

Template for preparing your research report submission to PNAS using RMarkdown

Pénélope Robert^a, Édouard Nadon-Beaumier^a, Alexis Matte^a, and Nadia Tardy^a

^aUniversité de Sherbrooke, Département de bio, 2500 Boulevard de l'Université, Sherbrooke, Québec, G1V 0A9

This manuscript was compiled on April 29, 2022

Please provide an abstract of no more than 250 words in a single paragraph. Abstracts should explain to the general reader the major contributions of the article. References in the abstract must be cited in full within the abstract itself and cited in the text.

1. Introduction

La notion de réseau écologique d'un écosystème est souvent déformée et confondue avec celle de réseau trophique. Nous avons régulièrement tendance à imaginer les interactions entre les différentes espèces avec une belle pyramide qui retrace les réseaux trophiques et la chaîne alimentaire en plaçant le plus important prédateur au sommet du prisme. En réalité, un réseau écologique inclut toutes les multitudes interactions possibles entre les espèces comme le mutualisme, la compétition, le commensalisme et le parasitisme (1). Évidemment, les réseaux trophiques ont une grande influence sur le réseau écologique, mais ce dernier ne devrait pas être représenté par une pyramide. Il devrait plutôt être décrit comme une toile ("web") avec les espèces clés de l'écosystème au centre, c'est-à-dire les espèces qui ont le plus grand nombre de connexions interspécifiques. On retrouve également dans la littérature les propriétés de petit monde que l'on peut découvrir dans les réseaux (2). C'est dans ce contexte que nous avons voulu comparer le réseau de collaboration entre les étudiants en écologie et le réseau écologique. Afin d'approfondir notre réflexion, nous avons voulu vérifier si le nombre de collaborations par étudiant de BIO500 sont semblables ou bien s'il y a des "étudiants clés," c'est-à-dire qui possèdent le plus de collaborations différentes au même titre qu'une espèce clé.

2. Méthode

Avant même de se pencher sur nos questions, il a fallu que chacun des étudiants de BIO500 recense l'entièreté des étudiants avec qui il ou elle a collaboré dans un travail d'équipe pendant son parcours académique en écologie à l'Université de Sherbrooke. Une fois les données recueillies et traitées, nous avons toutes les informations en main pour commencer notre analyse. Afin de comparer en général le réseau étudiant et les réseaux écologiques, nous avons généré deux figures: le réseau de liens des étudiants de BIO500 et leur collaboration Figure 1 ainsi qu'un diagramme à bande du nombre de collaboration pour chacun des étudiants de BIO500 Figure 2.

Ensuite, pour répondre à notre question, nous avons tout d'abord émis une hypothèse. Selon nous, le nombre de collaborations par personne devrait être environ semblable d'un étudiant à l'autre puisque le parcours académique et les cours sont relativement les mêmes. Nous nous attendons donc à ce que le nombre de collaborations par étudiant suive une distribution normale. En fait, c'est la méthode utilisée par

les écologistes pour savoir si le réseau qu'ils étudient ont un nombre "normal" de liens (3). Si ce n'est pas le cas, nous nous pencherons sur l'hypothèse des "étudiants clés." L'objectif sera de les identifier et de déterminer la cause de leur haut nombre de collaborateurs.

En premier lieu, nous avons effectué une analyse visuelle et graphique, puis en deuxième lieu une analyse statistique de notre distribution. Les deux analyses, comme les traitements de données ci-haut, ont été effectuées sur le logiciel R version 4.0.3. L'analyse visuelle est représentée par un histogramme, voici 4 résultats possibles ainsi que leur signification (3):

Une courbe symétrique et mince

C'est à quoi nous nous attendions. La variance entre le nombre de collaboration entre les élèves de BIO500 est faible et très concentré autour de la moyenne. Les étudiants ont donc eux un parcours académique semblable.

Une courbe symétrique et large

Cette courbe corrobore également notre hypothèse: une distribution normale. Par contre, la variation entre les élèves est plus grande qu'attendu. Il a une petite divergence entre les parcours, mais qui sans débalancer les données.

