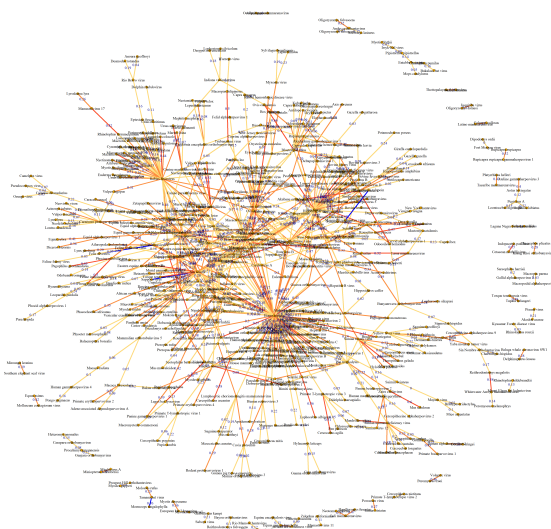

NOUVELLE APPROCHE POUR QUANTIFIER LES INTERACTIONS DE RÉSEAUX HÔTES VIRUS



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 parcours modélisation en
écologie (MODE)

avril-juin 2023 soutenus le : 19 juin 2023

Référent universitaire : **Marie-Pierre Etienne**

encadrant

superviseur

Introduction

L'époque géologique actuelle, l'Anthropocène, est synonyme de forte homogénéisation de la composition géographique des espèces à l'échelle planétaire [Capinha et al., 2015, Helmus et al., 2014]. Ce brassage global du monde vivant est induit notamment par une migration généralisée des espèces vers des latitudes plus élevées [], des déplacements humains importants [] et une destruction soutenue et systémique des habitats []. Ces mouvements provoquent une réduction de la compartimentation des méta-réseaux, qui était préalablement établie par des barrières bio-géographiques de grande échelle [Fricke and Svenning, 2020]. En d'autres termes, des espèces ne pouvant précédemment pas interagir pour des raisons de séparation géographique se retrouvent aujourd'hui dans les mêmes écosystèmes. Les réseaux d'interactions hôtes-virus ne déroge pas à la règle et sont eux aussi fortement touchés par ces changements []. Ce n'est pas anodin car l'introduction de nouveaux parasites [], ou d'hôtes vecteurs [] dans un milieu peut avoir des effets considérables []. De plus, l'enjeu n'est pas seulement écologique mais touche aussi à la santé publique [], car une majeure partie des maladies sont zoonotiques [Belay et al., 2017]. La destruction d'habitats [], les migrations à petite [] ou grande [] échelle ainsi que l'augmentation de la densité de population [] rend les contacts virus-hôte-humain plus prévalant que jamais. Il y a donc un besoin considérable de comprendre par quel mécanisme certaines espèces vont interagir ou non ? quels sont ceux qui structurent les réseaux d'interaction ? quels sont les canaux potentiels de transmission ? Répondre à ces questions permettrait de mieux comprendre le rôle des interactions dans les écosystèmes et d'améliorer la conservation. Cela rendrait aussi possible l'anticipation de futures maladies zoonotiques ou des sources d'émergence plus probables sur lesquels il faudrait apporter un effort de recherche plus important [].

Les causes de transmission de maladies zoonotiques sont multi-niveaux [Belay et al., 2017] et lutter contre leurs apparitions requiert donc une approche multi-sectorielle. La distribution, l'intensité et les routes d'exposition [Belay et al., 2017], mais aussi les distances géographiques et phylogénétiques [Albery et al., 2020] sont quelques exemples des nombreuses variables jouant sur la probabilité de diffusion des virus. Des approches prédictives de partage viral de paires d'hôtes ont été fructueuses [Albery et al., 2020] mais se basent seulement sur des dites paires. Il a été montré que les réseaux apportent des informations sur la sépara-

tion géographique[], la distance phylogénétique [] et les meta-réseaux probabiliste peuvent prendre en compte la probabilité de réalisation des interactions [Strydom et al., 2022]. C'est objets renferment donc de nombreuse information écologique, spatiales, historique et phylogénétiques, or ils sont généralement incomplets et énormément biaisé []. Cependant de nouvelles méthodes de prédictions d'interactions permettent de limité ces problèmes []. Nous proposons donc ici, en utilisant un réseaux prédits et des quantification d'importance de chemin de transmission, une nouvelle approche prenant en compte l'ensemble du réseaux pour identifié les potentiel espèces et groupes d'espèces susceptible de propagé des maladies zoonotiques, ainsi qu'une identification des canaux de transmission les plus importants au sein de l'ensemble de la classe des mammifères. **(Fin introduction)**

