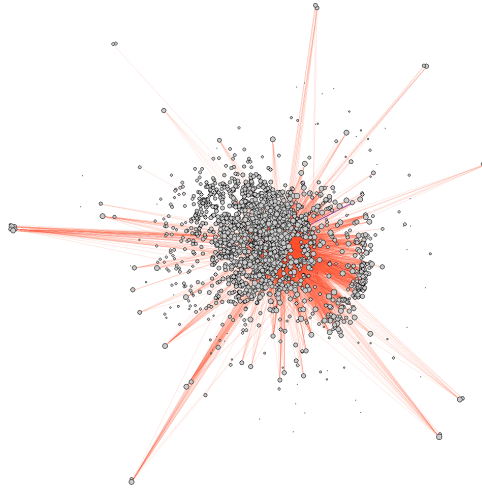

ESTIMER LE PARTAGE VIRAL ET L'IMPORTANCE DES ASSOCIATIONS A TRAVERS LA STRUCTURE GLOBALE DES MÉTA-RÉSEAUX HÔTE VIRUS



Rapport de stage dans le cadre de 1^{er} année de master

Mathis Gheno

Supervisé par :

Timothée Poisot

Université de Rennes

Master 1 mention biodiversité, écologie, évolution (BEE) parcours modélisation en
écologie (MODE)

avril-juin 2023 Soutenus le : 19 juin 2023

Référant universitaire : **Marie-Pierre Etienne**

Remerciement

Je remercie Timothée Poisot de m'avoir accueilli au sein de son laboratoire. Je remercie aussi Gabriel Dansereau et toutes les personnes du *Poisot Lab* avec qui j'ai pu échanger tout au long de mon stage.

Encadrant : Timothée Poisot : Département de Sciences Biologiques, Université de Montréal. timothee.poisot@umontreal.ca

Superviseur : Marie-Pierre Etienne : UP Mathématiques Appliquées - UMR IRMAR marie-pierre.etienne@agrocampus-ouest.fr

Introduction

L'époque géologique actuelle, l'Anthropocène, est synonyme de forte homogénéisation de la composition géographique des espèces (Capinha *et al.*, 2015; Fricke et Svenning, 2020; Helmus *et al.*, 2014; Saladin *et al.*, 2020), y compris virales (Carlson *et al.*, 2022a). Ce brassage global du monde vivant est induit notamment par une migration généralisée des espèces vers des latitudes plus élevées. Ces mouvements sont causés par le réchauffement climatique (Saladin *et al.*, 2020), des déplacements humains importants (Capinha *et al.*, 2015; Helmus *et al.*, 2014) et une destruction soutenue et systémique des habitats (Pimm, 2008). Cela provoque une réduction de la compartimentation des méta-réseaux, qui était préalablement établie par des barrières bio-géographiques de grande échelle (Fricke et Svenning, 2020). En d'autres termes, des espèces ne pouvant pas interagir pour des raisons de séparation géographique se retrouvent aujourd'hui dans les mêmes écosystèmes. Les réseaux d'interactions hôtes-virus ne dérogent pas à la règle et sont, eux aussi, fortement touchés par ces changements (Carlson *et al.*, 2022a; Hoberg et Brooks, 2015; Morales-Castilla *et al.*, 2021). Une majeure partie des maladies humaines sont zoonotiques (Belay *et al.*, 2017). Ce n'est pas anodin, car l'introduction de nouveaux parasites ou d'hôtes vecteurs peut donc avoir des effets considérables sur la santé publique. De plus la diversité des habitat et des espèces qui permettaient une réduction des probabilités de rencontre de pathogène perd de son efficacité à cause de leurs dégradation (White et Razgour, 2020; Pongsiri *et al.*, 2009).

Il y a donc un besoin considérable de comprendre par quels mécanismes certaines espèces vont interagir ou non ? Quelles sont celles qui structurent les réseaux grâce à leurs d'interactions ? Ensuite, quels sont les canaux potentiels de transmissions virales ? Répondre à ces questions permettrait l'anticipation de futures maladies zoonotiques ou des sources d'émergences les plus probables sur lesquels un effort de recherche plus important devra être effectué (Belay *et al.*, 2017; Plowright *et al.*, 2017).

Il a été montré que les réseaux contiennent des informations sur la séparation géographique (Fricke et Svenning, 2020), la distance phylogénétique (Rohr et Bascompte, 2014) et les méta-réseaux apportent une approche probabiliste de réalisation des interactions (Strydom *et al.*, 2022). S'ils sont bien construits, ces objets renferment de nombreuses informa-

tions écologiques, spatiales, historiques et phylogénétiques. Cependant, ils sont généralement incomplets. Une estimation montre que 1.67 million des virus infectant les mammifères sont inconnus (Carroll *et al.*, 2018). Cependant, une étude plus récente montre que cette diversité manquante est une surestimation d'ordre 10^2 (Carlson *et al.*, 2022b) ce qui reste tout de même une part d'ombre considérable. Il est estimé que seulement 1% des interactions hôtes virus au sein des mammifères ont été observées (Carlson *et al.*, 2022b). De plus, ces observations sont soumises à des biais d'échantillonnages. Ces biais sont fonctions des hôtes ou virus d'intérêt, de leurs potentiels zoonotiques, du nombre de citations dans la littérature scientifique, de leurs localisations géographiques et taxonomiques (Albery *et al.*, 2021; Mollentze et Streicker, 2020; Poisot *et al.*, 2023). De nouveaux modèles de prédiction très performants utilisant des méthodes de *machin learning* (Pichler *et al.*, 2020) ou de *graph embedding* (Strydom *et al.*, 2022) on fait leurs apparitions, ces méthodes semblent très robustes aux biais cités précédemment (Poisot *et al.*, 2023). Ainsi — mis à part un manque de connaissances sur la diversité viral — l'utilisation de réseaux d'interactions prédits pourrait être une bonne solution pour anticiper l'émergence de maladie zoonotique (Poisot *et al.*, 2023), ou d'estimer le partage viral entre hôtes.

Les causes de transmission de maladies zoonotiques sont multiniveaux (Plowright *et al.*, 2017), lutter efficacement contre leurs apparitions requiert donc une approche transdisciplinaire (Belay *et al.*, 2017; Plowright *et al.*, 2017). L'écologie peut apporter des démarches prédictives. Par exemple, Albery *et al.* (2020) proposent une méthode de prédiction de partage viral entre paires d'hôtes se basant sur leurs distances géographiques et phylogénétiques. Carlson *et al.* (2022a) ajoute à cette méthode une dimension spatiale en utilisant des modèles de distribution d'espèces. Leurs résultats de prédiction sont très encourageants. Cependant, les distances géographique et phylogénétique ne sont pas les seules caractéristiques qui définissent le risque de transmission entre hôtes. Au niveau écologique, on retrouve aussi la distribution et la densité des réservoirs d'hôtes ainsi que les routes d'exposition (Plowright *et al.*, 2017). Il semble donc implorant de prendre en compte le réseau dans sa globalité pour mieux quantifier cette transmission.

