Application of network theory to mark recapture data allows insights into population structure of two *Heliconius* species

Luciana L.F. de Lima¹, Gilberto Corso², Márcio Z. Cardoso³

¹Programa de Pós-Graduação em Ecologia, Universidade Federal do Rio Grande do Norte, Natal, Brazil
²Departamento de Biofísica e Farmacologia, Universidade Federal do Rio Grande do Norte, Natal, Brazil
³Departamento de Ecologia, Universidade Federal do Rio Grande do Norte, Natal, Brazil

E-mail: mzc@cb.ufrn.br, marciozikan@gmail.com

Received December 5 2014; Accepted January 10 2015; Published online June 1 2015

Abstract

By noting the spatial location of captured individuals mark-recapture studies create a collection of discrete events spread in space and time. This setup is appropriate for network modeling where the vertices (or nodes) are the points of capture and links are established whenever a recapture occurs. Applying network analytical tools, it is possible to ascertain aspects of spatial structure and generate predictions regarding the likely causes of structure in the network. We studied the spatial network of two tropical butterfly species, *Heliconius erato* and *H. melpomene*, using a mark-recapture database from a 2-year survey in an Atlantic Forest remnant in Brazil. The overall network structure of both species was similar in number of vertices and average connectivity. *Heliconius erato* had a smaller, more disconnected network structure, suggesting shorter traveling paths. The distribution of connectivity of both species was better adjusted by a power-law distribution. We found hubs in both species; hubs are points of high capture and their location is correlated with the location of flowering plants visited by adults. In complex systems, hub elimination can have a notable collapsing effect in network structure. Because resource hubs are important for butterfly network organization we suggest management as well as experimental tests with regards to the role of resource hotspots for population structure.

Keywords: network theory; functional habitat; butterflies; Atlantic forest; movement ecology.

1 Introduction

Studies of networks seek to describe and understand patterns related to the way discrete entities interact in a dimensionless space (Albert and Barabási, 2002). Application of network theory in areas of biology is increasingly common (May, 2006; Zhang, 2011, 2012a, 2012b). One relevant example of network thinking in population biology is the analysis of pattern of infection and disease in injected-drug users (Neaigus et al., 1994;
The elements that compose such a network are made of users and patterns of friendship and needle exchange can be used to link any two participants, thus creating a network. The intensity of any interaction between two units can be expressed by the width of a link. Subsequently, disease spread can be modeled using varying degrees of network construction that can direct public health policies on disease control. In the needle exchange example, one could look for well-connected users for instance, to follow the path of the disease.

Network structure can be random, homogeneous or complex, and one way of assessing the general this is to use the degree distribution of the number of interactions (or links) between the individual components of the network (Zhang, 2012a, 2012c). Links in a random network are stochastically established among participants and could be a useful null model for analysis of network structure (Bornholdt and Schuster, 2002). In homogeneous networks, every participant has the same linkage density as the remaining participants in the network. In complex networks link frequency distributions are well described by a power function, with a straight line in a log plot. In these kinds of networks, a few components have a disproportionally higher number of connections than the rest of the network. Complex networks strike an intuitive appeal in ecology because it is easy to recognize the existence of key species in interaction networks (Paine, 1966; Gilbert, 1980) but also elements of the habitat that contain resources that are used by a significant portion of the population, such as nesting grounds and water holes. In network terminology, these are called hubs. Hubs are not seen either in homogeneous or in random networks and are important elements in network stability. It is hypothesized that complex networks can be formed via a mechanism called preferential attachment, where new links tend to be formed where other links are already formed, giving rise to clusters. A nice biological example of hubs can be seen in the patterns of roost use by urban bats in the city of Brisbane, Australia (Rhodes et al., 2006). Although most trees with available holes are used for roosting in the city parks, tagging studies revealed that roost use was concentrated in one key tree (the hub). Loss of the hub would have implied in disaggregation of the bat population, showing the importance of hubs for network organization.

Three useful parameters can be used to describe network structure. The clustering coefficient (C) is used to quantify the overall degree to which components of the network organize themselves into subsets within the overall network structure. Higher values of C indicate the presence of subsets (also known as cliques). Not surprisingly, this variable is positively correlated with the complexity of the network. While the distance between two elements in a network is the shortest path between them, the diameter of a network (diam) measures the maximum distance of any two elements in it. Diameter scales up with network size and complexity. Finally, the number of components (nC) describes the number of disconnected sets of nodes in a network. The higher the values of nC, the more discrete units are formed and the less structured a network will be. Many other network parameters are available and biologically friendly reviews on the use of networks and network metrics can be found in Proulx et al. (2005) and May (2006).