La gestion d'une maladie avant sont apparition sur des sujets humain est la méthode la plus efficace. L'écologie prédictive permet de telle approches. Un des frein majeur dans la prédiction d'interaction est le manque de données fiable [Strydom et al., 2021]. Il est difficile d'observer des interactions qui sont des évènements ayant une limite dans le temps. Il y'a aussi qualitativement plus interactions rare qui font suite a des rencontre fortuite que d'interaction commune []. Il faut aussi pouvoir dissocié la co-occurrence et l'interaction []. Tous ces facteur font que la collection de données qualitative sur les interaction entre espèces est très complexe. c'est pour cette raison que des initiative de regroupement de données ont récemment vu le jour [Carlson et al., 2022].

Or les interactions biotiques font partie intégrantes des mécanismes structurant et équilibrant les écosystèmes. Changer la structure des relations entre espèces a donc des effet direct sur les systèmes écologiques. Prédire les interactions potentielles afin d'anticiper les effets de l'introduction d'espèces dans les écosystèmes est donc d'une importance considérable pour la conservation, mais aussi pour d'autre problématique tel que l'émergence de maladie. Effectivement l'application de ces méthode sur les réseaux hôtes virus est plus que jamais une nécessité. D'une part ils ont une grand importance pour les écosystème (*ou* non pas que les interactions hôtes virus sont d'une importance primordial dans la stabilité des écosystèmes), mais aussi est surtout car cela apport des informations d'intérêt majeur pour la santé public.

Les interactions entres espèces sont structurantes au sein des écosystèmes. Leurs im-

portances dans la stabilité font d'elle un champs d'études de premier plan en écologie. C'est notamment les méthodes de prédictions de ces interactions qui récemment, dans la littérature scientifique, prennent de plus en plus de place. De nombreuses approches font leurs émergences.

La prédictions d'interaction entre espèces et plus précisément entre hôtes et parasite est un enjeu car une problématique actuelle est l'homogénéisation par l'Homme de la composition géographique des espèces[Fricke and Svenning, 2020, Helmus et al., 2014, Capinha et al., 2015] (*voir les 3 premier article citer*) + *trouve que les espèces introduites sont plus prévalentes dans les zones fortement modifié par les humains*. Quantifié l'impact de l'introduction de nouvelles espèces dans un réseaux d'interactions est primordial pour prévoir si cela aura un effets négligeable ou non sur la structure du réseaux, l'abondance et la diversité des espèces (**quelles pourrait être les relation entre l'importance des interactions et des mesures telle que l'abondance, la diversité ... ? link prediction may have subtle conceptual differences. For example, a 'link' can be hard to define : is the aim to predict all existing hosts of a virus or all potentially compatible hosts ?**

- Homogénéisation par delà les barrière géologique - a un effet sur la structure des réseaux locaux - a un effet sur la structure du méta réseaux -
 1. Les interactions structurent l'écosystème
 2. Les interactions dans la communauté
 3. Comment les interactions entre espèces sont représentées dans la littérature, les avantages et les inconvénient de chaque méthode
 4. Challenge des prédictions d'interaction, biais, sources d'erreur et utilité - Prédire des interaction entre espèces = difficile - Grosse quantité de données mais pas standardisée + beaucoup de biais - Besoin de méthodes qui prennent en compte ces contraintes - Différentes approches : A partir des traits (plus facile pour généraliser). A partir des réseaux interactions
 5. Interaction hôtes virus, enjeux important. - La prédiction d'interaction potentiel entre hôte virus → challenge important
 6. Les réseaux d'interaction
 7. *Graph embedding* appliqué aux réseaux

8. Pourquoi s'intéresser aux interactions structurantes d'un réseaux, quelles impacts à la prédiction des interaction sur l'importance de celle-ci au sein du réseaux. - Des méthodes pour prédire les associations entre espèce commence a faire apparition Besoin de s'assurer que ces méthodes ne prédisent pas seulement les interactions fortement structurante du réseaux d'une part Besoin de connaitre quelle interaction sont structurantes et lesquelles le sont moins afin d'orienter les études Peut potentiellement amélioré la prédiction d'interaction grâce a une meilleur connaissance des réseaux

Mat et meth

0.1 Obtentions des données

Les données utilisées proviennent de 2 méta-réseaux. CLOVER est composé de 5477 associations mammifère virus observer, il est tiré de multiples bases de données [Gibb et al., 2021]. TREFLE est un réseaux comprenant les même espèces que CLOVER mais avec 75901 associations supplémentaires, celle-ci ont été prédites par des méthodes de réductions de dimension de réseaux [Poisot et al., 2023] et viennent avec la probabilité qu'elles soient effectivement réalisées.

