

La centralité dans les réseaux de collaboration étudiant ; une comparaison des similitudes avec les réseaux écologiques

Philippe Bourassa^a, Simon Bourgeois^a, Justine Labelle^a, Kayla Trempe-Kay^a

^a Université de Sherbrooke Département de biologie 2500 Boulevard de l'Université Sherbrooke Québec J1K 2R1

Abstract

Les propriétés des réseaux écologiques fournissent des indications sur les interactions écologiques et la résilience du réseau face aux changements et aux perturbations. Parmi ces propriétés, la centralité a été utilisée pour comparer les similarités d'un réseau d'interaction entre les étudiants du cours BIO500 et un réseau écologique. Il a été possible d'identifier les étudiants avec les plus grandes centralités et de les associer à des espèces "clés de voûte" et généralistes. De plus, les étudiants qui ne sont pas visés par l'analyse du réseau de collaboration des étudiants du cours ont tendance à avoir une centralité plus faible. Nous avons aussi tenté d'observer l'influence de la session de début de baccalauréat et la formation préalable des étudiants sur la centralité. Toutefois, aucun de ces facteurs ne l'impacte de manière significative, entre autres dû à la faible taille d'échantillon.

Keywords: Réseau écologique, Collaboration, Centralité

Introduction

Les écologistes ont comme pratique l'étude des interactions entre les espèces qui peut être traduite et analysée à l'aide de réseaux écologiques. Les différentes propriétés de ces réseaux peuvent fournir des indications sur l'organisation des processus écologiques, mais aussi sur la résilience du réseau face aux changements et aux perturbations (Delmas et al., 2019). L'étude des réseaux écologiques peut se faire autant à l'échelle des écosystèmes qu'au niveau des espèces. Dans le dernier cas, ils permettent d'observer le rôle et l'importance de celles-ci à travers leurs interactions au sein du réseau (Delmas et al., 2019). Plusieurs facteurs peuvent influencer ces interactions tels que l'abondance locale des espèces, leurs traits ainsi que les conditions environnementales (Delmas et al., 2019). En tant qu'écologistes si bien formés par l'Université de Sherbrooke, nous avons tenté de déterminer si les propriétés d'un réseau écologique sont similaires à ceux d'un réseau d'interactions entre les étudiants du cours BIO500 à l'hiver 2023. Au cours de leur parcours universitaire, les étudiants ont été confrontés à réaliser de nombreux travaux d'équipes qui représentent des collaborations avec différents individus. Ces collaborations permettent d'analyser les interactions entre les étudiants et les facteurs qui peuvent les influencer. L'objectif est donc d'observer l'importance des étudiants au sein du réseau de collaboration à l'aide de la centralité, mais aussi de comprendre comment l'année du début du baccalauréat et la formation préalable influencent cette centralité.

Méthode

Échantillonnage

Durant l'hiver 2023, les étudiants du cours de BIO500 ont été divisés en 10 équipes de 4 personnes. Chaque équipe a complété trois fichiers csv concernant les cours auxquels ils ont réalisé des travaux d'équipes

*Corresponding author

Email addresses: Philippe.Bourassa4@USherbrooke.ca (Philippe Bourassa), Simon.Bourgeois@USherbrooke.ca (Simon Bourgeois), Justine.Labelle@USherbrooke.ca (Justine Labelle), Kayla.Trempe-Kay@USherbrooke.ca (Kayla Trempe-Kay)

au courant de leur baccalauréat. Le premier fichier permet d'identifier les collaborations effectuées, c'est-à-dire les liens d'interactions avec leurs différents coéquipiers en fonction du cours. Le deuxième fichier représente des informations relatives aux étudiants de l'équipe. Plus spécifiquement, leur région administrative, leur participation au régime coopératif, leur année de début du baccalauréat, leur formation préalable et leur numéro du programme. Puis, le dernier identifie les cours dans lesquels des travaux ont été réalisés et d'autres données pertinentes.