There is very large body of literature of network in ecology dealing with community structure (Jordano et al., 2003; Bascompte and Jordano, 2007; Zhang, 2011; Zhang et al., 2014), but relatively fewer studies on the use of network theory on population biology. Recent studies using molecular data (Fortuna et al., 2009a) or tagging studies (Rhodes et al., 2006; Jacoby et al., 2012) show that the use of the network approach can be very useful to reveal population processes (Jacoby et al., 2012).

By marking, releasing, and subsequently recapturing individuals, tagging studies generate a set of data that can be used to assess movement, survival and useful parameters in population biology (White and Burnham, 1999). Because mark recapture databases are composed of discrete events (captures and recaptures) on discrete entities (individuals), treating these data with network analysis is intuitive. When an individual is tagged (or recaptured) the spatial location of the capture can be recorded on a map. It is customary to have research trails
(or similar sampling units) divided into $n_i$ stations, which are referred to when an individual is captured. These stations can be transformed into a set of $N$ units. Whenever an individual is captured at $n_i$ and subsequently recaptured at an $n_j$ site, a link has been created in the network. Places where several individuals are recaptured can be the source of hubs in the network and suggest important features of the population organization in space. Thus individual movement provides the links to establish the connectivity in a mark-recapture network.

The distribution of individuals and populations in space is a result of the interaction between habitat distribution, demography, dispersal ability and stochasticity (Turchin, 1998; Thomas and Kunin, 1999). Mark recapture studies have been instrumental in providing parameters for functional models in attempts to model population dynamics with regards to patch size and distance (Hanski, 1997; Fagan and Lutscher, 2006). Because of their popularity, ease of manipulation and relatively well-known biology, butterflies have become models in studies of spatial structure (Haddad, 1999; Van Dyck and Baguette, 2005).

The two butterfly species used in the study, *Heliconius melpomene* (A) and *Heliconius erato* (B). The two species bear high resemblance due to Müllerian mimicry, but are easy to distinguish upon close inspection and field experience. Photos by Luciana Lima.

The use of space by mobile species such as butterflies is influenced by resource distribution and habitat characteristics such as microclimate, flowering plants (for species that forage for nectar), sites for mate searching and plants where females lay eggs (Weiss et al., 1988; Dennis et al., 2006; Dennis and Sparks, 2006). In the tropical butterfly *Heliconius*, distribution of nectar and pollen plants plays an important role in population structuring (Ehrlich and Gilbert, 1973) and mark-recapture studies show that long-lived adults establish home ranges that encompass plants where adults seek pollen and nectar (Turner, 1971; Ehrlich and Gilbert, 1973; Mallet 1986a; Murawski and Gilbert, 1986). In this paper, we used network theory to evaluate the spatial structure of two related but phylogenetically distinct co-mimics, *Heliconius erato* (L. 1758) and *Heliconius melpomene* (L. 1758) (Fig. 1). We asked whether this approach is useful in identifying network structure and detection of potential hubs (expected because of individual attachments to resources such as adult food and larval foods plants) and whether it provides insights into patterns of habitat use.

We found evidence for hub formation and complexity in *Heliconius* networks and subtle differences in network structure between the two species. As a proof-of-concept, the mark-recapture network approach is a useful tool for understanding population structure.
2 Material and Methods

2.1 Mark-recapture

The mark-recapture study was conducted in the Mata do Jiqui, an 80-ha Atlantic forest remnant (5°55'45"S, 35°10'58"W) located near the city of Natal, on the northeastern coast of Brazil. Plantations, open habitats, and a forested river floodplain border the site. We surveyed a 1.5-km trail that mixed border and lowland (near the floodplain) forested habitats, with stations marked at every 25 meters and an 800-m long central trail within the forest remnant, with stations located at every 20 meters. Butterflies were sought extensively along these trails between 7:30 and 12:00 and their capture location was recorded by referring to the nearest station. Upon capture, individuals were marked on the forewing with a unique number using a marking pen (Sanford Sharpie Ultrafine) and immediately released. Mark recapture was conducted 2-3 times a week in January/February and July/August in 2007 and 2008, during the dry and rainy season, respectively.