0.2 Mesure d'importances de paire d'espèces a partir de réseaux d'interactions

La quantification d'importance des paires d'espèces utilise la communicabilité entre les noeuds du réseaux. Cette mesure est introduite par [Estrada and Hatano, 2008] et se base sur la caractéristique des puissance de matrices d'adjacence, qui permet d'obtenir, pour une puissance k , le nombre de chemin possible en k étapes entre chaque paire de noeud. En ajoutant chaque puissance et en pénalisant les chemin de grand taille, 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. Puisqu'il y a une infinité d'étape possible, et que la pénalisations se fait avec les factoriel de k . G est donc l'exponentiel de A , ce qui leurs 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 vecteur propre associé aux valeurs propre λ_j de la matrice $A_{(n \times n)}$. C'est a partir de l'équation 2 que la communicabilité est calculé.

Les matrices G sont aussi normalisées par rapport a leurs valeurs minimum et maximum. Cela rend possible la comparaison d'association d'espèces présentes dans des réseaux ayant des noeuds ou des liens différents.

TREFLE et CLOVER en tant que réseaux d'interaction bipartie, sont caractérisés par des matrices 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.

0.3 Probabilité et importance

0.4 Comparaison d'importance entre CLOVER et TREFLE

Les 5477 d'associations que partagent CLOVER et TREFLE sont mit en relation

où φ_j sont les vecteurs propres de la matrice A associés au valeurs propres λ_j

Plusieur approches : - Perturbations du réseaux puis comparaison des Standard Value
Décomposition + Random Graph Dot Product

- Perturbations puis analyse spectral et mesure communicabilité

$$G_{pq} = \sum_{j=1}^n \varphi_j(p) \varphi_j(q) e^{\lambda_j}$$

- Stockastique graphs models - Stockastick block models - Latent block models - Approche exploratoire de backpropagation ?

Résultats

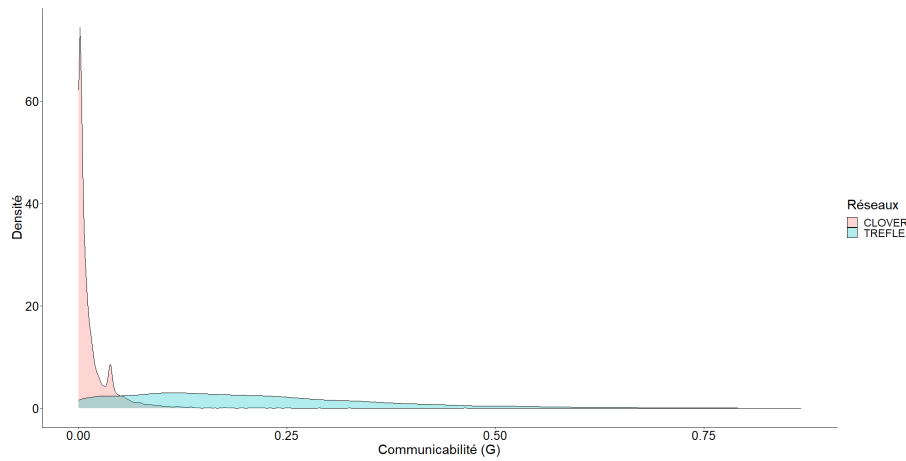


FIGURE 1 – distributions des valeurs de communicabilité pour chaque association observées

