
**PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA E
BIODIVERSIDADE**

**ASSOCIAÇÕES ENTRE BORBOLETAS FRUGÍVORAS EM ÁREAS DE
FLORESTA COM DIFERENTES HISTÓRICOS DE PERTURBAÇÃO ANTRÓPICA**

Rodrigo Vieira Guidelli

Dissertação apresentada ao Instituto de Biociências da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como requisito para obtenção do título de Mestre em Ecologia e Biodiversidade.

Rodrigo Vieira Guidelli

**ASSOCIAÇÕES ENTRE BORBOLETAS FRUGÍVORAS EM
ÁREAS DE FLORESTA COM DIFERENTES HISTÓRICOS DE
PERTURBAÇÃO ANTRÓPICA**

Dissertação apresentada ao Instituto de
Biociências da Universidade Estadual
Paulista “Júlio de Mesquita Filho”, como
requisito para obtenção do título de
Mestre em Ecologia e Biodiversidade.

ORIENTADOR: Prof. Dr. Claudio José Von Zuben
COORIENTADOR: Prof. Dr. Fernando José Von Zuben

Rio Claro - SP
2016

591.5 Guidelli, Rodrigo Vieira
G946a Associações entre borboletas frugívoras em áreas de
floresta com diferentes históricos de perturbação antrópica /
Rodrigo Vieira Guidelli. - Rio Claro, 2016
125 f. : il., figs., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista,
Instituto de Biociências de Rio Claro
Orientador: Claudio José Von Zuben
Coorientador: Fernando José Von Zuben

1. Ecologia animal. 2. Dados biológicos. 3. Nymphalidae.
4. Biclusterização. 5. Árvores de decisão. 6. Redes
bayesianas. I. Título.

*Dedico este trabalho aos meus pais,
Antonio e Roseli, e a minha irmã Giovanna,
pelo apoio incondicional em todos os momentos.*

AGRADECIMENTOS

Agradeço a todos os pesquisadores envolvidos neste projeto e à minha família e amigos por sempre estarem comigo, nos bons e nos maus momentos.

RESUMO

Em 2009 Uehara Prado *et al.*, coletaram uma grande quantidade de dados para avaliar o papel das borboletas da família Nymphalidae como bioindicadoras, porém esses dados não foram utilizados em sua totalidade. O presente estudo está direcionado à experimentação e modelagem de interações ecológicas, a partir dos dados obtidos por Uehara-Prado *et al.* (2009), juntamente com aqueles não previamente utilizados que, no intuito de extrair o máximo de informação de relevância biológica e ecológica. Para tanto, foram utilizados três diferentes tipos de abordagens: (1) Biclusterização (Cheng & Church, 2000; Madeira & Oliveira, 2004); (2) Árvores de decisão (Quinlan, 1986; Bell, 1999; De'ath & Fabricius, 2000; Olden *et al.*, 2008) e (3) Redes Bayesianas (Korb & Nicholson, 2003; McCann *et al.*, 2006; Chen & Pollino, 2012; Pearl, 2014). Os resultados se mostraram bastante promissores, e as três ferramentas atingiram as expectativas; em biclusterização, conseguimos identificar todos os padrões de correlação dentro dos cenários apresentados, árvores de decisão se mostraram extremamente eficazes na classificação das variáveis apresentadas e as Redes Bayesianas conseguiram identificar quais variáveis influenciavam ou eram influenciadas pelas outras. Com este trabalho esperamos incentivar outros pesquisadores à revisitarem antigas bases de dados com ferramentas computacionais mais modernas, pois seu potencial é extraordinário.

Palavras-chave: Borboletas frugívoras, Biclusterização, Árvores de Decisão, Redes Bayesianas.

ABSTRACT

Elucidating the complex interactions networks in ecological systems is not an easy task (Proulx *et al.*, 2005) and, in order to extract information in an efficient way, powerful computational tools and the right approach, to the types of scenario to be studied, are required. In 2009 Uehara-Prado *et al.*, collected a great amount of data to assess the role of the Nymphalidae family of butterflies as bio-indicators, but these data were not used in its entirety. This study is aimed at experimentation and modeling of the ecological interactions from the data obtained by Uehara-Prado *et al.* (2009), along with those not previously used, in order to extract the maximum information of biological and ecological significance. Therefore, three different approaches were used: (1) BiclustORIZATION (Cheng & Church, 2000; Wood & Olive, 2004); (2) Decision Trees (Quinlan, 1986; Bell, 1999; De'ath & Fabricius., 2000; Olden *et al.*, 2008) and (3) Bayesian Networks (Korb & Nicholson, 2003; McCann *et al.*, 2006; Chen & Pollino, 2012; Pearl, 2014). The results were very promising, and the three tools reached our expectations; with BiclustORIZATION we managed to identify all the correlation patterns inside the scenarios presented, Decision Trees proved to be extremely effective in the classification of the variables and the Bayesian Networks were able to identify what variables influenced or were influenced by the others. With this work, we hope to encourage other researchers to revisit old databases with more modern computational tools, because its potential is extraordinary.

Keywords: Fruit-feeding butterflies, Biclustering, Decision Trees, Bayesian Networks.

SUMÁRIO

1 INTRODUÇÃO.....	8
2 MATERIAL E MÉTODOS.....	10
2.1 ÁREA DE ESTUDO	10
2.2 GRUPO DE ESTUDO	11
2.3 PROCEDIMENTOS DE AMOSTRAGEM.....	11
3 FERRAMENTAS MATEMÁTICAS E COMPUTACIONAIS	11
3.1 GRAFOS	12
3.2 CLUSTERIZAÇÃO	13
3.3 BICLUSTERIZAÇÃO	13
3.4 O ALGORITMO INCLOSE	16
3.5 EXEMPLO DE ESTRATÉGIA DE ANÁLISE DE DADOS EMPREGANDO GRAFOS E BICLUSTERIZAÇÃO.....	17
4 APRENDIZADO DE MÁQUINA	20
4.1 ÁRVORES DE DECISÃO	23
4.2 REDES BAYESIANAS.....	31
4.3 INFERÊNCIA EM REDES BAYESIANAS	36
4.4 APRENDIZADO EM REDES BAYESIANAS.....	37
5 RESULTADOS E DISCUSSÃO	38
5.1 BICLUSTERIZAÇÃO.....	38
5.2 ÁRVORES DE DECISÃO.....	41
5.3 REDES BAYESIANAS.....	46
6 CONCLUSÃO.....	53
7 PERSPECTIVAS FUTURAS.....	54
8 REFERÊNCIAS BIBLIOGRÁFICAS.....	55

ASSOCIAÇÕES ENTRE BORBOLETAS FRUGÍVORAS EM ÁREAS DE FLORESTA COM DIFERENTES HISTÓRICOS DE PERTURBAÇÃO ANTRÓPICA

1 INTRODUÇÃO

Apresentar modelos que elucidem as intrincadas redes de interações dentro dos diferentes habitats não é uma tarefa fácil (Proulx *et al.*, 2005). Enquanto redes de interações em nível molecular têm estado na vanguarda da biologia moderna, devido à quantidade cada vez maior de dados pós genômicos disponíveis, redes de interações em outras escalas só agora vêm recebendo a devida atenção da comunidade científica, em particular, redes ecológicas (Aderhold *et al.*, 2013). Ecossistemas são sistemas dinâmicos complexos, com intrincadas redes de interações entre as espécies e os elementos abióticos que os compõem, apresentando novos desafios e oportunidades à modelagem e estatística computacional (Aderhold *et al.*, 2013).

Entre estes desafios, está o fato de redes de interações em sistemas ecológicos se apresentarem em um ambiente espacialmente distribuído e poderem variar dentro deste ambiente, dependendo da composição dos elementos (fatores bióticos e abióticos) presentes, o que pode levar a determinados comportamentos que, à primeira vista, parecem imprevisíveis e difíceis de serem analisados, como a mudança no habitat influenciando o número de indivíduos de uma espécie que, por sua vez, influenciam outras mudanças (Henneman & Memmott, 2001), ou mesmo todo o sistema, alternando entre diferentes estados dinâmicos (Beisner *et al.*, 2003).

Para extrair informações de maneira eficiente a partir deste tipo de sistema complexo, faz-se necessária uma abordagem adequada ao tipo de cenário em estudo e a utilização de ferramentas que possam inferir sobre a estrutura das redes de interações formadas, com base nos dados coletados (Milns *et al.*, 2010).

Em 2009, foi realizada uma avaliação da resposta de grupos de borboletas frugívoras a três tipos de perturbações antrópicas em reservas da Floresta Atlântica (com diferentes cenários, envolvendo extração de madeira, efeitos do fogo, manejo de trilhas e efeito de borda) utilizando o mesmo design amostral, com o objetivo de verificar o papel desses insetos como bioindicadores e a existência de parâmetros úteis e consistentes que possam melhorar a avaliação e o monitoramento nesses tipos de cenário (Uehara-prado *et al.*, 2009)

Os resultados obtidos pelos autores indicaram que, através da aplicação de técnicas de ordenação com base em dados de abundância, presença e ausência e frequência de ocorrência

(IndVal) (Dufrêne & Legendre, 1997), a composição de espécies de borboletas difere entre áreas perturbadas e não perturbadas dentro de todas as reservas analisadas. Os padrões observados nas ordenações não se alteraram quando espécies foram agrupadas em subfamílias. O estudo também revelou 12 espécies indicadoras típicas para os diferentes tipos de perturbação dentro das reservas. A riqueza de espécies observada foi maior nas áreas perturbadas em todas as três reservas (Uehara-Prado *et al.*, 2009).

Além dos dados referentes ao papel das espécies como bioindicadoras, esses autores também coletaram uma grande quantidade de dados adicionais sobre as espécies (Uehara Prado, comunicação pessoal): sexagem, tamanho e idade, que, devido à natureza do estudo desenvolvido, ainda não haviam sido utilizados.

O presente estudo tem por objetivo a experimentação e modelagem de interações ecológicas, a partir dos dados obtidos por Uehara-Prado *et al.* (2009), juntamente com aqueles não previamente utilizados que, justamente por envolverem uma complexidade grande de fatores bióticos e abióticos relacionados a cada tipo de cenário de perturbação antrópica, requerem abordagens computacionais que permitam uma melhor compreensão da associação entre as espécies, níveis de perturbação antrópica e fatores bióticos e abióticos relacionados.

Para tanto, foram utilizados no presente trabalho três diferentes tipos de abordagens: (1) Biclusterização, com o agrupamento simultâneo de linhas e colunas de uma matriz, as quais apresentam algum tipo de coerência interna (Cheng & Church, 2000; Madeira & Oliveira, 2004); (2) Árvores de decisão, para identificação de associações não-lineares dependentes de contexto entre múltiplas variáveis preditoras correlacionadas (Quinlan, 1986; Bell, 1999; De'ath & Fabricius, 2000; Olden *et al.*, 2008) e (3) Redes Bayesianas, modelos gráficos para análises sob condições de incerteza, que representam de forma simples as relações de causalidade das variáveis de um sistema, aptas a lidar com dados faltantes e informações provenientes de diferentes fontes (Korb & Nicholson, 2003; McCann *et al.*, 2006; Chen & Pollino, 2012; Pearl, 2014).

Sendo assim, os resultados de interesse não estarão vinculados às consequências isoladas da atuação de cada fator individualmente, mas sim ao comportamento emergente resultante de todas as possíveis interações, levando em conta todos os fatores envolvidos e disponíveis junto aos dados coletados, com o intuito de identificar e analisar quaisquer possíveis associações de relevância biológica que venham a surgir.

2 MATERIAL E MÉTODOS

2.1 ÁREA DE ESTUDO

Os dados utilizados no presente projeto foram coletados por Uehara-Prado em três diferentes reservas, situadas em uma região de relevo montanhoso dentro da Serra do Mar, na Floresta Atlântica Montana, sudeste do Brasil. Esta área possui um clima úmido, sem estação seca, com uma pluviosidade anual de 1.860 a 4.400 mm dentro das reservas (Sigrh, 2008).

As três reservas estudadas estão alocadas dentro de uma área de Mata Atlântica de 300.000 há de comprimento possuindo diferentes históricos de perturbação antrópica e dentro de cada uma foram selecionadas duas áreas, com perturbações distintas, para coleta de dados (Uehara-Prado *et al.*, 2009). As unidades amostrais foram alocadas de modo a cada uma abranger um dos três tipos diferentes de condição: floresta secundária juntamente com vegetação secundária em estágios iniciais (Núcleo Santa Virgínia - NSV), somente floresta secundária (Estação Biológica de Boracéia - EBB) e apenas vegetação secundária em estágios iniciais (Reserva Biológica de Paranapiacaba - RBP).

Núcleo Santa Virgínia – NSV – A primeira área sofreu manejo através do corte da vegetação e fogo, com posterior estabelecimento de pastagens. Atualmente, esta área se encontra ocupada por vegetação secundária em estágios iniciais (Tabarelli & Mantovani, 1999). A segunda área desta reserva foi seletivamente explorada para extração de madeira e, atualmente, é uma área de floresta secundária razoavelmente bem preservada, com alguns remanescentes próximos de floresta primária (Uehara-Prado *et al.*, 2009).

Estação Biológica de Boracéia – EBB – área de floresta secundária com pequenas estradas e trilhas. Há um manejo periódico da vegetação através de poda. Porém, algumas trilhas abandonadas a partir de 1970 permaneceram sem manejo algum (Uehara-Prado *et al.*, 2009).

Reserva Biológica de Paranapiacaba – RBP – a remoção de vegetação nesta área para a construção de estradas vem desde o início da colonização do Brasil e se agravou a partir dos anos 1950, com a construção de complexos industriais e posterior abertura de rodovias no local durante os anos de 1970 (SMA, 1990). Atualmente, a reserva se encontra, em sua maior parte, ocupada por vegetação secundária em estágios iniciais (Kirisawa *et al.*, 2003).

2.2 GRUPO DE ESTUDO

No Brasil, existem cerca de 3.200 espécies conhecidas de borboletas, que são divididas em dois grupos, segundo os hábitos alimentares dos adultos: as nectarívoras, que se alimentam de néctar durante a vida adulta; e as frugívoras (família Nymphalidae), que obtêm a maior parte de seus nutrientes de frutas fermentadas e seiva de plantas, além de animais em decomposição (DeVries, 1987; Lamas, 2004). Estes grupos de organismos, juntamente com as mariposas, são essenciais para a manutenção de importantes processos ecossistêmicos, como polinização, herbivoria e decomposição (Lomov *et al.*, 2006, Furlanetti *et al.*, 2010).

Como o objetivo de Uehara-Prado *et al.* (2009) foi a avaliação desse grupo no papel de bioindicadores, foi necessário que se levassem em conta alguns critérios para a alocação de insetos neste grupo. Algumas das principais características de um bom bioindicador são: taxonomia relativamente bem resolvida; conhecimento biológico; diversidade conveniente para a pesquisa que se pretende desenvolver; ciclo de vida curto; diversidade ecológica; fidelidade de habitat; associação estreita com recursos ou outras espécies; “sedentarismo” relativo; fidelidade na amostragem, triagem e identificação e pouco uso humano (Freitas *et al.*, 2006).

Além de cumprir a maior parte destes critérios, as Nymphalidae são facilmente amostráveis através do método de armadilhas atrativas com iscas (DeVries, 1988; Pinheiro & Ortiz, 1992), possibilitando aos pesquisadores o mínimo de manuseio durante a identificação e soltura de seus indivíduos, evitando possíveis danos aos mesmos. Através das armadilhas, também se torna possível, nas amostragens simultâneas, padronizar o esforço em diferentes áreas (Freitas *et al.*, 2004; Uehara-Prado *et al.*, 2004).

2.3 PROCEDIMENTOS DE AMOSTRAGEM (UEHARA-PRADO ET AL., 2009)

Em cada reserva, duas áreas com perturbações contrastantes foram escolhidas para comparação, sendo uma mais perturbada que a outra: NSV – corte e fogo × extração seletiva de madeira. EBB – trilhas manejadas × trilhas não manejadas. RBP – região de borda × região de interior. Daqui em diante, estas áreas serão chamadas apenas de "perturbadas" e "não perturbadas", respectivamente, para cada reserva.

Doze unidades amostrais foram instaladas em cada reserva, seis em áreas não perturbadas e seis em áreas perturbadas. A unidade amostral é definida por cinco armadilhas combinadas. As armadilhas usadas são cilindros feitos de rede com um funil interno, dentro do qual foram colocados como isca, banana amassada juntamente com caldo de cana, fermentado por pelo menos 48 h. As armadilhas foram posicionadas de maneira enfileirada no sub-bosque de cada área, juntamente com as trilhas pré-existent, a aproximadamente 1,5 – 2 m de altura do solo, com uma distância de pelo menos 23 m entre armadilhas adjacentes.

O comprimento total dos transectos de amostragem variou de 115 a 160 m, e não diferiu entre as áreas perturbadas e não perturbadas dentro das reservas (*t-test*, $P > 0.05$ em todos os casos). As armadilhas eram checadas a cada 48 horas, e as iscas substituídas a cada visita. Cada borboleta encontrada e que pudesse ser identificada em campo era liberada após receber uma marca individual. Recapturas não foram reincluídas na análise.

As armadilhas foram mantidas em campo aberto durante oito dias por mês. A amostragem foi feita mensalmente de Novembro a Maio, período mais favorável à captura de borboletas na região sudeste do Brasil (Brown, 1972). A amostragem em cada reserva foi feita em diferentes anos: NSV, 2004 - 2005; EBB, 2005 - 2006 e RBP, 2006 - 2007. Sabe-se que variações anuais podem representar um empecilho às comparações entre as reservas. Porém, em sua maior parte, este estudo é focado em comparações entre áreas dentro das reservas. Soma-se a isso o fato de que, por restrições logísticas, não foi possível coletar amostras em todas as reservas simultaneamente.

Uehara-Prado *et al.* (2009) descreveram 1.886 indivíduos de 72 espécies de borboletas coletadas. No presente estudo, além da identificação até espécie e quantificação de todos os exemplares coletados, também foram considerados dados (ainda não publicados e que não haviam sido alvo de estudos de qualquer natureza) referentes à sexagem, idade estimada e tamanho das borboletas.

3 FERRAMENTAS MATEMÁTICAS E COMPUTACIONAIS

3.1 GRAFOS

A abordagem mais usada na modelagem de associações e na descrição de relações é o emprego de grafos (Bondy & Murty, 2008). Segundo Zhang (2012), um grafo é uma entidade matemática cujos elementos são vértices e arestas, as quais conectam os vértices. No contexto

deste trabalho, os vértices podem corresponder às espécies e aos fatores bióticos e abióticos. Por outro lado, as arestas que ligam pares de vértices são utilizadas para descrever as relações e associações entre os vértices.

Representar as mais relevantes relações e associações entre entidades de um processo ecológico, na forma de um grafo, permite o emprego de técnicas computacionais para análise do grafo resultante (Kuznetsov, 1999).

3.2 CLUSTERIZAÇÃO

A detecção de padrões, agrupamentos e tendências presentes no grafo que descreve um processo ecológico geralmente requer o emprego de ferramentas de aprendizado de máquina. Terão destaque neste projeto técnicas de organização e agrupamento de dados (Everitt, 1993; Xu & Wunsch, 2010), para reconhecimento de padrões e grupos, técnicas essas que serão descritas mais em detalhe a seguir.

Segundo DiMaggio *et al.* (2008), problemas envolvendo organização e agrupamento de dados são predominantes em uma série de disciplinas e incluem, entre outras aplicações, o reconhecimento de padrões e a extração de conhecimento, seja diretamente a partir de dados coletados do processo em estudo, seja a partir de modelos matemáticos na forma de grafos, como será o caso do presente projeto. Segundo esses mesmos autores, o objetivo do agrupamento de dados, independente da aplicação, é o de organizar os dados de tal maneira que objetos que exibam atributos “similares” sejam agrupados juntos. O termo entre aspas indica que a definição de similaridade depende da aplicação que se pretende fazer em cada estudo e pode corresponder à comparação direta de valores ou ao grau de coerência entre tendências ou padrões de valores.

3.3 BICLUSTERIZAÇÃO

Vários métodos têm sido propostos para a análise de grafos, dentre eles destaca-se a biclusterização ou, de forma mais genérica, a co-clusterização (Govaert & Nadif, 2014). O termo biclusterização foi introduzido pela primeira vez por Cheng & Church (2000) na análise de dados de expressão gênica, e representa o agrupamento simultâneo de linhas e colunas de uma matriz, as quais apresentam algum tipo de coerência interna. A ideia é realizar a

biclusterização da matriz de conectividade do grafo extraído do processo em estudo, visando detectar padrões de ocorrência. Exemplo: se determinada espécie está sempre presente em certo cenário ecológico e está sempre ausente em outro certo cenário ecológico, então a conectividade do grafo que associa tais espécies aos cenários ecológicos vai indicar isso e a biclusterização da matriz de conectividade do grafo vai permitir identificar quaisquer eventuais associações que possam apresentar alguma relevância biológica. Esta última etapa exige do(s) pesquisador(es) um conhecimento prévio de fatores que possam interferir no sistema biológico sendo estudado (no caso, comunidade de espécies), permitindo assim uma interpretação correta dos *biclusters* encontrados, feita com bom embasamento biológico/ecológico.

Considerando o grupo das borboletas frugívoras, foram obtidos agrupamentos de espécies através do emprego de técnicas de biclusterização. Para tanto, foram aplicadas regras de agrupamento, considerando as linhas e colunas das matrizes de dados, para definir padrões e gerar *biclusters* empregando o algoritmo InClose (Andrews, 2009). Este algoritmo garante encontrar todos os *biclusters* existentes, em suas versões maximais. Um *bicluster* é maximal quando nenhuma linha ou coluna pode ser adicionada sem violar as condições de coerência interna do *bicluster*.

Os *biclusters*, e seus respectivos grafos, foram obtidos a partir da aplicação do algoritmo InClose na matriz binária de presença e ausência (Planilha 1) encontrada no Apêndice A. Esta planilha é a mesma utilizada por Uehara-Prado *et al.* (2009) para testar o grupo das Nymphalidae como bioindicadoras. Porém, ele se utilizou principalmente do *Indicator Species Analysis* (IndVal) (Dufrêne & Legendre, 1997). Já no presente estudo, a disposição dos valores na planilha precisou ser levemente modificada para a aplicação do InClose.

Finalmente, cada agrupamento obtido foi analisado em termos biológicos, buscando encontrar relação entre os aspectos da bionomia e comportamento das borboletas envolvidas, e os locais de ocorrência de cada espécie.

Tomando como exemplo uma matriz X binária arbitrária, apresentada na Figura 1 a seguir, podem ser extraídos *biclusters* de 1's dessa matriz, sendo apresentadas quatro propostas na Figura 1. Para ilustrar como pode ser feita a interpretação de cada *bicluster*, se a matriz X binária exemplificada representasse as conexões de amizade num grafo, cada *bicluster* revelaria uma comunidade, ou seja, um grupo de indivíduos que compartilham amizade entre si, de modo que todos são amigos de todos dentro da comunidade.

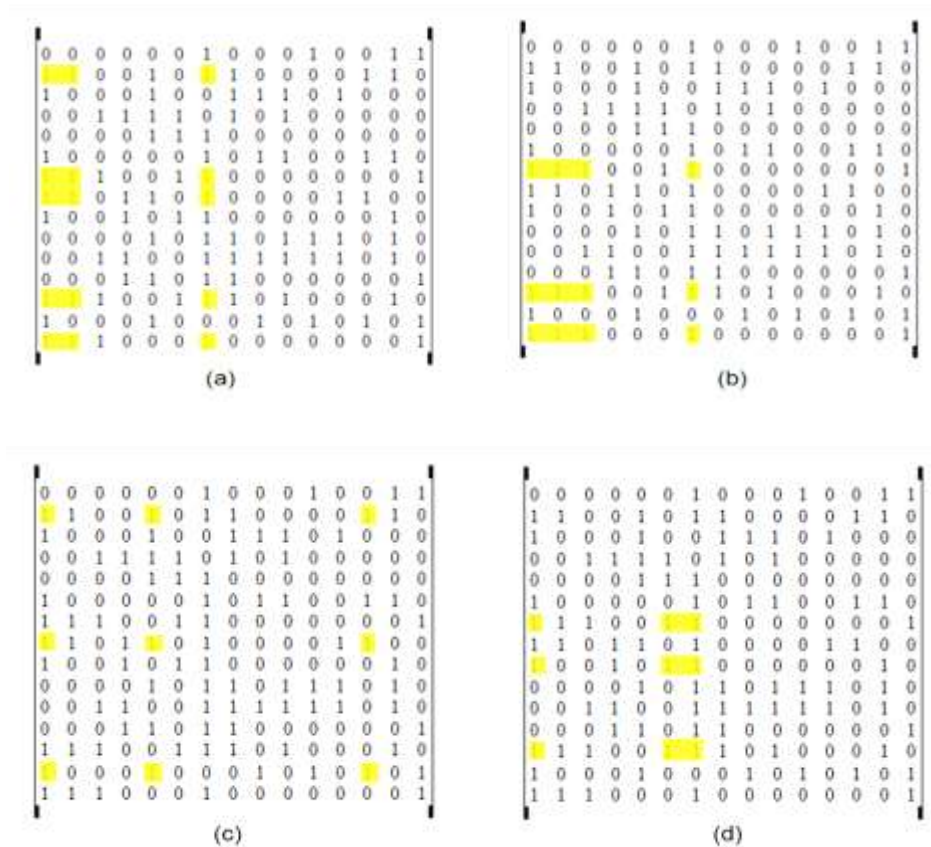


Figura 1 – Exemplo de quatro *biclusters* (em amarelo) formados por um subconjunto de objetos e um subconjunto de atributos não-contíguos na matriz de dados original.

No presente estudo, os dados coletados foram agrupados em uma matriz em que as colunas representam os locais de coleta (e suas respectivas perturbações) e as linhas, as características das espécies coletadas, semelhante ao exemplo abaixo.

Espécies	Reserva 1		Reserva2		Reserva 3	
	P1	NP 1	P 2	NP 2	P 3	NP 3
Esp 1						
Esp 2						
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Esp _n						

Tabela 1: Cada linha representa uma espécie, com P e NP representando as áreas Perturbadas e Não Perturbadas de cada reserva. Cada campo da tabela pode representar: presença ou ausência da espécie, número de indivíduos da espécie, razão entre sexos para a espécie, ou atributos morfológicos dos indivíduos coletados.

Quando se consideram dados binários, as matrizes de dados são compostas por 0 e 1, indicando respectivamente, ausência e presença de cada espécie de borboleta. Na sequência, foram feitos testes com as matrizes, em que constará o número real de indivíduos coletados para cada uma das espécies de borboletas. Posteriormente à análise inicial, foram acrescentados os dados referentes à sexagem, idade estimada e tamanho das borboletas. A biclusterização não é restrita a dados binários e pode ser aplicada a matrizes com elementos pertencentes ao conjunto dos inteiros e dos reais, mas neste caso deve-se recorrer a uma extensão do algoritmo InClose, denominada RInClose (Veroneze *et al.*, 2015).