Traitement et analyse des données

Une fois la compilation des données réalisée par chaque équipe, les données ont été partagées à l'ensemble du groupe et celles-ci ont été fusionnées à l'aide du logiciel R sous forme de trois fichiers csv. Par la suite, les données des différents fichiers ont été nettoyées et standardisées afin d'obtenir une conformité au sein des différentes tables. Ces données ont ensuite été injectées à l'aide du logiciel R dans le système de gestion de données SQLite3 (Müller et al., 2023). Ce système nous a permis d'extraire les données d'intérêt via des requêtes qui sont nécessaires à la création des figures. Il est à noter que ce n'est pas toutes les données qui ont été utilisées lors des analyses. Les représentations visuelles du réseau d'interaction, ainsi que la centralité ont été effectuées grâce au package "Igraph", ainsi qu'à l'aide du package "Scales" du logiciel R (Csárdi et al., 2023; Wickham et al., 2022). Le package "Vioplot" du logiciel R a aussi été utilisé pour la création du diagramme en violon (Adler et al., 2022). Ensuite, le package "targets" a été utilisé afin d'automatiser l'ensemble du processus et d'augmenter la reproductibilité de cette étude (Landau et al., 2023a,b). Finalement, le gabarit PNAS dans Rmarkdown a été utilisé grâce au package "Rrticles" (Allaire et al., 2022).

Résultats

La Figure 1 représente le réseau de collaborations des individus du cours BIO500 donné à l'hiver 2023 et leurs collègues. On y observe une variation au niveau de l'importance des étudiants grâce au score de centralité. En effet, la taille des nœuds, qui représente chaque étudiant, varie de manière à ce qu'elle augmente plus le score de centralité est haut, soit proche de 1. Les individus de faible centralité sont retrouvés en marge du réseau alors que les individus présentant une forte centralité sont retrouvés plus au centre. On y décèle également plusieurs occurrences d'interactions entre des étudiants. Effectivement, des lignes plus épaisses sont dénotées dans le réseau.

La Figure 2 permet de visualiser l'impact de l'année de début sur la centralité des individus. Les étudiants NA sont les étudiants qui ne sont pas le cours BIO500 à l'hiver 2023 qui ont effectués des travaux d'équipes avec les étudiants du cours lors de leur parcours. Il est à noter que pour les sessions hiver 2019, été 2021, automne 2021, hiver 2022 et automne 2022, on y retrouve seulement un étudiant ayant débuté son parcours. Ces années ne possèdent donc pas un assez grand échantillonnage pour obtenir des résultats significatifs. On observe que les individus ayant commencé à l'automne 2020 possèdent une médiane similaire aux étudiants ayant commencé à l'hiver 2020. Les étudiants de l'automne 2019 possèdent une médiane plus basse que l'automne 2020 et l'hiver 2020. Les étudiants qui ne sont pas dans le cours possèdent aussi une médiane plus faible que l'automne 2020 et l'hiver 2020, mais elle est similaire aux étudiants de l'automne 2019. Par contre, les différences ne sont pas nécessairement significatives étant donné que les intervalles de 95% s'entrecroisent.

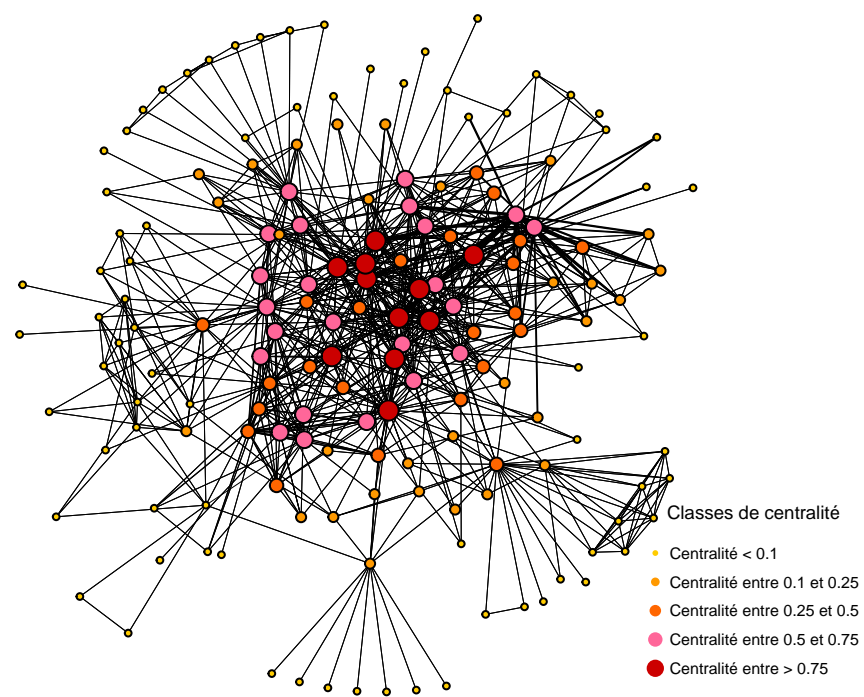


Figure 1: Distribution de la centralité selon la session de début de programme des individus du réseau. La taille des nœuds et leur couleur varient selon la centralité des étudiants tel qu'expliqué par la légende. L'épaisseur des connexions entre les nœuds reflètent l'occurrence d'interactions entre deux étudiants. Plus les lignes sont épaisses, plus les étudiants ont collaboré un grand nombre de fois.