Ainsi, grâce à une mesure de communicabilité entre les nœuds d'un réseau introduite par Estrada et Hatano (2008), nous proposons une nouvelle approche prenant en compte l'ensemble de la structure des réseaux pour quantifier l'influence des associations et des

paires d'espèces, ainsi que des prédictions de partage viral potentiel. Une approche similaire a déjà été explorée avec l'utilisation de la centralité (Gómez *et al.*, 2013), mais cela ne prend en considération seulement le chemin le plus court entre deux espèces. En utilisant cette méthode, nous comparons l'influence des associations d'un réseau observé et de son méta-réseau prédit associé (Poisot *et al.*, 2023). Étant donné que le réseau prédit est plus dense en lien et est beaucoup moins biaisé autour de certaines espèces, notre première hypothèse est que l'influence des associations et des paires d'espèces de ce réseau sera plus homogène et plus faible que le réseau observé. La mesure de partage viral potentiel a été appliquée sur le méta-réseau prédit et compare les potentiels intra- et extra-ordre des 10 ordres d'hôtes majoritaire. Notre seconde hypothèse est que les ordres très représentés dans le réseau auront un partage viral intra-ordre plus important que les ordres moins représentés. Cette supposition équivaut à faire l'hypothèse nulle que les liens sont répartis aléatoirement dans le réseau et donc que leur partage viral potentiel dépend seulement du nombre d'espèces des ordres. Notre troisième hypothèse est que le partage viral potentiel intra-ordre sera toujours plus élevé que le partage viral potentiel extra-ordre car les hôtes similaires devraient partager les mêmes virus.

Méthode

Obtentions des données

Les données utilisées proviennent de 2 méta-réseaux. Le premier est Clover, composé de 5 477 associations mammifère-virus observé, il est tiré de multiples bases de données (Gibb *et al.*, 2021). Le second est Trefle, un réseau comprenant les mêmes espèces que Clover, mais avec 75 901 associations supplémentaires, celles-ci ont été prédites par des méthodes de réductions de dimension de réseau (Poisot *et al.*, 2023).

Mesures d'influence des paires et interactions

La quantification d'influence des paires d'espèces utilise la communicabilité entre les nœuds du réseau. Cette mesure a été introduite par Estrada et Hatano (2008). Elle se base sur la caractéristique des puissances de matrices d'adjacences qui permet d'obtenir pour

une puissance k , le nombre de chemin possible en k étapes entre chaque paire de nœuds. En ajoutant chaque k puissances et en pénalisant les chemins de grandes tailles, ils arrivent à l'expression suivante :

$$G = \sum_{k=0}^{\infty} \frac{(\mathbf{A}^k)}{k!} = (e^{\mathbf{A}}) \quad (1)$$

avec A la matrice d'adjacence, k le nombre d'étapes et G la matrice de communicabilité. La pénalisation avec des factorielles de k permet de définir G comme l'exponentiel de A , ce qui leur permet ensuite de déduire la forme spectral suivante :

$$G = \sum_{j=1}^n \varphi_j \varphi_j^T e^{\lambda_j} \quad (2)$$

avec φ_j les vecteurs propres associés aux valeurs propres λ_j de la matrice A . C'est à partir de l'équation 2 que la communicabilité a été calculé. La matrice G est une quantification de l'influence des paires d'espèces pour le réseau (voir encadré 1), ainsi si chaque espèce dispose d'au moins une association dans le réseau la matrice G ne présentera que des valeurs strictement supérieur à 0. Pour cette partie les seules valeurs utilisées de la matrice G sont celles des couple hôtes (h) virus (v) ayant une association réalisée dans Trefle et dans Clover (Donc l'ensemble des associations observé hv_{obs}).

La différence d'ordres de grandeurs entre la communicabilité des deux réseaux est de 10^{50} , Trefle ayant des valeurs plus élevé car plus dense en lien. Les matrices G ont donc été normalisées par rapport à leurs valeurs minimum et maximum entre 0 et 1. Cette normalisation laisse les valeurs de G d'un réseau aux mêmes rangs et rend possibles les comparaisons de paires d'espèces présentes dans des réseaux ayant des structures et des dimensions différentes.

La quantification d'importance des associations (I) a été calculée à partir des valeurs de communicabilité normalisées :

$$\widetilde{I}_{hv} = \widetilde{G}_{hv} - \zeta_{hv} \widetilde{G}_{hv} \quad (3)$$

où $\zeta \widetilde{G}_{hv}$ est la communicabilité normalisée entre h et v sans l'association hv et \widetilde{I}_{hv} est l'importance normalisée de l'association hv .

Trefle et Clover en tant que réseaux d'interaction bipartie, sont caractérisés par des

matrices d'incidence potentiellement non carrée et non-symétrique, rendant le calcul de G impossible. Ce problème est simplement résolu par la création de matrices d'adjacences avec hôtes et virus présent à la fois sur les lignes et colonnes.

Partage viral potentiel à l'échelle de l'ordre

Pour chacune des 81 378 associations de Trefle une matrice ζG a été calculée. Deux groupes ont été distingués : l'ordre de l'hôte auquel il a été retiré une interaction (intra-ordre) et l'ensemble des autres ordres (extra-ordre). Intra- et extra- dénote donc ici la position de l'association retirée lors du calcul de ζG . Une comparaison a ensuite été effectuée sur la partie hôte-hôte de la matrice G via la formule suivante :

$$\mu_{\Delta G_{grp}} = \overline{G_{grp} - \zeta_{hv} G_{grp}} \quad (4)$$

ΔG_{grp} est la différence entre les valeurs de la matrice G_{grp} et ζG_{grp} , μ est la moyenne (par groupe) de l'ensemble des différences. Cette moyenne est forcément positive, car la communicabilité ne peut que diminuer lorsque l'on retire un lien. Un z-score a été calculé pour chaque $\mu_{\Delta G_{grp}}$ avec comme points de référence l'impact de la modification sur l'ensemble de la partie hôte-hôte :

$$zscore_{grp} = \frac{\mu_{\Delta G_{grp}} - \mu_{\Delta G_h}}{\sigma_{\Delta G_h}} \quad (5)$$