3.4 O ALGORITMO InClose

Andrews (2009) propôs um algoritmo capaz de encontrar todos os *biclusters* maximais de 1's em uma matriz binária, de forma simples e rápida. Esses *biclusters* são ditos serem maximais no sentido de não ser possível acrescentar nenhuma linha ou coluna adicional que preserve a propriedade de valor constante 1 para todos os seus elementos. É evidente que, havendo interesse em *biclusters* de 0's, o algoritmo pode ser diretamente adaptado para extrair esses *biclusters*.

O algoritmo InClose foi proposto no contexto de análise de conceitos formais (Poelmans *et al.*, 2013a, b), que estão diretamente associados à informação contida nas sub-matrizes da Figura 1, por exemplo. Ao ser executado, o InClose fornece todos os *biclusters* maximais presentes na matriz de dados, sem repetição. Para tanto, o InClose aplica o conceito de ordem lexicográfica. Por exemplo, combinações matemáticas possuem uma ordem lexicográfica quando a sequência {1, 2, 3} vem antes de {1, 2, 4} e a sequência {1, 2, 6} antecede {1, 3}. Se os *biclusters* forem gerados na ordem lexicográfica, é possível dizer, ainda durante sua construção, se ele é ou não é redundante com algum outro *bicluster* já gerado.

Considere como exemplo uma base de dados denominada Mushroom: registros do *Audubon Society Field Guide to North American Mushrooms* (1981). Esta base de dados contém 8.124 cogumelos descritos por 125 atributos binários. Ao ser executado junto à base Mushroom, o algoritmo InClose levou 1,54 segundos para encontrar todos os 226.920 *biclusters*, em uma máquina Intel i7 2.90 GHz com 6 GB de memória RAM, em ambiente Windows 8. O número de *biclusters* pode ser menor caso se imponha restrições de tamanho mínimo, seja para número de linhas ou número de colunas.

O grupo de pesquisa já dispunha de um programa computacional com a implementação do algoritmo InClose para uso imediato, o qual foi fornecido a partir de uma parceria com o LBiC (Laboratório de Bioinformática e Computação Bioinspirada) da Unicamp. Também encontrava-se à disposição, para pronto uso, uma extensão do algoritmo InClose, denominada RInClose e desenvolvida por Veroneze *et al.* (2015) do LBiC - UNICAMP. O RInClose opera com matrizes cujos elementos são inteiros, binários ou reais, apresentando um enorme potencial para utilizações futuras em análises de banco de dados ecológicos.

3.5 EXEMPLO DE ESTRATÉGIA DE ANÁLISE DE DADOS EMPREGANDO GRAFOS E BICLUSTERIZAÇÃO

Para exemplificar como se dará o emprego da biclusterização na análise de grafos gerados a partir de dados experimentais, adotam-se aqui estratégias de análise de dados similares às aquelas empregadas em Pavlopoulos *et al.* (2011), embora esses autores não tenham empregado diretamente o conceito de biclusterização, como foi feito ao longo desta pesquisa.

Suponha que a interação de borboletas frugívoras junto a diversos ambientes ecológicos admita uma descrição na forma da matriz de conectividade da Figura 1. O grafo associado à matriz de conectividade da Figura 1 é apresentado na Figura 2 a seguir.

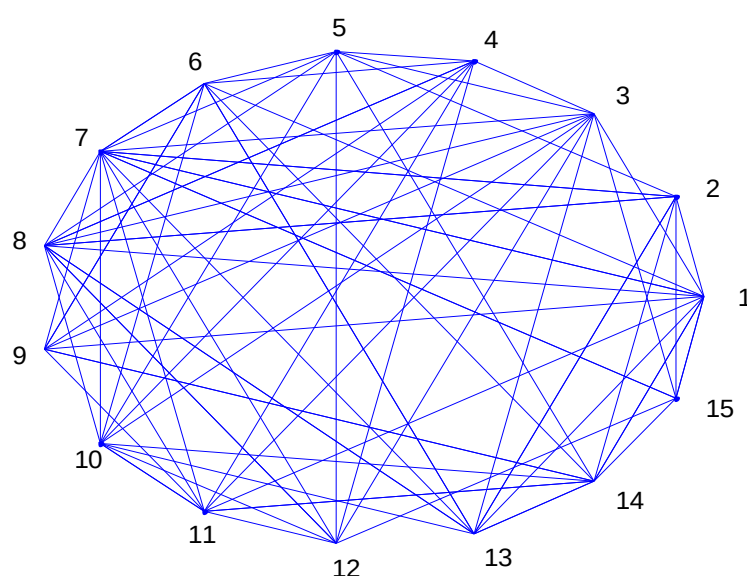


Figura 2 – Grafo associado à matriz de conectividade da Figura 1.

Como a matriz de conectividade da Figura 1 não é simétrica, é possível considerar um grafo direcionado; mas é apresentada, na Figura 2, uma versão não-direcionada do grafo, em

que também não estão sendo consideradas as auto-conexões (conexão de um vértice consigo próprio), como é o caso, por exemplo, dos vértices 2 e 4. Essas simplificações apenas viabilizam o estudo de caso com uma matriz de conectividade sintética e arbitrária já apresentada na Figura 1.

Repare que a presença de 1's em elementos (i, j) da matriz da Figura 1 indica a existência de uma aresta entre os vértices i e j no grafo da Figura 2.

Através de inspeções visuais, seja da matriz de conectividade da Figura 1, seja do grafo com vértices e arestas da Figura 2, não é possível extrair conhecimento relevante presente no modelo matemático, o qual supostamente representa interações existentes em um processo ecológico em estudo.

No entanto, com a aplicação do algoritmo InClose junto à matriz de conectividade da Figura 1, obtém-se os quatro *biclusters* destacados nas Figuras 1(a), 1(b), 1(c) e 1(d). No *bicluster* da Figura 1(c), por exemplo, fica evidente a existência de uma coerência de conexões entre os vértices 2, 8 e 14. Esses são os únicos vértices que se conectam simultaneamente aos vértices 1, 5 e 13. Isolando-se este *bicluster* “camuflado” na Figura 2, resulta o grafo da Figura 3. Trata-se de associações relevantes entre vértices da Figura 2, mas que não estavam evidentes ainda e foram identificadas pelo algoritmo de biclusterização InClose, ao ser executado. Em teoria de grafos, o grafo apresentado na Figura 3 é denominado um *biclique* do grafo apresentado na Figura 2 (Bondy & Murty, 2008), o que indica que o raciocínio simétrico, iniciando pelos vértices 1, 5 e 13, é igualmente válido. Repare que, na Figura 3, o trio de vértices 2, 8 e 14 não se conecta diretamente entre si, assim como o trio de vértices 1, 5 e 13.

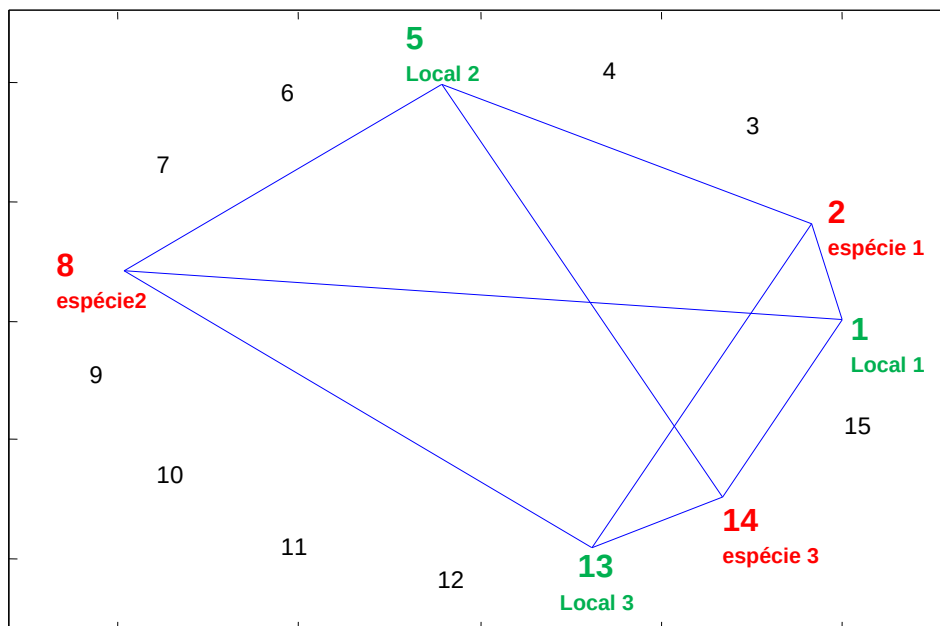


Figura 3 – Grafo associado ao *bicluster* da Figura 1(c), fornecido pelo algoritmo InClose.

Supondo que os vértices 2, 8 e 14 se referem a espécies de borboletas frugívoras e que os vértices 1, 5 e 13 se referem a todos os ambientes com acentuada atividade humana que foram monitorados, e se as arestas existentes indicam que cada uma destas espécies (2, 8, 14) de borboleta ocorre em cada um dos ambientes (1, 5, 13), logo, este conjunto de três espécies possui algum tipo de associação com aquele conjunto de ambientes.

A análise nesta seção envolve simplificações e dados fictícios, tendo apenas o propósito de ilustrar o tipo de conhecimento que pode ser extraído após monitorar a ocorrência de espécies frugívoras em múltiplos ambientes devidamente classificados.

A análise real será inteiramente baseada na matriz de dados da Planilha 2, Apêndice B, que contém os dados coletados por Uehara Prado *et al.* (2009). A disposição dos dados presentes na planilha foi adaptada para se parecer mais com a Tabela 1, tornando a aplicação do algoritmo InClose possível.

4 APRENDIZADO DE MÁQUINA

Uma vez realizadas as associações entre espécies e locais, através da ferramenta InClose aplicada sobre os dados de presença e ausência, o próximo passo é a avaliação do conjunto de dados inéditos, que contém as demais variáveis, no intuito de encontrar um padrão de associação de espécies dentro de determinado ambiente. Com base nos atributos de cada indivíduo, classificam-se as espécies de acordo com seus atributos e local de ocorrência.

O estudo desenvolvido por Uehara-Prado *et al.* (2009), teve como principal objetivo a análise do papel de espécies de borboletas frugívoras como bioindicadoras. Logo, não houve uma ênfase na descrição dos demais atributos dos indivíduos, pois trabalhou-se com um conjunto de dados ainda incompleto.

Além disso, observando o conjunto de dados utilizado na biclusterização, e o conjunto inédito dos demais atributos, são visíveis mais algumas diferenças entre ambos: os dados de presença e ausência são compostos apenas por zeros e uns, já o conjunto de dados dos atributos possui números inteiros. Há também mais atributos no conjunto inédito do que no conjunto revisitado. A soma destas três principais características do conjunto inédito torna a análise e modelagem desse conjunto de atributos ainda mais promissora que a anterior.

Atualmente, tem-se atribuído cada vez mais ênfase à predição de variáveis, e não apenas sua descrição e explicação, devido não apenas à comum falta de dados de campo em estudos ecológicos, mas à crescente destruição de habitats e problemas com mudanças climáticas em escala global. Com isso, a importância de tais predições tem sido, em muitos casos, o principal objetivo das análises de interações entre espécies e ambiente (Franklin, 1998; Iverson and Prasad, 1998; Vayssières *et al.*, 2000). Muitos métodos baseados em algoritmos de aprendizado de máquina, desenvolvidos para predição e classificação de variáveis, têm sido aplicados para essa tarefa (Franklin, 1995; Guisan & Zimmermann, 2000).

De acordo com Basgalupp (2010), algoritmos de aprendizado de máquina são ferramentas poderosas que muitas vezes se utilizam de Inteligência Artificial e Estatística para criar modelos capazes de extrair conhecimento a partir de determinado conjunto de dados.

Na utilização de algoritmos de aprendizado, considera-se que o conjunto de dados no qual o algoritmo será aplicado (no presente caso, o grupo das borboletas frugívoras) é composto por objetos descritos por um conjunto de atributos (propriedades, características), sendo que a cada objeto

deve ser associada uma classe, dentre um conjunto de classes possíveis (Fayyad & Keki 1992; Rokach & Maimon, 2005; Tsang *et al.*, 2011; Otero *et al.*, 2012).

Atributos são variáveis independentes e observáveis, e seus domínios podem ser tanto contínuos (numéricos, como tamanho e idade) como discretos (categóricos, como sexo, por exemplo) (Fayyad & Keki, 1992; Rokach & Maimon, 2005; Basgalupp, 2010; Tsang *et al.*, 2011; Otero *et al.*, 2012). Os atributos contínuos assumem valores numéricos em intervalos no eixo dos números reais; os discretos são divididos em dois tipos: os categóricos ordinais, que assumem um conjunto finito de valores, que podem ser ordenados; e os categóricos não-ordinais, que assumem um conjunto finito de valores que não podem ser ordenados (Quinlan, 1986; Basgalupp, 2010). As classes, por sua vez, têm seu valor determinado a partir dos valores das variáveis independentes (atributos). Os dados desses conjuntos são chamados de dados de treinamento, e a partir deles se faz a inferência indutiva para a construção do classificador (Monard & Baranauskas, 2003; Alpaydin, 2004)

O algoritmo, também chamado de indutor ou algoritmo de indução, pode utilizar um de dois principais métodos para o aprendizado indutivo: método supervisionado e não supervisionado.

Métodos supervisionados são métodos que buscam encontrar relações entre os atributos de entrada e os atributos alvo (classes); uma vez encontrada, tal relação é descrita na forma de uma estrutura chamada de *modelo* (Fayyad & Keki, 1992; Rokach & Maimon, 2005). Ao aplicar o algoritmo a grandes conjuntos de dados pré-classificados, espera-se que ele descubra como realizar o mapeamento da entrada para a saída (Fayyad & Keki, 1992).

Há dois principais tipos de modelos de aprendizado supervisionado: modelos de classificação (ou classificadores) e modelos de regressão (Rokach & Maimon, 2005).

Modelos de regressão são adequados para se lidar com variáveis dependentes que possuem valores contínuos ou discretos ordenados, retornando uma resposta numérica; neste tipo de modelo, o erro de predição é geralmente medido pela somatória do quadrado da diferença entre os valores preditos e os observados (Loh, 2011). Já os modelos de classificação são projetados para variáveis dependentes que tenham um número finito de valores não ordenados, retornando uma resposta nominal (classe), como "verdadeiro" ou "falso"; seu erro de predição é medido em termos de custo de classificação errônea (Loh, 2011).

No aprendizado supervisionado, os valores da classe (atributo classe) do conjunto são conhecidos; se o domínio do atributo classe é discreto, tem-se um problema de classificação;

se é numérico, tem-se um problema de regressão ou aproximação de funções (Basgalupp, 2010). No presente estudo, o modelo abordado foi o de Classificação.

A tarefa de um classificador consiste na atribuição de objetos (valores de entrada) a uma das classes de categorias pré-definidas (valores de saída) com base nos atributos dos objetos (Loh, 2011; Tsang *et al.*, 2011; Otero *et al.*, 2012). A Figura 4 mostra um exemplo fictício de como uma ferramenta de classificação funciona. Tan *et al.* (2005) definem classificação como a tarefa de aprender uma função alvo f que mapeia cada conjunto de atributos x , ou seja, cada exemplo, a uma das y classes pré-definidas.

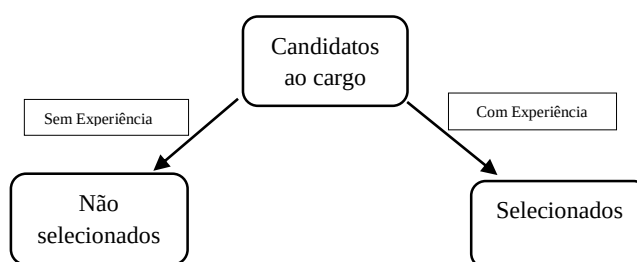


Figura 4: Exemplo de uma ferramenta de classificação analisando um grupo de pessoas interessadas em uma vaga de emprego para professor de uma faculdade. Após mapeado o conjunto de objetos, e seus respectivos atributos, aqueles sem experiência foram alocados para a classe “Não selecionados”. Já aqueles com experiência foram alocados na classe “Selecionados”.

Na construção de modelos computacionais de classificação, existem essencialmente duas abordagens principais: *Top Down* e *Bottom up*.

Top Down: também chamada de *teste de hipótese*; parte-se do princípio de que existe uma hipótese, e se deseja confirmá-la ou refutá-la, sendo o modelo construído a partir de informações concedidas por especialistas (Cortes *et al.*, 2002; Basgalupp, 2010).

Bottom up: também chamada de *busca de conhecimento*, inicia-se o processo de exploração dos dados na tentativa de descobrir algo que ainda não é de conhecimento, em que o modelo é obtido através da identificação das relações entre as variáveis da base de dados (Cortes *et al.*, 2002; Basgalupp, 2010).

4.1 ÁRVORES DE DECISÃO

Um dos modelos de classificação mais populares é o de árvores de decisão (Tsang *et al.*, 2011). Isso se deve a suas inúmeras qualidades, dentre elas: a flexibilidade necessária para lidar com uma ampla gama de tipos de resposta, incluindo numéricas e categóricas; facilidade e robustez de construção; o fato de produzir um resultado altamente interpretável; grande capacidade de predição; adaptabilidade em lidar com conjuntos de dados que contêm erros ou dados faltantes (De'ath & Fabricius, 2000).

Árvores de Decisão (Figura 5) são diagramas *bottom up* construídos de modo sequencial, com aprendizagem indutiva; representam um algoritmo que auxilia a tomada de decisão para a resolução de um determinado problema complexo (Mitchel 1999; Vayssieres *et al.*, 2000; Baisen *et al.*, 2006; Coelho e Leite, 2015), funcionando como um grafo de decisões representado na forma de ramos e nós, e oferecendo um modelo com alto nível de capacidade de predição e classificação, quando bem treinado (Baisen *et al.*, 2006).

Como o modelo de árvore utilizado no presente estudo emprega a abordagem *Bottom up*, há algumas condições a serem seguidas para sua implementação na base de dados. São elas:

- A informação sobre cada objeto a ser classificado deve poder ser expressa através de uma coleção fixa de propriedades ou atributos.
- O número de classes pode ser definido a priori (treinamento supervisionado) ou automaticamente a partir dos dados disponíveis (treinamento não-supervisionado).
- São possíveis dois tipos de classes: as discretas e as contínuas.
- Deve haver um número maior de objetos do que classes.
- A tarefa de classificação deve ser implementada usando uma base de regras de decisão, de forma lógica.

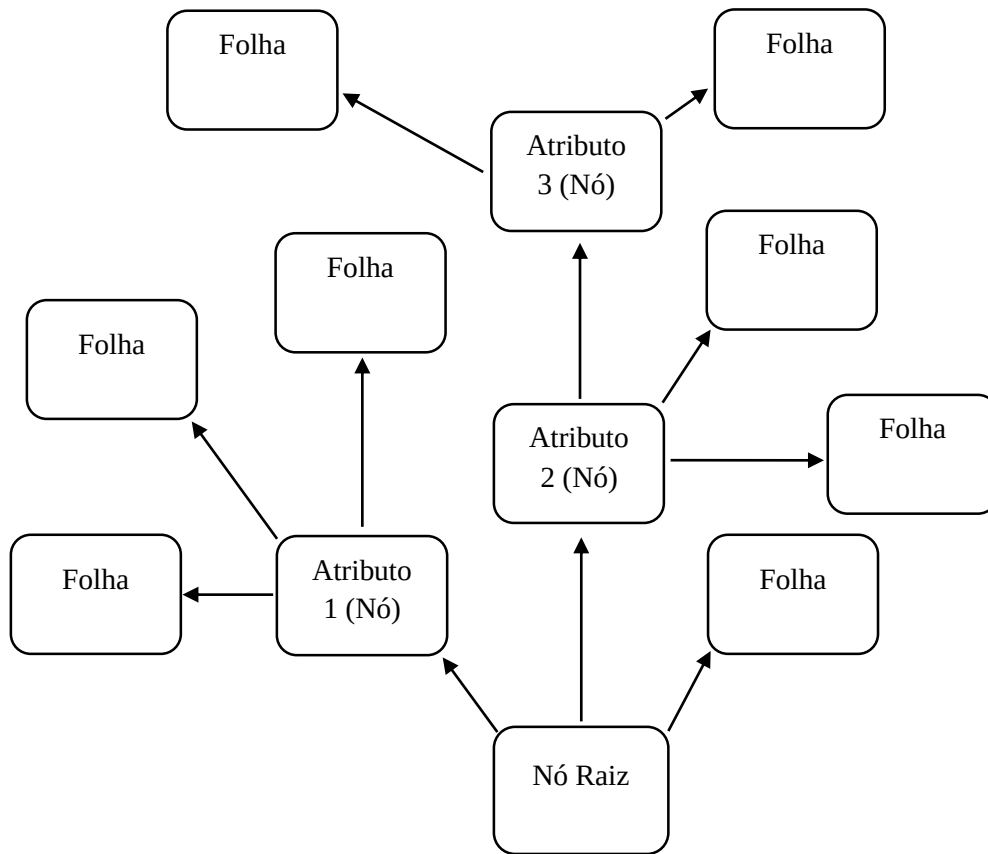


Figura 5: Exemplo fictício de uma árvore de decisão, com folha(s) (ou nós terminais) indicando uma classe, nó(s) internos contendo alguma pergunta/teste sobre o valor de um atributo específico (ou de um subconjunto de atributos) e ramos para cada um dos valores possíveis do teste.

Esses diagramas têm como entrada um objeto ou situação, descrito por um conjunto de atributos (ou propriedades), e como saída uma decisão (Figura 6), que é o valor de saída, predito de acordo com a entrada (Kingsford & Salzberg, 2008; Mitchel 1997; Pozzer, 2006). O resultado é uma estrutura de hierarquia chamada de árvore de decisão, com nós internos e folhas (nós terminais), conectados por ramos (Breiman *et al.*, 1984; Dunham, 2002).

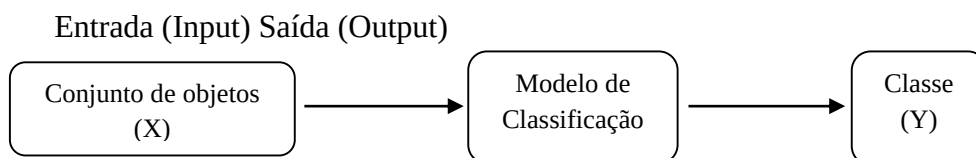


Figura 6: Diagrama do processo de classificação

Uma árvore de decisão chega a sua "decisão" pela execução de uma sequência de testes, em que os itens analisados são classificados de acordo com uma série de perguntas sobre as propriedades de cada atributo (Mitchel, 1999). Em geral, o procedimento de uma árvore de decisão é o seguinte: apresenta-se um conjunto de dados ao nó inicial (ou nó raiz, que também

é um nó interno) da árvore e, dependendo do resultado do teste lógico usado pelo algoritmo de indução escolhido, este nó ramifica-se para um dos nós filhos, ou descendentes, e este procedimento é repetido até que um nó terminal (folha) seja alcançado (Kingsford & Salzberg, 2008; Barbosa *et al.*, 2009; Silva *et al.*, 2011). Cada folha está associada a uma classe.

Um nó interno contém uma pergunta, ou teste, sobre o valor de um atributo, e conduza um novo nó para cada possível resposta à pergunta. Os ramos são descritos como os possíveis caminhos, ou um valor para cada atributo, para os resultados dos testes de cada nó (Russel & Novig, 2003); e as folhas especificam o valor de saída, caso a folha tenha sido atingida (Pozzer, 2006). Nas árvores de classificação, cada nó terminal ou folha contém uma única aresta de entrada, nenhuma de saída, e um rótulo que indica a classe predita. Por definição de árvore, existe um caminho único entre o nó raiz e qualquer nó folha. Logo, cada caminho na árvore (do nó raiz até um nó folha) corresponde a uma regra de classificação, facilitando a interpretação do modelo pelo usuário (Sun & Zhang, 2004; Kingsford & Salzberg, 2008; Coelho e Leite, 2015).

A árvore não tem hipóteses restritas em relação à distribuição da variável-alvo (dependente) e pode facilmente incorporar atributos categóricos ou numéricos (Breiman *et al.*, 1984; Thuraisingham, 1999; Vayssieres *et al.*, 2000).

Esses modelos de classificação podem ser usados para inúmeras tarefas. Porém, são especialmente adequados para os seguintes cenários (Mitchel, 1997):

- Há dados de treinamento disponíveis;
- A função utilizada tem valores de saída discretos;
- Conjunto de dados pode conter erros;
- Presença de dados faltantes no conjunto de dados.

O processo de construção e treinamento de uma árvore de decisão se chama indução, podendo exigir alta demanda computacional, e geralmente se utiliza de processos automáticos a partir dos dados disponíveis (Tan *et al.*, 2005). Todo processo de indução possui um *bias* indutivo, associado à preferência de uma hipótese sobre outras, uma vez que existam hipóteses igualmente consistentes (Kingsford & Salzberg, 2008; Otero *et al.*, 2012).

Para ilustrar o processo de indução, tomemos como exemplo a Figura 7, que representa um conjunto de dados fictício, semelhante àquele que foi foco do presente estudo: espécies de insetos sendo alocadas em determinados locais com base em suas características (tamanho, idade etc.). Para cada espécie, são definidos atributos contínuos (Ex: idade, tamanho etc.) e categóricos não-ordinais (Ex: sexo). O classificador deve mapear os atributos e, de acordo com eles, alocar as espécies em determinados locais.

Primeiro, o conjunto de treinamento, em que as classes dos objetos são conhecidas, é utilizado por um algoritmo de aprendizado para construir um modelo. Uma vez construído, o classificador pode ser aplicado para prever os rótulos de classes dos exemplos de um conjunto teste, ou seja, exemplos cujas classes são desconhecidas.

