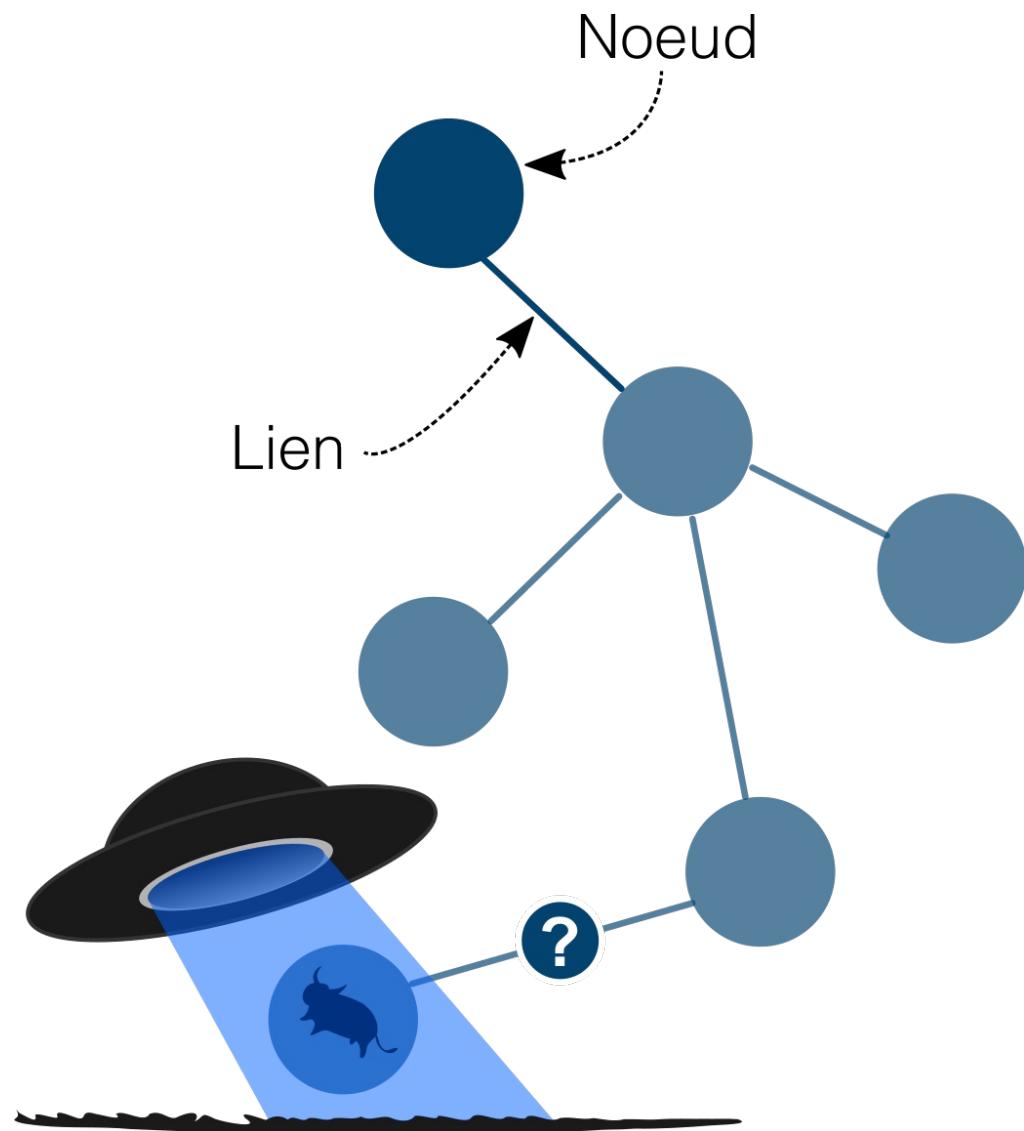
Problématique 2

- Pas toutes implémentées dans R
- Structures de sorties divergeantes, rendant difficile la comparaison méthodologique
- Intensives

Le Projet ALien

Le Projet ALien

Pourquoi alien?



Être en mesure d'évaluer les potentiels liens générés par l'arrivée d'une nouvelle espèce dans un réseau (*Invasive species*).

