

# graph4lg: un package R pour construire et analyser des graphes en écologie du paysage et génétique des populations

Paul Savary\*

29 avril 2021

## Résumé

Modéliser la connectivité écologique des habitats et son influence sur la biodiversité est nécessaire pour améliorer la compréhension des processus écologiques et guider les mesures de conservation de la biodiversité. Cela requiert d'analyser des données spatiales afin de cartographier les espaces contribuant à cette connectivité. Les réseaux d'habitat des espèces étudiées peuvent être représentés sous la forme de graphes paysagers dont les nœuds correspondent aux taches d'habitat et les liens aux chemins de dispersion potentiels entre ces taches [Urban and Keitt, 2001]. Ces graphes permettent de calculer des métriques pour guider la gestion des espaces naturels ou pour des analyses statistiques ultérieures [Baranyi et al., 2011]. En effet, les graphes paysagers peuvent être confrontés à des données empiriques issues du terrain pour (i) valider leur réalisme écologique ou (ii) tester des hypothèses sur le lien entre la connectivité des habitats et des réponses biologiques [Foltête et al., 2020]. Les données génétiques (marqueurs microsatellites, SNPs) sont particulièrement adaptées à ces analyses dans la mesure où la structure génétique des populations est influencée par la connectivité de leurs habitats [Keyghobadi, 2007]. Ces données peuvent également être analysées sous forme de graphes génétiques, dont les nœuds correspondent à des populations composées de plusieurs individus génotypés et les liens aux échanges génétiques les plus significatifs entre populations [Greenbaum and Fefferman, 2017, Savary et al., 2021b]. Ces liens sont généralement pondérés par la distance génétique entre populations.

Des progrès conséquents ont amélioré l'utilisation des graphes génétiques et paysagers mais un outil informatique réunissant une large gamme de paramètres de construction et d'analyse de ces deux types de graphes faisait défaut. Par ailleurs, malgré la complémentarité de ces graphes, peu de méthodes d'analyse permettent de les comparer. Nous avons donc développé un package R pour faciliter et encourager l'utilisation de ces graphes [Savary et al., 2021a].

Il intègre des fonctions dédiées à la conversion et à l'import de données génétiques à partir des formats de données utilisés par les principaux logiciels et packages R de génétique des populations (`gstudio`, `pegas`, `genepop` ou `STRUCTURE` [Pritchard et al., 2000]). Il permet à l'utilisateur de calculer différentes distances génétiques, des distances géodésiques, mais aussi des distances-coût, censées intégrer la résistance du paysage au déplacement des espèces. Ce dernier type de distance se base sur des données spatiales ponctuelles et une couche raster représentant l'occupation du sol de la zone d'étude.

Une gamme importante de paramètres peuvent être utilisés pour créer les graphes génétiques et paysagers. En particulier, les différentes méthodes d'élagage se basent sur des seuils de distances génétiques ou paysagères, sur des contraintes topologiques ou sur des tests statistiques utilisant le principe d'indépendance conditionnelle. Nous avons également rendu accessible le logiciel Graphab [Foltête et al., 2012] aux utilisateurs de R pour faciliter la construction et l'analyse de graphes paysagers dans cet environnement. L'utilisation de programmes codés en Java à l'aide de R permet l'intégration de Graphab au package est sans que l'utilisateur ait à changer d'environnement.

---

\*ARP-Astrance, Paris; UMR 6049 ThéMA, Université Bourgogne Franche-Comté - CNRS, Besançon; UMR 6282 Biogéosciences, Université Bourgogne Franche-Comté - CNRS, Dijon [paul.savary@univ-fcomte.fr](mailto:paul.savary@univ-fcomte.fr)

Par ailleurs, les fonctions du package réalisent des analyses préliminaires ayant pour but d'adapter les méthodes de construction des graphes aux questions de recherche. Les graphes génétiques et paysagers ainsi créés peuvent être analysés avec des métriques calculées au niveau des nœuds telles que des indices de centralité ou des indices spécifiques aux graphes paysagers. Ces graphes peuvent également être partitionnés en modules à l'aide de plusieurs algorithmes de maximisation de la modularité. Les utilisateurs peuvent ensuite visualiser et/ou cartographier les résultats de ces analyses.

Une des principales originalités de `graph4lg` repose sur ses fonctions dédiées à la comparaison de graphes partageant les mêmes nœuds. Cette comparaison s'effectue au niveau des nœuds en étudiant la relation entre les métriques calculées, au niveau des liens en comparant les topologies des graphes ou en étudiant la relation entre des matrices d'adjacence pondérées, et en comparant les partitions en modules de deux graphes. L'utilisateur peut également exporter les graphes sous forme de couches shapefile pour faciliter leur intégration à un Système d'Information Géographique (SIG) et les analyses ultérieures. `graph4lg` contribue ainsi au potentiel des graphes génétiques et paysagers pour l'analyse de la connectivité écologique des habitats, tout en encourageant de futures recherches sur les aspects méthodologiques relatifs à ces outils.

**Mots-clés :** R – Connectivité écologique – Théorie des graphes – Génétique des populations – Écologie du paysage

## Références

- Gabriella Baranyi, Santiago Saura, János Podani, and Ferenc Jordán. Contribution of habitat patches to network connectivity : redundancy and uniqueness of topological indices. *Ecological Indicators*, 11(5) :1301–1310, 2011.
- Jean-Christophe Foltête, Céline Clauzel, and Gilles Vuidel. A software tool dedicated to the modelling of landscape networks. *Environmental Modelling & Software*, 38 :316–327, 2012.
- Jean-Christophe Foltête, Paul Savary, Céline Clauzel, Marc Bourgeois, Xavier Girardet, Yohan Sahaoui, Gilles Vuidel, and Stéphane Garnier. Coupling landscape graph modeling and biological data : a review. *Landscape Ecology*, pages 1–18, 2020.
- Gili Greenbaum and Nina H Fefferman. Application of network methods for understanding evolutionary dynamics in discrete habitats. *Molecular Ecology*, 26(11) :2850–2863, 2017.
- Nusha Keyghobadi. The genetic implications of habitat fragmentation for animals. *Canadian Journal of Zoology*, 85(10) :1049–1064, 2007.
- Jonathan K Pritchard, Matthew Stephens, and Peter Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2) :945–959, 2000.
- Paul Savary, Jean-Christophe Foltête, Hervé Moal, Gilles Vuidel, and Stéphane Garnier. `graph4lg` : a package for constructing and analysing graphs for landscape genetics in r. *Methods in Ecology and Evolution*, 12(3) :539–547, 2021a. doi : <https://doi.org/10.1111/2041-210X.13530>. URL <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13530>.
- Paul Savary, Jean-Christophe Foltête, Hervé Moal, Gilles Vuidel, and Stéphane Garnier. Analysing landscape effects on dispersal networks and gene flow with genetic graphs. *Molecular Ecology Resources*, 21(4) :1167–1185, 2021b. doi : <https://doi.org/10.1111/1755-0998.13333>.
- Dean Urban and Timothy Keitt. Landscape connectivity : a graph-theoretic perspective. *Ecology*, 82(5) :1205–1218, 2001.