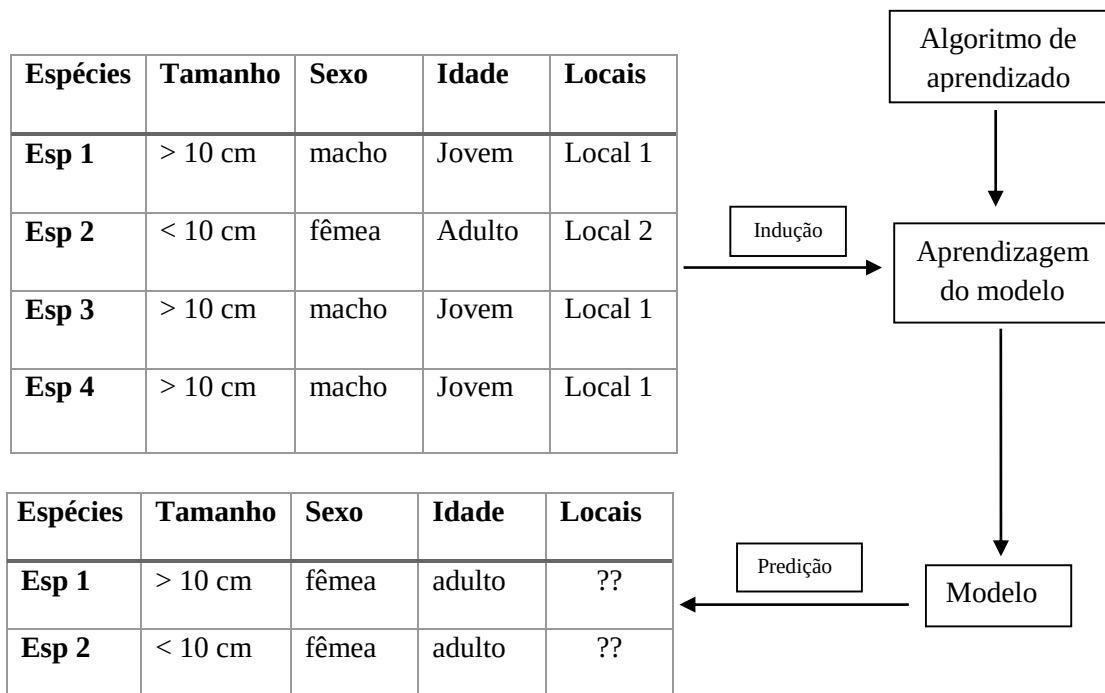


Figura 7: Processo de indução de um classificador e predição para novos objetos.

Diferente de outros modelos de classificação (redes neurais artificiais, *k*-vizinhos mais próximos etc.), a grande vantagem das árvores de decisão é que o conhecimento adquirido é representado por meio de regras, que podem facilmente ser expressas em linguagem natural, facilitando o entendimento por parte do leitor.

No presente estudo, é desejável que o modelo seja de fácil interpretação, deixando claro ao usuário o porquê de cada objeto estar alocado em determinada classe, e este é exatamente nosso

objetivo: a interpretação. Não aplicamos árvores de decisão apenas para termos um bom classificador, mas também porque elas nos dão regras que explicam o motivo de determinado fato ocorrer sob determinadas circunstâncias.

É importante mencionar que geralmente não é possível encontrar uma estrutura ótima de árvores de decisão para determinado problema, já que o número de árvores de decisão possíveis cresce fatorialmente com o número de atributos, logo, se torna inviável a busca por uma "árvore ideal" para o problema que se quer resolver (Tan *et al.*, 2006; Basgalupp, 2010; Otero *et al.*, 2012). Com isso, na busca por um algoritmo para indução de árvores, têm-se utilizado aqueles baseados em heurística, que é uma técnica em que se busca encontrar a melhor solução para determinado problema, num período de tempo relativamente curto (com relação aos métodos tradicionais). Mesmo que a solução não seja a melhor possível, os resultados apresentados são satisfatórios e o período de tempo despendido no processo é viável.

Há muitos algoritmos utilizados para a construção de árvores de decisão, entre eles os mais populares são C4.5 (Quinlan, 1993), ID3 (Quinlan, 1986) e CART (Breiman *et al.*, 1984). Neste estudo, foi utilizado o algoritmo C4.5 (Quinlan, 1993), que representa uma significativa evolução do algoritmo ID3 (Quinlan, 1986).

As principais diferenças do C4.5 (Quinlan, 1993) em relação ao ID3 (Quinlan, 1986) são: o primeiro aceita tanto valores contínuos como categóricos, consegue lidar com valores faltantes, usa medidas de razão de ganho para selecionar atributos que melhor dividem os exemplos e possui métodos de pós-poda da árvore gerada.

O C4.5 está disponível na ferramenta de mineração de dados WEKA (Hall *et al.*, 2009) (<http://www.cs.waikato.ac.nz/~ml/weka/index.html>) (*Waikato Environment for Knowledge Analysis*). No WEKA (Hall *et al.*, 2009), o algoritmo C4.5 é chamado de J48, que é uma implementação em Java do algoritmo C4.5.

O C4.5 (Quinlan, 1993) utiliza como base o paradigma TDIT (*Top-down induction of decision trees*) que produz regras de decisão de forma implícita numa árvore de decisão, a qual é construída por sucessivas divisões dos exemplos de acordo com os valores de seus atributos preditivos (Quinlan, 2014). C4.5 (Quinlan, 1993) é um algoritmo de busca recursiva do tipo "guloso", executando sempre o melhor passo avaliado localmente, sem se preocupar se este passo, junto à sequência completa de passos, vai produzira melhor solução ao final (Quinlan, 2014; Tan *et al.*, 2006; Otero *et al.*, 2012; Giasson *et al.*, 2013). Para gerar árvores de classificação, a partir de um conjunto de dados de treinamento, o algoritmo se utiliza da técnica

"dividir para conquistar", que gera sub-árvores da raiz até as folhas, criando subdivisões dos problemas (teste) presentes em cada nó, estas sendo mais simples que o problema original (Otero *et al.*, 2012; Giasson *et al.*, 2013). Este processo é repetido até que todos os objetos tenham sido classificados, ou todos os atributos preditivos utilizados (Tan *et al.*, 2006).

Um único atributo é testado em cada nó interno da árvore. Para fazer a escolha do melhor atributo preditor (teste) para cada nó teste, o C4.5 (Quinlan, 1993) se utiliza de um critério baseado em entropia, chamado de ganho de informação (Kingsford & Salzberg, 2008; Apté & Weiss 1997; Quinlan, 1993). Primeiramente, a entropia nada mais é do que a medida de heterogeneidade (impureza) de um conjunto de objetos de entrada, relativa à sua classificação (Mitchell 1997; Otero *et al.*, 2012). Em árvores de decisão, se utiliza a entropia para estimar o quão aleatória é a variável a se prever, no caso, a classe. A entropia de um conjunto de entrada S com m classes é dada por:

$$Entropia(S) = \sum_{c=1}^m - P_c \cdot \log_2 P_c$$

Em que P_c é a proporção de dados em S , que pertencem à classe c , e m é o número total de classes (Otero *et al.*, 2012). Dado um conjunto de objetos, qual atributo preditor escolher? Os valores do atributo vão definir como é dividido o conjunto de objetos e, para determinar o quão bom é o atributo, basta comparar o grau de entropia do nó de origem (antes da divisão) com aquele resultante (após a divisão); o que apresentar uma maior diferença e maximizar o ganho de informação é escolhido como atributo teste daquele nó (Basgalupp, 2010). O ganho de informação é dado por:

$$Ganho\ de\ informação(S, A) = Entropia(S) - \sum_{v=1}^T \frac{|S_v|}{|S|} \cdot Entropia(S_v)$$

O ganho de informação de um atributo A corresponde à redução de entropia esperada, conseguida ao se dividir o conjunto de objetos em T subconjuntos, em que T é o número de diferentes valores no domínio do atributo A , ou seja, o número de nós filhos. Sé o número total de objetos do nó origem e S_v é o número de exemplos associados ao nó resultante v .

Para melhor compreender como funciona a escolha do atributo preditor, tomemos como exemplo o seguinte problema de classificação (Figura 8), envolvendo duas classes:

Suponha que um nó interno esteja com a entropia máxima, ou seja 50% das amostras dele são da Classe A e 50% são da classe B. Isso produz um valor de entropia para o nó igual a 1. Vamos considerar 20 objetos, o que produz 10 para a Classe A e 10 para a Classe B. Suponha

que existam 2 atributos candidatos a serem escolhidos para a decisão deste nó. E cada um desses atributos admite 2 valores possíveis apenas, o que leva a apenas 2 nós-filho. O atributo 1 particiona os 20 objetos, conforme a situação ilustrada abaixo, e o atributo 2 faz uma outra partição, também indicada abaixo.

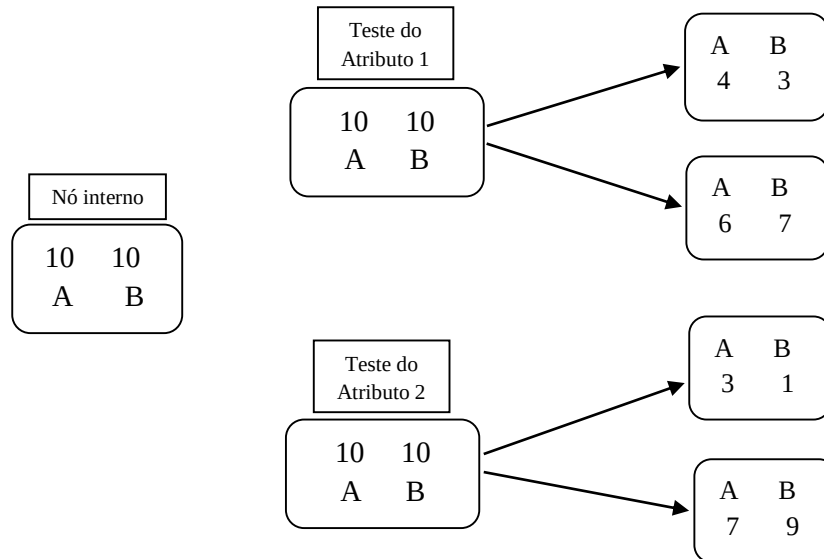


Figura 8: Teste para a escolha de um atributo preditor, mostrando o particionamento do conjunto de objetos por ambos os atributos.

A entropia máxima do Nó é:

$$Entropia = -(0.5 \cdot \log_2 0.5 + 0.5 \cdot \log_2 0.5) = 1$$

O ganho de informação de cada atributo é:

$$\begin{aligned} \text{Ganho de informação (Atributo 1)} &= 1 - \frac{7}{20} \left[\frac{4}{7} \log_2 \left(\frac{4}{7} \right) + \frac{3}{7} \log_2 \left(\frac{3}{7} \right) \right] + \\ &\frac{13}{20} \left[\frac{6}{13} \log_2 \left(\frac{6}{13} \right) + \frac{7}{13} \log_2 \left(\frac{7}{13} \right) \right] = 0.007947 \end{aligned}$$

$$\begin{aligned} \text{Ganho de informação (Atributo 2)} &= 1 - \frac{4}{20} \left[\frac{3}{4} \log_2 \left(\frac{3}{4} \right) + \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \right] + \\ &\frac{16}{20} \left[\frac{7}{16} \log_2 \left(\frac{7}{16} \right) + \frac{9}{16} \log_2 \left(\frac{9}{16} \right) \right] = 0.046785 \end{aligned}$$

Qual atributo escolher dentre os dois? Com base no ganho de informação, a escolha recai sobre o atributo 2, pois o ganho é maior.

A construção de uma árvore de decisão busca diminuir a entropia, ou seja, a aleatoriedade (dificuldade de previsão) da variável alvo.

Contudo, há um problema em escolher o ganho de informação, pois ele dá preferência aos atributos com muitos valores, que particionam o conjunto de dados em muitos subconjuntos pequenos (muitas arestas). É mais provável que o subconjunto seja puro (sem entropia) se existir um grande número de valores. Selecionar esses tipos de atributos pode resultar em baixo desempenho (Mitchel, 1997).

Para contornar esse problema, ajusta-se o ganho de informação com a entropia de particionamento, que nada mais é do que a entropia do nó, conhecida também como equação SplitINFO, que usa o ganho de informação relativo como critério de avaliação. Desta forma, particionamentos de alta entropia (que resultam em um grande número de pequenos subconjuntos) são penalizados (Mitchel 1997; Tan *et al.*, 2006; Otero *et al.*, 2012).

$$SplitINFO(S, A) = \sum_{v=1}^T - \frac{|Sv|}{S} \cdot \log_2 \frac{|Sv|}{S}.$$

Fazendo uso de ambas as equações, consegue-se a Razão de Ganho, uma modificação do Ganho de Informação, que leva em conta a entropia de particionamento, reduzindo, desta forma, o erro associado a atributos com muitas ramificações (Quinlan 1993).

$$Razão\ de\ Ganho(S, A) = \frac{Ganho\ de\ informação(S, A)}{SplitINFO(S, A)}$$

A Razão de ganho, cuja fórmula encontra-se acima, leva em conta o número de ramos quando se escolhe um atributo.

Após o atributo escolhido ter particionado o conjunto de treino de acordo com seu valor, sucessivamente outros atributos são selecionados para particionar os subconjuntos gerados. Esse processo perdura enquanto houver subconjuntos contendo objetos de classes diferentes, ou enquanto houver atributos a serem testados (Lee *et al.*, 2005). Uma vez obtido um subconjunto uniforme (onde todos, ou quase todos, os objetos pertençam à mesma classe) um nó folha é criado com o nome da respectiva classe (Lee *et al.*, 2005).

Para mais detalhes das distinções de abordagem para os diferentes tipos de atributos, consultar Quinlan (1993). Atributos numéricos e categóricos ordinais recebem um tratamento similar, onde se define um limiar de bipartição dos dados, enquanto que atributos categóricos não ordinais geram tantos nós filhos quanto categorias.

4.2 REDES BAYESIANAS

Uma vez interpretadas as diferentes interações entre os locais e as espécies, nosso próximo passo foi encontrar uma maneira de determinar quais das variáveis analisadas condicionavam e/ou eram condicionadas pelas outras. Para isso, utilizamos as redes bayesianas, uma ferramenta que nos trouxe informações extremamente promissoras acerca dos dados coletados. Atualmente estas ferramentas estão entre as mais modernas técnicas existentes para a análise e modelagem de dados (Laskey & Mahoney, 2000; Korb & Nicholson, 2004), e vem sendo aplicadas nos mais diversos campos de pesquisa, desde análise de expressões gênicas para elucidar relações entre os genes (Friedman et al., 2000), até problemas de diagnóstico médico e análises de falhas em sistemas computacionais (Antal *et al.*, 2003; Horvitz *et al.*, 2001).

De acordo com William Bolstad (2007), algumas das vantagens de se usar Redes Bayesianas em relação à estatística frequentista são:

1. A "objetividade" da estatística frequentista é frequentemente obtida através da desconsideração qualquer conhecimento prévio sobre o processo a ser medido. No entanto, na ciência há geralmente algum conhecimento prévio sobre o processo a ser medido. Estatística Bayesiana utiliza duas fontes de informação: o prior informação que temos sobre o processo e as informações sobre o processo de contido nos dados. Eles são combinados usando o teorema de Bayes.
2. A abordagem bayesiana permite avaliações diretas de probabilidade sobre os parâmetros. Isso é muito mais útil para um pesquisador do que as declarações de confiança permitidos pela estatística frequentista.
3. As estatísticas Bayesianas tem uma única ferramenta, o teorema de Bayes, o qual é utilizado em todas as situações.
Isto contrasta com os procedimentos frequentistas, que requerem muitas ferramentas diferentes.
4. Métodos Bayesianos muitas vezes superam os métodos frequentistas, mesmo quando julgado por critérios frequentistas.

5. Estatística Bayesiana tem uma maneira simples de lidar com parâmetros de perturbação. Eles estão sempre marginalizados fora da distribuição conjunta a posteriori.
6. O teorema de Bayes nos dá o caminho para encontrar a distribuição preditiva de observações futuras. Isso nem sempre é facilmente feito de uma maneira frequentista.

As Redes Bayesianas são definidas como modelos estatísticos multivariados para um conjunto de variáveis $A = \{A_1, \dots, A_n\}$ (Jensen & Nielsen, 2007), em que sua estrutura (ou, topografia) é a de um Grafo acíclico dirigido (DAG), com os nós representando as variáveis aleatórias (eventos que queremos modelar) e as arestas direcionadas indicando a dependência estatística que as variáveis possuem entre si (Figura 9); um nó afeta, ou causa, o outro de acordo com a direção das arestas (Charniak 1991; Heckerman, 1995; Friedman, 2003).

Para facilitar o entendimento, podemos dizer que um nó é o “pai” de um “nó filho”, quando há uma aresta partindo do primeiro e chegando no segundo.

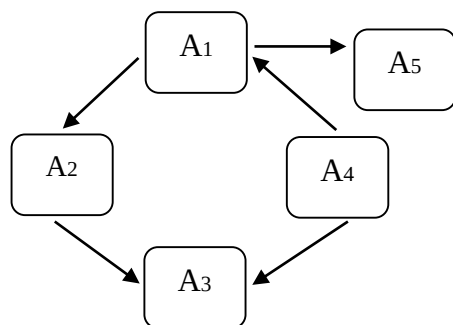


Figura 9: Representação fictícia de um Grafo Acíclico Dirigido, com as variáveis e suas relações de dependência. O mais comum é que Redes Bayesianas representem relações de condicionalidade entre variáveis. No exemplo acima, podemos considerar que A1 é a "causa" de A5.

Cada nó possui uma tabela de probabilidades, dizendo as chances de ocorrer o evento representado por ele (Pearl, 1988; Barber, 2012; Kajerlff, 2005; Darwiche, 2009). Se for um nó filho, essas probabilidades são condicionadas aos eventos associados aos nós pais.

Através da estrutura do DAG associado, é possível determinar a dependência e independência das relações entre as variáveis, de modo que se pode descobrir, sem a necessidade de realizar quaisquer cálculos numéricos, quais variáveis são relevantes ou irrelevantes para alguma outra variável de interesse (Aguilera *et al.*, 2011).

Há muitas outras vantagens associadas a esse tipo de modelo (Uusitalo, 2007; Aguilera *et al.*, 2010), entre elas: são adequados a conjuntos de dados incompletos, possibilitam um aprendizado estrutural, combinam diferentes fontes de conhecimento, promovem um tratamento explícito de incertezas e fornecem suporte para análises de decisão, além de permitirem respostas rápidas.

Estas redes modelam a força quantitativa das conexões entre variáveis, permitindo que as crenças probabilísticas (acreditar em determinado evento baseado em suas chances de ocorrência) sobre elas sejam automaticamente atualizadas conforme novas informações forem disponibilizadas (Korb & Nicholson, 2010)

Uma vez definida a estrutura, precisamos quantificar as relações entre os nós conectados. Isso é feito especificando-se uma distribuição de probabilidade condicional para cada nó (Pearl, 1988; Korb & Nicholson, 2010) dada por: $p(X_i/pa(X_i))$, para cada variável X_i , $i = 1, 2, 3, \dots, n$, dados seu “nós pais” (de origem) no grafo, denotados como $pa(X_i)$. Suponhamos, por exemplo, que o grafo da Figura 9, fosse o componente qualitativo de uma Rede Bayesiana e se quisesse encontrar a distribuição condicional de cada nó, tal distribuição seria $p(A1/A4)$, $p(A2/A1)$, $p(A3/A2, A4)$, $p(A4)$, $p(A5/A1)$ (Aguilera *et al.*, 2011).

Para entendermos melhor como funciona esse processo, tomemos como exemplo as seguintes questões sobre um problema adaptado de Koller e Friedman (2009):

A - Com base na Rede Bayesiana abaixo (Figura 10), qual a probabilidade de uma pessoa ter câncer, sabendo que ela fuma?

$$P(C|F) = 0.7$$

B - Qual a probabilidade de uma pessoa ter bronquite, sabendo que ela fuma?

$$P(B|F) = 0.8$$

C - Qual a porcentagem de pessoas que fumam?

$$P(F) = 0.4$$

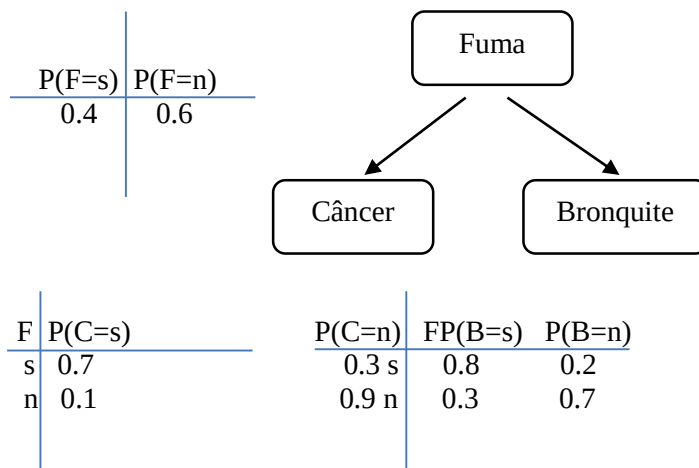


Figura 10: Rede bayesiana representando um problema envolvendo duas doenças associadas ao hábito de fumar (Koller e Friedman, 2009)

Podemos notar que cada nó tem sua tabela de probabilidades, e as arestas estão indicando as relações entre eles.

Há um conceito bastante importante no processo de construção de Redes Bayesianas que é preciso levar em consideração, o de Markov Blanket do nó. Ele consiste no conjunto de nós formados pelos “pais” e “filhos” de um nó e pelos outros pais dos filhos do nó (Friedman & Koller, 2003). A propriedade markoviana afirma que estados anteriores são irrelevantes para a predição dos estados seguintes, desde que o estado atual seja conhecido, ou seja, as variáveis não descendentes não fornecem informações adicionais sobre a variável em questão (Friedman *et al.*, 1997, Cheng *et al.*, 1999).

Em geral, nas modelagens de Redes Bayesianas, se requer que a condição de Markov Blanket seja verdadeira, não havendo dependências diretas no sistema que está sendo modelado que não estejam sendo representadas através de arcos (Korb & Nicholson, 2010).

Com essa condição satisfeita, concluímos que a distribuição de probabilidade conjunta da rede é igual ao produto das probabilidades condicionais de todos os nós, dados os valores dos seus ‘pais’ (Cheng *et al.*, 1999; Barber, 2012). Sendo assim, utilizando o componente quantitativo da rede, podemos definir a distribuição de probabilidade conjunta de toda a rede como:

$$P(A_1, A_2, \dots, A_n) = \prod_{i=1}^n P(A_i | pa(A_i))$$

Sendo, A_i uma variável (nó) da rede e $pa(A_i)$ os seus ‘pais’.

As redes bayesianas podem ser vistas como uma representação compacta da distribuição de probabilidades conjuntas do universo do problema.

Tomemos como exemplo o problema descrito por Pearl (1988):

Você instalou um alarme contra roubos na sua casa, que dispara em caso de invasão. Infelizmente, o alarme é sensível a terremotos e, quando o alarme disparar, seus 2 vizinhos, João e Maria, disseram que vão te ligar. Porém, João, às vezes, confunde o som do seu alarme com o despertador, e Maria ouve música num volume alto e nem sempre escuta o alarme.

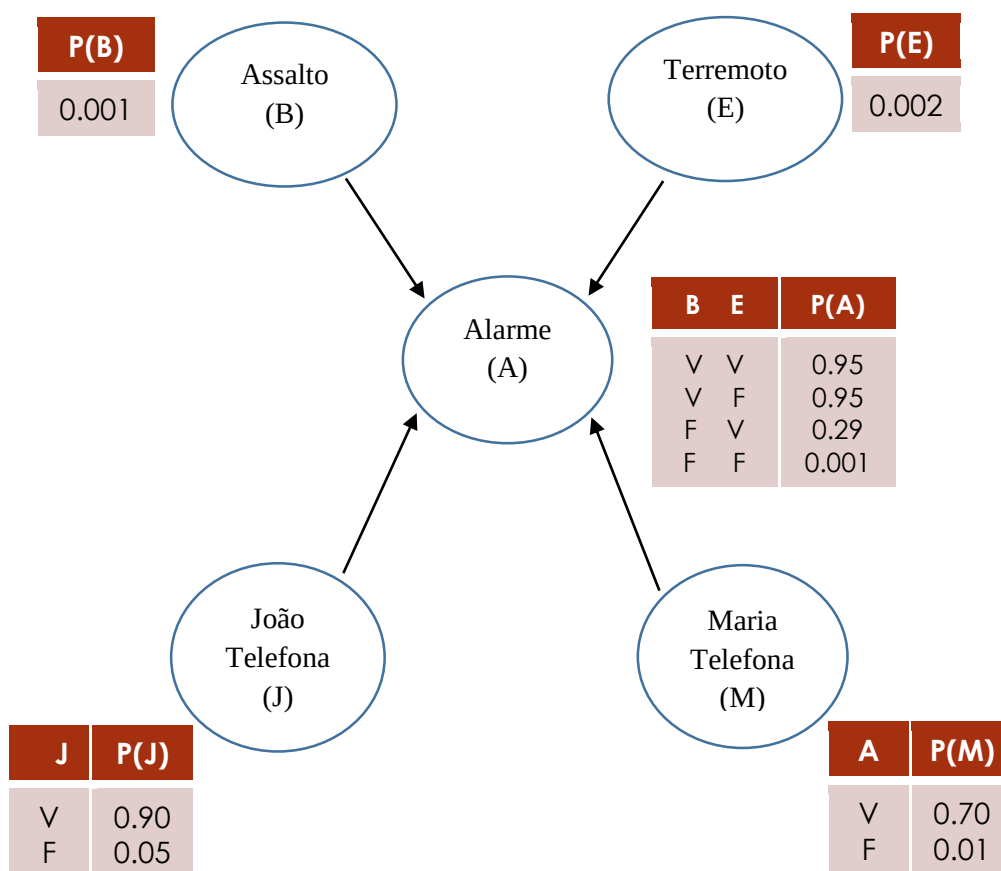


Figura 11: Tabelas de probabilidades associadas à ocorrência de cada evento em uma rede Bayesiana, representada por um grafo.

Qual é a probabilidade de não haver roubo, nem terremoto, o alarme tocar, João ligar e Maria ligar?

$$P(J|A) * P(M|A) * P(A|\neg L \wedge \neg T) * P(\neg L) * P(\neg T) = 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = 0.00062 \text{ ou } 0,062\%$$

4.3 INFERÊNCIA EM REDES BAYESIANAS

Inferência em Redes Bayesianas significa computar a probabilidade condicional para algumas variáveis, dadas informações (evidências) de outras variáveis (Jensen *et al.*, 1990), o que é uma tarefa fácil quando toda evidência disponível está presente nos “nós pais” da variável de interesse (como no exemplo anterior). Mas quando a evidência está presente apenas no “nó filho” da variável de interesse, temos que inferir sobre as variáveis anteriores, partindo da variável alvo. Para isso, basta aplicarmos o Teorema de Bayes (por isso o nome Redes Bayesianas), que pode ser descrito através da seguinte equação:

Dados dois eventos, **A** e **B**:

$$P(A|B) = P(B|A) P(A) / P(B)$$

Em que **P(A)** e **P(B)** são as probabilidades *a priori* de A e de B, respectivamente. **P(A|B)** é a probabilidade *a posteriori* de que ocorra A, uma vez que B tenha ocorrido. Já **P(B|A)** é a probabilidade *a posteriori* de ocorrência de B, dado que A tenha acontecido.

