



---

**PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOLOGIA**

---

**ESTUDO DE COMUNIDADES DE DÍPTEROS NECRÓFAGOS SOB O  
FORMALISMO DE BICLUSTERS E ÁRVORES DE DECISÃO**

**CAROLINE RODRIGUES DE SOUZA**

Tese apresentada ao Instituto de Biociências do Câmpus de Rio Claro, Universidade Estadual Paulista, como parte dos requisitos para obtenção do título de Doutora em Ciências Biológicas (Área de Concentração – Zoologia).

**Agosto - 2016**

**CAROLINE RODRIGUES DE SOUZA**

**ESTUDO DE COMUNIDADES DE DÍPTEROS NECRÓFAGOS SOB O  
FORMALISMO DE GRAFOS E ÁRVORES DE DECISÃO**

Tese apresentada ao Instituto de  
Biotecnologia do Campus de Rio Claro,  
Universidade Estadual Paulista Júlio  
de Mesquita Filho, como parte dos  
requisitos para obtenção do título de  
Doutora em Ciências Biológicas (Área  
de Concentração – Zoologia)

Orientador: Prof. Dr. Claudio José Von Zuben

Rio Claro – SP  
2016

595.77 Souza, Caroline Rodrigues de  
S729e       Estudo de comunidades de dípteros necrófagos sob o  
formalismo de biclusters e árvores de decisão / Caroline  
Rodrigues de Souza. - Rio Claro, 2016  
181 f. : il., figs., tabs.

Tese (doutorado) - Universidade Estadual Paulista,  
Instituto de Biociências de Rio Claro  
Orientador: Claudio José Von Zuben

1. Díptero. 2. Sinantropismo. 3. Moscas. 4. Ambientes  
antrópicos. 5. Zoologia. I. Título.

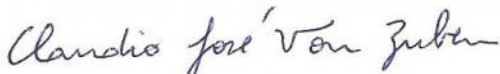
CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: ESTUDO DE COMUNIDADES DE DÍPTEROS NECRÓFAGOS SOB O FORMALISMO DE *BICLUSTERS* E ÁRVORES DE DECISÃO

AUTORA: CAROLINE RODRIGUES DE SOUZA

ORIENTADOR: CLAUDIO JOSÉ VON ZUBEN

Aprovada como parte das exigências para obtenção do Título de Doutora em CIÊNCIAS BIOLÓGICAS (ZOOLOGIA), pela Comissão Examinadora:



Prof. Dr. CLAUDIO JOSÉ VON ZUBEN

Departamento de Zoologia / Instituto de Biociências de Rio Claro - SP



Prof. Dra. PATRICIA JACQUELINE THYSSEN

Departamento de Biologia Animal - Instituto de Biologia / Universidade Estadual de Campinas - SP



Prof. Dr. WESLEY AUGUSTO CONDE GODOY

Departamento de Entomologia e Acarologia / USP - Escola Superior de Agricultura Luiz de Queiroz - Piracicaba/SP



Prof. Dra. MARIA JOSE DE OLIVEIRA CAMPOS

Departamento de Ecologia / Instituto de Biociências de Rio Claro - SP



Prof. Dr. EDILBERTO GIANNOTTI

Departamento de Zoologia / Instituto de Biociências de Rio Claro - SP

Rio Claro, 05 de agosto de 2016

*Aos meus pais, Osmar e Floracy, pela dedicação, sustentação, coragem e amor à mim oferecidos durante esta caminhada.*

*À minha irmã, Leticia e ao meu irmão, Leonardo, pelo carinho, amor e exemplos de vida.*

*Ao meu marido, Ivanildo, pelo amor e paciência nos momentos mais difíceis.*

## **Agradecimentos**

Agradeço, primeiramente a Deus, que me criou e guiou os meus caminhos até aqui.

Agradeço, do fundo do meu coração, aos meus pais, Osmar e Floracy, que tanto se esforçaram para que eu chegasse até aqui. Também sempre me aconselharam a estudar e foram a base e a fortaleza nas horas mais difíceis. Também sou muito grata aos meus irmãos, Leonardo e Letícia, que sempre me apoiaram e estiveram torcendo por mim! Amo muito todos vocês!

Sou muito grata ao apoio, à paciência e ao amor que meu marido Ivanildo tanto dedicou à mim. Nas horas complicadas, era ele que sempre me acalmava e me dava a paz que eu tanto necessitava. Te amo! E não poderia me esquecer da nossa cachorrinha Penellope, que sempre me fez companhia nos vários dias de estudo em casa.

Agradeço também ao apoio de todos os meus amigas ( Meire, Daliane, Stefania, Ana Cláudia, Carolinne, Marcela, Ana Paula...), estando perto ou longe, sempre estiveram presentes e preocupadas em todos esses anos de estudo. Também sou grata aos meus amigos e colegas de laboratório, que, de uma forma ou de outra, fizeram parte desses anos de aprendizado.

À UNESP – Universidade Estadual “Júlio Mesquita Filho”, pela oportunidade da realização deste trabalho.

Ao Prof. Dr. Cláudio José Von Zuben, pela orientação, pelos conhecimentos transmitidos, pelo apoio e pela amizade que se fortaleceu nesses anos de convivência.

Ao pesquisador Prof. Fernando José Von Zuben, agradeço, imensamente, pelo ensinamento das ferramentas matemáticas, pelo auxílio e pela paciência comigo. Também sou grata à Doutoranda Rosana Veronese que me ajudou muito na compreensão dos biclustres.

À FAPESP, pela bolsa de estudo concedida para realização deste estudo.

Portanto, agradeço a todos aqueles que de alguma forma puderam me auxiliar na realização deste trabalho.

## RESUMO

O reconhecimento de padrões em comunidades ecológicas é um dos mais antigos e persistentes desafios da ciência ecológica. Nesse sentido, através da modelagem matemática a partir de grafos, utilizando *biclusters* e árvores de decisão, buscou-se encontrar padrões na ocorrência de dípteros necrófagos coletados em três diferentes áreas (rural, urbana e mata) e com três diferentes iscas (sardinha, fígado e carne moída) na cidade de Rio Claro-SP. A partir dos algoritmos (InClose e RInClose – *biclusters*, C4.5 – árvores de decisão através da análise dos *biclusters*, podemos destacar, por exemplo, aqueles que relacionaram *Atherigona orientalis* (Schiner) e *Musca domestica* Linnaeus em todos os ambientes de coleta, menos na área de floresta, provavelmente devido ao fato de serem espécies conhecidas por sua preferência por áreas habitadas pelo homem. Por outro lado, mesmo sendo espécies que possui alta dependência pela antropobiocenose, outros *biclusters* indicaram uma associação entre *Lucilia eximia* (Wiedemann) e *Chrysomya megacephala* (Fabricius) ocorrendo com maior frequência na floresta, atraídas pelas iscas de carne ou fígado, já conhecidas pela sua importância na atividade de califorídeos. Isso pode ser explicado devido ao crescimento urbano ocorrido ao redor da floresta estudada, dando-lhe características próprias de local habitado pelo homem. No que se refere às árvores de decisão, foram obtidas no total 19 árvores de decisão (sendo 8 para a família Sarcophagidae, 4 para Muscidae e 7 para Calliphoridae) e a partir delas pode-se detectar vários padrões como: *A. orientalis*, *Synthesiomyia nudiseta* (Wulp, 1883) e *M. domestica* estando presente conjuntamente se a estação for Primavera, a isca for sardinha e o local for Urbana. *Oxysarcodexia thornax* (Walker) foi a única espécie, que para atingir a condição de presença, percorreu somente um nó, a estação Primavera, sendo, portanto, considerada generalista neste estudo, já que as variáveis local e isca foram indiferentes para determinar sua localização. Por tudo isso, considerou-se que os *biclusters* e as árvores de decisão são ferramentas úteis na estimativa de padrão na relação existentes entre a comunidade de dípteros necrófagos.

**PALAVRAS-CHAVE:** Sinantropismo, moscas e ambientes antrópicos.

## ABSTRACT

The pattern recognition in ecological communities is one of the most antique and persistent challenges of ecological science. In this sense, through mathematical modelling from graphs, using biclusters and decision trees, the objective of the present work was to find patterns in the occurrence of necrophagous dipterans collected in three different areas (rural, urban and forest) and with three different baits (sardines, liver and beef) in the city of Rio Claro-SP. Algorithms (InClose and RInClose-biclusters, C 4.5-decision trees) were used to obtain the results. Through the analysis of biclusters, it can be highlighted those that related *Atherigona orientalis* (Schiner) and *Musca domestica* Linnaeus in all environments, less forest area, probably due to the fact of being species known for their preference for areas inhabited by humans. On the other hand, even being species possessing high dependence by antropobiocenose, other biclusters indicated an association between *Lucilia eximia* (Wiedemann) and *Chrysomya megacephala* (Fabricius) occurring with greater frequency in the forest, attracted by the bait of meat or liver, both known for their importance in the attractiveness of calliphorid fly. This may be explained due to the population growth occurred around the forest, giving the characteristics of place inhabited by man. With regard to decision trees, were obtained in total 19 decision trees (8 to the family Sarcophagidae, 4 for Muscidae and 7 to Calliphoridae) and from them various patterns can be detected such as: *A. orientalis*, *Synthesiomyia nudiseta* (Wulp, 1883) and *M. domestica* being present together if the season is Spring, the bait is sardines and the location is urban. *Oxysarcodexia thornax* (Walker) was the only specie in which to achieve the condition of presence, it is necessary to walk only one node (the spring), therefore, it was considered to be generalist in this study, since the local variables and bait were indifferent to determine this location. For all that, it was considered that the decision trees and biclusters are useful tools in the elucidation of existing relationship patterns between the community of dipterans collected.

**KEY-WORDS :** Synanthropism, Flies, Anthropic environments.



## LISTA DE ILUSTRAÇÕES

Página

<b>Figura 1</b> – Grafo com conjunto de vértices $V = \{a,b,c,d\}$ e arestas $A = \{ab,bc,cd\}$ .....	17
<b>Figura 2</b> - Diagrama das sete pontes de Königsberg.....	17
<b>Figura 3</b> - Representação gráfica à esquerda de um grafo não direcionado (apenas uma aresta existe entre dois vértices) e à direita de um grafo direcionado (mais de uma aresta é permitida entre dois vértices). Os círculos representam os vértices (nós) e os segmentos de reta representam as ligações (arestas). Os números referem-se aos vértices (nós).....	18
<b>Figura 4</b> - A aresta em vermelho é um laço e as arestas em verde são paralelas múltiplas.....	18
<b>Figura 5</b> - Dado um grafo $G$ (a), os grafos $G_1$ (b) e $G_2$ (c) são subgrafos de $G$ .....	19
<b>Figura 6</b> - Grafos (a) bipartido e (b) grafo bipartido completo.....	20
<b>Figura 7</b> - Grafo $F$ , para visualização de diferentes caminhos entre os nós 1 e 4.....	21
<b>Figura 8</b> - Comparação entre um grafo cíclico e uma árvore.....	22
<b>Figura 9</b> - Estrutura de uma árvore.....	23
<b>Figura 10</b> - Técnicas <i>bottom-up</i> e <i>top-down</i> .....	25
<b>Figura 11</b> - Representação de uma árvore de decisão e sua respectiva representação no Espaço.....	26
<b>Figura 12</b> - Uma árvore de decisão simples para o diagnóstico de um paciente.....	27
<b>Figura 13</b> - Matriz $X$ de Dados.....	46
<b>Figura 14</b> – Multigrafo bipartido.....	47
<b>Figura 15</b> - Estruturas dos <i>biclusters</i> . (a) <i>bicluster</i> único; (b) biclusters com linhas e colunas exclusivas; (c) estrutura de xadrez; (d) biclusters com linhas exclusivas; (e) biclusters com colunas exclusivas; (f) biclusters não-sobrepostos com estrutura de árvore; (g) biclusters não sobrepostos não-exclusivos; (h) biclusters sobrepostos com estrutura hierárquica; e (i) biclusters sobrepostos arbitrariamente posicionados.....	48
<b>Figura 16</b> - Exemplo de um bicluster formado por um subconjunto de objetos e um subconjunto de atributos não-contíguos na matriz de dados original.....	50
<b>Figura 17</b> - Exemplos de tipos diferentes de biclusters. (a) BVC. (b) BVCL. (c) BVCC. (d) BVCo, aditivo. (e) BVCo, multiplicativo. (f) BEC.....	51
<b>Figure 18</b> – Final configuration of the decision tree designed to predict absence or presence of the species <i>Sarcodexia lambens</i> . Notice that the decision tree is automatically generated	

from the collected datasets, by a learning from data approach. The ellipsoidal nodes (internal nodes) perform a test over an attribute, thus being a decision point. The decision outcome will indicate which edge to follow. The square nodes (leaves of the tree) are the available classes. Starting from the root node, there is a single path to each leaf node, and the sequence of decisions from the root to each leaf may be expressed as a single IF-THEN classification rule, making the process highly interpretable.....119

## LISTA DE TABELAS

---

<b>Tabela 1</b> - Espécies de três famílias de Diptera, coletadas em três ambientes diferentes utilizando três tipos de iscas, na cidade de Rio Claro-SP (continua).....	40
<b>Tabela 1</b> - Espécies de três famílias de Diptera, coletadas em três ambientes diferentes utilizando três tipos de iscas, na cidade de Rio Claro-SP.....	41
<b>Tabela 2</b> – Ocorrência das espécies da família Muscidae, atraídas por três tipos de isca nos três ambientes de coleta Os números (1-9), presentes tanto nas linhas quanto nas colunas, são úteis para a geração dos <i>biclusters</i> e posteriormente sua interpretação.....	54
<b>Tabela 3</b> – Ocorrência das espécies da família Calliphoridae, atraídas por três tipos de isca nos três ambientes de coleta. Os números (1-9), presentes tanto nas linhas quanto nas colunas, são úteis para a geração dos <i>biclusters</i> e posteriormente sua interpretação.....	54
<b>Tabela 4</b> – Ocorrência das espécies da família Sarcophagidae, atraídas por três tipos de isca nos três ambientes de coleta Os números (1-9), presentes tanto nas linhas quanto nas colunas, são úteis para a geração dos <i>biclusters</i> e posteriormente sua interpretação.....	55
<b>Tabela 5</b> - Cinco espécies mais abundantes coletadas de Muscidae, distribuídas em seus locais de coleta e isca utilizadas.....	56
<b>Tabela 6</b> - Cinco espécies mais abundantes coletadas de Sarcophagidae, distribuídas em seus locais de coleta e isca utilizadas.....	56
<b>Tabela 7</b> - Cinco espécies mais abundantes coletadas de Calliphoridae, distribuídas em seus locais de coleta e isca utilizadas.....	56
<b>Tabela 8</b> - Cinco espécies mais abundantes coletadas de cada família, distribuídas em seus locais de coleta e isca utilizadas.....	58
<b>Tabela 9</b> – Demonstração em formato de tabela do bicluster OPSM de número 1, obtido através da Tabela 3.....	65
<b>Tabela 10</b> - Demonstração em formato de tabela do bicluster OPSM de número 2, obtido através da Tabela 3.....	66
<b>Tabela 11</b> - Demonstração em formato de tabela do bicluster OPSM de número 3, obtido através da Tabela 3.....	66
<b>Tabela 12</b> - Demonstração em formato de tabela do bicluster OPSM de número 1, obtido através da Tabela 4.....	67
<b>Tabela 13</b> - Demonstração em formato de tabela do bicluster OPSM de número 2, obtido através da Tabela 4.....	67

<b>Tabela 14</b> - Demonstração em formato de tabela do bicluster OPSM de número 3, obtido através da Tabela 4.....	68
<b>Tabela 15</b> - Demonstração em formato de tabela do bicluster OPSM de número 4, obtido através da Tabela 4.....	68
<b>Tabela 16</b> - Demonstração em formato de tabela do bicluster OPSM de número 5, obtido através da Tabela 4.....	69
<b>Tabela 17</b> - Demonstração em formato de tabela do bicluster OPSM de número 6, obtido através da Tabela 4.....	69
<b>Tabela 18</b> - Demonstração em formato de tabela do bicluster OPSM de número 7, obtido através da Tabela 4.....	70
<b>Tabela 19</b> - Demonstração em formato de tabela do bicluster OPSM de número 8, obtido através da Tabela 4.....	70
<b>Tabela 20</b> - Demonstração em formato de tabela do bicluster OPSM de número 1, obtido através da Tabela 5.....	71
<b>Tabela 21</b> - Demonstração em formato de tabela do bicluster OPSM de número 2, obtido através da Tabela 5.....	71
<b>Tabela 22</b> - Demonstração em formato de tabela do bicluster OPSM de número 3, obtido através da Tabela 5.....	72
<b>Tabela 23</b> - Demonstração em formato de tabela do bicluster OPSM de número 4 obtido através da Tabela 5.....	72
<b>Tabela 24</b> - Demonstração em formato de tabela do bicluster OPSM de número 5, obtido através da Tabela 5.....	73
<b>Tabela 25</b> - Demonstração em formato de tabela do bicluster OPSM de número 6, obtido através da Tabela 5.....	73
<b>Tabela 26</b> - Demonstração em formato de tabela do bicluster OPSM de número 7, obtido através da Tabela 5.....	74
<b>Tabela 27</b> - Demonstração em formato de tabela do bicluster OPSM de número 8, obtido através da Tabela 5.....	74
<b>Tabela 28</b> - Demonstração em formato de tabela do bicluster OPSM de número 9, obtido através da Tabela 5.....	75
<b>Tabela 29</b> - Demonstração em formato de tabela do bicluster OPSM de número 10, obtido através da Tabela 5.....	75
<b>Tabela 30</b> - Demonstração em formato de tabela do bicluster OPSM de número 11, obtido através da Tabela 5.....	76

<b>Tabela 31</b> - Demonstração em formato de tabela do bicluster OPSM de número 12, obtido através da Tabela 5.....	76
<b>Tabela 32</b> - Demonstração em formato de tabela do bicluster OPSM de número 13, obtido através da Tabela 5.....	77
<b>Tabela 33</b> - Demonstração em formato de tabela do bicluster OPSM de número 14, obtido através da Tabela 5. ....	77
<b>Tabela 34</b> - Demonstração em formato de tabela do bicluster OPSM de número 1, considerando o zero como dado faltante, obtido através da Tabela 6.....	79
<b>Tabela 35</b> - Demonstração em formato de tabela do bicluster OPSM de número 2, considerando o zero como dado faltante, obtido através da Tabela 6.....	80
<b>Tabela 36</b> - Demonstração em formato de tabela do bicluster OPSM de número 1, considerando o zero como dado faltante, obtido através da Tabela 7.....	80
<b>Tabela 37</b> - Demonstração em formato de tabela do bicluster OPSM de número 2, considerando o zero como dado faltante, obtido através da Tabela 7.....	81
<b>Tabela 38</b> - Demonstração em formato de tabela do bicluster OPSM de número 3, considerando o zero como dado faltante, obtido através da Tabela 7.....	81
<b>Tabela 39</b> - Demonstração em formato de tabela do bicluster OPSM de número 4, considerando o zero como dado faltante, obtido através da Tabela 7.....	82
<b>Tabela 40</b> - Demonstração em formato de tabela do bicluster OPSM de número 5, considerando o zero como dado faltante, obtido através da Tabela 7.....	82
<b>Tabela 41</b> - Demonstração em formato de tabela do bicluster OPSM de número 5, considerando o zero como dado faltante, obtido através da Tabela 7.....	83
<b>Tabela 42</b> - Demonstração em formato de tabela do bicluster OPSM de número 1, considerando o zero como dado faltante, obtido através da Tabela 9.....	83
<b>Tabela 43</b> - Demonstração em formato de tabela do bicluster OPSM de número 2, considerando o zero como dado faltante, obtido através da Tabela 9.....	84
<b>Tabela 44</b> - Demonstração em formato de tabela do bicluster OPSM de número 3, considerando o zero como dado faltante, obtido através da Tabela 9.....	84
<b>Tabela 45</b> - Demonstração em formato de tabela do bicluster OPSM de número 4, considerando o zero como dado faltante, obtido através da Tabela 9.....	85
<b>Tabela 46</b> - Demonstração em formato de tabela do bicluster OPSM de número 5, considerando o zero como dado faltante, obtido através da Tabela 9.....	85

<b>Tabela 47</b> - Demonstração em formato de tabela do bicluster OPSM de número 6, considerando o zero como dado faltante, obtido através da Tabela 9.....	86
<b>Tabela 48</b> - Demonstração em formato de tabela do bicluster OPSM de número 7, considerando o zero como dado faltante, obtido através da Tabela 9.....	86
<b>Tabela 49</b> - Demonstração em formato de tabela do bicluster OPSM de número 8 considerando o zero como dado faltante, obtido através da Tabela 9.....	87
<b>Tabela 50</b> - Demonstração em formato de tabela do bicluster OPSM de número 9 considerando o zero como dado faltante, obtido através da Tabela 9.....	87
<b>Tabela 51</b> - Demonstração em formato de tabela do bicluster OPSM de número 10 considerando o zero como dado faltante, obtido através da Tabela 9.....	88
<b>Tabela 52</b> - Demonstração em formato de tabela do bicluster OPSM de número 11 considerando o zero como dado faltante, obtido através da Tabela 9.....	88
<b>Tabela 53</b> - Demonstração em formato de tabela do bicluster OPSM de número 12 considerando o zero como dado faltante, obtido através da Tabela 9.....	89
<b>Tabela 54</b> - Demonstração em formato de tabela do bicluster OPSM de número 13 considerando o zero como dado faltante, obtido através da Tabela 9.....	89
<b>Tabela 55</b> - Demonstração em formato de tabela do bicluster OPSM de número 14 considerando o zero como dado faltante, obtido através da Tabela 9.....	90
<b>Tabela 56</b> - Demonstração em formato de tabela do bicluster OPSM de número 15 considerando o zero como dado faltante, obtido através da Tabela 9.....	90
<b>Tabela 57</b> - Demonstração em formato de tabela do bicluster OPSM de número 16 considerando o zero como dado faltante, obtido através da Tabela 9.....	91
<b>Tabela 58</b> - Demonstração em formato de tabela do bicluster OPSM de número 17 considerando o zero como dado faltante, obtido através da Tabela 9.....	91
<b>Tabela 59</b> - Demonstração em formato de tabela do bicluster OPSM de número 18 considerando o zero como dado faltante, obtido através da Tabela 9.....	92
<b>Tabela 60</b> - Demonstração em formato de tabela do bicluster OPSM de número 19 considerando o zero como dado faltante, obtido através da Tabela 9.....	92
<b>Tabela 61</b> - Demonstração em formato de tabela do bicluster OPSM de número 20 considerando o zero como dado faltante, obtido através da Tabela 9.....	93
<b>Table 62</b> – Abundance of species of blowflies collected along the four seasons, in three different locations, and using three types of baits. (L: liver, M: minced meat, S: Sardine, S1: <i>S. lambens</i> , S2: <i>O. thornax</i> , S3: <i>P (E) collusor</i> , S4: <i>O. avuncula</i> , S5: <i>P (S) ingens</i> , S6: <i>P (E) anguilla</i> , S7: <i>P (P) intermutans</i> , S8 <i>P: (S) florencioi</i> , S9 : <i>A. orientalis</i> , S10: <i>S. nudiseta</i> , S11:	

*M. domestica*, S12: *O. chalcogaster*, S13: *L. eximia*, S14: *H. segmentaria*, S15: *L. cuprina*, S16: *H. semidiaphana*, S17: *C. megacephala*, S18: *C. albiceps*, S19: *M. peregrina*).....116

**Table 63** - Tendencies in collections involving species of the Sarcophagidae family.....117

**Table 64**- Tendencies in collections involving species of the Muscidae family.....117

**Table 65**- Tendencies in collections involving species of the Calliphoridae family.....118

## SUMÁRIO

---

	Página
<b>INTRODUÇÃO</b> .....	14
DÍPTEROS NECRÓFAGOS.....	15
NOÇÕES SOBRE GRAFOS.....	16
ÁRVORES DE DECISÃO.....	22
REFERÊNCIAS.....	32
<b>OBJETIVOS</b> .....	38
<b>DADOS UTILIZADOS</b> .....	39
REFERÊNCIAS .....	42
<b>CAPÍTULO I – Aplicação de biclusters para modelar a ocorrência de dípteros necrófagos, no sudeste do Brasil</b> .....	43
RESUMO.....	43
INTRODUÇÃO.....	44
MATERIAL E MÉTODOS.....	52
RESULTADOS.....	57
DISCUSSÃO.....	93
CONCLUSÃO.....	95
REFERÊNCIAS.....	96
<b>CAPÍTULO II - Application of decision trees to model the occurrence of dipterans in a monitored ecosystem</b> .....	106
ABSTRACT .....	106
INTRODUCTION.....	107
MATERIAL AND METHODS.....	108
EXPERIMENTAL RESULTS.....	111
DISCUSSION.....	112
REFERENCES.....	120
<b>CONSIDERAÇÕES FINAIS</b> .....	127



APÊNDICE A – Biclusters com valores constantes iguais a 0, obtidos através da Tabela 3, ao se usar dados binários.....	128
APÊNDICE B – Biclusters do tipo BVCC e BVCL, obtidos a partir da Tabela 3.....	130
APÊNDICE C – <i>Biclusters</i> do tipo OPSM, obtidos a partir da Tabela 3 (Muscidae).....	133
APÊNDICE D – <i>Biclusters</i> do tipo OPSM, obtidos a partir da Tabela 4 (Calliphoridae).....	136
APÊNDICE E – <i>Biclusters</i> do tipo OPSM, obtidos a partir da Tabela 5 (Sarcophagidae)....	140
APÊNDICE F – <i>Biclusters</i> do tipo OPSM, obtidos a partir da Tabela 6 (Muscidae).....	146
APÊNDICE G – <i>Biclusters</i> do tipo OPSM, obtidos a partir da Tabela 7 (Calliphoridae).....	151
APÊNDICE H – <i>Biclusters</i> do tipo OPSM, obtidos a partir da Tabela 8 (Sarcophagidae)....	158
APÊNDICE I – Árvores de decisão das espécies de dípteros necrófagos obtidas durante o presente trabalho.....	163

## INTRODUÇÃO

---

O reconhecimento de padrões em comunidades ecológicas é um dos mais antigos e persistentes desafios da ciência ecológica. Por vários anos, ecólogos de comunidades concentraram-se em duas tarefas fundamentais: primeiro, desenvolver métodos adequados para a detecção de padrões em conjuntos de espécies; e segundo, identificar processos capazes de gerar determinados padrões (LEWINSOHN et al., 2006).

Uma grande variedade de sistemas, que se estendem desde redes de genes até redes sociais, tem sido modelada matematicamente a partir de grafos, os quais são estruturas de dados formadas por nós e arestas que interligam par-a-par esses nós (AGNARSSON; GREENLAW, 2006). A grande vantagem desses modelos está na possibilidade de evidenciar não apenas os elementos que compõem o sistema em estudo, mas também e principalmente as interações que eles estabelecem entre si (MASON; VERWOERD, 2007). De acordo com esta visão, os elementos do sistema são nós (genes, pessoas, espécies) conectados por ligações (corregulação de genes, interação de pessoas ou de espécies). Em ecologia, o uso de grafos tem sido aplicado há algum tempo, como nos estudos clássicos de teias tróficas de Connell (1961) e de Paine (1966). Como um processo biológico não é executado somente por um elemento, mas pela interação de múltiplos elementos (EISENBERG et al., 2000), a sua modelagem em grafos permite caracterizar adequadamente os principais aspectos de processos biológicos bastante complexos (BARABÁSI, 2002).

Além de compreender os processos ecológicos através da interação de seus elementos constituintes, muitas vezes busca-se desenvolver modelos de predição e de classificação a partir de dados observados. Dentre as abordagens mais empregadas, encontram-se os mecanismos indutivo-observacionais de obtenção de modelos matemáticos a partir de dados amostrados de fenômenos biológicos (EASTMAN, 2006). Sendo assim, em lugar de uma abordagem *top-down* de modelagem fenomenológica, adota-se aqui uma perspectiva *bottom-up*, centrada em dados e em ferramentas de aprendizado de máquina (DIEKS, 2009).

As árvores de decisão se destacam neste processo de modelagem a partir de dados, por produzirem modelos diretamente interpretáveis, por apresentarem baixo custo computacional, permitindo assim o tratamento de grandes bases de dados, e por obterem alto desempenho em tarefas de predição e classificação (BREIMAN et al., 1984).

Por outro lado, há também a tarefa de agrupamento de dados (*clustering*) que consiste em separar amostras de dados relativas a um determinado domínio de estudo em grupos (*clusters*) distintos, sendo que instâncias pertencentes a um mesmo grupo devem apresentar

características semelhantes (JAIN; MURTY; FLYNN, 1999). Já o agrupamento bidimensional de dados (*biclustering*), que foi aplicado neste trabalho, permite fazer esse agrupamento simultaneamente em linhas e colunas de uma Tabela Relacional e pode ser uma tarefa útil em diversos domínios, tais como em Bioinformática, Previsão de Séries Temporais, Reconhecimento de Padrões, Análise e Processamento de Imagens e Recuperação de Informação. Nomes como biclusterização, agrupamento duplo de dados, *simultaneous clustering*, *co-clustering*, *two-wayclustering*, *blockclustering*, *bidimensional clustering*, entre outros, também são frequentemente usados na literatura para referenciar a tarefa de *biclustering* (BARKOW et al., 2006; CHARRAD; BEM AHMED, 2011; CHENG; CHURCH, 2000; DE FRANÇA, 2010; MADEIRA; OLIVEIRA, 2004; NIE et al., 2011).

Sendo assim, o emprego de modelagem matemática em grafos (*biclusters* e árvores de decisão) tende a ampliar o conhecimento disponível sobre processos ecológicos e, particularmente, sobre a biodiversidade de insetos, além de permitir reconhecer possíveis padrões de relação em suas comunidades. Foram utilizados, no presente estudo, os dados de espécies de dípteros necrófagos coletados na cidade de Rio Claro – SP (SOUZA, 2011; SOUZA; ZUBEN, 2012; SOUZA; ZUBEN, 2016) e, a partir deles, estão sendo gerados modelos de grafos e árvores de decisão. Com os grafos gerados, o objetivo foi compreender melhor a interação dos seguintes fatores: local de coleta, tipo de isca e estação do ano, na determinação de qual (quais) espécie(s) de dípteros necrófagos podem ser mais facilmente encontradas conjuntamente.

## **DÍPTEROS NECRÓFAGOS**

Os dípteros compõem uma das quatro ordens megadiversas de insetos, com mais de 160 mil espécies descritas (PAPE et al., 2011). Esta diversidade está associada a uma enorme variedade de hábitos alimentares, incluindo a necrofagia, em que os adultos e/ou imaturos de dípteros utilizam o recurso encontrado (matéria orgânica de origem animal) como fonte alimentar ou estimulante para cópula e oviposição (SMITH, 1986). Muitas espécies de dípteros necrófagos possuem importância médica e veterinária, uma vez que podem atuar na veiculação de patógenos, tais como bactérias, fungos, vírus, protozoários e ovos de helmintos ao homem e aos animais, além de serem causadoras de miíases (CHOW, 1940; GREENBERG, 1971; GUIMARÃES; PAPAVERO, 1999). Adicionalmente,, sua associação ao ambiente humano ocorre devido às moscas explorarem substâncias e resíduos orgânicos

produzidos pela atividade humana e animal, especialmente fezes e resíduos vegetais como fonte energética (MONTEIRO, 1995).

Apesar dos hábitos necrófagos de alguns dípteros muscóides constituírem uma ameaça à saúde humana e animal, este comportamento torna-os consumidores importantes de carcaças de vertebrados e recicladores de nutrientes na natureza (BYRD; CASTNER, 2001). A maioria das fêmeas das espécies ovipositam em carcaças, tornando-as importantes agentes consumidoras desses recursos (VARGAS; WOOD, 2010). Os representantes de dípteros necrófagos mais comumente associados à colonização de carcaças são os muscóides, dentre eles podem-se citar as famílias: Calliphoridae, Sarcophagidae e Muscidae (PUJOL-LUZ et al., 2008).

Em virtude do hábito alimentar descrito acima, tais insetos podem ser de grande utilidade nos processos de investigação criminal, permitindo aos peritos estimarem o intervalo pós-morte, através da presença e frequência desses insetos e também da bionomia dos seus estágios imaturos (AMENDT et al., 2007). Essa ciência é conhecida como Entomologia Forense, relacionando o conhecimento sobre insetos e outros artrópodes às diversas questões criminais ( OLIVEIRA-COSTA, 2003; PUJOL-LUZ et al., 2008)

A distribuição, abundância e composição da fauna necrófaga e seus padrões de dinâmica populacional, diferem de acordo com a área geográfica (ANDERSON; VANLAERHOVEN, 1996; ARNALDOS et al., 2001, CARVALHO et al. 2004). As causas dessas diferenças certamente estão ligadas a fatores ambientais, tais como a temperatura, além dos fatores biológicos intrínsecos do organismo (SERRA et al., 2007). É relevante observar a influência das interações entre esses fatores sobre a composição da fauna de dípteros necrófagos, ainda mais no que se refere às moscas-varejeiras que são os primeiros artrópodes a colonizarem corpos em decomposição (SMITH, 1986).

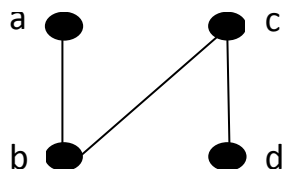
## **NOÇÕES SOBRE GRAFOS**

Um grafo é uma estrutura de abstração bastante útil na representação e na solução de diversos tipos de problemas (GOLDBARG; GOLDBARG, 2012), sendo essas suas maiores vantagens. O termo grafo provém de uma expressão de notação gráfica e foi introduzido pela primeira vez pelo químico Edward Frankland e posteriormente usado por Alexander Crum Brown (1884) (FOULDS, 1992).

Usando uma definição simples, tem-se que um grafo é um modelo matemático que representa uma coleção de objetos (chamados vértices), que são ligados aos pares por outra

coleção de objetos (chamados arestas ou arcos). Na forma de ilustração, os vértices podem ser representados por pontos, caixas ou círculos, e as arestas por linhas conectando os vértices (Figura 1).

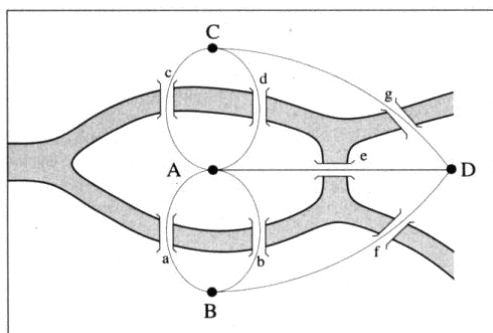
Figura 1 – Grafo com conjunto de vértices  $V = \{a,b,c,d\}$  e arestas  $A = \{ab,bc,cd\}$ .



Grafo é um par ordenado de conjuntos disjuntos  $(V; A)$  tal que  $A$  é um subconjunto do conjunto de pares não ordenados de  $V$ . O conjunto  $V$  é chamado de conjunto dos vértices (ou nós) e o conjunto  $A$  de conjunto das ligações (ou arestas). Diz-se que um elemento  $(v1; v2)$  de  $A$  liga os vértices  $v1$  e  $v2$  e que  $v1$  e  $v2$  são vértices adjacentes. Duas ligações são adjacentes se elas compartilham um mesmo vértice (BALAKRISHNAN; RANGANATHAN, 2012).

O primeiro e mais famoso problema em Teoria dos Grafos foi o das Pontes de Königsberg (na Antiga Prússia) resolvido por Euler (1736). O quebra-cabeça famoso na época era encontrar um passeio que visitasse todas as pontes da cidade de Königsberg (Figura 2), passando uma única vez em cada ponte. Analisando esse diagrama abstrato, Euler provou que tal passeio era impossível.

Figura 2 - Diagrama das sete pontes de Königsberg



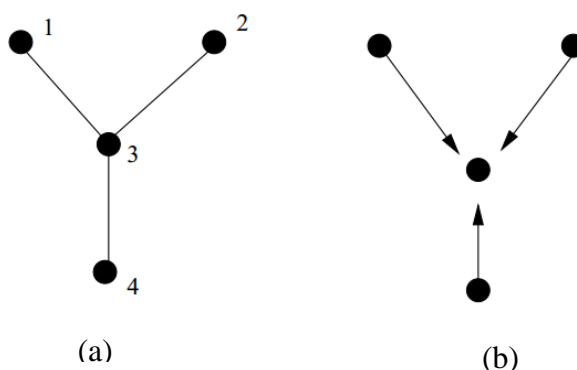
Fonte: Arquivo Escolar (2010).

A seguir, encontram-se algumas definições sobre grafos.

a) Grafos orientados e não orientados

Em alguns casos, os grafos podem possuir uma orientação específica em cada aresta. Os que incluem essa informação são ditos **grafos orientados** (Figura 3b) e aqueles que não a apresentam, são ditos **não orientados** (Figura 3a). Em um grafo orientado, as arestas são pares ordenados, ou seja, um vértice é considerado a “origem” da aresta e o outro seu “destino”. Nesse caso, as arestas possuem um sentido marcado por uma seta e recebem o nome de **arcos** (GOLDBARG; GOLDBARG, 2012).