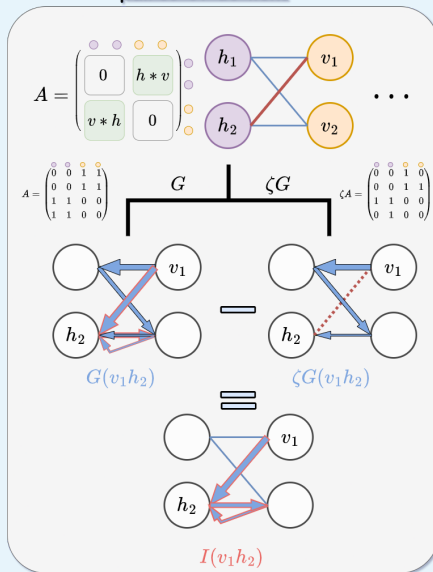
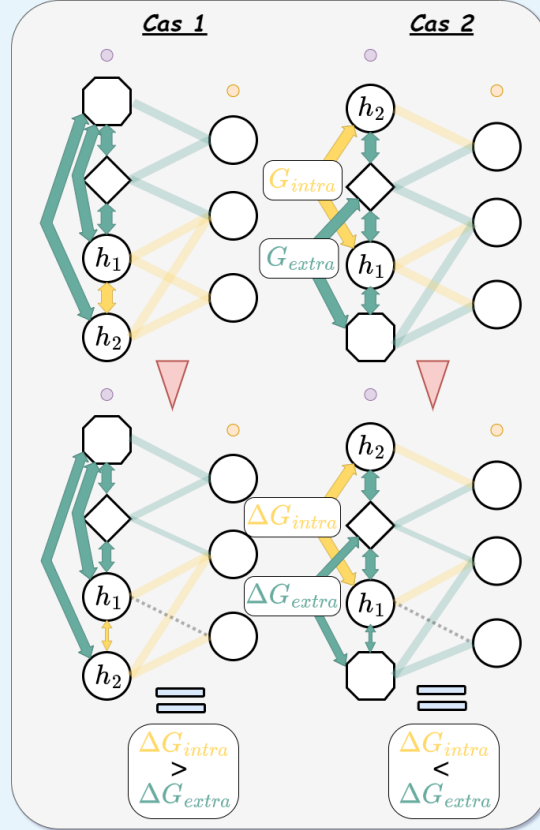
$\mu_{\Delta G_h}$ est donc une version de l'eq. 4 avec pour groupe l'ensemble de la partie hôte-hôte, $\sigma_{\Delta G_{grp}}$ est l'écart-type de $(G_{grp} - \zeta_{hv} G_{grp})$

Modèles

Un GAM a été utilisé pour mettre en relation l'importance des associations de Clover et de Trefle. Pour le potentiel viral, un glm avec un terme multiplicatif entre ordre et position de retrait du lien de a été utilisé. Les calculs et analyses, on été effectués sous R 3.1.2. .

Encadré 1 – Détails schématiques de la communicabilité, de l'importance et du partage viral potentiel.

En mettant de côté pour un instant l'interprétation biologique et en se plaçant sur une interprétation plus schématique et simplifiée de graphes, on peut expliquer pourquoi la communicabilité semble être un bon outil pour exprimer l'influence des liens ou des nœuds dans un réseau. La boîte du bas présente le procédé suivi. Si deux nœuds ont une communicabilité élevée, cela indique que beaucoup de chemins courts les relient. Ils ont donc des interactions directes et intermédiaires avec un grand nombre de nœuds du réseau. Cette idée est représentée par le graphe $G(v_1, h_2)$. Le nombre d'intermédiaires maximal est de 3 et certains chemins de communications ne sont pas représentés pour des questions de visibilité (valable pour tout l'encadrer). Les flèches contourées en rouge sont issues de chemins empruntant le lien $h_2 v_1$. On notera que c'est une approximation de l'effet de la paire sur l'ensemble du réseau, car pour être comptabilisé un chemin doit forcément "revenir" à un des deux nœuds. Afin d'avoir la valeur de communicabilité qui passe par un lien, nous calculons l'ensemble des chemins qui ne passe pas par ce lien en le supprimant de A lors du calcul de ζG . On effectue ensuite une soustraction, qui nous donne notre valeur de communicabilité du lien, l'importance I . Pour revenir à une interprétation plus biologique, un chemin peut représenter au sein d'un réseau, une cascade d'interaction qui s'atténue à chaque intermédiaire (ex. Épidémie Hôte A par virus 1, diminution de Hôte A, diminution de virus 2 de Hôte A, augmentation de Hôtes B car moins virus 2, ...Etc).

Influence des paires et de liens

Partage viral potentiel


L'encadrer du haut montre notre approche pour le partage viral. Nous nous intéressons pour cette partie à l'impact qu'a eu le retrait sur des groupes d'intérêt biologique. Les formes indiquent l'ordre auquel les hôtes appartiennent, les liens jaunes sont dits intra-ordre, car ils comprennent un hôte appartenant au même ordre que celui avec le lien retiré. Et inversement pour les liens verts qui sont dits extra-ordre. Les flèches indiquent la communicabilité entre les paires d'hôtes. Pour les deux réseaux du bas, les changements d'épaisseur des flèches et des liens montrent comment la communicabilité entre deux espèces a été impactée par le retrait de l'association en pointillé noir. Le cas 1 est un exemple de partage intra-ordre supérieur à extra-ordre. Le cas 2 est un exemple de partage extra-ordre supérieur à intra-ordre. Nous avons nommé cette mesure partage viral potentiel, car malgré le fait qu'elle quantifie principalement des partages effectifs, elle prend aussi en compte des partages non réalisés avec des hôtes et virus intermédiaires. De plus, la métrique se base sur un réseau prédit.