Para melhor compreender as relações entre os eventos, tomemos como exemplo o seguinte problema:

Um biólogo sabe que em um local A determinada espécie de insetos ocorre no sexo masculino (M) em 50% dos casos. Em sua coleta, ele sabe que a probabilidade *a priori* de um inseto ser proveniente do local A é 1/50000 e a probabilidade *a priori* de que qualquer inseto coletado seja do sexo masculino (M) é 1/20. Tem-se que:

$$P(M|A) = 1/2$$

$$P(A) = 1/50000$$

$$P(M) = 1/20$$

Um inseto do sexo masculino é coletado. Qual a probabilidade dele ocorrer no local A -

$$P(A|M) = ?$$

$$P(A|M) = P(M|A)P(A) / P(M) = (1/2 * 1/50000) / (1/20) = 1 / 5000 = 0,0002$$

4.4 APRENDIZADO EM REDES BAYESIANAS

No aprendizado em redes bayesianas busca-se, em um conjunto de dados (no nosso caso, os valores das variáveis referentes às borboletas frugívoras), encontrar uma rede que melhor represente as relações de dependência condicional entre as variáveis envolvidas. Para tanto, usa-se a abordagem de busca e pontuação, através de uma meta heurística de busca (mecanismos como: hill climbing, algoritmos genéticos e sistemas imunológicos artificiais) e, em seguida, calcula-se a verossimilhança da rede (pontuação) a partir dos dados amostrados (algumas penalidades de complexidade do modelo podem ser incorporadas em modelos mais complexos). Entre alguns dos métodos mais utilizados estão: *Bayesian Information Criterion* (BIC), *Akaike Information Criterion* (AIC) e *Minimum Description Length* (MDL).

Para a análise do conjunto de dados do presente estudo, usou-se o *Bayes Net Toolbox* (Murphy, 2001), que é uma extensão do software MATLAB (abreviatura de Matrix Laboratory). O MATLAB é um programa para desenvolvimento e implementação de algoritmos numéricos ou simbólicos, conhecido mundialmente como uma excelente ferramenta para a solução de problemas matemáticos, científicos e tecnológicos (Trindade & Sampaio, 2002). O MATLAB ainda possui uma série de *toolboxes*, ou seja, conjuntos abrangentes de funções MATLAB cujo objetivo é resolver problemas de áreas específicas, tais como: processamento de sinais, projeto de sistemas de controle, simulação dinâmica de sistemas, identificação de sistemas, redes neurais, lógica fuzzy (nebulosa ou difusa), otimização de sistemas, *wavelets* e cálculo simbólico, e outras áreas (Alves *et al.*, 2007).

Uma vez que o conjunto de dados tenha sido apresentado ao software, a Rede Bayesiana é sintetizada automaticamente pela análise direta dos dados e é apresentada na forma de um grafo acíclico. Para evitar Redes Bayesianas muito complexas, geralmente limita-se o número de filhos associados a cada nó pai, sendo usual adotar o número máximo de dois nós filhos para cada nó pai.

5 RESULTADOS E DISCUSSÃO

5.1 BICLUSTERIZAÇÃO

Os resultados obtidos foram inteiramente baseados na Planilha 1, Apêndice A, com algumas modificações para aplicação do ferramental. Trata-se de uma matriz binária (seus elementos são 0 ou 1), onde as linhas representam as espécies de borboletas e as colunas representam os locais de coleta. Cada elemento da matriz, portanto, indica se a espécie correspondente àquela linha está ausente (0) ou presente (1) no local correspondente a coluna. Os biclusters extraídos dessa matriz de dados representam um subconjunto de linhas e colunas com todos os elementos iguais a zero (0) ou todos os elementos iguais a (1)

Uma vez que o algoritmo tenha sido aplicado sobre os dados, conseguimos encontrar, de maneira extremamente rápida, uma série de agrupamentos e associações exclusivas de determinadas espécies com algumas das áreas estudadas. Com isso, conseguimos identificar espécies mais e menos generalistas com relação aos cenários abordados e aquelas que ocorrem em apenas um tipo de cenário (perturbado ou não perturbado) pois, como mostrado em estudos anteriores, as subfamílias de Nymphalidae e reagem de maneira diferente aos distúrbios (Uehara-Prado *et al.*, 2007; Barlow *et al.*, 2008; Uehara-Prado *et al.*, 2009).

Para facilitar a interpretação dos resultados, vamos denominar as áreas do Núcleo Santa Virgínia (NSV), manejo da vegetação através de corte e queimada (área mais perturbada) de A1, já a área que sofreu processo de extração seletiva de madeira (perturbação mais branda, se comparada com a primeira) chamaremos de A2. Na Estação Biológica de Boracéia (EBB), a área mais perturbada (trilhas manejadas) foi denominada B1, e a área onde não há o manejo de trilhas, B2. E finalmente, na Reserva Biológica de Paranapiacaba (RBP), a área de borda (mais perturbada) foi denominada C1, e a área de interior, C2.

Espécies que aparecem simultaneamente e exclusivamente em A1, B1 e C1: *Catonephele acontius* (Biblidinae), *Epiphileorea* (Biblidinae), *Hamadryas feronia* (Biblidinae) e *Blepolenis batea* (Brassolinae).

Por aparecerem simultânea e exclusivamente apenas nessas áreas, pode-se inferir uma forte associação destas espécies com os locais descritos, e potencial para alocação destas na categoria de indicadores das áreas perturbadas. Vale lembrar que o foco deste estudo não foi a busca por bioindicadores, mas apresentar uma nova metodologia de elucidação das relações entre espécies e os locais onde foram coletadas, com uma abordagem diferente sobre dados não

previamente utilizados. Logo, é natural que a metodologia utilizada por Uehara-Prado *et al.* (2009), o *Indicator Species Analysis* (IndVal) (Dufrêne & Legendre, 1997), desenvolvida exclusivamente para validação de bioindicadores, apresente um outro enfoque e uso de diferentes tipos de dados (como abundância, por exemplo, um fator decisivo na validação de indicadores biológicos). Também não foi possível quantificar os graus de perturbação em cada área (do mais perturbado ao menos perturbado), o que poderia interferir na amostragem de espécies e posterior análise dos cenários. Por isso optamos por seguir a metodologia de Uehara-Prado, mantendo as áreas divididas em perturbadas e não perturbadas.

Fazendo uma comparação das espécies encontradas com aquelas apresentadas por Uehara-Prado *et al.* (2009), em relação às espécies consideradas bons indicadores de locais perturbados, não encontramos menção das espécies descritas acima como indicadores biológicos de áreas perturbadas. Atribuímos essa diferença na identificação de grupos de espécies às metodologias, aos diferentes tipos de dados utilizados por ambos os trabalhos e suas respectivas abordagens.

Com relação às espécies acima, membros da subfamília Biblidinae são geralmente encontrados em áreas de mata menos conservadas (Uehara-Prado *et al.*, 2009; da Silva, 2011), porém não restritos a elas. Logo, era de se esperar que encontrássemos representantes dessa família nestas áreas. A presença de espécies de borboletas está intimamente ligada a sua planta hospedeira (Shahabuddin & Terborgh, 1999; Brown & Freitas, 2000; Shahabuddin & Ponte, 2005; Barlow *et al.*, 2008). No caso das espécies acima, suas larvas alimentam-se de plantas das famílias: Anacardiaceae, Euphorbiaceae, Lauraceae, Leguminosae, Moraceae, Rutaceae, Sapindaceae, Ulmaceae, Urticaceae e Verbenaceae (SILVA *et al.*, 1968; JENKINS, 1983; 1984; 1985; 1986; 1990; ACKERY *et al.*, 1999), muitas típicas de regiões de borda, clareiras ou matas secundárias. Quanto aos indivíduos de Brassolinae, foi mostrado por Uehara-Prado *et al.* (2009), que os membros desta subfamília mostraram abundância elevada em apenas um cenário, A1 (Uehara-Prado *et al.*, 2009) o que é plausível, se considerarmos que suas larvas se alimentam de Arecaceae (palmeiras), Poaceae (bambus e gramíneas), Musaceae (bananas) e Cyperaceae (ciperáceas) (Srygley *et al.*, 1999; Beccaloni *et al.*, 2008), que são geralmente encontradas em locais de vegetação secundária em áreas perturbadas (Uehara-Prado *et al.*, 2009). No caso de *Blepolenis batea*, já foi sugerido previamente que a alimentação de suas larvas está restrita à família das Arecaceae (Srygley *et al.*, 1999).

Os cenários mais relevantes encontrados foram:

Espécies que sempre ocorrem em todos os cenários A1, A2, B1, B2, C1, C2:

Caligo beltrao (Brassolinae), *Dasyophthalma rusina* (Brassolinae), *Eryphanis reevesi* (Brassolinae), *Prepona amphimachus* (Charaxinae), *Morpho catenarius* (Satyrinae), *Forsterinaria necys* (Satyrinae), *Taydebis peculliaris* (Satyrinae).

Estas espécies foram consideradas como as mais generalistas, ocorrendo em todos os cenários. Uehara-Prado *et al.* (2007), descreveu Charaxinae e Satyrinae como possuindo maior abundância de indivíduos em áreas perturbadas, porém alguns trabalhos também descrevem as Satyrinae como indicadores de locais não perturbados, ou apenas associadas a locais não perturbados (Fucilini *et al.*, 2009; Silva, 2011).

Espécies que nunca aparecem nos cenários A1, B1, C1:

Caligo illioneus (Brassolinae), *Narope cyllarus* (Brassolinae), *Opoptera aorsa* (Brassolinae), *Opsiphanes quiteria* (Brassolinae), *Euptychia ernestina* (Brassolinae), *Paryphthimoides interjecta* (Satyrinae), *Pierella nereis* (Satyrinae), *Pseudodebis euptychidia* (Satyrinae), *Splendeuptychia hygina* (Satyrinae), *Taygetis mermeria* (Satyrinae).

Este cenário é inteiramente composto por espécies das subfamílias Brassoliane e Satyrinae que, embora tenham sido previamente descritas como indicadores de locais perturbados (Uehara-Prado *et al.*, 2009), apresentam aqui alguns exemplares que nunca ocorrem em áreas perturbadas. Algumas Satyrinae já foram associadas a locais não perturbados (Fucilini *et al.*, 2009; Silva, 2011) e Uehara-Prado *et al.*, (2007) também encontraram uma maior abundância de Brassoliane em locais menos perturbados, o que não pode ser explicado pela presença de suas plantas hospedeiras, em sua maioria espécies que ocorrem em áreas de mata secundária e áreas perturbadas (Penz *et al.*, 1999; Beccaloni *et al.*, 2008). Algumas hipóteses, como a adaptação de larvas de algumas Brassolinae aos microclimas mais amenos de áreas não perturbadas (Koh, 2007; Uehara-Prado *et al.*, 2009), o papel da territorialidade nos indivíduos adultos (Freitas *et al.*, 1997; Srygley & Penz, 1999) e alguns tipos de distúrbios ambientais desencadeando gatilhos fenológicos (Ribeiro *et al.*, 2010), embora não inteiramente testadas, parecem ser as que melhor explicam esse fenômeno (Uehara-Prado *et al.*, 2009). Não podemos descartar também a possibilidade de ocorrência dessas espécies em áreas perturbadas, mas sem nenhuma captura na coleta.

Espécies que nunca aparecem nos cenários A2, B2, C2:

Biblishyperia (Biblidinae), *Catonephele acontius* (Biblidinae), *Catonephele numilia* (Biblidinae), *Catonepheles abrina* (Biblidinae), *Diaethriacan drena* (Biblidinae), *Epiphile oreia* (Biblidinae), *Hamadryas februa* (Biblidinae), *Hamadryas feronia* (Biblidinae), *Temenisla othoe* (Biblidinae), *Blepolepis batea* (Brassolinae), *Caligo brasiliensis* (Brassolinae), *Memphis appias* (Charaxinae), *Memphis arginussa* (Charaxinae), *Memphis ryphea* (Charaxinae), *Prepona chalciope* (Charaxinae), *Morpho aega* (Satyrinae), *Smyrna blomfieldia* (Nymphalinae), *Manataria hercina* (Satyrinae), *Moneuptychia soter* (Satyrinae), *Pedaliodes phanias* (Satyrinae), *Praepedaliodes amusis* (Satyrinae), *Praepedaliodes phanias* (Satyrinae).

Estas espécies não foram encontradas em locais de vegetação mais conservada, com a subfamília Biblidinae apresentando o maior número de indivíduos, seguida de Satyrinae. Uehara-Prado *et al.* (2009) encontraram que Biblidinae possuiu uma abundância maior apenas em um cenário perturbado (NSV), Satyrinae é sempre mais numerosa em áreas perturbadas e Brassolinae foi registrada com mais indivíduos em áreas menos perturbadas. Nymphalinae não foi citada, provavelmente por não possuir uma abundância significativa.

A falta de associação destas espécies aos locais mais preservados se deve, provavelmente, ao fato de muitas Biblidinae, Satyrinae e Charaxinae possuírem indivíduos que se utilizam de plantas encontradas em locais de mata secundária como hospedeiras (Uehara-Prado *et al.*, 2007; Uehara-Prado *et al.*, 2009). Como dito anteriormente, apenas Brassolinae parece não ter seu comportamento explicado pelas plantas do local no qual foi encontrada, embora neste caso ela tenha sido a segunda que menos apresentou indivíduos não associados a locais menos perturbados, ficando atrás apenas de Nymphalinae.

Os demais cenários encontrados podem ser conferidos no Apêndice C.

5.2 ÁRVORES DE DECISÃO

Nesta seção, será descrita a variação de resposta das variáveis perante os diferentes cenários, utilizando todas simultaneamente ou então removendo algumas, identificando assim quais as mais apropriadas para a tarefa de classificação. Desta forma, conseguimos classificar as variáveis e descobrir quais estão especificamente associadas a cada cenário. Com isso pretendemos mostrar o potencial desta ferramenta para estudos de modelagem em Ecologia e

áreas afins, servindo tanto para classificação de objetos de diferentes classes quanto para a predição de classes para dados desconhecidos (Tan *et al.*, 2006).

Esta análise foi realizada com o algoritmo C4.5 (Quinlan, 1993) presente na ferramenta para mineração de dados WEKA (Hall *et al.*, 2009) (<http://www.cs.waikato.ac.nz/~ml/weka/index.html>) (*Waika to Environment for Knowledge Analysis*). Logo, a árvores construídas estão em inglês.

As variáveis utilizadas foram:

- (1) Região (são 3 regiões ao todo)
- (2) Local (Perturbado e Não perturbado, ou seja, são 2 ao todo)
- (3) Espécie (são 72 ao todo)
- (4) Subfamília (são 6 ao todo)
- (5) Idade estimada (são 5 níveis ao todo)
- (6) Sexo (macho ou fêmea, ou seja, são 2 ao todo)
- (7) Tamanho da Asa (foi quantizado em 6 níveis, conforme indicado a seguir)

Cada variável será discutida em detalhes no tópico seguinte (Redes Bayesianas).

Devido ao tamanho da árvore construída, esta foi alocada no Apêndice C. Abaixo, segue uma parte da primeira árvore elaborada, contendo todas as variáveis exceto por subfamília. O tamanho da asa é dado em milímetros e a divisão da idade se deu com a letra A representando os indivíduos mais novos e a letra E os mais velhos. Aqui, todos os locais foram divididos em perturbados e não perturbados, "managed" e "unmanaged trails", respectivamente.

```
Tamanho_da_asa <= 30 mm
|  Especie = Arc_cluena
| |  Idade_Estimada = A: unmanaged_trails (0.42)
| |  Idade_Estimada = B: managed_trails (0.0)
| |  Idade_Estimada = C: managed_trails (2.22/0.22)
| |  Idade_Estimada = D: managed_trails (1.11/0.11)
| |  Idade_Estimada = E: managed_trails (0.0)
|  Especie = Ble_batea: managed_trails (0.37)
```

- | Especie = Cal_arisbe: managed_trails (4.75/0.75)
- | Especie = Cal_beltrao: unmanaged_trails (3.24/0.37)
- | Especie = Car_griseldis

Figura 12: Fragmento de uma árvore de decisão construída utilizando-se de todas as variáveis, exceto por Subfamília.

Esta foi a árvore mais complexa, contendo o maior número de variáveis.

Para interpretá-la, vamos primeiramente recordar que um caminho completo entre a raiz e cada folha da árvore de decisão representa uma regra completa que leva à classificação de um subconjunto dos indivíduos em um dos rótulos de classificação possíveis. Sendo assim, ao se percorrer a árvore de decisão da raiz até uma folha específica, se chega a uma regra que leva a uma classe, dentre as classes existentes para os dados. Uma árvore, em estrutura de dados, é definida como um grafo com um caminho único entre quaisquer pares de nós. Logo, existe um caminho único (uma regra única) entre a raiz e qualquer das folhas da árvore. O classificador resultante, portanto, é composto por tantas regras quanto o número de folhas da árvore de decisão.

Os números entre parênteses, associados a cada folha indicam:

Primeiro número - Quantos indivíduos foram classificados por aquela regra com o rótulo da folha.

Segundo Número - Quando existente, indica quantos desse total de indivíduos do primeiro número foi erroneamente classificado pela regra.

Por se tratarem de indivíduos, os campos entre parênteses deveriam corresponder a números inteiros. Porém, o conjunto de dados possui campos faltantes e, dependendo do valor deste campo, o indivíduo vai para uma classe ou para outra. O algoritmo estima a probabilidade de pertencer a esta ou àquela classe pela frequência de cada valor do atributo (usando os outros campos em que o atributo está preenchido). Se não houvesse dados faltantes, os dois campos entre parênteses seriam sempre inteiros. Quando o segundo campo não aparece é porque aquela regra não produziu erros de classificação.

Tomemos como exemplo a primeira folha, A regra que ela representa é: Se [Tamanho_da_asa ≤ 30], [Especie = Arc_cluena] e [Idade_Estimada = D], então [Classe = managed_trails].

Essa regra representa 1.11 indivíduos, sendo que ela classifica corretamente 1.0 indivíduos e incorretamente 0.11 indivíduos. Ou seja, 0.11 indivíduos que atendem ao perfil do antecedente da regra foram na verdade coletados em [unmanaged_trails]. Essa única regra consegue rotular, portanto, apenas 1.11 dos indivíduos, e erra na rotulação de 0.11. As demais regras cuidam de indivíduos com outros perfis de atributos. Cada árvore de decisão possui uma tabela mostrando as instâncias classificadas correta e incorretamente.

Com este exemplo, é possível ver o grande potencial desta ferramenta para estudos ecológicos. A base de dados foi inteiramente organizada de maneira que o conhecimento possa ser extraído de forma prática e eficiente. Desta forma, podemos, por exemplo, determinar de forma simples qual o tamanho, idade e sexo de uma determinada espécie que ocorre em uma determinada região apenas seguindo as ramificações da árvore formada.

Tomando como exemplo outro cenário, onde não foram considerados os atributos [Especie] e [Idade_estimada], temos a seguinte árvore:

```
Tamanho_da_asa <= 30: managed_trails (704.12/89.97)
Tamanho_da_asa > 30
| Subfamilia = Satyrinae: managed_trails (162.29/57.13)
| Subfamilia = Brassolinae
| | Tamanho_da_asa <= 75
| | | Tamanho_da_asa <= 58
| | | | Regiao = E.B._Boraceia
| | | | Tamanho_da_asa <= 51: managed_trails (72.47/26.54)
| | | | Tamanho_da_asa > 51: unmanaged_trails (32.73/12.68)
| | | | Regiao = Núcleo_Santa_Virginia: unmanaged_trails (82.01/24.0)
| | | | Regiao = R.B._Paranapiacaba
```

Figura 13: Fragmento de árvore construída desconsiderando-se Espécie e Idade estimada.

Aqui podemos corroborar que a distribuição de indivíduos das subfamílias segue o padrão descrito por Uehara-Prado *et al.* (2009). Porém, com a árvore de decisão, conseguimos uma distribuição mais refinada, por conta dos atributos adicionados. Podemos identificar não somente onde está alocada a maior quantidade de indivíduos de cada subfamília, mas também seu tamanho, sexo e idade, tudo de maneira simples, já que os resultados são altamente interpretáveis.

Da mesma forma com que elaboramos árvores com locais perturbados (managed) e não perturbados (unmanaged), podemos construir uma árvore considerando cada local distintamente, como mostrado abaixo. Desta forma, refina-se ainda mais a classificação.

```

Especie = Arc_cluena
| Idade_Estimada = A: unmanaged_trails (1.25)
| Idade_Estimada = B: managed_trails (0.0)
| Idade_Estimada = C: managed_trails (2.5/0.5)
| Idade_Estimada = D: managed_trails (1.25/0.25)
| Idade_Estimada = E: managed_trails (0.0)
Especie = Ble_batea
| Idade_Estimada = A
| | Sexo = 0: managed_trails (2.8/1.4)
| | Sexo = 1
| | | Tamanho_da_asa<= 44: slash_and_burn (1.0)
| | | Tamanho_da_asa> 44: managed_trails (3.2/0.6)
| Idade_Estimada = B
| | Sexo = 0: managed_trails (2.8/1.4)
| | Sexo = 1
| | | Tamanho_da_asa<= 45
| | | | Tamanho_da_asa<= 44: slash_and_burn (1.0)
| | | | Tamanho_da_asa> 44: edge (1.0)
| | | Tamanho_da_asa> 45: managed_trails (2.2/0.6)
| Idade_Estimada = C: slash_and_burn (1.0)
| Idade_Estimada = D: managed_trails (0.0)
| Idade_Estimada = E: managed_trails (0.0)

```

Figura 14: Fragmento de árvore construída considerando cada local como distinto. Não foram utilizados os atributos Subfamília e Região

Nesta árvore foram desconsiderados os atributos Subfamília e Região.

Notamos que cada um dos tipos de perturbação está especificado, o que ajuda a melhorar ainda mais quaisquer análises ecológicas sobre esses dados, pois podemos identificar a qual perturbação específica cada uma das espécies está associada.

Sob determinadas condições, alguns atributos não se mostraram aptos à tarefa de classificadores. Na árvore apresentada no Apêndice C, vemos que quando considerados os seis locais como distintos, apenas os atributos Subfamília e Tamanho da Asa não são capazes de classificar as seis regiões.

Cabe salientar que os atributos mais próximos da raiz, que são os primeiros a serem testados, são os de maior poder discriminante das classes.

5.3 REDES BAYESIANAS

Primeiramente, vamos denominar as variáveis e seu significado.

- (1) Região (são 3 regiões ao todo)
- (2) Local (convertidos apenas para Perturbado e Não Perturbado, ou seja, são 2 ao todo)
- (3) Espécie (são 72 ao todo)
- (4) Subfamília (são 6 ao todo)
- (5) Idade estimada (são 5 níveis ao todo)
- (6) Sexo (macho ou fêmea, ou seja, são 2 ao todo)
- (7) Tamanho da Asa (foi quantizado em 6 níveis, conforme indicado a seguir)

A rede bayesiana foi gerada automaticamente por análise direta dos dados. Logo, as conclusões a seguir foram extraídas dos dados da Planilha 2, Apêndice B, e somente deles, pois a rede bayesiana não usou nenhuma outra informação. Todas as linhas com dados faltantes e linhas com Tamanho da Asa nulo foram desconsideradas, sobrando 1593 amostras.

A variável “Tamanho da Asa” foi quantificada em 6 níveis. Com exceção do Nível 1 (todos os integrantes pertencentes à Subfamília Satyrinae), que contém poucos indivíduos, os demais níveis possuem um número bastante equilibrado de indivíduos (este foi o critério para a definição do intervalo).

0mm =< Nível 1 < 18 mm

18mm =< Nível 2 < 25 mm

25mm =< Nível 3 < 37 mm

37mm =< Nível 4 < 48mm

48mm =< Nível 5 < 60mm

60 mm =< Nível 6

Cabe mencionar que, como dito anteriormente, uma rede bayesiana trata cada variável como sendo aleatória e indica a probabilidade de uma certa variável assumir um certo valor dado que as variáveis que a condicionam assumem certos valores.

Quanto as variáveis, efeitos ecológicos da heterogeneidade de habitats podem variar consideravelmente entre grupos de espécies, dependendo se os atributos estruturais do meio são percebidos como perturbação (Tews *et al.*, 2004)

O que pode ser interpretado da Rede Bayesiana apresentada na Figura 14 é:

- As variáveis 2, 4 e 6 não condicionam ninguém e são condicionadas por outras variáveis. Em outras palavras, as variáveis 2, 4 e 6 são as variáveis que podem ser estimadas a partir de outras variáveis.
- A variável (3) – Espécie condiciona (ou seja, torna mais provável o acerto na definição do valor de) as variáveis (1), (2) (4), (6) e (7). Ou seja, sabendo-se o valor da variável (3), eu posso estimar melhor o valor das variáveis (1), (2), (4), (6) e (7).
- A variável (5) – Idade estimada é independente das demais variáveis. Ou seja, não condiciona e não é condicionada por elas. Logo, o seu conhecimento não ajuda a entender o que ocorre com as demais variáveis.
- A variável (2) – Local, além de ser condicionada pela variável (3) – Espécie, também é condicionada pela variável (1) – Região. Em outras palavras, conhecendo a região de coleta e a espécie coletada, torna-se mais provável o acerto na definição do local (Perturbado ou Não Perturbado).
- A variável (6) – Sexo, além de ser condicionada pela variável (3) – Espécie, também é condicionada pela variável (7) – Tamanho da Asa. Em outras palavras, conhecendo a espécie e o tamanho da asa, torna-se mais provável o acerto na definição do sexo.

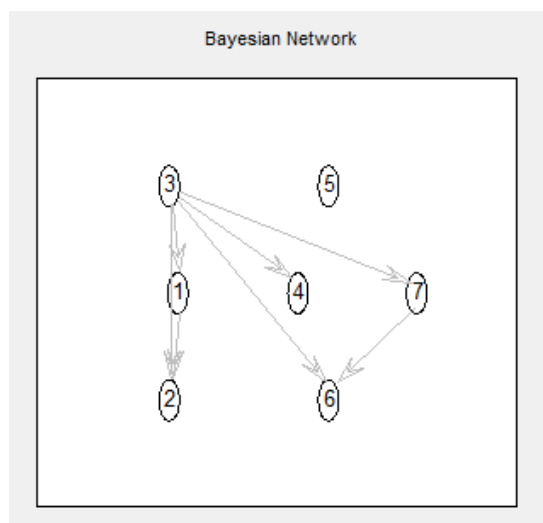


Figura 15: Rede Bayesiana representando todas as variáveis do problema em questão, quais condicionam e quais são condicionadas por outras. O desenho foi feito automaticamente pelo software *Bayes Net Toolbox*.

