Short communication

Linking habitats for multiple species

Raul Brás [a], J. Orestes Cerdeira [b,c,*], Diogo Alagador [d], Miguel B. Araújo [d,e,f]

[a] Centre for Applied Mathematics and Economics (CEMAPRE), Instituto Superior de Economia e Gestão, Technical University of Lisbon (TULisbon), Rua do Quelhas 6, 1200-781 Lisbon, Portugal
[b] Department of Mathematics, Faculdade de Ciências e Tecnologia, New University of Lisbon, Campus de Caparica, 2829-516 Caparica, Portugal
[c] Forest Research Centre (CEF), Instituto Superior de Agronomia, Technical University of Lisbon (TULisbon), Tapada da Ajuda 1, 1349-017 Lisboa, Portugal
[d] Rui Nabeiro Biodiversity Chair, CIBIO, University of Évora, Casa Cordovil 2° Andar, Rua Dr. Joaquim Henrique da Fonseca, 7000-890 Évora, Portugal
[e] Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales, CSIC, C/ José Gutiérrez Abascal 2, 28006 Madrid, Spain
[f] Center for Macroecology, Evolution and Climate, University of Copenhagen, DK-2100, Denmark

ARTICLE INFO

ABSTRACT

The establishment of linkages between habitats is of great importance to avert the detrimental impacts of land fragmentation and climate change on biodiversity. Linkages need to be cost-efficient, and should account for specific dispersal requirements of species. Since cost-efficient linkages defined independently for each individual species are more costly than linkages optimised for multiple species, there is need for methods specifically designed to retrieve efficient linkages for multiple species. *MulTyLink* (Multiple Type Linkages) is a C++ open source program that defines cost-efficient linkages free of barriers for the species considered, and that allows species-specific dispersal requirements to be considered. Here we present, discuss and illustrate the algorithms used by *MulTyLink* to identify cost-efficient linkages for multiple species.

Software availability

Name of the software: *MulTyLink* − Multiple Type Linkages
Developers: Raul Brás, J. Orestes Cerdeira, Diogo Alagador
Contact address: Instituto Superior de Economia e Gestão,
        Technical University of Lisbon (TULisbon), Rua do Quelhas
        6, 1200-781 Lisbon, Portugal
Email: rbars@iseg.utl.pt
Supported systems: Windows OS, Linux, Macintosh
Programming language: C++ and Qt libraries
Available since: October, 2011
Availability: Free download with supporting material at http://purl.oclc.org/multylink
License: GPL v3

* Corresponding author. Forest Research Centre (CEF), Instituto Superior de Agronomia, Technical University of Lisbon (TULisbon), Tapada da Ajuda 1, 1349-017 Lisboa, Portugal.
    E-mail addresses: rbras@iseg.utl.pt (R. Brás), orestes@isa.utl.pt (J.O. Cerdeira), alagador@uevora.pt (D. Alagador), maraujo@mncn.csic.es (M.B. Araújo).

1. Introduction

Habitat fragmentation is a key driver of biodiversity loss (Brooks et al., 2002; Hanski, 2005). Linking once connected natural habitats is imperative to maintain biological diversity (Luque et al., 2012). However, given the limited resources available for conservation and the existing conflicts between conservation and exploitative uses, location-allocation efficiency is a desirable property of any conservation plan (Pressey and Nicholls, 1989).

Graph theory is recognised as being a convenient framework to incorporate spatial criteria into conservation planning (Urban and Keitt, 2001; Fall et al., 2007), and a number of methods use graph theory for identifying cost-efficient areas to promote connectivity between habitats. Some of these methods address connectivity and species' representation simultaneously. Examples are the selection of areas to achieve representation targets for species, forming a unique contiguous network (Önal and Briers, 2005; Cerdeira et al., 2005; Fuller et al., 2006; Cerdeira et al., 2010), or permitting less strict forms of spatial coherence, selecting areas that are spatially clustered but not necessarily connected by contiguous corridors (Önal and Briers, 2002, 2003; Alagador and Cerdeira, 2007). Other methods solely involve connecting existing sets of reserves or

habitats. Table 1 reports open-source software implementing these latter approaches.