Une courbe asymétrique à gauche

Ce genre de distribution non normal indique que la majorité des élèves se retrouve malgré tout proche de la moyenne. Cependant, un certain nombre non négligeable d'élèves ont plus que collaboration et qu'aucun ou très peu d'élèves ont peu collaboré. Cette courbe nous indique qu'il a plusieurs "étudiants clés" plus ou moins important.

Une courbe symétrique avec des données aberrantes vers la droite

On observe une distribution normale, mais avec ce qui semble être des données aberrantes. Cela représente précisément l'autre hypothèse "d'étudiants clés" peu nombreux et très important. Leur divergence distinctive pourrait signifier une grande différence de leur parcours.

L'analyse statistique est effectuée par le test de Shapiro-Wilk dans lequel les hypothèses sont les suivantes:

Hypothèse nulle (H0): les données ont une distribution normale
Hypothèse alternative (H1): les données n'ont pas une distribution normale

La commande `shapiro.test()` est facile d'utilisation et sa sortie nous informe sur la valeur de W et le p-value. C'est seulement le p-value qui nous intéresse. Si p-value est plus petit que 0.05, on rejette l'hypothèse nulle: les données sont non normales. Si p-value est plus grand que 0,05, on ne rejette

Significance Statement

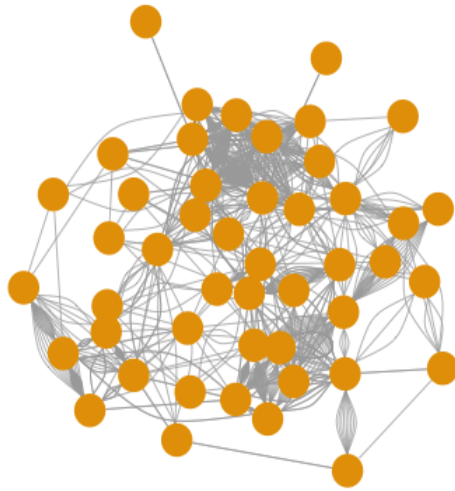


Fig. 1. Réseau de collaborations entre les élèves de BIO500.

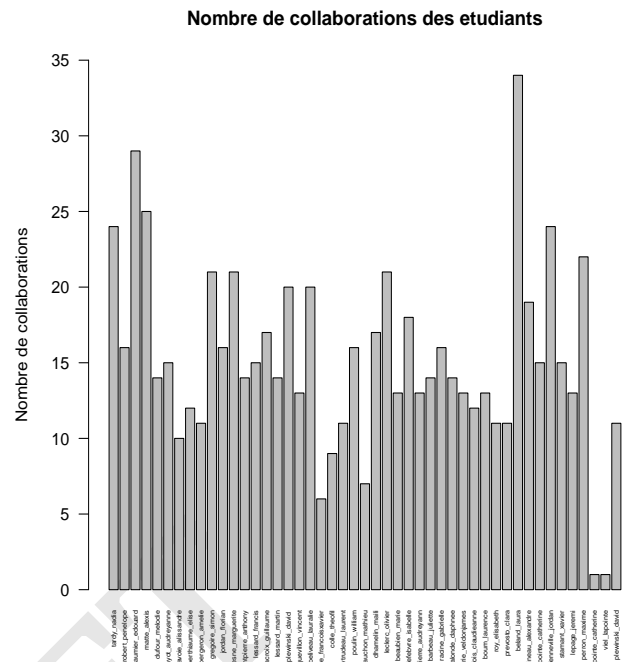


Fig. 2

pas l'hypothèse nulle: les données ont donc une distribution normale.

3. Analyse et discussion

La Figure 1 nous permet d'observer la complexité du réseau de collaboration et la dynamique entre les étudiants.

Dans la figure 2, on remarque que la moyenne semble être autour de 15 collaborateurs. Il semble également y avoir deux étudiants qui se démarque du lot. Nous pouvons les considérer comme des "étudiants clés" potentiels.

L'histogramme de la fréquence du nombre de collaborations des étudiants de BIO500, la figure 3, semble plutôt balancé. La majorité des étudiants de BIO500 ont eu entre 10 et 15 collaborations comme observer dans la figure 2. Par contre, l'allure de la courbe rend notre analyse mitigée. Notre courbe n'est pas tout à fait symétrique, mais son asymétrie vers la gauche n'est pas totalement importante.