Figura 3 - Representação gráfica à esquerda de um grafo não orientado (apenas uma aresta existe entre dois vértices) e à direita de um grafo orientado (mais de uma aresta é permitida entre dois vértices). Os círculos representam os vértices (nós) e os segmentos de reta representam as ligações (arestas). Os números referem-se aos vértices (nós).



b) Laço

Os nós constituintes de uma aresta podem ou não ser diferentes. No caso de não serem, a aresta forma um **laço** (Figura 4 - aresta no número 3). Algumas definições permitem laços no grafo; outras proíbem laços, exigindo que os dois extremos de cada aresta sejam vértices distintos. E quando as arestas ligam os mesmos pares de nós, elas são conhecidas como **arestas paralelas** (Figura 4 – arestas ligando os nós 4 e 5).

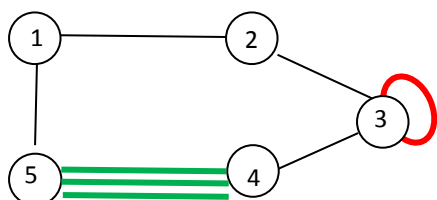


Figura 4 – A aresta no número 3 é um laço e as arestas entre os números 4 e 5 são paralelas múltiplas.

c) Grafo simples, pseudografo, multigrafo

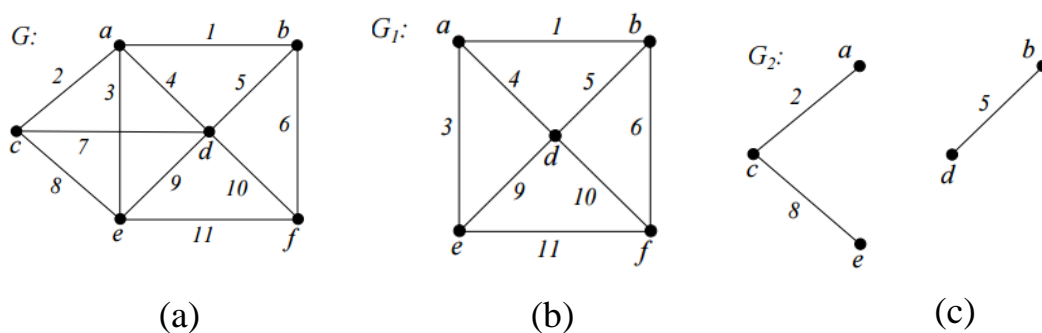
Quando um grafo não possui laços nem arestas paralelas, ele é conhecido como **grafo simples**. Se houver no mínimo um laço, o grafo pode ser chamado de **pseudografo**. Se esse possuir todos os vértices com um laço associado, ele é denominado **pseudografo reflexivo**.

Um grafo não direcionado que possui no mínimo duas arestas paralelas é denominado **multigrafo**. Quando o multigrafo possuir dois ou mais arcos de mesma direção ligando um mesmo par de vértices, ele é conhecido como **multigrafo direcional** (GOLDBARG;GOLDBARG, 2012).

d) Subgrafo

Um grafo  $H(V', A')$  é um subgrafo de um grafo  $G(V, A)$  se todos os vértices e todas as arestas de  $H$  pertencem a  $G$  ( $V' \subseteq V, A' \subseteq A$ ), e cada aresta de  $H$  possui as mesmas extremidades que em  $G$  (Figura 5). Denotamos um subgrafo através da mesma notação usada para conjuntos, isto é  $H \subset G$ .

Figura 5 – Dado um grafo  $G(a)$ , os grafos  $G_1(b)$  e  $G_2(c)$  são subgrafos de  $G$ .



Fonte: OLIVEIRA; RANGEL, 2013.

Nesse contexto, podemos afirmar que:

- Todo grafo é um subgrafo de si próprio.
- Um subgrafo de um subgrafo de um grafo  $G$  também é um subgrafo de  $G$ .
- Um vértice de um grafo  $G$  é um subgrafo de  $G$ .

- Uma aresta de um grafo  $G$  é um subgrafo de  $G$ .

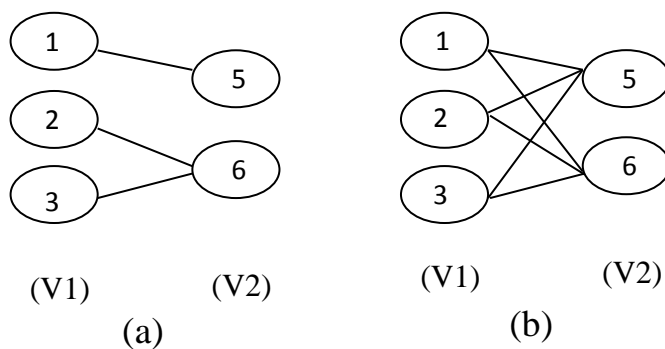
e) Grau do vértice

Dado um grafo  $G$  e um vértice  $v \in V(G)$ , designa-se por *grau* ou *valência* de  $v$  e denota-se por  $d_G(v)$  como sendo o número de arestas de  $G$  incidentes em relação a  $v$ . Nesta definição, cada laço deve ser contado duas vezes.

f) Grafos bipartidos

Um grafo bipartido é aquele em que o conjunto  $V$  de vértices pode ser dividido em dois subconjuntos disjuntos  $V_1$  e  $V_2$ , tais que todas as arestas têm uma terminação em  $V_1$  e uma em  $V_2$ . Não há arestas entre elementos do mesmo conjunto. E um grafo bipartido completo ocorre quando todos os vértices de  $V_1$  são ligados a todos os vértices de  $V_2$  (Figura 6).

Figura 6 - Grafos (a) bipartido e (b) grafo bipartido completo.



g) Grafo Conexo

Um grafo é dito **conexo** se para todo par de vértices  $i$  e  $j$  existe pelo menos um caminho entre  $i$  e  $j$  (GOLDBARG ;GOLDBARG, 2012). Caso contrário o grafo é **desconexo**. Um grafo é **totalmente desconexo** quando não existe nenhuma aresta.



Em virtude da simetria dos grafos, a existência de um caminho de  $i$  a  $j$  equivale à existência de um caminho de  $j$  a  $i$ . Portanto, um grafo é conexo se e somente se quaisquer dois de seus vértices são ligados por um caminho.

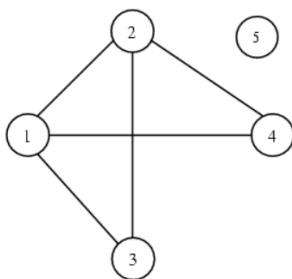
#### h) Passeio, Caminho e Ciclo

Um **passeio** em um grafo  $F = (V, E)$  é uma seqüência alternada de vértices e arestas que começa e termina com vértices.

Um **caminho** é uma cadeia sem repetição de vértices (GOLDBARG; GOLDBARG, 2012). Ou seja, um caminho é um passeio que não possui nós repetidos. Na Figura 7, podemos notar que no grafo  $F$ , entre os nós 1 e 4, temos os seguintes caminhos  $(1,4)$ ,  $(1,2,4)$ ,  $(1,3,2,4)$ . E o **comprimento de um caminho** é o número de arcos que ele contém. Ciclos de comprimento 1 são laços (*loops*).

Um **caminho** considerado **fechado** é conhecido por **ciclo**, isto é, um passeio que contém exatamente dois nós iguais: o primeiro e o último. Um grafo sem ciclos é dito **acíclico**.

Figura 7 - Grafo  $F$ , para visualização de diferentes caminhos entre os nós 1 e 4.



Fonte: Prestes (2012)

#### i) Árvore

A **árvore** é um tipo especial de **grafo**, útil na representação de dados. Cada elemento (nó) tem zero ou mais sucessores, mas tem apenas um predecessor, exceto o primeiro nó, a

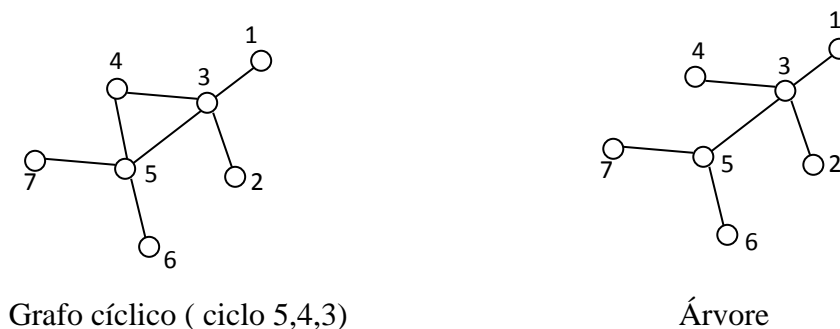
raiz da árvore. Ela é constituída por grafos que não admitem ciclos (**grafos acíclicos**) e que são **conexos** (GOLDBARG; GOLDBARG, 2012). Na Figura 8, têm-se uma comparação entre árvore e um grafo cíclico.

Para um grafo  $G=(V, A)$  de  $n$  - vértices ( $n > 0$ ), as seguintes afirmações são verdadeiras (e caracterizam uma árvore com  $n$  vértices):

- $G$  é conexo e não possui ciclos;
- $G$  é conexo e tem  $n-1$  arestas;
- $G$  tem  $n-1$  arestas e nenhum ciclo;
- Para dois vértices  $u, v$ , há exatamente um caminho entre  $u$  e  $v$ .

Em outras palavras, uma árvore é um grafo constituído por um conjunto de nós e um conjunto de arcos que ligam pares de nós, em que: cada arco liga um **nó-pai** a um ou mais **nós-filhos** e todos os nós, com exceção da raiz, têm um nó-pai. Ao predecessor (único) de um nó, chama-se **nó-pai** e os seus sucessores são os **nós-filhos**. O **grau de um nó** é o número de sub-árvores (ou nós-filhos) que descendem desse nó. Um **nó-folha** não tem filhos, tem grau 0. Um **nó-raiz** não tem pai. Os arcos que ligam os nós chamam-se **ramos** (Figura 9).

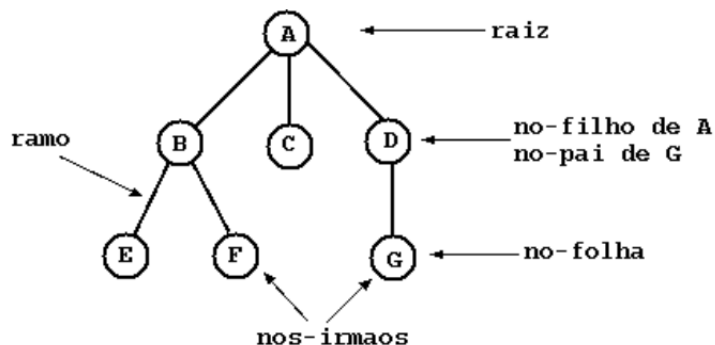
Figura 8- Comparação entre um grafo cíclico e uma árvore



## ÁRVORES DE DECISÃO

Árvores de decisão são modelos estatísticos que utilizam um treinamento supervisionado para a classificação e previsão de dados. Em outras palavras, em sua construção é utilizado um conjunto de treinamento formado por entradas e saídas (são conhecidas como classes).

Figura 9 – Estrutura de uma árvore.



Embora diversos modelos matemáticos tenham surgido nas diferentes áreas do conhecimento relacionadas à classificação, as árvores de decisão têm sido consideradas como um dos modelos mais adequados para a extração do conhecimento para a Aprendizagem Automática, pois são simples e de fácil compreensão; e podem ser expressas numa sub-rotina em qualquer linguagem de programação (MICHIE et al., 1994).

Amplamente utilizadas em algoritmos de classificação, as árvores de decisão são representações simples do conhecimento, utilizam a estratégia de dividir para conquistar: um problema complexo é decomposto em sub-problemas mais simples e recursivamente esta técnica é aplicada a cada sub-problema. A classificação, através de uma árvore de decisão, ocorre à medida que são percorridos os caminhos descritos pelos nós (nodos), até ser encontrado um nó que contém a característica determinante do caminho seguido, recebendo então o nome de folha (GAMA, 2004).

O ponto principal das Árvores de Decisão é a tomada de decisões considerando para isso os atributos mais relevantes e compreensíveis para a maioria das pessoas. Ao escolher e apresentar os atributos em uma ordem de importância, as Árvores de Decisão permitem conhecer quais fatores que mais influenciam na sua construção.

Elas são muito úteis em atividades de mineração de dados, isto é, o processo de extração de informações previamente desconhecida, a partir de grandes bases de dados. Aplicações desta técnica podem ser vista em diversas áreas, desde cenários de negócios até sistemas de piloto automático de aeronaves e diagnósticos médicos. São também uma forma

de construir classificadores que predizem ou revelam classes ou informações úteis baseadas nos valores de atributos de um conjunto de dados.

### **Construção de uma árvore de decisão**

A construção dos modelos computacionais de classificação emprega comumente um dos paradigmas relacionados abaixo (HEIJST et al., 1997) e eles podem ser melhor compreendidos através da Figura 10.

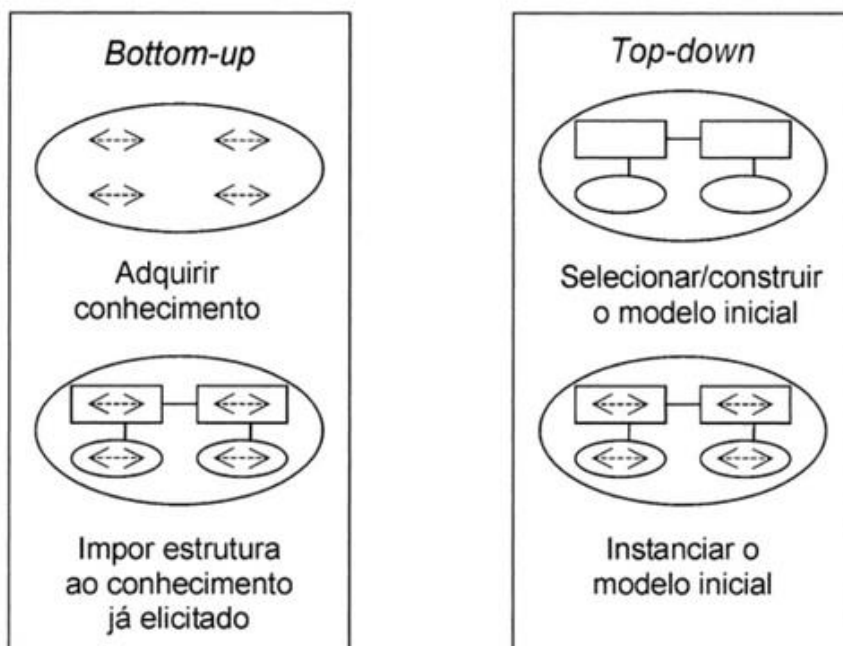
- *Top-down*: refere-se ao processo de construção do modelo de classificação a partir de informações fornecidas por especialistas;
- *Bottom-up*: refere-se ao processo de refinamento em que um modelo abstrato é selecionado ou construído, e depois instanciado com conhecimento específico da aplicação.

Como afirmado por Rezende (2003), os algoritmos que induzem árvores de decisão pertencem à família de algoritmos *Top Down Induction of Decision Trees* – TDIDT, ou seja, do nó raiz em direção às folhas. Embora haja diferenças significativas na forma de efetuar os passos, qualquer algoritmo desta categoria utiliza a técnica de dividir para conquistar. Como já explicado anteriormente, esta filosofia baseia-se na sucessiva divisão do problema em vários sub-problemas de menores dimensões, até que uma solução para cada um dos problemas mais simples seja encontrada (CASTANHEIRA, 2008). O TDIDT 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. De acordo com BRAMER (2007), esse processo é conhecido como particionamento recursivo (uma árvore de decisão particiona recursivamente um conjunto de dados até que cada subconjunto obtido deste particionamento possua casos de uma única classe).

Para atingir este objetivo, esse tipo de algoritmo examina e compara a distribuição de classes durante o processo de construção da árvore de decisão. Através disso, os dados ficam organizados de maneira compacta, sendo utilizados para classificar novos casos (GOLDSCHMIDT, 2010). Uma vez que o número de árvores de decisão possíveis aumenta à medida que o número de atributos também aumenta, torna-se impraticável buscar a estrutura

da árvore de decisão ótima para um determinado problema, devido ao elevado custo computacional envolvido (problema NP-Completo). Nesse sentido, há algoritmos baseados em heurísticas que, mesmo não garantindo uma solução ótima, apresentam resultados aceitáveis (VON ZUBEN; ATTUX, 2010)

Figura 10 – Técnicas *bottom-up* e *top-down* .

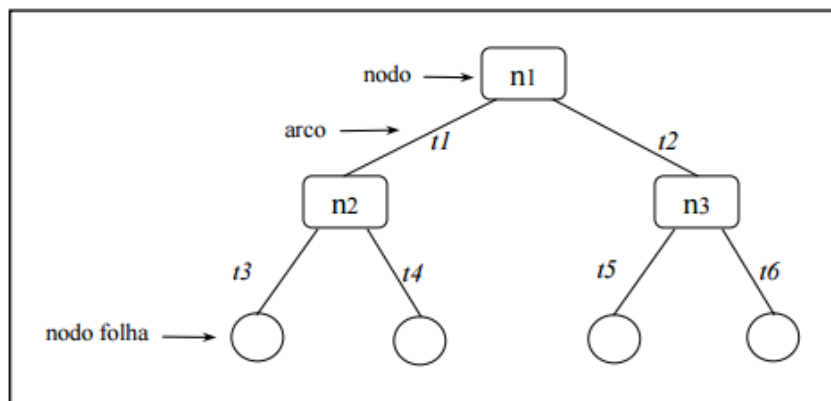


Fonte : Heijst; Shreiber; Wielinga (1997)

### Representação e exemplo de uma árvore de decisão

A Figura 11, a seguir, representa uma árvore de decisão. Nota-se que cada nó (nodo) de decisão possui um teste para algum atributo e cada ramo descendente corresponde a um possível valor deste atributo. Além disso, o conjunto de ramos é distinto e cada nó folha está associada a uma classe. E a regra de classificação é cada percurso que árvore faz da raiz à folha (GAMA, 2004; GARCIA, 2003).

Figura 11 - Representação de uma árvore de decisão e sua respectiva representação no espaço.



Fonte: Gama (2004)

Na Figura 12, têm-se um exemplo de uma árvore de decisão para o diagnóstico de um paciente. Cada elipse é uma pergunta (teste) em um atributo para um dado conjunto de pacientes. Cada retângulo representa uma classe, nesse caso, o diagnóstico. Para classificar (diagnosticar) um paciente, é necessário começar pela raiz e seguir os testes até alcançar a folha.

Como as regras que representam uma árvore de decisão são disjuntas, isto é, apenas uma única regra dispara quando um novo exemplo é classificado, uma outra forma de representar tais regras é escrevê-las de forma separada para cada nó-folha, iniciando pela raiz (REZENDE, 2003). As regras de decisão, ou regras de classificação, são estruturas do tipo “Se <condição> então <conclusão>”, em que <condição> é um conjunto de atributos e seus valores e <conclusão> é uma classe do conjunto de dados.

Abaixo, têm-se as regras da árvore exemplificada anteriormente:

**Se** Paciente se sente bem = Sim **Então**

Classe = Saudável

**Fim Se**

**Se** Paciente se sente bem = não **and** Paciente tem dor= não

**And** Temperatura do Paciente  $\leq$  37 **Então**

Classe=Saudável

**Fim Se**

**Se** Paciente se sente bem = não **and** Paciente tem dor= não

**And** Temperatura do Paciente  $> 37$  **Então**

Classe = Doente

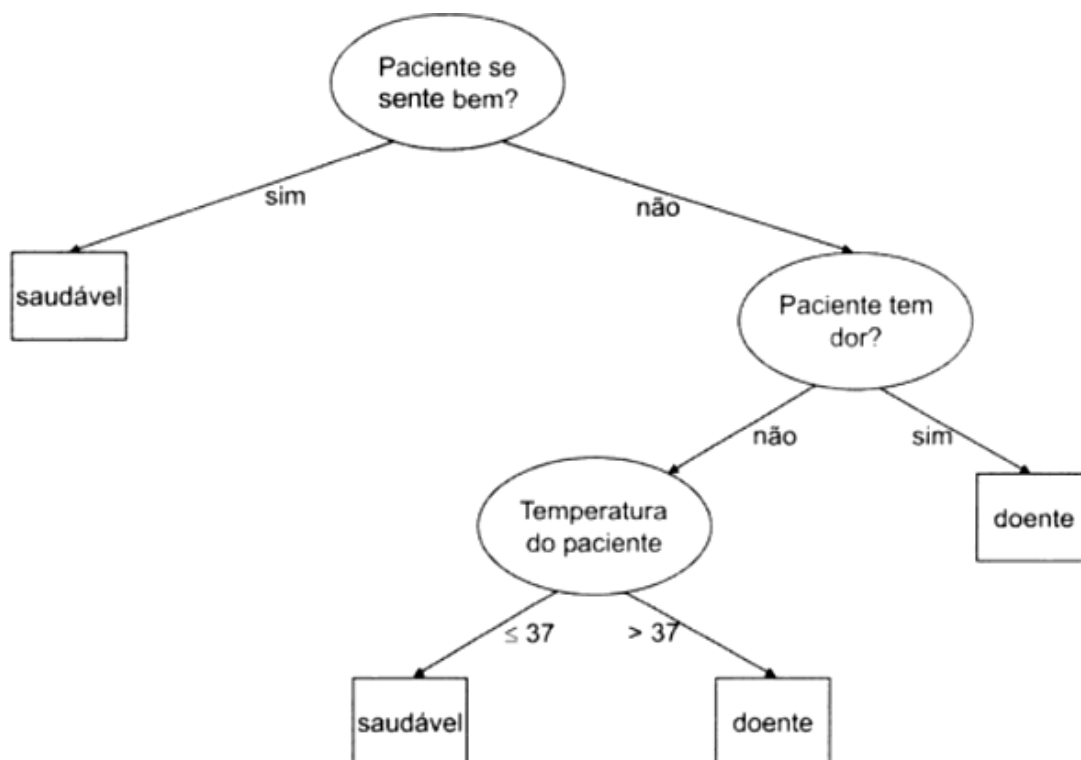
**Fim Se**

**Se** Paciente se sente bem = não **and** Paciente tem dor= sim

Classe = Doente

**Fim Se**

Figura 12- Uma árvore de decisão simples para o diagnóstico de um paciente



Fonte: Rezende (2003)

### Indução de árvores de decisão

De uma forma geral, Quinlan (1986) afirma que se os atributos de uma árvore de decisão estão adequados, é sempre possível construir uma árvore de decisão que classifica de forma correta cada objeto no conjunto de treinamento. De acordo com Tan et al. (2005), durante o processo de construção de modelo através de um algoritmo de aprendizado, é necessário construí-lo com uma boa capacidade de generalização, ou seja, que consiga prever, com alta

taxa de acerto, rótulos de classes para exemplos que não foram utilizados na construção do modelo.

No intuito de fazer isso, a árvore de decisão deve capturar alguma relação significativa entre a classe de um objeto e os valores dos seus atributos. Dada a escolha entre duas árvores de decisão, sendo ambas corretas em relação ao conjunto de treinamento, parece sensato preferir a mais simples, que é a provável em melhor capturar a estrutura inerente ao problema. É esperado que árvore mais simples seria, portanto, a melhor em classificar corretamente mais objetos fora do conjunto de treinamento.

### **Poda**

No processo de construção de árvores de decisão, algumas arestas ou sub-árvores podem refletir ruídos ou erros, gerando um problema conhecido como sobreajuste, que significa um aprendizado muito específico do conjunto de treinamento, não permitindo ao modelo generalizar. No intuito de se evitar isso, recorre-se à limitação e à redução do tamanho da árvore, respectivamente, através das abordagens de pré-poda e pós-poda (ESPOSITO et al., 1997), cujo objetivo é melhorar a taxa de acerto do modelo para novos exemplos, os quais não foram utilizados no conjunto de treinamento (HAN, 2001).

De acordo com Von Zuben; Attux (2010), o processo de podagem pode ser feito a partir dos seguintes passos:

1. Percorre a árvore em profundidade.
2. Para cada nó de decisão calcula: erro no nó + soma dos erros nos nós descendentes
3. Se o erro do nó é menor ou igual à soma dos erros dos nós descendentes então o nó é transformado em folha.

A medida da diferença dada por uma função baseada nas proporções das classes entre o nó corrente e seus descendentes valoriza a pureza das partições. Na pós-podagem a árvore é gerada no tamanho máximo e então a árvore é podada aplicando métodos de evolução confiáveis. A sub-árvore com o melhor desempenho será a escolhida. Este processo pode ser computacionalmente ineficiente pelo fato de gerar uma árvore muito grande e depois esta mesma árvore é reduzida a uma árvore mínima. Para interromper o crescimento da árvore, verifica-se se a divisão é confiável ou não. Caso seja confiável, interrompe-se o crescimento



da árvore. Este processo é conhecido como pré-podagem da árvore. A pré-podagem é mais rápida porém menos eficiente que a pós-podagem pelo fato do risco de interromper o crescimento da árvore ao selecionar uma árvore sub-ótima (BREIMAN, 1984).

A pós-poda é a abordagem mais utilizada e mais confiável, mas requer um processo mais lento, enquanto que a pré-poda tem a vantagem de não gastar tempo na construção de uma estrutura que não será utilizada no final da árvore. Nem sempre a árvore podada é mais precisa que a correspondente gerada, mas a poda ajuda a simplificar a árvore, o que é essencial em árvores muito complexas.

## Algoritmos

Há muitos algoritmos de classificação que elaboram as árvores de decisão. Não há uma forma de determinar qual é o melhor, tendo em vista que o seu desempenho pode variar de acordo com a situação (GOLDSCHMIDT, 2010).

Os estudos que permitiram o aparecimento das árvores de decisão tiveram início com o professor Ross Quinlan da Universidade de Sidney. A sua contribuição foi a elaboração de um algoritmo chamado ID3 – Iterative Dichotomiser 3 - desenvolvido em 1983 (QUINLAM, 1993), baseado em sistemas de inferência e em conceitos de sistemas de aprendizagem. A partir de um conjunto de exemplos, ele constrói árvores de decisão que serão usadas para classificar amostras futuras. Além disso, o ID3 separa um conjunto de treinamento em subconjuntos, de forma que estes contenham exemplos de uma única classe. Essa divisão é feita através de um único atributo selecionado a partir de uma propriedade estatística, conhecida como **ganho de informação**, que mede quanto informativo é um atributo. Um ponto importante desse algoritmo é que ele é limitado, já que ele não trabalha com atributos do tipo contínuo (CASTANHEIRA, 2008).

O ID3 usa o **ganho de informação** para selecionar, entre os candidatos, os atributos que serão utilizados a cada passo, enquanto constrói a árvore. E para se obter o ganho de informação, é necessário, primeiramente, calcular a **entropia** que caracteriza a pureza/impureza dos dados: em um conjunto de dados e é uma medida da falta de homogeneidade dos dados de entrada em relação a sua classificação (MITCHELL, 1997). Nos casos em que a árvore é usada para classificação, os critérios de partição mais conhecidos são baseados na entropia.

Para determinar se uma condição de teste realizada é realmente boa, é preciso comparar o grau de entropia do nó-pai (antes da divisão) com o grau de entropia dos nós-

filhos (após a divisão). O atributo que gerar uma maior diferença é escolhido como condição de teste. O ganho é dado pela Equação (1), na forma:

$$\text{Ganho} = \text{entropia (pai)} - \sum_{j=1}^n \left[ \frac{N(v_j)}{N} \text{entropia}(v_j) \right] \quad (1)$$

onde  $n$  é o número de valores do atributo, ou seja, o número de nós-filhos,  $N$  é o número total de objetos do nó-pai e  $N(v_j)$  é o número de exemplos associados ao nó filho  $v_j$ .

O grau de **entropia** é definido pela Equação (2) a seguir:

$$\text{Entropia (nó)} = - \sum_{i=1}^c p(i/\text{nó}) \cdot \log_2 [p(i/\text{nó})] \quad (2)$$

onde  $p(i/\text{nó})$  é a fração dos registros pertencentes à classe  $i$  no nó, e  $c$  é o número de classes.

A entropia terá valor máximo (igual a 1) quando o conjunto de dados for heterogêneo, ou seja, quando  $x$  predizer totalmente  $y$ , e será 0 quando  $x$  e  $y$  não apresentarem nenhuma associação (MITCHEL, 1997). O critério de ganho seleciona como atributo-teste aquele que maximiza o ganho de informação. Ao se utilizar o ganho de informação, pode-se ter um grande problema: dar preferência a atributos com muitos valores possíveis (número de arestas). Buscando evitar isso, Quinlan (1993) criou a **Razão de Ganho** (do inglês Gain Ratio), que nada mais é do que o ganho de informação relativo (ponderado) como critério de avaliação. A **razão de ganho** é definida pela Equação (3), na forma:

$$\text{Razão de ganho (nó)} = \frac{\text{ganho}}{\text{entropia (nó)}} \quad (3)$$

Quinlan (1988) sugere ainda que isso seja feito em duas etapas: (1) calcular o ganho de informação para todos os atributos, considerando apenas aqueles atributos que obtiveram um ganho de informação acima da média, (2) a partir disso, escolher aquele que apresentar a melhor razão de ganho. Sendo assim, Quinlan mostrou que a razão de ganho supera o ganho de informação tanto em termos de acurácia quanto em termos de complexidade das árvores de decisão geradas.

O algoritmo C4.5 (QUINLAN, 1993), utilizado no presente trabalho, representa uma significativa evolução do ID3 (QUINLAN, 1986). Por ter mostrado excelentes resultados em problemas de classificação, ele é um dos algoritmos mais utilizados na literatura. As principais contribuições que ele trouxe em relação ao ID3 são (DE ANDRADE,2013):

- Trabalha tanto com atributos categóricos (ordinais ou não-ordinais) como também com atributos contínuos. Para lidar com os últimos, esse algoritmo C4.5 define um limiar e divide os exemplos de forma binária: aqueles cujo valor do atributo é maior que o limiar e aqueles cujo valor do atributo é menor ou igual ao limiar;
- O algoritmo C4.5 permite a utilização de valores desconhecidos que são representados como '?' e ele trata esses valores de forma especial. Esses valores não são utilizados nos cálculos de ganho e entropia;
- Utiliza a medida de razão de ganho para selecionar o atributo que melhor divide os exemplos. Essa medida se mostrou superior ao ganho de informação, gerando árvores mais precisas e menos complexas;
- Lida com problemas em que os atributos possuem custos diferenciados;
- Apresenta um método de pós-poda das árvores geradas. O algoritmo C4.5 faz uma busca na árvore, de baixo para cima, e transforma em nós folha aqueles ramos que não apresentam nenhum ganho significativo.

A ferramenta de mineração de dados WEKA (WITTEN & FRANK, 1999) (<http://www.cs.waikato.ac.nz/~ml/weka/index.html>) (*Waikato Environment for Knowledge Analysis*) disponibiliza a implementação do algoritmo C4.5, porém o mesmo é chamado de J48 nessa ferramenta. De acordo com De Andrade (2013), o C4.5 é do tipo:

- Guloso: executa sempre o melhor passo avaliado localmente, sem se preocupar se este passo, junto à sequência completa de passos, vai produzir a melhor solução ao final;
- “Dividir para conquistar”: partindo da raiz, criam-se sub-árvores até chegar nas folhas, o que implica em uma divisão hierárquica em múltiplos subproblemas de decisão, os quais tendem a ser mais simples que o problema original.

---

**REFERÊNCIAS**

---

AGNARSSON, G.; GREENLAW, R. **Graph Theory: Modeling, Applications, and Algorithms** Prentice Hall; 1 edition ,2006.

AMENDT, J. et al. Best practice in forensic entomology – standards and guidelines. **Int. J. Legal Med.**, v.121, p.90-104, 2007.

ANDERSON, G.; VANLAERHOVEN, S.H.. Initial studies on succession on carrion in the carrion in Southwestern British Columbia. **J. Foren. Sci.** 41: 617-625,1996.

ARAÚJO, J. M. F. R. **Inteligência Artificial I - Aprendizagem (Parte II)**, Disponível em [http://www.dsc.ufcg.edu.br/~joseana/IAPos\\_NA16\\_2.pdf](http://www.dsc.ufcg.edu.br/~joseana/IAPos_NA16_2.pdf) Acesso em 6 de janeiro de 2015.

ARNALDOS I, et al. An initial study on the succession of sarcosaprophagous Diptera (Insecta) on carrion in the southeastern Iberian Peninsula. **Int. J. of Legal Medi.** 114, 156-162, 2001.

ARQUIVO ESCOLAR. **Capítulo 3: Introdução a Teoria de Grafos.** 2010. Disponível em : [http://arquivoescolar.org/bitstream/arquivo-e/45/2/metodos\\_finitos\\_II.pdf](http://arquivoescolar.org/bitstream/arquivo-e/45/2/metodos_finitos_II.pdf). Acesso em 10 jan 2014.

BALAKRISHNAN R; RANGANATHAN K (2012). **A textbook of graph theory.** Springer, Berlin.

BARABÁSI, A-L; et al. **Evolution of the Social Network of Scientific Collaborations.** Physica A, 311. 2002. pág. 590-614.

BARKOW, S. et al. **BicAT: a biclustering analysis toolbox.** Bioinformatics 22, 1282-1283, 2006.

BRAMER, M. **Principles of data mining.** Springer, London, 2007.

BREIMAN L, FRIEDMAN JH, OLSHEN RA, STONE CJ. **Classification and regression trees**. Pacific Grove: Cole Advanced Books and Software; 1984.

BYRD, J. H.; CASTNER, J. L. **Insects of forensic importance**, p. 43–80. In: J. H. Byrd & J. L. Castner (eds.). *Forensic Entomology: the utility of arthropods in legal investigations*. Boca Raton, CRC Press, xvi + 418 p, 2001

CARVALHO L.M.L., et al..Observations on the succession patterns of necrophagous insects on a pig carcass in an urban area of Southeastern Brazil. **Aggrawal's Int. J. Foren. Med. Toxic.** 5, 33-39, 2004.

CASTANHEIRA, L. C. **Aplicação de Técnicas de Mineração de Dados em Problemas de Classificação de Padrões**. UFMG, Belo Horizonte, 2008 (Dissertação). Disponível em: <https://online.unisc.br/seer/index.php/tecnologia/article/viewFile/3283/2692>, Acesso em 10 de abril de 2014.

CHARRAD M.; BEN AHMED, M.**Simultaneous clustering: A survey**. In: Kuznetsov SO, Mandal DP, Kundu MK, Pal SK (eds) *Pattern Recognition and Machine Intelligence, Lecture Notes in Computer Science*, vol 6744, Springer, Berlin Heidelberg, pp 370–375, DOI 10.1007/978-3-642-21786-9 60, 2011.

CHENG, Y.; CHURCH, G. M. **Biclustering of expression data**. In Proc. 8th Intl. Conf. on Intelligent Systems for Molecular Biology, pages 93–103. AAAI Press, 2000.

CHOW, C.Y. The commom blue bottle fly *Chrysomya megacephala* as a carrier of pathogenic bacteria in Peiping. **China Chin. Med.**, v.57, p.145-153, 1940.

DE ANDRADE, CESAR A. B. **Análise Automática de Malwares Utilizando as Técnicas de Sandbox e Aprendizado de Máquina**. Dissertação (Mestrado em Sistemas e Computação) - Instituto Militar de Engenharia, 2013.

DE FRANÇA, F. O. **Biclusterização na análise de dados incertos**. Tese de Doutorado, Faculdade de Engenharia Elétrica e de Computação, UNICAMP, Campinas, 2010.

DIEKS, D. (2009), “Bottom-Up and Top-Down: The Plurality of Explanation and Understanding in Science,” in H. de Regt, K. Eigner & S. Leonelli (eds.). **Scientific Understanding: Philosophical Perspectives**. Pittsburgh: University of Pittsburgh Press.

EASTMAN, J.R. Idrisi 15: The Andes Edition. Worcester, MA: Clark University, 2006.

EULER, L. Solutio problematis ad geometriam situs pertinentis, Comment. **Acad. Sci. Imp. Petropol.** 8 (1736), 128–140.

EISENBERG, D.; MARCOTTE, E. M.; XENARIOS, J.; YEATES, T. O. Protein function in the post-genomic era. **Nature**, 405:823826, 2000.

ESPOSITO, F.; MALERBA, D.; SEMERARO, G. A comparative analysis of methods for pruning decision trees. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 19, n. 5, p. 476-491, May 1997.

FOULDS, L. R. **Graph Theory Applications**. Springer-Verlag, 1992.

GAMA, J. (2004) **Árvores de Decisão, Universidade do Porto**. Disponível em [http://www.liaad.up.pt/~jgama/Aulas\\_ECD/arv.pdf](http://www.liaad.up.pt/~jgama/Aulas_ECD/arv.pdf), acesso em 14/06/14.

GARCIA, S.C. **O Uso de Árvores de Decisão na Descoberta de Conhecimento na Área da Saúde**. UFRGS. Dissertação, 2003.

GOLDBARG, M.; GOLDBARG, E. **Grafos : conceitos, algoritmos e aplicações**. Tradução . Rio de Janeiro: Elsevier, 2012.