Some of the reported applications deliver a unique (or a small set of) "best" linkage(s) (i.e., the areas promoting connectivity between habitats), while others produce a quantitative evaluation of how suitable (or probable) each area is for the established connectivity goals.

None of the above methods was specifically designed to efficiently link habitats occupied by multiple species, presenting distinct distribution patterns, i.e., occurring in distinct sets of suitable isolated habitats, and having different dispersal constraints, i.e., distinct dispersal distances and dissimilar suitable areas to disperse. One size fits all would not be a valid approach since areas suitable for the dispersal of some species may be barriers for others.

Although the existing methods for identification of linkages can be used for multiple species (by merging individual species solutions into a single solution), the resulting configuration would most probably be far from a minimum cost (or area) solution. To overcome this limitation Lai et al. (2011) used graphs to take into account species dispersal specificities and proposed procedures which they applied to trace links for populations of wolverines and Canadian lynx in Western Montana. Alagador et al. (2012) also developed an approach to this problem, which they applied to link Iberian Peninsula protected areas, clustered in four environmental-similar classes.

We introduce *MulTyLink*, a software to efficiently link the habitats occupied by multiple species with distinct distributions and/or dispersal requirements, and specifically designed to deal with large data sets. *MulTyLink* constructs a graph for each group (of "similar") species, taking into account the areas acting as barriers and the dispersal capacities of these species. When selecting areas for a group of species in a graph, *MulTyLink* deems the possibility of using these areas for other groups, thus reducing costs and the number of selected areas. After having obtained a solution ensuring the connection of all the habitats for each group of species, the last step of *MulTyLink*'s procedure consists of removing all areas whose removal would not affect the required linkages for all species in analysis.

2. Methods

It is assumed that the study region is divided into cells. Cells that for whatever reason are not suitable for conservation action are filtered out from the analysis, leaving a set T of candidate cells for usage as linkage units.

For each (group of "similar") species, k ($k = 1, 2, …, m$), a subset of cells T_k, called terminal cells of type k, represent the habitats where the species occurs and that need to be linked. A distance threshold, d_k, is used to define adjacency rules between cells (i.e., cells distancing d_k or less from a cell t are considered directly reachable from t), for species k. In addition, a set of cells, B_k, is given to identify the dispersal barriers for species k. With this information a graph is constructed for each species k, where nodes are cells of $T\backslash B_k$ and the existence of an edge (u, v), connecting cells u and v, means that the distance between u and v is less than or equal to d_k (expressing that species k can move directly between cells u and v).

Solutions are sets of cells of T such that, on the graph of each species k, the terminal cells in T_k are linked. We call these solutions feasible linkages.

A cost is associated to every non-terminal cell, and the goal is to find a minimum cost feasible linkage (i.e., minimising the sum of costs of non-terminal cells). If all costs are equal to one, then the number of cells is minimised.

Table 1
Open-source software for linking natural habitats.

Software	Reference
Circuitscape	McRae and Beier (2007)
	McRae and Shah (2011)
Conefor Sensinode	Saura and Torné (2009)
Connectivity Analysis Toolkit	Carroll et al. (2012)
Linkage Mapper	McRae and Shah (2011)
LQGraph	Fuller and Sarkar (2006)
UNICOR	Landguth et al. (2012)

If only one type of terminal cells exists and all species have the same dispersal barriers and the same adjacency rule, i.e., $m = 1$, finding a minimum cost feasible linkage is the node-weighted version of the well known minimum Steiner tree problem in graphs (Hwang et al., 1992). Minimum Steiner trees were introduced in the context of conservation ecology by Sessions (1992). Later, Williams (1998) proposed a minimum Steiner tree flow model to link existing reserves in order to simultaneously maximise cost-efficiency and minimise unsuitable area. Solving minimum Steiner tree problems is the key procedure of several software presented in Table 1. There are many variants and extensions of the minimum Steiner tree problem (see Du and Hu, 2008). To identify minimum cost linkages for multiple species, with distinct distribution patterns and distinct dispersal requirements is another generalization of the node-weighted minimum Steiner tree (and of the node-weighted minimum Steiner forest) problem(s).