2.2 Network analysis

In order to describe and analyze the network structure of the two butterfly species we first constructed a square matrix whose entries corresponded to the number of stations of the mark-recapture study. Each time an individual captured at station \( i \) was recaptured at station \( j \) we assigned a number 1 at the \( a_{ij} \) matrix element. In the case of no recapture between station \( k \) and \( l \) we assign \( a_{kl} = 0 \). We interpreted this matrix as an adjacency matrix of a network.

To describe the network structure we calculated the following metrics: average number of links \( [k] \), the distribution of connectivity \( [P(k)] \), the clustering coefficient \( [C] \), and network diameter \( [diam] \) (Zhang, 2012a, 2012c). We explain these metrics below.

Contrary to complex networks, there is no preferential attachment in a random network and the probability of any two vertices to be linked is the same. In any network, the average number of links per vertex is given by \( <k> = 2L/N \), where \( L \) is the number of links and \( N \) the total number of vertices in the network. In a random network, the frequency or distribution of connectivity \( P(k) \) of the network can be described by the binomial:

\[
P(k) = C^k_N \cdot p^k (1-p)^{N-k} = \lambda^k e^{-\lambda} / k!
\]

where \( p \) is the probability of any two vertices being connected and \( \lambda \) is the average number of connections. With large \( N \), this formula can be approximated by the Poisson distribution on the right side of the equation, allowing for the generation of an expected distribution under a random model. A complex network, on the other hand, has a connectivity distribution that follows the power-law equation

\[
P(k) = Ak^{-\gamma}
\]

where \( A \) is a normalization constant and \( \gamma \) is the exponent parameter of the distribution (Newman et al., 2001).

The clustering coefficient \( C \), normally used as a measure of network complexity, will have a null value in a tree-like network such as a phylogenetic tree, and high values in complex networks, with rich clique patterns (Watts and Strogatz, 1998). For the random Poisson network the value of \( C \) is typically small, in which case it has a well defined analytical value of \( C_{null} = <k>/N \) (Newman et al., 2001).

The shortest path between vertices \( i \) and \( j \) in the network is \( d_{ij} \), and the diameter of the network is given by \( diam \), which corresponds to the maximal \( d_{ij} \). The number of components \( n_c \) expresses the number of disconnected subsets of the network. In the case where \( n_c = 1 \), the network is fully connected, i.e., there is a link connecting any two vertices in the network. A weight \( w_l \) can be added to a vertex \( l \) to express the number of times a butterfly was recaptured at a certain point in space.
In order to test whether the network was random or followed a different distribution we fitted our data to the expected curves and used the error term to evaluate the goodness of fit. Visualization of network topology was done using the software Pajek (http://vlado.fmf.uni-lj.si/pub/networks/pajek).

3 Results
During our two-year survey we captured 745 individual butterflies, with a preponderance of H. melpomene (N= 410 individuals) over H. erato (N= 335 individuals). Recapture rates were similar for both species, with the overall recapture rate near 30% and more individuals were caught in the rainy season.

3.1 Mark recapture network
The overall network structure for the mark-recapture data was similar for the two species (Fig. 2). This is because the two networks shared similar number of nodes (N) and links (L) (Table 1). Nodes refer to the places where an individual butterfly was captured (i.e., the stations) and links refer to a recapture event (or several recaptures). The average connectivity <k> of both networks was similar, with 2.36 and 2.12 links between nodes for H. erato and H. melpomene, respectively. However, because of the higher recapture frequency for H. melpomene its weighed network (<k>_W = 9.11) had 50% more connections than H. erato (<k>_W = 7.35) which can be seen in the thicker links in the H. melpomene network (Fig. 3B). The formation of hubs was easily identified in both networks, with some vertices showing as many as 5 or 6 links, well above the average of 2 connections (Fig. 2).

The network metrics suggest that H. erato has more compartmentalized, restricted movement than H. melpomene. This can be visualized in the many disconnected clusters vertices in the network (Fig. 2A). This led to a large number of disconnected components [nC] and a smaller diameter, suggesting that H. erato has shorter traveling paths than H. melpomene (Table 1 and Fig. 2A). Additionally, the short paths of H. erato create a high probability of recapture around the same group of points, leading to the formation of cliques in the network. The higher value of the clustering coefficient for the H. erato network suggests a complex network structure. In fact, the C value for Herato is five times larger than expected under a null model for a randomly assorted network with the same number of elements (Table 1).