GOLDSCHMIDT, R. R.. **Uma Introdução à Inteligência Computacional: Fundamentos, Ferramentas e Aplicações**. 1 ed. Rio de Janeiro: IST-Rio, 2010. v. 1, p. 142.

GOMES, L. Entomologia Forense: Novas tendências e tecnologias nas ciências criminais . Rio de Janeiro. Ed. Technical Books, 2010.

GREENBERG, B. 1971. **Flies and disease**, Vol. 1. Princeton University Press, 856 pp.

GUIMARÃES J. H.; PAPAVERO, N. 1999. **Myiasis in man and animals in the Neotropical region**. Bibliographic database. São Paulo, FAPESP/ Editora Plêiade, 308p.

HAN, J. **Data Mining: Concepts and Techniques**. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2001.

HEIJST, V. G.; SCHREIBER, A. T.; WIELINGA, B. J. Using explicit ontologies in KBS development. **International Journal of Human-Computer Studies** , v. 46, n. 2-3, p. 183-192, feb./mar 1997

JAIN, A. K., MURTY, M. N., FLYNN, P. J. **Data Clustering: A Review**. ACM Computing Surveys, vol. 31, no. 3, pp. 254-323, Sep., 1999.

LEWINSOHN, T. M.; PRADO, P. I.; JORDANO, P.; BASCOMPTE, J.; OLESEN, J. Structure in plant animal interactions assemblages. **Oikos**, 113: 174-184, 2006.

MADEIRA, S. C.; OLIVEIRA, A. L. (2004). **Biclustering algorithms for biological data analysis: A survey**. IEEE/ACM Trans. Comput. Biol. Bioinformatics, 1(1):24–45.

MASON, O.; VERWOERD, M.. Graph theory and networks in Biology. **IET Systems Biology** . 1, 89-119, 2007.

MICHIE D., SPIEGELHALTER D. J., TAYLOR C. C. **Machine learning, Neural and Statistical Classification** . Ellis Horwood, 1994

MITCHELL, T.M. **Machine Learning**, WCB/McGraw-Hill, 1997.

MONTEIRO, R.M. **Microhimenópteros (Insecta: Hymenoptera) parasitóides e insetos predadores de moscas sinantrópicas (Insecta: Diptera) na Granja Capuavinha, Monte-Mor, SP**. Campinas: UNICAMP, 1995. 99p. (Dissertação, Mestrado).

NIE D.; YAN FU; JUNLIN ZHOU; YUKE FANG. Predicting Time Series with Multiple Mixed Models. **J. Comput. Informat. Systems** 7:4 (2011) 1092-1099.

OLIVEIRA, V A; RANGEL, S. 2013. **Teoria dos Grafos: Subgrafos, Operações com Grafos.** Disponível em:

[http://www.ibilce.unesp.br/Home/Departamentos/MatematicaAplicada/socorro4029/aula2\\_operacoesrev2014.pdf](http://www.ibilce.unesp.br/Home/Departamentos/MatematicaAplicada/socorro4029/aula2_operacoesrev2014.pdf). Acesso em ago 2015.

OLIVEIRA-COSTA, J. 2003. **Entomologia Forense: Quando os insetos são vestígios.** Editora Millenium, Campinas, 258p.

PAPE, T.; BLAGODEROV, V.; MOSTOVSKI, M. B. 2011. Order DIPTERA Linnaeus, 1758. In: Zhang, Z. Q. (Ed.) Animal biodiversity: An outline of higherlevel classification and survey of taxonomic richness. **Zootaxa**, 3148: 1-237.

PAINE, R.T. 1966. **Food web complexity and species diversity.** The American Naturalist 100: 65 - 75.

PRESTES, E. **Teoria dos Grafos.** (2012). Disponível em: <http://www.inf.ufrgs.br/~prestes/Courses/Graph%20Theory/GrafosA2.pdf>. Acesso em 10 de jan de 2014.

PUJOL-LUZ, J. R.; ARANTES, L. C.; CONSTANTINO, R. Cem anos da entomologia forense no Brasil (1908-2008). **Rev. Bras. Entomol.**, v.52, p.485-492, 2008.

REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações.** Ed. Manole, São Paulo, 2003.

QUINLAN, J. R. **Induction of Decision Trees**, in Machine Learning, Volume 1 , pages 81-106, 1986.

QUINLAN, J. **Decision trees and multivalued attributes.** Machine Intelligence, 11:305-318, 1988.

QUINLAN, J. R. C4.5: **Programs for machine learning.** San Mateo, CA: Morgan Kaufmann Publishers, 1993.



SERRA H, GODOY WAC, VON ZUBEN FJ, VON ZUBEN CJ, REIS SF (2007.) Sex ratio and dynamic behavior in populations of the exotic blowfly *Chrysomya albiceps* (Diptera, Calliphoridae). **Braz. J. Biol.** 67, 347-353.

SMITH, KGV, 1986. **A manual of forensic entomology**. Cornell Univ. Press, Ithaca, NT.

SOUZA, C.R. **Sazonalidade, sinantropia e preferência por iscas de dípteros necrófagos da região de Rio Claro, SP**. 2011. 60 f. Dissertação - (mestrado) - Universidade Estadual Paulista, Instituto de Biociências de Rio Claro, 2011. Disponível em: <<http://hdl.handle.net/11449/99577>>. Acesso em 10 fev 2013.

SOUZA , C. R.; ZUBEN , C. J. V. 2012. Diversity and synanthropy of Calliphoridae (Diptera) in the region of Rio Claro, SP, Brazil. **Neot. Entomol.** 41 :243-248.

SOUZA , C. R.; ZUBEN , C. J. V. 2016. Synanthropy of Sarcophagidae (Diptera) in southeastern Brazil. **Neot. Entomol.** Doi: 10.1007/s13744-016-0411-0.

TAN, P. N.; STEINBACH, M.; KUMAR,V. 2005. **Introduction to Data Mining** (First Edition). Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

VARGAS, J.; WOOD, D. M. 2010. Calliphoridae, p. 1297–1304. In: B. V. Brown, A. Borkent, J. M. Cumming, D. M. Wood, N. E. Woodley & M. A. Zumbado (eds.). **Manual of Central American Diptera**. Vol. 2. Canada, Ontario, NCR Research Press, 728 p.

VON ZUBEN, F. J.; ATTUX, R. R. F. **Árvores de Decisão**, Tópico 7 das Notas de Aula do Curso de Pós-Graduação IA004 - Redes Neurais II, FEEC/Unicamp, 2010. Disponível em [ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia004\\_1s10/notas\\_de\\_aula/topico7\\_IA004\\_1s10.pdf](ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia004_1s10/notas_de_aula/topico7_IA004_1s10.pdf), Acesso em 10 de setembro de 2012.

WITTEN, I.; FRANK, E. **Practical machine learning tools and techniques with Java implementations**. San Francisco, CA: Morgan Kaufmann, 1999.

## OBJETIVOS

---

Os objetivos deste trabalho foram:

- Através dos ferramentais matemáticos utilizados, *biclusters* e árvores de decisão, buscar compreender melhor um padrão de distribuição das espécies de dípteros necrófagos em três locais diferentes de coleta;
- Buscar um padrão de distribuição das espécies em relação às três diferentes iscas utilizadas;
- Avaliar possíveis relações ecológicas entre as espécies coletadas, e tentar interpretá-las biologicamente.

## DADOS UTILIZADOS

---

Para o presente trabalho, foram utilizados os dados referentes à dissertação de Mestrado em Ciências Biológicas (Zoologia) pela Universidade Estadual “Júlio Mesquita Filho” intitulada “Sazonalidade, sinantropia e preferência por iscas de dípteros necrófagos da região de Rio Claro, SP”. Eles foram base para a aplicação das ferramentas matemáticas nesta Tese de Doutorado.

Neste trabalho foram coletados 2.782 exemplares de dípteros (Tabela 1), sendo 1.253 espécimes (45,04%) pertencentes à família Calliphoridae (SOUZA; VON ZUBEN, 2012), seguida por Muscidae com 1.089 (39,14%) e Sarcophagidae (SOUZA; VON ZUBEN, 2016), com 440 exemplares (15,82%), somando um total de 34 espécies identificadas e alguns morfotipos determinados (SOUZA, 2011)

O estudo foi efetuado em três localidades distintas do município de Rio Claro, SP: na área urbana, na área rural e na área de floresta, no intuito de se determinar o índice de sinantropia (NUORTEVA, 1963) das principais espécies.

Foram utilizadas como iscas: carne bovina moída, fígado bovino e sardinha. Em cada armadilha, foram colocados 100 gramas de uma determinada isca. Em todos os locais de coleta foram colocadas três armadilhas que continham um dos tipos de iscas utilizados. As armadilhas feitas com garrafas plásticas tipo PET de dois litros foram penduradas em árvores a uma distância aproximada de 1,5 metros do chão e distantes igualmente umas das outras (aproximadamente 13 metros).

As coletas foram feitas mensalmente, entre os meses de setembro de 2009 e agosto de 2010. As armadilhas foram expostas durante cinco dias consecutivos (sem a troca das iscas a cada 24 horas) e as coletas foram realizadas a cada 24 horas, a partir do segundo dia, através da retirada da organza com os insetos atraídos.

A identificação dos espécimes foi realizada com auxílio de estereoscópio Zeiss Stemi modelo SV11, utilizando chaves dicotômicas, segundo Carvalho; Ribeiro (2000) e Carvalho; Mello-Patiu (2008) e também contou com a colaboração dos seguintes pesquisadores: Profa. Dr. Patrícia J. Thyssen da Universidade Federal de Pelotas - RS, Prof. Dr. Júlio Mendes da Universidade Federal de Uberlândia-MG e Sr. Carlos Alexandre Alves da Universidade Estadual de Campinas-SP.

Tabela 1 - Espécies de três famílias de Diptera, coletadas em três ambientes diferentes utilizando três tipos de iscas, na cidade de Rio Claro-SP.

(continua)

FAMÍLIA	Espécies	URBANA												RURAL												FLORESTA											
		FÍGADO				CARNE				SARDINHA				FÍGADO				CARNE				SARDINHA				FÍGADO				CARNE				SARDINHA			
		P	V	O	I	P	V	O	I	P	V	O	I	P	V	O	I	P	V	O	I	P	V	O	I	P	V	O	I	P	V	O	I	P	V	O	I
MUSCIDAE	<i>A. orientalis</i>	7	7	0	0	191	29	5	0	131	1	0	0	4	1	0	0	310	15	0	4	11	0	2	0	0	0	0	0	134	3	0	0	39	0	0	0
	<i>S.nudiseta</i>	1	0	0	0	12	2	3	0	12	0	3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	<i>M. domestica</i>	0	1	0	0	9	7	10	0	4	1	0	0	5	24	0	0	33	6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>O. chalcogaster</i>	0	0	0	0	3	7	11	0	1	2	1	0	0	9	0	0	12	5	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	<i>Muscina</i> sp	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>Ophyra</i> sp	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CALLIPHORIDAE	<i>L. eximia</i>	55	54	18	0	13	6	7	0	4	2	4	0	39	5	5	2	19	1	0	3	0	0	0	3	336	40	35	58	11	40	4	0	8	13	6	3
	<i>H. segmentaria</i>	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	10	0	0	45	1	4	0	6	2	0	0	23	0	2	0
	<i>L. cuprina</i>	2	9	0	0	1	0	0	0	1	3	0	0	3	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>H. semidiphana</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	11	2	0	0	1	0	0	0
	<i>C. megacephala</i>	0	17	0	0	8	1	0	0	10	1	0	0	0	15	0	0	1	2	0	0	0	0	0	0	0	27	0	0	33	9	0	0	5	5	0	0
	<i>C. albiceps</i>	0	0	0	0	2	0	0	0	25	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	4	9	2	0	36	27	0	0	30	9	0	0
	<i>C. idioidea</i>	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	2	0	0	0	1	0	0
	<i>C. macellaria</i>	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	<i>M. peregrina</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	5	4	0	0	1	0	0	0
SARCOPHAGIDAE	<i>P. (S) lambens</i>	0	0	0	0	8	0	0	0	17	4	6	0	1	0	0	0	2	0	0	0	3	1	2	0	7	0	0	0	15	0	0	1	9	0	2	2
	<i>Oxysarcodexia</i> sp	3	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	6	1	0	0	3	3	0	0	7	3	2	3
	<i>O. thornax</i>	8	2	0	0	5	1	0	0	9	5	5	0	2	0	0	0	4	0	0	0	16	6	0	1	5	2	0	0	3	0	0	0	3	0	1	1
	<i>O. cumiforcipes</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
	<i>P (Euboetcheria)</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	4	0	1	0	12	7	1	0	11	4	1	4
	<i>P (E) collusor</i>	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	1	0	0	19	3	0	1	14	11	0	0
	<i>O. paulistanensis</i>	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0



**REFERÊNCIAS**

---

CARVALHO, C. J. B.; MELLO-PATIU, C.A. Key to adults of the most common forensic species of Diptera in South America. **Rev. Bras. Entomol.** v.52, n.3, p.390- 406, 2008

CARVALHO, C. J. B.; RIBEIRO, P. B. Chave de identificação das espécies de Calliphoridae (Diptera) do sul do Brasil. **Rev. Bras. Parasitol.** 9: 255-268, 2000.

NUORTEVA, P. Synanthropy of blowflies (Dipt., Calliphoridae) in Finland. **An. Entomol. Fennici.** 29: 1-49, 1963.

SOUZA, C.R. **Sazonalidade, sinantropia e preferência por iscas de dípteros necrófagos da região de Rio Claro, SP.** 2011. 60 f. Dissertação - (mestrado) - Universidade Estadual Paulista, Instituto de Biociências de Rio Claro, 2011. Disponível em: <<http://hdl.handle.net/11449/99577>>. Acesso em 10 fev 2013.

SOUZA , C. R.; ZUBEN , C. J. V. 2012. Diversity and synanthropy of Calliphoridae (Diptera) in the region of Rio Claro, SP, Brazil. **Neot.I Entomol.** 41 :243-248.

SOUZA , C. R.; ZUBEN , C. J. V. 2016. Synanthropy of Sarcophagidae (Diptera) in southeastern Brazil. **Neot. Entomol.** Doi: 10.1007/s13744-016-0411-0.

## CAPÍTULO I

---

### Aplicação de *biclusters* para modelar a ocorrência de dípteros necrófagos, no sudeste do Brasil

CR DE SOUZA<sup>1</sup>, FJV ZUBEN<sup>2</sup> & CJV ZUBEN<sup>1</sup>

<sup>1</sup>Univ Estadual Paulista, Instituto de Biociências, Depto de Zoologia, Rio Claro, SP, Brasil.; <sup>2</sup> Univ Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação, Depto de Engenharia da Computação e Automação Industrial, Campinas, SP, Brasil.

#### RESUMO

Um dos grandes problemas a serem resolvidos na área de entomologia é encontrar padrões oriundos nas bases de dados já conhecidas. Buscando-se soluções, utilizou-se nesse trabalho a aplicação de uma ferramenta matemática chamada de *biclusters* para modelar a ocorrência de dípteros necrófagos coletadas em três locais distintos (área rural, urbana e de mata), através de três tipos de iscas (sardinha, fígado e carne moída), no período de um ano, na cidade de Rio Claro-SP.. Para as análises efetuadas no presente trabalho, foram utilizados dois algoritmos: In-Close2 (ANDREWS, 2011) para dados binários e RInClose (VERONEZE et al., 2014) para dados inteiros, buscando-se encontrar associações entre as espécies. Uma bastante recorrente foi a entre os muscídeos *Atherigona orientalis* (Schiner) e *Musca domestica* Linnaeus, talvez por serem ambas consideradas eusinantrópicas. Além disso, *Lucilia eximia* (Wiedemann) e *Chrysomya megacephala* (Fabricius) foram encontradas associadas no ambiente de mata, mesmo sendo espécies freqüentadoras de ambientes antrópicos. Isso pode ter ocorrido devido à modificações encontradas nesse local que o caracterizam com um certo grau de ocupação humana. Também foram demonstradas outras associações inclusive entre as espécies do gênero *Peckia* ocorrendo com maior freqüência na área de mata, como entre *Peckia (Squamatodes) ingens* (Walker) e *Peckia (Euboettcheria) anguilla* (Curran & Walley., sendo que há relatos da presença das espécies aqui mencionadas no ambiente de floresta. A partir disso, conclui-se que ao se analisar os resultados obtidos através dos *biclusters*, nota-se a relevância dessa ferramenta na busca por relações entre as espécies aqui consideradas.

## INTRODUÇÃO

Nos últimos tempos, nota-se um aumento na capacidade de coletar, gerar e armazenar dados em formato digital, devido ao avanço nas áreas de Engenharia de Computação. E para que esses dados possam dar origem a conhecimentos e informações úteis, eles precisam ser processados. Nesse contexto, tem-se a Mineração de Dados (MD) que é uma das etapas do processo de extração de conhecimento de bases de dados (KDD) que busca extrair conhecimento relevante (i.e., padrões e tendências) embutido em grandes bases de dados, através da aplicação de algoritmos de classificação, agrupamento, sumarização, regressão, dentre outros (FAYYAD et al., 1996).

Um dos exemplos de uma técnica de Mineração de Dados multivariados que é capaz de explicitar correlações simultâneas entre objetos e atributos é a Técnica de Agrupamento de Dados ou Clusterização (JAIN et al., 1999). O conceito de clusterização está ligado a toda operação que divide dados em subconjuntos (*clusters*), seguindo algum critério de similaridade entre eles. De uma forma geral, obter a solução para um problema de clusterização corresponde ao processo de agrupar os elementos (objetos) de uma base de dados (conjunto) de tal forma que os grupos formados, ou *clusters*, representem uma configuração em que cada elemento possua uma maior similaridade com qualquer elemento do mesmo *cluster* do que com elementos de outros *clusters* (ANKERST et al., 1999, HRUSCHKA; EBECKEN, 2001; CARLANTONIO, 2001). Na Literatura, a análise de *clusters* pode ser chamada também de *Clustering*, *Q-analysis*, *Typology*, *Classification Analysis* ou *Numerical Taxonomy*.

Apesar de essa técnica possuir um histórico de casos de sucesso em suas aplicações (JAIN et al., 1999; BERKHIN, 2006), ela possui algumas limitações. Esta técnica não agrupa objetos e atributos simultaneamente e também não aponta os objetos que são semelhantes a outros com base apenas em um subconjunto de atributos. Em aplicações reais, amostras de dados são, comumente, correlatas a vários subconjuntos de amostras distintas, quando são considerados subconjuntos distintos de atributos (DE FRANÇA; VON ZUBEN 2009). Uma técnica que possui tal característica, aplicando os critérios de similaridade às linhas e às colunas das matrizes de dados, é denominada biclusterização (MADEIRA; OLIVEIRA, 2004; DE FRANÇA et al., 2006).

Considerando que os conjuntos de dados são estruturados como uma matriz de objetos (linhas) e atributos correspondentes (colunas), a biclusterização consegue buscar vários subconjuntos de linhas e colunas da matriz de dados sendo que cada um desses subconjuntos



(*biclusters*) possui objetos que compartilham alguma similaridade, definida pela aplicação com os atributos especificamente selecionados. Sendo assim, um único objeto pode pertencer a vários *biclusters* desde que associados a um subconjunto distinto de atributos, permitindo a extração de informações adicionais do conjunto de dados (COELHO et al., 2009).

O termo *biclusterização* foi introduzido por MIRKIN (1996) para descrever a clusterização simultânea dos conjuntos de linhas e colunas de uma matriz de dados, mas seu primeiro uso ocorreu no contexto da biologia computacional com o trabalho de CHENG; CHURCH (2000). A partir desse momento, algoritmos têm sido propostos e utilizados em vários campos de aplicação (ABDULLAH; HUSSAIN, 2006; BISSON; HUSSAIN, 2008; FAURE et al., 2007). Os termos co-clusterização, clusterização bidimensional, agrupamento duplo de dados, *simultaneous clustering*, *co-clustering*, *two-way clustering*, *block clustering*, *bidimensional clustering*, são comumente citados na literatura para se designar biclusterização. Em síntese, este conceito denota a existência de submatrizes “significativas” em uma dada matriz, ou seja, um subconjunto de linhas e de colunas que possuem um certo padrão de homogeneidade (PINHEIRO et al., 2012).

A técnica de biclusterização tem sido aplicada de diferentes formas, mas a que se destaca pela intensa aplicação é o estudo de expressão gênica. Ela busca encontrar como os genes são regulados sob certas condições, utilizando dados de expressão gênica por microarrays (BEN-DOR et al., 2002; BLEULER et al., 2004; CHENG; CHURCH, 2000; DAS; BORAH, 2014; DIVINA; AGUILAR-RUIZ, 2006; HANCZAR; NADIF, 2011; IHMELS et al., 2004; MITRA; BANKA, 2006; MURALI; KASIF, 2003; NEPOMUCENO et al., 2010; PRELIĆ et al., 2006; XIE et al., 2007; YAN; WANG, 2013;).

Além disso, ela também é utilizada para prever os fenótipos de tumor de seus genótipos (DAO et al., 2010), para a classificação de documentos, agrupando os documentos e palavras relacionadas simultaneamente (BISSON; HUSSAIN, 2008), na recuperação de informação e mineração de texto (DE CASTRO et al., 2007b; FELDMAN; SANGER, 2006), na análise de dados eleitorais (HARTIGAN, 1972) e também na filtragem colaborativa (DE CASTRO *et al.*, 2007a; DE CASTRO et al., 2007c; SYMEONIDIS et al., 2007), além de muitas outras áreas.

Trabalhar com a técnica de biclusterização ou coclusterização permite organizar informações, extrair conhecimentos e mostrar as tendências presentes em grandes massas de dados. Ou seja, adquirir o domínio desta metodologia representa um passo significativo no processo de incorporação de inteligência em análise de dados, que é um dos principais objetivos das técnicas de inteligência artificial.

De forma inovadora, aplicou-se, no presente trabalho, os *biclusters* no contexto dos dípteros necrófagos, buscando-se encontrar padrões de associações entre as espécies de estudadas aqui.

### GRAFOS E *BICLUSTERS*

Considere uma matriz de dados  $X$ ,  $n \times m$ , definida por um conjunto de linhas  $L = \{l_1, l_2, \dots, l_n\}$  e o conjunto de colunas é dado por  $C = \{c_1, c_2, \dots, c_m\}$ . Cada elemento  $x_{ij}$  dessa matriz é um valor que representa a relação existente entre a linha  $i$  e a coluna  $j$  (Figura 13). Na realidade, esta matriz pode representar diferentes tipos de dados numéricos, como objetos e atributos (compreendendo às linhas e colunas da matriz) ou um mapa de cores de uma imagem (DE FRANÇA et al., 2006).

Figura 13 - Matriz X de Dados

	$c_1$	...	$c_j$	...	$c_m$
$l_1$	$x_{11}$	...	$x_{1j}$	...	$x_{1m}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$l_i$	$x_{i1}$	...	$x_{ij}$	...	$x_{im}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$l_n$	$x_{n1}$	...	$x_{nj}$	...	$x_{nm}$

Será utilizado  $X(L, C)$  para denotar a matriz  $X$ . Considerando que  $I \subseteq L$  e  $J \subseteq C$  são, respectivamente, subconjuntos de linhas e colunas,  $X(I, J)$  denota a submatriz de  $X$  com o conjunto de linhas  $I$  e o conjunto de colunas  $J$ , em que  $I = \{i_1, i_2, \dots, i_k\}$  é um subconjunto de linhas ( $I \subseteq L$  e  $k \leq M$ ) e  $J = \{j_1, j_2, \dots, j_s\}$  é um subconjunto de colunas ( $J \subseteq C$  e  $s \leq N$ ).

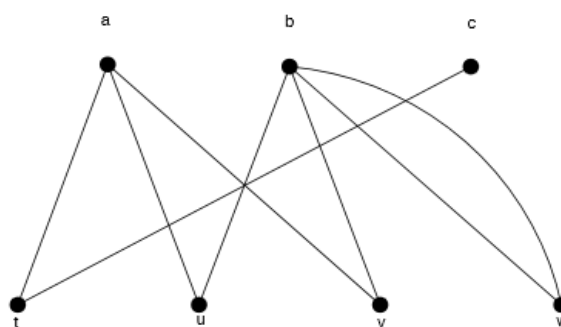
De acordo com Madeira; Oliveira (2004), devido à similaridade entre eles, uma matriz de dados pode ser vista como um **grafo bipartido com pesos**, considerando as linhas e colunas como vértices, no qual cada conjunto de vértices (linhas e colunas) pertence a bipartições diferentes. Um grafo bipartido é um multigrafo  $G = (V, E)$ , cujo conjunto dos  $v$  vértices admite uma partição  $V = X \cup Y$  em dois subconjuntos  $X$  e  $Y$ , de tal forma que cada

aresta de  $G$  tenha uma extremidade em  $X$  e outra em  $Y$ . O par  $X, Y$  representa uma bipartição de  $G$ . Assim,  $G$  é bipartido com bipartição  $X, Y$  se e somente se

$$\forall \{ a, b \} \in E \quad a \in X \Leftrightarrow b \in Y.$$

Pode-se ter como exemplo o multigrafo representado abaixo (Figura 14), que é bipartido, com bipartição  $X = \{ a, b, c \}$  e  $Y = \{ t, u, v, w \}$ .

Figura 14 – Multigrafo bipartido



Fonte: Prestes (2012)

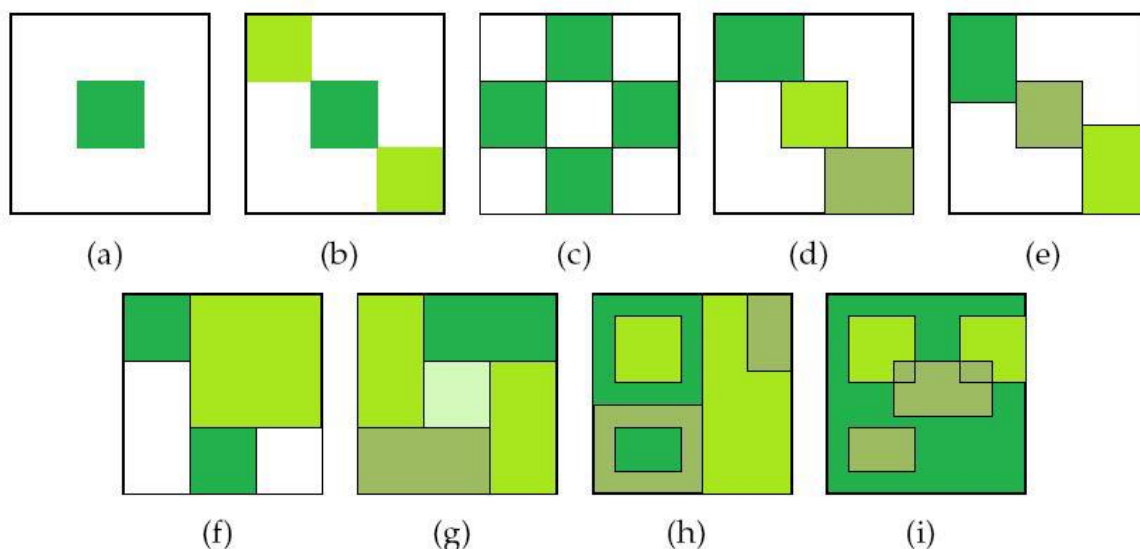
A matriz de dados  $X(L, C)$  pode ser vista como um **grafo bipartido ponderado**, em que cada nó  $n_i \in L$  corresponde a uma linha e cada nó  $n_j \in R$  corresponde a uma coluna. A aresta entre os nós  $n_i$  e  $n_j$  possui um peso  $a_{ij}$ , denotando o elemento da matriz na intersecção entre a linha  $i$  e a coluna  $j$ . Dessa forma, nota-se que encontrar um *bicluster* de tamanho máximo é o mesmo que encontrar um **grafo bipartido completo (biclique)**, em que todos os nós  $n_i \in L$  são conectados a todos os nós  $n_j \in R$  (CHENG; CHURCH, 2000; MADEIRA; OLIVEIRA, 2004; TANAY et al., 2002), um problema conhecido como NP-completo (PEETERS, 2003). A presença de bicliques indicaria um alto grau de similaridade entre os dados agrupados e nesse caso, o grafo pode ser considerado como um grafo biclusterizado.

### ESTRUTURA DOS *BICLUSTERS* E SEUS DIFERENTES TIPOS

Atualmente, encontram-se muitos algoritmos para a resolução do problema da biclusterização. Eles podem assumir dois tipos de pesquisa: buscam apenas um único *bicluster* na matriz de dados (Figura 15a) como por exemplo o proposto por Cheng; Church,

(2000), ou procuram por múltiplos *biclusters* (Figuras 15.b, 15.c, 15.d, 15.e, 15.f, 15.g, 15.h, 15.i), em que o número de *biclusters* que se pretende encontrar pode ou não ser predefinido. A maioria dos algoritmos segue o último critério e têm-se como exemplos os seguintes trabalhos: de Hartigan (1972) e Santamaria (2009). Além disso, os algoritmos se diferenciam na forma como identificam cada *bicluster* e no tipo de *biclusters* que eles podem buscar. Para avaliá-los, é necessário saber a necessidade do usuário e levar em consideração os itens acima citados (MADEIRA; OLIVEIRA, 2004).

Figura 15: Estruturas dos *biclusters*. (a) *bicluster* único; (b) *biclusters* com linhas e colunas exclusivas; (c) estrutura de xadrez; (d) *biclusters* com linhas exclusivas; (e) *biclusters* com colunas exclusivas; (f) *biclusters* não-sobrepostos com estrutura de árvore; (g) *biclusters* não-sobrepostos não-exclusivos; (h) *biclusters* sobrepostos com estrutura hierárquica; e (i) *biclusters* sobrepostos arbitrariamente posicionados



Fonte: Madeira; Oliveira (2004).

É interessante ressaltar que um *bicluster* é formado por um subconjunto de objetos e um subconjunto de atributos, os quais apresentam similaridade alta entre si e não precisam ser contíguos na matriz original. Na Figura 16, tem-se um exemplo de um *bicluster* (fictício) obtido a partir de uma matriz de dados, demonstrando que ele não é contíguo na matriz original.

Em relação ao tipo de *biclusters*, Madeira; Oliveira, 2004 os dividiu em cinco tipos diferentes:

1. *Biclusters* com Valores Constantes (BVC);

2. *Biclusters* com Valores Constantes nas Linhas (BVCL);
3. *Biclusters* com Valores Constantes nas Colunas (BVCC);
4. *Biclusters* com Valores Coerentes (BVCo), do tipo aditivo ou multiplicativo; e
5. *Biclusters* com Evoluções Coerentes (BEC).

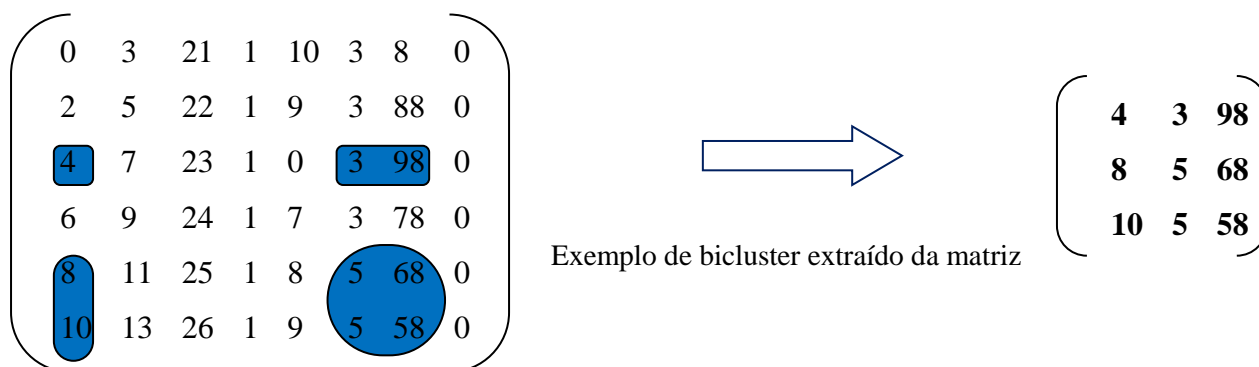
Os *biclusters*, BVC, BVCL, BVCC e BVCo, ilustrados nas Figuras 15.a a 15.d, respectivamente, mostram subconjuntos de linhas e colunas com comportamentos similares. O mais geral entre eles é o BVCo, sendo que os outros são apenas casos especiais deste. O quinto tipo, BEC, ilustrado na Figura 15.e, apresenta comportamento coerente, não buscando valores coerentes dentro dos *biclusters*, mas sim propriedades coerentes nos valores (DE FRANÇA, 2010).

Há também um caso especial de *bicluster* do tipo BEC (Figura 15.e), conhecido como OPSM (*Order-preserving* sub-matriz) que foi definido por Ben-Dor et al. (2002). De acordo com os autores, o *Bicluster* OPSM é uma submatriz  $(I, J)$  de uma matriz de dados  $An \times m$  de tal modo que existe uma permutação  $P = \{p_1, p_2, \dots, p_s\}$  das colunas do conjunto  $J$ , em que  $a_{ip1} \leq a_{ip2} \leq \dots \leq a_{ips}, \forall i \in I$ . Em outras palavras, ele é definido como uma submatriz que preserva o ordem das colunas selecionadas para todas as linhas também selecionadas. Só será um OPSM se houver uma permutação de uas colunas em que a sequência de valores em cada linha seja estritamente crescente.

Os algoritmos mais simples identificam *biclusters* formados por valores constantes (BVC) e podem-se citar como exemplos os algoritmos de Cho et al. (2004), Hartigan (1972) e Tibshirani et al. (1999). Um outro tipo considera *biclusters* em que o valor de cada linha ou coluna é sempre o mesmo (BVCL/BVCC), o que é o caso dos seguintes trabalhos : Califano et al (2000), Getz, et al. (2000), Segal, et al. (2001), Sheng et al. (2003) e Yang; Wang (2002).

Nos estudos de Cheng; Church (2000), Cho (2004), Getz ,et al (2000), Hartigan (1972), Kluger et al. (2003) e Lazzeroni; Owen (2000), encontram-se algoritmos que buscam *biclusters* com valores coerentes BVCo, usando um modelo aditivo ou multiplicativo. Finalmente, podem-se citar os *biclusters* com base em uma evolução coerente de seus valores (BEC). Os algoritmos que buscam esse tipo são visualizados nos trabalhos de Ben-Dor et al. (2002), Liu; Wang (2003) e Murali; Kasif (2003). A Figura 17 sumariza os diferentes tipos de *biclusters*.

Figura 16 - Exemplo de um bicluster formado por um subconjunto de objetos e um subconjunto de atributos não-contíguos na matriz de dados original.



É necessário ressaltar que, dependendo da matriz de dados utilizada, não é possível encontrar *biclusters* perfeitamente constantes ou coerentes, e assim, é necessário admitir pequenas variações ou ruídos em seus valores (DE FRANÇA, 2010).

#### DIFERENTES PARÂMETROS NA BUSCA DE *BICLUSTERS*

Outro ponto de variação nos algoritmos é a forma com que eles buscam os *biclusters*, já que nem todos seguem os mesmos parâmetros. Madeira; Oliveira (2004) os dividem em cinco tipos diferentes:

a) Combinação iterativa de *clustering* em linhas e colunas (*iterative row and column clustering combination*): consiste em aplicar os métodos usuais de *clustering* nas linhas e nas colunas da matriz, combinando posteriormente os resultados, de modo a obter os *biclusters* finais, através de um processo iterativo. Foi utilizada nos seguintes trabalhos: Getz et al. (2000) e Tang et al.(2001).

b) Divisão e Conquista (*divide and conquer*): segue a abordagem de “dividir para conquistar”. A instância dada do problema é dividida em duas ou mais instâncias menores e cada instância menor é resolvida. No final, as soluções das instâncias menores são combinadas para produzir uma solução da instância original. Estes algoritmos possuem a vantagem de serem muito rápidos, mas podem perder bons *biclusters* devido às divisões do problema original. O primeiro algoritmo deste tipo foi o “Block Clustering”, introduzido por

Hartigan (1972) e desenvolvido posteriormente por Duffy; Quiroz (1991) e Tibshirani et al. (1999).

c) Pesquisa “Gulosa” (*greedy search*): segue uma heurística gananciosa em que se criam os *biclusters* através da adição/remoção de colunas/linhas dos mesmos, com base em um critério de maximização do ganho local. O primeiro trabalho em que se utilizou esse tipo de parâmetro de busca foi o de Cheng e Church (2000).

d) Enumeração exaustiva de *biclusters* (*exhaustive bicluster enumeration*): procura exaustivamente todos os *biclusters*. Fazem uma procura exaustiva de todos os *biclusters* mediante um conjunto de restrições, como o tamanho e a homogeneidade dos *biclusters*. Os algoritmos demonstrados nos trabalhos de Tanay et al. (2002) e Liu; Wang (2003) constituem exemplos de algoritmo de procura exaustiva.

e) Identificação de parâmetros de uma dada distribuição (*distribution parameter identification*): assume que os *biclusters* são gerados segundo um determinado modelo estatístico e procura identificar os parâmetros da distribuição que aproximam da melhor forma a distribuição observada nos dados reais, minimizando iterativamente um determinado critério de qualidade.