- Pertinent dans un contexte de changements climatiques où les espèces vont être amenées à se déplacer.
- On parle d'un ré-assemblage des communautés par perte ou gain en espèce

Le Projet ALien...

Une initiative d'un réseau de collaborateurs

Développer un package R permettant de prédire les interactions entre espèces

Les objectifs:

- Rendre les analyses de réseaux davantage reproductible
- Permettre la comparaison méthodologique



Le Projet ALien...

Une initiative d'un réseau de collaborateurs

Développer un package R permettant de prédire les interactions entre espèces

Les défis:

1. Générer un format standard de données propre au package
2. Implémenter les différentes méthodes avec une structure de sortie uniforme
3. Développer des outils d'évaluation/diagnostique des modèles



Le Projet ALien...

Une initiative d'un réseau de collaborateurs

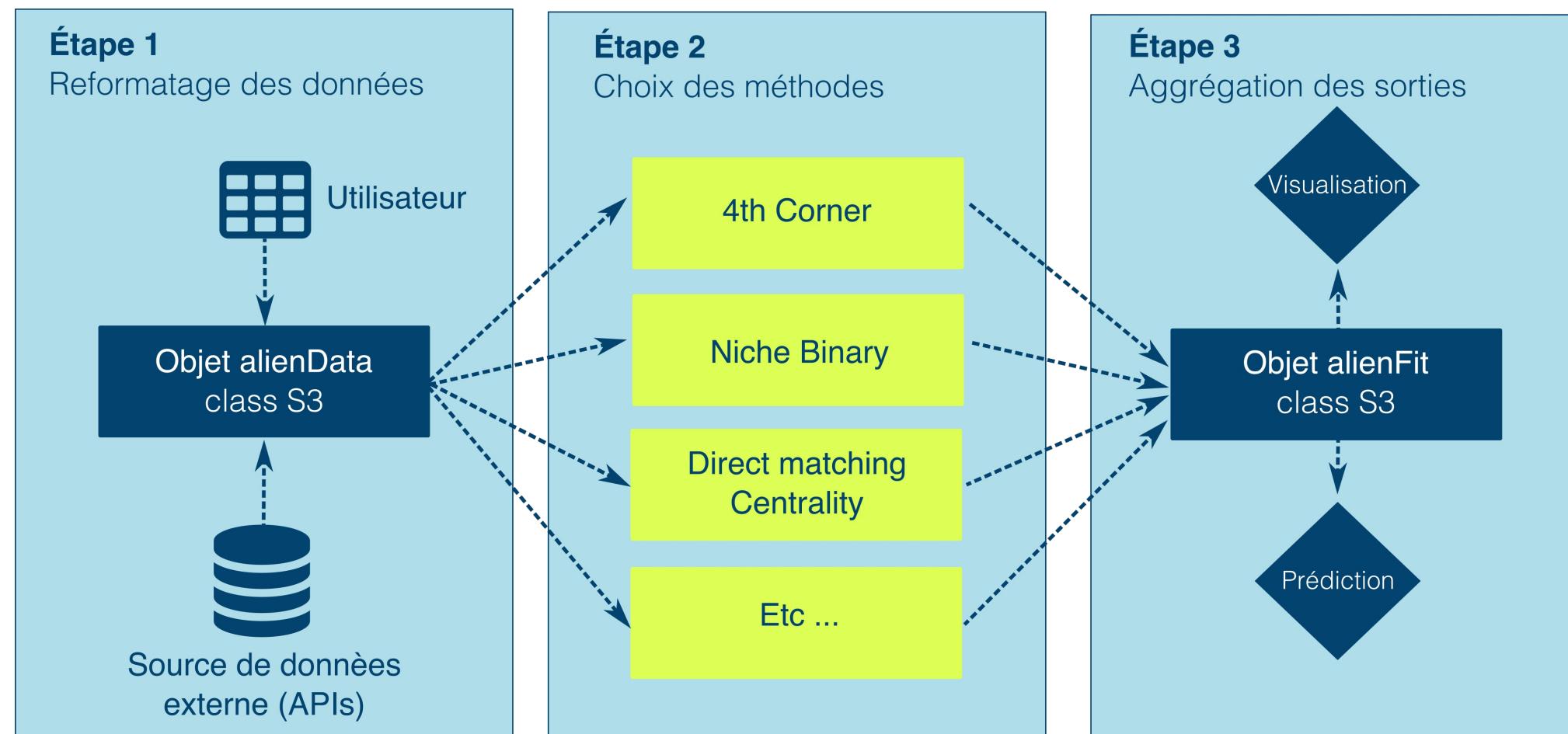
Les bonnes pratiques du développeur:

- Intégration d'un système de contrôle de version (Git)
- Utilisation de tests unitaires pour le contrôle de qualité du code (package `testthat`)
- Construction du package à chaque modifications du code (`codecov`, `appveyor`, `travis CI`)
- Dépot sur GitHub



Le projet ALien...

Le package R



Le Projet ALien...

Étape 1: Formatage des données

Structure de l'objet `alienData`

```
> str(bartomeus)
List of 10
 $ id0bs      : 'data.frame'
 $ interactPair:'data.frame'
 $ interactSp  : int [1:331, 1:331] ...
 $ interactInd : int [1:2938, 1:2938] ...
 $ co0cc       : num [1:331, 1:331] ...
 $ coAbund     : NULL
 $ siteEnv     : NULL
 $ traitSp     : 'data.frame'
 $ traitInd    : 'data.frame'
 $ phy         : NULL
 - attr(*, "co0ccSource")= chr "id0bs"
 - attr(*, "scaleSiteEnv")= logi FALSE
 - attr(*, "scaleTrait")= logi FALSE
 - attr(*, "interceptSiteEnv")= logi FALSE
 - attr(*, "interceptTrait")= logi FALSE
 - attr(*, "class")= chr "alienData"
```

2 jeux de données disponibles:

1. Bartomeus *et al*, 2017

- Données collectées en Espagne (Séville)
- 273 polliniseurs
- 54 plantes
- 2938 individus
- 1469 interactions
- Phylogénie des espèces disponible

Le Projet ALien...

Étape 1: Formatage des données

Structure de l'objet `alienData`

```
> str(bartomeus)
List of 10
 $ id0bs      : 'data.frame'
 $ interactPair:'data.frame'
 $ interactSp  : int [1:331, 1:331] ...
 $ interactInd : int [1:2938, 1:2938] ...
 $ co0cc       : num [1:331, 1:331] ...
 $ coAbund     : NULL
 $ siteEnv     : NULL
 $ traitSp     : 'data.frame'
 $ traitInd    : 'data.frame'
 $ phy         : NULL
 - attr(*, "co0ccSource")= chr "id0bs"
 - attr(*, "scaleSiteEnv")= logi FALSE
 - attr(*, "scaleTrait")= logi FALSE
 - attr(*, "interceptSiteEnv")= logi FALSE
 - attr(*, "interceptTrait")= logi FALSE
 - attr(*, "class")= chr "alienData"
```

2 jeux de données disponibles:

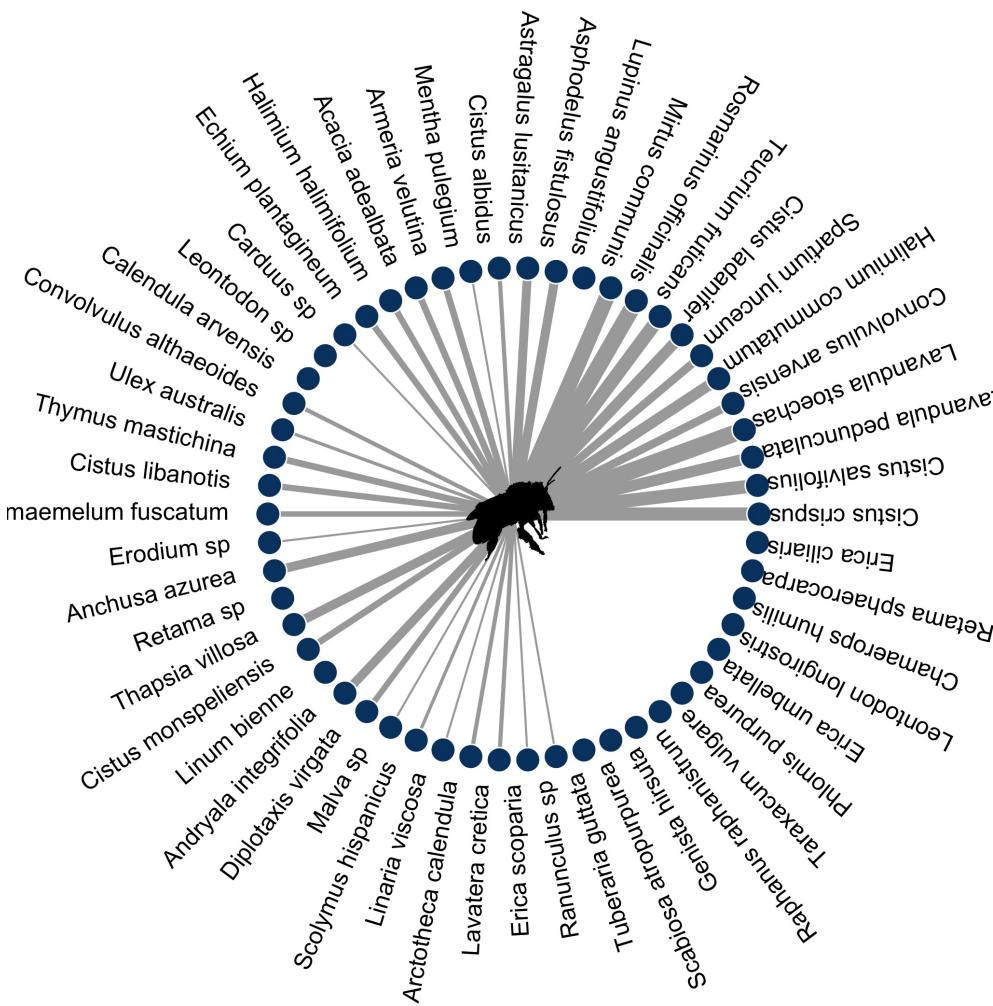
2. Kolpeke *et al*, 2017

- 30 ans de données collectées à travers l'Europe