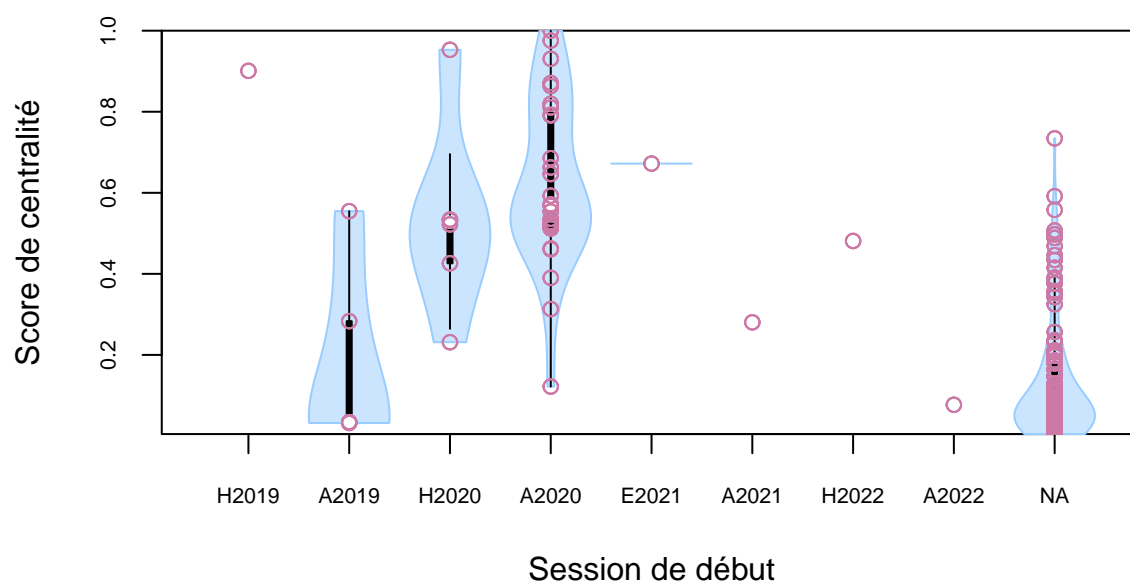


Figure 2: Distribution de la centralité selon la session de début de programme des individus du réseau. Les points mauves dans le graphique correspondent au score de centralité de chaque individu. Le point blanc représente la médiane du groupe et les boîtes noires correspondent à l'intervalle entre les quartiles. La ligne noir plus mince montre l'étendu où 95% des données se situent. Les côtés montrent la distribution des données.