To solve this problem we developed two alternative heuristics: *Type by Type* and *Grasp*.

2.1. Type by Type

For a given permutation $P = (t_1, t_2, …, t_m)$ of $\{1, 2, …m\}$, the *Type by Type* algorithm starts by linking the terminal cells of T_k on the graph of species k, with $k = t_1$, using (a node-weighted adaptation of) the minimum spanning tree approximation to the minimum Steiner tree problem (Vazirani, 2003, chap. 2). The minimum spanning tree approximation consists of finding a minimum spanning tree on a complete graph whose nodes are terminals and where the weight of every edge (u, v) is the cost of the minimum cost path connecting terminal u to v. Every edge of the minimum spanning tree is replaced by the corresponding minimum cost path, and a minimal solution is obtained by sequentially removing non-terminal nodes, whenever removal do not disconnect terminals.

Next, the cost of every cell used to link terminals of T_k, with $k = t_1$, is redefined to zero, and the linkage of terminals of T_k, with $k = t_2$ is determined using the above procedure. The procedure is repeated for types $k = t_3, …, t_m$ Finally, the solution consisting of all linkages is pruned from redundant cells (i.e., cells whose removal do not increase the number of components of any type) until the solution becomes minimal.

When running the algorithm with different permutations, *Type by Type* outcomes a solution with the smallest cost among the obtained solutions.

The *Type by Type* algorithm was successfully tested by Alagador et al. (2012) to identify cost-efficient linkages between environmentally-similar protected areas in the Iberian Peninsula. A similar approach, with a different strategy to link the terminals of T_k, was also proposed by Lai et al. (2011).

2.2. Grasp

Grasp is a heuristic of the general family known as Greedy Randomized Adaptive Search Procedure (Feo and Resende, 1995).

Starting with the (unfeasible) solution consisting only of all terminal cells, at each iteration *Grasp* links a pair of not yet linked terminals of the same type. The type of terminals to link is uniformly selected among the types for which terminals are not connected yet. To link terminals of the chosen type k, a terminal s_k is uniformly selected from T_k and the path of minimum cost among the minimum cost paths (excluding barriers) between s_k and every other terminal of T_k, not yet linked with s_k, is determined. The path is added to the current solution, and the costs of its cells are redefined to zero. When all terminals of the same type are linked (or when no more linkages are possible), the same procedure implemented in the *Type by Type* approach eliminates redundant cells (i.e., cells whose removal do not increase the number of components of any type) until the solution becomes minimal.

Given the randomized nature of *Grasp*, it is expectable to obtain different solutions from different runs. *Grasp* selects a solution of minimum cost among the solutions obtained from an arbitrary predefined number of runs.

It is likely that when the number of runs on *Grasp* is equal to the number of permutations on *Type by Type*, solutions obtained from *Grasp* have smaller costs than those produced with *Type by Type*. However, *Grasp* is computationally more demanding.

It should be noted that identifying minimum cost linkages, even when only one type of habitat exists, and species have common barriers and the same dispersal rule (i.e., the node-weighted version of the minimum Steiner tree problem), is a difficult, i.e., NP-hard problem (Garey and Johnson, 1979). It is widely conjectured that no efficient (i.e., polynomially-time) algorithm exists for NP-hard problems. Therefore, the solutions produced by *Type by Type* and *Grasp* have no guarantee of being optimal, and for large instances (say a hundred thousand of cells and a dozen of different types of terminals) the algorithms can take long time to find a solution, and may even abort due to space overflow.

3. Program description

MulTyLink is a dialog-based application integrating mapping capabilities and visualization routines, that implements *Type by*
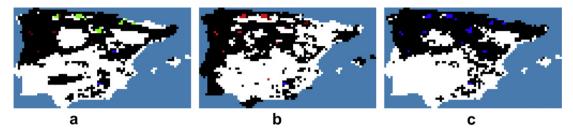
Fig. 1. Maps depicting the protected habitats and barriers (white cells) for three threatened reptile species in Iberian Peninsula. (a) Green cells represent the habitats for *Lacerta bilineata*, Western green lizard (blue and dark red cells are habitats for the other species); (b) Red cells represent the habitats for *Lacerta schreiberi*, the Iberian emerald lizard (blue and dark red cells are habitats for the other species); (c) dark blue cells represent the habitats for *Coronella austriaca*, Smooth snake. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Type and *Grasp* to identify cost-efficient linkages for multiple species.