L'analyse statistique peut nous en dire plus. Le test de Shapiro-Wilk nous relève un p-value de 0.08508. Puisque qu'il est au-dessus de 0.5 nous ne pouvons donc pas rejeter l'hypothèse nulle: la distribution est donc normale. L'hypothèse de notre question nous est donc confirmée: le nombre de collaborations entre les élèves de BIO500 sont relativement semblables et suivent une distribution normale. Par contre, notre p-value est très proche du seuil d'acceptation. Nous voulons donc pousser notre réflexion plus loin. Pourquoi est-ce que notre réseau est plus ou moins normal selon les standards de nombre de liens entre les espèces? Selon Arthur et ses collaborateurs (3), même en écologie vérifier la normalité du nombre de liens dans un réseau n'est pas tout à fait adéquate. Effectivement, il faut savoir que cette méthode a été développée à l'aide de simulation de distribution de réseaux "parfaits" et sans biais. Alors, cette analyse suppose que les connexions

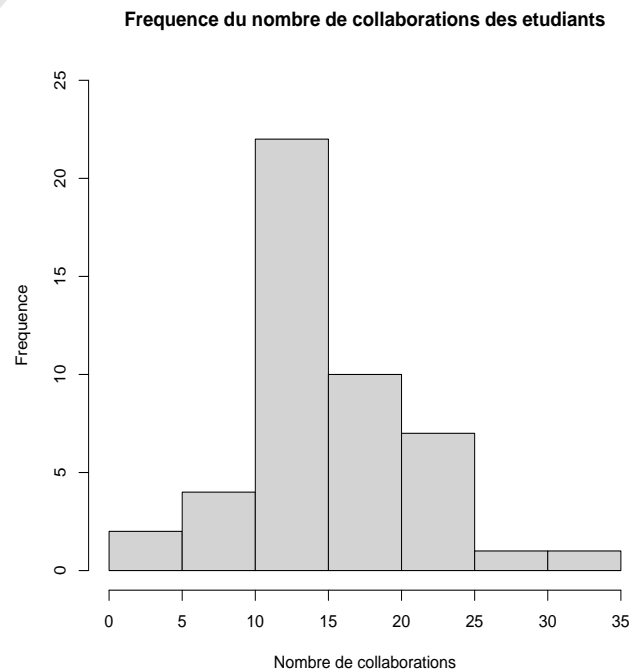


Fig. 3

ont toutes des propriétés fixes dans le réseau. En d’autres mots, la méthode ne prend pas en compte la stochasticité d’un écosystème, dans notre cas, la stochasticité des parcours académiques. Nous avons assumé, tout comme la méthode que nous avons utilisé, que chaque étudiant de BIO500 ont eu le même parcours. Ce n’est effectivement pas le cas. Le cours de BIO500 est le cours obligatoire de 6ème session (3ème années) des étudiants en écologie. Puisque nous sommes en hiver 2022, la majorité des étudiants devraient être rentrer à l’université en automne 2019. Effectivement, 39 étudiants sur 47 sont de la même cohorte. Le reste des étudiants ont des parcours divergents, avec plus ou moins de cours de terminer alors inévitablement un nombre de collaborations différents, c’est-à-dire de la stochasticité.

“Étudiants-clés”

Malgré avoir prouver la normalité de notre distribution, notre curiosité nous a poussé à identifier les deux “étudiants-clé” potentiels: Laura Béland et Édouard Nadon-Beaumier avec 34 et 29 collaborations. On peut considérer ces personnes les plus connectées dans le réseau: ils ont le plus de lien différents avec la totalité du réseau. Dans un réseau écologique, ils seraient les espèces les plus importantes pour garder l’intégrité du système puisqu’ils ont un large étendue de connexions.

1. Fortuna LAW Miguel A. AND Zaman (2013) Evolving digital ecological networks. *PLOS Computational Biology* 9:1–6.
2. Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393(6684):440–442.
3. MacDonald AAM, Banville F, Poisot T (2020) Revisiting the links-species scaling relationship in food webs. *Patterns* 1(7):100079.