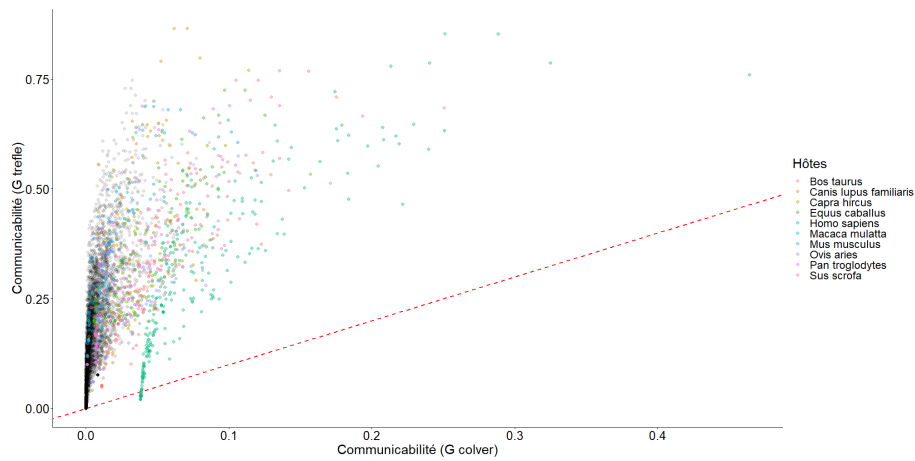


FIGURE 2 – comparaison des valeurs de communicabilité pour chaque association observées

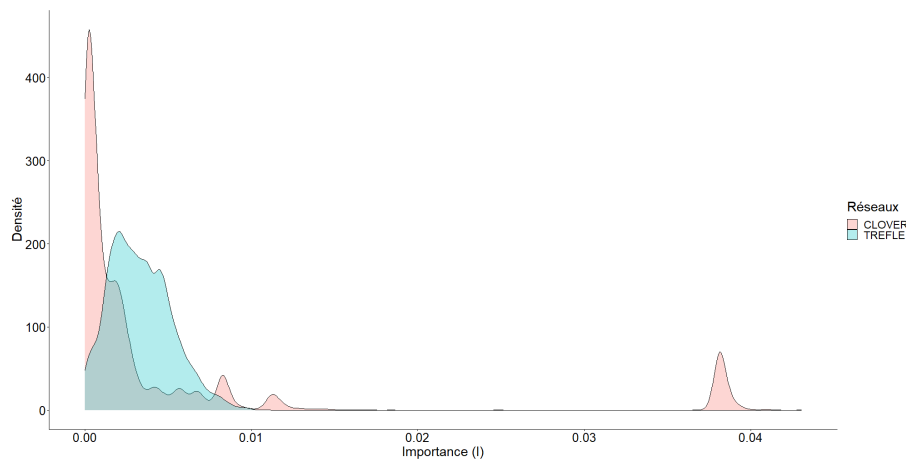


FIGURE 3 – distributions des valeurs d'importance pour chaque association observées

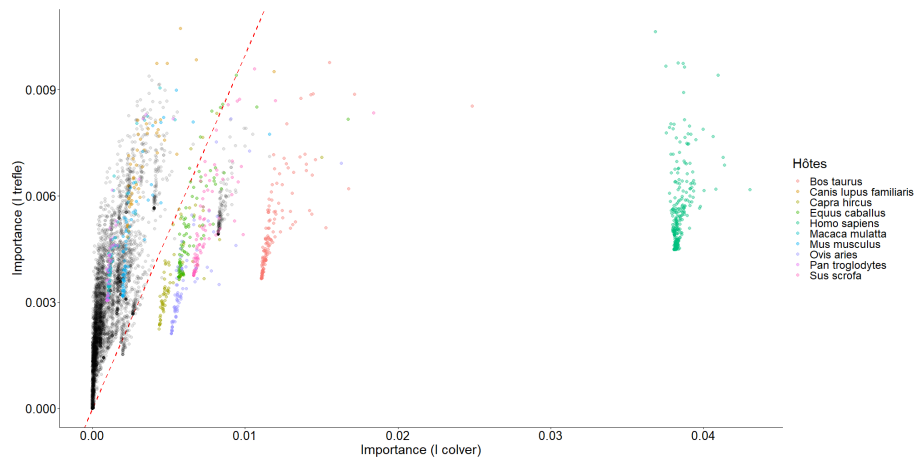


FIGURE 4 – comparaison des valeurs d'importance pour chaque association observées