Na primeira análise, está clara a dependência das demais variáveis em relação à variável "Espécie", apontando esta como a mais significativa na determinação das outras. Tal resultado é previsível se levarmos em conta que a presença de borboletas frugívoras, bem como dos demais insetos, está associada à bionomia de cada espécie (Price *et al.*, 2011; Papaj *et al.*, 2012; Davies *et al.*, 2004; Fischer *et al.*, 2013; Souza-Silva, 2008; Mopper *et al.*, 2013).

Em se tratando da variável "Local", diretamente condicionada pela espécie, uma das principais características relacionadas à sua escolha é a planta hospedeira (Shahabuddin & Terborgh, 1999; Brown & Freitas, 2000; Shahabuddin & Ponte, 2005; Barlow *et al.*, 2008), que muitas vezes é uma característica que liga diferentes grupos de espécies (Tews *et al.*, 2004). Na maioria dos habitats, as comunidades de plantas determinam a estrutura física do ambiente e, portanto, têm uma influência considerável sobre as distribuições e interações de espécies animais (Lawton, 1983; McCoy & Bell, 1991). Quando na forma larval, as Nymphalidae são herbívoras. Logo, apresentam uma estreita relação com a vegetação de determinada área (Silva, 2011). Cada espécie busca um tipo diferente de local e região de acordo com suas preferências alimentares. Muitas espécies têm preferência por determinada família ou espécie de plantas (Singer, 1984; DeVries, 1987). A maioria das Biblidines e Charavines, por exemplo, foram encontradas em locais de mata secundária, onde ocorreram distúrbios, já que suas larvas se alimentam de plantas típicas de mata secundária da mata Atlântica (Uehara-Prado *et al.*, 2007). Tais distúrbios podem, muitas vezes, alterar a fenologia e os recursos alimentares das Nymphalidae (adultos e larvas). Esses efeitos são mais pronunciados em espécies de grande porte, como muitas Brassolinae (Ribeiro *et al.*, 2010). Ribeiro *et al.* (2010) também

encontraram que a temperatura média é outra importante característica que, além de estar ligada aos distúrbios ocorridos no local, afeta fortemente a ocorrência de espécies de borboletas frugívoras. Alguns estudos anteriores indicam haver uma correlação entre a variável tamanho (e morfologia geral do corpo) e o local de ocorrência (Breuker *et al.*, 2007; Norberg & Leimar, 2002) em populações de borboletas, o que provavelmente está associado às diferentes respostas.

Outra explicação é a territorialidade: indivíduos adultos de muitas espécies optam por se utilizar das características de determinado cenário para estabelecer seu território (Uehara-Prado *et al.*, 2009); tais características podem mudar mediante algum distúrbio, fazendo com que as espécies precisem realocar seus territórios.

A determinação da variável "Sexo" com base nas variáveis "Tamanho da Asa" e "Espécie", está associada ao forte dimorfismo sexual presente em Nymphalidae (Silva, 2011; Breuker *et al.*, 2007; Uehara-Prado *et al.*, 2004; Uehara-Prado *et al.*, 2005; Santos *et al.*, 2011). Tal dimorfismo, expresso principalmente nas asas, é muito comum em Lepidoptera, e decorre dos diferentes papéis ecológicos desempenhados por cada sexo, que requerem diferentes tipos de padrões de voo para diferentes propósitos biológicos (Wickman, 1992; Van Dyck & Wiklund, 2002; Berwaerts *et al.*, 2006).

A maioria das fêmeas apresenta um tamanho maior em relação aos machos da mesma espécie, o que provavelmente está associado à sua alta dispersão no ambiente em busca de plantas hospedeiras, tendo necessidade de voar por distâncias maiores (Bets & Wootton, 1988; Ehrlich 1984, Freitas 1996, Ramos & Freitas 1999; Uehara-Prado *et al.*, 2005; Breuker *et al.*, 2007), embora alguns trabalhos não tenham encontrado diferença entre as taxas de dispersão de machos e fêmeas (Breuker *et al.*, 2007). Já o design do corpo dos machos parece ser mais fortemente influenciado pelo sistema de acasalamento de cada espécie do que pela dispersão (Breuker *et al.*, 2007; Wickman, 1992; Wahlberg *et al.*, 2002).

A segunda análise foi feita desconsiderando-se a variável "Espécie". O resultado está apresentado na Figura 15 a seguir:

Variáveis envolvidas:

- (1) Região (são 3 regiões ao todo)
- (2) Local (convertidos apenas para Perturbado e Não Perturbado, ou seja, são 2 ao todo)
- (3) Subfamília (são 6 ao todo)
- (4) Idade estimada (são 5 níveis ao todo)

- (5) Sexo (macho ou fêmea, ou seja, são 2 ao todo)
- (6) Tamanho da Asa (foi quantizado em 6 níveis, conforme já apresentado)

Com a saída da variável “Espécie”, fica evidente que o papel condicionante das demais variáveis foi assumido pela variável (3) – “Subfamília”, que condiciona as variáveis (1), (2), (5) e (6). Como foi retirada uma variável fortemente condicionante de quase todas as outras, condicionamentos desprezados antes afloraram e passaram a ser considerados relevantes agora.

Semelhante ao papel da variável Espécie na análise anterior, aqui Subfamília representa praticamente o mesmo papel, já que indivíduos dentro de uma subfamília apresentam características semelhantes.

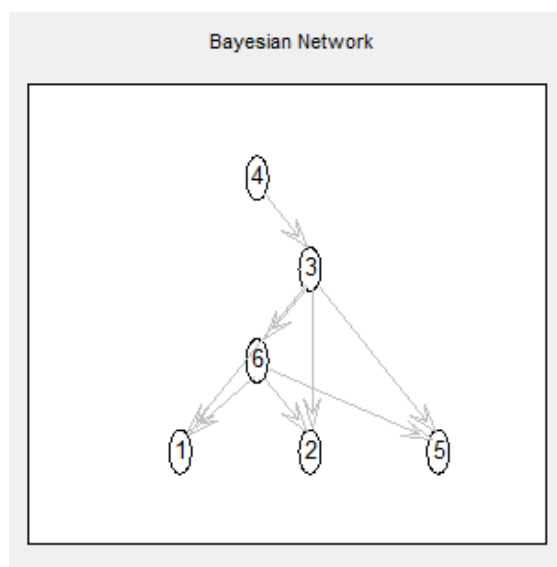


Figura 16: Rede Bayesiana representando todas as variáveis, exceto Espécie. O desenho foi feito automaticamente pelo software *Bayes Net Toolbox*.

É fundamental observar que os dados apontam que as variáveis 1, 2 e 5 não condicionam ninguém. Porém, são condicionadas por outras variáveis. Em outras palavras, as variáveis 1, 2 e 5 são as variáveis que podem ser estimadas a partir de outras variáveis.

Tanto “Região” quanto “Local” estão condicionadas pelas variáveis “Subfamília” e tamanho da asa; “Sexo” continua condicionada por tamanho da asa e agora também por subfamília. “Idade estimada” é uma variável independente (não é condicionada por ninguém, como já havia sido constatado antes), mas agora passou a condicionar a subfamília. Ou seja, conhecendo-se a idade estimada pode ajudar a indicar a subfamília. Isso não foi evidenciado na rede bayesiana anterior por não representar uma relação condicionante relevante, comparada às

demais. Mas é importante deixar claro que, com o conhecimento da subfamília (variável 3), a rede bayesiana indica que as variáveis 1, 2, 5 e 6 não dependem da variável 4.

A terceira análise foi feita desconsiderando-se a variável “Região” e a variável “Espécie”. O resultado é apresentado na Figura 17.

Variáveis envolvidas:

- (1) Local (convertidos apenas para [*managed*] e [*unmanaged*], ou seja, são 2 ao todo)
- (2) Subfamília (são 6 ao todo)
- (3) Idade estimada (são 5 níveis ao todo)
- (4) Sexo (macho ou fêmea, ou seja, são 2 ao todo)
- (5) Tamanho da Asa (foi quantizado em 6 níveis, conforme já apresentado)

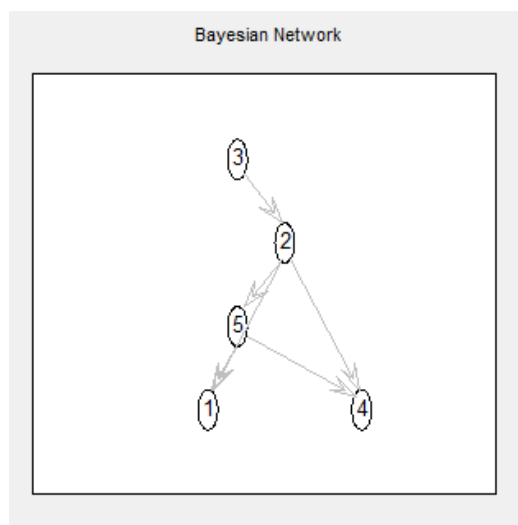


Figura 17: Rede Bayesiana representando todas as variáveis, exceto “Região” e “Espécie”. O desenho foi feito automaticamente pelo software *Bayes Net Toolbox*.

Neste caso, apenas “Subfamília” determina o local. Sexo continua sendo determinado por subfamília e tamanho da asa.

A quarta análise foi feita desconsiderando-se a variável “Região” e a variável “Subfamília”. O resultado é apresentado na Figura 18.

Variáveis envolvidas:

- (1) Local (convertidos apenas para Perturbado e Não Perturbado, ou seja, são 2 ao todo)
- (2) Espécie (são 72 ao todo)
- (3) Idade estimada (são 5 níveis ao todo)
- (4) Sexo (macho ou fêmea, ou seja, são 2 ao todo)
- (5) Tamanho da Asa (foi quantificado em 6 níveis, conforme já apresentado)

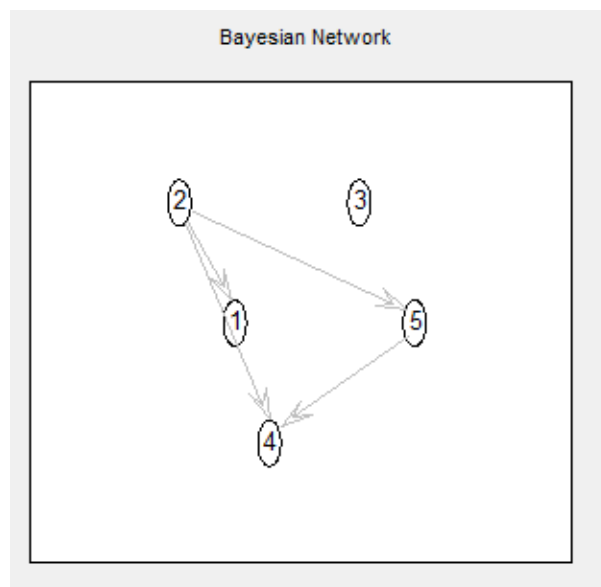


Figura 18: Rede Bayesiana representando todas as variáveis, exceto “Região” e “Subfamília”. O desenho foi feito automaticamente pelo software *Bayes Net Toolbox*.

Chegou-se apenas a uma sub-rede já obtida na análise 1, mostrando a sua robustez à retirada de 2 variáveis do estudo. Local e sexo continuam sendo as variáveis dependentes de outras.

Na quinta análise desconsiderou-se a variável "Tamanho da Asa". O resultado é apresentado na Figura 19.

Variáveis envolvidas:

- (1) Região (convertidos apenas para Perturbado e Não Perturbado, ou seja, são 2 ao todo)
- (2) Local
- (3) Espécie (são 72 ao todo)
- (4) Subfamília
- (5) Idade estimada (macho ou fêmea, ou seja, são 2 ao todo)
- (6) Sexo (foi quantificado em 6 níveis, conforme já apresentado)

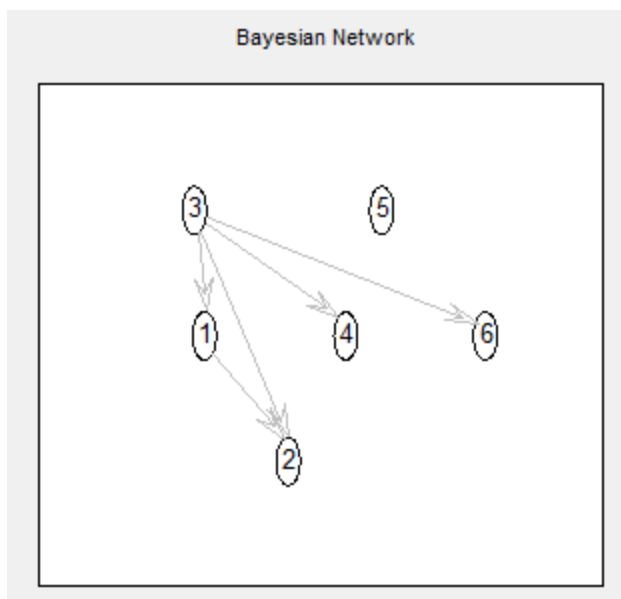


Figura 19: Rede Bayesiana representando todas as variáveis, exceto “Tamanho da Asa”. O desenho foi feito automaticamente pelo software *Bayes Net Toolbox*.

Notamos novamente a dependência da rede em relação à “Espécie”. Também é interessante ressaltar que nenhuma das análises apresentou uma relação direta entre “Idade Estimada” e “Tamanho da Asa”.

Há uma tabela de probabilidade associada a cada variável, fornecida junto com a Rede Bayesiana, mas elas têm dimensões enormes, pois as variáveis têm muitos valores possíveis, particularmente a variável (3). Logo, elas não foram apresentadas e nem utilizadas para fundamentar as conclusões apresentadas acima.

6 CONCLUSÃO

O modelo de biclusterização utilizado se mostrou bastante rápido e eficiente, com o InClose sendo capaz de encontrar todos os padrões de correlação condicional “escondidos” dentro da matriz de dados, através da análise dos *biclusters* maximais. Os *clusters* identificados estão de acordo com a maioria dos dados sobre ocorrência dessas espécies, descritas previamente na literatura, embora a falta de dados sobre a bionomia de muitas espécies não permita um entendimento maior nas relações entre elas e seus locais de ocorrência (Uehara-Prado *et al.*, 2009).

Com o modelo de árvores de decisão, conseguimos classificar todas as variáveis de tal maneira que é possível prever onde dados faltantes estariam alocados, e com um alto grau de confiança. Nem todas as variáveis se mostraram bons classificadores.

Em Redes Bayesianas, foi possível identificar quais variáveis condicionam ou eram condicionadas por outras. Com isso é possível determinar cada uma delas, mesmo que nem todas estejam presentes nos dados coletados.

7 PERSPECTIVAS FUTURAS

Os três diferentes modelos matemáticos, e suas aplicações, descritos neste trabalho foram elaborados com o intuito de extrair o máximo de conhecimento, em termos de interações entre as variáveis, que fosse relevante para as áreas de ecologia e biologia, em contraste com a muitos dos trabalhos descritos na literatura (em áreas semelhantes ou iguais), que se utilizam de ferramentas semelhantes (Friedman *et al.*, 2000), porém direcionados na tentativa de encontrar um único modelo capaz de elucidar os dados.

Muitas vezes, conjuntos de dados coletados ao longo de meses, ou anos, ficam estagnados após a conclusão de um único experimento e se perdem em meio a outros projetos. O potencial das ferramentas abordadas para aplicações futuras nas áreas ecológicas e biológicas é gigantesco, e com este trabalho esperamos encorajar outros pesquisadores a revisitarem bases de dados antigas, completas ou incompletas, com uma ou mais abordagens computacionais modernas, no intuito de extrair o máximo de conhecimento possível.

REFERÊNCIAS BIBLIOGRÁFICAS

- ACKERY, P. R.; DE JONG, R.; VANE-WRIGHT, R. I. "The butterflies: Hedyloidea, Hesperioidea and Papilionoidea "In: KRISTENSEN, N. P. (Ed.). "Lepidoptera, Moths and Butterflies Volume I: Evolution, Systematic, and Biogeography" In: FISCHER, M. (Ed.). "Handbook of Zoology" Volume IV: Arthropoda: Insecta. Berlin: Gruyter, 1999. p. 263-300.
- ADERHOLD, ANDREJ, DIRK HUSMEIER, AND V. ANNE SMITH. "Reconstructing ecological networks with hierarchical Bayesian regression and Mondrian processes", *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, 2013.
- AGUILERA, P. A., FERNÁNDEZ, A., RECHE, F., & RUMÍ, R. "Hybrid Bayesian network classifiers: Application to species distribution models", *Environmental Modelling & Software*, 25, 12, p. 1630-1639, 2010.
- ALPAYDIN, E. "Introduction to machine learning", *Cover, Copyright Page, Table of Contents for*: p. 1-327, 2004.
- ANDREWS, S. "In-Close, a fast algorithm for computing formal Concepts", *Proceedings of the International Conference on Conceptual Structures (ICCS)*, 2009.
- ANTAL, P., FANNES, G., TIMMERMAN, D., DE MOOR, B., & MOREAU, Y. "Bayesian applications of belief networks and multilayer perceptrons for ovarian tumor 24 classification with rejection", *Artificial Intelligence in Medicine*, 29, p. 39–60, 2003.
- APTÉ, CHIDANAND, AND SHOLOM WEISS. "Data mining with decision trees and decision rules", *Future generation computer systems*, 13, 2, p. 197-210, 1997.
- ASPINALL, W.P., CARNIEL, R., JAQUET, O., WOO, G., HINCKS, T. "Using hidden multistate Markov models with multi-parameter volcanic data to provide empirical evidence for alert level decision-support", *Journal of Volcanology and Geothermal Research*, 153, p. 112 e 124, 2006.
- BADII, R. AND POLITI, A. "Thermodynamics and complexity of cellular automata", *Phys. Rev. Lett.*, 78, 3, p. 444-447, 1997.

- BADII, R. AND POLITI, A. "Complexity, Hierarchical Structures and Scaling in Physics, volume 6 of Cambridge Nonlinear Science Series", *Cambridge University Press*, 1997.
- BAISEN, Z., VALENTINE, I., KEMP, P., & LAMBERT, G. "Predictive modelling of hill-pasture productivity: integration of a decision tree and a geographical information system", *Agricultural Systems*, 87, p. 1-17, 2006.
- BARBER D. "Bayesian Reasoning and Machine Learning", *Cambridge University Press*, 2012.
- BARBOSA, J. M., DE SENNA CARNEIRO T. G., AND TAVARES, A. I. "Métodos de Classificação por Árvores de Decisão Disciplina de Projeto e Análise de Algoritmos."
- BARLOW, J., ARAUJO, I. S., OVERAL, W. L., GARDNER, T. A. MENDES, F. S., LAKE, I. R., PERES, C. A. "Diversity and composition of fruit-feeding butterflies in tropical Eucalyptus plantations", *Biodiversity and Conservation*, 17, p. 1089–1104, 2008.
- BARLOW, J., GARDNER, T. A., ARAUJO, I. S., AVILA - PIRES, T. C., BONALDO, A. B., COSTA, J. E., ESPOSITO, M. C., FERREIRA, L. V., HAWES, J, HERNANDEZ, M. M., HOOGMOED, M. S. LEITE, R. N., LO - MAN - HUNG, N. F., MALCOLM, J. R., MARTINS, M. B., MESTRE, L. A. M., MIRANDA - SANTOS, R., NUNES - GUTJAHR, A. L., OVERAL, W. L., PARRY, L., PETERS, S. L., RIBEIRO - JUNIOR, M. A., DA SILVA, M. N. F., MOTTA, C. S., PERES, C. A. "Quantifying the biodiversity value of tropical primary, secondary, and plantation forests", *Proceedings of the National Academy of Sciences of the United States of America*, 104, p. 18555-18560, 2007a.
- BARLOW, J., OVERAL, W. L., ARAUJO, I. S., GARDNER, T. A., PERES, C. A. "The value of primary, secondary and plantation forests for fruit-feeding butterflies in the Brazilian Amazon", *Journal of Applied Ecology*, 44, p. 1001-1012, 2007b.
- BASGALUPP, M.P. "LEGAL-Tree: Um algoritmo genético multi-objetivo lexicográfico para indução de árvores de decisão", Tese de Doutorado, ICMC-USP, São Carlos, 2010.
- BEISNER, BEATRIX E., DANIEL T. HAYDON, AND KIM CUDDINGTON. "Alternative stable states in ecology", *Frontiers in Ecology and the Environment*, 1, 7, p. 376-382, 2003.
- BELL, JOHN F. "Tree-based methods", *Machine learning methods for ecological applications*, Springer US, p. 89-105, 1999.

- BERWAERTS K. & VAN DYCK H. "Take-off performance under optimal and suboptimal thermal conditions in the butterfly *Parargeaegeria*", *Oecologia*, 141, p. 536–545, 2004.
- BERWAERTS K., AERTS P. & VAN DYCK H. "On the sex specific mechanisms of butterfly flight: flight performance relative to flight morphology, wing kinematics, and sex in *Parargeaegeria*", *Biol. J. Linn. Soc.*, 89, p. 675–687, 2006.
- BETTS C.R. & WOOTTON R.J. "Wing shape and flight behaviour in butterflies (Lepidoptera: papilionoidea and hesperioidea): a preliminary analysis" *J. Exp. Biol.* 138, p. 271–288, 1988.
- BIERREGAARD, R. O. JR, LAURANCE, W. F., GASCON, C., BENITEZ-MALVIDO, J., FEARNSIDE, P. M., FONSECA, C. R., GANADE, G., MALCOLM, J. R., MARTINS, M. B., MORI, S., OLIVEIRA, M., RANKIN-DE MÉRONA, J., SCARIOT, A., SPIRONELLO, W. & WILLIAMSON, B. "Principles of forest fragmentation and conservation in the Amazon. Lessons from Amazonia: the ecology and conservation of a fragmented forest" (ed. by R. O. Bierregaard, C. Gascon, T. E. Lovejoy and R. Mesquita), *Yale University Press*, New Haven, p. 371–385, 2001.
- BOLSTAD, WILLIAM M. "Logic, probability, and uncertainty." *Introduction to Bayesian Statistics, Second Edition*, p. 55-76, 2007.
- BONDY, J. A., MURTY, U. S. R. "Graph theory with applications", *Springer*, 2008.
- BOUYER, J. "Identification of ecological indicators for monitoring ecosystem health in the trans-boundary W Regional park: a pilot study", *Biological Conservation*, 38, p.73-88, 2007.
- BREIMAN, L, FRIEDMAN, J., OLSHEN. R. & STONE, C. "Classification and Regression Trees", *Wadsworth*, 1984.
- BREIMAN, L. "Heuristics of instability in model selection", Technical Report, Statistics Department, University of California at Berkeley (to appear, *Annals of Statistics*), 1994.
- BREIMAN, L., FRIEDMAN, J., STONE, C. J., & OLSHEN, R. A. "Classification and regression trees", *CRC press*, 1984.

- BREUKER, C. J., BRAKEFIELD P. M., AND GIBBS, M. "The association between wing morphology and dispersal is sex specific in the glanville fritillary butterfly *Melitaea cinxia* (Lepidoptera: Nymphalidae)", *European Journal of Entomology*, 104, 3, p. 445-452, 2007.
- BROWN JR., K. S., FREITAS, A. V. L. "Atlantic Forest butterflies: indicators for landscape conservation", *Biotropica*, 32, p. 934–956, 2007.
- BROWN JR, K. S., FREITAS, A. V. L. "Butterfly communities of urban forest fragments in Campinas, São Paulo, Brazil: structure, instability, environmental correlates, and conservation", *Journal of Insect Conservation*, 6, p.217–231, 2003.
- BROWN JR., K. S. "Maximizing daily butterfly counts", *Journal of the Lepidopterists' Society*, 26, p.183–196, 1972.
- BROWN JR., K. S. "Conservation of Neotropical environments: insects as indicators", The conservation of insects and their habitats. Entomological Society Symposium XV, *Academic Press*, London, 349-404, 1991.
- BROWN JR., K. S. & Freitas, A. V. L. "Atlantic Forest butterflies: indicators for landscape conservation", *Biotropica*, 32, p. 934-956, 2000.
- Brown, K. S. Jr. & Freitas, A. V. L. "Butterfly communities of urban forest fragments in Campinas, São Paulo, Brazil: structure, instability, environmental correlates, and conservation", *Journal of Insect Conservation*, 6, p.217-231, 2003.
- CAIN, J. "Planning improvements in natural resources management: guidelines for using Bayesian networks to support the planning and management of development programmes in the water sector and beyond", Center for Ecology and Hydrology, Crowmarsh Gifford, Wallingford, UK, 2001.
- CAIN, J., BATCHELOR, C., AND WAUGHTRY, D. "Belief networks: a framework for the participatory development of natural resource management strategies" *Environ. Dev. Sustain*, 1, p. 123–133, 1999.
- CASTELLETTI, A., SONCINI-SESSA, R. "Bayesian networks in environmental management and planning" *Environ. Model. Softw.*, submitted for publication-a.
- CASTELLETTI, A., & SONCINI-SESSA, R. "Coupling real-time control and socio-economic issues in participatory river basin planning", *Environmental Modelling & Software*, 22, 8, p.1114-1128, 2007

- DE CASTRO, P. A. D. “Sinergia entre Sistemas Imunológicos Artificiais e Modelos Gráficos Probabilísticos”, Tese de doutorado, Doutorado em Engenharia Elétrica, Unicamp, Campinas, 2009.
- CHAPELLE, O., SCHÖLKOPF, B., & ZIEN, A. “Semi-supervised learning”, *MIT Press*, 2006.
- CHARNIAK, E. “Bayesian Networks without Tears”, *I. A. Magazine*, 1991.
- CHEN, SERENA H., AND CARMEL A. POLLINO. "Good practice in Bayesian network modelling", *Environmental Modelling & Software*, 37, p. 134-145, 2012.
- CHENG, J., AND GREINER R. "Comparing Bayesian network classifiers", *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc., 1999.
- CHENG, Y., CHURCH, G. M. “Biclustering of expression data”, *Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology*, p. 93-103, 2000.
- COELHO & LEITE. “Árvores de Decisão” Pós Graduação –Turma APS-30 – FATEC-SP, 2015.
- DA COSTA C. S., PORCARO R. M., AND LIFSCHITZ S. "Mineração de dados - Funcionalidades, técnicas e abordagens", IN: *THE BRAZILIAN GOVERNMENT USING CREDIT SCORING*, Leonardo Sales, PUC-Rio Inf, Rio de Janeiro, 2002.
- DA SILVA, C. C., OLIVEIRA, S. R. M., ADAMI, S. F., CRIVELANTI, R. C., COELHO, R. M., DE ITAPURA, A. B., ... & GERALDO, B. "Diferentes classificadores na predição de classes de solos em mapeamento digital", *Embrapa Informática Agropecuária-Artigo em anais de congresso (ALICE)*. In: *SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO*, 15, Curitiba. Anais... São José dos Campos: Inpe, 2011.
- DA SILVA, GABRIELA CORSO. “Distribuição temporal e espacial de borboletas Nymphalidae em áreas de Mata Atlântica do Parque Estadual da Serra do Tabuleiro, SC: um aporte à conservação e ao manejo de áreas protegidas”, Diss. Universidade Federal de Santa Catarina, Centro de Ciências Biológicas, Programa de Pós-Graduação em Ecologia, Florianópolis, 2011.