MulTyLink requires an input file with geographical information for mapping purposes and the identification of terminal cells, friction values (cells with friction values greater than a given threshold are excluded from consideration), costs associated to non-terminal cells and, for each type k, the barriers and an adjacent distance d_k. The value d_k defines which cells are considered adjacent when linking terminals of T_k (these distance values are expressed in step units). Cells having centres within a circle of radius d_k centred on cell j are adjacent to j. Typically, cells are squares from a grid, and setting $d_k = 1$ defines two cells adjacent if they have a common edge in the grid. Setting $d_k = \sqrt{2}$ makes two cells adjacent if they have a common edge or corner. When $d_k \geq 2$, some pairs of cells with no common edges and corners are also considered adjacent. In this case, optimal feasible linkages will look like stepping stones.

The spatial distribution of barriers and high friction cells will strongly constrain the shape of feasible linkages, and may even prevent terminals of the same type to be linked. *MulTyLink* delivers solutions with the minimum possible number of linkage fragments for each type. That is, defining a *connected component* as the (maximal) linkage of terminals of the same type, *MulTyLink* computes the minimum number of connected components for each type. This number is indicative of how fragmented are habitats of each type, k.

4. Example

Here we exemplify how *MulTyLink* operates using a simple example where the habitats occupied by three threatened reptile species (*Lacerta bilineata*, the Western green lizard; *Lacerta schreiberi*, the Iberian emerald lizard; and *Coronella austriaca*, the Smooth snake) within Iberian Peninsula protected areas are to be linked.
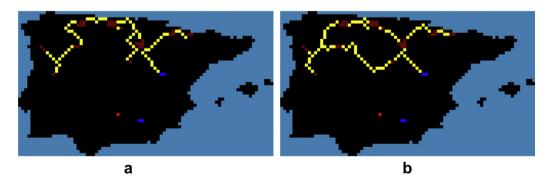
Terminals were defined as those cells, among 2310 $10' \times 10'$ (c. 18 km lat. × 15 km long.) cells representing the Iberia Peninsula, where any of the three species occurs and with more than 75% of surface area protected (Fig. 1).

Barriers for each species were identified from ensembles of bioclimatic models (Araújo and New, 2007), obtained using a small set of climatic variables, quantifying the climatic suitability of each cell (see Araújo et al., 2011; for a detailed description of the methods). The 50% of the cells with the lowest climatic suitability for each of the species were taken as barriers for the movement of that species (Fig. 1). Moreover, for each species, adjacency rules were settled to $\sqrt{2}$, indicating the need to define "continuous" linkages (see above the effect of d_k over solutions).
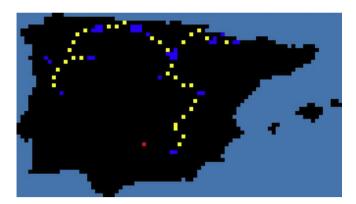
When a cost of 1 was assigned to every non-terminal cell the linkages obtained with the *Grasp* and *Type by Type* algorithms (using default parametrisations) included 68 and 72 cells, respectively (Fig. 2a).

This specific formulation (all cells with equal cost) seeks to minimise the number of linkage sites. Alternatively, costs can be heterogeneously distributed among cells, each cell having a cost that quantifies the value incurred to allocate it to conservation purposes. We used cell coverage by Natura 2000 network (European Community Directive 92/43/EEC) to assign costs to cells, defining the cost of a cell as the fraction of the cell's surface not covered by Natura 2000 areas. With these costs, *Grasp* produced a solution with total cost of 40.48 (i.e., the equivalent of 37.52 cells totally covered by Natura 2000), and 78 cells (Fig. 2b). Solutions from *Type by Type* displayed a total cost of 40.95 and 79 cells. In this case linkages are longer than when the algorithms were used simply to minimise the number of cells.