Résultats

Mesures d'influences

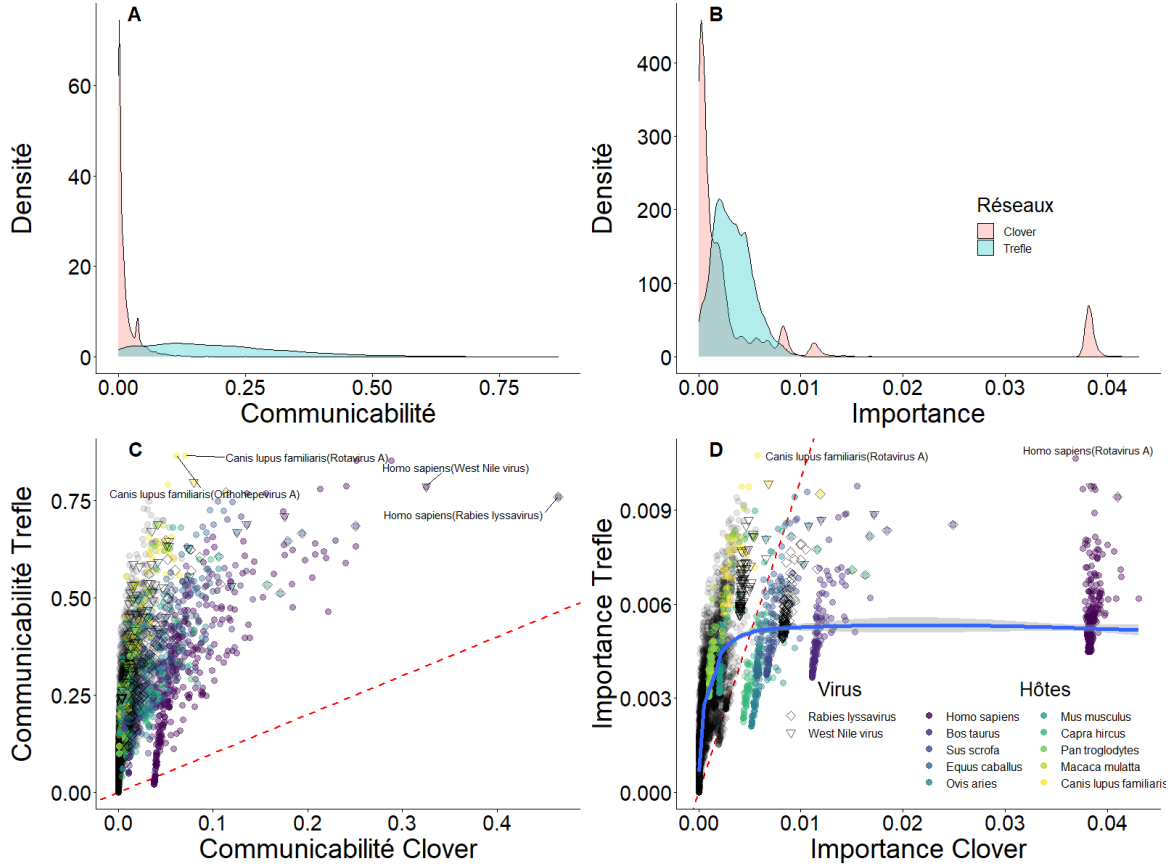


FIGURE 1 – A : Distributions des valeurs de communicabilité pour chaque association observées. B : Distributions des valeurs d'importance pour chaque association observées. C : Comparaison des valeurs de communicabilité entre Clover et Trefle pour chaque association observées. D : comparaison des valeurs d'importance entre Clover et Trefle pour chaque association observées. La légende représente les hôtes et virus ayant le plus d'associations. Hôtes et virus sont ordonnés (séparément) de haut en bas puis de droite à gauche par rapport au nombre d'association dont ils disposent dans Clover. Les lignes en pointillé rouge sont les droites 1 : 1. La courbe bleu représente la régression associée au GAM.

Clover et Trefle ont respectivement une communicabilité relative moyenne de 0.02 et 0.2, un écart-type de 0.03 et 0.1, une erreur standard de 10^{-5} et 10^{-4} . La \widetilde{G}_{hv} de Trefle est donc plus homogène et plus grand que la \widetilde{G}_{hv} de Clover (fig. 1 A). Cependant quelques associations de *Homo sapiens* diminuent dans Trefle par rapport à Clover (fig. 1 C). Sur 5 494 associations, 178 ont une \widetilde{G}_{hv} dans Trefle inférieure à la \widetilde{G}_{hv} de Clover et 149 sont des associations comprenant *Homo sapiens*. 44% des paires d'espèces d' *Homo sapiens* ont subi une réduction de leurs influences relatives dans Trefle. Il reste tout de même des paires d'espèces dont la communicabilité est forte dans les deux réseaux, c'est le cas pour

TABEAU 1 – Nombre d'association et moyenne des z-score intra- et extra- pour les 10 ordres disposant de plus d'associations dans le réseaux Trefle. Les valeurs entres parenthésés représentent les écarts types.

ordres	N	Moyennes intra-	Moyennes extra-
<i>Rodentia</i>	19074	-0.12 (± 0.03)	0.05 (± 0.01)
<i>Primates</i>	14637	0.37 (± 0.04)	-0.05 (± 0.01)
<i>Cetartiodactyla</i>	13374	0.16 (± 0.03)	-0.03 (± 0.01)
<i>Chiroptera</i>	13083	-0.17 (± 0.05)	0.05 (± 0.02)
<i>Carnivora</i>	11623	0.31 (± 0.03)	-0.04 (± 0.01)
<i>Perissodactyla</i>	2012	2.9 (± 0.4)	-0.02 (± 0.003)
<i>Lagomorpha</i>	1761	1.5 (± 0.3)	-0.02 (± 0.004)
<i>Diprotodontia</i>	1383	0.3 (± 0.2)	-0.01 (± 0.01)
<i>Didelphimorphia</i>	1156	1.7 (± 0.5)	-0.02 (± 0.01)
<i>Eulipotyphla</i>	923	1.2 (± 0.4)	-0.014 (± 0.004)

Homo sapiens x *Rabies lyssavirus* et *Homo sapiens* x *West Nile virus*. D'autre vois leurs influences drastiquement augmentées comme *Canis lupus familiaris* x *Rotavirus A* et *Canis lupus familiaris* x *Orthohepevirus A*.

Clover et Trefle ont respectivement une importance relative moyenne de 0.005 et 0.003, un écart-type de 0.010 et 0.002, et une erreur standard de 10^{-5} et 10^{-6} . L'importance est plus homogène dans Trefle que dans Clover (fig. 1 B). Certaines espèces d'hôtes ont des associations très importantes dans Clover qui sont limité dans Trefle (ex : *Homo sapiens*, *Bos taurus*)(fig. 1 B et D). Ce sont notamment les 6 hôtes ayant le plus d'associations dans Clover qui voit leurs importances relatives diminuées dans Trefle (*Homo sapiens*, n = 402 ; *Bos taurus*, n = 141 ; *Sus scrofa*, n = 117 ; *Equus caballus*, n = 97 ; *Ovis aries*, n = 83 ; *Mus musculus*, n = 74) ainsi que d'une espèce de virus (*Rabies lyssavirus*, n = 210)(fig. 1 D). L'importance des associations dans Clover est très dépendante des l'espèce avec un grand nombre de lien. Ce regroupement par espèces est très fortement limité dans Trefle. Le modelé additif généralisé comprend un lissage cubique. Ce terme de lissage présente les caractéristiques suivantes : edf >8, Ref.df = 9, F = 1038, p-valeur<0.001. Le modèle au global explique 63% de la variance (GVC< 10^{-6} , esti.echelle< 10^{-6}).