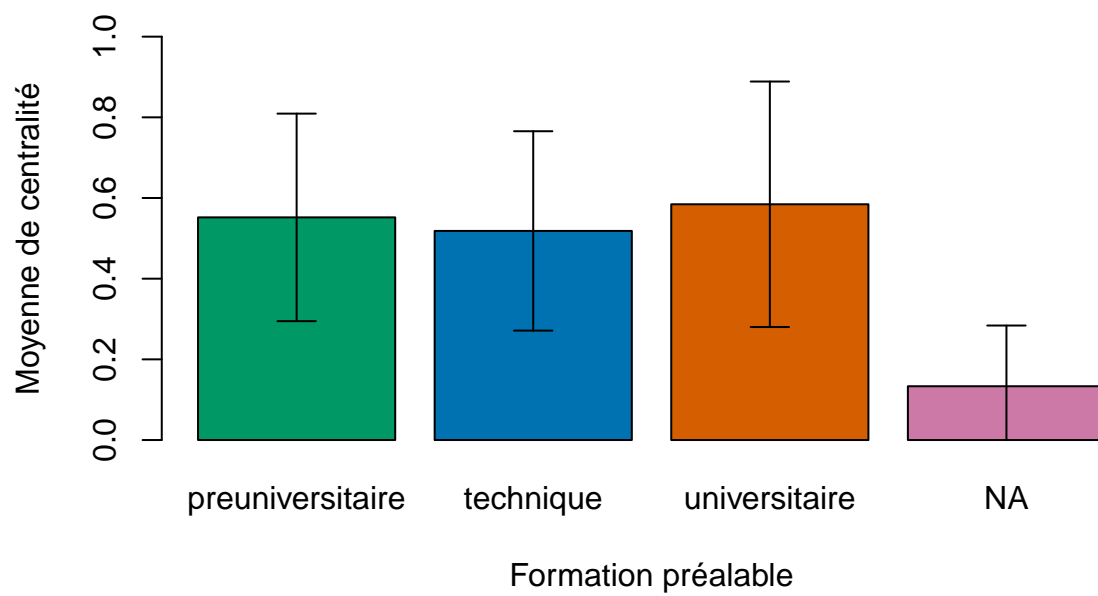


Figure 3: Moyenne de centralité selon la formation préalable des individus du réseau. Les barres d'erreurs correspondent aux écart-types associés à la moyenne de chaque classe soit : préuniversitaire, technique, universitaire et NA (individus hors cours BIO500 hiver 2023).

En ce qui concerne la Figure 3 elle illustre la moyenne de centralité retrouvée pour chacune des formations préalables effectués par les étudiants du cours de BIO500 de l'hiver 2023 par rapport aux individus qui ne sont pas dans le cours. On observe très peu de variation par rapport à la formation préalable au niveau des étudiants au sein du cours, mais celle des étudiants hors du cours est très faible. Cependant, les barres d'erreurs se chevauchent ce qui indique que les différences entre les moyennes ne sont pas significatives.

Discussion

Les propriétés du réseau de collaboration (Figure 1) peuvent être utilisées pour identifier l'importance des individus à l'aide de la centralité. En effet, une grande centralité signifie que les individus ont eu une grande diversité d'interaction, donc des collaborations avec plusieurs personnes différentes. Inversement, les personnes ayant une faible centralité sont celles qui ont eu des interactions très peu diversifiées, donc des collaborations avec les mêmes individus. Cependant, il faut garder en tête que la centralité est influencée par le nombre de collaborations effectué par les étudiants. Par conséquent, il est normal pour les individus hors cours BIO500 à l'hiver 2023 d'avoir une très faible centralité. Les individus avec une grande centralité (ceux en rouge au centre du réseau) peuvent être comparés à des espèces "clés de voûte" d'un réseau écologique en raison que leur disparition compromettrait la structure et le fonctionnement d'un écosystème (Cagua et al., 2019). De plus, cette différence dans la centralité peut aussi être comparée aux individus généraliste et spécialiste d'un réseau écologique (Cagua et al., 2019). Les individus qui ont une faible centralité seraient considérés des spécialistes, tandis que les individus avec une forte centralité seraient considérés des généralistes. Ainsi, la centralité nous confirme que le réseau de collaboration comporte des propriétés similaires à un réseau écologiste, ce qui répond à notre premier objectif. Contrairement à ce qu'on aurait pu s'attendre, l'année de début n'a pas réellement d'impact sur la centralité (Figure 3). L'automne 2020 représente la majorité des étudiants du cours BIO500, ce qui signifie qu'il s'agit des individus ayant fait la plus diversité d'interactions. Ainsi, on s'attendait à ce qu'il ait une plus grande centralité. Toutefois, les étudiants des autres années ont des centralités similaires aux étudiants de l'automne 2020. Ceci peut être expliqué par un manque de données ou simplement que l'année de début n'influence pas la centralité. Il serait possible de le comparer ces individus à des migrants arrivant dans une nouvelle population et qui finissent par s'intégrer parfaitement à la population les rendant indissociables.

Comme l'année de début, la formation préalable n'a pas d'impact sur la centralité (Figure 3). Nous pensions que la majorité des étudiants avait suivi une formation préuniversitaire et que ceux de techniques allaient avoir tendance à travailler ensemble donc être des individus spécialistes. Leur petit nombre et leur spécialisation auraient pu diminuer leur centralité ce qui n'est pas le cas. Contrairement à nos attentes, la formation préalable n'influence pas la centralité, puisque les étudiants provenant de technique ont agi comme des généralistes dans notre réseau. Il est intéressant de noter que la centralité des étudiants qui sont à l'extérieur du réseau semble être beaucoup plus faible comme il est possible de voir dans les Figures 2 et 3. Ce résultat est normal étant donné que ces étudiants ont effectué très peu de collaborations dans le réseau dû au fait qu'ils ne sont pas les individus visés par l'analyse, soit les étudiants du cours de BIO500 à l'hiver 2023. Ils sont donc moins importants que les étudiants qui sont considérés dans le réseau.