- 374 sites
- 52 Saules
- 91 parasites du saule
- 125 Parasitoïdes

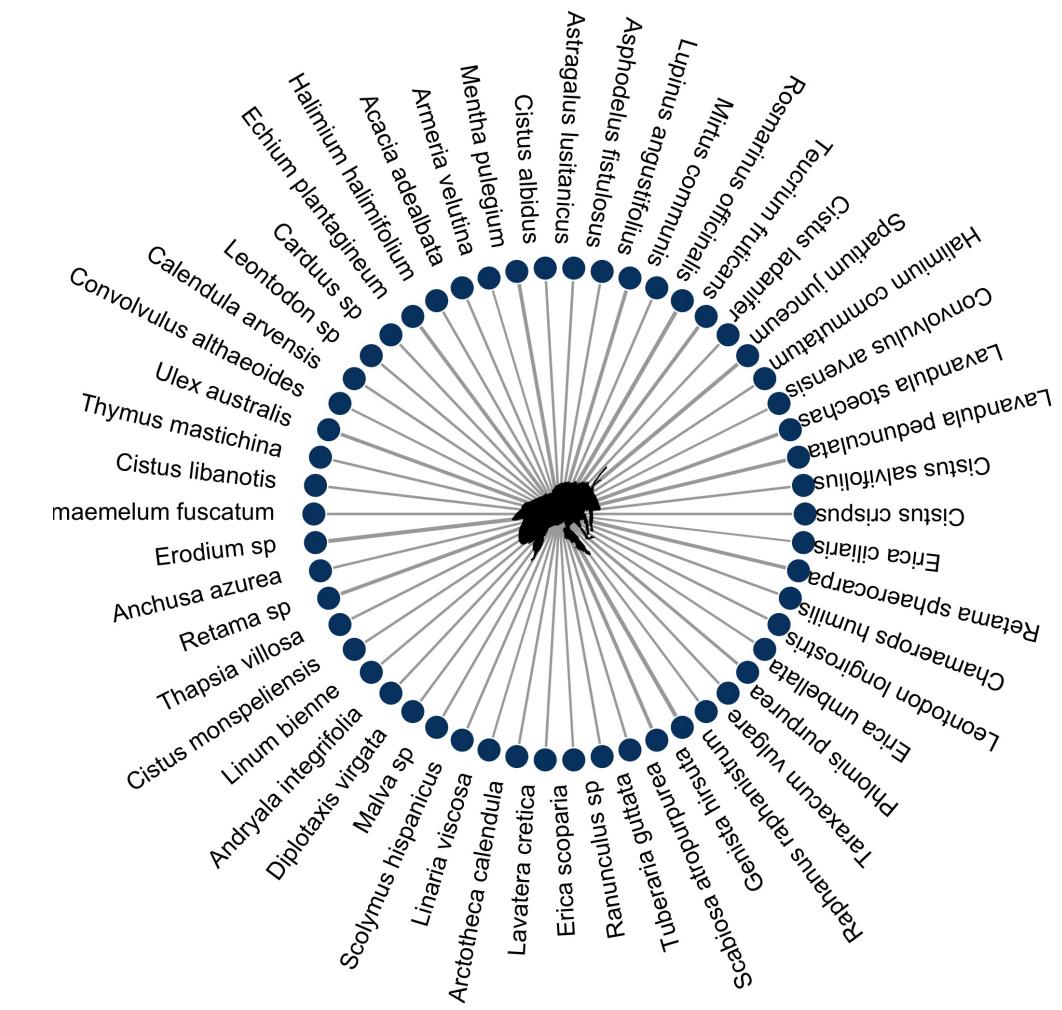
Le Projet Alien...