- DARWICHE, A. "Modeling and Reasoning with Bayesian Networks" Cambridge University Press, 2009.
- DARWICHE, A. & HUANG, C. "Inference in Belief Networks: A procedural guide" International Journal of Approximate Reasoning, 1994.
- DAUD, N., AND CORNE D. W. "Human readable rule induction in medical data mining", *Proceedings of the European Computing Conference*, Springer US, 2009.
- DAVIS, J. M., AND JUDY A. S. "The effect of natal experience on habitat preferences", *Trends In Ecology & Evolution*, 19, 8, p. 411-416, 2004.
- DE'ATH, G., AND FABRICIUS K. E. "Classification and regression trees: a powerful yet simple technique for ecological data analysis", *Ecology*, 81, 11, p. 3178-3192, 2000.
- DEBINSKI, D. M. & HOLT, R. D. "Habitat fragmentation experiments: a global survey and overview", *Conservation Biology*, 14, p. 342-355, 2000.
- DEVRIES, P. J. "Stratification of fruit-feeding nymphalidae butterflies in a Costa Rican rainforest", *Journal of Research on the Lepidoptera*, 26, p. 98-108, 1988.
- DEVRIES, P. J. "The Butterflies of Costa Rica and their Natural History. Papilionidae, Pieridae and Nymphalidae I", *Princeton University Press*, New Jersey, p. 327, 1987.
- DEVRIES, P. J. "The butterflies of Costa Rica and their natural history. Papilionidae, Pieridae, Nymphalidae", Princeton, *Princeton University Press*, p. 327, 1987.
- DIMAGGIO, P. A., MCALLISTER, S. R., FLOUDAS, C. A., FENG, X. J., RABINOWITZ, J. D., & RABITZ, H. A. "Biclustering via optimal re-ordering of data matrices in systems biology: rigorous methods and comparative studies", *BMC Bioinformatics*, 9, p.458, 2008.
- DUNHAM M. H. "Data mining: introduction and advanced topics", Pearson Education, Upper Saddle River, NJHenneman, M. Laurie, and Jane Memmott Infiltration of a Hawaiian community by introduced biological control agents, *Science*, 293.5533 (2001), p. 1314-1316, 2002.
- DUNHAM, MARGARET H. "Data mining: Introductory and advanced topics", *Pearson Education India*, 2006.
- ENDERS, CRAIG K. "Applied missing data analysis", *Guilford Publications*, 2010.

- EVERITT, B. "Cluster Analysis", *John Wiley*, 3rd. edition, 1993.
- FAYYAD, USAMA M., AND KEKI B. IRANI. "On the handling of continuous-valued attributes in decision tree generation" *Machine learning*, 8, 1, p. 87-102, 1992.
- FEINSINGER, P. "Designing Field Studies for Biodiversity Conservation", The Fischer, Walter, Friedhelm Krupp, and Wolfgang Schneider, *Plantas e invertebrados*, 2013.
- FONSECA, J. "Indução de árvores de decisão Hist Class–Proposta de um algoritmo não paramétrico", Diss. Faculdade de Ciências e Tecnologia da UNL, 1994.
- FRANKLIN, J. "Predicting the distribution of shrub species in southern California from climate and terrain-derived variables", *Journal of Vegetation Science*, p. 733-748, 1998.
- FRANKLIN, J. "Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients", *Progress in physical geography*, 19, 4, p. 474-499, 1995.
- FREITAS, A. V. L.; BROWN, K. S. JR. "Phylogeny of the Nymphalidae (Lepidoptera)", *Systematic Biology*, 53, p.363-383, 2004.
- FREITAS, A. V. L.; FRANCINI, R. B.; BROWN, K. S. JR., "Métodos de estudos em biologia da conservação e manejo da vida silvestre", Fundação O Boticário de Proteção à Natureza, Editora da UFPR, Curitiba, 5, p. 125-151, 2003.
- FREITAS, A. V. L.; LEAL I. R.; UEHARA-PRADO & IANNUZZI. "Insetos como indicadores de conservação da paisagem", *Biologia da Conservação*, Editora da UERJ, Rio de Janeiro, p.201-225, 2006.
- FREITAS, A. A., WIESER D. C. AND APWEILER, R. "On the importance of comprehensible classification models for protein function prediction", *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 7, 1, p. 172-182, 2010.
- FRIEDMAN, N. "Learning belief networks in the presence of missing values and hidden variables", In Proc. *14th International Conference on Machine Learning*, Morgan Kaufmann, p. 125–133, 1997.

- FRIEDMAN, N., LINIAL, M., NACHMAN, I., & PE'ER, D. "Using Bayesian networks to analyze expression data" *Proc. Fourth Annual International Conference on Computational Molecular Biology*, 1, 127–135, 2000.
- FRIEDMAN, N. AND KOLLER D. "Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks" *Machine learning*, 50, 1-2, p. 95-125, 2003.
- FRIEDMAN, N., GEIGER D., ANDGOLDSZMIDT M. "Bayesian network classifiers", *Machine learning*, 29, 2-3, p. 131-163, 1997.
- FUCILINI, L. L., AND ISERHARD C. A. "Padrões de Diversidade de Satyrinae (Lepidoptera, Nymphalidae) no Rio Grande do Sul, Brasil" *Salão de Iniciação Científica (21.: 2009 out. 19-23: Porto Alegre, RS)*, Livro de resumos, UFRGS, Porto Alegre, 2009.
- FURLANETTI, P. R. R. "A comunidade de borboletas frugívoras de áreas em processo de restauração, fragmentos de floresta estacional semidecidual e pastagens", Universidade Estadual Paulista, 2010.
- GIASSON, E., HARTEMINK, A. E., TORNQUIST, C. G., TESKE, R., & BAGATINI, T. "Evaluation of five algorithms of decision trees and three digital elevation models for digital soil mapping at semidetall level at the Lageado Grande Watershed, RS, Brazil", *Ciência Rural*, 43, 11,p. 1967-1973, 2003.
- GOVAERT, G., NADIF, M. "Co-Clustering - Models, Algorithms and Applications", *John Wileyand Sons*, 2014.
- GRAHAM, J. W. "Missing data analysis: Making it work in the real world", *Annual review of psychology*, 60, p. 549-576, 2009.
- GUISAN, A., AND ZIMMERMANN N. E. "Predictive habitat distribution models in ecology" *Ecological modelling*, 135, 2, p. 147-186, 2000.
- HALL, M., FRANK, E., HOLMES, G., PFAHRINGER, B., REUTEMANN, P., & WITTEN, I. H. "The WEKA data mining software: an update", *ACM SIGKDD explorations newsletter*11, 1, p. 10-18, 2009.
- HARTIGAN, J. A. "Direct clustering of a data matrix", *Journal of the American Statistical Association*, 67, 337, p. 123-129, 1972.

HORVITZ, E., RUAN, Y., GOMES, C. P., KAUTZ, H., SELMAN, B., & CHICKERING, D.M. "A Bayesian approach to tackling hard computational problems", In *Proc. of 17th Conference on Uncertainty and Artificial Intelligence*, p. 235–244, 2001.

<http://www.mathworks.com/help/stats/classification-trees-and-regression-trees.html>.

INTRODUÇÃO AO PROGRAMA "MATLAB" COM APLICAÇÕES Material elaborado por:
Prof. Dr. Antônio César Baleeiro Alves (UCG/UFG) Prof. Dr. Enes Gonçalves Marra (UFG) Prof. Dr. José Wilson Lima Nerys (UFG) Goiânia – GO, 2007.

IVERSON, L. R., AND PRASAD, A. M. "Predicting abundance of 80 tree species following climate change in the eastern United States", *Ecological Monographs*, 68, 4, p. 465-485, 1998.

JENKINS, D. W. "Neotropical Nymphalidae I. Revision of *Hamadryas*", *Bulletin of the Allyn Museum*, Sarasota, 81, p. 1-146, 1983.

JENKINS, D. W. "Neotropical Nymphalidae II. Revision of *Myscelia*", *Bulletin of the Allyn Museum*, Sarasota, 87, p. 1-64, 1984.

JENKINS, D. W. "Neotropical Nymphalidae VIII. Revision of *Eunica*", *Bulletin of the Allyn Museum*, Sarasota, 131, p. 1-177, 1990.

JENKINS, D. W. "Neotropical Nymphalidae. III. Revision of *Catonephele*", *Bulletin of the Allyn Museum*, Sarasota, 92, p. 1-65, 1985.

JENKINS, D. W. "Neotropical Nymphalidae. V. Revision of *Epiphile*", *Bulletin of the Allyn Museum*, Sarasota, 101, p. 1-70, 1986.

JENSEN, F. V., NIELSEN, T.D., "Bayesian Networks and Decision Graphs", *Springer*, 2007.

JENSEN, F.V., NIELSEN, T.D., "Bayesian Networks and Decision Graphs", *Springer*, 2007.

JENSEN, V. F. "Bayesian Networks and Decision Graphs", *Springer-Verlag*, 2001.

KINGSFORD, C., ANDSALZBERG S. L. "What are decision trees?", *Nature biotechnology* 26, 9, p. 1011-1013, 2008.

KIRIZAWA, M. S., M., LOPES, E. A., CUSTODIO FILHO, A. "Flora Fanerogâmica da Reserva Biológica do Alto da Serra de Paranapiacaba", 2003, Disponível em: <http://www.ibot.sp.gov.br/PESQUISA/paranapiacaba.htm>

- KJAERULFF, U. B., ANDMADSEN A. L. "Probabilistic networks-an introduction to bayesian networks and influence diagrams", *Aalborg University*, 2005.
- KOH, L. P. "Impacts of land use change on South-east Asian forest butterflies: a review", *Journal of Applied Ecology*, 44, 703–713, 2007.
- KOLLER, D. & FRIEDMAN, N. "Probabilistic Graphical Models: Principles and Techniques", *The MIT Press*, 2009.
- KORB, K. B., NICHOLSON, A. E., "Bayesian Artificial Intelligence", *Chapman and Hall/CRC Press*, London, p. 364, 2004.
- KREMEN, C., COLWELL, R., ERWIN, T. L., MURPHY, D. D., NOSS, R. F. & SANJAYAN, M. A. "Terrestrial arthropod assemblages: their use in conservation planning", *Conservation Biology*, 7, p. 796-808, 1993.
- KUZNETSOV, S. O. "Learning of Simple Conceptual Graphs from Positive and Negative Examples" *Proceedings of the Third European Conference on Principles of Data Mining and Knowledge Discovery*, Lecture Notes in Computer Science, Springer, 1704, p. 384-391, 1999.
- LAMAS, G., HEPPNER J. B. "Atlas of Neotropical Lepidoptera", Association for Tropical Lepidoptera, *Scientific Publishers*, Gainesville, EUA, p. 439, 2004.
- LANDRES, P. B., VERNER, J., THOMAS, J. W. "Ecological uses of vertebrate indicator species: a critique", *Conservation Biology*, 2, p. 316-328, 1988.
- LAPLACE, P. S. "Memoir on the probability of the causes of events", *Statistical Science*, p. 364-378, 1986.
- LAURANCE, W. F., LOVEJOY, T. E., VASCONCELOS, H. L., BRUNA, E. M., DIDHAM, R. K., STOUFFER, P. C., GASCON, C., BIERREGAARD, R. O., LAURANCE, S. G. & SAMPAIO, E. "Ecosystem decay of Amazonian forest fragments: a 22-year investigation", *Conservation Biology*, 16, p. 605-618, 2002.
- LAWTON, J. H. "Plant architecture and the diversity of phytophagous insects", *Annual Review of Entomology*, 28, p. 23–39, 1983.

- Lee, H. D., Monard, M. C., Voltolini, R. F., & Wu, F. C. “Avaliação experimental e comparação de algoritmos de seleção de atributos importantes com o algoritmo FDimBF baseado na dimensão fractal”, *Technical Report*, 264, ICMC-USP, 2005.
- LEWINSOHN, T. M., FREITAS, A. V. L., PRADO, P. I. "Conservation of terrestrial invertebrates and their habitats in Brazil", *Conservation Biology*, 19, p. 640-645, 2005.
- LOH, WEI-YIN. "Classification and regression trees" *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1, 1, p. 14-23, 2011.
- LOMOV, B. “Are butterflies and moths useful indicators for restoration monitoring?: a pilot study in Sydney’s Cumberland Plain Woodland”, *Ecological Management & Restoration*, Austrália, 7, 3, p.204-210, 2006.
- MADEIRA, S. C., OLIVEIRA, A. L. “Biclustering algorithms for biological data analysis: a survey”, *IEEE Transactions on Computational Biology and Bioinformatics*, 1, 1, p. 24-45, 2004.
- MCCOY, E. D. & BELL, S. S. “Habitat structure: the evolution and diversification of a complex topic”, *Habitat structure: the physical arrangement of objects in space* (ed. by S. S. Bell, E.D. McCoy and H.R. Mushinsky), Chapman & Hall, London, p. 3-27, 1991.
- MAGO, V. K., PRASAD, B., BHATIA, A., ANDMAGO, A., “A decision making system for the treatment of dental caries”, in *Soft computing applications in business*, Prasad B., Springer, Berlin, (Ed.), 230, p. 231-242, 2008.
- HALL M., FRANK E., HOLMES G., PFAHRINGER B., REUTEMANN P., WITTEN I. H. “The WEKA Data Mining Software: An Update”, *SIGKDD Explorations*, 11, 1, 2009.
- MCCANN, R. K., MARCOT B. G., AND ELLIS, R. "Bayesian belief networks: applications in ecology and natural resource management", *Canadian Journal of Forest Research*, 36, 12, p. 3053-3062, 2006.
- MCGEACH, M. A. "The selection, testing and application of terrestrial insects as bioindicators" *Biol. Rev. Camb. Philos. Soc.*, 73, p. 181-201, 1998.
- MESTRE, L. A. M., SANTOS M., R., NUNES, GUTJAHR, A. L., OVERAL, W. L., PARRY, L., PETERS, S. L., RIBEIRO - JUNIOR, M. A., DA SILVA, M. N. F., MOTTA, C. S., PERES, C. A. "Quantifying the biodiversity value of tropical

- primary, secondary, and plantation forests" *Proceedings of the National Academy of Sciences of the United States of America*, 104, p. 18555-18560, 2007a.
- MILNS, I., BEALE C. M., ANDANNE SMITH, V. "Revealing ecological networks using Bayesian network inference algorithms", *Ecology*, 91, 7, p. 1892-1899, 2010.
- MITCHELL, T. "Machine Learning", McGraw-Hill, 1997.
- MITCHELL, T. M. "Machine learning and data mining", *Communications of the ACM*, 42, 11, p. 30-36, 1999.
- MITCHELL, T. M. "Machine learning", *Burr Ridge, IL: McGraw Hill*, 45, 1997.
- MONARD, M., AND BARANAUSKAS, J. "Conceitos sobre aprendizado de máquinas em sistemas inteligentes: Fundamentos e aplicações, Cap. 4.", 2003.
- MONARD, M. C., ANDBARANAUSKAS J. A. "Indução de Regras e Árvores de Decisão", 2003.
- MOPPER, S., AND STRAUSS S. Y., eds. "Genetic structure and local adaptation in natural insect populations: effects of ecology, life history, and behavior", *Springer Science & Business Media*, 2013.
- MURPHY, K. "The bayes net toolbox for matlab" *Computing science and statistics*, 33, 2, p. 1024-1034, 2001. Disponível em: <http://www.cs.berkeley.edu/~murphyk/Bayes/bayes.html>.
- MURPHY, K. "A brief introduction to graphical models and Bayesian networks", 1998.
- NEAPOLITAN, R. E. "Learning bayesian networks" *Upper Saddle River: Prentice Hall*, 38, 2004.
- NEOPOLITAN, R. E. "Learning Bayesian networks" *Prentice Hall*, Englewood Cliffs, N.J., 2003.
- NEW, T. R. "Are Lepidoptera an effective umbrella group' for biodiversity conservation?" *Journal of Insect Conservation*, 1, p. 5-12, 1997.
- NICHOLSON, A., AND K. KORB. "Bayesian artificial intelligence", 2003.

- NIEDERMAYER, D. “An Introduction to Bayesian Networks and their Contemporary Applications”, 1998.
- NIEMI, G. J., MCDONALD, M. E. “Application of ecological indicators”, *Annual Review of Ecology Evolution and Systematics*, 35, p. 89–111, 2004.
- OLDEN, J. D., AND JACKSON, D. A. "A comparison of statistical approaches for modelling fish species distributions", *Freshwater biology*, 47, 10, p. 1976-1995, 2002.
- OLDEN, J. D., LAWLER, J. J., AND N. LEROYPOFF. "Machine learning methods without tears: a primer for ecologists", *The Quarterly review of biology*, 83, 2, p. 171-193, 2008.
- OLIVEROS, A. Q., CARNIEL, R., TÁRRAGA, M., ASPINALL, W., “On the Application of Hidden Markov Model and Bayesian Belief Network to Seismic Noise at Las Canadas Caldera”, *Chaos Solutions & Fractals*, Tenerife, Spain, 37, 849 e 857, 2008.
- OTERO, F. E. B., FREITAS A. A., AND JOHNSON, C. G. "Inducing decision trees with an ant colony optimization algorithm", *Applied Soft Computing*, 12, 11, p. 3615-3626, 2012.
- PANG-NING, T., STEINBACH M., AND KUMAR V. "Introduction to data mining", *Library of Congress*, 2006.
- PAPAJ, D. R., AND LEWIS, A. C. eds. “Insect Learning: Ecology and Evolutionary Perspectives”, Springer Science & Business Media, 2012.
- PAVLOPOULOS, G. A., SECRIER, M., MOSCHOPOULOS, C. N., SOLDATOS, T. G., KOSSIDA, S., AERTS, J., SCHNEIDER, R., BAGOS, P. G. “Using graph theory to analyze biological networks”, *Bio Data Mining*, 4, p. 1-27, 2011.
- PEARL, J. “Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference”, *Morgan Kaufmann*, 1988.
- PEARL, J. “Probabilistic Reasoning in Intelligent Systems”, *Morgan Kaufmann*, San Mateo, CA, 1988.
- PEARL, J. “Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference”, *Morgan Kaufmann*, San Mateo, CA, 1998.
- PEARL, J. “Probabilistic reasoning in intelligent systems: networks of plausible inference”, *Morgan Kaufmann*, 2014.

- PENZ, C. M., AIELLO, A., SRYGLEY, R. B. "Early stages of *Caligo illioneus* and *C. idomeneus* (Nymphalidae, Brassolinae) from Panama, with remarks on larval food plants for the Subfamily", *Journal of the Lepidopterists' Society*, 53, p. 142-152, 1999.
- PINHEIRO, C. E. G.; ORTIZ, J. V. C. "Communities of fruit-feeding butterflies along a vegetation gradient in central Brazil", *Journal Biogeography*, 19, p. 505-511, 1992.
- POELMANS, J., IGNATOV, D. I., KUZNETSOV, S. O., DEEDENE, G. "Formal concept analysis in knowledge processing: A survey on applications", *Expert Systems with Applications*, 40, p. 6538-6560, 2013a.
- POELMANS, J., KUZNETSOV, S. O., IGNATOV, D. I., DEEDENE, G. "Formal concept analysis in knowledge processing: A survey on models and techniques", *Expert Systems with Applications*, 40, p. 6601-6623, 2013b.
- PÖYRY, J., LUOTO, M., HEIKKINEN, R. K., KUUSSAARI, M., & SAARINEN, K. "Species traits explain recent range shifts of Finnish butterflies." *Global Change Biology*, 15, 3, p. 732-743, 2009.
- POZZER, C. T. "Aprendizado por árvores de decisão", Universidade Federal de Santa Maria, Rio Grande do Sul, 2006.
- PRASAD, A. M., IVERSON L. R., AND LIAW A. "Newer classification and regression tree techniques: bagging and random forests for ecological prediction", *Ecosystems*, 9, 2, p. 181-199, 2006.
- PRICE, P. W., DENNO, R. F., EUBANKS, M. D., FINKE, D. L., & KAPLAN, I. "Insect ecology: behavior, populations and communities", *Cambridge University Press*, 2011.
- PROULX, S. R., PROMISLOW D. E., AND PHILLIPS, P. C. "Network thinking in ecology and evolution", *Trends in Ecology & Evolution*, 20, 6, p. 345-353, 2005.
- QUINLAN, J. "Decision trees and multivalued attributes", *Machine Intelligence*, 11, p. 305-318, 1988.
- QUINLAN, J. R. "Induction of decision trees", *Machine Learning*, 1, 1, p. 81-106, 1986.
- QUINLAN, J. R. "C4.5: programs for machine learning", *Morgan Kaufmann Publishers Inc.*, San Francisco, CA, USA, 1993.

- QUINLAN, J. R. "Induction of decision trees" *Machine learning*, 1, 1, p. 81-106, 1986.
- QUINLAN, J. R. "C4. 5: programs for machine learning", *Elsevier*, 2014.
- RAGHUNATHAN, T. E. "What do we do with missing data? Some options for analysis of incomplete data", *Annu. Rev. Public Health*, 25, p. 99-117, 2004.
- RAVAGNANI, F. G., SAIDEMBERG, D. M., FARIA, A. L. C., SARTOR, S. B., OLIVEIRA, D.N., CATHARINO, R.R. "Quality Control of Biomarkers: From the Samples to Data Interpretation", in *Akyar, I. (ed.) Latest Research in to Quality Control*, 22, 2012.
- RIBEIRO, D. B., PRADO, P. I., BROWN JR, K. S., & FREITAS, A. V. "Temporal Diversity Patterns and Phenology in Fruit-feeding Butterflies in the Atlantic Forest" *Biotropica*, 42, 6, p. 710-716, 2010.
- RICHARD, E. "Varieties of Learning", *In: AI Game programming wisdom*, Charles River Media, 2002.
- ROKACH, L., & MAIMON, O. "Top-down induction of decision trees classifiers-a survey", *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 35, 4, p. 476-487, 2005.
- ROSEN, K.H. "Discrete mathematics and its applications", *William C Brown Pub*; 4th edition, 1998.
- RUGGIERI, S. "Efficient C4. 5 [classification algorithm]" *Knowledge and Data Engineering, IEEE Transactions on*, 14, 2, p. 438-444, 2002.
- RUSSEL, S., & NORVIG, P. "Artificial Intelligence: A Modern Approach", *Prentice-Hall*, 3rd. Edition, 2009.
- RUSSELL, S. AND NOVIG, P. "Artificial Intelligence: a Modern Approach", 2nd Edition, *Prentice-Hall*, 2003.
- RYAN HOULETTE, D. F. "Construction a Decision Tree Based on Past Experience", *In: AI Game programming wisdom 2*, Charles River Media, 2004.
- SANTOS, J. P. D., ISERHARD, C. A., TEIXEIRA, M. O., & ROMANOWSKI, H. P. "Fruit-feeding butterflies guide of subtropical Atlantic Forest and Araucaria Moist Forest in State of Rio Grande do Sul, Brazil", *Biota Neotropica*, 11, 3, p. 253-274, 2011.

- SCHLUETER, P. M., AND HARRIS S. A. "Analysis of multi locus fingerprinting data sets containing missing data", *Molecular Ecology Notes*, 6, 2, p. 569-572, 2006.
- SHAHABUDDIN, G., PONTE, C.A. "Frugivorous butterfly species in tropical forest fragments: correlates of vulnerability to extinction", *Biodiversity and Conservation*, 14, p. 1137-1152, 2005.
- SHAHABUDDIN, G., TERBORGH, J. W. "Frugivorous butterflies in Venezuelan forest fragments: abundance, diversity and the effects of isolation", *Journal of Tropical Ecology*, 15, p. 703-722, 1999.
- SHAHABUDDIN, G., AND PONTE C. A. "Frugivorous butterfly species in tropical forest fragments: correlates of vulnerability to extinction" *Biodiversity & Conservation*, 14, 5, p. 1137-1152, 2005.
- SIGRH Sistema de Informação para o Gerenciamento de Recursos Hídricos do Estado de São Paulo, Banco de Dados Pluviométricos do Estado de São Paulo. Acessado em 07/04/2008. www.sigrh.sp.gov.br
- SILVA, A. D. A., GONÇALVES, C. R., GALVÃO, D. M., GONÇALVES, A. J. L., GOMES, J., SILVA, M. N., & SIMONI, L. D. "Quarto catálogo dos insetos que vivem nas plantas do Brasil, seus parasitos e predadores", Rio de Janeiro: Ministério da Agricultura, 1, 2, p. 622, 1998.
- SINGER, M. C. "Butterfly-Hostplant Relationships: Host Quality, Adult Choice and Larval Success", In: The Biology of Butterflies (R. I. Vane-Wright & P. E. Ackery, eds.), *Florida: Academic Press Inc.*, p. 81-88, 1984.
- CUSTODIO, A., TIMONI J. L., AND JUNQUEIRA AGUIAR, L. S. "Rain-forest of the Serra do Mar: degradation and reconstitution", *Rain-forest of the Serra do Mar: degradation and reconstitution*, SMA, 1990.
- SOUZA-SILVA, M. "Ecologia e conservação das comunidades de invertebrados cavernícolas na Mata Atlântica Brasileira", Tese de Doutorado em Ecologia Conservação e Manejo da Vida Silvestre, ICB-UFMG, p. 226, 2008.
- SPARROW, H. R., SISK, T.D., EHRLICH, P. R. & MURPHY, D. D. "Techniques and guidelines for monitoring Neotropical butterflies", *Conservation Biology*, 8, p. 800-809, 1994.