In order to appreciate the spatial spread of the network we overlaid the vertices of the network with the approximate location of the collections points in the Mata do Jiqui (Fig. 3). The more localized and disconnected structure of the H. erato network is clearly visible by the lack of lines connecting clusters of stations (Fig. 4A).
Fig. 2 Network structure for (A) *Heliconius erato* and (B) *Heliconius melpomene* based on mark-recapture data. The dots represent places where butterflies were collected. Links between any two points represent the recapture of at least one individual. Thickness of lines represents the relative number of recaptures between any two points. A well-defined hub (a point with several other points linking to it) is seen in the center left for the network of *Heliconius melpomene*. The network for *Heliconius erato* network has a higher number of components and is less well connected than *H. melpomene*. 
Fig. 3 Spatial location of the two *Heliconius* networks. These figures were drawn using the same data presented in Figure 3 but redrawn so as to allow the vertices to be located in the corresponding spatial position as in the study site. Almost all the vertices are located in the perimeter trail and only a few are located in the central trail, although both trails systems were inspected for butterflies. Hubs are identified by the thicker linkage lines. A. *Heliconius erato*, B. *Heliconius melpomene*.

The presence of nodes with higher linkage and the high values of $C$ reject the hypothesis of either a random or a homogeneous network system. This is supported by the fact that the power law had a better fit than the random Poisson distribution adjusted to frequency distribution of network connectivity for both species (Fig. 4). Formal curve fitting (Table 2) confirms that power-law provides a better fit than Poisson for both species,
especially for *H. erato*, where the fit was far superior than Poisson. This result results from the smaller movement paths that *H. erato* makes in the network. As *H. erato* moves less than *H. melpomene* it stays around the same places, creating well connected hubs and several vertices with only one connection. These two ingredients generate a good power law.

![Figure 4](image-url)  
*Fig. 4* Frequency distribution of network connectivity (squares) for *Heliconius erato* (A) and *Heliconius melpomene* (B). The fit for the power law (continuous line) and the Poisson (dashed line) are shown.
Table 1 Network parameters for the mark-recapture data for *Heliconius erato* and *H. melpomene* in Mata do Jiqui, Brazil. \( N \) = number of vertices; \( L \) = number of links, \( L_W \) = weighed links, \( \text{diam} \) = network diameter (dimensionless), \( n_C \) = number of components of the network, \( \langle k \rangle \) = average number of links in the network, \( \langle k \rangle_W \) = average number of links in the weighed network, \( C \) = clustering coefficient, \( C/C_{\text{null}} \) = normalized clustering coefficient.

<table>
<thead>
<tr>
<th>Species</th>
<th>N</th>
<th>L</th>
<th>( L_W )</th>
<th>( \text{diam} )</th>
<th>( n_C )</th>
<th>( \langle k \rangle )</th>
<th>( \langle k \rangle_W )</th>
<th>C</th>
<th>( C/C_{\text{null}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>H. erato</em></td>
<td>33</td>
<td>28</td>
<td>103</td>
<td>5</td>
<td>8</td>
<td>2.36</td>
<td>7.35</td>
<td>0.25</td>
<td>4.9</td>
</tr>
<tr>
<td><em>H. melpomene</em></td>
<td>34</td>
<td>32</td>
<td>155</td>
<td>9</td>
<td>4</td>
<td>2.12</td>
<td>9.11</td>
<td>0.036</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Table 2 Power-law (\( \gamma \)) and Poisson (\( \lambda \)) fitting parameters for the distribution of connectivity \( P(k) \) in the mark-recapture networks for *Heliconius erato* and *H. melpomene*. Connectivity distribution was calculated using the weighed network parameters.

<table>
<thead>
<tr>
<th>Curve fitting parameters</th>
<th>Power-law</th>
<th>Poisson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species</td>
<td>( \gamma )</td>
<td>( \text{error} )</td>
</tr>
<tr>
<td><em>H. erato</em></td>
<td>1.85</td>
<td>0.458</td>
</tr>
<tr>
<td><em>H. melpomene</em></td>
<td>1.76</td>
<td>6.51</td>
</tr>
</tbody>
</table>

4 Discussion

The application of network methodology to our mark recapture data allowed us to successfully characterize the population structure of two *Heliconius* species. By doing so, we were able to effectively detect population traits that biological intuition and descriptive studies have suggested in the past. This ability to numerically describe network attributes suggests not only a heuristic but also a predictive value in this approach.