Exemples de visualisation

GLM avec distribution de poisson (16.3%)

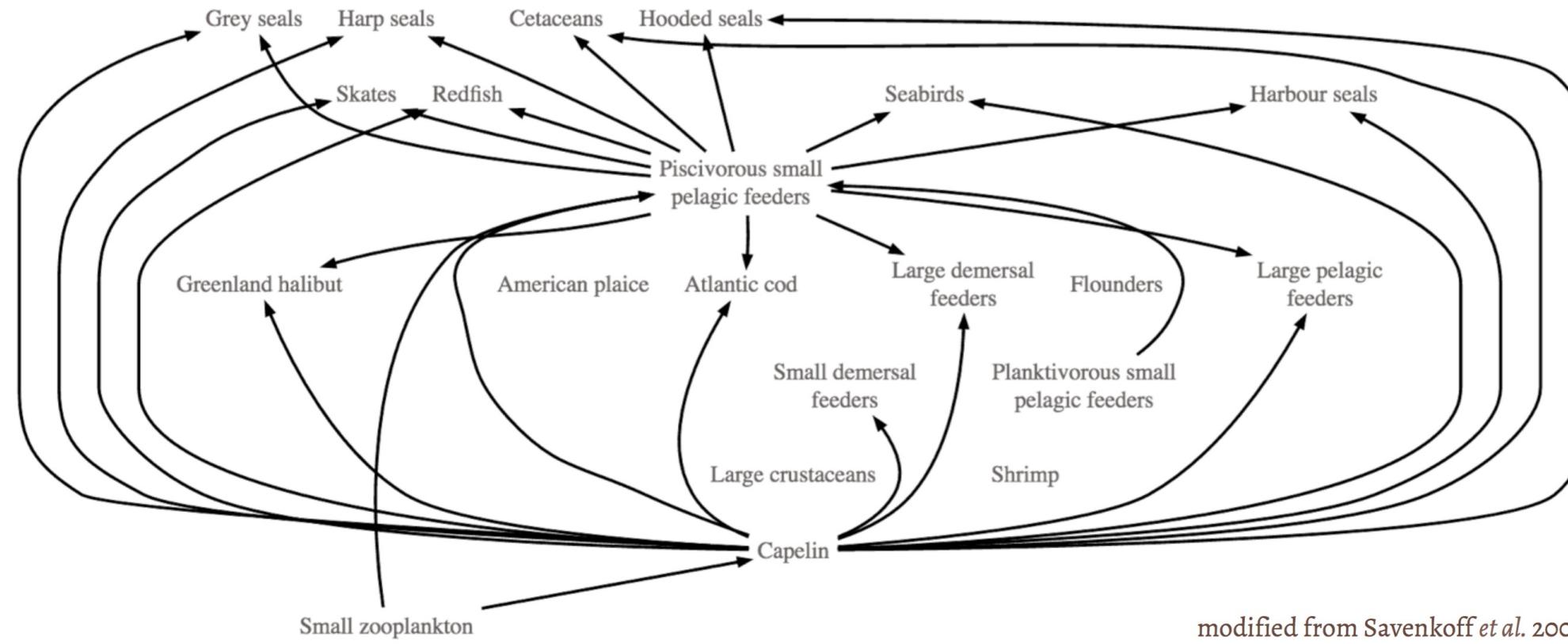


Random Forest (0.7%)



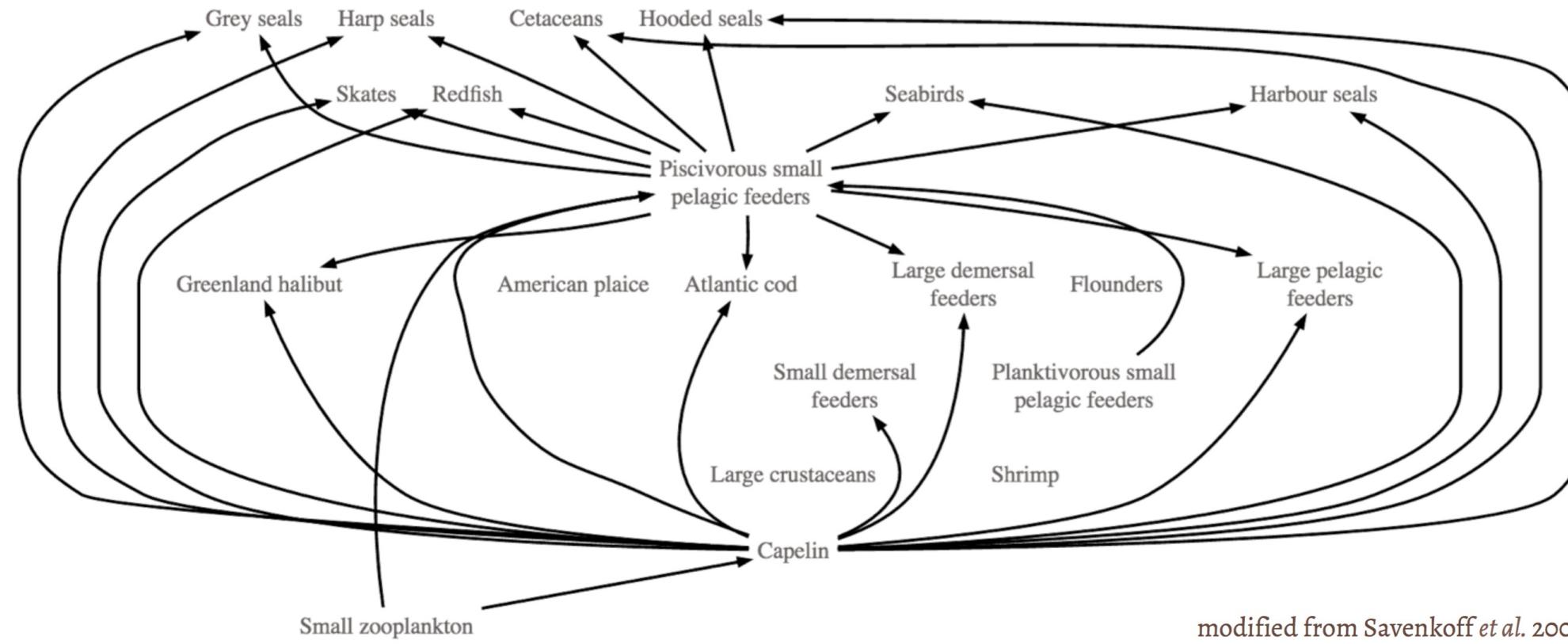
Le projet ALien...