Pour l'ensemble des 10 ordres présentés sur la figure 2 les *Rodentia*, les *Chiroptera*, les *Primates*, les *Cetartiodactyla*, les *Carnivora* et les *Diprotodontia* ont un partage viral potentiel qui différent du global d'un ordre de grandeur de 10^{-2} . Ils ont un z-score respectif de -0.12, -0.17, 0.37, 0.16, 0.31, 0.3 (tab. 1). Les *Rodentia*, les *Chiroptera* sont les seuls groupes à avoir un partage viral intra-ordre plus important qu'extra-ordre, ainsi que d'un z-score intra- négatif (tab. 1, fig. 2). Les *Perissodactyla*, les *Lagomorpha*, les *Didelphimorphia* et les *Eulipotyphla* ont un partage viral intra-ordre plus important par rapport au global d'un

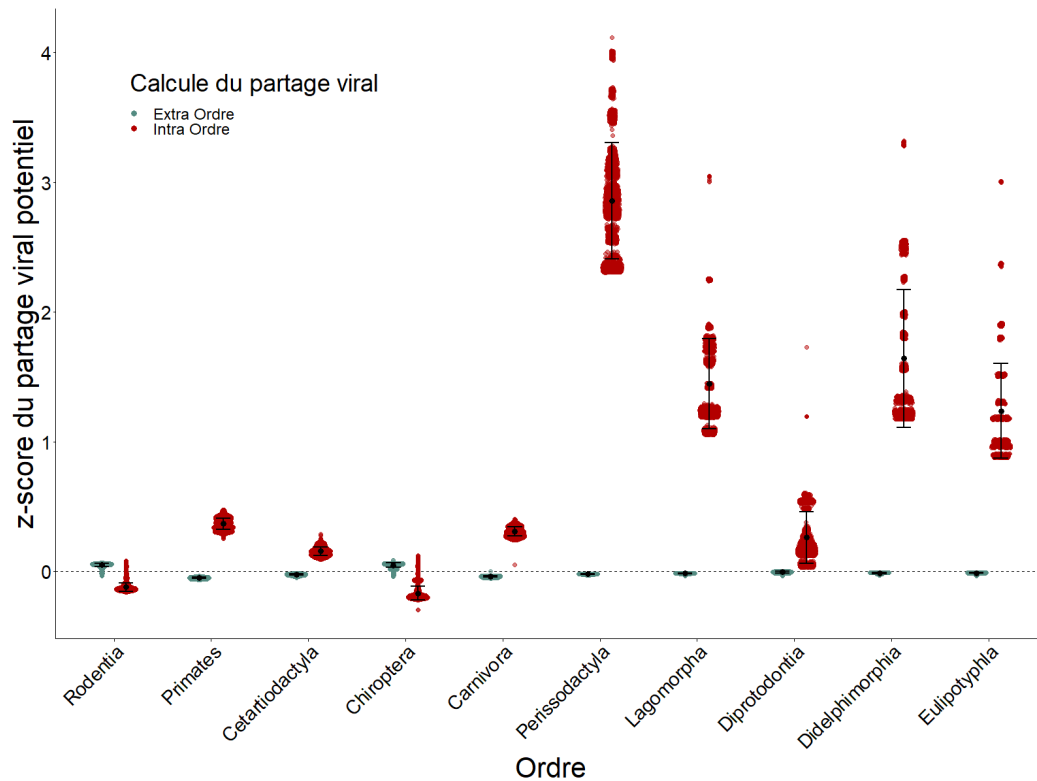


FIGURE 2 – Z-score de partage viral potentiel entre hôtes compris ou non dans le même ordre. Chaque point (duplicé entre intra- et extra-) représente une des 81 395 associations de Trefle. Les barres d'erreurs représentent la moyenne et l'écart-type. Les points sont agencés en fonction de leur courbe de densité. Les échelles de densité ne sont pas identiques entre les ordres.

ordre de grandeur de 10^{-1} . Ils ont un z-score respectif de 2.9, 1.5, 1.6 et 1.2 (tab. 1). Les z-scores de partage viral extra-ordre sont tous très proche du partage viral global (tab. 1, fig. 2). Les résultats du glm ne sont pas présentés, car l'ensemble des facteurs "ordre" et "position de l'association retirée", ont des p-values très significative, c'est cependant une conséquence du nombre considérable d'observation, cela n'indique pas forcément un effet biologique réel. Le glm n'apporte donc pas d'informations supplémentaires aux valeurs du tab. 1.

Discussions

Les prédictions augmentent l'influence des paires

La communicabilité et l'importance comme montrées dans l'encadré 1 sont respectivement des quantifications de l'influence d'une paire d'espèces ou d'un lien sur le réseau. Nous avons trouvé que l'influence des paires d'espèces est mieux réparti dans Trefle que

dans Clover, mais aussi que ces influences (relatives aux réseaux) sont plus importante dans Trefle. Ces observations sont causées par la densité de lien importante du réseau prédit. Chaque paire d'espèces présente beaucoup plus d'association. Il est donc plus simple pour ces paires d'impacter le réseau, car l'ensemble est mieux connecté. Cette observation n'est donc pas en accord avec une partie de notre hypothèse 1 qui supposais que les pair d'espèces allaient avoir une influence moindre dans Trefle par rapport à Clover. Avec la communicabilité, on distingue légèrement le biais d'échantillonnage de Clover qui est surtout centré sur *Homo sapiens*, cet effet disparaît totalement dans Trefle.

Il n'est pas surprenant que les paires d'espèces (ex : *Homo sapiens* x *Rabies lyssavirus* et *Homo sapiens* x *West Nile virus*) qui sont très influentes dans le réseau Clover le reste dans Trefle, car ce dernier est construit sur la base de Clover avec des associations supplémentaire (Poisot *et al.*, 2023), les espèces hautement connectées dans Clover le reste donc dans Trefle. Les paires d'espèces qui méritent une attention plus grande sont celles qui sont peu influentes dans Clover, mais beaucoup plus dans Trefle car ce sont celles qui ont été le plus réintégrer au méta-réseau par les prédictions des nouvelles interactions. (ex : *Canis lupus familiaris* x *Rotavirus A* et *Canis lupus familiaris* x *Orthohepevirus A* ; qui augmentes d'un facteur 10).