Além disso, os algoritmos podem ser divididos em determinísticos e não-determinísticos. Os algoritmos não-determinísticos podem encontrar diferentes soluções a cada execução, enquanto os determinísticos sempre encontram a mesma solução (DE FRANÇA et al., 2006).

Figura 17. Exemplos de tipos diferentes de *biclusters*. (a) BVC. (b) BVCL. (c) BVCC. (d) BVCo, aditivo. (e) BVCo, multiplicativo. (f) BEC.

1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1

(a)

1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4

(b)

3	4	5	6
3	4	5	6
3	4	5	6
3	4	5	6

(c)

1	2	5	0
2	3	6	1
4	5	8	3
5	6	9	4

(d)

1	4	2	6
2	8	4	12
3	12	6	18
4	12	8	24

(e)

57	23	34	9
93	47	56	20
39	20	27	11
80	30	47	12

(f)

## MATERIAIS E MÉTODOS

### ALGORITMOS UTILIZADOS

Para as análises efetuadas no presente trabalho, foram utilizados dois algoritmos: In-Close2 (ANDREWS, 2011) e RInClose (VERONEZE et al., 2014). A seguir, têm-se mais informações sobre cada um.

#### a) In-Close2

As tabelas são comumente usadas para representar dados de incidência na forma de relações binárias. Uma das formas de representar os dados é através da Análise Formal de Conceitos- AFC (*Formal Concept Analysis*), em que as colunas de uma tabela representam atributos ou itens, linhas representam objetos, instâncias ou transações e as interseções entre linhas e colunas representam a relação de incidência existente entre ambas. Nesse caso, a presença do valor 1 (verdadeiro) na linha  $i$ , coluna  $j$ , diz que objeto possui o atributo  $j$ . Uma presença do valor 0 (falso) na linha  $i$ , coluna  $j$ , diz que o objeto não tem o atributo  $j$  (ANDREWS, 2009).

A emergência da Análise Formal de Conceitos (AFC) como uma tecnologia de análise de dados (ANDREWS; ORPHANIDES, 2010) aumentou a necessidade de que os algoritmos calculem os conceitos formais de forma rápida. Um conceito formal é determinado por sua *extensão* e *intensão*. A *extensão* refere-se ao conjunto de objetos que são instâncias do conceito, enquanto a *intensão*, ao conjunto de características comuns a todos os objetos (SOWA, 1984). Matematicamente, estas entidades são pares ordenados  $(A; B)$  em que  $A$  e  $B$  são respectivamente subconjuntos do conjunto de objetos e do conjunto de atributos. Em relação aos dados binários, com *biclusters* sendo definidos com valor constante, o problema de enumeração é equivalente ao problema de listar todos os cliques bipartidos maximais em um gráfico bipartido sobre as linhas e colunas, para as quais existem algoritmos eficientes para enumeração (VERONEZE et al., 2014).

Nesse contexto, pode-se citar um algoritmo baseado na AFC denominado de In-Close2 (ANDREWS, 2011). Considerando também seu precursor In-Close, ambos foram baseados em outro algoritmo conhecido como Close-By-One (KUZNETSOV, 1999). O In-Close2 utiliza uma abordagem lexicográfica para mineração de conceitos formais evitando as despesas gerais de computar dados repetidos (VERONEZE et al., 2014). De acordo com



Freitas (2004), a abordagem lexicográfica é conceitualmente simples, fácil de usar, de implementar e de estender para objetivos adicionais. A idéia básica desta abordagem não é atribuir pesos, mas sim diferentes prioridades a objetivos distintos, e então focar na otimização dos objetivos de acordo com as suas respectivas prioridades.

Os dados utilizados neste trabalho foram convertidos em dados binários, identificando ausência ou presença das espécies e o algoritmo In-Close2 foi utilizado para gerar os *biclusters*.

## **b) RInClose**

O algoritmo In-Close2 foi desenvolvido especificamente para extrair todos os *biclusters* constantes maximais de uma matriz de dados binários. A partir dele, foi proposta uma nova família de algoritmos denominada de RIn-Close. Ambos operam basicamente da mesma forma, mas com a seguinte diferença: no In-Close2, cada *bicluster* pode gerar apenas um descendente por atributo e no RIn-Close, cada *bicluster* pode gerar vários descendentes por atributo (possivelmente com sobreposição entre suas extensões). Quando não houver sobreposição, as adaptações são simples. Caso contrário, o teste da canonicidade não é suficiente para evitar *biclusters* redundantes (VERONEZE et al., 2014).

Essa nova família proposta de algoritmos é a primeira em atingir uma performance completa, correta e não redundante na obtenção de todos os *biclusters* máximos a partir de uma matriz numérica (VERONEZE et al., 2014). Por completo, queremos dizer que ela busca todos os *biclusters* máximos presentes no conjunto de dados; por correta, esperamos que ela busque apenas *biclusters* que atendem o padrão de coerência informado; e por não redundante, queremos dizer que ela não busca o mesmo *bicluster* mais de uma vez (OLIVEIRA, 2015).

## **ANÁLISES**

As análises foram feitas de algumas formas, a partir das tabelas de dados. Na **Análise 1**, os dados foram considerados como binários e posteriormente como números inteiros. No primeiro caso, considerou-se o zero como a ausência de uma determinada espécie numa dada circunstância e o 1 como a presença. As demais análises (**2 e 3**) foram feitas apenas considerando os dados como números inteiros.

## Análise 1

### a) Análise com dados binários

Em relação aos dados binários, buscaram-se tanto *biclusters* com valores constantes iguais a 0 quanto *biclusters* com valores iguais a 1. Ou seja, a partir dessa ferramenta, pôde-se perceber a relação entre as espécies no que se refere à ausência destas ou à presença.

Para a essa análise, as espécies foram divididas em três tabelas distintas, cada uma referente a uma família (Tabela 2 - Muscidae, Tabela 3 - Calliphoridae, Tabela 4 – Sarcophagidae – convertidas para 0 e 1). Vale ressaltar que nessas tabelas há números (1-9) tanto nas linhas quanto nas colunas que facilitam a interpretação dos *biclusters*, cujos resultados serão gerados considerando esses números. Foram utilizados aqui apenas os três tipos de isca e o ambiente em que as espécies foram coletadas.

Tabela 2 – Ocorrência das espécies da família Muscidae, atraídas por três tipos de isca nos três ambientes de coleta. Os números (1-9), presentes tanto nas linhas quanto nas colunas, são úteis para a geração dos *biclusters* e posteriormente sua interpretação.

Espécies Muscidae	URBANA			RURAL			FLORESTA		
	FÍGADO	CARNE	SARDINHA	FÍGADO	CARNE	SARDINHA	FÍGADO	CARNE	SARDINHA
	1	2	3	4	5	6	7	8	9
<i>A. orientalis</i>	14	225	132	5	329	13	0	137	39
<i>S.nudiseta</i>	1	17	15	0	0	1	0	0	0
<i>M. domestica</i>	1	26	5	29	39	1	0	0	0
<i>O. chalcogaster</i>	0	21	4	9	17	1	0	1	0
<i>Muscina sp</i>	0	1	0	0	0	0	0	0	0
<i>Ophyra sp</i>	0	1	0	0	0	0	0	0	0

Tabela 3 – Ocorrência das espécies da família Calliphoridae, atraídas por três tipos de isca nos três ambientes de coleta. Os números (1-9), presentes tanto nas linhas quanto nas colunas, são úteis para a geração dos *biclusters* e posteriormente sua interpretação.

Espécies - Calliphoridae	URBANA			RURAL			FLORESTA		
	FÍGADO	CARNE	SARDINHA	FÍGADO	CARNE	SARDINHA	FÍGADO	CARNE	SARDINHA
	1	2	3	4	5	6	7	8	9
<i>L. eximia</i>	127	26	10	51	23	3	469	55	30
<i>H. segmentaria</i>	1	0	0	0	1	10	50	8	25
<i>L. cuprina</i>	11	1	4	3	4	0	0	0	0
<i>H. semidiphana</i>	0	0	0	0	0	0	26	13	1
<i>C. megacephala</i>	17	9	11	15	3	0	27	42	10
<i>C. albiceps</i>	0	2	25	0	1	0	15	63	39
<i>C. idioidea</i>	0	1	0	0	1	0	1	2	1
<i>C. macellaria</i>	0	0	0	1	0	0	1	0	0
<i>M. peregrina</i>	0	0	0	0	0	0	4	9	1

Tabela 4 – Ocorrência das espécies da família Sarcophagidae, atraídas por três tipos de isca nos três ambientes de coleta. Os números (1-9), presentes tanto nas linhas quanto nas colunas, são úteis para a geração dos *biclusters* e posteriormente sua interpretação.

Espécies- Sarcophagidae	URBANA			RURAL			FLORESTA			
	FÍGADO	CARNE	SARDINHA	FÍGADO	CARNE	SARDINHA	FÍGADO	CARNE	SARDINHA	
	1	2	3	4	5	6	7	8	9	
<i>S. lambens</i>	1	0	8	27	1	2	6	7	16	13
<i>Oxysarcodexia sp</i>	2	3	0	2	0	1	0	7	6	15
<i>O. thornax</i>	3	10	6	19	2	4	23	7	3	5
<i>O. cumiforcipes</i>	4	0	0	0	0	0	1	0	0	2
<i>P (Euboetcheria)</i>	5	0	0	0	0	0	1	5	20	20
<i>P (E) collusor</i>	6	0	2	1	0	0	0	5	23	25
<i>O. paulistanensis</i>	7	2	1	0	0	1	0	1	0	0
<i>O. avuncula</i>	8	0	0	0	0	0	1	3	1	5
<i>H. aurescens</i>	9	0	0	0	0	0	0	1	0	0
<i>P (S) ingens</i>	10	0	1	0	0	2	0	1	7	5
<i>P (E) anguilla</i>	11	0	1	2	0	1	1	1	11	9
<i>P (P) intermutans</i>	12	0	0	0	0	0	0	23	10	8
<i>P (E) australis</i>	13	0	0	0	0	0	0	0	2	7
<i>R. belforti</i>	14	0	0	2	1	0	2	0	0	0
<i>P (S) florencioi</i>	15	0	0	0	0	0	0	3	5	5
<i>P (P) chrysostoma</i>	16	0	0	0	0	0	0	0	1	0
<i>L. crispula</i>	17	0	0	2	0	0	0	0	0	0
<i>Peckia sp</i>	18	0	0	0	0	0	0	0	0	1
<i>O. angrensis</i>	19	0	0	0	0	0	0	2	1	0
<i>O. adminixta</i>	20	0	0	0	0	0	0	0	0	1
<i>P (P) pexata</i>	21	0	0	0	0	0	0	0	1	0
<i>O. parva</i>	22	0	0	0	0	0	0	0	1	0
<i>Helicobia sp</i>	23	0	0	1	0	0	0	0	0	0
<i>B. orientata</i>	24	1	0	0	0	0	0	0	0	0
<i>O. diana</i>	25	0	0	1	0	0	0	0	0	0

### b.) Análise com dados inteiros

As Tabelas 2, 3 e 4 da análise dos dados binários também foram utilizadas para a primeira busca de *biclusters* considerando os dados como inteiros. Nesse contexto, buscou-se *biclusters* do tipo BVCC, BVCL, BVCo e OPSM (BEC).

Como a nossa base de dados contém muitos zeros, a maioria dos *biclusters* buscados nessa etapa possuem valores iguais a 0. Do ponto de vista biológico, dependendo do caso, a interpretação desses valores possivelmente não traz grandes avanços. A ausência de respostas prejudica a análise dos dados, mas por outro lado esses dados faltantes não podem ser ignorados. Essa ausência geralmente implica em um desequilíbrio nas propriedades estatísticas da amostra realizada da população, causando um viés na média das respostas. Nessa situação, é importante que a base de dados seja tratada de forma a manter essas propriedades estatísticas as mais próximas possíveis de seus valores reais. Nesse sentido, foi

feita uma nova análise buscando-se outros *biclusters* apenas do tipo OPSM, considerando os zeros como dados faltantes, utilizando **uma versão do RInClose específica para isso**.

## Análise 2

As cinco espécies mais abundantemente coletadas de cada família formaram uma tabela tendo dados de isca e local associados (produzindo as Tabelas 5, 6 e 7). A partir de cada tabela, foram gerados apenas *biclusters* do tipo OPSM.

Tabela 5 - Cinco espécies mais abundantes coletadas de Muscidae, distribuídas em seus locais de coleta e isca utilizadas.

Espécies - Muscidae	URBANA			RURAL			FLORESTA			
	FÍGADO	CARNE	SARDINHA	FÍGADO	CARNE	SARDINHA	FÍGADO	CARNE	SARDINHA	
	1	2	3	4	5	6	7	8	9	
<i>A. orientalis</i>	1	14	225	132	5	329	13	0	137	39
<i>S. nudiseta</i>	2	1	17	15	0	0	1	0	0	0
<i>M. domestica</i>	3	1	26	5	29	39	1	0	0	0
<i>O. chalcogaster</i>	4	0	21	4	9	17	1	0	1	0
<i>Muscina sp</i>	5	0	1	0	0	0	0	0	0	0

Tabela 6- Cinco espécies mais abundantes coletadas de Sarcophagidae, distribuídas em seus locais de coleta e isca utilizadas.

Espécies - Sarcophagidae	URBANA			RURAL			FLORESTA			
	FÍGADO	CARNE	SARDINHA	FÍGADO	CARNE	SARDINHA	FÍGADO	CARNE	SARDINHA	
	1	2	3	4	5	6	7	8	9	
<i>S. lambens</i>	1	0	8	27	1	2	6	7	16	13
<i>O. thornax</i>	2	10	6	19	2	4	23	7	3	5
<i>P (Euboetcheria)</i>	3	0	0	0	0	0	1	5	20	20
<i>P (E) collusor</i>	4	0	2	1	0	0	0	5	23	25
<i>P (P) intermutans</i>	5	0	0	0	0	0	0	23	10	8

Tabela 7 - Cinco espécies mais abundantes coletadas de Calliphoridae, distribuídas em seus locais de coleta e isca utilizadas.

Espécies - Calliphoridae	URBANA			RURAL			FLORESTA			
	FÍGADO	CARNE	SARDINHA	FÍGADO	CARNE	SARDINHA	FÍGADO	CARNE	SARDINHA	
	1	2	3	4	5	6	7	8	9	
<i>L. eximia</i>	1	127	26	10	51	23	3	469	55	30
<i>H. segmentaria</i>	2	1	0	0	0	1	10	50	8	25
<i>H. semidiphana</i>	3	0	0	0	0	0	0	26	13	1
<i>C. megacephala</i>	4	17	9	11	15	3	0	27	42	10
<i>C. albiceps</i>	5	0	2	25	0	1	0	15	63	39

### Análise 3

As Tabelas 5, 6 e 7 foram unidas em única Tabela (Tabela 8) e a partir dela, foram gerados biclusters do tipo OPSM, considerando o zero como dado faltante. Nessa etapa, a análise foi feita diretamente dessa forma, por ter-se avaliado que os resultados oriundos de uma análise considerando o zero como dado faltante têm se apresentado mais viáveis utilizando-se uma base de dados com muitos zeros, como a que está sendo utilizada no presente trabalho.

## RESULTADOS

Como visto anteriormente, o *bicluster* é formado por um subconjunto de objetos e um subconjunto de atributos não-contíguos na matriz de dados original. No presente trabalho, o conjunto de objetos é representado pelas espécies de dípteros necrófagos e o conjunto de atributos pode ser, dependendo do caso, as iscas utilizadas, o local de coleta ou ambos simultaneamente.

Em vista disso, para uma melhor visualização e uma posterior interpretação, em alguns *biclusters* obtidos aqui, foi colocada também a sua demonstração via tabela. Porém, em alguns casos, é necessário voltar à tabela original para uma melhor compreensão. Além disso, os resultados mais relevantes foram grifados em cinza.

A busca dos *biclusters* foi dividida em **Análises 1, 2 e 3**, sendo a primeira contendo também uma análise considerando os dados como binários.

### Análise 1

#### Análise com dados binários

Através da **Tabela 2** convertida para 0 e 1, foram encontrados cinco *biclusters* com valores constantes iguais a 0, como por exemplo o  $A\{2\} = [5\ 6]$ ;  $B\{2\} = [1\ 3\ 4\ 5\ 6\ 7\ 8\ 9]$ , que mostra a ausência de *Muscina* sp e *Ophyra* sp, em todos os ambientes, com exceção da área urbana (isca de carne bovina moída).

Tabela 8 - Cinco espécies mais abundantes coletadas de cada família, distribuídas em seus locais de coleta e isca utilizadas.

Espécies	URBANA			RURAL			FLORESTA			
	FÍGADO	CARNE	SARDINHA	FÍGADO	CARNE	SARDINHA	FÍGADO	CARNE	SARDINHA	
	1	2	3	4	5	6	7	8	9	
<i>A. orientalis</i>	1	14	225	132	5	329	13	0	137	39
<i>S.nudiseta</i>	2	1	17	15	0	0	1	0	0	0
<i>M. domestica</i>	3	1	26	5	29	39	1	0	0	0
<i>O. chalcogaster</i>	4	0	21	4	9	17	1	0	1	0
<i>Muscina sp</i>	5	0	1	0	0	0	0	0	0	0
<i>L. eximia</i>	6	127	26	10	51	23	3	469	55	30
<i>H. segmentaria</i>	7	1	0	0	0	1	10	50	8	25
<i>H. semidiphana</i>	8	0	0	0	0	0	0	26	13	1
<i>C. megacephala</i>	9	17	9	11	15	3	0	27	42	10
<i>C. albiceps</i>	10	0	2	25	0	1	0	15	63	39
<i>P. (S) lambens</i>	11	0	8	27	1	2	6	7	16	13
<i>O. thornax</i>	12	10	6	19	2	4	23	7	3	5
<i>P (Euboetcheria)</i>	13	0	0	0	0	0	1	5	20	20
<i>P (E) collusor</i>	14	0	2	1	0	0	0	5	23	25
<i>P (P) intermutans</i>	15	0	0	0	0	0	0	23	10	8

Cinco *biclusters* com valores maiores que zero também foram encontrados, sendo que quatro destes possuem associação entre *A. orientalis* e *M. domestica*. Eles estão listados a seguir:

$$* A\{1\} = [1\ 2\ 3]; B\{1\} = [1\ 2\ 3\ 6];$$

que refere-se à presença *A. orientalis*, *S. nudiseta* e *M. domestica* em todas as iscas expostas na área urbana e na área rural, apenas na isca de sardinha.

$$* A\{2\} = [1\ 3]; B\{2\} = [1\ 2\ 3\ 4\ 5\ 6],$$

refere-se à presença *A. orientalis* e *M. domestica*, em todas as iscas expostas tanto na área urbana quanto na área rural.

$$* A\{3\} = [1\ 2\ 3\ 4]; B\{3\} = [2\ 3\ 6];$$

refere-se à presença *A. orientalis*, *S. nudiseta*, *M. domestica* e *O. chalcogaster* na área urbana (isca de carne e sardinha) e na área rural, na isca de sardinha.

$$* A\{4\} = [1\ 3\ 4]; B\{4\} = [2\ 3\ 4\ 5\ 6];$$

refere-se à presença *A. orientalis*, *M. domestica* e *O. chalcogaster* na área urbana (isca de carne e sardinha) e na área rural, em todas as iscas,

$$* A\{5\} = [1\ 4]; B\{5\} = [2\ 3\ 4\ 5\ 6\ 8];$$

refere-se à presença *A. orientalis* e *O. chalcogaster* na área urbana (isca de carne e sardinha), na área rural (em todas as iscas) e na área de floresta, com a isca de carne.

Através da **Tabela 4** convertida para 0 e 1, ao se usar dados binários, encontraram-se 10 *biclusters* com valores constantes iguais a 0. Todos eles estão relacionados no **Apêndice A** e para uma melhor compreensão, é necessário avaliá-los juntamente com a tabela, já que os números que eles contêm são aqueles que estão grifados em vermelho na tabela.

Tendo como objetivo analisar a importância biológica desses resultados, notou-se que um *bicluster* com todos os valores iguais a zero não traz nenhuma informação e por isso, para as análises das demais tabelas, eles não serão demonstrados aqui.

Além disso, para a **Tabela 4** foram encontrados 12 *biclusters* com valores maiores que zero, sendo que 11 destes possuem uma associação entre *L. eximia* e *C. megacephala*. Eles estão listados a seguir:

$$* A\{1\} = [1\ 2\ 3\ 5]; B\{1\} = [1\ 5];$$

que refere-se à presença das espécies *L. eximia*, *H. segmentaria*, *L. cuprina* e *C. megacephala*, na área urbana com a isca fígado e na área rural com a isca de carne.

$$* A\{2\} = [1\ 3\ 5]; B\{2\} = [1\ 2\ 3\ 4\ 5];$$

refere-se à presença das espécies *L. eximia*, *L. cuprina* e *C. megacephala*, na área urbana (todas as iscas) e na área rural com a isca de fígado e carne.

$$* A\{3\} = [1\ 5]; B\{3\} = [1\ 2\ 3\ 4\ 5\ 7\ 8\ 9];$$

refere-se à presença das espécies *L. eximia* e *C. megacephala* em todos os ambientes, exceto na área rural com a isca sardinha.

$$* A\{4\} = [1\ 2\ ]; B\{4\} = [1\ 5\ 6\ 7\ 8\ 9\ ];$$

refere-se à presença das espécies *L. eximia* e *H. segmentaria* na área urbana (isca de fígado), área rural (iscas de carne e sardinha) e na área de floresta, em todas as iscas.

$$* A\{5\} = [1\ 2\ 5\ ]; B\{5\} = [1\ 5\ 7\ 8\ 9\ ];$$

refere-se à presença das espécies *L. eximia*, *H. segmentaria* e *C. megacephala* na área urbana (isca fígado), na área rural (isca carne) e na floresta, em todas as iscas.

$$* A\{6\} = [1\ 3\ 5\ 6\ 7\ ]; B\{6\} = [2\ 5\ ];$$

refere-se à presença das espécies *L. eximia*, *L. cuprina*, *C. megacephala*, *C. albiceps* e *C. idioidea*, na área urbana e na área rural com a isca de carne.

$$* A\{7\} = [1\ 3\ 5\ 6\ ]; B\{7\} = [2\ 3\ 5\ ];$$

refere-se à presença das espécies *L. eximia*, *L. cuprina* e *C. megacephala* e *C. albiceps* na área urbana (isca de carne e sardinha) e na área rural com a isca de carne.

$$* A\{8\} = [1\ 5\ 6\ ]; B\{8\} = [2\ 3\ 5\ 7\ 8\ 9\ ];$$

refere-se à presença das espécies *L. eximia*, *C. megacephala* e *C. albiceps* na área urbana (isca carne e sardinha), na área rural (isca de carne) e na área de floresta, em todas as iscas.

$$* A\{9\} = [1\ 5\ 6\ 7\ ]; B\{9\} = [2\ 5\ 7\ 8\ 9\ ];$$

refere-se à presença das espécies *L. eximia*, *C. megacephala*, *C. albiceps* e *C. idioidea* nas áreas urbana e rural com a isca de carne e na área da floresta, em todas as iscas.

$$* A\{10\} = [1\ 5\ 8\ ]; B\{10\} = [4\ 7\ ];$$

refere-se à presença das espécies *L. eximia*, *C. megacephala* e *C. macellaria* na área rural e de floresta, com a isca de fígado.



$$* A\{11\} = [1\ 2\ 5\ 6\ 7]; B\{11\} = [5\ 7\ 8\ 9];$$

refere-se à presença das espécies *L. eximia*, *H. segmentaria*, *C. megacephala*, *C. albiceps* e *C. idioidea* na área rural com a isca de carne, e na área de floresta em todas as iscas.

$$* A\{12\} = [1\ 2\ 4\ 5\ 6\ 7\ 9]; B\{12\} = [7\ 8\ 9];$$

refere-se à presença das espécies *L. eximia*, *H. segmentaria*, *H. semidiaphana*, *C. megacephala*, *C. albiceps*, *C. idioidea* e *M. peregrina*, em todas as iscas na área de mata.

É importante constar que, pela **Tabela 5** convertida para 0 e 1, ao se usar dados binários, encontraram-se 84 biclusters com valores constantes iguais a 0, como por exemplo:  $A\{15\} = [17, 18, 20, 23, 25]$ ;  $B\{15\} = [1, 2, 4, 5, 6, 7, 8]$ , que se referem à ausência de exemplares de *Sarcophaga (Lipoptilocnema) crispula* Lopes, *Peckia* sp, *Oxysarcodexia admixta* (Lopes), *Helicobia* sp e *Oxysarcodexia diana* (Lopes) em todos os locais, exceto na isca de sardinha presente na área urbana e na área de floresta.

Além disso, foram encontrados 21 biclusters com valores maiores que 0, sendo que 18 destes mostraram associação entre *P. (S) lambens* e *O. thornax*. Eles estão listados a seguir:

$$* A\{1\} = [3\ 7]; B\{1\} = [1\ 2\ 5\ 7];$$

que refere-se à presença das espécies *O. thornax* e *O. paulistanensis* na área urbana (isca de fígado e carne), na área rural (carne) e na área de floresta (isca fígado).

$$* A\{2\} = [2\ 3]; B\{2\} = [1\ 3\ 5\ 7\ 8\ 9];$$

refere-se à presença das espécies *Oxysarcodexia* sp e *O. thornax*, na área urbana (iscas de fígado e sardinha), na área rural (isca de carne) e em todas as iscas na área de floresta.

$$* A\{3\} = [2\ 3\ 7]; B\{3\} = [1\ 5\ 7];$$

refere-se à presença das espécies *Oxysarcodexia* sp, *O. thornax* e *O. paulistanensis* na área urbana (isca de fígado), na área rural (carne) e na área de floresta (isca fígado).

$$* A\{4\} = [1\ 3\ 6\ 7\ 10\ 11]; B\{4\} = [2\ 7];$$

refere-se à presença das espécies *P. (S) lambens*, *O. thornax*, *P. (E) collusor*, *O. paulistanensis*, *P. (S) ingens* e *P. (E) anguilla* na área urbana (isca de carne) e na área de floresta ( isca fígado).

$$* A\{5\} = [1\ 3\ 6\ 11\ ]; B\{5\} = [2\ 3\ 7\ 8\ 9\ ];$$

refere-se à presença das espécies *P. (S) lambens*, *O. thornax*, *P. (E) collusor* e *P. (E) anguilla* na área urbana (isca de carne e sardinha)e em todas as iscas na área de floresta.

$$* A\{6\} = [1\ 3\ ]; B\{6\} = [2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ ];$$

refere-se à presença das espécies *P. (S) lambens* e *O. thornax*, na área urbana (isca de carne e sardinha) e em todas as iscas na área rural e na de floresta.

$$* A\{7\} = [1\ 3\ 11\ ]; B\{7\} = [2\ 3\ 5\ 6\ 7\ 8\ 9\ ];$$

refere-se à presença das espécies *P. (S) lambens*, *O. thornax* e *P. (E) anguilla* nas áreas urbana e rural, sendo ambas com isca de carne e sardinha e em todas as iscas na área rural e na de floresta.

$$* A\{8\} = [1\ 3\ 7\ 10\ 11\ ]; B\{8\} = [2\ 5\ 7\ ];$$

refere-se à presença das espécies *P. (S) lambens*, *O. thornax*, *O. paulistanensis*, *P. (S) ingense* *P. (E) anguilla* nas áreas urbana e rural, sendo ambas com isca de carne e na de floresta, com isca de fígado.

$$* A\{9\} = [1\ 3\ 10\ 11\ ]; B\{9\} = [2\ 5\ 7\ 8\ 9\ ];$$

refere-se à presença das espécies *P. (S) lambens*, *O. thornax*, *P. (S) ingens* e *P. (E) anguilla* nas áreas urbana e rural, sendo ambas com isca de carne e na de floresta, em todas as iscas.

$$* A\{10\} = [1\ 3\ 6\ 10\ 11\ ]; B\{10\} = [2\ 7\ 8\ 9\ ];$$

refere-se à presença das espécies *P. (S) lambens*, *O. thornax*, *P. (E) collusor*, *P. (S) ingens* e *P. (E) anguilla* nas áreas urbana (isca de carne) e na área de floresta, em todas as iscas.

$$* A\{11\} = [1\ 3\ 14\ ]; B\{11\} = [3\ 4\ 6\ ];$$

refere-se à presença das espécies *P. (S) lambens*, *O. thornax* e *R. belforti* na área urbana (isca de sardinha) e na área rural (isca de fígado e sardinha).

$$* A\{12\} = [1\ 2\ 3\ 11]; B\{12\} = [3\ 5\ 7\ 8\ 9];$$

refere-se à presença das espécies *P. (S) lambens*, *Oxysarcodexia* sp., *O. thornax* e *P. (E) anguila* na área urbana (isca sardinha), na área rural (isca carne) e na área de floresta, em todas as iscas.

$$* A\{13\} = [1\ 3\ 11\ 14]; B\{13\} = [3\ 6];$$

refere-se à presença das espécies *P. (S) lambens*, *O. thornax*, *P. (E) anguila* e *R. belforti* nas áreas urbana e rural com isca de carne, ambas na isca de sardinha.

$$* A\{14\} = [1\ 2\ 3\ 6\ 11]; B\{14\} = [3\ 7\ 8\ 9];$$

refere-se à presença das espécies *P. (S) lambens*, *Oxysarcodexia* sp, *O. thornax*, *P. (E) collusor* e *P. (E) anguila* na área urbana rural com isca de sardinha e na área de floresta, em todas as iscas.

$$* A\{15\} = [1\ 2\ 3\ 7\ 10\ 11]; B\{15\} = [5\ 7];$$

refere-se à presença das espécies *P. (S) lambens*, *Oxysarcodexia* sp, *O. thornax*, *O. paulistanensis*, *P. (S) ingens* e *P. (E) anguila* na área rural com isca de carne e na área de floresta, com a isca de fígado.

$$* A\{16\} = [1\ 2\ 3\ 10\ 11]; B\{16\} = [5\ 7\ 8\ 9];$$

refere-se à presença das espécies *P. (S) lambens*, *Oxysarcodexia* sp., *O. thornax*, *P. (S) ingens* e *P. (E) anguila* na área rural com isca de carne e na área de floresta, em todas as iscas.

$$* A\{17\} = [1\ 3\ 5\ 8\ 11]; B\{17\} = [6\ 7\ 8\ 9];$$

refere-se à presença das espécies *P. (S) lambens*, *O. thornax*, *P. (Euboetcheria)* sp, *O. avuncula*, e *P. (E) anguila* na área rural com a isca de sardinha e na área de floresta, em todas as iscas,

$$* A\{18\} = [1\ 3\ 4\ 5\ 8\ 11]; B\{18\} = [6\ 9];$$

refere-se à presença das espécies *P. (S) lambens*, *O. thornax*, *O. culmiforcipes*, *P. (Euboetcheria)* sp, *O. avuncula*, e *P. (E) anguilla* nas áreas rural e de floresta, ambas na isca de sardinha.

\* A{19} = [1 2 3 5 6 8 10 11 12 15 19 ]; B{19} = [7 8 ];

refere-se à presença das espécies *P. (S) lambens*, *Oxysarcodexia* sp., *O. thornax*, *P. (Euboetcheria)* sp, *P. (E) collusor*, *O. avuncula*,, *P. (S) ingens* *P. (E) anguilla*, *P. (P) intermutans* *P (S) florencioi* e *O. angrensis* na área de floresta, nas iscas de fígado e carne.

\* A{20} = [1 2 3 5 6 8 10 11 12 15 ]; B{20} = [7 8 9 ];

refere-se à presença das espécies *P. (S) lambens*, *Oxysarcodexia* sp., *O. thornax*, *P. (Euboetcheria)* sp, *P. (E) collusor*, *O. avuncula*,, *P. (S) ingens* *P. (E) anguilla*, *P. (P) intermutans* e *P (S) florencioi* na área de floresta, em todas as iscas.

\* A{21} = [1 2 3 5 6 8 10 11 12 13 15 ]; B{21} = [8 9 ];

refere-se à presença das espécies *P. (S) lambens*, *Oxysarcodexia* sp., *O. thornax*, *P. (Euboetcheria)* sp, *P. (E) collusor*, *O. avuncula*,, *P. (S) ingens* *P. (E) anguilla*, *P. (P) intermutans*, *P. (E) australis* e *P (S) florencioi* na área de floresta, nas iscas de carne e sardinha.

### **Análise com dados inteiros**

#### **BVCC/BVCL**

Durante a busca por *biclusters* do tipo BVCC (valores constantes nas colunas) e BVCL (valores constantes nas linhas), notou-se que foram obtidos *biclusters* contendo zero ou um, de forma constante em suas linhas ou colunas. Do ponto de vista biológico, esse tipo de resultado não é válido. Para exemplificar, alguns exemplos foram listados no **Apêndice B**.

## OPSM

A partir da **Tabela 2 (Muscidae)**, obteve-se um total de 34 *biclusters* do tipo OPSM que estão listados no **Apêndice C**. Observando-os, notou-se que eles possuem muitos zeros e por isso, apenas alguns foram demonstrados e os demais foram somente citados.

No que se refere à **Tabela 3 (Calliphoridae)**, obteve-se um total de 85 *biclusters* do tipo OPSM que estão listados no **Apêndice D** e para a **Tabela 4 (Sarcophagidae)**, encontraram-se um total de 128 *biclusters* do tipo OPSM, listados no **Apêndice E**.

Devido à falta de informação que um *bicluster* contendo muitos zeros fornece, foi feita uma análise considerando os zeros como dados faltantes. A partir disso, notou-se uma queda no número de *biclusters* para todas as famílias. No caso da **Tabela 2**, foi de 34 para 3 *biclusters* do tipo OPSM:

$$* A\{1\} = [1\ 2\ 3]; B\{1\} = [1\ 2\ 3\ 6];$$

Tabela 9 – Demonstração em formato de tabela do *bicluster* OPSM de número 1, obtido através da Tabela 3.

Espécies		URBANA		RURAL	
		FÍGADO	CARNE	SARDINHA	SARDINHA
		<b>1</b>	<b>2</b>	<b>3</b>	<b>6</b>
<i>A. orientalis</i>	<b>1</b>	14	225	132	13
<i>S. nudiseta</i>	<b>2</b>	1	17	15	1
<i>M. domestica</i>	<b>3</b>	1	26	5	1

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *A. orientalis*, *S. nudiseta* e *M. domestica*: 6 - Rural (sardinha) ≤ 1 - Urbana (fígado) ≤ 3 - Urbana (sardinha) ≤ 6 - Rural (carne).

$$* A\{2\} = [1\ 2\ 3\ 4]; B\{2\} = [2\ 3\ 6];$$

Tabela 10 - Demonstração em formato de tabela do bicluster OPSM de número 2, obtido através da Tabela 3.

Espécies	URBANA		RURAL
	CARNE	SARDINHA	SARDINHA
	<b>2</b>	<b>3</b>	<b>6</b>
<i>A. orientalis</i>	<b>1</b>	225	132
<i>S.nudiseta</i>	<b>2</b>	17	15
<i>M. domestica</i>	<b>3</b>	26	5
<i>O. chalcogaster</i>	<b>4</b>	21	4

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *A.orientalis*, *S. nudiseta*, *M. domestica* e *O. chalcogaster*: 6 - Rural (sardinha)  $\leq$  3 - Urbana (sardinha)  $\leq$  2 - Urbana (carne).

$$* A\{3\} = [1\ 3\ 4]; B\{3\} = [3\ 5\ 6];$$

Tabela 11- Demonstração em formato de tabela do bicluster OPSM de número 3, obtido através da Tabela 3.

Espécies	URBANA		RURAL	
	SARDINHA	CARNE	SARDINHA	
	<b>3</b>	<b>5</b>	<b>6</b>	
<i>A. orientalis</i>	<b>1</b>	132	329	13
<i>M. domestica</i>	<b>3</b>	5	39	1
<i>O. chalcogaster</i>	<b>4</b>	4	17	1

que por sua vez, refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *A.orientalis*, *M. domestica* e *O. chalcogaster*: 5 – Rural (carne)  $\leq$  3 - Urbana (sardinha)  $\leq$  6 - Rural (sardinha).

Observando os três biclusters mostrados anteriormente, nota-se que nenhum deles possui alguma associação que envolva o ambiente de floresta e que todos envolveram as espécies *A. orientalis* e *M. domestica*.

Considerando ainda o zero como dado faltante, para a **Tabela 3 (Calliphoridae)**, encontraram-se apenas 8 *biclusters* do tipo OPSM, número bem inferior ao encontrado anteriormente ( 85 *biclusters*). Eles estão listados a seguir.

$$* A\{1\} = [1\ 3\ 5]; B\{1\} = [1\ 2\ 4];$$

Tabela 12- Demonstração em formato de tabela do bicluster OPSM de número 1, obtido através da Tabela 4.

Espécies	URBANA		RURAL
	FÍGADO	CARNE	FÍGADO
	<b>1</b>	<b>2</b>	<b>4</b>
<i>L. eximia</i>	<b>1</b>	127	26
<i>L. cuprina</i>	<b>3</b>	11	1
<i>C. megacephala</i>	<b>5</b>	17	9

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *L. eximia*, *L. cuprina* e *C. megacephala* : 2 – Urbana (carne) ≤ 4 - Rural (fígado) ≤ 1 - Urbana (fígado).

$$* A\{2\} = [1\ 5\ 6\ 7]; B\{2\} = [2\ 5\ 7];$$

Tabela 13 - Demonstração em formato de tabela do bicluster OPSM de número 2, obtido através da Tabela 4.