For instance, newborn *Heliconius erato* individuals disperse and eventually settle in home ranges (Turner, 1971; Mallet, 1986a), leading to highly localized individual movements determined by the location and exploitation of adult feeding sites, mate location and oviposition sites (Ehrlich and Gilbert, 1973). Our network data has identified these localized movements. Further, we have been able to detect subtle differences in network configuration between *H. erato* and *H. melpomene*. These are probably due to a higher movement rate of *H. melpomene* within the fragment and suggests a more open structure in adult dispersion, at least at the more local scale of our study. It would be instructive to learn how much these results are specific to our study site or whether they are more universal.

Both species showed the formation of hubs in their networks. Hubs are nodes with high linkage density formation and are characteristic of scale free networks, classically illustrated by airport networks, where a few main airports receive most of the flights, leaving peripheral airports with less links (Albert and Barabási, 2002). Our field observations suggest that the hubs are probably correlated with the location of adult feeding plants, *Lantana* shrubs in most cases. Thus, as expected, adult resources play an important structuring force in individual spatial location. The location of hubs around nectar/pollen plants and the high capture densities near these spots are useful in indicating the location of functional butterfly habitat (Ehrlich and Gilbert, 1973; Dennis et al., 2003). In our study this corresponds to those areas in the perimeter trail where most of the *Lantana* bushes were located and the notable absence of captures in the central trail. This is the essential habitat types for...
Heliconius erato and H. melpomene, which are typical of forest margin and second growth habitats, and absent from forest canopy and and shady forest interior (Benson, 1978; Estrada and Jiggins, 2002).

In addition to adult resources, host plant location is also essential for any butterfly, Heliconius being no exception. We had less frequent captures at or around Passiflora hosts, probably because of the more inconspicuous nature of the vines. Nevertheless, Passiflora hosts were also present in these edge habitats albeit at lower densities that the nectar plants.

Given the patchy and fragmentary nature of tropical forest remnants, we suggest that knowledge of hub dynamics will be extremely valuable to understand and predict population dynamics and guide conservation and management efforts (Rhodes et al., 2006; Fortuna et al., 2009a, b). For example, most of the hub-forming Lantana plants and Passiflora hosts are located in border areas that are easily impacted by human activity such as vegetation cutting and fire. As yet, we do not know how a hub collapse might affect Heliconius population structure and whether there is a threshold for hub collapse before population density is impacted. In a short time scale we predict that hub collapse will divert visitors to other plants or hubs, modifying the size of the network and perhaps increasing linkage density. In the long term however, hub extinction will likely lead to reduction in population size and population collapse, as suggested in network theory (Solé and Bascompte, 2006). Evidence for shifts in spatial distribution of adult Heliconius caused by changes in resource availability have been reported by Ehrlich and Gilbert (1973), who observed that cutting of flowers a road cleaning crew modified the movement patterns of the butterflies. This confirms our expectation that changes in resource availability will modify individual movement patterns and network structure. Based on the natural history of Heliconius butterflies, we envision other likely hubs, such as those relating individual faithfulness to roosting sites (Mallet, 1986b). Because roosting sites are aggregations, sites where butterflies roost are conceptually similar in functional terms as those sites used by bats in Australia (Rhodes et al., 2006).

The use of network methods in population biology has highlighted several phenomena such as hub dynamics, network stability and percolation (May, 2006; Brooks, 2006; Rozenfeld et al., 2008). In our study system it allowed us to visualize spatial structure, quantify hub importance and predict threats to population viability using mark-recapture data. In particular, we highlight the importance of identifying hubs in population networks and on assessment of their importance to population dynamics and viability.

Acknowledgement
We would like to thank Cristiana Engelmann, Clarissa Felipe and other undergraduate assistants for their help in fieldwork. We appreciate the financial support of Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior (CAPES) for a master’s fellowship to Luciana Lima, Conselho Nacional de Pesquisas (CNPq) for undergraduate research support to C. Engelmann and C. Felipe and productivity grants to G. Corso and M. Z. Cardoso. We thank Empresa de Agropecuária do Estado do Rio Grande do Norte (Emparn) management for the research permit.

References
Dennis RLH, Sparks TH. 2006. When is a habitat not a habitat? Dramatic resource use changes under differing weather conditions for the butterfly Plebejus argus. Biological Conservation, 129: 291-301
Fagan WF, Lutscher F. 2006. Average dispersal success: linking home range, dispersal, and metapopulation dynamics to reserve design. Ecological Applications, 16: 820-828


Rhodes M, Wardell-Johnson GW, Rhodes MP, Raymond B. 2006. Applying network analysis to the conservation of habitat trees in urban environments: a case study from Brisbane, Australia. Conservation Biology, 20: 861-870