Solutions were obtained in 1.2 (1.3) and 0.13 (0.13) elapsed time seconds, for homogeneous (heterogeneous) costs, with *Grasp* and *Type by Type*, respectively, on a linux (Kubuntu) 64-bit machine with Intel Core 2 Quad Q9550 2.83 Ghz CPU.

Fig. 2. Optimised linkages (yellow cells) for the three reptile species (dark red cells are habitats for more than one species; the red cell is habitat for the Iberian emerald lizard; the dark blue cells are habitats for the Smooth snake) occurring in Iberian Peninsula protected areas. (a) *Grasp* solution obtained by setting costs equal to 1: 68 cells were selected; (b) *Grasp* solution obtained using heterogeneous costs measured as function of the existence of the Natura 2000 areas in the cell: 78 cells were selected that encompass the equivalent to 37.52 cells with Natura 2000. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 3. Optimised linkages (yellow cells) to connect Smooth snake's occurring habitats (dark blue cells) in Iberian Peninsula protected areas, obtained with the *Grasp* heuristic, using Natura-coverage to define costs and establishing adjacency distances of $\sqrt{2}$ to Western green lizard and Iberan emerald lizard and $\sqrt{8}$ for Smooth snake. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

It is important to note that, because of the location of barriers (Fig. 1) and the dispersal distance rule ($d_k = \sqrt{2}$, for the three species), not all terminals occupied by the same species were connected (Fig. 2). Solutions for the Iberian emerald lizard and for the Smooth snake presented two connected components, one of them composed only by terminal cells (i.e., no linkage cells). Terminal cells occupied by the Western green lizard were linked into one connected component in both solutions.

Setting Smooth snake's linkage adjacencies to $\sqrt{8}$ (i.e., two cells are adjacent if there is no more than one cell between them) allowed the linkages for this species to define one connected component (Fig. 3).

When grid cell's costs were defined by Natura 2000 coverage, 78 was also the required number of cells to establish these less fragmented linkages when applying *Grasp* (80 cells when using *Type by Type*), but now passing through the equivalent of 36.59 cells totally covered by Natura 2000 (37.58 cells for *Type by Type*).

We also resettled the original example to compare solutions produced by *MulTyLink* with the outcomes of merging the solutions obtained independently for each of the three species. The "species-by-species" approach, using *Grasp* algorithm for each species, found a solution with 109 cells (when minimising the number of cells) and with a cost of 52.57 (when minimising cost), which represents 60.3% more cells and 29.9% higher cost than the "three-in-one" *MulTyLink* solutions obtained with *Grasp*. Using *Type by Type*, the "species-by-species" minimum cells' solution contained 118 cells and the minimum cost solution had a cost of 49.03, representing 63.9% more cells and 19.7% higher cost than *MulTyLink* solutions obtained with *Type by Type*.

References

Alagador, D., Cerdeira, J.O., 2007. Designing spatially-explicit ecological reserve networks in the presence of mandatory sites. Biological Conservation 137, 254–262.

Alagador, D., Triviño, M., Cerdeira, J.O., Brás, R., Araújo, M.B., 2012. Linking like with like: optimising connectivity between environmentally-similar habitats. Landscape Ecology 27, 291–301.

Araújo, M.B., Alagador, D., Cabeza, M., Nogués-Bravo, D., Thuiller, W., 2011. Climate change threatens European conservation areas. Ecology Letters 14, 484–492.

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. Trends in Ecology and Evolution 22, 42–47.

Brooks, T.M., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Rylands, A.B., Konstant, W.R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G., Hilton-Taylor, C., 2002. Habitat loss and extinction in the hotspots of biodiversity. Conservation Biology 16, 909–923.

Carroll, C., McRae, B.H., Brookes, A., 2012. Use of linkage mapping and centrality analysis across habitat gradients to conserve connectivity of gray wolf populations in western North America. Conservation Biology 26, 78–87.

Cerdeira, J.O., Gaston, K.J., Pinto, L.S., 2005. Connectivity in priority area selection for conservation. Environmental Modelling and Assessment 10, 183–192.