Les prédictions augmentent l'influence des associations

L'hypothèse que l'importance des associations diminue dans Trefle est valide, en moyenne elles diminuent très légèrement. Cette diminution est principalement causée par *Homo sapiens* qui voit l'importance de ses associations fortement réduite. L'importance des liens de Clover est très fortement contrôlée par les espèces d'hôtes, mais aussi par quelques virus. Cette mesure d'importance met donc très bien en avant les biais d'échantillonnage de Clover et montre que dans Trefle cet effet espèces est fortement atténué. Effectivement dans le méta-réseau prédit, l'influence d'une interaction est beaucoup moins dépendante des espèces qui la compose. La relation cubique entre les importances de Clover et de Trefle montre que les associations prédites ont fortement pénalisé les interactions de grande importance. Ces résultats sont cohérents avec les objectifs des prédictions effectuées par Poisot *et al.* (2023). Identifié les interactions de faible influence dans Clover et de forte influence dans Trefle

pourrais peut-être permettre d'identifier des canaux de transmission viraux potentiels peu étudiés. Cela pourrait surtout être intéressant de croiser ces valeurs avec la métrique de partage viral afin d'identifier avec quels groupes d'espèces ces interactions ont un potentiel de partage viral élevé.

Partage viral potentiel

Le partage viral potentiel va à l'encontre de notre hypothèse 2 qui était que les espèces très représentées auraient un partage viral intra-ordre plus élevé que les espèces moins représentées. C'est d'autant plus surprenant que parmi les 5 groupes avec le plus d'espèces, 2 d'entre eux (les *Rodentia* et les *Chiroptera*) ont un partage viral intra-ordre inférieur au partage viral potentiel extra-ordre. Cependant, leur z-score montre que l'effet ne dépasse pas l'intervalle d'un écart-type du partage viral global. Ces deux groupes ont donc un potentiel de partage viral intra-ordre plus faible que les autres ordres et un partage viral extra-ordre plus fort, mais ces effets ne diffèrent que très faiblement du potentiel global. Nous observons donc ici une tendance, aucune conclusion claire ne peut être tirée. Le même raisonnement peut être effectué pour les *Primates*, les *Cetartiodactyla*, les *Carnivora*, les *Diprotodontia* mais ceux-ci ont des z-score positifs pour l'intra-groupe. Pour *Perissodactyla*, *Lagomorpha*, *Didelphimorphia* et *Eulipotyphla* l'histoire est différente, chacun de ces ordres présente un partage viral intra-groupe potentiel supérieur d'au moins un écart-type du partage global. C'est un signal très fort qui montre que ces espèces partagent un ensemble de virus similaire.

Nos résultats sont en contradiction avec Albery *et al.* (2020) qui prédisent des partages viraux intra- et extra-ordre pour la plupart opposés aux nôtres. Notamment pour les *Rodentia* et les *Chiroptera* ils prédisent en moyenne par espèces plus de lien intra- que extra-ordre ou encore pour les *Perissodactyla* et *Lagomorpha* des moyennes extra- plus importantes qu'intra-ordre. Nos résultats ont tendance à être plus en accord avec Carlson *et al.* (2022b) qui en combinant des modèles de partage viraux et de distribution d'espèces montre que les *Chiroptera* seront des vecteurs importants dans la dissémination virale.

On notera que puisque l'Ordre des Primates a une tendance positive de partage viral potentiel intra-ordre, et que les *Chiroptera* et *Rodentia* ont une tendance positive de partage viral potentiel extra-ordre, nous apportons un indice que les risques d'apparition de maladies

zoonotiques auraient tendance à provenir de notre propre ordre ou de ces deux derniers.

Communicabilité : pistes d'améliorations

Dans l'état actuel, la méthode présentée ici reste assez compliquée à interpréter biologiquement. Premièrement, nous avons choisi d'explorer la communicabilité sous la forme spectrale proposée par (Estrada et Hatano, 2008), cette version a l'avantage d'obtenir des valeurs de communicabilité asymptotique. Mais des valeurs asymptotiques n'ont pas forcément de sens, considère seulement les 6 ou 8 premières étapes pourrais être largement suffisant car la transmission virale inter-espèces ce fait généralement sur des hôtes proche phylogénétiquement et géographiquement (Davies et Pedersen, 2008). Dans la même optique, la pénalisation utilisée n'est potentiellement pas adaptée à notre problématique car probablement trop faible. D'autres formes de pénalisations ont déjà été proposées (Estrada et Higham, 2010), mais aucune d'elles - a notre connaissance - dans une optique écologique. Une variation de la communicabilité utilisant la *Singular Value Decomposition* a aussi été proposée (Crofts et Higham, 2009; Crofts *et al.*, 2010) il est possible qu'elle soit plus adaptée pour les réseaux hôte virus. Trouver une forme de la communicabilité ayant une meilleure signification biologique devrais donc faire partie des prochaines phases de développement de cette approche. Deuxièmement, nous avons utilisé un méta-réseau, la caractéristique de ceux-ci est d'avoir une très forte densité de lien, car ils comprennent des associations potentielles qui ne sont pas forcément réalisées (Strydom *et al.*, 2021). La structure des réseaux peut drastiquement changer nos résultats car la communicabilité est entièrement basée sur celle-ci. Afin d'améliorer cette partie, deux pistes principales peuvent être proposées. La première est d'utiliser des réseaux géographiquement réalistes, mais cela limiterait la méthode aux espèces de même milieu. Cela ferait gagner en précision mais perdre en pouvoir prédictif global. Une seconde option consisterait à construire des méta-réseaux prenant en compte la probabilité d'association, mais aussi la probabilité de rencontre entre deux espèces, avec par exemple une pénalisation de cette probabilité avec la distance géographique.

Dans la partie quantification d'importance, nous avons montré que l'approche de la communicabilité est fortement influencé par la densité de lien global ou par la centralité

des nœuds. Il est possible que d'autres caractéristiques d'agencement impactent cette mesure. Utilisées, des interactions dont la prédiction est basée sur la structure du réseau pourraient mener à des biais de validation qui ne sont pas flagrants dans nos résultats. Cette mesure doit donc pour le moment être interprétée avec précaution et testée plus amplement. Il est fortement possible que les résultats obtenus par la communicabilité soient fortement influencés par la méthode de prédiction ou qu'ils aient un comportement peu évident à comprendre ou à percevoir. Ce type de comportement pourrait potentiellement biaiser nos résultats de partage viral. Il y a un par exemple un gros décalage de valeurs entre les ordres très peu représentés du réseau et les ordres les plus représentés. L'abondance des espèces dans le réseau a clairement eu un impact sur la mesure du partage viral potentiel, il reste à savoir si cet impact est bénéfique ou préjudiciable pour une interprétation biologique. Une première piste pour mettre à l'épreuve la communicabilité est donc de la tester sur des méta-réseaux ou réseaux prédit à partir de différentes méthodes et avec diverses structures. Si des résultats similaires sont obtenus, cela pourrait indiquer non seulement que la méthode est robuste, mais aussi que les modèles de prédiction sont cohérents entre eux.