Espécies	URBANA	RURAL	FLORESTA
	CARNE	CARNE	FÍGADO
	<b>2</b>	<b>5</b>	<b>7</b>
<i>L. eximia</i>	<b>1</b>	26	23
<i>C. megacephala</i>	<b>5</b>	9	3
<i>C. albiceps</i>	<b>6</b>	2	1
<i>C. idioidea</i>	<b>7</b>	1	1

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *L. eximia*, *C. megacephala*, *C. albiceps* e *C. idioidea* : 5 – Rural (carne) ≤ 2 - Urbana (carne) ≤ 7 - Floresta (fígado).

$$* A\{3\} = [1\ 5\ 6\ 7]; B\{3\} = [2\ 5\ 8\ 9];$$

Tabela 14 - Demonstração em formato de tabela do bicluster OPSM de número 3, obtido através da Tabela 4.

Espécies	URBANA	RURAL	FLORESTA	
	CARNE	CARNE	CARNE	SARDINHA
	<b>2</b>	<b>5</b>	<b>8</b>	<b>9</b>
<i>L. eximia</i>	<b>1</b> 26	23	55	30
<i>C. megacephala</i>	<b>5</b> 9	3	42	10
<i>C. albiceps</i>	<b>6</b> 2	1	63	39
<i>C. idioidea</i>	<b>7</b> 1	1	2	1

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *L. eximia*, *C. megacephala*, *C. albiceps* e *C. idioidea* : 5 – Rural (carne) ≤ 2 - Urbana (carne) ≤ 9 - Floresta (sardinha) ≤ 8 - Floresta (carne).

$$* A\{4\} = [1\ 2\ 5\ ]; B\{4\} = [1\ 5\ 7\ ];$$

Tabela 15 - Demonstração em formato de tabela do bicluster OPSM de número 4, obtido através da Tabela 4.

Espécies	URBANA	RURAL	FLORESTA
	FÍGADO	CARNE	FÍGADO
	<b>1</b>	<b>5</b>	<b>7</b>
<i>L. eximia</i>	<b>1</b> 127	23	469
<i>H. segmentaria</i>	<b>2</b> 1	3	50
<i>C. megacephala</i>	<b>5</b> 17	1	27

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *L. eximia*, *H. segmentaria* e *C. megacephala* : 5 – Rural (carne) ≤ 1 - Urbana (fígado) ≤ 7 - Floresta (fígado).

$$* A\{5\} = [5\ 6\ 7\ ]; B\{5\} = [2\ 5\ 7\ 8\ ];$$



Tabela 16 - Demonstração em formato de tabela do bicluster OPSM de número 5, obtido através da Tabela 4.

Espécies	URBANA		RURAL	FLORESTA	
	CARNE	CARNE	FÍGADO	CARNE	
	<b>2</b>	<b>5</b>	<b>7</b>	<b>8</b>	
<i>C. megacephala</i>	<b>5</b>	9	3	27	42
<i>C. albiceps</i>	<b>6</b>	2	1	15	63
<i>C. idioidea</i>	<b>7</b>	1	1	1	2

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *C. megacephala*, *C. albiceps* e *C. idioidea* : 5 – Rural (carne) ≤ 2 - Urbana (carne) ≤ 7 - Floresta (fígado) ≤ 8- Floresta (carne).

$$* A\{6\} = [1 \ 5 \ 7]; B\{6\} = [2 \ 5 \ 7 \ 9];$$

Tabela 17 - Demonstração em formato de tabela do bicluster OPSM de número 6, obtido através da Tabela 4.

Espécies	URBANA		RURAL	FLORESTA	
	CARNE	CARNE	FÍGADO	SARDINHA	
	<b>2</b>	<b>5</b>	<b>7</b>	<b>9</b>	
<i>L. eximia</i>	<b>1</b>	26	23	469	30
<i>C. megacephala</i>	<b>5</b>	9	3	27	10
<i>C. idioidea</i>	<b>7</b>	1	1	1	1

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *L. eximia*, *C. megacephala* e *C. idioidea* : 5 – Rural (carne) ≤ 2 - Urbana (carne) ≤ 9 – Floresta (sardinha) ≤ 7 - Floresta (fígado).

$$* A\{7\} = [1 \ 2 \ 5 \ 7]; B\{7\} = [5 \ 7 \ 9];$$

Tabela 18 - Demonstração em formato de tabela do bicluster OPSM de número 7, obtido através da Tabela 4.

Espécies	RURAL		FLORESTA	
	CARNE	FÍGADO	SARDINHA	
	<b>5</b>	<b>7</b>		<b>9</b>
<i>L. eximia</i>	<b>1</b>	23	469	30
<i>H. segmentaria</i>	<b>2</b>	1	50	25
<i>C. megacephala</i>	<b>5</b>	3	27	10
<i>C. idioidea</i>	<b>7</b>	1	1	1

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *L. eximia*, *H. segmentaria*, *C. megacephala* e *C. idioidea* : 5 – Rural (carne) ≤ 9 – Floresta (sardinha) ≤ 7 - Floresta (fígado).

$$* A\{8\} = [5\ 7\ 9]; B\{8\} = [7\ 8\ 9];$$

Tabela 19 - Demonstração em formato de tabela do bicluster OPSM de número 8, obtido através da Tabela 4.

Espécies	FLORESTA		
	FÍGADO	CARNE	SARDINHA
	<b>7</b>	<b>8</b>	<b>9</b>
<i>C. megacephala</i>	<b>5</b>	27	42
<i>C. idioidea</i>	<b>7</b>	1	2
<i>M. peregrina</i>	<b>9</b>	4	9

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *C. megacephala*, *C. idioidea* e *M. peregrina* : 9 – Floresta (sardinha) ≤ 7 - Floresta (fígado) ≤ 8 – Floresta (carne).

Desses oito *biclusters* citados anteriormente, seis mostraram associação entre as espécies *L. eximia* e *C. megacephala*. Além disso, seis também mostraram que o ambiente rural utilizando-se a isca de carne foi onde as espécies de califorídeos encontradas associadas ocorreram em menor número. Por outro lado, sete *biclusters* dos oito encontrados denotam a maior incidência das espécies dessa família no ambiente de Floresta com a isca de fígado ou carne.

Finalizando a análise considerando o zero como dado faltante, obteve-se um número bem inferior ao encontrado anteriormente (128 *biclusters*), que foi um total de 14 *biclusters* do tipo OPSM para a **Tabela 4 (Sarcophagidae)**. Eles estão listados a seguir.

$$* A\{1\} = [1 \ 3 \ 11]; B\{1\} = [2 \ 3 \ 5];$$

Tabela 20 - Demonstração em formato de tabela do bicluster OPSM de número 1, obtido através da Tabela 5.

Espécies	URBANA		RURAL	
	CARNE	SARDINHA	CARNE	
		<b>2</b>	<b>3</b>	<b>5</b>
<i>P. (S) lambens</i>	<b>1</b>	8	27	2
<i>O. thornax</i>	<b>3</b>	6	19	4
<i>P (E) anguilla</i>	<b>11</b>	1	2	1

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *P. (S) lambens*, *O. thornax* e *P (E) anguilla*: 5 – Rural(carne) ≤ 2 – Urbana (carne) ≤ 3 – Urbana (sardinha).

$$* A\{2\} = [1 \ 3 \ 11]; B\{2\} = [3 \ 5 \ 7];$$

Tabela 21 - Demonstração em formato de tabela do bicluster OPSM de número 2, obtido através da Tabela 5.

Espécies	URBANA	RURAL	FLORESTA	
	SARDINHA	CARNE	FÍGADO	
		<b>3</b>	<b>5</b>	<b>7</b>
<i>P. (S) lambens</i>	<b>1</b>	27	2	7
<i>O. thornax</i>	<b>3</b>	19	4	7
<i>P (E) anguilla</i>	<b>11</b>	2	1	1

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *P. (S) lambens*, *O. thornax* e *P. (E) anguilla*: 5 – Rural(carne) ≤ 7 – Floresta (fígado) ≤ 3 – Urbana (sardinha).

\*  $A\{3\} = [7\ 10\ 11]$ ;  $B\{3\} = [2\ 5\ 7]$ ;

Tabela 22 - Demonstração em formato de tabela do bicluster OPSM de número 3, obtido através da Tabela 5.

		URBANA	RURAL	FLORESTA
Espécies		CARNE	CARNE	FÍGADO
		<b>2</b>	<b>5</b>	<b>7</b>
<i>O. paulistanensis</i>	<b>7</b>	1	1	1
<i>P. (S) ingens</i>	<b>10</b>	1	2	1
<i>P. (E) anguilla</i>	<b>11</b>	1	1	1

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *O. paulistanensis*, *P. (S) ingens* e *P. (E) anguilla*: 7 – Floresta (fígado) ≤ 2 – Urbana (carne) ≤ 5 – Rural (carne)

\*  $A\{4\} = [1\ 5\ 8\ 11]$ ;  $B\{4\} = [6\ 7\ 9]$ ;

Tabela 23- Demonstração em formato de tabela do bicluster OPSM de número 4 obtido através da Tabela 5.

		RURAL	FLORESTA	
Espécies		SARDINHA	FÍGADO	SARDINHA
		<b>6</b>	<b>7</b>	<b>9</b>
<i>P. (S) lambens</i>	<b>1</b>	6	7	13
<i>P. (Euboettcheria) sp</i>	<b>5</b>	1	5	20
<i>O. avuncula</i>	<b>8</b>	1	3	5
<i>P. (E) anguilla</i>	<b>11</b>	1	1	9

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *P. (S) lambens*, *P. (Euboettcheria) sp*, *O. avuncula* e *P. (E) anguilla*: 6 – Rural (sardinha) ≤ 7 – Floresta (fígado) ≤ 9 – Floresta (sardinha).

\*  $A\{5\} = [3\ 7\ 11]$ ;  $B\{5\} = [2\ 5\ 7]$ ;

Tabela 24 - Demonstração em formato de tabela do bicluster OPSM de número 5, obtido através da Tabela 5.

Espécies	RURAL		FLORESTA	
	SARDINHA	FÍGADO	SARDINHA	
		<b>2</b>	<b>5</b>	<b>7</b>
<i>O. thornax</i>	<b>3</b>	6	4	7
<i>O. paulistanensis</i>	<b>7</b>	1	1	1
<i>P (E) anguilla</i>	<b>11</b>	1	1	1

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *O. thornax*, *O. paulistanensis* e *P. (E) anguilla*: 5 – Rural (carne) ≤ 2 – Urbana (carne) ≤ 7 – Floresta (fígado).

\*  $A\{6\} = [1\ 7\ 11]$ ;  $B\{6\} = [2\ 5\ 7]$ ;

Tabela 25 - Demonstração em formato de tabela do bicluster OPSM de número 6, obtido através da Tabela 5.

Espécies	URBANA	RURAL	FLORESTA
	CARNE	CARNE	FÍGADO
		<b>2</b>	<b>5</b>
<i>S.lambens</i>	<b>1</b>	8	2
<i>O. paulistanensis</i>	<b>7</b>	1	1
<i>P (E) anguilla</i>	<b>11</b>	1	1

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *P. (S) lambens*, *O. paulistanensis* e *P. (E) anguilla*: 5 – Rural (carne) ≤ 7 – Floresta (fígado) ≤ 2 – Urbana (carne).

$$* A\{7\} = [6 \ 10 \ 11]; B\{7\} = [2 \ 7 \ 8];$$

Tabela 26- Demonstração em formato de tabela do bicluster OPSM de número 7, obtido através da Tabela 5.

Espécies	URBANA		FLORESTA	
	CARNE	FÍGADO	CARNE	
		<b>2</b>	<b>7</b>	<b>8</b>
<i>P (E) collusor</i>	<b>6</b>	2	5	23
<i>P (S) ingens</i>	<b>10</b>	1	1	7
<i>P (E) anguilla</i>	<b>11</b>	1	1	11

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *P (E) collusor*, *P. (S) ingens* e *P. (E) anguilla*: 2 – Urbana (carne) ≤ 7 – Floresta (figado) ≤ 8 – Floresta (carne).

$$* A\{8\} = [6 \ 10 \ 11]; B\{8\} = [2 \ 7 \ 9];$$

Tabela 27 - Demonstração em formato de tabela do bicluster OPSM de número 8, obtido através da Tabela 5.

Espécies	URBANA		FLORESTA	
	CARNE	FÍGADO	SARDINHA	
		<b>2</b>	<b>7</b>	<b>9</b>
<i>P (E) collusor</i>	<b>6</b>	2	5	25
<i>P (S) ingens</i>	<b>10</b>	1	1	5
<i>P (E) anguilla</i>	<b>11</b>	1	1	9

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *P (E) collusor*, *P. (S) ingens* e *P. (E) anguilla*: 2 – Urbana (carne) ≤ 7 – Floresta (figado) ≤ 9 – Floresta (sardinha).

\*  $A\{9\} = [1\ 10\ 11\ ]$ ;  $B\{9\} = [2\ 7\ 8\ 9\ ]$ ;

Tabela 28 - Demonstração em formato de tabela do bicluster OPSM de número 9, obtido através da Tabela 5.

Espécies	URBANA		FLORESTA		
	CARNE	FÍGADO	CARNE	SARDINHA	
		<b>2</b>	<b>7</b>	<b>8</b>	<b>9</b>
<i>S lambens</i>	<b>1</b>	8	7	16	13
<i>P (S) ingens</i>	<b>10</b>	1	1	7	5
<i>P (E) anguilla</i>	<b>11</b>	1	1	11	9

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *P. (S) lambens*, *P. (S) ingens* e *P. (E) anguilla*: 2 – Urbana (carne) ≤ 7 – Floresta (fígado) ≤ 8 – Floresta (carne) ≤ 9 – Floresta (sardinha).

\*  $A\{10\} = [1\ 10\ 11\ ]$ ;  $B\{10\} = [5\ 8\ 9\ ]$ ;

Tabela 29- Demonstração em formato de tabela do bicluster OPSM de número 10, obtido através da Tabela 5.

Espécies	RURAL	FLORESTA		
	CARNE	CARNE	SARDINHA	
		<b>5</b>	<b>8</b>	<b>9</b>
<i>S lambens</i>	<b>1</b>	2	16	13
<i>P (S) ingens</i>	<b>10</b>	2	7	5
<i>P (E) anguilla</i>	<b>11</b>	1	11	9

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *P. (S) lambens*, *P. (S) ingens* e *P. (E) anguilla*: 5 – Rural (carne) ≤ 9 – Floresta (sardinha) ≤ 8 – Floresta (carne).

\*  $A\{11\} = [1\ 2\ 11\ ]$ ;  $B\{11\} = [5\ 7\ 9\ ]$ ;

Tabela 30 - Demonstração em formato de tabela do bicluster OPSM de número 11, obtido através da Tabela 5.

Espécies	RURAL		FLORESTA	
	CARNE	FÍGADO	SARDINHA	
		<b>5</b>	<b>7</b>	<b>9</b>
<i>S lambens</i>	<b>1</b>	2	7	13
<i>Oxysarcodexia</i> sp	<b>2</b>	1	7	15
<i>P (E) anguilla</i>	<b>11</b>	1	1	9

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *P. (S) lambens*, *Oxysarcodexia* sp e *P. (E) anguilla*: 5 – Urbana (carne) ≤ 7 – Floresta (fígado) ≤ 9 – Floresta (sardinha).

\*  $A\{12\} = [1\ 5\ 11\ ]$ ;  $B\{12\} = [6\ 7\ 8\ 9\ ]$ ;

Tabela 31 - Demonstração em formato de tabela do bicluster OPSM de número 12, obtido através da Tabela 5.

Espécies	RURAL		FLORESTA		
	SARDINHA	FÍGADO	CARNE	SARDINHA	
		<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<i>S lambens</i>	<b>1</b>	6	7	16	13
<i>P (Euboettcheria)</i> sp	<b>5</b>	1	5	20	20
<i>P (E) anguilla</i>	<b>11</b>	1	1	11	9

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *P. (S) lambens*, *Oxysarcodexia* sp e *P. (E) anguilla*: 6 – Rural (sardinha) ≤ 7 – Floresta (fígado) ≤ 9 – Floresta (sardinha) ≤ 8 – Floresta (carne).



\* A{13} = [5 6 15 ]; B{13} = [7 8 9 ];

Tabela 32 - Demonstração em formato de tabela do bicluster OPSM de número 13, obtido através da Tabela 5.

Espécies	FLORESTA			
	FÍGADO	CARNE	SARDINHA	
		<b>7</b>	<b>8</b>	<b>9</b>
<i>P (Euboettcheria) sp</i>	<b>5</b>	5	20	20
<i>P (E) collusor</i>	<b>6</b>	5	23	25
<i>P (S) florencioi</i>	<b>15</b>	3	5	5

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *P (Euboettcheria) sp*, *P (E) collusor* e *P (S) florencioi*: 6– Rural (sardinha) ≤ 7 – Floresta (fígado) ≤ 9 – Floresta (sardinha) ≤ 8 – Floresta (carne).

\* A{14} = [1 5 10 11 15 ]; B{14} = [7 8 9 ];

Tabela 33 - Demonstração em formato de tabela do bicluster OPSM de número 14, obtido através da Tabela 5.

Espécies	FLORESTA			
	FÍGADO	CARNE	SARDINHA	
		<b>7</b>	<b>8</b>	<b>9</b>
<i>P. (S) lambens</i>	<b>1</b>	7	16	13
<i>P (Euboettcheria) sp</i>	<b>5</b>	5	20	20
<i>P (S) ingens</i>	<b>10</b>	1	7	5
<i>P (E) anguilla</i>	<b>11</b>	1	11	9
<i>P (S) florencioi</i>	<b>15</b>	3	5	5

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *P. (S) lambens*, *P. (Euboettcheria) sp.*, *P. (S) ingens*, *P. (E) anguilla* e *P (S)*

*florencioi*: 6 – Rural (sardinha) ≤ 7 – Floresta (fígado) ≤ 9 – Floresta (sardinha) ≤ 8 – Floresta (carne).

Observando-se os 14 *biclusters* citados acima, nota-se que:

- Nove *biclusters* mostraram associação entre *P. (S) lambens* e *P. (E) anguilla* e destes, seis indicaram que o ambiente floresta com a isca carne ou sardinha seria onde houve a maior incidências dessas espécies de sarcófagídeos.

- Seis *biclusters* possuem a associação entre *P. (S) ingens* e *P. (E) anguilla* e destes, cinco mostram que o ambiente de floresta é o local onde houve a maior ocorrência dessas espécies conjuntamente.

- Os *biclusters* de número um e o de número dois relacionam as mesmas espécies (*P. (S) lambens*, *O. thornax* e *P. (E) anguilla*), sendo que em ambos o ambiente rural com a isca de carne foi onde ocorreu a menor quantidade de espécimes coletado. Por outro lado, a área rural com a isca de sardinha foi onde ocorreu a maior.

- Os *biclusters* de número sete e oito relacionam as mesmas espécies (*P.(E) collusor*, *P. (S) ingens* e *P. (E) anguilla*). A única diferença entre eles é no tipo de isca do local com maior número de indivíduos coletados ( floresta), pois em um é a carne e no outro é a sardinha.

- Os *biclusters* de número nove e dez também relacionam espécies iguais (*P. (S) lambens*, *P. (S) ingens* e *P. (E) anguilla*), mostrando que elas ocorreram em maior número no ambiente de floresta.

## **Análise 2**

Cabe relembrar aqui que a diferença entre a Análise 2 e a anterior: esta considera apenas as cinco espécies mais abundantes de cada família.

Sobre a Análise 2, a primeira família escolhida foi Muscidae. Os dados contidos na **Tabela 5** foram rodados utilizando o *Rin-Close* e foram gerados 25 *biclusters* OPSM .Todos estão listados no **Apêndice F**

A segunda família considerada foi Calliphoridae. Os dados contidos na **Tabela 6** foram rodados utilizando o *Rin-Close* e foram gerados 41 *biclusters* do tipo OPSM, que estão listados no **Apêndice G**.

E por último, os dados contidos na **Tabela 7** referentes as espécies de Sarcophagidae foram rodados utilizando o *Rin-Close* e foram gerados 28 *biclusters* do tipo OPSM, que estão descritos no **Apêndice H**.

Comparando-se a análise descrita acima com àquela feita considerando-se o zero como dado faltante, notou-se uma queda no número de *biclusters* tanto para **Muscidae** (de 25 para 2 *biclusters*) quanto para **Calliphoridae** (de 41 para 6 *biclusters*). No que se refere aos dados de *Sarcophagidae*, não foi encontrado nenhum *bicluster*.

Abaixo, seguem os *biclusters* OPSM para a família **Muscidae**. Nesse caso, apesar da quantidade pequena obtida de *biclusters* (dois), notou-se a repetição da associação entre *A. orientalis* e *M. domestica* vista na **Análise 1**, mas aqui também há a presença de *S. nudiseta*. No intuito de facilitar a interpretação, apresenta-se também a tabela evidenciando a ausência do número zero:

$$* A\{1\} = [1\ 2\ 3]; B\{1\} = [1\ 2\ 3\ 6];$$

Tabela 34 - Demonstração em formato de tabela do *bicluster* OPSM de número 1, considerando o zero como dado faltante, obtido através da Tabela 6.

Espécies	URBANA			RURAL	
	FÍGADO	CARNE	SARDINHA	SARDINHA	
	<b>1</b>	<b>2</b>		<b>3</b>	<b>6</b>
<i>A. orientalis</i>	<b>1</b>	14	225	132	13
<i>S. nudiseta</i>	<b>2</b>	1	17	15	1
<i>M. domestica</i>	<b>3</b>	1	26	5	1

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *S. nudiseta* e *M. domestica* : 6- Rural (sardinha)  $\leq$  1 – Urbana (fígado)  $\leq$  3 – Urbana (sardinha)  $\leq$  2 – Urbana (carne).

$$* A\{2\} = [1\ 2\ 3\ 4]; B\{2\} = [2\ 3\ 6];$$

Tabela 35 - Demonstração em formato de tabela do bicluster OPSM de número 2, considerando o zero como dado faltante, obtido através da Tabela 6.

Espécies	URBANA		RURAL
	CARNE	SARDINHA	SARDINHA
	<b>2</b>	<b>3</b>	<b>6</b>
<i>A. orientalis</i>	<b>1</b>	225	132
<i>S. nudiseta</i>	<b>2</b>	17	15
<i>M. domestica</i>	<b>3</b>	26	5
<i>O. chalcogaster</i>	<b>4</b>	21	4

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *S. nudiseta*, *M. domestica* e *O. chalcogaster* : 6 - Rural (sardinha)  $\leq$  3 - Urbana (sardinha)  $\leq$  2 - Urbana (carne).

Abaixo, seguem os seis *biclusters* do tipo OPSM considerando o zero como dado faltante para a família **Calliphoridae**. Como observou-se na **Análise 1**, nota-se a associação entre *L. eximia* e *C. megacephala* em todos os *biclusters* encontrados.

$$* A\{1\} = [1\ 2\ 4]; B\{1\} = [1\ 5\ 7]$$

Tabela 36 - Demonstração em formato de tabela do bicluster OPSM de número 1, considerando o zero como dado faltante, obtido através da Tabela 7.

Espécies	URBANA		FLORESTA
	RURAL	CARNE	FÍGADO
	<b>1</b>	<b>5</b>	<b>7</b>
<i>L. eximia</i>	<b>1</b>	127	23
<i>H. segmentaria</i>	<b>2</b>	1	1
<i>C. megacephala</i>	<b>4</b>	17	3

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. segmentaria* e *C. megacephala* : 5 - Rural (carne)  $\leq$  1 - Urbana (fígado)  $\leq$  7 - Floresta (fígado).

$$* A\{2\} = [1\ 2\ 4]; B\{2\} = [1\ 5\ 9];$$

Tabela 37 - Demonstração em formato de tabela do bicluster OPSM de número 2, considerando o zero como dado faltante, obtido através da Tabela 7.

Espécies	URBANA		FLORESTA	
	RURAL	FÍGADO	CARNE	SARDINHA
		<b>1</b>	<b>5</b>	<b>9</b>
<i>L. eximia</i>	<b>1</b>	127	23	30
<i>H. segmentaria</i>	<b>2</b>	1	1	25
<i>C. megacephala</i>	<b>4</b>	17	3	10

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. segmentaria* e *C. megacephala* : 5 - Rural (carne) ≤ 9 - Floresta (sardinha) ≤ 1 - Urbana (fígado).

$$* A\{3\} = [1\ 2\ 4]; B\{3\} = [5\ 8\ 9];$$

Tabela 38 - Demonstração em formato de tabela do bicluster OPSM de número 3, considerando o zero como dado faltante, obtido através da Tabela 7.

Espécies	RURAL		FLORESTA	
	FLORESTA	CARNE	CARNE	SARDINHA
		<b>5</b>	<b>8</b>	<b>9</b>
<i>L. eximia</i>	<b>1</b>	23	55	30
<i>H. segmentaria</i>	<b>2</b>	1	8	25
<i>C. megacephala</i>	<b>4</b>	3	42	10

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. segmentaria* e *C. megacephala* : 5 - Rural (carne) ≤ 9 - Floresta (sardinha) ≤ 8 - Floresta (carne).

$$* A\{4\} = [1 \ 4 \ 5]; B\{4\} = [2 \ 5 \ 8];$$

Tabela 39 - Demonstração em formato de tabela do bicluster OPSM de número 4, considerando o zero como dado faltante, obtido através da Tabela 7.

Espécies	URBANA		RURAL FLORESTA	
	CARNE	CARNE	CARNE	CARNE
		<b>2</b>	<b>5</b>	<b>8</b>
<i>L. eximia</i>	<b>1</b>	26	23	55
<i>C. megacephala</i>	<b>4</b>	9	3	42
<i>C. albiceps</i>	<b>5</b>	2	1	63

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *C. megacephala* e *C. albiceps* : 5 - Rural (carne)  $\leq$  2 – Urbana (carne)  $\leq$  8 – Floresta (carne).

$$* A\{5\} = [1 \ 4 \ 5]; B\{5\} = [2 \ 3 \ 5 \ 9];$$

Tabela 40 - Demonstração em formato de tabela do bicluster OPSM de número 5, considerando o zero como dado faltante, obtido através da Tabela 7.

Espécies	URBANA				
	RURAL		FLORESTA		
	CARNE	SARDINHA	CARNE	SARDINHA	
		<b>2</b>	<b>3</b>	<b>5</b>	<b>9</b>
<i>L. eximia</i>	<b>1</b>	26	10	23	30
<i>C. megacephala</i>	<b>4</b>	9	11	3	10
<i>C. albiceps</i>	<b>5</b>	2	25	1	39

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *C. megacephala* e *C. albiceps* : 3 – Urbana (sardinha)  $\leq$  5 - Rural (carne)  $\leq$  2 – Urbana (carne)  $\leq$  9 – Floresta (sardinha).

$$* A\{6\} = [1 \ 4 \ 5]; B\{6\} = [2 \ 5 \ 7 \ 9];$$

Tabela 41 - Demonstração em formato de tabela do bicluster OPSM de número 5, considerando o zero como dado faltante, obtido através da Tabela 7.

Espécies	URBANA		RURAL		FLORESTA	
	1	2	3	4	5	6
<i>L. eximia</i>	1	26	23	469	30	
<i>C. megacephala</i>	4	9	3	27	10	
<i>C. albiceps</i>	5	2	1	151	39	

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *C. megacephala* e *C. albiceps* : 5 - Rural (carne) ≤ 2 – Urbana (carne) ≤ 9 – Floresta (sardinha) ≤ 7 – Floresta (fígado).

Os *biclusters* de número um, dois e três relacionam as mesmas espécies (*L. eximia*, *H. segmentaria* e *C. megacephala*), sendo que o ambiente Rural com a isca carne é onde há a menor incidência dessas espécies. Da mesma forma, os demais *biclusters* ( quatro, cinco e seis) também apresentaram as mesmas espécies ( *L. eximia*, *C. megacephala* e *C. albiceps*) e nesse caso, o ambiente de Floresta foi o local onde elas foram encontradas em maior número.

### Análise 3

Através da Tabela 9, foram obtidos um total de **20 biclusters do tipo OPSM, considerando o zero como dado faltante**. Eles estão listados abaixo:

$$* A\{1\} = [1\ 2\ 3\ ]; B\{1\} = [1\ 2\ 3\ 6\ ];$$

Tabela 42 - Demonstração em formato de tabela do bicluster OPSM de número 1, considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	URBANA			RURAL	
	1	2	3	4	5
<i>A. orientalis</i>	1	14	225	132	13
<i>S. nudiseta</i>	2	1	17	15	1
<i>M. domestica</i>	3	1	26	5	1

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *S. nudiseta* e *C. albiceps* : 6 – Rural (sardinha) ≤ 1 – Urbana (fígado) ≤ 3- Urbana (sardinha) ≤ 2 – Urbana (carne).

$$* A\{2\} = [6 \ 9 \ 12]; B\{2\} = [1 \ 2 \ 5];$$

Tabela 43 - Demonstração em formato de tabela do bicluster OPSM de número 2, considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	URBANA		RURAL	
	FÍGADO	CARNE	CARNE	
		<b>1</b>	<b>2</b>	<b>5</b>
<i>L.eximia</i>	<b>6</b>	127	26	23
<i>C. megacephala</i>	<b>9</b>	17	9	3
<i>O. thornax</i>	<b>12</b>	10	6	4

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L.eximia*, *C. megacephala* e *O. thornax* : 5 – Rural (carne) ≤ 2 – Urbana (carne) ≤ 1- Urbana (fígado).

$$* A\{3\} = [6 \ 9 \ 12]; B\{3\} = [1 \ 5 \ 9];$$

Tabela 44 - Demonstração em formato de tabela do bicluster OPSM de número 3, considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	URBANA	RURAL	FLORESTA	
	FÍGADO	CARNE	CARNE	
		<b>1</b>	<b>5</b>	<b>9</b>
<i>L. eximia</i>	<b>6</b>	127	23	30
<i>C. megacephala</i>	<b>9</b>	17	3	10
<i>O. thornax</i>	<b>12</b>	10	4	5

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *C. megacephala* e *O. thornax* : 5 – Rural (carne) ≤ 9 –Floresta (carne) ≤ 1- Urbana (fígado).



\*  $A\{4\} = [6\ 7\ 9\ 12\ ]$ ;  $B\{4\} = [5\ 7\ 9\ ]$ ;

Tabela 45 - Demonstração em formato de tabela do bicluster OPSM de número 4, considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	RURAL		FLORESTA	
	CARNE	FÍGADO	SARDINHA	
	<b>5</b>	<b>7</b>		<b>9</b>
<i>L. eximia</i>	<b>6</b>	23	469	30
<i>H. segmentaria</i>	<b>7</b>	1	50	25
<i>C. megacephala</i>	<b>9</b>	1	15	39
<i>O. thornax</i>	<b>12</b>	4	7	5

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. segmentaria*, *C. megacephala* e *O. thornax*: 5 – Rural (carne) ≤ 9 – Floresta (sardinha) ≤ 7- Floresta (fígado).

\*  $A\{5\} = [9\ 10\ 11\ 12\ ]$ ;  $B\{5\} = [2\ 3\ 5\ ]$ ;

Tabela 46 - Demonstração em formato de tabela do bicluster OPSM de número 5, considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	URBANA		RURAL
	CARNE	SARDINHA	CARNE
	<b>2</b>	<b>3</b>	<b>5</b>
<i>C. megacephalax</i>	<b>9</b>	9	11
<i>C. albiceps</i>	<b>10</b>	2	25
<i>P. (S) lambens</i>	<b>11</b>	8	27
<i>O. thornax</i>	<b>12</b>	6	19

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *C. megacephala*, *C. albiceps*, *P. (S) lambens* e *O. thornax*: 5 – Rural (carne) ≤ 2 – Urbana (carne) ≤ 3- Urbana (sardinha).

$$* A\{6\} = [6\ 7\ 9]; B\{6\} = [1\ 5\ 7];$$

Tabela 47 - Demonstração em formato de tabela do bicluster OPSM de número 6, considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	URBANA RURAL FLORESTA		
	FÍGADO	CARNE	CARNE
	<b>1</b>	<b>5</b>	<b>7</b>
<i>L. eximia</i>	<b>6</b>	9	11
<i>H. segmentaria</i>	<b>7</b>	2	25
<i>C. megacephala</i>	<b>9</b>	8	27

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. segmentaria*, *C. megacephala*: 7 – Floresta (carne) ≤ 1 – Urbana (fígado) ≤ 5- Rural (carne).

$$* A\{7\} = [9\ 10\ 11]; B\{7\} = [5\ 7\ 8];$$

Tabela 48 - Demonstração em formato de tabela do bicluster OPSM de número 7, considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	RURAL FLORESTA		
	CARNE	FÍGADO	CARNE
	<b>5</b>	<b>7</b>	<b>8</b>
<i>C. megacephala</i>	<b>9</b>	3	27
<i>C. albiceps</i>	<b>10</b>	1	15
<i>P. (S) lambens</i>	<b>11</b>	2	7

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *C. megacephala*, *C. albiceps*, *P. (S) lambens*: 5 – Rural (carne) ≤ 7 – Floresta (fígado) ≤ 8- Floresta (carne).

$$* A\{8\} = [10 \ 11 \ 12]; B\{8\} = [3 \ 5 \ 7];$$

Tabela 49 - Demonstração em formato de tabela do bicluster OPSM de número 8 considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	URBANA	RURAL	FLORESTA
	SARDINHA	CARNE	CARNE
	<b>3</b>	<b>5</b>	<b>7</b>
<i>C. albiceps</i>	<b>10</b>	25	1
<i>P. (S) lambens</i>	<b>11</b>	27	2
<i>O. thornax</i>	<b>12</b>	19	4

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *C. albiceps*, *P. (S) lambens* e *O. thornax*: 5 – Rural (carne) ≤ 7 – Floresta (carne) ≤ 3- Urbana (sardinha).

$$* A\{9\} = [9 \ 11 \ 12]; B\{9\} = [3 \ 5 \ 9];$$

Tabela 50 - Demonstração em formato de tabela do bicluster OPSM de número 9 considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	URBANA	RURAL	FLORESTA
	SARDINHA	CARNE	SARDINHA
	<b>3</b>	<b>5</b>	<b>9</b>
<i>C. megacephala</i>	<b>9</b>	11	3
<i>P. (S) lambens</i>	<b>11</b>	27	2
<i>O. thornax</i>	<b>12</b>	19	4

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *C. megacephala*, *P. (S) lambens* e *O. thornax*: 5 – Rural (carne) ≤ 9 – Floresta (sardinha) ≤ 3- Urbana (sardinha).

\*  $A\{10\} = [1\ 2\ 3\ 4\ 6]$ ;  $B\{10\} = [2\ 3\ 6]$ ;

Tabela 51 - Demonstração em formato de tabela do bicluster OPSM de número 10 considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	URBANA		RURAL	
	CARNE	SARDINHA	SARDINHA	
		<b>2</b>	<b>3</b>	<b>6</b>
<i>A. orientalis</i>	<b>1</b>	225	132	13
<i>S.nudiseta</i>	<b>2</b>	17	15	1
<i>M. domestica</i>	<b>3</b>	26	5	1
<i>O. chalcogaster</i>	<b>4</b>	21	4	1
<i>L. eximia</i>	<b>6</b>	26	10	3

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *S.nudiseta*, *M. domestica*, *O. chalcogaster* e *L. eximia*: 6 – Rural (sardinha) ≤ 3 – Urbana (sardinha) ≤ 2- Urbana (carne).

\*  $A\{11\} = [3\ 4\ 6]$ ;  $B\{11\} = [3\ 4\ 6]$ ;

Tabela 52 - Demonstração em formato de tabela do bicluster OPSM de número 11 considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	URBANA		RURAL	
	SARDINHA	FÍGADO	SARDINHA	
		<b>3</b>	<b>4</b>	<b>6</b>
<i>M. domestica</i>	<b>3</b>	5	29	1
<i>O. chalcogaster</i>	<b>4</b>	4	9	1
<i>L. eximia</i>	<b>6</b>	10	51	3

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *M. domestica*, *O. chalcogaster* e *L. eximia*: 6 – Rural (sardinha) ≤ 3 – Urbana (sardinha) ≤ 4- Rural (fígado).

$$* A\{12\} = [1\ 3\ 4\ 6]; B\{12\} = [3\ 5\ 6];$$

Tabela 53 - Demonstração em formato de tabela do bicluster OPSM de número 12 considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	URBANA		RURAL	
	SARDINHA	CARNE	SARDINHA	
		<b>3</b>	<b>5</b>	<b>6</b>
<i>A. orientalis</i>	<b>1</b>	132	329	13
<i>M. domestica</i>	<b>3</b>	5	39	1
<i>O. chalcogaster</i>	<b>4</b>	4	17	1
<i>L. eximia</i>	<b>6</b>	10	23	3

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *M. domestica*, *O. chalcogaster* e *L. eximia*: 6 – Rural (sardinha) ≤ 3 – Urbana (sardinha) ≤ 4- Rural (carne).

$$* A\{13\} = [4\ 11\ 12]; B\{13\} = [2\ 4\ 5];$$

Tabela 54 - Demonstração em formato de tabela do bicluster OPSM de número 13 considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	URBANA		RURAL	
	CARNE	FÍGADO	CARNE	
		<b>2</b>	<b>4</b>	<b>5</b>
<i>O. chalcogaster</i>	<b>4</b>	21	9	17
<i>P. (S) lambens</i>	<b>11</b>	8	1	2
<i>O. thornax</i>	<b>12</b>	6	2	4

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *O. chalcogaster*, *P. (S) lambens* e *O. thornax*: 4 – Rural (fígado) ≤ 5– Rural (carne) ≤ 2- Urbana (carne).