Exemple de résultat: prédire avec Netflix



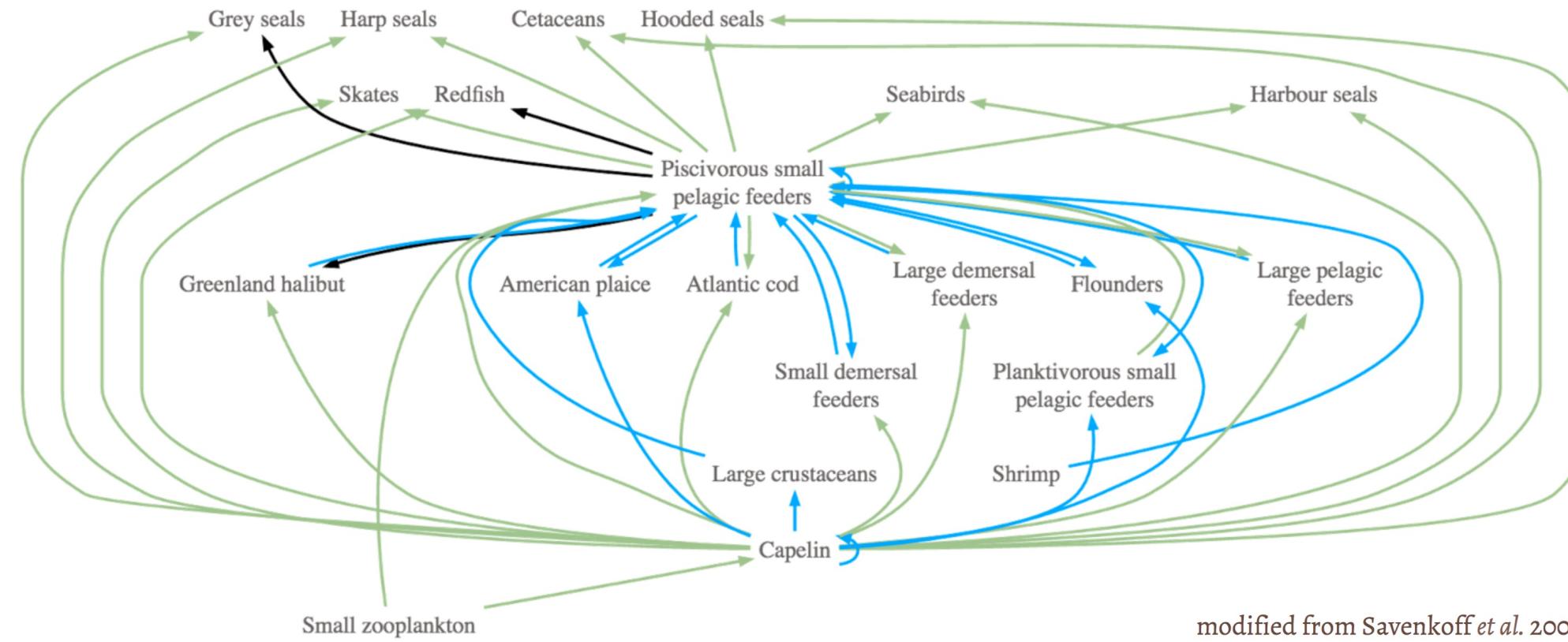
Le projet ALien...

Exemple de résultat: prédire avec Netflix



Le projet ALien...

Exemple de résultat: prédire avec Netflix



Empirical and predicted – Predicted only – Empirical only

Beauchesne *et al.*, 2017.

En résumé

Le projet Alien...

- Un effort de groupe (10 écologistes)
- Package pour prédire les interactions dans un réseau écologique (11 méthodes)
- Permet de reconstruire un réseau suite à l'ajout et/ou retrait d'espèces dans une communauté

Développements futurs

- Décisions à venir: Objet de type S4?
- Intégration des méthodes des 10 méthodes restantes
- Définir la structure de sortie (objet `alienFit`)
- En bref, beaucoup de travail en perspective...

Merci !