Cerdeira, J.O., Pinto, L.S., Cabeza, M., Gaston, K.J., 2010. Species specific connectivity in reserve-network design using graphs. Biological Conservation 143, 408–415.

Du, D., Hu, X., 2008. Steiner Tree Problems in Computer Communication Networks. World Scientific Publishing Company.

Fall, A., Fortin, M.-J., Manseau, M., O'Brien, D., 2007. Spatial graphs: principles and applications for habitat connectivity. Ecosystems 10, 448–461.

Feo, T.A., Resende, M.G.C., 1995. Greedy randomized adaptive search procedures. Journal of Global Optimization 6, 109–133.

Fuller, T., Mungua, M., Mayfield, M., Sánchez-Cordero, V., Sarkar, S., 2006. Incorporating connectivity into conservation planning: a multi-criteria case study from central Mexico. Biological Conservation 133, 131–142.

Fuller, T., Sarkar, S., 2006. LQGraph: a software package for optimizing connectivity in conservation planning. Environmental Modelling and Software 21, 750–755.

Garey, M.R., Johnson, D.S., 1979. Computers and Intractability: a Guide to the Theory of NP Completeness. W. H. Freeman & Company, San Francisco.

Hanski, I., 2005. The Shrinking World: Ecological Consequences of Habitat Loss. In: Excellence in Ecology, vol. 14. International Ecology Institute, Oldendorf/Luhe, Germany.

Hwang, F., Richards, D., Winter, P., 1992. The Steiner Tree Problem. In: Annals of Discrete Mathematics. North-Holland Mathematics Studies, Amsterdam, The Netherlands.

Lai, K.J., Gomes, C.P., Schwartz, M.K., McKelvey, K.S., Calkin, D.E., Montgomery, C.A., 2011. The Steiner multigraph problem: wildlife corridor design for multiple species. In: Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI-11); San Francisco, CA, USA; August 7–11, 2011. AAAI Press, 8 p.

Landguth, E.L., Hand, B.K., Glassy, J., Cushman, S.A., Sawaya, M.A., 2012. Unicor: a species connectivity and corridor network simulator. Ecography 35, 9–14.

Luque, S., Saura, S., Fortin, M.-J., 2012. Landscape connectivity analysis for conservation: insights from combining new methods with ecological and genetic data. Landscape Ecology 27, 153–157.

McRae, B.H., Beier, P., 2007. Circuit theory predicts gene flow in plant and animal populations. Proceedings of the National Academy of Sciences 104, 19885–19890.

McRae, B.H., Shah, V.B., 2011. Circuitscape User Guide. The University of California, Santa Barbara, CA, USA.

Önal, H., Briers, R.A., 2002. Incorporating spatial criteria in optimum reserve network selection. Proceedings of the Royal Society of London Series B-Biological Sciences 269, 2437–2441.

Önal, H., Briers, R.A., 2003. Selection of minimum-boundary reserve network. Proceedings of the Royal Society of London Series B-Biological Sciences 270, 1487–1491.

Önal, H., Briers, R.A., 2005. Designing a conservation reserve network with minimal fragmentation: a linear integer programming approach. Environmental Modelling and Assessment 10, 193–202.

Pressey, R.L., Nicholls, A.O., 1989. Efficiency in conservation evaluation: scoring versus iterative approaches. Biological Conservation 50, 199–218.

Saura, S., Torné, J., 2009. Conefor sensinode 2.2: a software package for quantifying the importance of habitat patches for landscape connectivity. Environmental Modelling and Software 24, 135–139.

Sessions, J., 1992. Solving for habitat connections as a Steiner network problem. Forest Science 38, 203–207.

Urban, D., Keitt, T., 2001. Landscape connectivity: a graph-theoretic perspective. Ecology 82, 1205–1218.

Vazirani, V.V., 2003. Steiner forest. In: Approximation Algorithms. Springer–Verlag, pp. 197–211.

Williams, J.C., 1998. Delineated protected wildlife corridors with multi-objective programming. Environmental Modelling and Assessment 3, 77–86.