\*  $A\{14\} = [1\ 11\ 12\ ]$ ;  $B\{14\} = [3\ 4\ 9\ ]$ ;

Tabela 55 - Demonstração em formato de tabela do bicluster OPSM de número 14 considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	URBANA		RURAL	FLORESTA
	SARDINHA		FÍGADO	SARDINHA
		<b>3</b>	<b>4</b>	<b>9</b>
<i>A. orientalis</i>	<b>1</b>	132	5	39
<i>P. (S) lambens</i>	<b>11</b>	27	1	13
<i>O. thornax</i>	<b>12</b>	19	2	5

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *P. (S) lambens* e *O. thornax*: 4 – Rural (fígado) ≤ 9 – Floresta (sardinha) ≤ 4- Rural (fígado).

\*  $A\{15\} = [6\ 9\ 10\ 12\ ]$ ;  $B\{15\} = [2\ 5\ 7\ ]$ ;

Tabela 56 - Demonstração em formato de tabela do bicluster OPSM de número 15 considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	URBANA		RURAL	FLORESTA
	CARNE		CARNE	FÍGADO
		<b>2</b>	<b>5</b>	<b>7</b>
<i>L. eximia</i>	<b>6</b>	26	23	469
<i>C. megacephala</i>	<b>9</b>	9	3	27
<i>C. albiceps</i>	<b>10</b>	2	1	15
<i>O. thornax</i>	<b>12</b>	6	4	7

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *C. megacephala*, *C. albiceps* e *O. thornax*: 5 – Rural (carne) ≤ 2 – Urbana (carne) ≤ 7- Floresta (fígado).

$$* A\{16\} = [6\ 9\ 10\ 11]; B\{16\} = [2\ 5\ 8\ 9];$$

Tabela 57 - Demonstração em formato de tabela do bicluster OPSM de número 16 considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	URBANA		RURAL		FLORESTA	
		CARNE		CARNE	CARNE	SARDINHA
		<b>2</b>		<b>5</b>	<b>8</b>	<b>9</b>
<i>L. eximia</i>	<b>6</b>	26	23	55	30	
<i>C. megacephala</i>	<b>9</b>	9	3	42	10	
<i>C. albiceps</i>	<b>10</b>	2	1	63	39	
<i>P. (S) lambens</i>	<b>11</b>	8	2	16	13	

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *C. megacephala*, *C. albiceps* e *P. (S) lambens*: 5 – Rural (carne) ≤ 2 – Urbana (carne) ≤ 9- Floresta (sardinha) ≤ 8 – Floresta (carne).

$$* A\{17\} = [9\ 10\ 14]; B\{17\} = [2\ 7\ 8];$$

Tabela 58 - Demonstração em formato de tabela do bicluster OPSM de número 17 considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	URBANA		FLORESTA	
		CARNE	FÍGADO	CARNE
		<b>2</b>	<b>7</b>	<b>8</b>
<i>C. megacephala</i>	<b>9</b>	9	3	27
<i>C. albiceps</i>	<b>10</b>	2	1	15
<i>P (E) collusor</i>	<b>14</b>	6	4	7

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *C. megacephala*, *C. albiceps* e *P. (E) collusor*: 7 – Floresta (fígado) ≤ 2 – Urbana (carne) ≤ 8- Floresta (carne).

\*  $A\{18\} = [1\ 6\ 11\ 13]$ ;  $B\{18\} = [6\ 8\ 9]$ ;

Tabela 59 - Demonstração em formato de tabela do bicluster OPSM de número 18 considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	RURAL		FLORESTA	
	SARDINHA	CARNE	SARDINHA	CARNE
		<b>6</b>	<b>8</b>	<b>9</b>
<i>A. orientalis</i>	<b>1</b>	13	137	39
<i>L. eximia</i>	<b>6</b>	3	55	30
<i>P. (S) lambens</i>	<b>11</b>	6	16	13
<i>P (Euboettcheria) sp</i>	<b>13</b>	1	20	20

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *L. eximia*, *P. (S) lambens* e *P. (Euboettcheria) sp*: 6 – Rural (sardinha) ≤ 9 – Floresta (sardinha) ≤ 8- Floresta (carne).

\*  $A\{19\} = [10\ 11\ 13]$ ;  $B\{19\} = [7\ 8\ 9]$ ;

Tabela 60 - Demonstração em formato de tabela do bicluster OPSM de número 19 considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies	FLORESTA			
	FÍGADO	CARNE	SARDINHA	
		<b>7</b>	<b>8</b>	<b>9</b>
<i>C. albiceps</i>	<b>10</b>	15	63	39
<i>P. (S) lambens</i>	<b>11</b>	7	16	13
<i>P (Euboettcheria) sp</i>	<b>13</b>	5	20	20



que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *C. albiceps*, *P. (S) lambens* e *P. (Euboettcheria) sp*: 7 – Floresta (fígado) ≤ 9 – Floresta (sardinha) ≤ 8- Floresta (carne).

\* A{20} = [6 8 15 ]; B{20} = [7 8 9 ];

Tabela 61 - Demonstração em formato de tabela do bicluster OPSM de número 20 considerando o zero como dado faltante, obtido através da Tabela 9.

Espécies		FLORESTA		
		FÍGADO	CARNE	SARDINHA
		<b>7</b>	<b>8</b>	<b>9</b>
<i>L. eximia</i>	<b>6</b>	469	55	30
<i>H. semidiaphana</i>	<b>8</b>	26	13	1
<i>P (P) intermutans</i>	<b>15</b>	23	10	8

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. semidiaphana* e *P (P) intermutans* : 9 – Floresta (sardinha) ≤ 8 – Floresta (carne) ≤ 7- Floresta (fígado).

Analisando os 20 *biclusters* encontrados, notou-se que:

- Seis mostraram associação entre *C. megacephala* e *O. thornax* e o local Rural com isca de carne foi onde essas espécies foram encontradas em menor número.
- Seis mostraram associação entre *L. eximia* e *C. megacephala*.
- Cinco apresentaram as espécies *C. albiceps* e *P. (S) lambens* conjuntamente.
- Os biclusters de número um e dois relacionam as mesmas espécies ( *L. eximia*, *C. megacephala* e *O. thornax*), sendo os locais de menor e maior coleta iguais em ambos.

## DISCUSSÃO

Os resultados obtidos através dos *biclusters*, utilizando tanto dados binários como quantitativos, demonstraram a relevância dessa ferramenta na busca por relações entre as

espécies aqui consideradas. Uma bastante recorrente foi a entre os muscódeos *A. orientalis* e *M. domestica*. Por serem espécies conhecidas por sua preferência por ambientes modificados pelo homem (LINHARES, 1979; D'ALMEIDA, 1983, SOUZA, 2011), os *biclusters* encontrados não as relacionaram conjuntamente em ambiente florestal. Em um estudo sobre nichos tróficos, D'almeida & D'almeida (1998) perceberam que o nicho de *A. orientalis* foi mais extenso na área urbana e o da *M. domestica* foi na área rural, observando que não houve sobreposição de nichos entre elas. Além disso, na penúltima análise considerada nesse trabalho, a espécie eussintrópica *S. nudiseta* (CARVALHO et al., 2002; D'ALMEIDA, 1992; LEANDRO & D'ALMEIDA, 2005, LINHARES, 1981; MENDES & LINHARES, 1993;) foi encontrada associada à essas espécies somente na área urbana e rural.

As espécies da família Calliphoridae são típicas de diferentes ambientes podendo ser rurais, urbanos ou florestais, muitas delas exibindo alta sinantropia. Como exemplo daquelas com uma alta dependência pela antropobiocenose, citam-se *L. eximia* (FERREIRA, 1983; D'ALMEIDA; LOPES, 1983), e *C. megacephala* (LINHARES, 1981, D'ALMEIDA; LOPES, 1983; VIANNA et al, 1998; RODRIGUES-GUIMARÃES et al , 2008). Por outro lado, no presente estudo, os *biclusters* indicaram uma associação entre essas espécies ocorrendo com maior frequência na mata, atraídas pelas isca de carne ou fígado. Isso possivelmente ocorreu devido ao crescimento populacional ocorrido ao redor da floresta estudada, dando-a características próprias de local habitado pelo homem. Em relação às iscas, ambas já são conhecidas pela sua importância na atratividade de califorídeos (D'ALMEIDA; FRAGA, 2007; CASTRO et al, 2010; SOUZA, 2011).

Em outra análise, *C. albiceps* foi encontrada conjuntamente com *L. eximia* e *C. megacephala*, também em maior quantidade na área de floresta, ambiente ao qual ela já foi encontrada (ALVES et al, 2014). Além disso, os *biclusters* da última análise encontraram associação de *O. thornax* também com *L. eximia* e *C. megacephala* em maior grau na área urbana. Cabe aqui salientar a importância dessa espécie na entomologia forense, visto que ela já foi coletada visitando cadáveres humanos no Brasil (BARROS et al., 2008). Também observou-se uma associação de *L. eximia* e *C. megacephala* com *H. segmentaria*, espécies cujo conhecimento de suas biologias são importantes na datação no intervalo pós morte (SALVIANO et al., 1996).

Também em maior ocorrência na floresta, alguns *biclusters* encontraram associadas *P. (S.) lambens* e *P. (E.) anguilla* sendo atraídas pelas iscas de carne e sardinha. Apesar da última isca não ser algo típico de ser encontrado em ambiente florestal, sabe-se que ela têm sido muito utilizada em trabalhos de levantamentos faunísticos de muscóides (FERREIRA,

1978; LINHARES, 1981; D'ALMEIDA; LOPES, 1983; CARRARO; MILWARD-DE-AZEVEDO, 1999, MARINHO et al., 2003, OLIVEIRA et al, 2002; SILVA et al, 2005; D'ALMEIDA; FRAGA, 2007), inclusive na mata (LOPES, 1973; D'ALMEIDA; LOPES, 1983). A presença de *P. (E) anguilla* na floresta já foi citada por muitos autores (MONTEIRO-FILHO; PENNEREIRO, 1987; CARVALHO; LINHARES, 2001; BARROS et al. 2008; MORETTI et al. 2008; ROSA et al. 2011) e *S. lambens* já foi coletada em reserva florestal (CARVALHO, LINHARES, 2001).

Outras associações entre as espécies do gênero *Peckia* ocorrendo com maior frequência na área de mata, como entre *P. (S) ingens* e *P. (E) anguilla*, entre essas espécies com *P. (E) collusor* e também com *P. (S) lambens* foram demonstradas através das análises dos *biclusters*. Em uma compilação dos registros das principais espécies de moscas encontradas em carcaças e cadáveres na região Neotropical, Alves et al (2014) demonstraram que há relatos da presença das espécies aqui mencionadas no ambiente de floresta. Por outro lado, quando *P. (S) lambens* e *P. (E) anguilla* foram encontradas relacionadas a *O. thornax*, essas espécies tiveram maior ocorrência na área urbana.

## CONCLUSÃO

A aplicação dos *biclusters* em dados de dípteros, apesar de ser inovadora, trouxe para o presente trabalho informações relevantes sobre como as espécies de dípteros necrófagos, no contexto estudado, se relacionam.

A partir da ferramenta aplicada, pôde-se encontrar padrões de associações entre as espécies de dípteros, mediante a isca e o local em que elas foram coletadas. Por isso, os *biclusters* podem ser considerados importantes na busca de informações, tendências e padrões dentro de uma base de dados de dípteros.

## REFERÊNCIAS

ABDULLAH, A.; HUSSAIN, A. (2006), ‘A new biclustering technique based on crossing minimization’, **Neurocomputing** 69 (16–18), 1882 – 1896.

ALVES, A.C.F.; SANTOS, W.E.; CREÃO-DUARTE, A.J., 2014. Diptera (Insecta) de importância forense da região Neotropical. **Entomotropica** 29, 77-94.

ANDREWS, S. 2009. In-Close, a fast algorithm for computing formal concepts. In: **International Conference on Conceptual Structures (ICCS)**, Moscow, 2009.

ANDREWS, S. **In-close2, a high performance formal concept miner**. ICCS 2011, 2011.

ANDREWS, S.; ORPHANIDES, C. (2010). Analysis of large data sets using formal concept lattices. In: KRYSZKIEWICZ, M. and OBIEDKOV, S, (eds.) **Proceedings of the 7th International Conference on Concept Lattices and Their Applications**. Seville, University of Seville, 104-115.

ANKERST, M., BREUNIG, M., M., KRIEGEL, H.-P., et al. OPTICS: Ordering Points to Identify the Clustering Structure , In: **Proceedings of the ACM SIGMOD Conference on Management of Data**, pp. 49-60, Philadelphia, PA, USA, June, 1999

BARROS, R.M., MELLO-PATIU, C.A. & PUJOL-LUZ, J.R. 2008. Sarcophagidae (Insecta, Diptera) associados à decomposição de carcaças de *Sus scrofa* Linnaeus (Suidae) em área de Cerrado do Distrito Federal, Brasil. **Rev. Bras. Entomol.** 52: 606–609

BEN-DOR, A.; CHOR, B.; KARP, R.; YAKHINI, Z. “Discovering local structure in gene expression data: the order-preserving submatrix problem.” **Proc. of the 6th Annual International Conference on Computational Biology**, 2002: 49-57.

BERKHIN, P. 2006. **A survey of clustering data mining techniques**. Grouping Multidimensional Data, (pp. 25–71).

BISSON G, HUSSAIN F. (2008), Chi-sim: A new similarity measure for the co-clustering task, in ‘**Proceedings of the 2008 Seventh International Conference on Machine Learning and Applications**’, ICMLA '08, IEEE Computer Society, Washington, DC, USA, pp. 211–217.

BLEULER, S., PRELIĆ, A.; ZITZLER, E E. “An EA framework for biclustering of gene expression data.” *Proc. of IEEE Congress on Evolutionary Computation (CEC)*, 2004: 166-173.

CALIFANO, A., STOLOVITZKY, G.; TU, Y. Analysis of gene expression microarrays for phenotype classification. **Proceedings of the International Conference on Intelligent Systems for Molecular Biology**, 2000. 8: p. 75-85.

CARLANTONIO, L. M. **Novas Metodologias para Clusterização de Dados**, Coordenação de Programas de Pós-Graduação em Engenharia, COPPE/UFRJ, 2001.

CARRARO, V.M.; MILWARD-DE-AZEVEDO, E.M.V. 1999. Quantitative description of calliphorid dipterans captured on Campus of the Federal Rural University of Rio de Janeiro using sardine bait. **Rev. Bras. Zootecni.**,v. 1, n. 1, p. 77-89,

CARVALHO, L. M. L.; A. X. LINHARES. 2001. Seasonality of insect succession and pig carcass decomposition in a natural forest area in southeastern Brazil. **J. Forensic Sciences** 46: 604608.

CARVALHO, C. J. B. DE; MOURA, M. O. & RIBEIRO, P. B. 2002. Chave para adultos de dípteros (Muscidae, Fanniidae, Anthomyiidae) associados ao ambiente humano no Brasil. **Rev. Brasil. Entomol.** 46(1):107-114.

CASTRO, R.C.S. et al. Levantamento de dípteros de interesse forense em uma área de cerrado em Brasília. **Universitas: Ciências da Saúde**, Brasília, v.8, p.1-16, 2010.

CHENG, Y.; CHURCH, G. M. (2000). Biclustering of expression data. Em *Proc. of the 8th Int. Conf. on Intelligent Systems for Molecular Biology*, (pp. 93–103).

CHO, H. et al. "Minimum Sum-Squared Residue Cocustering of Gene Expression Data," **Proc. Fourth SIAM Int'l Conf. Data Mining**, 2004.

COELHO, G.; DE FRANÇA, F.; VON ZUBEN, F. Multi-Objective Biclustering: When Non-dominated Solutions are not Enough. **J. Math. Model. Algor.**, v. 8, n. 2, p. 175-202, 2009.

D'ALMEIDA, J. M. **Sinantropia de dípteros caliptratos na área metropolitana do Rio de Janeiro**. Tese de Mestrado, Universidade Federal Rural do Rio de Janeiro, 1983. 193p.

D'ALMEIDA, J.M.; LOPES, H.S. Sinantropia de Dípteros Muscóides (Calliphoridae) no Estado do Rio de Janeiro. **Arq. Univ. Fed. Rural Rio de Jan.**, v. 6, n. 1, p. 39-48, 1983

D'ALMEIDA, J. M. Calyptrate Diptera (Muscidae and Anthomyidae) of the State of Rio de Janeiro - I. Synanthropy. **Memórias do Instituto Oswaldo Cruz** 87: 381- 386, 1992.

D'ALMEIDA J.M.; FRAGA, M.B. (2007) Efeito de diferentes iscas na atração de califorídeos (Diptera) no campus do Valonguinho, Universidade Federal Fluminense, Niterói, RJ, Brasil. **Rev Bras Parasitol Vet** 16:199–204

DAS, M.; BORAH, B. Biclustering of Gene Expression Data using a Two - Phase Method. **Internat. J. Comp. Applic.**, v. 103, n. 13, p. 6-10, 2014.

DAO, P, et al. "Inferring cancer subnetwork markers using density-constrained biclustering." **Bioinformatics** 26.18 (2010): i625-i631

DE CASTRO, P. A. D. et al (2007a). Applying Biclustering to Perform Collaborative Filtering. In **Proc. of the 7th International Conference on Intelligent Systems Design and Applications** (pp. 421–426). Rio de Janeiro, Brazil.

DE CASTRO, P. A. D. et al (2007b.) Applying Biclustering to Text Minin

g: An Immune-Inspired Approach. In L. N. de Castro, F. J. Von Zuben, & H. Knidel (Eds.), **Artificial Immune Systems, Proc. of the 6th International Conference on Artificial**

**Immune Systems (ICARIS)**, volume 4628 of *Lecture Notes in Computer Science* (pp. 83–94). Santos, Brazil.

DE CASTRO, P. A. D. et al (2007c). Evaluating the Performance of a Biclustering Algorithm Applied to Collaborative Filtering: A Comparative Analysis. In **Proc. of the 7th International Conference on Hybrid Intelligent Systems** (pp. 65–70). Kaiserslautern, Germany.

DE FRANÇA, F. O.; BEZERRA, G. P.; VON ZUBEN, F. J. (2006). New Perspectives for the Biclustering Problem. **Em Evolutionary Computation**, 2006. CEC 2006. IEEE Congress on, (pp. 753–760). IEEE.

DE FRANÇA, F. O., E VON ZUBEN F. J. **Algoritmo multi-objetivo e multipopulacional para biclusterização**. Anais do Segundo Encontro dos Alunos e Docentes do Departamento de Engenharia de Computação e Automação Industrial, Campinas, 2009. Disponível em >[https://www.academia.edu/456366/Algoritmo\\_Multi-Objetivo\\_E\\_Multipopulacional\\_Para\\_Biclusteriza%C3%A7%C3%A3o](https://www.academia.edu/456366/Algoritmo_Multi-Objetivo_E_Multipopulacional_Para_Biclusteriza%C3%A7%C3%A3o). Acesso em 10 ago 2015.

DE FRANÇA, F. O. **Biclusterização na análise de dados incertos**. Tese de Doutorado, Faculdade de Engenharia Elétrica e de Computação, UNICAMP, Campinas, 2010.

DIVINA, F.; AGUILAR-RUIZ, E J. S. Biclustering of expression data with evolutionary computation. **IEEE Transactions on Knowledge and Data Engineering** 18, 2006: 590-602.

DUFFY, D.; QUIROZ, A.. A permutation-based algorithm for block clustering. **J. Classific.**, 1991. 8(1).

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; & UTHURUSAMY, R. “Advances in Knowledge Discovery and Data Mining” 1996, AAAIPress, The Mit Press.

FAURE, N., CHRETIENNE, P., GOURDIN, E.; SOURD, F. (2007), ‘Biclique completion problems for multicast network design’, *Discrete Optimization* 4 (3-4), 360–377.

FELDMAN, R; SANGER, J. (2006). *The Text Mining Handbook*. Cambridge University Press.

FERREIRA, M. J. M. Sinantropia de dípteros muscóides de Curitiba, Paraná. I: Calliphoridae. **Rev. Bras. Biol.**, 38: 445-454, 1978.

FERREIRA, M. J. M. Sinantropia de Calliphoridae (Diptera) em Goiânia, Góias. **Revista Brasileira de Biologia**, 43: 199-210, 1983.

FREITAS, A. A. A critical review of multi-objective optimization in datamining: a position paper. *SIGKDD Explor. Newsl.* , 6(2):77–86, 2004

GETZ, G., LEVINE, E.; DOMANY, E. Coupled two-way clustering analysis of gene microarray data. *Proc Natural Academy of Sciences USA*, 2000. 97(22): p. 12079-12084.

HANCZAR, B., NADIF, E M. Using the bagging approach for biclustering of gene expression data. *Neurocomputing 74*, 2011: 1595-1605.

HARTIGAN, J. A. (1972). Direct clustering of a data matrix. *Journal of the American Statistical Association (JASA)*, 67(337), 123–129.

HRUSCHKA, E. R., EBECKEN, N. F. F. A Genetic algorithm for cluster analysis , Submitted to: *IEEE Transactions on Evolutionary Computation* , January 2001

JAIN, A.K.; MURTY, M.N.; FLYNN, P.J. (1999) Data Clustering: A Review, *ACM Computing Surveys*, vol. 31, no. 3, pp. 264-323.

KLUGER Y; et al. Spectral biclustering of microarray data: coclustering genes and conditions. *Genome Res*, 2003. 13(4): p. 703-716.

KUZNETSOV, S. “Learning of simple conceptual graphs from positive and negative examples,” in *Principles of Data Mining and Knowledge Discovery*. Springer, 1999, pp. 384–391.



IHMELS J., FRIEDLANDER, G.; BERGMANN, S.; SARIG, O; ZIV, Y; BARKAI, N. Revealing modular organization in the yeast transcriptional network. *Nat. Genet.* 2002;31:370-377.

LAZZERONI, L.; OWEN, A. (2000). Plaid models for gene expression data. Technical report, Stanford University.

LEANDRO, M. J. F.; D'ALMEIDA, J. M. Levantamento de Calliphoridae, Fanniidae, Muscidae e Sarcophagidae em um fragmento de mata na Ilha do Governador, Rio de Janeiro, Brasil. *Iheringia*, Série Zoologia, 95(4):377-381, 2005

LINHARES, A. X. **Sinantropia de dípteros muscóides de Campinas.** 1979. Tese de Mestrado. Universidade Estadual de Campinas, Campinas, SP, 129 p.

LINHARES, A. X. Synanthropy of Calliphoridae and Sarcophagidae (Diptera) in the city of Campinas, São Paulo, Brazil. *Rev. Brasil. Entomol.*, 25 (3) : 189-215, 1981.

LIU, J.; WANG, W. OP-Cluster: **Clustering by Tendency in High Dimensional Space**, in Proceedings of the Third IEEE International Conference on Data Mining. 2003, IEEE Computer Society. p. 187.

LOPES, H.S. Collecting and rearing Sarcophagidae flies (Diptera) in Brazil, during forty ears. *An. Acad. Bras. Ciênci.*, v. 45, n. 2, p. 279-291, 1973.

MADEIRA, S. C.; OLIVEIRA, A. L. (2004). Biclustering algorithms for biological data analysis: A survey. *IEEE Transactions on Computational Biology and Bioinformatics*, 1(1), 24-45.

MARINHO, C.R.; AZEVEDO, A.C.G.; AGUIAR-COELHO, V.M.A. Diversidade de Califorídeos (Díptera:Calliphoridae) em área urbana. Rio de Janeiro. *Entomol. Vector.*, v.10, n. 2, p. 185-199, 2003.

MENDES, J.; LINHARES, A.X. Atratividade por iscas, sazonalidade e desenvolvimento ovariano em várias espécies de Muscidae (Diptera). *Rev. Bras. Entomol.* 37: 289-297, 1993.

MIRKIN, B. *Mathematical Classification and Clustering*. Springer, 1996.

MITRA, S., BANKA, E H.. “Multi-objective evolutionary biclustering of gene expression data.” **Pattern Recognition Journal** 39, 2006: 2464-2477.

MONTEIRO-FILHO, E. L. A.; PENEIREIRO, J. L. 1987. Estudo da decomposição e sucessão sobre uma carcaça animal numa área do estado de São Paulo, Brasil. **Rev. Brasil. Bio.** 47: 289–295.

MORETTI, T.C., RIBEIRO, O.B., THYSSEN, P.J., SOLIS, D.R., 2008. Insects on decomposing carcasses of small rodents in a secondary forest in Southeastern Brazil. **Eur. J. Entomol.** 105, 691-696.

MURALI, T. M.; KASIF, E S.. “Extracting conserved gene expression motifs from gene expression data.” **Proc. of the 8th Pacific Symposium on Biocomputing**, 2003: 77-88.

NEPOMUCENO, J. A., TRONCOSO, A. ; AGUILAR-RUIZ, E J. S. “Evolutionary metaheuristic for biclustering based on linear correlations among genes.” **Proc. of the 25th ACM Symposium On Applied Computing (SAC)**, 2010: 1143-1147.

OLIVEIRA, S. H. G. **On bicluster aggregation and its benefits for enumerative solutions.** 2015.94 f. Dissertação de mestrado apresentada a Faculdade de Engenharia Elétrica de Computação da Universidade Estadual de Campinas para obtenção do título de Mestre em Engenharia Elétrica, na área de Engenharia de Computação. Disponível em: [http://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/theses/Dissertacao\\_Mestrado\\_Saul\\_lo.pdp#page=3&zoom=auto,-99,632](http://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/theses/Dissertacao_Mestrado_Saul_lo.pdp#page=3&zoom=auto,-99,632). Acesso em 22 jul 2015

OLIVEIRA, V. C.; MELLO, R. P. D'ALMEIDA, J. M.. Dípteros muscóides como vetores mecânicos de ovos de helmintos em jardim zoológico, Brasil. **Rev. Saúde Pública**, São Paulo , v. 36, n. 5, p. 614-620, Oct. 2002

PEETERS, R. The maximum edge biclique problem is NP-complete. **Discrete Applied Mathematics**, 131(3):651–654, 2003.

PINHEIRO, R G S; MARTINS, I. C.; OCHI, L. S.; SIMONETTI, L. G. **Métodos Exatos e Heurísticos para Biclusterização em Grafos.** XIV CLAIO/XLIV SBPO - Congresso Latino

Iberoamericano de Investigación Operativa/Simpósio Brasileiro de Pesquisa Operacional.  
Rio de Janeiro, Brazil, September 2012 .

PRELIĆ, A; BLEULER, S.; ZIMMERMANN, P.; WILLE, A.; BÜHLMANN, P.; Gruissem W.; Hennig, L.; Thiele, L.; Zitzler, E.; A systematic comparison and evaluation of biclustering methods for gene expression data. **Bioinformatics** 22(9), 2006: 1122-1129.

PRESTES, E. **Teoria dos Grafos**. [2012]. Disponível em: <http://www.inf.ufrgs.br/~prestes/Courses/Graph%20Theory/GrafosA2.pdf>. Acesso em 10 de jan de 2014.

RODRIGUES-GUIMARÃES R, et al. Sinantropia da fauna de califorídeos (Diptera, Calliphoridae) na Baixada Fluminense, Rio de Janeiro, Brasil. **Rev. Ciênc. Tecnol.** 8(1):22–33, 2008.

ROSA, T.A.; BABATA, M.L.Y.; SOUZA, C.M.; SOUSA, D., MELLO-PATIU, C.A.; VAZ-DE-MELLO, F.Z.; MENDES, J. 2011. Arthropods associated with pig carrion in two vegetation profiles of Cerrado in the State of Minas Gerais, Brazil. **Rev. Brasil. Entomol.** 55 : 424–434.

SALVIANO, R.J.B.; et al. 1996. Calliphoridae (Diptera) associated with human corpses in Rio de Janeiro, Brazil. **Entomol.Vect.**3:145-146.

SANTAMARIA, R. (2009). Análisis **Visual de Datos de Expresion Génica Mediante Biclustering**. Tesis Doctoral. Universidad de Salamanca.

SEGAL, E. et al. (2001). Rich probabilistic models for gene expression. **Bioinformatics**, 17:243–252.

SHENG, Q.; MOREAU, Y.; DE MOOR, B. **Biclustering microarray data by Gibbs sampling**. **Bioinformatics**, 2003. 19 Suppl 2: p. ii196-205.

SILVA, A.S.; HECK, C.A.; DOYLE, R.L.; MONTEIRO, S.G Levantamento das espécies de dípteras na região de Santa Maria baseado em diferentes substratos. **Revista da Faculdade de Zootecnia, Veterinária e Agronomia de Uruguaiana**, v.12, n.1, p.51-58, 2005.

SOUZA, C.R. **Sazonalidade, sinantropia e preferência por iscas de dípteros necrófagos da região de Rio Claro**, SP. 2011. 60 f. Dissertação - (mestrado) - Universidade Estadual Paulista, Instituto de Biociências de Rio Claro, 2011. Disponível em: <<http://hdl.handle.net/11449/99577>>. Acesso em 10 fev 2013.

SOWA, J. F. **Conceptual Structures: Information Processing in Mind and Machine**. Addison-Wesley, 1984.

SYMEONIDIS, P., et al. (2007). Nearest**biclusters** collaborative filtering with constant values. In **Advances in Web Mining and Web Usage Analysis, volume 4811 of Lecture Notes in Computer Science** (pp. 36– 55). Philadelphia, USA: Springer-Verlag.

TANAY, A. ; SHARAN, R.; SHAMIR, R. Discovering statistically significant *biclusters* in gene expression data. **Bioinformatics**, vol. 18, pp. 136-144, 2002.

TANG, C., ZHANG L.; ZHANG A. **Interrelated Two-way Clustering: An Unsupervised Approach for Gene Expression Data Analysis**, in **Proceedings of the 2nd IEEE International Symposium on Bioinformatics and Bioengineering 2001**, IEEE Computer Society. p. 41.

TIBSHIRANI, R. et al. "Clustering Methods for the Analysis of DNA Microarray Data," technical report, Dept. of Health Research and Policy, Dept. of Genetics, and Dept. of Biochemistry, Stanford Univ., 1999.

VERONEZE, R., BANERJEE, A.; ZUBEN, F. J. V. (2014). Enumerating all maximal *biclusters* in real-valued datasets. **CoRR**, *abs/1403.3562*.

VIANNA, E.E.S.; BRUM, J.G.; RIBEIRO, P.B. Synanthropy of Calliphoridae (Diptera) in Pelotas, Rio Grande do Sul State, Brazil. **Rev. Brasil. Parasitol. Vet.**, v. 7, n. 2, p. 141-147, 1998.

XIE, B., CHEN, S.; LIU, E F. **Biclustering of gene expression data using PSO-GA hybrid. The 1st International Conference on Bioinformatics and Biomedical Engineering, 2007:** 302-305.

YAN, D.; WANG, J. Biclustering of gene expression data based on related genes and conditions extraction. **Pattern Recognition**, v. 46, n. 4, p. 1170-1182, 2013.

## CAPÍTULO II

---

### Application of decision trees to model the occurrence of dipterans in a monitored ecosystem

CR DE SOUZA<sup>1</sup>, FJ VON ZUBEN<sup>2</sup> & CJ VON ZUBEN<sup>1</sup>

Univ Estadual Paulista (UNESP), Biosciences Institute, Department of Zoology, Rio Claro, São Paulo state, Brazil.; <sup>3</sup> Univ Estadual de Campinas (Unicamp), Faculty of Electrical Engineering and Computation, Campinas, São Paulo state, Brazil.

\*Article submitted for publication in the journal *Ecological Informatics*

**ABSTRACT:** Due to advances in information technology, several approaches are now available for information extraction and for better understanding of ecological processes, simply by monitoring relevant variables, thus guiding to the synthesis of prediction or classification models. Employing a machine learning paradigm denoted learning from data, decision trees were synthesized from datasets generated by the frequency of occurrence of three dipterous families (Muscidae, Sarcophagidae and Calliphoridae), with samples collected for a period of one year in the city of Rio Claro, São Paulo state, southeastern Brazil. The occurrence was converted to a binary state (absent or present) and the decision trees operate as high-performance binary classifiers. The interpretability of the obtained decision trees is explored in this work, automatically producing reports such as: (1) *Musca domestica* Linnaeus, *Synthesiomyia nudiseta* (Wulp), *Atherigona orientalis* Schiner and *Ophyra chalcogaster* (Wiedmann) are species that were found during the autumn season in the urban environment, all attracted by the minced meat bait, which seems to be favourable to the development of their larvae; and (2) *Peckia (Euboettcheria) anguilla* (Curran & Walley, 1934) and *Peckia (Peckia) intermutans* (Walker, 1861) were found in the forest area, attracted by the minced meat bait. Concerning report number 2, both species were already considered non-synanthropic and the preference for meat bait for the latter species had been also confirmed and tested in laboratory by some authors. Possibly due to the fact that they belong to the same genus, the species *Lucilia eximia* (Wiedmann) and *Lucilia cuprina* (Wiedmann) guided to trees covering the same sequence of attributes (season, location and bait type), and also sharing the same attribute values for the correct classification. In the current analysis, the season appears as the most relevant attribute to characterize absence or presence of a given species, thus indicating the apparent influence of climatic factors on the population dynamics of these dipterans. The relevance of such machine learning approach for wildlife surveys was shown to be distinguished, automatically unveiling aspects of ecological processes that are intricate and hard to be directly detected by a careful inspection of the monitored variables.

**KEYWORDS:** Necrophagous Diptera, graph, decision trees, flies

## Introduction

Identifying patterns in ecological communities is one of the most persistent challenges of ecological science (Lewinsohn et al., 2006). Increased complexity and dimensionality of the collected information requires advanced computational approaches to store, organize and analyze data before they can be transformed into a new biological knowledge (Geurts et al., 2009). Aiming at understanding ecological processes on a given system, through the interaction of its components, it is usual to monitor the time evolution of several variables and to ask for mathematical models for prediction and classification, which are synthesized directly from the observed data.

Decision trees are machine learning devices characterized by high performance and high-level of interpretability of the obtained classifier (Mitchell, 1997; Kingsford; Salzberg, 2008). The decision tree is synthesized using a learning from data approach to discover structure and conditional dependence in datasets (Han, 2001; Tan et al., 2005). When dealing with classification tasks, mathematical models, such as decision trees, are induced to automatically assign a class label, among several previously defined labels, to unlabelled objects in the dataset, with each object being characterized by a set of attributes. Labeled objects form the training data, used as input to the learning algorithm, which will provide as output the mentioned mathematical model, here a decision tree.

Greenberg (1971) and Furlanetto et al. (1984) highlighted the great importance of the study of caliptrated blowflies for sanitarian purposes, given that they act as vectors of etiological agents such as amoeba cysts, helminth eggs, pathogenic enteric bacteria, viruses and fungi, with many of their representatives exhibiting a high synanthropy index. Caliptrated blowflies develop on various substrates, ranging from living tissue, carcasses, to urban waste and faeces. These substrates are used to feed adults and larvae, allowing the subsequent placement of a series of harmful micro-organisms, as mentioned above, when in contact with the substrate. Additionally, larvae of these blowflies occasionally cause myiasis on humans and domestic animals (Guimarães et al., 1983;. Guimarães; Papavero, 1999).

Moreover, the dipterans are one of the most important Hexapoda orders in the study of forensic entomology (Keh 1985, Tomberlin et al 2011.), being the first insects to colonize carcasses and corpses (Catts; Goff 1992), especially the Calliphoridae family, which acts in all stages of decomposition and includes species that can be used as a tool for calculating the PMI (Goff 1993).

This study investigates the effects of some factors, including the season, the type of bait, and the data collection site, on the occurrence of certain species of necrophagous blowflies in three different ecological niches. It's the first time that this technique was applied to the carrion flies. The collected data form a two-dimensional matrix that is used as input to a machine learning algorithm designed to learn from data. The technical purpose was to correctly classify each species as absent or present, given the environmental conditions. On the other hand, the practical purpose was to interpret the obtained classification model, aiming at extracting useful knowledge for healthcare policies and for forensic entomology, since many of the monitored species are important in both contexts.

## **Material and Methods**

### **The dataset configuration**

For this study, the available dataset is composed of a collection of 2,648 specimens of Diptera, with 1,245 specimens (47.02%) belonging to the family Calliphoridae (Souza; Von Zuben, 2012), followed by Muscidae with 1,082 specimens (40.86%), and Sarcophagidae, with 321 specimens (12.12%), for a total of 19 species identified (see Table 63).

The study was carried out in three different locations in the city of Rio Claro, São Paulo state, Southeastern Brazil: in urban, rural and forest areas. The motivation was to determine the synanthropy index (Nuorteva, 1963) of the major species.