Conclusion

Nous avons montré en prenant en compte la structure globale des méta-réseaux Trefle et Clover, que les prédictions d'associations faites par (Poisot *et al.*, 2023) nous permettent, comme voulu, de limiter les biais d'échantillonnage. Certaines espèces et certaines associations qui sont centrales dans les réseaux peuvent être détectées grâce à cette méthode. Nous avons trouvé que certains ordres de mammifères ont un partage viral intra-groupe plus important que la moyenne. Nous avons aussi détecté certaines tendances de partage viral pour les ordres les plus représentés du réseau. Cependant, la métrique de partage viral potentiel basée sur la communicabilité manque pour l'instant de maturité. Nos résultats restent tout de même encourageants et ouvrent la voie pour prendre en considération l'ensemble des associations dans les prédictions de partage viral.

Matériels complémentaires

Données et codes disponibles en ligne : <https://github.com/GhMathis/edges>

Illustration de la page de couverture : Réseau Trefle, associations de plus grandes Importances partagés avec Clover. Package R utilisé : *Igraph*.

Références

- G. F. Albery, E. A. Eskew, N. Ross, and K. J. Olival. Predicting the global mammalian viral sharing network using phylogeography. *Nature Communications*, 11(1) :2260, 2020. ISSN 2041-1723. doi : 10.1038/s41467-020-16153-4.
- G. F. Albery, D. J. Becker, L. Brierley, C. E. Brook, R. C. Christofferson, L. E. Cohen, T. A. Dallas, E. A. Eskew, A. Fagre, M. J. Farrell, E. Glennon, S. Guth, M. B. Joseph, N. Mol-lentze, B. A. Neely, T. Poisot, A. L. Rasmussen, S. J. Ryan, S. Seifert, A. R. Sjodin, E. M. Sorrell, and C. J. Carlson. The science of the host–virus network. *Nature Microbiology*, 6(12) :1483–1492, 2021. ISSN 2058-5276. doi : 10.1038/s41564-021-00999-5.
- E. D. Belay, J. C. Kile, A. J. Hall, C. Barton-Behravesh, M. B. Parsons, S. Salzer, and H. Walke. Zoonotic Disease Programs for Enhancing Global Health Security. *Emerging Infectious Diseases*, 23(13) :S65–70, 2017. ISSN 1080-6059. doi : 10.3201/eid2313.170544.
- C. Capinha, F. Essl, H. Seebens, D. Moser, and H. M. Pereira. The dispersal of alien species redefines biogeography in the Anthropocene. *Science*, 348(6240) :1248–1251, 2015. doi : 10.1126/science.aaa8913.
- C. J. Carlson, G. F. Albery, C. Merow, C. H. Trisos, C. M. Zipfel, E. A. Eskew, K. J. Olival, N. Ross, and S. Bansal. Climate change increases cross-species viral transmission risk. *Nature*, 607(7919) :555–562, 2022a. ISSN 1476-4687. doi : 10.1038/s41586-022-04788-w.
- C. J. Carlson, R. J. Gibb, G. F. Albery, L. Brierley, R. P. Connor, T. A. Dallas, E. A. Eskew, A. C. Fagre, M. J. Farrell, H. K. Frank, R. L. Muylaert, T. Poisot, A. L. Rasmussen, S. J. Ryan, and S. N. Seifert. The Global Virome in One Network (VIRION) : An Atlas of Vertebrate-Virus Associations. *mBio*, 13(2) :e02985–21, 2022b. doi : 10.1128/mbio.02985-21.
- D. Carroll, P. Daszak, N. D. Wolfe, G. F. Gao, C. M. Morel, S. Morzaria, A. Pablos-Méndez, O. Tomori, and J. A. K. Mazet. The Global Virome Project. *Science*, 359(6378) :872–874, 2018. doi : 10.1126/science.aap7463.
- J. J. Crofts and D. J. Higham. A weighted communicability measure applied to complex brain networks. *Journal of The Royal Society Interface*, 6(33) :411–414, 2009. doi : 10.1098/rsif.2008.0484.
- J. J. Crofts, E. Estrada, D. J. Higham, and A. Taylor. Mapping directed networks. *Electronic Transactions in Numerical Analysis*, 37 :337–350, 2010. ISSN 1068-9613.
- T. J. Davies and A. B. Pedersen. Phylogeny and geography predict pathogen community similarity in wild primates and humans. *Proceedings. Biological Sciences*, 275(1643) :1695–1701, 2008. ISSN 0962-8452. doi : 10.1098/rspb.2008.0284.
- E. Estrada and N. Hatano. Communicability in complex networks. *Physical Review E*, 77 (3) :036111, 2008. doi : 10.1103/PhysRevE.77.036111.

- E. Estrada and D. J. Higham. Network Properties Revealed through Matrix Functions. *SIAM Review*, 52(4) :696–714, 2010. ISSN 0036-1445. doi : 10.1137/090761070.
- E. C. Fricke and J.-C. Svenning. Accelerating homogenization of the global plant–frugivore meta-network. *Nature*, 585(7823) :74–78, 2020. ISSN 1476-4687. doi : 10.1038/s41586-020-2640-y.
- R. Gibb, G. F. Albery, D. J. Becker, L. Brierley, R. Connor, T. A. Dallas, E. A. Eskew, M. J. Farrell, A. L. Rasmussen, S. J. Ryan, A. Sweeny, C. J. Carlson, and T. Poisot. Data Proliferation, Reconciliation, and Synthesis in Viral Ecology. *BioScience*, 71(11) : 1148–1156, 2021. ISSN 0006-3568. doi : 10.1093/biosci/biab080.
- J. M. Gómez, C. L. Nunn, and M. Verdú. Centrality in primate–parasite networks reveals the potential for the transmission of emerging infectious diseases to humans. *Proceedings of the National Academy of Sciences*, 110(19) :7738–7741, 2013. doi : 10.1073/pnas.1220716110.
- M. R. Helmus, D. L. Mahler, and J. B. Losos. Island biogeography of the Anthropocene. *Nature*, 513(7519) :543–546, 2014. ISSN 1476-4687. doi : 10.1038/nature13739.
- E. P. Hoberg and D. R. Brooks. Evolution in action : Climate change, biodiversity dynamics and emerging infectious disease. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 370(1665) :20130553, 2015. ISSN 0962-8436. doi : 10.1098/rstb.2013.0553.
- N. Mollentze and D. G. Streicker. Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proceedings of the National Academy of Sciences*, 117(17) :9423–9430, 2020. doi : 10.1073/pnas.1919176117.
- I. Morales-Castilla, P. Pappalardo, M. J. Farrell, A. A. Aguirre, S. Huang, A.-L. M. Gehman, T. Dallas, D. Gravel, and T. J. Davies. Forecasting parasite sharing under climate change. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 376(1837) :20200360, 2021. ISSN 1471-2970. doi : 10.1098/rstb.2020.0360.
- M. Pichler, V. Boreux, A.-M. Klein, M. Schleuning, and F. Hartig. Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks. *Methods in Ecology and Evolution*, 11(2) :281–293, 2020. ISSN 2041-210X. doi : 10.1111/2041-210X.13329.
- S. L. Pimm. Biodiversity : Climate Change or Habitat Loss — Which Will Kill More Species? *Current Biology*, 18(3) :R117–R119, 2008. ISSN 0960-9822. doi : 10.1016/j.cub.2007.11.055.
- R. K. Plowright, C. R. Parrish, H. McCallum, P. J. Hudson, A. I. Ko, A. L. Graham, and J. O. Lloyd-Smith. Pathways to zoonotic spillover. *Nature Reviews Microbiology*, 15(8) : 502–510, 2017. ISSN 1740-1534. doi : 10.1038/nrmicro.2017.45.
- T. Poisot, M.-A. Ouellet, N. Mollentze, M. J. Farrell, D. J. Becker, L. Brierley, G. F. Albery, R. J. Gibb, S. N. Seifert, and C. J. Carlson. Network embedding unveils the hidden interactions in the mammalian virome. *Patterns*, page 100738, 2023. ISSN 2666-3899. doi : 10.1016/j.patter.2023.100738.