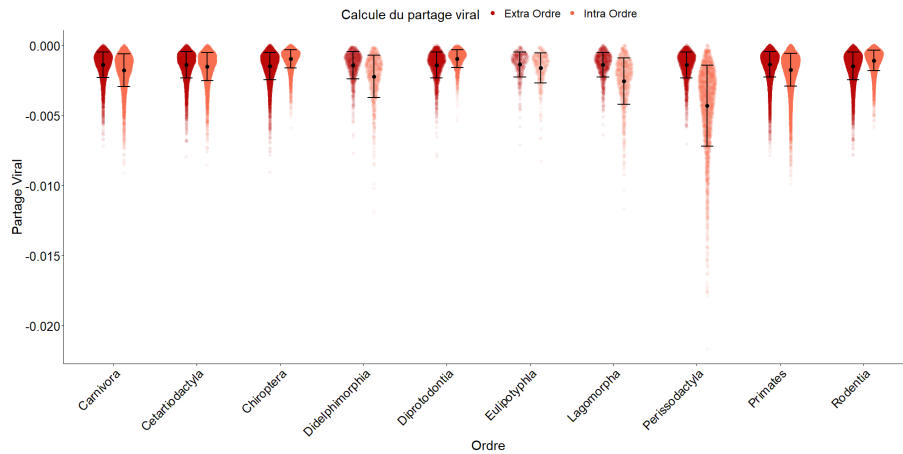


FIGURE 5 – Score de partage viral entre pour des hôtes compris ou non dans le même ordre

- Importance des interactions n'a pas de lien avec la probabilité de les observer

Discutions

Conclusion

Références

[Albery et al., 2020] Albery, G. F., Eskew, E. A., Ross, N., and Olival, K. J. (2020). Predicting the global mammalian viral sharing network using phylogeography. *Nature Communications*, 11(1) :2260.

- [Belay et al., 2017] Belay, E. D., Kile, J. C., Hall, A. J., Barton-Behravesh, C., Parsons, M. B., Salyer, S., and Walke, H. (2017). Zoonotic Disease Programs for Enhancing Global Health Security. *Emerging Infectious Diseases*, 23(13) :S65–70.
- [Capinha et al., 2015] Capinha, C., Essl, F., Seebens, H., Moser, D., and Pereira, H. M. (2015). The dispersal of alien species redefines biogeography in the Anthropocene. *Science*, 348(6240) :1248–1251.
- [Carlson et al., 2022] Carlson, C. J., Gibb, R. J., Albery, G. F., Brierley, L., Connor, R. P., Dallas, T. A., Eskew, E. A., Fagre, A. C., Farrell, M. J., Frank, H. K., Muylaert, R. L., Poisot, T., Rasmussen, A. L., Ryan, S. J., and Seifert, S. N. (2022). The Global Virome in One Network (VIRION) : An Atlas of Vertebrate-Virus Associations. *mBio*, 13(2) :e02985–21.
- [Estrada and Hatano, 2008] Estrada, E. and Hatano, N. (2008). Communicability in complex networks. *Physical Review E*, 77(3) :036111.
- [Fricke and Svenning, 2020] Fricke, E. C. and Svenning, J.-C. (2020). Accelerating homogenization of the global plant–frugivore meta-network. *Nature*, 585(7823) :74–78.
- [Gibb et al., 2021] Gibb, R., Albery, G. F., Becker, D. J., Brierley, L., Connor, R., Dallas, T. A., Eskew, E. A., Farrell, M. J., Rasmussen, A. L., Ryan, S. J., Sweeny, A., Carlson, C. J., and Poisot, T. (2021). Data Proliferation, Reconciliation, and Synthesis in Viral Ecology. *BioScience*, 71(11) :1148–1156.
- [Helmus et al., 2014] Helmus, M. R., Mahler, D. L., and Losos, J. B. (2014). Island biogeography of the Anthropocene. *Nature*, 513(7519) :543–546.
- [Poisot et al., 2023] Poisot, T., Ouellet, M.-A., Mollentze, N., Farrell, M. J., Becker, D. J., Brierley, L., Albery, G. F., Gibb, R. J., Seifert, S. N., and Carlson, C. J. (2023). Network embedding unveils the hidden interactions in the mammalian virome. *Patterns*, page 100738.
- [Strydom et al., 2022] Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M. J., Fortin, M.-J., Hemming, V., Mercier, B., and Pollock, L. J. (2022). Graph embedding and transfer learning can help predict potential species interaction networks despite data limitations.

[Strydom et al., 2021] Strydom, T., Catchen, M. D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., Forero-Muñoz, N. R., Higinio, G., Mercier, B., Gonzalez, A., Gravel, D., Pollock, L., and Poisot, T. (2021). 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.