Three types of baits were used: minced meat, liver and sardine. Each trap received 100 grams of a particular bait, which was not replaced until the end of collection. Every sampling site received three traps, each one with a distinct type of bait. Traps made with PET plastic bottles of two liters were hung on trees at a distance of 1.5 meters from the ground and also about 13 meters distant from each other.

The collections were made monthly, between the months of September 2009 and August 2010. The traps were exposed for five consecutive days and samples were taken every 24 hours, starting from the second day. The organza was removed and placed into a pot with ether. Thereafter, the insects were removed and organza was placed in the trap again.

### **Conceptual aspects of decision trees**



Decision trees are well-known machine learning tools for inducing interpretable classification models, also known as predictive functions, from available datasets (Breiman et al., 1984; Buntine, 1994). The high interpretability of the obtained classifier is supported by the fact that each path from the root node of the tree to a leaf node generates a propositional rule in the form: IF ⟨The attributes of the object obey a given set of conditions⟩ THEN ⟨The object belongs to a specific class⟩. The set of conditions is obtained taking a decision at each node, thus producing a sequence of decisions from the root node to each leaf node.

Notice that, by definition (Edwards, 2000), there is a single path from the root to any of the leaves, which guarantees a single rule associated with each leaf node of the obtained tree. Besides being highly interpretable, the obtained decision tree may be used to indicate the degree of relevance of each attribute to the corresponding learning task and may also propose a conditional dependence among attributes and among attributes and the classes. Generally, attributes closer to the root node are more relevant to the classification task (Quinlan, 1993).

To better formalize a classifier algorithm, let us first define an attribute as the value that describes a given characteristic of an individual. The finite set of all attributes that fully characterize an individual is called an object. Adding the label of the already classified objects as the independent attribute of the object and organizing these labelled objects in a matrix having the objects in the rows and the attributes in the columns, we arrive to a plain file denoted the input-output dataset or the training matrix.

Then, a classifier algorithm is a function that maps all labelled object to its corresponding class, thus being a discrete-valued function (Mitchell, 1997). A decision tree is a classifier algorithm that uses a graphical model: it considers one of the dependent attributes at each node of the tree (we will mention further the case of multiple attributes being considered at the same node), starting at the root node, and make a test for that attribute. Depending on the outcome of the test, one of the subsequent nodes will be achieved and another test is performed there, considering another attribute or even the same one, until a leaf node is achieved. Each leaf node is associated with a class label. Therefore, each test leads either to another test or to a target class. As a consequence, a decision tree may be interpreted as a hierarchically defined sequence of tests to be applied to objects toward predicting their associated class label (Quinlan, 1993).

There are three types of attributes that a machine learning algorithm should be prepared to deal with (Mitchell, 1997): (1) discrete unordered, also known as categorical; (2) discrete ordered; and (3) continuous, which is always ordered. Decision trees have a specific

treatment for type 1 and a single and distinct treatment for types (2) and (3), as will be further explained.

The purpose of the classifier algorithm is to produce the best achievable accuracy. For decision trees, interpretability, dimension reduction (discarding a subset of attributes), and low computational cost are additional aspects that influence the quality of the obtained solution. The synthesis of a decision tree can be performed efficiently by established approaches in the literature, such as C4.5 (Quinlan, 1993) and CART (Breiman et al., 1984). Here we adopt C4.5, because CART may consider multiple attributes at each test, tending to be less interpretable for the purpose of the current application.

Starting at the root node, the C4.5 algorithm should subsequently define the best test to be performed at each node. To do so, each attribute of the objects (notice that all objects share the same number of attributes) is considered as a candidate, and the one that reduces the most the impurity after the decision is taken as the selected one. The impurity is measured by an entropy index (Quinlan, 1993) and can be interpreted as the degree of mixture of distinct labels in the subsequent nodes, after the test is applied. For example, consider a two-class classification problem, two candidate continuous attributes (A1 and A2), and 100 objects at the test node, being 50 from class C1 and 50 from class C2, and also suppose that each attribute, when tested, will produce the following situations:

Situation 1: The test involving attribute A1 will guide to two descendant nodes, one with the division of 20 objects from class C1 and 25 objects from class C2, and the other with the division of 30 objects from class C1 and 25 objects from class C2.

Situation 2: The test involving attribute A2 will guide to two descendant nodes, one with the division of 45 objects from class C1 and 10 objects from class C2, and the other with the division of 5 objects from class C1 and 40 objects from class C2.

Given that each leaf node should have objects from a single class to avoid classification errors, Situation 2 is more favourable, because it guides to a stronger reduction in the impurity of the descendant nodes; more specifically, the test produces partitions of the dataset characterized by the most accentuated reduction in the impurity. Therefore, attribute A2 should be preferred over attribute A1 to perform the test at this specific tree node. There is no guarantee, however, of obtaining the best tree topology, because a greedy selection is performed locally and there is no backtracking to revert locally optimal selection, which eventually proves to be not globally optimal in further steps of the algorithm. The index that defines the best test in C4.5 is denoted gain ratio (Quinlan, 1993). For categorical attributes, there will be one descendant node per category, and for discrete ordered and continuous

attributes a threshold value is optimized and always two descendant nodes are produced, one for objects for which the value of the corresponding attribute is below the threshold, and the other for objects for which the value of the corresponding attribute is above or equal to the threshold.

When part of the training matrix is reserved as a validation dataset, pruning steps can be performed and parameters of the C4.5 algorithm can be tuned, both aiming at improvement in performance (Quinlan, 1993). The implementation of the C4.5 algorithm adopted here was the J48 package available in the WEKA toolbox (Hall et al., 2009), and no pruning was allowed, guiding to classifiers with 100% of correct classification in all the scenarios considered. Relevant parameters used to run the J48 software will be presented together with the experimental results. One that deserves attention here is the minimum number  $M$  of objects to be associated with a node, which is used as a stopping criterion. For  $M$  closer to 1, the algorithm is free to produce large trees. Otherwise, for  $M$  much larger than 1, the final number of nodes in the tree tends to be reduced, even when the impurity in the node remains at a high level.

The authors are conscious of the potential risk of overfitting when no pruning is allowed and when 100% of correct classification is achieved. However, the particularities of the present qualitative approach should be emphasized to legitimate the learning parameter setting:

- (1) Given that few ecological traits are involved in the experiments, the final decision trees are parsimonious enough to remain fully interpretable;
- (2) The decision trees are generally not composed of all the available ecological traits, indicating that a relevant ecological association is suitably described by a subset of traits, again guiding to parsimonious decision trees and, consequently, simple association rules;

- (3) A significant part of the findings and associations require 100% of correct classification, because they do not accept samples that somehow violate the raised hypothesis.
- (4) Noise in the collected data is not necessarily of great concern, given the experimental conditions, characterized by several replications of field collections that are close in space and time. It is evident that the traps conceived are not able to guarantee that species not collected are indeed absent from that site in nature. Nonetheless, the absence of some species in the experiment can be at least an indication of a very low occurrence of that species in the natural sampled site. Depending on the practical purpose of the analysis, very low occurrence may be approximately interpreted as absence.
- (5) Even if item (4) above does not apply in certain scenarios, it is an indisputable fact that the findings and associations to be presented in this paper are clearly conditioned and fully supported by the available dataset. Being this way, it will be always possible to automatically revise the established findings and associations when they supposedly become inconsistent after incorporating sensible novelties into the dataset.

## Experimental results

One decision tree was obtained for each of the 19 species, being 8 for the family Sarcophagidae, 4 for Muscidae, and 7 for Calliphoridae – **Supplementary Material**). For illustrative purposes, Figure 18 presents the tree related to the species *Sarcodexia lambens* Wiedemann (1861) (Diptera: Sarcophagidae).

For creation of decision trees, we used the data mining system free WEKA in version 3.6.13, which can be acquired for free at <http://www.cs.waikato.ac.nz/ml/weka/>. All other relevant parameters used to run the J48 software are taken with default values, less than M (minNumobj) that takes the value of 1 and the U (unpruned) was marked as True.

Considering the Sarcophagidae species, it was noted that several groups followed essentially the same classification pattern (see Table 64). Among them, we can mention that *P (E) collusor* and *P (P) intermutans* will be present if the site is forest and the season is summer. If the season is spring and the site is forest, it will be detected *S. lambens*, *P(E)*

*collusor*, *O. avuncula*, *P(E) anguilla*, *P(P) intermutans*: *S. lambens*, *O. Thornax* and *O. avuncular*. On the other hand, *S. lambens*, *O. Thornax* and *O. avuncula*, *P(E) florencioi* will be jointly collected if the season is winter, the site is forest and the bait is sardine.

Now focusing on Muscidae, some species also follow similar patterns (see Table 65). *A. orientalis*, *S. nudiseta*, *M. Domestica* and *O. chalcogaster* will be present if the seaason is autumn, the site is urban and the bait is minced meat. On the other hand, *M. domestica* and *O. chalcogaster* will be found when the season is spring, the site is urban, and the bait is sardine.

With regard to Calliphoridae species (see Table 66), *L. eximia* and *L. cuprina* are present if the season is spring, the site is rural and the bait is liver. *L. eximia*, *L. cuprina* and *C. albiceps* will be detected when the season is spring, the site is rural and the bait is minced meat. If the season is summer, the site is rural and the bait is minced meat, *L. cuprina* and *C. megacephala* will be present.

Making a joint analysis of the three families, it was noted that the decision trees of the species *S. lambens*, *P (E) collusor*, *M. domestica* and *O. chalcogaster* produced the same sequence of nodes, more specifically spring season, urban site and sardine bait. Taking this as a rule basis, but changing the season to autumn, we find again *S. lambens* and *P (E) collusor*, together with the following species: *O. avuncula*, *P (E.) anguilla*, *H. segmentaria*, *H. semidiaphana*, *M. peregrina*, *P. (P.) intermutans* and *O. thornax*.

Among the three variables considered in this study (season, location and bait type), the season was the most frequently selected at the root node of the decision tree (10 in 19 possibilities), making it the most influential variable. For the remaining nine species, the site was placed at the root. The bait type was never selected at the root node.

*Oxysarcodexia thornax* was the only species to achieve a final decision tree composed of solely one test: no matter the site, and no matter the bait type, if the season is spring this species is detected. Other species require two sequential tests. For example, no matter the site, if the season is autumn and the bait is sardine, *S. lambens* is present. Also, no matter the site, if the season is spring and the bait is minced meat, *O. chalcogaster* will be detected.

Without going through the bait type, the presence of *A. orientalis* can be explained if the season is spring and the site is urban or rural, or alternatively if the season is summer and the site is urban. Also, if the season is spring and the site is urban, *S. nudiseta* is present; if the site is rural and the season is summer, *M. domestica* will be detected; and *H. segmentaria*, *H. semidiaphana* and *M. peregrina* will be present if the site is forest and the season is spring.

## Discussion

Individuals belonging to the species *P.(E) anguilla* and *P.(P.) intermutans* were found in the forest area and attracted by the minced meat bait. The preference for minced meat has been confirmed and tested in laboratory for *P(P) intermutans* (Loureiro et al., 2005) and the preference of *P(P) intermutans* for frequenting forest sites has also been reported by several authors (Linhares, 1979; 1981; Dias et al., 1984), being thus considered a non-synanthropic species.

Because of the food supply offered by human activity in urban areas, *M. domestica* (Linhares, 1979; D'Almeida, 1992; Uribe-M et al., 2010), *S. nudiseta* (Gregor, 1975; Uribe-M et al., 2010), *A. orientalis* (Linhares, 1979; D'Almeida, 1983) and the species of the genus *Ophyra* (Costa et al., 2000) have been reported as synanthropic species. This was confirmed by Souza (2011) by means of synanthropic indices with positive values found for these species. In the present study, all collected Muscidae species were found in the urban area and were attracted by minced meat, known to be favorable to the development of their larvae (Krüger et al., 2002; D'Almeida et al., 1999).

*M. domestica* and *O. chalcogaster* were detected simultaneously in urban areas and were attracted by the sardine bait, widely used in studies of Muscidae (Leandro; D'Almeida, 2005). The close relation of *M. domestica* with species of the genus *Ophyra* is already known and the latter are used in the biological control of the former on poultry and pig farming in Germany (Betke et al., 1989) and also in the United States (Hogsette; Jacobs, 1999), although this practice is not yet widespread in Brazil.

In a general view of the three families, it was noted that many species also occurred along the urban sites and were attracted by the sardine bait. An intuitive explanation is that, being a white meat, biochemical reactions trigger faster decomposition processes in sardine when compared to what happens with red meat (Santos; Moya-Borja, 1997), promoting an intense attraction for different Muscidae species.

Possibly due to the fact that they belong to the same genus, the species *L. eximia* and *L. cuprina* occurred under the same conditions (spring season, rural site and liver bait). It is common sense that *L. cuprina* is a species with rural origins, being responsible for myiasis in sheep (Mariconi et al., 1999; Norris, 1990; Stevens; Wallman, 2006; Zumpt, 1965). The rural environment has also been reported as a place where *L. eximia* has been found in Brazil, Peru and Argentina (Baumgartner; Greenberg, 1984; Ferreira, 1978; Linhares, 1981;. Madeira et al., 1982; Mariluis; Schnack, 1989).

For the *Lucilia* species, liver has been shown to be effective not only as a convenient substrate for larvae breeding, but also as an attraction bait (Quattro; Wasti, 1978; D'Almeida; Lopes, 1983; D'Almeida, 1989; Delves; Browne, 1989). In contrast, D'Almeida; Mello (1996) observed that the minced meat was the preferred substrate for the oviposition of *L. cuprina* and *C. megacephala*, which could explain the occurrence of these two species together in this bait, as noted by the decision tree.

The simultaneous occurrence of the genus *Lucilia* and the species introduced in Brazil of the genus *Chrysomya* (Baumgartner; Greenberg, 1984;. Guimarães et al., 1978; 1979) under the same conditions (season as spring or summer, site as rural, and bait as minced meat) may be due to similarity of ecological niches between them, suggesting the occurrence of an interspecific overlap (Ferreira, 1983; Guimarães et al., 1979; Guimarães, 1984; Wells; Greenberg, 1992b; Ulyett, 1950). In addition, the species of the genus *Chrysomya* have shown increased competitive and / or predatory skills during the confrontation with native species Wells; Greenberg, 1992a, b, c;. Reis et al., 1999) and are known to perform intragild predation, i.e. pass the trophic level of competitors (omnivorous) to assume the role of predator of another competitor during the larval stage (Faria et al., 2007; Rosa et al., 2006).

For dipterans, both the temperature as the relative humidity have been linked to changes in their abundance (Batista-da-Silva et al., 2010), besides precipitation (Gomes et al., 2000, Mello et al., 2007). In this study, the fact that the season has been most frequently selected at the root node of the decision tree indicates that the influence of climatic factors on the population dynamics of blowflies is really great. Indeed, climatic factors can be restrictive, thus influencing the development, reproduction and behavior of these insects (Silveira Neto et al., 1976).

Mendes; Linhares (1993) and D'Almeida; Mello (1996) found *O. thornax* as being attracted to different types of baits, such as liver, chicken entrails and fish, showing that this species is generalist. In this study, although this species has been collected only in spring season, it can still be considered a generalist, being indifferent to site and bait type. Furthermore, the absence or presence of several species could be fully determined by the use of only two attributes (season and site, or season and bait type, or site and bait type), with the other attribute being considered irrelevant for the correct classification.

In summary, the automatic synthesis of decision trees directly from the available datasets provided support to explain several interaction profiles among species and the existing climatic and environmental conditions. This is of utmost importance as the basis for forensic applications and entomological surveys. The inherent complexity of ecological

systems is treated here with the aid of a machine learning tool, capable of automatically extracting meaning from raw and unstructured datasets.

Though not being the case here, the computational tools are scalable to deal with huge datasets, composed of millions of rows and thousands of columns, keeping the feasibility of the current methodology of analysis in scenarios characterized by learning matrices of such size. Given that the decision trees will become larger, pruning steps and lower rates of correct classification should be admitted to favour interpretability and generality of the obtained results.

### **Acknowledgments**

The authors wish to thank the group of the Laboratory of Bioinformaticas and Bio-inspired Computing that both helped us in using this mathematical tool. Thanks also to FAPESP (2013/00201-3) for scholarships and to the Post Graduate Program in Biological Sciences (Department of Zoology), Biosciences Institute, UNESP, Rio Claro Campus, São Paulo, Brazil.



Table 62 – Abundance of species of blowflies collected along the four seasons, in three different locations, and using three types of baits. (L: liver, M: minced meat, S: Sardine, S1: *S. lambens*, S2: *O. thornax*, S3: *P (E) collusor*, S4: *O. avuncula*, S5: *P (S) ingens*, S6: *P (E) anguilla*, S7: *P (P) intermutans*, S8 *P: (S) florencioi*, S9 : *A. orientalis*, S10: *S. nudiseta*, S11: *M. domestica*, S12: *O. chalcogaster*, S13: *L. eximia*, S14: *H. segmentaria*, S15: *L. cuprina*, S16: *H. semidiaphana*, S17: *C. megacephala*, S18: *C. albiceps*, S19: *M. peregrina*).

	Spring									Summer									Autumn									Winter									T
	Forest			Urban			Rural			Forest			Urban			Rural			Forest			Urban			Rural			Forest			Urban			Rural			
	L	M	S	L	M	S	L	M	S	L	M	S	L	M	S	L	M	S	L	M	S	L	M	S	L	M	S	L	M	S	L	M	S	L	M	S	
S1	7	15	9	0	8	17	1	2	3	0	0	0	0	0	4	0	0	1	0	0	2	0	0	6	0	0	2	0	1	2	0	0	0	0	0	0	800
S2	5	3	3	8	5	9	2	4	16	2	0	0	2	1	5	0	0	6	0	0	1	0	0	5	0	0	0	0	1	0	0	0	0	0	1	79	
S3	4	19	14	0	2	1	0	0	0	1	3	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	56	
S4	3	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	10		
S5	0	5	4	0	1	0	0	2	0	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	
S6	1	6	4	0	1	0	0	1	1	0	3	4	0	0	0	0	0	0	0	2	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	26	
S7	17	4	2	0	0	0	0	0	0	6	5	5	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	41	
S8	0	3	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	2	0	3	0	0	0	0	0	0	2	1	0	0	0	0	0	0	13		
S9	0	134	39	7	191	131	4	310	11	0	3	0	7	29	1	1	15	0	0	0	0	0	5	0	0	0	2	0	0	0	0	0	0	4	0	8944	
S10	0	0	0	1	12	12	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	1	34	
S11	0	0	0	0	9	4	5	33	0	0	0	0	1	7	1	24	6	1	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	101	
S12	0	1	0	0	3	1	0	12	1	0	0	0	0	7	2	9	5	0	0	0	0	0	11	1	0	0	0	0	0	0	0	0	0	0	0	53	
S13	336	11	8	55	13	4	39	19	0	40	40	13	54	6	2	5	1	0	35	4	6	18	7	4	5	0	0	58	0	3	0	0	0	2	3	794	
S14	45	6	23	0	0	0	0	0	0	1	2	0	0	0	0	0	1	10	4	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	95	
S15	0	0	0	2	1	1	3	2	0	0	0	9	0	3	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	
S16	26	11	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	40	
S17	0	33	5	0	8	10	0	1	0	27	9	5	17	1	1	15	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	134	
S18	4	36	30	0	2	25	0	1	0	9	27	9	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	145	
S19	4	5	1	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	
T	452	293	146	73	256	215	54	387	32	88	100	49	90	53	19	54	32	19	43	7	16	19	36	21	5	0	4	58	4	9	0	0	0	2	7	5	2648

Table 63– Tendencies in collections involving species of the Sarcophagidae family

Species of the <i>Sarcophagidae</i> family	Season	Area	Bait
<i>P(E) collusor</i> e <i>P(E) florencioi</i>	Winter	Forest	Minced meat
<i>O. thornax</i> e <i>P(E) florencioi</i>	Summer	Forest	Liver
<i>S. lambens</i> e <i>P(E) collusor</i>	Spring	Urban	Sardine
<i>S. lambens</i> , <i>P(E) collusor</i> e <i>P(E) anguilla</i> :	Spring	Urban	Minced meat
<i>O. thornax</i> e <i>P(E) anguilla</i>	Autumn	Urban	Sardine
<i>P(E) anguilla</i> e <i>P(P) intermutans</i> :	Autumn	Forest	Minced meat
<i>O. thornax</i> , <i>P(E) anguilla</i> , <i>P(P) intermutans</i> e <i>P(E) florencioi</i> :	Autumn	Forest	Sardine
<i>S. lambens</i> , <i>O. thornax</i> , <i>O. avuncula</i> , <i>P(E) florencioi</i>	Winter	Forest	Sardine
<i>S. lambens</i> , <i>P(E) collusor</i> , <i>O. avuncula</i> , <i>P(E) anguilla</i> , <i>P(P) intermutans</i> , <i>O. thornax</i>	Spring	Forest	-
<i>P(E) collusor</i> e <i>P(P) intermutans</i>	Summer	Forest	-

Table 64- Tendencies in collections involving species of the Muscidae family

Species of the Muscidae family	Area	Spring	Bait
<i>A. orientalis</i> , <i>S. nudiseta</i> , <i>M. domestica</i> , <i>O. chalcogaster</i>	Urban	Autumn	Minced meat
<i>M. domestica</i> e <i>O. chalcogaster</i>	Spring	Urban	Sardine
<i>A. orientalis</i> , <i>S. nudiseta</i>	Spring	Urban	-

Table 65- Tendencies in collections involving species of the Calliphoridae family

Species of the Calliphoridae family	Season	Area	Bait
<i>L. eximia e L. cuprina</i>	Spring	Rural	Liver
<i>L. eximia, L. cuprina e C. albiceps</i>	Spring	Rural	Minced meat
<i>L. cuprina e C. megacephala</i>	Summer	Rural	Minced meat
<i>H. segmentaria, H. semidiaphana e M. peregrina</i>	Spring	Forest	-

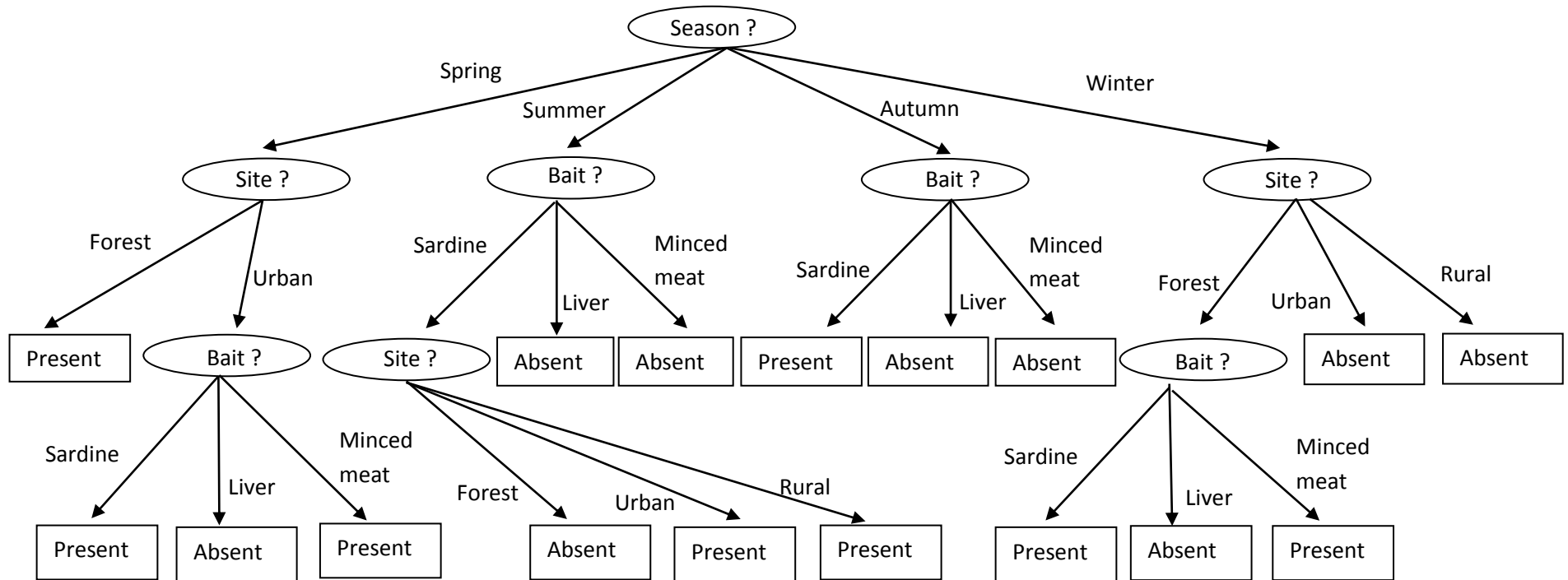


Figure 18 – Final configuration of the decision tree designed to predict absence or presence of the species *Sarcodexia lambens*. Notice that the decision tree is automatically generated from the collected datasets, by a learning from data approach. The ellipsoidal nodes (internal nodes) perform a test over an attribute, thus being a decision point. The decision outcome will indicate which edge to follow. The square nodes (leaves of the tree) are the available classes. Starting from the root node, there is a single path to each leaf node, and the sequence of decisions from the root to each leaf may be expressed as a single IF-THEN classification rule, making the process highly interpretable.

**REFERENCES**

---

Batista-da-Silva, J. A., Moya-Borja, G. E.; Queiroz, M. M. C. Ocorrência e Sazonalidade de Muscóides (Diptera, Calliphoridae) de Importância Sanitária no município de Itaboraí, RJ, Brasil. **EntomoBrasilis** , 3 , 16-21. 2010.

Baumgartner, D.L.; Greenberg, B. The genus *Chrysomya* (Diptera: Calliphoridae). **J Med Entomol** 21: 105-113, 1984.

Betke , P .; Hiepe, T . ; Muller, P.; Schultka, H.; Schumann, H.. Biological control of *Musca domestica* with *Ophyra aenescens* on pig production enterprises. **Monatshefte for Veterinarmedizin** 44: 842-844, 1989.

Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. **Classification and regression trees**. Wadsworth International Group, 1984.

Buntine, W. Operations for learning with graphical models. **Journal of Artificial Intelligence Research**, vol. 2, pp. 159-225, 1994.

Catts EP, Goff ML. Forensic entomology in criminal investigations. **Annual Review of Entomology** 37: 253-272, 1992.

Costa, P. R. P.; R. L. Franz; E. E. S. Vianna; Ribeiro, P. B. 2000. Synanthropy of *Ophyra* spp (Diptera, Muscidae) in Pelotas, RS, Brazil. **Rev. Brasil. Parasitol.Veter.** 9 (2): 165-168.

D'Almeida, J. M. **Sinantropia de dípteros caliptrados na área metropolitana do Rio de Janeiro**. Tese de Mestrado, Universidade Federal Rural do Rio de Janeiro, 1983. 193p.

D'Almeida, J.M. Substratos utilizados para a criação de dípteros caliptrados no Jardim Zoológico do Rio de Janeiro (RIO- ZOO). **Mem Inst Oswaldo Cruz** 84: 257-264, 1989.

D'Almeida, J. M. Calyptrate Diptera (Muscidae and Anthomyidae) of the State of Rio de Janeiro - I. Synanthropy. **Memórias do Instituto Oswaldo Cruz** 87: 381- 386, 1992.

D'Almeida, J.M.; Lopes, H.S. Sinantropia de Dípteros Muscóides (Calliphoridae) no Estado do Rio de Janeiro. **Arq. Univ. Fed Rural do Rio de Janeiro**, v. 6, n. 1, p. 39-48, 1983

D'Almeida, J.M.; Mello, R.P. Comportamento de dípteros muscóides frente substrato de oviposição, em laboratório, no Rio de Janeiro, RJ, Brasil. **Mem. Inst. Oswaldo Cruz** 91: 131-136, 1996.

D'Almeida, J. M.; Borges, C.; Gonçalves, C. A. Desenvolvimento Pós-embrionário de *Ophyra aenescens* (Wiedemann, 1830) (Diptera: Muscidae) em Diferentes Dietas, sob Condições de Laboratório. *Memórias do Instituto Oswaldo Cruz*, 94(1), 1999.

Delves, R.I.; Browne, L.B. Changes in reproduction performace of the australian sheep blowfly, *Lucilia cuprina* (Diptera, Calliphoridae) following laboratory colonization. **J appl Entomol** 107: 228-237, 1989.

Dias, E .S.; Neves, D. P., Lopes, H. S. Estudos sobre a fauna de Sarcophagidae (Diptera) de Belo Horizonte, Minas Gerais. I. Levantamento taxonômico e sinantrópico. **M. Inst. Oswaldo Cruz** 79: 83-91, 1984.

Edwards, D. **Introduction to Graphical Modelling**, Springer, 2nd. Edition, 2000.

Faria, L. D. B; Reigada, C.; Trinca L. A.; Godoy, W. A. C. 2007. Foraging behaviour by an intraguild predator blowfly, *Chrysomya albiceps* (Diptera : Calliphoridae). **J. Ethology**. Tokyo: Springer Tokyo, v. 25, n. 3, p. 287-294.

Ferreira, M. J. M. Sinantropia de dípteros muscóides de Curitiba, Paraná. I: Calliphoridae. **Revista Brasileira de Biologia**, 38: 445-454, 1978.

Ferreira, M. J. M. Sinantropia de Calliphoridae (Diptera) em Goiânia, Góias. **Revista Brasileira de Biologia**, 43: 199-210, 1983.

Furlanetto, S. M. P. et al. Microrganismos enteropatogênicos em moscas africanas pertencentes ao gênero *Chrysomya* (Diptera: Calliphoridae) no Brasil. **Rev.Microbiol.**, v. 15, p. 170-174, 1984.

Geurts, P.; Irrthum, A.; Wehenkel, L. Supervised Learning with Decision Tree-Based Methods in Computational and Systems Biology. **Molecular BioSystems**, 5, 1593-1605, 2009.

Goff, M.L. Estimation of postmortem interval using arthropod development and succession patterns. **Forensic Science Review**. vol V, p. 81- 94, 1993.

Gomes, A.; Koller, W. W.; Barros, A. T. M. Sazonalidade da mosca-varejeira, *Cochliomyia macellaria* (Diptera: Calliphoridae), na região dos cerrados, Campo Grande, MS. **Rev. Brasil. Parasitol. Veter.**, 9: 125-128, 2000.

Greenberg, B. 1971. **Flies and disease**, Vol. 1. Princeton University Press, 856 pp.

Gregor, F. Synanthropy of Muscidae and Calliphoridae (Diptera) in Cuba. **Folia Parasitologica** **22**: 57–61, 1975.

Guimarães J. H.; Papavero, N. 1999. **Myiasis in man and animals in the Neotropical region**. Bibliographic database. São Paulo, FAPESP/ Editora Plêiade, 308p.

Guimarães, J.H.; Papavero, N.A.; Prado, A.P. As míases na região Neotropical (identificação, biologia e bibliografia). **Rev. Bras. Zool** ., v. 1, n. 4, p. 293-416; 1983.

Guimarães, J.H.; Prado, A.P.; Buralli, M. Dispersal and distribution of three newly introduced species of *Chrysomya* Robineau-Desvoidy in Brasil (Diptera, Calliphoridae). **Rev. Bras. Entomol** **23**: 245-255, 1979.

Guimarães, J.H.; Prado, A.P; Linhares, A. X. Three newly introduced blowfly species in Southern Brazil (Diptera: Calliphoridae). **Rev. Bras. Entomol** **22**: 53-60, 1978.

Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA Data Mining Software: An Update, **SIGKDD Explorations**, vol. 11, Issue 1, 2009.

Han, J. **Data Mining: Concepts and Technique**. San Francisco, CA: Morgan Kaufmann, 2001.

Hogsette, J.A.; Jacobs, R.D. Failure of *Hydrotaea aenescens*, a larval predator of the house fly, *Musca domestica* L., to establish in wet poultry manure on a commercial farm in Florida USA. **Med. Vet. Entomol.**, v.13, p.349-354, 1999.

Keh, B. 1985. Scope and applications of forensic entomology. **Annual of Review Entomology** 30: 137- 154.

Kingsford, C.; Salzberg, S.L. What are decision trees? **Nat. Biotechnol.** 26: 1011–1013, 2008.

Krüger, R. F.; Ribeiro, P. B.; Carvalho, C. J. B.; Costa, P. R. P. Desenvolvimento de *Synthesiomyia nudiseta* (Diptera, Muscidae) em laboratório. **Iheringia**. Série Zoologia, 92(4), 25-30, 2002.

Leandro, M. J. F.; D'Almeida, J. M. Levantamento de Calliphoridae, Fanniidae, Muscidae e Sarcophagidae em um fragmento de mata na Ilha do Governador, Rio de Janeiro, Brasil. **Iheringia**, Série Zoologia, 95(4):377-381, 2005

Lewinsohn T M; Prado P I; Jordano, P.; Bascompte, J.; Olesen, J. Structure in plant animal interactions assemblages. **Oikos**, 113: 174-184, 2006.

Linhares, A. X. **Sinantropia de dípteros muscóides de Campinas**. 1979. Tese de Mestrado. Universidade Estadual de Campinas, Campinas, SP, 129 p.

Linhares, A. X. Synanthropy of Calliphoridae and Sarcophagidae ( Diptera) in the city of Campinas, São Paulo, Brazil. **Revista Brasileira de Entomologia**, 25 (3) : 189-215, 1981.

Loureiro, M.; Oliveira, V.; D'almeida, J. Desenvolvimento pós-embrionário de *Pattonella intermutans* (Thomson) (Diptera: Sarcophagidae) em diferentes dietas. **Rev. Bras. Entomol.** 49(1): 127-129, 2005.



Madeira, N.G.; Dias, F.S.; Mascarenhas, C.S. Contribuição ao conhecimento da fauna de Calliphoridae (Diptera) sinantrópicos de Pampulha, Belo Horizonte, Minas Gerais. **Rev. Bras. Entomol.**, v.26, p.137-140, 1982.

Mariconi, F. A. M., Guimarães, J. H.; Filho, E. B. **A mosca doméstica e outras moscas nocivas**. Piracicaba: FEALQ, 135 p., 1999.

Mariluis, J.C.; Schnack, J.A. Ecology of the blow flies of an eusynanthropic habitat near Buenos Aires (Diptera: Calliphoridae). **Eos**, v.65, p.93-101, 1989.

Mello, R. S., Queiroz, M. M. C. ; Aguiar-Coelho, V. M. Population fluctuations of calliphorid species (Diptera, Calliphoridae) in the Biological Reserve of Tinguá, state of Rio de Janeiro, Brazil. **Iheringia, Série Zoologia**, 97: 481-485, 2007.

Mendes, J.; Linhares, A.X. 1993. Atratividade por iscas, sazonalidade e desenvolvimento ovariano em várias espécies de Muscidae (Diptera). **Rev. Bras. Entomol.** 37: 289-297.

Mitchell, T. **Machine Learning**, McGraw-Hill, 1997.

Norris, K. R. Evidence for the multiple exotic origin of Australia populations of sheep blowfly, *Lucilia cuprina* (Wiedemann) (Diptera; Calliphoridae). **Aust. J. Zool.**, 38: 635-648, 1990.

Nuorteva, P. 1963. Synanthropy of blowflies (Diptera, Calliphoridae) in Finland. **Ann. Entomol. Fenn.** 29: 1-49.

Quattro, M.J.; Wasti, S.S. Olfactory and oviposition responses of the green bottle fly, *Phaenicia sericata* to a variety of natural baits (Diptera: Calliphoridae). **Rev Brasil Biol** 8: 115-119, 1978.

Quinlan, J.R. **C4.5 Programs for Machine Learning**, Morgan Kaufmann, 1993.

Reis, S. F., Von Zuben, C.J.; Godoy, W.A.C. Larval aggregation for food in experimental population of *Chrysomya putoria* (Wied.) and *Cochliomyia macellaria* (F.) (Dipt. Calliphoridae). **J. Appl. Entomol.** 123: 485-489, 1999.

Rosa ,G.S. et al. The dynamics of intraguild predation in *Chrysomya albiceps* Wied. (Diptera:Calliphoridae): interactions between instars and species under different abundances of food. **Neotrop Entomol**, 35(6):775-80, 2006.

Santos, M.J.P.; Moya-borja, G.E. Estudo comparado do desenvolvimento pós-embrionário de *Fannia pusio* (Wiedemann, 1830) (Diptera: Fanniidae), em diferentes substratos. **Arq Biol Tecnol** 40: 253-261, 1997.

Silveira Neto, S.; Nakano, O.; Barbin, D.; Villa Nova , N. A. **Manual de ecologia dos insetos**. São Paulo: Agronômica Ceres. 419 p, 1976.

Souza, C.R. Sazonalidade, sinantropia e preferência por iscas de dípteros necrófagos da região de Rio Claro, SP. 2011. 60 f. Dissertação - (mestrado) - Universidade Estadual Paulista, Instituto de Biociências de Rio Claro, 2011. Disponível em: <<http://hdl.handle.net/11449/99577>>. Acesso em 10 fev 2013.

Souza , C. R.; Zuben , C. J. V. 2012. Diversity and synanthropy of Calliphoridae (Diptera) in the region of Rio Claro, SP, Brazil. **Neotrop. Entomol.** 41 :243-248.

Stevens, J. R.; Wallman , J. F. The evolution of Myiases in humans and other animal in the Old and New Worlds (part I): phylogenetic analyses. **Tren. Paras.**, 22: 129-136, 2006.