- M. J. Pongsiri, J. Roman, V. O. Ezenwa, T. L. Goldberg, H. S. Koren, S. C. Newbold, R. S. Ostfeld, S. K. Pattanayak, and D. J. Salkeld. Biodiversity Loss Affects Global Disease Ecology. *BioScience*, 59(11) :945–954, 2009. ISSN 0006-3568. doi : 10.1525/bio.2009.59.11.6.
- R. P. Rohr and J. Bascompte. Components of Phylogenetic Signal in Antagonistic and Mutualistic Networks. *The American Naturalist*, 184(5) :556–564, 2014. ISSN 0003-0147. doi : 10.1086/678234.
- B. Saladin, L. Pellissier, C. H. Graham, M. P. Nobis, N. Salamin, and N. E. Zimmermann. Rapid climate change results in long-lasting spatial homogenization of phylogenetic diversity. *Nature Communications*, 11(1) :4663, 2020. ISSN 2041-1723. doi : 10.1038/s41467-020-18343-6.
- T. Strydom, M. D. Catchen, F. Banville, D. Caron, G. Dansereau, P. Desjardins-Proulx, N. R. Forero-Muñoz, G. Higinio, B. Mercier, A. Gonzalez, D. Gravel, L. Pollock, and T. Poisot. A roadmap towards predicting species interaction networks (across space and time). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 376(1837) :20210063, 2021. ISSN 1471-2970. doi : 10.1098/rstb.2021.0063.
- T. Strydom, S. Bouskila, F. Banville, C. Barros, D. Caron, M. J. Farrell, M.-J. Fortin, V. Hemming, B. Mercier, and L. J. Pollock. Graph embedding and transfer learning can help predict potential species interaction networks despite data limitations. 2022.
- R. J. White and O. Razgour. Emerging zoonotic diseases originating in mammals : A systematic review of effects of anthropogenic land-use change. *Mammal Review*, 50(4) : 336–352, 2020. ISSN 1365-2907. doi : 10.1111/mam.12201.

Résumé. Les changements environnementaux induisent des mouvements d'espèces importants sur l'ensemble du globe. Des espèces ne co-occurent pas se retrouvent dans les mêmes écosystèmes. Les virus et leurs hôtes vecteurs sont aussi touchés. Dans l'objectif d'anticiper les maladies zoonotiques, ces interactions méritent un intérêt particulier. Des initiatives pour prédire les interactions entre espèces voient le jour. Nous utilisons un réseau hôte virus observé de l'ensemble des mammifères et son méta-réseau prédit associé (Poisot *et al.*, 2023) pour identifier les espèces structurantes du méta-réseau. Nous cherchons aussi à quantifier le partage viral potentiel par ordres d'hôtes en se basant sur la structure globale du méta-réseau.

Nous supposons que : -1- Les espèces très structurantes du réseau observé verront leur influence fortement diminuée dans le méta-réseau prédit et que globalement les associations seront moins influentes dans le méta-réseau prédit. -2- Les ordres les plus représentés dans le réseau auront un partage viral intra-ordre plus important que les moins représentés. -3- Le partage viral intra-ordre sera toujours plus élevé que le partage viral extra-ordre.

Nous trouvons que la prédiction a rééquilibré le poids des interactions et limite les effets de biais d'échantillonnage. Comparé au méta-réseau global les *Chiroptera* et les *Rotendia* ont un partage viral potentiel intra-ordre plus faible alors que les *Primate*, les *Cetartidocyla* et les *Carnivora* ont un potentiel plus élevé. Les ordres avec peu de représentants ont un partage viral intra-ordre très élevé par rapport à la moyenne. Nos résultats indiquent que les maladies zoonotiques sont légèrement plus susceptibles de provenir des *Chiroptera*, *Rotendia* ou *Primate*.

mots clés : Communicabilité, Structure global, Méta-réseaux, Partage viral, Maladies zoonotiques, Hôtes, Virus.

ESTIMATE VIRAL SHARING AND IMPORTANCE OF ASSOCIATIONS THROUGH THE GLOBAL STRUCTURE OF HOST VIRUS META-NETWORK

Abstract. Environmental changes induce major displacements of species around the globe. Species that don't co-occur may share same ecosystems in the future. This is also the case for virus and their host. These interactions are of particular interest, especially in view of zoonotic diseases. Some initiative to predict potential species interactions are developed. To identify species that structure the network, we are using an observed host virus meta-network that encompasses all the mammals and its linked predicted meta-network (Poisot *et al.*, 2023). We also quantify potential viral sharing per host orders based on the global structure of meta-network.

We hypothesise that : -1- Species that structure a lot the observed meta-network will see their influence strongly diminish and that overall, associations on the predicted network will be less influential. -2- Most represented host species in the meta-network will have a viral sharing greater than less represented species. -3- The intra-order viral sharing will always be greater than the extra-order viral sharing.

We observed that the predictions rebalanced interactions influences and limited the sampling bias. Compare to the global meta-network, we observed small trends for *Chiroptera* and *Rotendia* with less intra-order viral sharing and more for *Primate*, *Cetartidocyla* and *Carnivora*. Orders with fewer species had a greater intra-order viral sharing than overall viral sharing. Our results indicate that zoonotic diseases are more likely to originate from *Chiroptera*, *Rotendia* or *Primate*.

key-words : Communicability, Global structure, Meta-network , Viral sharing, Zoonotic disease, Host, Virus.