- SRYGLEY, ROBERT B. "Early stages of *Caligo Illioneus* and *C. Idomeneus* (Nymphalidae, Brassolinae) from Panama, with remarks on larval food plants for the subfamily", 1999.
- SUN, R., AND ZHANG X. "Top-down versus bottom-up learning in cognitive skill acquisition", *Cognitive Systems Research*, 5, 1, p. 63-89, 2004.
- SUN, W., C. JIN, AND LI J. "Decision tree and PCA-based fault diagnosis of rotating machinery", *Mechanical Systems and Signal Processing*, 21, 3, p. 1300-1317, 2007.
- TABARELLI, M., MANTOVANI, W. "A regeneração de uma floresta tropical montana após corte e queima", *Revista Brasileira de Biologia*, São Paulo, Brasil, 59, p. 239-250, 1999.
- TAN, P.-N., STEINBACH, M., & KUMAR, V. "Introduction to Data Mining", *Addison-Wesley Longman Publishing Co.*, (First Edition), Inc., Boston, MA, USA, 2005.
- TRINDADE, M., SAMPAIO, R. "Introdução ao Matlab", PUC-Rio, 2002.
- TSAMARDINOS, I., ALIFERIS, C. F., STATNIKOV, A. R., & STATNIKOV, E. "Algorithms for Large Scale Markov Blanket Discovery", *FLAIRS Conference*, 2, 2003.
- TSANG, S., KAO, B., YIP, K. Y., HO, W. S., & LEE, S. D. "Decision trees for uncertain data", *IEEE Transactions on Knowledge and Data Engineering*, 23, 1, p. 64-78, 2011.
- TSCHARNTKE, T., STEFFAN-DEWENTER, I., KRUESS, A., & THIES, C. "Characteristics of insect populations on habitat fragments: a mini review", *Ecological research*, 17, 2, p. 229-239, 2002.
- TURNER, I.M. "Species loss in fragments of tropical rainforest: a review of the evidence", *Journal of Applied Ecology*, 33, p. 200-209, 1996.
- Tutorial: Machine Learning. Disponível em:
<http://www.gamasutra.com/gdc2005/features/20050307/postcard-sanchez-crespo.htm>.
- UEHARA-PRADO, M. "Artrópodes como indicadores biológicos de perturbação antrópica", 2009.
- UEHARA-PRADO, M. "Diversidade e composição de borboletas frugívoras em cerradão e plantio de eucalipto em três lagoas Mato Grosso do Sul Brasil", 2009.
- UEHARA-PRADO, M., BROWN JR., K.S. & FREITAS, A.V.L. "Species richness, Composition and abundance of fruit-feeding butterflies in the Brazilian Atlantic Forest:

- comparison between a fragmented and a continuous landscape", *Global Ecology and Biogeography*, 16, p. 43-54, 2007.
- UEHARA-PRADO, M., BROWN JR., K.S. & FREITAS, A.V.L. "Biological traits of frugivorous butterflies in a fragmented andba continuous landscape in the South Brazilian Atlantic Forest", *Journal of the Lepidopterists' Society*, 59, p. 96-106, 2005.
- UEHARA-PRADO, M., FERNANDES, J. O., BELLO, A. M., MACHADO, G., SANTOS, A. J., VAZ - DE - MELLO, F. Z. & FREITAS, A. V. L. "Selecting terrestrial arthropods as indicators of small - scale disturbance: A first approach in the Brazilian Atlantic Forest", *Biology Conservation*, 142, p. 1220-1228, 2009.
- UEHARA-PRADO, M., K. S. BROWN, AND A. V. L. FREITAS. "Biological traits of frugivorous butterflies in a fragmented and a continuous landscape in the South Brazilian Atlantic Forest." *JOURNAL-LEPIDOPTERISTS SOCIETY*, 59, 2, p. 96, 2005.
- UEHARA-PRADO, M.; FREITAS, A. L.; FRANCINI, R. B. "Guias das borboletas frugívoras da reserva estadual do Morro Grande e região de Caucaia do alto, Cotia (SP)", *Biota Neotropica*, 4, 1, p. 1-9, 2004. Disponível em:
<<http://www.biotaneotropica.org.br/v4n1/pt/abstract?inventory+BN00504012004>>.
Acessado em: 5 fev. 2006.
- UEHARA-PRADO, M., FREITAS, A. V. L., FRANCINI, R. B., & BROWN JR, K. S. "Guia das borboletas frugívoras da reserva estadual do Morro Grande e região de Caucaia do Alto, Cotia (São Paulo)", *Biota Neotropica*, 4, 1, p. 1-5, 2004.
- UUSITALO, L., "Advantages and challenges of Bayesian networks in environmental modelling", *Ecological Modelling*, 203, p. 312e318, 2007.
- VAN DYCK H. & WIKLUND C. "Seasonal butterfly design: morphological plasticity among three developmental pathways relative to sex, flight and thermo regulation", *J. E. Biol.*, vol. 15, p. 216–225, 2002.
- VAYSSIÈRES, M. P., PLANT R. E., ANDALLEN-DIAZ B. H. "Classification trees: An alternative non-parametric approach for predicting species distributions", *Journal of vegetation science*, 11, 5, p. 679-694, 2000.

- VERONEZE, R. “Tratamento de Dados Faltantes Empregando Biclusteração com Imputação Múltipla”, Dissertação de Mestrado, Faculdade de Engenharia Elétrica e de Computação (FEEC), Unicamp, 2011.
- WAHLBERG N., KLEMETTI, T. & HANSKI, I. “Metapopulation structure and movements in five species of checkerspot butterflies”, *Oecologia*, 130, p. 33–43, 2002.
- WEIBULL, ANN-CHRISTIN, ÖSTMAN Ö., AND GRANQVIST, Å. "Species richness in agro ecosystems: the effect of landscape, habitat and farm management", *Biodiversity & Conservation*, 12, 7, p. 1335-1355, 2003.
- WICKMAN P. O. “Sexual selection and butterfly design – a comparative design”, *Evolution* 46, p. 1525–1536, 1992.
- WITTEN, I. H. AND E. FRANK. “Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations”, California: Morgan Kaufmann, 2005.
- XU, R., WUNSCH, D.C. II. “Clustering algorithms in biomedical research: a review”, *IEEE Review on Biomedical Engineering*, 3, p. 120-154, 2010.
- ZHANG, W. “Graphs, Networks and Agent-based Modeling”, *Computational Ecology*, World Scientific, New Jersey, 2012.

Apêndice A – Planilha 1, dados de presença e ausência coletados por Uehara-Prado *et al.*, (2009).

UC	Perturbação vide Magrini <i>et al.</i> , 2011)	Espécie	Subfamília	Total
E.B. Boracéia	managed trails	Archeuptychia cluena	Satyrinae	1
E.B. Boracéia	managed trails	Archeuptychia cluena	Satyrinae	1
E.B. Boracéia	managed trails	Archeuptychia cluena	Satyrinae	1
E.B. Boracéia	managed trails	Blepolenis batea	Brassolinae	1
E.B. Boracéia	managed trails	Blepolenis batea	Brassolinae	1
E.B. Boracéia	managed trails	Blepolenis batea	Brassolinae	1
E.B. Boracéia	managed trails	Blepolenis batea	Brassolinae	1
E.B. Boracéia	managed trails	Blepolenis batea	Brassolinae	1
E.B. Boracéia	managed trails	Blepolenis batea	Brassolinae	1
E.B. Boracéia	managed trails	Blepolenis batea	Brassolinae	1
E.B. Boracéia	managed trails	Caligo arisbe	Brassolinae	1
E.B. Boracéia	managed trails	Caligo arisbe	Brassolinae	1
E.B. Boracéia	managed trails	Caligo beltrao	Brassolinae	1
E.B. Boracéia	managed trails	Caligo beltrao	Brassolinae	1
E.B. Boracéia	managed trails	Caligo beltrao	Brassolinae	1
E.B. Boracéia	managed trails	Caligo beltrao	Brassolinae	1
E.B. Boracéia	managed trails	Caligo beltrao	Brassolinae	1
E.B. Boracéia	managed trails	Caligo beltrao	Brassolinae	1
E.B. Boracéia	managed trails	Caligo beltrao	Brassolinae	1
E.B. Boracéia	managed trails	Caligo beltrao	Brassolinae	1
E.B. Boracéia	managed trails	Caligo beltrao	Brassolinae	1
E.B. Boracéia	managed trails	Caligo beltrao	Brassolinae	1
E.B. Boracéia	managed trails	Caligo beltrao	Brassolinae	1
E.B. Boracéia	managed trails	Caligo beltrao	Brassolinae	1
E.B. Boracéia	managed trails	Carminda griseldis	Satyrinae	1
E.B. Boracéia	managed trails	Catonephele acontius	Biblidinae	1
E.B. Boracéia	managed trails	Catonephele acontius	Biblidinae	1
E.B. Boracéia	managed trails	Catonephele numilia	Biblidinae	1
E.B. Boracéia	managed trails	Catonephele numilia	Biblidinae	1
E.B. Boracéia	managed trails	Catonephele numilia	Biblidinae	1
E.B. Boracéia	managed trails	Dasyophthalma creusa	Brassolinae	1
E.B. Boracéia	managed trails	Dasyophthalma creusa	Brassolinae	1
E.B. Boracéia	managed trails	Dasyophthalma creusa	Brassolinae	1
E.B. Boracéia	managed trails	Dasyophthalma creusa	Brassolinae	1
E.B. Boracéia	managed trails	Dasyophthalma creusa	Brassolinae	1
E.B. Boracéia	managed trails	Dasyophthalma creusa	Brassolinae	1
E.B. Boracéia	managed trails	Dasyophthalma creusa	Brassolinae	1
E.B. Boracéia	managed trails	Dasyophthalma creusa	Brassolinae	1
E.B. Boracéia	managed trails	Dasyophthalma creusa	Brassolinae	1
E.B. Boracéia	managed trails	Dasyophthalma creusa	Brassolinae	1
E.B. Boracéia	managed trails	Dasyophthalma creusa	Brassolinae	1
E.B. Boracéia	managed trails	Dasyophthalma creusa	Brassolinae	1
E.B. Boracéia	managed trails	Dasyophthalma rusina	Brassolinae	1
E.B. Boracéia	managed trails	Dasyophthalma rusina	Brassolinae	1
E.B. Boracéia	managed trails	Dasyophthalma rusina	Brassolinae	1
E.B. Boracéia	managed trails	Dasyophthalma rusina	Brassolinae	1

[illegible]

[illegible]

[illegible]

[illegible]

E.B. Boracéia	managed trails	Narope cyllene	Brassolinae	1
E.B. Boracéia	managed trails	Narope cyllene	Brassolinae	1
E.B. Boracéia	managed trails	Narope cyllene	Brassolinae	1
E.B. Boracéia	managed trails	Opoptera syme	Brassolinae	1
E.B. Boracéia	managed trails	Opsiphanes invirae	Brassolinae	1
E.B. Boracéia	managed trails	Pierella nereis	Satyrinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	managed trails	Prepona demophon	Charaxinae	1
E.B. Boracéia	managed trails	Prepona demophon	Charaxinae	1
E.B. Boracéia	managed trails	Prepona demophon	Charaxinae	1
E.B. Boracéia	managed trails	Prepona demophon	Charaxinae	1
E.B. Boracéia	managed trails	Prepona demophon	Charaxinae	1
E.B. Boracéia	managed trails	Prepona demophon	Charaxinae	1
E.B. Boracéia	managed trails	Prepona demophoon	Charaxinae	1
E.B. Boracéia	managed trails	Prepona demophoon	Charaxinae	1
E.B. Boracéia	managed trails	Splendeptychia hygina	Satyrinae	1
E.B. Boracéia	managed trails	Splendeptychia hygina	Satyrinae	1
E.B. Boracéia	managed trails	Taygetis acuta	Satyrinae	1
E.B. Boracéia	managed trails	Taygetis acuta	Satyrinae	1

[illegible]

[illegible]

E.B. Boracéia	unmanaged trails	Dasyophthalma rusina	Brassolinae	1
E.B. Boracéia	unmanaged trails	Eryphanis reevesi	Brassolinae	1
E.B. Boracéia	unmanaged trails	Eryphanis reevesi	Brassolinae	1
E.B. Boracéia	unmanaged trails	Eryphanis reevesi	Brassolinae	1
E.B. Boracéia	unmanaged trails	Eryphanis reevesi	Brassolinae	1
E.B. Boracéia	unmanaged trails	Eryphanis reevesi	Brassolinae	1
E.B. Boracéia	unmanaged trails	Eryphanis reevesi	Brassolinae	1
E.B. Boracéia	unmanaged trails	Eryphanis reevesi	Brassolinae	1
E.B. Boracéia	unmanaged trails	Eryphanis reevesi	Brassolinae	1
E.B. Boracéia	unmanaged trails	Eryphanis reevesi	Brassolinae	1
E.B. Boracéia	unmanaged trails	Eryphanis reevesi	Brassolinae	1
E.B. Boracéia	unmanaged trails	Euptychia ernestina	Satyrinae	1
E.B. Boracéia	unmanaged trails	Forsterinaria necys	Satyrinae	1
E.B. Boracéia	unmanaged trails	Forsterinaria necys	Satyrinae	1
E.B. Boracéia	unmanaged trails	Forsterinaria necys	Satyrinae	1
E.B. Boracéia	unmanaged trails	Forsterinaria necys	Satyrinae	1
E.B. Boracéia	unmanaged trails	Morpho catenarius	Morphinae	1
E.B. Boracéia	unmanaged trails	Morpho catenarius	Morphinae	1
E.B. Boracéia	unmanaged trails	Morpho catenarius	Morphinae	1
E.B. Boracéia	unmanaged trails	Morpho catenarius	Morphinae	1
E.B. Boracéia	unmanaged trails	Morpho catenarius	Morphinae	1
E.B. Boracéia	unmanaged trails	Morpho catenarius	Morphinae	1
E.B. Boracéia	unmanaged trails	Morpho catenarius	Morphinae	1
E.B. Boracéia	unmanaged trails	Myscelia orsis	Biblidinae	1
E.B. Boracéia	unmanaged trails	Myscelia orsis	Biblidinae	1
E.B. Boracéia	unmanaged trails	Myscelia orsis	Biblidinae	1
E.B. Boracéia	unmanaged trails	Myscelia orsis	Biblidinae	1
E.B. Boracéia	unmanaged trails	Myscelia orsis	Biblidinae	1
E.B. Boracéia	unmanaged trails	Myscelia orsis	Biblidinae	1
E.B. Boracéia	unmanaged trails	Myscelia orsis	Biblidinae	1
E.B. Boracéia	unmanaged trails	Myscelia orsis	Biblidinae	1
E.B. Boracéia	unmanaged trails	Myscelia orsis	Biblidinae	1
E.B. Boracéia	unmanaged trails	Myscelia orsis	Biblidinae	1
E.B. Boracéia	unmanaged trails	Narope cyllene	Brassolinae	1
E.B. Boracéia	unmanaged trails	Narope cyllene	Brassolinae	1
E.B. Boracéia	unmanaged trails	Narope cyllene	Brassolinae	1
E.B. Boracéia	unmanaged trails	Opoptera aorsa	Brassolinae	1
E.B. Boracéia	unmanaged trails	Paryphthimoides interjecta	Satyrinae	1
E.B. Boracéia	unmanaged trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	unmanaged trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	unmanaged trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	unmanaged trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	unmanaged trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	unmanaged trails	Prepona amphimachus	Charaxinae	1
E.B. Boracéia	unmanaged trails	Prepona demophon	Charaxinae	1
E.B. Boracéia	unmanaged trails	Prepona demophon	Charaxinae	1
E.B. Boracéia	unmanaged trails	Taygetis acuta	Satyrinae	1
E.B. Boracéia	unmanaged trails	Taygetis acuta	Satyrinae	1
E.B. Boracéia	unmanaged trails	Taygetis acuta	Satyrinae	1
E.B. Boracéia	unmanaged trails	Taygetis mermeria	Satyrinae	1

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

R.B. Paranapiacaba	edge	Memphis morvus	Charaxinae	1
R.B. Paranapiacaba	edge	Memphis morvus	Charaxinae	1
R.B. Paranapiacaba	edge	Memphis morvus	Charaxinae	1
R.B. Paranapiacaba	edge	Memphis morvus	Charaxinae	1
R.B. Paranapiacaba	edge	Memphis otrere	Charaxinae	1
R.B. Paranapiacaba	edge	Memphis otrere	Charaxinae	1
R.B. Paranapiacaba	edge	Memphis otrere	Charaxinae	1
R.B. Paranapiacaba	edge	Memphis otrere	Charaxinae	1
R.B. Paranapiacaba	edge	Memphis ryphea	Charaxinae	1
R.B. Paranapiacaba	edge	Memphis ryphea	Charaxinae	1
R.B. Paranapiacaba	edge	Moneuptychia soter	Satyrinae	1
R.B. Paranapiacaba	edge	Morpho catenarius	Morphinae	1
R.B. Paranapiacaba	edge	Myscelia orsis	Biblidinae	1
R.B. Paranapiacaba	edge	Myscelia orsis	Biblidinae	1
R.B. Paranapiacaba	edge	Myscelia orsis	Biblidinae	1
R.B. Paranapiacaba	edge	Myscelia orsis	Biblidinae	1
R.B. Paranapiacaba	edge	Myscelia orsis	Biblidinae	1
R.B. Paranapiacaba	edge	Myscelia orsis	Biblidinae	1
R.B. Paranapiacaba	edge	Narope cyllene	Brassolinae	1
R.B. Paranapiacaba	edge	Narope cyllene	Brassolinae	1
R.B. Paranapiacaba	edge	Narope cyllene	Brassolinae	1
R.B. Paranapiacaba	edge	Narope cyllene	Brassolinae	1
R.B. Paranapiacaba	edge	Narope cyllene	Brassolinae	1
R.B. Paranapiacaba	edge	Opoptera syme	Brassolinae	1
R.B. Paranapiacaba	edge	Opoptera syme	Brassolinae	1
R.B. Paranapiacaba	edge	Opsiphanes invirae	Brassolinae	1
R.B. Paranapiacaba	edge	Opsiphanes invirae	Brassolinae	1
R.B. Paranapiacaba	edge	Paryphthimoides phronius	Satyrinae	1
R.B. Paranapiacaba	edge	Paryphthimoides phronius	Satyrinae	1
R.B. Paranapiacaba	edge	Paryphthimoides phronius	Satyrinae	1
R.B. Paranapiacaba	edge	Paryphthimoides phronius	Satyrinae	1
R.B. Paranapiacaba	edge	Paryphthimoides phronius	Satyrinae	1
R.B. Paranapiacaba	edge	Paryphthimoides phronius	Satyrinae	1
R.B. Paranapiacaba	edge	Pedaliodes phanias	Satyrinae	1
R.B. Paranapiacaba	edge	Praepedaliodes phanias	Satyrinae	1
R.B. Paranapiacaba	edge	Prepona amphimachus	Charaxinae	1
R.B. Paranapiacaba	edge	Prepona amphimachus	Charaxinae	1
R.B. Paranapiacaba	edge	Prepona amphimachus	Charaxinae	1
R.B. Paranapiacaba	edge	Prepona amphimachus	Charaxinae	1
R.B. Paranapiacaba	edge	Prepona amphimachus	Charaxinae	1
R.B. Paranapiacaba	edge	Prepona demophon	Charaxinae	1
R.B. Paranapiacaba	edge	Prepona demophon	Charaxinae	1
R.B. Paranapiacaba	edge	Prepona demophon	Charaxinae	1
R.B. Paranapiacaba	edge	Pseudodebis euptychidia	Satyrinae	1
R.B. Paranapiacaba	edge	Smyrna blomfildia	Nymphalinae	1
R.B. Paranapiacaba	edge	Smyrna blomfildia	Nymphalinae	1
R.B. Paranapiacaba	edge	Taygetis acuta	Satyrinae	1
R.B. Paranapiacaba	edge	Temenis laothoe	Biblidinae	1
R.B. Paranapiacaba	edge	Yphthimoides castrensis	Satyrinae	1
R.B. Paranapiacaba	edge	Yphthimoides castrensis	Satyrinae	1
R.B. Paranapiacaba	edge	Yphthimoides castrensis	Satyrinae	1

[illegible]

R.B. Paranapiacaba	interior	Opoptera syme	Brassolinae	1
R.B. Paranapiacaba	interior	Opoptera syme	Brassolinae	1
R.B. Paranapiacaba	interior	Opoptera syme	Brassolinae	1
R.B. Paranapiacaba	interior	Opoptera syme	Brassolinae	1
R.B. Paranapiacaba	interior	Opoptera syme	Brassolinae	1
R.B. Paranapiacaba	interior	Opoptera syme	Brassolinae	1
R.B. Paranapiacaba	interior	Opoptera syme	Brassolinae	1
R.B. Paranapiacaba	interior	Opoptera syme	Brassolinae	1
R.B. Paranapiacaba	interior	Opoptera syme	Brassolinae	1
R.B. Paranapiacaba	interior	Opsiphanes invirae	Brassolinae	1
R.B. Paranapiacaba	interior	Opsiphanes invirae	Brassolinae	1
R.B. Paranapiacaba	interior	Opsiphanes quiteria	Brassolinae	1
R.B. Paranapiacaba	interior	Opsiphanes quiteria	Brassolinae	1
R.B. Paranapiacaba	interior	Paryphthimoides phronius	Satyrinae	1
R.B. Paranapiacaba	interior	Paryphthimoides phronius	Satyrinae	1
R.B. Paranapiacaba	interior	Prepona amphimachus	Charaxinae	1
R.B. Paranapiacaba	interior	Prepona amphimachus	Charaxinae	1
R.B. Paranapiacaba	interior	Prepona amphimachus	Charaxinae	1
R.B. Paranapiacaba	interior	Prepona amphimachus	Charaxinae	1
R.B. Paranapiacaba	interior	Prepona amphimachus	Charaxinae	1
R.B. Paranapiacaba	interior	Prepona demophon	Charaxinae	1
R.B. Paranapiacaba	interior	Prepona demophon	Charaxinae	1
R.B. Paranapiacaba	interior	Splendeptychia doxes	Satyrinae	1
R.B. Paranapiacaba	interior	Splendeptychia hygina	Satyrinae	1
R.B. Paranapiacaba	interior	Taydebis peculliaris	Satyrinae	1
R.B. Paranapiacaba	interior	Taygetis acuta	Satyrinae	1
R.B. Paranapiacaba	interior	Yphthimoides grymon	Satyrinae	1
R.B. Paranapiacaba Total				706
Total Geral				1886

Apêndice B – Planilha 2, dados completos, sem padronização, coletados por Uehara-Prado *et al.*, (2009).

UC	Perturbação (vide Magrini <i>et al.</i> , 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
E.B. Boracéia	managed trails	Archeuptychia cluena	Satyrinae	C	0	23	1
					1	25	1
				D	1	26	1
		Blepopenis batea	Brassolinae	A	0	46	1
					?	48	1
					1	45	2
				B	0	?	1
					1	46	1
					?	47	1
		Caligo arisbe	Brassolinae	B	1	62	1
				C	0	68	1
		Caligo beltrao	Brassolinae	B	0	83	1
					?	84	1
					?	85	1
					1	71	1
					?	75	1
				C	1	75	2
					?	?	1
				D	0	75	1
		Carmina griseldis	Satyrinae	B	0	(vazio)	1
		Catonephele acontius	Biblidinae	A	0	(vazio)	1
				C	0	37	1
		Catonephele numilia	Biblidinae	A	0	?	1
					1	34	1
					?	35	1
		Dasyophthalma creusa	Brassolinae	-	1	47	1
				A	1	48	1
				B	0	48	1
					1	48	1
					?	49	1
				C	0	54	2
					?	56	1
					1	46	1
					?	48	1
					?	?	1
				D	1	47	1
		Dasyophthalma rusina	Brassolinae	A	1	45	1
UC	Perturbação (vide Magrini <i>et al.</i> , 2011)	Espécie	Subfamília	Idade estimada	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total

				(A=mais nova)			
				B	1	45	1
					?	46	1
					?	47	1
					?	?	1
				C	0	49	1
					?	50	1
					1	42	1
					?	44	1
					?	45	3
					?	48	1
					?	49	1
					?	?	3
				D	0	47	1
					1	42	1
					?	44	1
		Epiphile oreia	Biblidinae	D	0	30	1
		Eryphanis reevesi	Brassolinae	A	1	51	1
					?	52	1
				B	0	56	1
					?	59	2
					?	?	1
					1	51	2
					?	52	1
					?	53	1
				C	0	?	1
					1	?	1
				D	0	57	1
					?	61	1
					1	52	1
					?	53	1
					?	?	1
				E	1	50	1
		Eteona tisiphone	Satyrinae	A	0	28	1
		Forsterinaria necys	Satyrinae	-	1	25	1
				A	1	24	1
				B	0	24	1
					?	25	2
					?	26	2
					?	27	2
					?	?	1
					1	23	3
					?	24	5
					?	?	3
				C	0	24	1
					?	25	2
UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total

				(A=mais nova)			
					?	26	2
					?	27	2
					?	?	1
					1	23	2
					?	24	2
					?	25	1
					?	26	3
					?	?	1
				D	0	25	1
					?	27	1
					?	?	1
					1	24	2
		Forsterinaria quantius	Satyrinae	?	?	?	1
				A	1	25	2
					?	26	1
					?	27	1
					?	?	2