Conclusion

En conclusion, les étudiants du cours BIO500 à l'hiver 2023 ont un réseau d'interaction similaire à un réseau écologique. En effet, la centralité a permis d'identifier que les individus avec la plus grande centralité agissent comme des espèces "clés de voûte" en plus d'être considérés comme des généralistes. Toutefois, il n'a pas été possible d'observer les facteurs qui influencent la centralité. La formation préalable et l'année de début n'ont pas eu d'effet significatif sur la centralité. Il serait intéressant de venir observer des variables qui pourraient réellement impacter la dynamique de notre réseau et la centralité des individus telles que la participation aux événements sociaux.

Remerciement

Nous aimerions prendre le temps de remercier Hugo Morin Brassard et Félix-Olivier Dufour qui nous ont grandement aidé lors de nos nombreuses difficultés rencontrées principalement celles dans notre code R.

References

- Daniel Adler, S. Thomas Kelly, Tom M. Elliott, and Jordan Adamson. `vioplot`: Violin Plot, December 2022. URL <https://CRAN.R-project.org/package=vioplot>.
- J. J. Allaire, Yihui Xie, Christophe Dervieux, RStudio, R. Foundation, Hadley Wickham, Journal of Statistical Software, Ramnath Vaidyanathan, Association for Computing Machinery, Carl Boettiger, Elsevier, Karl Broman, Kirill Mueller, Bastiaan Quast, Randall Pruim, Ben Marwick, Charlotte Wickham, Oliver Keyes, Miao Yu, Daniel Emaasit, Thierry Onkelinx, Alessandro Gasparini, Marc-Andre Desautels, Dominik Leutnant, MDPI, Taylor Francis, , Oğuzhan Öğreden, Dalton Hance, Daniel Nüst [aut, cph, Petter Uvesten, Elio Campitelli, John Muschelli, Alex Hayes, Zhian N. Kamvar, Noam Ross [aut, cph, Robrecht Cannoodt (rcannoodt), Duncan Luguern, David M. Kaplan (dmkaplan2000), Sebastian Kreutzer, Shixiang Wang, Jay Hesselberth, Alfredo Hernández, Stefano Coretta (stefanocoretta), Greg Macfarlane (gregmacfarlane), Matthias Templ (matthias da), Alvaro Uzaleta (auzaleta), JooYoung Seo, Callum Arnold (arnold c), Rob Hyndman, Dmytro Perepolkin (dmi3kno), and Tom Palmer (remlapmot). `rticles`: Article Formats for R Markdown, August 2022. URL <https://CRAN.R-project.org/package=rticles>.
- Edgar Fernando Cagua, Kate L Wootton, and Daniel B Stouffer. Keystoneness, centrality, and the structural controllability of ecological networks. *Journal of Ecology*, 107(4):1779–1790, 2019.
- Gábor Csárdi, Tamás Nepusz (See also AUTHORS file.), Vincent Traag, Szabolcs Horvát, Fabio Zanini, Daniel Noom, Kirill Müller, Maëlle Salmon, and Chan Zuckerberg Initiative `igraph` author details. `igraph`: Network Analysis and Visualization, April 2023. URL <https://CRAN.R-project.org/package=igraph>.
- Eva Delmas, Mathilde Besson, Marie-Hélène Brice, Laura A Burkle, Giulio V Dalla Riva, Marie-Josée Fortin, Dominique Gravel, Paulo R Guimarães Jr, David H Hembry, Erica A Newman, et al. Analysing ecological networks of species interactions. *Biological Reviews*, 94(1):16–36, 2019.
- William Michael Landau, Samantha Oliver, Tristan Mahr, and Eli Lilly and Company. `tarchetypes`: Archetypes for Targets, March 2023a. URL <https://CRAN.R-project.org/package=tarchetypes>.
- William Michael Landau, Matthew T. Warkentin, Mark Edmondson, Samantha Oliver, Tristan Mahr, and Eli Lilly and Company. `targets`: Dynamic Function-Oriented 'Make'-Like Declarative Pipelines, March 2023b. URL <https://CRAN.R-project.org/package=targets>.
- Kirill Müller, Hadley Wickham, David A. James, Seth Falcon, D. Richard Hipp (for the included SQLite sources), Dan Kennedy (for the included SQLite sources), Joe Mistachkin (for the included SQLite sources), SQLite Authors (for the included SQLite sources), Liam Healy (for the included SQLite sources), R. Consortium, and RStudio. `RSQLite`: SQLite Interface for R, April 2023. URL <https://CRAN.R-project.org/package=RSQLite>.
- Hadley Wickham, Dana Seidel, and RStudio. `scales`: Scale Functions for Visualization, August 2022. URL <https://CRAN.R-project.org/package=scales>.