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
					?	26	1
				B	0	27	2
					?	28	1
					?	29	1
					1	24	1
					?	25	3
					?	26	5
					?	27	3
					?	28	2
					?	29	1
					?	23	4
				C	0	27	2
						28	3
						-	1
					1	23	1
						24	2
						25	3
						26	3
						27	1
				D	1	25	2
						26	1
				E	0	26	1
						27	1
					1	26	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
		Godartiana muscosa	Satyrinae	B	0	28	1
		Hamadryas epinome	Biblidinae	-	-	-	1
				B	0	36	1
						37	1
					1	35	1
						37	1
						38	1
						39	1
						40	1
				C	1	38	1
						39	1
				D	1	34	1
						-	1
		Hamadryas feronia	Biblidinae	A	1	-	1
		Hamadryas fornax	Biblidinae	A	1	37	1
						38	1
				B	1	38	1
						39	1
		Hermeuptychia hermes	Satyrinae	B	1	19	1
						-	1
				D	0	17	1
						18	1
		Memphis morvus	Charaxinae	B	1	28	1
		Memphis otrere	Charaxinae	B	0	34	1
						(vazio)	1
				C	0	32	1
						-	1
		Morpho aega	Morphinae	E	1	50	1
		Morpho catenarius	Morphinae	A	1	69	1
				B	0	74	2
						(vazio)	1
					1	66	2
						67	1
						68	1
						70	1
						-	2
				C	0	73	1
						75	1
					1	66	1
						71	1
						72	1
						-	1
				D	1	71	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 femea, 1 macho)	Tamanho asa anterior (mm)	Total
						-	2
		Myscelia orsis	Biblidinae	A	0	29	1
					1	-	1
				B	0	28	2
						29	4
						30	1
						-	1
					1	26	1
				C	0	27	1
						28	3
						29	2
						30	1
				D	0	25	1
						28	1
		Narope cyllene	Brassolinae	A	0	(vazio)	1
				B	0	32	2
						34	1
						(vazio)	2
					1	30	1
						31	1
				C	0	(vazio)	1
					1	(vazio)	1
		Opoptera syme	Brassolinae	B	1	43	1
		Opsiphanes invirae	Brassolinae	B	1	-	1
		Pierella nereis	Satyrinae	B	1	36	1
		Prepona amphimachus	Charaxinae	-	-	-	2
				A	0	51	1
					1	44	1
						46	1
						47	5
						50	1
				B	0	51	1
						52	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 femea, 1 macho)	Tamanho asa anterior (mm)	Total
					1	46	3
						47	1
						51	1
				C	0	52	2
						54	1
					1	45	1
						46	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
						50	1
						52	1
				D	1	44	1
						47	1
						-	1
		Prepona demophon	Charaxinae	B	0	57	1
					1	48	1
				C	0	55	1
				D	0	53	1
					1	53	1
						54	1
						-	1
		Prepona demophoon	Charaxinae	A	0	60	1
				B	1	44	1
		Splendeptychia hygina	Satyrinae	B	0	21	1
				C	0	21	1
		Taygetis acuta	Satyrinae	A	0	44	1
				B	0	42	1
						44	1
					1	41	1
						42	1
						43	1
				C	0	42	1
						44	1
					1	42	1
				D	0	46	1
						(vazio)	1
					1	41	3
						43	1
		Taygetis mermeria	Satyrinae	B	1	57	1
		Taygetis ypthima	Satyrinae	A	0	40	2
						41	1
						42	2
						43	1
				B	0	37	1
						40	3
						41	1
					1	36	1
						37	1
						38	1
						39	2
				C	1	37	1
						39	1
				D	1	35	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
		<i>Ypthimoides castrensis</i>	Satyrinae	B	1	20	1
	unmanaged trails	<i>Archeuptychia cluena</i>	Satyrinae	A	(vazio)	(vazio)	1
				(vazio)	(vazio)	(vazio)	1
		<i>Caligo arisbe</i>	Brassolinae	C	0	-	1
		<i>Caligo beltrao</i>	Brassolinae	A	0	81	1
						-	3
					1	78	2
				B	0	82	1
						85	2
						86	1
					1	74	2
						75	1
						77	3
						78	1
						84	1
						85	1
						-	1
				C	0	82	1
						84	1
						86	1
						-	1
					1	75	1
						77	4
						80	1
		<i>Caligo illioneus</i>	Brassolinae	C	1	84	1
				D	1	78	1
		<i>Carminda griseldis</i>	Satyrinae	B	0	18	1
		<i>Carminda paeon</i>	Satyrinae	D	0	22	1
		<i>Colobura dirce</i>	Nymphalinae	B	1	36	1
		<i>Dasyophthalma creusa</i>	Brassolinae	A	0	53	1
				B	0	54	1
						55	1
						56	1
						57	1
					1	48	1
				C	0	58	1
					1	47	3
						48	2
						49	2
						50	1
						-	1
				D	1	49	1
		<i>Dasyophthalma rusina</i>	Brassolinae	A	0	-	1
					1	45	1
				B	0	48	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
						51	2
						52	2
					1	44	1
				C	0	44	1
						52	2
					1	43	1
						44	1
						45	1
						-	1
				D	1	48	1
		Eryphanis reevesi	Brassolinae	A	1	54	2
						55	1
				B	0	57	1
						58	1
					1	53	1
						-	2
				D	0	57	1
					1	52	1
		Euptychia ernestina	Satyrinae	B	1	21	1
		Forsterinaria necys	Satyrinae	B	0	28	1
					1	25	1
				C	0	26	1
					1	24	1
		Morpho catenarius	Morphinae	-	-	-	1
				B	1	68	1
						-	2
				C	1	65	1
				D	0	-	1
		Myscelia orsis	Biblidinae	-	0	-	1
				B	0	28	1
						30	1
						-	1
					1	27	1
						28	1
				C	0	27	1
						29	3
						30	1
				D	0	28	1
		Narope cyllene	Brassolinae	B	1	32	1
				C	(vazio)	(vazio)	2
		Opoptera aorsa	Brassolinae	C	0	42	1
		Paryphthimoides interjecta	Satyrinae	C	0	22	1
		Prepona amphimachus	Charaxinae	A	1	45	1
				B	0	52	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
					1	52	1
				C	0	51	1
						54	1
					1	45	1
						52	1
		Prepona demophon	Charaxinae	C	0	53	2
		Taygetis acuta	Satyrinae	A	1	42	1
				B	1	44	1
				C	0	48	1
		Taygetis mermeria	Satyrinae	B	1	60	1
		Taygetis rectifascia	Satyrinae	A	1	33	1
		Taygetis ypthima	Satyrinae	A	0	41	1
						-	1
					1	39	1
						40	1
						42	1
				B	0	38	2
						39	1
						41	1
						-	1
					1	38	1
						39	2
						40	1
						41	1
				C	0	39	1
E.B. Boracéia Total							471
Núcleo Santa Virgínia	selective logging	Caligo arisbe	Brassolinae	A	0	67	2
					1	60	1
						63	1
				B	0	67	1
						70	2
					1	61	2
						63	1
						64	1
				C	0	64	1
						65	1
					1	57	1
						60	1
						61	2
				D	0	69	1
					1	59	1
						64	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
				(vazio)	(vazio)	(vazio)	1
		Caligo beltrao	Brassolinae	C	0	76	1
						88	1
					1	72	1
				D	1	0	1
		Catoblepia amphirrhoe	Brassolinae	A	1	57	1
		Dasyophthalma rusina	Brassolinae	B	0	45	1
						49	2
						52	1
					1	45	1
				C	0	52	1
					1	40	1
						44	2
						45	2
						46	1
				D	0	50	1
					1	44	2

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
						45	2
						46	2
		Eryphanis reevesi	Brassolinae	A	1	52	1
				B	0	54	1
						55	1
						56	2
					1	50	2
						51	1
						52	1
						53	1
						55	1
				C	0	57	2
						58	1
						59	1
					1	51	1
						52	2
				D	0	0	1
						55	2
					1	52	1
		Forsterinaria necys	Satyrinae	A	1	26	1
				B	0	20	1
						21	1
					1	27	1
						29	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
				C	0	26	1
						27	1
					1	26	2
		Forsterinaria quantius	Satyrinae	A	1	27	1
				B	1	26	1
				C	0	0	1
						28	1
					1	26	1
		Hamadryas fornax	Biblidinae	B	1	38	1
		Memphis morvus	Charaxinae	B	0	34	1
		Memphis otrere	Charaxinae	C	0	31	1
		Morpho catenarius	Morphinae	A	1	67	1
				B	0	73	1
						74	1
						75	1
					1	65	2
						66	1
						67	1
						68	1
						69	1
						70	4
						72	2
				C	0	67	1
						74	1
					1	0	2
						62	1
						65	1
						66	2
						67	2
						68	1
						69	2
						70	4
						71	2
						72	2
						73	3
						76	1
				D	1	69	1
		Narope cyllarus	Brassolinae	C	1	34	1
		Narope cyllene	Brassolinae	A	1	31	1
						34	1
				B	0	34	2
					1	32	1
						33	1
				C	0	montada	1
					1	32	1
		Opoptera syme	Brassolinae	A	1	43	1
				B	0	48	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
						49	1
					1	44	1
						46	1
						48	1
				C	0	44	1
				D	0	48	1
		Paryphthimoides poltys	Satyrinae	B	0	21	1
		Praepedaliodes amysis?	Satyrinae	B	0	31	1
		Prepona amphimachus	Charaxinae	A	1	47	1
				B	1	46	1
						48	1
		Splendeptychia hygina	Satyrinae	C	1	22	1
		Taygetis acuta	Satyrinae	A	1	40	1
						42	1
						43	1
				B	0	44	1
					1	41	1
						42	2
						43	1
				C	0	46	1
					1	44	1
				D	1	41	1
		Taygetis mermeria	Satyrinae	A	0	62	1
				B	1	57	1
		Taygetis yphthima	Satyrinae	A	0	39	1
				B	0	37	1
						38	1
						39	4
						40	2
						43	1
					1	35	2
						37	2
						39	1
				C	1	35	2
						37	1
						38	1
		Yphthimoides castrensis	Satyrinae	B	0	20	1
	slash and burn	Biblis hyperia	Biblidinae	B	1	29	1
		Blepopenis batea	Brassolinae	A	0	49	1
					1	43	1
						47	1
				B	0	49	1
					1	43	1
						46	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 femea, 1 macho)	Tamanho asa anterior (mm)	Total
				C	1	42	1
		Caligo arisbe	Brassolinae	A	0	67	1
						68	3
					1	57	1
						63	1
				B	0	69	1
					1	60	3
						61	2
						62	3
						63	5
						64	1
						65	1
						68	1
				C	0	65	1
						68	1
					1	58	1
						59	1
						60	1
						61	3
						62	2
						70	1
				D	1	0	4
						60	1
						62	2
				E	1	60	1
		Caligo beltrao	Brassolinae	B	1	74	1
						75	1
				D	1	78	1
		Caligo brasiliensis	Brassolinae	B	0	94	1
		Carminda griseldis	Satyrinae	B	0	17	1
						19	1
						20	1
						22	1
					1	19	1
				C	0	0	1
						18	1
						19	3
						20	2
					1	21	1
		Carminda paeon	Satyrinae	A	1	19	1
		Catoblepia amphirrhoe	Brassolinae	B	1	61	1
				C	1	52	1
				E	1	0	1
		Catonephele acontius	Biblidinae	A	0	38	1
		Catonephele sabrina	Biblidinae	B	0	36	1
		Colobura dirce	Nymphalinae	A	1	31	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
		<i>Dasyophthalma rusina</i>	Brassolinae	B	0	57	1
				C	0	52	1
					1	44	1
						53	1
				D	1	44	1
				E	1	47	1
		<i>Diaethria candrena</i>	Biblidinae	C	1	19	1
		<i>Ectima thecla</i>	Biblidinae	A	1	20	1
						23	1
				B	0	23	1
					1	21	1
						22	2
				C	0	21	1
					1	22	1
						(vazio)	1
				D	?	22	1
		<i>Epiphile oreia</i>	Biblidinae	A	0	26	1
				C	0	24	1
						26	1
						28	1
					1	26	1
				D	0	26	1
						28	1
		<i>Eryphanis reevesi</i>	Brassolinae	A	1	51	1
				C	1	53	1
						54	1
				D	1	0	1
						55	1
						58	1
		<i>Eteona tisiphone</i>	Satyrinae	A	1	24	2
						27	1
				B	0	28	2
						29	1
						30	3
					1	26	1
						29	1
				C	0	27	2
						29	1
						30	2

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
					1	0	1
						25	2
						26	1
						31	1
		Forsterinaria necys	Satyrinae	A	1	24	1
				B	0	27	1
					1	22	2
						23	1
						24	1
						25	1
				C	0	23	1
						26	2
						27	2
						28	1
					1	24	2
						27	1
						ENVELOPE	1
				D	0	23	1
						25	1
						26	1
					1	25	1
		Forsterinaria quantius	Satyrinae	A	1	24	1
						26	1
				B	0	26	1
						27	4
					1	25	1
				C	0	24	1
						27	6
						28	1
						29	1
				D	0	26	1
						27	2
						28	1
					1	24	1
		Hamadryas amphinome	Biblidinae	B	1	39	1
		Hamadryas epinome	Biblidinae	A	1	34	1
						36	1
						37	1
				B	1	0	1
						34	1
						35	3
						36	2
						37	1
						38	2
						39	1
				C	1	34	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
						36	2
						37	2
						38	1
						ENVELOPE	1
				ENVELOPE	1	ENVELOPE	1
		Hamadryas feronia	Biblidinae	A	1	40	1
		Hamadryas fornax	Biblidinae	A	1	35	1
						36	1
						38	1
				B	1	38	1
				C	1	36	1
				D	0	35	1
		Hermeuptychia hermes	Satyrinae	B	0	17	1
						18	2
						22	1
					1	21	2
				C	0	16	1
						18	6
						19	1
						21	1
					1	19	1
						20	1
						21	1
				D	0	18	1
						21	1
					1	20	1
				E	0	17	1
				ENVELOPE	ENVELOPE	ENVELOPE	1
		Manataria hercina	Satyrinae	C	0	47	1
		Memphis appias	Charaxinae	A	0	33	1
		Memphis arginus	Charaxinae	A	1	29	1
						30	1
						31	1
				B	0	32	1
				C	0	34	1
				D	0	33	1
		Memphis morvus	Charaxinae	A	0	32	1
		Memphis otrere	Charaxinae	A	1	30	1
				B	0	33	1
						34	1
						35	1
					1	30	1
						31	1
				C	0	0	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
						31	1
						32	2
						33	1
		Memphis philumena	Charaxinae	C	0	31	1
		Memphis ryphea	Charaxinae	?	0	33	1
				A	1	28	1
						30	1
				B	1	30	1
		Moneuptychia soter	Satyrinae	A	1	19	1
				B	1	17	2
						18	1
						19	2
				C	0	20	1
					1	19	3
						20	1
				D	0	19	1
					1	18	3
						20	1
				(vazio)	montada	(vazio)	1
		Morpho catenarius	Morphinae	?	1	0	1
				A	0	66	1
						77	1
					1	69	2
				B	0	0	1
						66	1
						69	2
						71	3
						72	1
						73	2
						74	3
						75	1
					1	0	1
						61	2
						62	1
						64	2
						65	1
						66	3
						68	1
						69	2
						70	3
						71	2
						72	2
						73	1
						75	2
					?	66	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
						72	1
				C	0	0	2
						63	1
						67	2
						70	3
						71	3
						72	3
						73	2
						74	3
						75	1
						76	1
						77	2
					1	0	7
						64	2
						65	5
						66	7
						67	5
						68	8
						69	2
						70	4
						71	4
						72	7
						74	1
						75	1
					?	68	1
				D	0	67	1
					1	0	8
						65	1
						66	1
						67	1
						70	1
						71	1
						74	1
				E	1	0	1
		Myscelia orsis	Biblidinae	A	0	25	1
						30	1
				B	0	27	2
						28	1
						29	2
					1	24	1
				C	0	28	3
		Opoptera syme	Brassolinae	B	1	41	1
				C	0	45	1
					1	46	1
		Paryphthimoides phronius	Satyrinae	?	?	?	1
				B	1	18	1
						19	2
						20	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
				C	0	18	3
						19	1
					1	19	1
				D	0	19	1
					1	18	1
						19	1
		Paryphthimoides poltys	Satyrinae	-	-	-	1
				B	0	20	1
						ENVELOPE	1
					1	21	2
				C	0	19	1
					1	0	1
						19	1
						(vazio)	1
					montada	21	1
				D	1	20	1
				ENVELOPE	ENVELOPE	ENVELOPE	1
		Praepedaliodes phanias	Satyrinae	D	1	31	1
				(vazio)	montada	31	1
		Prepona amphimachus	Charaxinae	B	0	54	1
					1	46	1
				C	1	47	1
		Prepona chalciope	Charaxinae	B	1	49	1
		Prepona demophon	Charaxinae	C	0	47	1
						54	1
					1	46	1
						62	1
				D	0	52	1
					1	47	1
						48	1
		Taygetis acuta	Satyrinae	B	1	42	1
						44	1
				C	0	44	1
		Taygetis mermeria	Satyrinae	A	1	56	1
						59	2
						60	2
						61	1
				B	0	60	1
						63	1
					1	55	1
						56	1
						57	3
						59	2
						60	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 femea, 1 macho)	Tamanho asa anterior (mm)	Total
						61	3
						62	1
						montada	1
				C	0	61	1
					1	63	1
				E	1	56	1
		Taygetis ypthima	Satyrinae	A	1	38	2
						40	1
				B	0	39	1
						40	2
					1	36	1
						38	1
						39	3
				C	0	36	1
					1	40	1
		Temenis laothoe	Biblidinae	D	1	28	1
		Yphthimoides castrensis	Satyrinae	B	0	0	1
						19	2
						20	2
						21	1
					1	19	1
						20	3
						21	2
				C	0	17	1
						18	1
						20	3
						21	1
					1	17	1
						18	1
						19	2
						20	1
						21	3
				D	0	19	1
				(vazio)	montada	(vazio)	3
		Yphthimoides grymon	Satyrinae	B	1	21	1
		Zaretis itys	Charaxinae	B	0	37	1
					1	32	1
Núcleo Santa Virgínia Total							709
R.B. Paranapiacaba	edge	Blepolenis batea	Brassolinae	B	1	45	1
		Caligo arisbe	Brassolinae	A	1	62	1
						65	1
				C	0	69	1
		Caligo beltrao	Brassolinae	A	1	71	1
				B	0	84	1
					1	75	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
						84	1
				D	1	80	1
		Carminda griseldis	Satyrinae	-	-	-	1
				A	1	20	1
				B	1	21	1
				C	1	21	1
						22	1
		Carminda paeon	Satyrinae	-	-	-	1
				A	0	19	1
						20	5
						22	1
					1	18	2
						19	3
				B	0	18	1
						19	9
						20	12
						21	1
					1	17	1
						18	5
						19	27
						20	3
						-	2
					-	18	2
				C	0	18	1
						19	6
						20	2
					1	17	2

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
						18	2
						19	7
						20	2
					-	20	1
				D	0	19	2
						20	1
					1	18	2
						19	4
				E	0	19	1
						20	1
						-	1
					1	20	1
		Catonephele acontius	Biblidinae	A	0	35	1
					1	30	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
		Catonephele numilia	Biblidinae	C	0	37	1
		Colobura dirce	Nymphalinae	A	1	35	1
				B	1	34	1
						36	1
				C	1	35	1
		Dasyophthalma creusa	Brassolinae	-	-	-	2
				A	0	50	1
						52	4
						53	2
						54	1
						55	3
						56	3
					1	44	1
						45	1
				B	0	52	2
						54	2
						55	4
						56	3
					1	47	1
				C	0	53	2
						54	6
						55	1
						56	2
						57	1
						60	1
					1	46	1
						48	1
				D	1	42	1
						46	1
						47	1
						48	1
		Dasyophthalma rusina	Brassolinae	-	-	-	1
				A	0	45	1
						49	2
						50	1
					1	44	1
						46	1
				B	0	50	4
						51	1
					1	42	2
						43	2
						44	4
						45	1
						47	1
					-		1
				C	0	44	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
						48	1
						49	1
						50	3
						51	1
						52	2
					1	41	2
						42	4
						43	6
						44	8
						45	7
						46	2
						47	2
						49	3
						50	2
						51	1
					-	-	1
				D	0	51	1
						-	1
					1	41	1
						42	4
						43	1
						45	5
						46	1
						47	1
						48	1
						51	1
						-	3
		Epiphile oreia	Biblidinae	A	1	27	1
						28	1
				B	0	29	1
						30	1
					1	26	1
				C	0	26	1
						27	1
		Eryphanis reevesi	Brassolinae	A	0	53	1
					1	49	1
						52	1
				B	0	55	1
						57	5
						62	1
					1	49	1
						51	1
						52	3
						53	1
						54	2
				C	0	55	2
						56	1
						58	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
					1	51	1
						53	1
						55	2
				D	1	52	1
		Eteona tisiphone	Satyrinae	A	0	31	1
				B	0	30	1
					1	31	1
				C	0	25	1
					1	28	1
						29	1
		Forsterinaria necys	Satyrinae	-	-	-	1
				A	0	23	1
						25	1
					1	23	1
						24	2
						25	1
					-	25	1
				B	0	22	1
						24	1
						25	1
						26	1
						27	1
					1	21	1
						23	4
						24	11
						25	3
						26	1
						28	1
					-	23	1
				C	0	23	1
						24	1
						25	5
						26	2
						27	4
						-	1
					1	19	1
						21	1
						22	1
						23	1
						24	4
						25	2
				D	1	23	1
						24	2
		Forsterinaria quantius	Satyrinae	E	0	27	1
		Godartiana muscosa	Satyrinae	B	0	26	1
				C	1	24	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
		Guaianaza pronophila	Satyrinae	B	0	23	1
					1	21	1
				C	0	23	1
		Hamadryas amphinome	Biblidinae	A	1	39	1
				B	1	36	1
						37	1
						41	1
						42	1
		Hamadryas februa	Biblidinae	B	1	35	1
		Hamadryas feronia	Biblidinae	B	1	36	1
						37	1
		Hamadryas fornax	Biblidinae	B	1	36	1
		Hermeuptychia hermes	Satyrinae	-	-	-	1
				A	0	18	2
					1	18	1
				B	0	17	1
						18	5
						19	1
						20	1
				C	0	18	3
					1	18	1
						19	1
					-	17	1
				D	0	17	1
						18	1
						19	1
						48	1
					1	20	1
		Memphis appias	Charaxinae	D	0	31	1
		Memphis morvus	Charaxinae	B	0	34	1
						35	1
						36	1
					1	28	1
						29	2
				C	0	33	2
						34	2
		Memphis otrere	Charaxinae	B	0	28	1
						34	2
					1	31	1
		Memphis ryphea	Charaxinae	A	1	31	1
				B	0	30	1
		Moneuptychia soter	Satyrinae	C	1	21	1
		Morpho catenarius	Morphinae	C	1	71	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
		Myscelia orsis	Biblidinae	A	1	24	1
				B	1	27	1
				C	1	25	1
						26	2
						27	1
		Narope cyllene	Brassolinae	B	1	33	1
				C	0	34	1
						35	1
					1	31	1
				(vazio)	(vazio)	(vazio)	1
		Opoptera syme	Brassolinae	A	1	39	1
				C	1	40	1
		Opsiphanes invirae	Brassolinae	B	1	40	1
						53	1
		Paryphthimoides phronius	Satyrinae	B	1	19	2
						20	1
						-	1
				C	0	20	1
					1	18	1
		Pedaliodes phanias	Satyrinae	A	1	30	1
		Praepedaliodes phanias	Satyrinae	A	1	29	1
		Prepona amphimachus	Charaxinae	A	1	53	1
				B	1	51	1
						53	2
				C	1	52	1
						54	1
		Prepona demophon	Charaxinae	C	0	47	1
						54	1
					1	54	1
		Pseudodebis euptychidia	Satyrinae	C	0	31	1
		Smyrna blomfieldia	Nymphalinae	A	0	45	1
					1	40	1
		Taygetis acuta	Satyrinae	C	1	-	1
		Temenis laothoe	Biblidinae	A	1	27	1
		Yphthimoides castrensis	Satyrinae	B	0	20	1
					1	18	2
						20	1
		Yphthimoides grymon	Satyrinae	A	1	17	1
				B	0	22	1
					1	21	2
				C	0	21	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
		Zaretis itys	Charaxinae	A	0	35	1
	interior	Caligo beltrao	Brassolinae	B	1	75	1
						76	1
		Carminda paeon	Satyrinae	B	1	18	1
				C	0	20	1
		Colobura dirce	Nymphalinae	B	1	34	2
						35	2
				C	1	31	1
						35	1
		Dasyophthalma creusa	Brassolinae	A	0	49	1
						52	1
						53	2
						54	1
						55	4
						56	2
					1	48	1
				B	0	52	1
						54	2
						55	3
					1	41	1
						46	1
						47	3
						48	2
						49	1
				C	0	52	1
						53	2
						54	2
						55	6
						56	1
						57	1
						58	2
					1	43	1
						46	1
						47	3
						48	1
						49	1
						51	1
				D	0	54	1
						55	1
					1	45	1
						46	1
				E	0	52	1
		Dasyophthalma rusina	Brassolinae	A	0	47	1
						49	1
						50	1
						51	3
						53	2

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
				B	0	47	1
						48	1
						49	1
						51	2
					1	43	2
						44	2
						46	1
						47	2
						52	1
				C	0	48	1
						49	4
						50	3
						51	2
						52	1
						53	1
					-		1
					1	42	2
						43	7
						44	9
						45	2
						46	1
						47	2
						48	1
						49	2
						50	1
					-		1
				D	0	46	1
						47	1
					1	41	1
						42	1
						43	1
						44	1
						45	1
						46	1
						49	1
					-		1
		Ectima thecla	Biblidinae	B	1	22	1
		Eryphanis reevesi	Brassolinae	A	0	55	1
						57	1
					1	54	1
						55	1
				B	0	57	1
					1	51	1
						52	2
				C	0	57	1
					1	52	1
						53	1
				D	0	57	1

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
					1	50	1
		Eteona tisiphone	Satyrinae	B	0	35	1
					1	26	1
				C	1	30	1
		Forsterinaria necys	Satyrinae	-	-	-	1
				A	1	24	3
						25	1
				B	0	23	1
						27	1
					1	23	1
				C	0	25	1
						26	1
					1	23	1
						24	1
				D	1	24	1
		Forsterinaria quantius	Satyrinae	A	0	28	1
					1	25	1
		Godartiana muscosa	Satyrinae	B	0	26	1
		Guaianaza pronophila	Satyrinae	C	0	22	1
		Hamadryas amphinome	Biblidinae	B	1	40	1
				C	1	43	1
		Hamadryas epinome	Biblidinae	A	1	36	1
		Hermeuptychia hermes	Satyrinae	B	1	17	1
		Memphis morvus	Charaxinae	A	0	35	1
				B	1	26	1
						28	1
						29	1
				C	0	31	1
		Memphis otrere	Charaxinae	A	0	35	1
					1	29	1
				C	0	33	1
		Memphis philumena	Charaxinae	B	1	28	3
						29	1
		Morpho catenarius	Morphinae	C	1	66	1
		Myscelia orsis	Biblidinae	A	0	31	1
					1	26	1
						28	1
				C	1	27	1
		Narope cyllene	Brassolinae	A	0	34	1
					1	33	1
				B	0	31	1
						34	2

UC	Perturbação (vide Magrini et al., 2011)	Espécie	Subfamília	Idade estimada (A=mais nova)	Sexo (0 fêmea, 1 macho)	Tamanho asa anterior (mm)	Total
					1	31	1
						33	2
				(vazio)	(vazio)	(vazio)	1
		Opoptera syme	Brassolinae	A	1	39	1
				B	0	44	1
						45	1
					1	40	2
						46	1
				C	1	42	2
				D	1	44	1
				(vazio)	(vazio)	(vazio)	1
		Opsiphanes invirae	Brassolinae	A	0	57	1
				B	1	-	1
		Opsiphanes quiteria	Brassolinae	A	1	43	1
						47	1
		Paryphthimoides phronius	Satyrinae	B	1	29	1
				C	0	20	1
		Prepona amphimachus	Charaxinae	B	1	42	1
						46	1
				C	1	46	1
						52	1
						54	1
		Prepona demophon	Charaxinae	B	1	47	1
		Prepona demophoon	Charaxinae	A	1	47	1
		Splendeuphychia doxes	Satyrinae	D	1	21	1
		Splendeuphychia hygina	Satyrinae	A	1	21	1
		Taydebis peculiaris	Satyrinae	A	1	23	1
		Taygetis acuta	Satyrinae	C	1	42	1
		Yphthimoides grymon	Satyrinae	B	0	21	1
R.B. Paranapiacaba Total							706
Total Geral							1886