Tan, P.-N.; Steinbach, M; Kumar, V. **Introduction to Data Mining** ,1<sup>st</sup> ed. Boston, MA: Addison-Wesley/Longman, 2005.

Tomberlin, J.K.; Mohr R.; Benbow, M.E.; Tarone, A.M., Vanlaerhoven, S. 2011. A roadmap for bridging basic and applied research in forensic entomology. **Annual Review of Entomology** 56: 401-421.

Ulyett, G.C. Competition for food and allied phenomena in sheep-blowfly populations. **Phil trans R Soc Lond**, v. 234, p. 77-174, 1950.

Uribe-M, N.; Wolff, M.; Carvalho, C. J. B. Synanthropy and ecological aspects of Muscidae (Diptera) in a tropical dry forest ecosystem in Colombia. **Rev. Brasil. Entomol.** 54 (3) : 462 – 470, 2010.

Wells, J.D.; Greenberg B. Rates of predation by *Chrysomya rufifacies* (Macquart) on *Cochliomyia macellaria* (Fabr.) (Diptera: Calliphoridae) in the laboratory: effect of predator and prey development. **Pan-Pacific Entomologist** 68: 12-14, 1992a

Wells, J.D.; Greenberg, B. Laboratory interaction between introduced *Chrysomya rufifacies* and native *Cochliomyia macellaria* (Diptera: Calliphoridae). **Environ. Entomol.** 21: 640-645, 1992b.

Wells, J.D.; Greenberg, B. Interaction between *Chrysomya rufifacies* and *Cochliomyia macellaria* (Diptera: Calliphoridae): the possible consequences of an invasion. **Bull. Entomol. Res.** 82: 133-137, 1992c.

Zumpt, F. **Myiasis in man and animals in the Old World.** Butterworths, 1965.

## CONSIDERAÇÕES FINAIS

---

Diante do que foi aqui apresentado, podemos concluir que a aplicação de *biclusters* e árvores de decisão em base de dados de dípteros necrófagos, apesar de nunca ter sido feita, é sim muito útil na detecção de padrões nas comunidades ecológicas estudadas nesse estudo. A primeira ferramenta demonstrou ser válida na busca de padrões de associações das espécies mediante a algumas variáveis consideradas (isca, local). Já a segunda, mostrou ser capaz de classificar cada espécie estudada mediante também a algumas variáveis ( estação, isca, local), conseguindo-se regras que informam quais condições determinada espécie estaria presente.

É importante salientar o caráter inovador desse trabalho e também demonstrar que cada uma das ferramentas aqui trabalhadas possuem um propósito. Dependendo dos resultados que se buscam em uma base de dados, vale apenas usar as duas ao mesmo tempo ou somente uma. Ou seja, são ferramentas com um grande potencial de uso e aplicação em base de dados entomológicos.

E os resultados aqui encontrados são válidos para serem aplicados em qualquer localidade que possua condições similares em termos de climatologia e vegetação à cidade de Rio Claro, no estado de São Paulo.

**APÊNDICE A – Biclusters com valores constantes iguais a 0, obtidos através da Tabela 3, ao se usar dados binários.**

$$* A\{1\} = [4\ 6\ 7\ 8\ 9]; B\{1\} = [1\ 6];$$

que refere-se à ausência das espécies *H. semidiaphana*, *C. albiceps*, *C. idioidea*, *M. peregrina* e *C. macellaria*, na área urbana (isca de fígado bovino) e na área rural (isca de sardinha).

$$* A\{2\} = [4\ 8\ 9]; B\{2\} = [1\ 2\ 3\ 5\ 6];$$

refere-se à ausência das espécies *H. semidiaphana*, *C. macellaria* e *M. peregrina* na área urbana (iscas: fígado, carne e sardinha) e na área rural (iscas: carne e sardinha).

$$* A\{3\} = [4\ 9]; B\{3\} = [1\ 2\ 3\ 4\ 5\ 6];$$

refere-se à ausência das espécies *H. semidiaphana* e *M. peregrina* na área urbana e na área rural, em todas as iscas utilizadas.

$$* A\{4\} = [4\ 7\ 8\ 9]; B\{4\} = [1\ 3\ 6];$$

refere-se à ausência das espécies *H. semidiaphana*, *C. idioidea*, *C. macellaria*, *M. peregrina*, na área urbana, em todas as iscas utilizadas.

$$* A\{5\} = [4\ 7\ 9]; B\{5\} = [1\ 3\ 4\ 6];$$

refere-se à ausência das espécies *H. semidiaphana*, *C. idioidea* e *M. peregrina*, na área urbana e na área rural, nas iscas de fígado e sardinha.

$$* A\{6\} = [4\ 6\ 7\ 9]; B\{6\} = [1\ 4\ 6];$$

refere-se à ausência das espécies *H. semidiaphana*, *C. albiceps*, *C. idioidea* e *M. peregrina*, na área urbana (isca fígado) e na área rural, nas iscas de fígado e sardinha.

$$* A\{7\} = [2\ 4\ 8\ 9]; B\{7\} = [2\ 3];$$

refere-se à ausência das espécies *H. segmentaria*, *H. semidiaphana*, *C. macellaria* e *M. peregrina*, na área urbana, nas iscas de carne e sardinha.

$$* A\{8\} = [2\ 4\ 9]; B\{8\} = [2\ 3\ 4];$$

refere-se à ausência das espécies *H. segmentaria*, *H. semidiaphana* e *M. peregrinae*, na área urbana e na área rural (nas iscas de fígado e sardinha) e na área rural, na isca de fígado.

$$* A\{9\} = [2\ 4\ 7\ 9]; B\{9\} = [3\ 4];$$

refere-se à ausência das espécies *H. segmentaria*, *H. semidiaphana*, *C. idioidea* e *M. peregrina*, na área urbana (isca de sardinha) e na área rural (isca de carne).

$$* A\{10\} = [3\ 8]; B\{10\} = [6\ 8\ 9];$$

refere-se à ausência das espécies *L. cuprina* e *C. macellaria*, na área rural (isca de sardinha) e na área de floresta, nas iscas de carne e sardinha.

### APÊNDICE B – Biclusters do tipo BVCC e BVCL, obtidos a partir da Tabela 3.

Para a Tabela 3, foram encontrados 7 biclusters do tipo BVCC. Alguns serão demonstrados com mais detalhes abaixo.

$$* A\{1\} = [4\ 5\ 6]; B\{1\} = [1\ 7\ 9];$$

Tabela 49 - Demonstração em formato de tabela do bicluster BVCC de número 1, obtido através da Tabela 3.

Espécies	URBANA		FLORESTA	
	FÍGADO	FÍGADO	SARDINHA	
		<b>1</b>	<b>7</b>	<b>9</b>
<i>O. chalcogaster</i>	<b>4</b>	0	0	0
<i>Muscina</i> sp	<b>5</b>	0	0	0
<i>Ophyra</i> sp	<b>6</b>	0	0	0

$$* A\{2\} = [2\ 3]; B\{2\} = [1\ 6\ 7\ 8\ 9];$$

Tabela 50 - Demonstração em formato de tabela do bicluster BVCC de número 2, obtido através da Tabela 3.

Espécies	URBANA RURAL		FLORESTA			
	FÍGADO	SARDINHA	FÍGADO	CARNE	SARDINHA	
		<b>1</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<i>S. nudiseta</i>	<b>2</b>	1	1	0	0	0
<i>M. domestica</i>	<b>3</b>	1	1	0	0	0

Nesse caso, o zero e o um foram encontrados conjuntamente, devido à utilização do  $\epsilon > 0$ .

$$* A\{3\} = [2\ 5\ 6]; B\{3\} = [4\ 5\ 7\ 8\ 9];$$

$$* A\{4\} = [2\ 3\ 4]; B\{4\} = [6\ 7\ 9];$$

$$* A\{5\} = [2\ 3\ 5\ 6]; B\{5\} = [7\ 8\ 9];$$

$$* A\{6\} = [2\ 3\ 4\ 5\ 6]; B\{6\} = [7\ 9];$$

$$* \quad A\{7\} = [5 \ 6]; \quad B\{7\} = [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9];$$

Na busca por *biclusters* BVCL, utilizando-se também a Tabela 3, foram encontrados um total de 7 *biclusters*. Nesse caso, para a busca, utiliza-se a transposta da matriz. Portanto, o conjunto A se refere ao número das colunas e o conjunto B se refere ao número

$$* \quad A\{1\} = [4 \ 5 \ 7 \ 8 \ 9]; \quad B\{1\} = [2 \ 5 \ 6];$$

Tabela 51- Demonstração em formato de tabela do bicluster BVCL de número 1, obtido através da Tabela 3.

			<i>S. nudiseta</i>	<i>Muscina</i> sp	<i>Ophyra</i> sp
			<b>2</b>	<b>5</b>	<b>6</b>
Rural	Fígado	<b>4</b>	0	0	0
	Carne	<b>5</b>	0	0	0
	Sardinha	<b>6</b>	1	0	0
Floresta	Fígado	<b>7</b>	0	0	0
	Carne	<b>8</b>	0	0	0
	Sardinha	<b>9</b>	0	0	0

$$* \quad A\{2\} = [1 \ 6]; \quad B\{2\} = [2 \ 3 \ 5 \ 6];$$

Tabela 52- Demonstração em formato de tabela do bicluster BVCL de número 2, obtido através da Tabela 3.

			<i>S. nudiseta</i>	<i>M.domestica</i>	<i>Muscina</i> sp	<i>Ophyra</i> sp
			<b>2</b>	<b>3</b>	<b>5</b>	<b>6</b>
Urbana	Fígado	<b>1</b>	1	1	0	0
Rural	Sardinha	<b>6</b>	1	1	0	0

$$* \quad A\{3\} = [1 \ 7 \ 9]; \quad B\{3\} = [4 \ 5 \ 6];$$

$$* \quad A\{4\} = [6 \ 8]; \quad B\{4\} = [4 \ 5 \ 6];$$



\*  $A\{5\} = [1\ 3\ 4\ 5\ 6\ 7\ 8\ 9]$ ;  $B\{5\} = [5\ 6]$ ;

\*  $A\{6\} = [7\ 8\ 9]$ ;  $B\{6\} = [2\ 3\ 5\ 6]$ ;

\*  $A\{7\} = [7\ 9]$ ;  $B\{7\} = [2\ 3\ 4\ 5\ 6]$ ;

**APÊNDICE C – Biclusters do tipo OPSM, obtidos a partir da Tabela 3 (Muscidae).**

$$* A\{1\} = [1\ 2\ 3\ 4\ 5\ 6]; B\{1\} = [1\ 2\ 3\ 7];$$

Tabela 53- Demonstração em formato de tabela do bicluster OPSM de número 1, obtido através da Tabela 3.

Espécies	URBANA			FLORESTA	
	FÍGADO	CARNE	SARDINHA	FÍGADO	
	<b>1</b>	<b>2</b>	<b>3</b>	<b>7</b>	
<i>A. orientalis</i>	<b>1</b>	14	225	132	0
<i>S.nudiseta</i>	<b>2</b>	1	17	15	0
<i>M. domestica</i>	<b>3</b>	1	26	5	0
<i>O. chalcogaster</i>	<b>4</b>	0	21	4	0
<i>Muscina</i> sp	<b>5</b>	0	1	0	0
<i>Ophyra</i> sp	<b>6</b>	0	1	0	0

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para todas as espécies: 7 - Floresta (fígado) ≤ 1 - Urbana (fígado) ≤ 3- Urbana (sardinha) ≤ 2 – Urbana (carne).

$$* A\{2\} = [1\ 2\ 3\ 4\ 5\ 6]; B\{2\} = [4\ 5\ 7];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para todas as espécies: 7 -Floresta (fígado) ≤ 4 - Rural (fígado) ≤ 5 - Rural (carne).

$$* A\{3\} = [1\ 2\ 3\ 4\ 5\ 6]; B\{3\} = [2\ 7\ 8\ 9];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para todas as espécies: 7 -Floresta (fígado) ≤ 9 - Floresta (sardinha) ≤ 8- Floresta (carne) ≤ 2 - Urbana (carne).

$$* A\{4\} = [1\ 2\ 3\ 4\ 5\ 6]; B\{4\} = [2\ 3\ 6\ 7];$$

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para todas as espécies: 7 - Floresta (fígado)  $\leq$  6 - Rural (sardinha)  $\leq$  3 - Urbana (sardinha)  $\leq$  2 - Urbana (carne).

$$* A\{5\} = [1\ 2\ 3\ 4\ 5\ 6]; B\{5\} = [2\ 3\ 7\ 9];$$

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para todas as espécies: 7 - Floresta (fígado)  $\leq$  9 - Floresta (sardinha)  $\leq$  3 - Urbana (sardinha)  $\leq$  2 - Urbana (carne).

$$* A\{6\} = [1\ 2\ 3\ 4\ 5\ 6]; B\{6\} = [5\ 7\ 8\ 9];$$

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para todas as espécies: 7 - Floresta (fígado)  $\leq$  9 - Floresta (sardinha)  $\leq$  8 - Floresta (carne)  $\leq$  5 - Rural (carne).

$$* A\{7\} = [1\ 2\ 5\ 6]; B\{7\} = [1\ 2\ 3\ 4\ 6\ 7];$$

$$* A\{8\} = [1\ 2\ 5\ 6]; B\{8\} = [4\ 5\ 7\ 8\ 9];$$

$$* A\{9\} = [1\ 2\ 5\ 6]; B\{9\} = [2\ 4\ 7\ 8\ 9];$$

$$* A\{10\} = [1\ 2\ 5\ 6]; B\{10\} = [2\ 3\ 4\ 7\ 9];$$

$$* A\{11\} = [3\ 4\ 5\ 6]; B\{11\} = [1\ 3\ 4\ 5\ 6\ 7\ 9];$$

$$* A\{12\} = [3\ 4\ 5\ 6]; B\{12\} = [3\ 4\ 5\ 6\ 7\ 8\ 9];$$

$$* A\{13\} = [1\ 3\ 4\ 5\ 6]; B\{13\} = [1\ 3\ 5\ 7];$$

$$* A\{14\} = [1\ 3\ 4\ 5\ 6]; B\{14\} = [3\ 5\ 6\ 7];$$

$$* A\{15\} = [1\ 3\ 4\ 5\ 6]; B\{15\} = [3\ 5\ 7\ 9];$$

$$* A\{16\} = [1\ 2\ 3\ 5\ 6]; B\{16\} = [1\ 2\ 3\ 6\ 7];$$

$$* A\{17\} = [2\ 3\ 4\ 5\ 6]; B\{17\} = [1\ 2\ 3\ 6\ 7\ 9];$$

$$* A\{18\} = [2\ 3\ 4\ 5\ 6]; B\{18\} = [4\ 5\ 7\ 8\ 9];$$

$$* A\{19\} = [2\ 3\ 4\ 5\ 6]; B\{19\} = [2\ 3\ 6\ 7\ 8\ 9];$$

$$* A\{20\} = [1\ 2\ 4\ 5\ 6]; B\{20\} = [2\ 4\ 7];$$

$$* A\{21\} = [2\ 5\ 6]; B\{21\} = [1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9];$$

$$* A\{22\} = [1\ 5\ 6]; B\{22\} = [1\ 2\ 3\ 4\ 6\ 7\ 8\ 9];$$

- \*  $A\{23\} = [1\ 5\ 6]$ ;  $B\{23\} = [1\ 3\ 4\ 5\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{24\} = [3\ 5\ 6]$ ;  $B\{24\} = [1\ 3\ 4\ 5\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{25\} = [4\ 5\ 6]$ ;  $B\{25\} = [1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{26\} = [1\ 3\ 5\ 6]$ ;  $B\{26\} = [1\ 3\ 5\ 6\ 7]$ ;
- \*  $A\{27\} = [1\ 4\ 5\ 6]$ ;  $B\{27\} = [1\ 2\ 3\ 7\ 9]$ ;
- \*  $A\{28\} = [1\ 4\ 5\ 6]$ ;  $B\{28\} = [1\ 2\ 7\ 8\ 9]$ ;
- \*  $A\{29\} = [1\ 4\ 5\ 6]$ ;  $B\{29\} = [1\ 3\ 5\ 7\ 9]$ ;
- \*  $A\{30\} = [1\ 4\ 5\ 6]$ ;  $B\{30\} = [1\ 5\ 7\ 8\ 9]$ ;
- \*  $A\{31\} = [1\ 4\ 5\ 6]$ ;  $B\{31\} = [2\ 6\ 7\ 8]$ ;
- \*  $A\{32\} = [1\ 4\ 5\ 6]$ ;  $B\{32\} = [5\ 6\ 7\ 8]$ ;
- \*  $A\{33\} = [2\ 3\ 5\ 6]$ ;  $B\{33\} = [1\ 2\ 3\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{34\} = [2\ 4\ 5\ 6]$ ;  $B\{34\} = [2\ 4\ 5\ 7\ 8\ 9]$ ;

**APÊNDICE D – Biclusters do tipo OPSM, obtidos a partir da Tabela 4 (Calliphoridae).**

\* A{1} = [4 6 7 8 9]; B{1} = [1 3 6 8 9];

Tabela 54- Demonstração em formato de tabela do bicluster OPSM de número 1, obtido através da Tabela 4.

Espécies	URBANA		RURAL	FLORESTA	
	FÍGADO	SARDINHA	SARDINHA	CARNE	SARDINHA
	<b>1</b>	<b>3</b>	<b>6</b>	<b>8</b>	<b>9</b>
<i>H. semidiaphana</i>	4	0	0	13	1
<i>C. albiceps</i>	6	0	25	63	39
<i>C. idioidea</i>	7	0	0	2	1
<i>C. macellaria</i>	8	0	0	0	0
<i>M. peregrina</i>	9	0	0	9	1

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *H. semidiaphana*, *C. albiceps*, *C. idioidea*, *C. macellaria* e *M. peregrina*: 1 - Urbana (fígado) ≤ 6 - Floresta (sardinha) ≤ 3 - Urbana (sardinha) ≤ 9 - Floresta (sardinha) ≤ 8 - Floresta (Carne).

\* A{2} = [4 6 7 8 9]; B{2} = [1 4 6 7];

\* A{3} = [4 6 7 8 9]; B{3} = [1 2 5 6 7];

\* A{4} = [4 6 7 8 9]; B{4} = [1 2 5 6 8 9];

\* A{5} = [2 4 8 9]; B{5} = [1 2 3 5 6 7 9];

\* A{6} = [2 4 8 9]; B{6} = [1 2 3 5 8];

\* A{7} = [1 3 4 5 9]; B{7} = [1 2 4 6];

\* A{8} = [1 2 4 5 9]; B{8} = [1 2 4 7];

\* A{9} = [1 4 5 9]; B{9} = [1 2 4 5 6 7];

\* A{10} = [1 4 5 9]; B{10} = [2 4 5 6 8];

\* A{11} = [2 4 5 9]; B{11} = [1 2 3 4 7];

\* A{12} = [2 4 5 9]; B{12} = [1 2 3 4 8];

\* A{13} = [1 2 4 9]; B{13} = [1 2 3 4 7];

- \*  $A\{14\} = [1\ 2\ 4\ 9]$ ;  $B\{14\} = [2\ 3\ 4\ 8]$ ;
- \*  $A\{15\} = [2\ 3\ 4\ 9]$ ;  $B\{15\} = [1\ 2\ 3\ 4\ 5]$ ;
- \*  $A\{16\} = [1\ 4\ 5\ 8\ 9]$ ;  $B\{16\} = [1\ 2\ 5\ 6\ 7]$ ;
- \*  $A\{17\} = [1\ 4\ 5\ 8\ 9]$ ;  $B\{17\} = [2\ 4\ 5\ 6\ 7]$ ;
- \*  $A\{18\} = [3\ 4\ 5\ 8\ 9]$ ;  $B\{18\} = [1\ 2\ 3\ 6]$ ;
- \*  $A\{19\} = [3\ 4\ 5\ 8\ 9]$ ;  $B\{19\} = [1\ 3\ 5\ 6]$ ;
- \*  $A\{20\} = [1\ 3\ 4\ 8\ 9]$ ;  $B\{20\} = [1\ 3\ 5\ 6]$ ;
- \*  $A\{21\} = [4\ 5\ 8\ 9]$ ;  $B\{21\} = [1\ 2\ 3\ 5\ 6\ 7]$ ;
- \*  $A\{22\} = [4\ 5\ 8\ 9]$ ;  $B\{22\} = [1\ 2\ 3\ 5\ 6\ 8]$ ;
- \*  $A\{23\} = [4\ 5\ 8\ 9]$ ;  $B\{23\} = [2\ 3\ 4\ 5\ 6\ 7]$ ;
- \*  $A\{24\} = [1\ 4\ 8\ 9]$ ;  $B\{24\} = [1\ 2\ 3\ 5\ 6\ 7]$ ;
- \*  $A\{25\} = [1\ 4\ 8\ 9]$ ;  $B\{25\} = [2\ 3\ 4\ 5\ 6\ 7]$ ;
- \*  $A\{26\} = [3\ 4\ 8\ 9]$ ;  $B\{26\} = [1\ 2\ 3\ 5\ 6]$ ;
- \*  $A\{27\} = [1\ 3\ 4\ 8]$ ;  $B\{27\} = [6\ 7\ 8\ 9]$ ;
- \*  $A\{28\} = [2\ 4\ 5\ 8\ 9]$ ;  $B\{28\} = [1\ 2\ 3\ 7]$ ;
- \*  $A\{29\} = [2\ 4\ 5\ 8\ 9]$ ;  $B\{29\} = [1\ 2\ 3\ 8]$ ;
- \*  $A\{30\} = [2\ 4\ 5\ 8\ 9]$ ;  $B\{30\} = [2\ 3\ 4\ 7]$ ;
- \*  $A\{31\} = [1\ 2\ 4\ 8\ 9]$ ;  $B\{31\} = [1\ 2\ 3\ 7]$ ;
- \*  $A\{32\} = [1\ 2\ 4\ 8\ 9]$ ;  $B\{32\} = [1\ 3\ 5\ 7]$ ;
- \*  $A\{33\} = [1\ 2\ 4\ 8\ 9]$ ;  $B\{33\} = [2\ 3\ 4\ 7]$ ;
- \*  $A\{34\} = [1\ 2\ 4\ 8]$ ;  $B\{34\} = [2\ 3\ 7\ 8]$ ;
- \*  $A\{35\} = [1\ 2\ 4\ 8]$ ;  $B\{35\} = [3\ 5\ 7\ 8]$ ;
- \*  $A\{36\} = [2\ 3\ 4\ 8\ 9]$ ;  $B\{36\} = [1\ 2\ 3\ 5]$ ;
- \*  $A\{37\} = [4\ 7\ 8\ 9]$ ;  $B\{37\} = [1\ 3\ 4\ 6\ 7]$ ;
- \*  $A\{38\} = [4\ 7\ 8\ 9]$ ;  $B\{38\} = [1\ 2\ 3\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{39\} = [4\ 7\ 8\ 9]$ ;  $B\{39\} = [1\ 2\ 3\ 5\ 6\ 8\ 9]$ ;
- \*  $A\{40\} = [4\ 6\ 8\ 9]$ ;  $B\{40\} = [1\ 2\ 3\ 5\ 6\ 8\ 9]$ ;
- \*  $A\{41\} = [4\ 6\ 7\ 9]$ ;  $B\{41\} = [1\ 3\ 4\ 6\ 8\ 9]$ ;
- \*  $A\{42\} = [4\ 6\ 7\ 9]$ ;  $B\{42\} = [1\ 2\ 4\ 5\ 6\ 7]$ ;
- \*  $A\{43\} = [4\ 6\ 7\ 9]$ ;  $B\{43\} = [1\ 2\ 4\ 5\ 6\ 8\ 9]$ ;
- \*  $A\{44\} = [2\ 4\ 7\ 8\ 9]$ ;  $B\{44\} = [1\ 3\ 6\ 7\ 9]$ ;
- \*  $A\{45\} = [2\ 4\ 7\ 8\ 9]$ ;  $B\{45\} = [1\ 3\ 5\ 7\ 9]$ ;

- \*  $A\{46\} = [2\ 4\ 7\ 8\ 9]$ ;  $B\{46\} = [1\ 3\ 5\ 8]$ ;
- \*  $A\{47\} = [2\ 4\ 7\ 8\ 9]$ ;  $B\{47\} = [2\ 3\ 5\ 7\ 9]$ ;
- \*  $A\{48\} = [2\ 4\ 7\ 8\ 9]$ ;  $B\{48\} = [2\ 3\ 5\ 8]$ ;
- \*  $A\{49\} = [2\ 4\ 7\ 9]$ ;  $B\{49\} = [1\ 3\ 4\ 6\ 7\ 9]$ ;
- \*  $A\{50\} = [2\ 4\ 7\ 9]$ ;  $B\{50\} = [1\ 3\ 4\ 5\ 7\ 9]$ ;
- \*  $A\{51\} = [2\ 4\ 7\ 9]$ ;  $B\{51\} = [1\ 3\ 4\ 5\ 8]$ ;
- \*  $A\{52\} = [2\ 4\ 7\ 9]$ ;  $B\{52\} = [2\ 3\ 4\ 5\ 7\ 9]$ ;
- \*  $A\{53\} = [2\ 4\ 7\ 9]$ ;  $B\{53\} = [2\ 3\ 4\ 5\ 8]$ ;
- \*  $A\{54\} = [1\ 2\ 4\ 5\ 7\ 9]$ ;  $B\{54\} = [1\ 3\ 4\ 7]$ ;
- \*  $A\{55\} = [1\ 4\ 5\ 7\ 9]$ ;  $B\{55\} = [1\ 3\ 4\ 6\ 7]$ ;
- \*  $A\{56\} = [1\ 4\ 5\ 7\ 9]$ ;  $B\{56\} = [3\ 4\ 6\ 8]$ ;
- \*  $A\{57\} = [3\ 4\ 7\ 9]$ ;  $B\{57\} = [1\ 3\ 4\ 6]$ ;
- \*  $A\{58\} = [3\ 4\ 7\ 9]$ ;  $B\{58\} = [3\ 4\ 5\ 6]$ ;
- \*  $A\{59\} = [3\ 5\ 7\ 9]$ ;  $B\{59\} = [6\ 7\ 8\ 9]$ ;
- \*  $A\{60\} = [4\ 5\ 7\ 9]$ ;  $B\{60\} = [1\ 3\ 4\ 6\ 8]$ ;
- \*  $A\{61\} = [2\ 4\ 5\ 7\ 9]$ ;  $B\{61\} = [1\ 3\ 4\ 8]$ ;
- \*  $A\{62\} = [1\ 4\ 5\ 7\ 8\ 9]$ ;  $B\{62\} = [1\ 3\ 6\ 7]$ ;
- \*  $A\{63\} = [1\ 4\ 5\ 7\ 8\ 9]$ ;  $B\{63\} = [2\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{64\} = [1\ 4\ 5\ 7\ 8\ 9]$ ;  $B\{64\} = [3\ 4\ 6\ 7]$ ;
- \*  $A\{65\} = [4\ 5\ 7\ 8\ 9]$ ;  $B\{65\} = [1\ 3\ 6\ 8]$ ;
- \*  $A\{66\} = [1\ 4\ 7\ 8\ 9]$ ;  $B\{66\} = [2\ 3\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{67\} = [1\ 4\ 7\ 8\ 9]$ ;  $B\{67\} = [2\ 3\ 5\ 6\ 8\ 9]$ ;
- \*  $A\{68\} = [1\ 2\ 4\ 7\ 8\ 9]$ ;  $B\{68\} = [2\ 3\ 7\ 9]$ ;
- \*  $A\{69\} = [1\ 2\ 4\ 7\ 8\ 9]$ ;  $B\{69\} = [3\ 5\ 7\ 9]$ ;
- \*  $A\{70\} = [2\ 4\ 6\ 7\ 9]$ ;  $B\{70\} = [1\ 4\ 5\ 7]$ ;
- \*  $A\{71\} = [2\ 4\ 6\ 7\ 9]$ ;  $B\{71\} = [1\ 4\ 5\ 8]$ ;
- \*  $A\{72\} = [2\ 4\ 6\ 7\ 9]$ ;  $B\{72\} = [1\ 4\ 5\ 9]$ ;
- \*  $A\{73\} = [2\ 4\ 6\ 7\ 9]$ ;  $B\{73\} = [1\ 4\ 6\ 7]$ ;
- \*  $A\{74\} = [2\ 4\ 6\ 7\ 9]$ ;  $B\{74\} = [1\ 4\ 6\ 9]$ ;
- \*  $A\{75\} = [2\ 4\ 6\ 9]$ ;  $B\{75\} = [2\ 3\ 4\ 8]$ ;
- \*  $A\{76\} = [2\ 4\ 6\ 9]$ ;  $B\{76\} = [2\ 3\ 4\ 9]$ ;
- \*  $A\{77\} = [1\ 4\ 5\ 6\ 7\ 9]$ ;  $B\{77\} = [1\ 4\ 6\ 7]$ ;

$$* A\{78\} = [4\ 5\ 6\ 7\ 9]; B\{78\} = [1\ 4\ 6\ 8];$$

$$* A\{79\} = [5\ 6\ 7\ 9]; B\{79\} = [1\ 4\ 6\ 7\ 8];$$

$$* A\{80\} = [5\ 6\ 7\ 9]; B\{80\} = [2\ 5\ 6\ 7\ 8];$$

$$* A\{81\} = [3\ 4\ 6\ 9]; B\{81\} = [3\ 4\ 5\ 6];$$

$$* A\{82\} = [1\ 4\ 5\ 6\ 7\ 8\ 9]; B\{82\} = [2\ 5\ 6\ 7];$$

$$* A\{83\} = [1\ 4\ 5\ 6\ 7\ 8\ 9]; B\{83\} = [2\ 5\ 6\ 8\ 9];$$

$$* A\{84\} = [1\ 4\ 6\ 7\ 8\ 9]; B\{84\} = [3\ 6\ 8\ 9];$$

$$* A\{85\} = [4\ 5\ 6\ 8\ 9]; B\{85\} = [2\ 3\ 5\ 6\ 8];$$



**APÊNDICE E – Biclusters do tipo OPSM, obtidos a partir da Tabela 5 (Sarcophagidae).**

\*  $A\{1\} = [4\ 5\ 8\ 9\ 12\ 13\ 15\ 16\ 18\ 19\ 20\ 21\ 22]$ ;  $B\{1\} = [1\ 2\ 3\ 4\ 5\ 6\ 9]$ ;

Tabela 55- Demonstração em formato de tabela do bicluster OPSM de número 1, obtido através da Tabela 5.

Espécies	URBANA			RURAL			FLORESTA
	FÍGADO	CARNE	SARDINHA	FÍGADO	CARNE	SARDINHA	SARDINHA
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>9</b>
<i>O. culmiforcipes</i>	<b>4</b>	0	0	0	0	1	2
<i>P (Euboettcheria) sp</i>	<b>5</b>	0	0	0	0	1	20
<i>O. avuncula</i>	<b>8</b>	0	0	0	0	1	5
<i>H. aurescens</i>	<b>9</b>	0	0	0	0	0	0
<i>P (P) intermutans</i>	<b>12</b>	0	0	0	0	0	8
<i>P (E) australis</i>	<b>13</b>	0	0	0	0	0	7
<i>P (S) florencioi</i>	<b>15</b>	0	0	0	0	0	5
<i>P (P) chrysostoma</i>	<b>16</b>	0	0	0	0	0	0
<i>Peckia sp</i>	<b>18</b>	0	0	0	0	0	1
<i>O. angrensis</i>	<b>19</b>	0	0	0	0	0	0
<i>O. adminixta</i>	<b>20</b>	0	0	0	0	0	1
<i>P (P) pexata</i>	<b>21</b>	0	0	0	0	0	0
<i>O. parva</i>	<b>22</b>	0	0	0	0	0	0

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *O. culmiforcipes*, *P (Euboettcheria) sp*, *O. avuncula*, *H. aurescens*, *P (P) intermutans*, *P (E) australis*, *P (S) florencioi*, *P (P) chrysostoma*, *Peckia sp*, *O. angrensis*, *O. adminixta*, *P (P) pexata* e *O. parva*: 1 - Urbana (fígado) ≤ 2 - Urbana (carne) ≤ 3 - Urbana (sardinha) ≤ 4 - Rural (fígado) ≤ 5 - Rural (carne) ≤ 6 - Rural (Sardinha) ≤ 9 - Floresta (sardinha).

\*  $A\{2\} = [7\ 13\ 16\ 18\ 20\ 21\ 22\ 24]$ ;  $B\{2\} = [1\ 2\ 3\ 4\ 5\ 6\ 7]$ ;

\*  $A\{3\} = [2\ 9\ 12\ 13\ 15\ 16\ 18\ 19\ 20\ 21\ 22]$ ;  $B\{3\} = [1\ 2\ 3\ 4\ 5\ 6\ 7]$ ;

\*  $A\{4\} = [2\ 9\ 12\ 13\ 15\ 16\ 18\ 19\ 20\ 21\ 22]$ ;  $B\{4\} = [1\ 2\ 3\ 4\ 5\ 6\ 8]$ ;

\*  $A\{5\} = [2\ 9\ 12\ 13\ 15\ 16\ 18\ 19\ 20\ 21\ 22]$ ;  $B\{5\} = [1\ 2\ 3\ 4\ 5\ 6\ 9]$ ;

\*  $A\{6\} = [7\ 9\ 16\ 19\ 21\ 22\ 24]$ ;  $B\{6\} = [1\ 2\ 3\ 4\ 5\ 6\ 9]$ ;

\*  $A\{7\} = [7\ 9\ 16\ 19\ 21\ 22\ 24]$ ;  $B\{7\} = [2\ 3\ 4\ 5\ 6\ 7\ 9]$ ;

- \* A{8} = [2 13 15 16 18 20 21 22 ]; B{8} = [1 2 3 4 5 6 7 9 ];
- \* A{9} = [9 12 16 19 21 22 24 ]; B{9} = [2 3 4 5 6 7 9 ];
- \* A{10} = [2 9 12 18 19 20 24 ]; B{10} = [2 3 4 5 6 7 8 ];
- \* A{11} = [2 13 15 16 18 20 21 22 24 ]; B{11} = [2 3 4 5 6 7 9 ];
- \* A{12} = [2 9 13 15 18 20 24 ]; B{12} = [2 3 4 5 6 8 9 ];
- \* A{13} = [4 13 16 18 20 21 22 ]; B{13} = [1 2 3 4 5 6 7 9 ];
- \* A{14} = [5 8 9 12 13 15 16 18 19 20 21 22 ]; B{14} = [1 2 3 4 5 6 7 ];
- \* A{15} = [5 8 9 12 13 15 16 18 19 20 21 22 ]; B{15} = [1 2 3 4 5 6 8 ];
- \* A{16} = [5 13 15 16 18 20 21 22 ]; B{16} = [1 2 3 4 5 6 7 8 ];
- \* A{17} = [5 8 13 15 16 18 20 21 22 ]; B{17} = [1 2 3 4 5 6 7 9 ];
- \* A{18} = [5 9 12 15 16 19 21 22 ]; B{18} = [1 2 3 4 5 6 8 9 ];
- \* A{19} = [5 8 9 13 15 18 20 ]; B{19} = [1 2 3 4 5 6 8 9 ];
- \* A{20} = [4 13 16 18 20 21 22 24 ]; B{20} = [2 3 4 5 6 7 9 ];
- \* A{21} = [2 4 8 9 12 18 19 20 ]; B{21} = [1 2 3 4 5 7 8 ];
- \* A{22} = [2 4 5 8 13 15 16 18 20 21 22 ]; B{22} = [1 2 3 4 5 7 9 ];
- \* A{23} = [2 4 5 8 9 13 15 18 20 ]; B{23} = [1 2 3 4 5 8 9 ];
- \* A{24} = [8 9 12 18 19 20 24 ]; B{24} = [2 3 4 5 6 7 8 ];
- \* A{25} = [5 13 15 16 18 20 21 22 24 ]; B{25} = [2 3 4 5 6 7 8 ];
- \* A{26} = [5 8 13 15 16 18 20 21 22 24 ]; B{26} = [2 3 4 5 6 7 9 ];
- \* A{27} = [5 9 12 15 16 19 21 22 24 ]; B{27} = [2 3 4 5 6 8 9 ];
- \* A{28} = [5 8 9 13 15 18 20 24 ]; B{28} = [2 3 4 5 6 8 9 ];
- \* A{29} = [3 4 17 18 20 23 25 ]; B{29} = [1 2 3 4 5 7 8 ];

Tabela 56- Demonstração em formato de tabela do bicluster OPSM de número 29, obtido através da Tabela 5

Espécies	URBANA			RURAL	FLORESTA			
	FÍGADO	CARNE	SARDINHA	FÍGADO CARNE	FÍGADO	FÍGADO	CARNE	
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>7</b>	<b>8</b>	
<i>O. thornax</i>	<b>3</b>	10	6	19	2	4	7	3
<i>O. culmiforcipes</i>	<b>4</b>	0	0	0	0	0	0	0
<i>L. crispula</i>	<b>17</b>	0	0	2	0	0	0	0
<i>Peckia</i> sp	<b>18</b>	0	0	0	0	0	0	0
<i>O. adminixta</i>	<b>20</b>	0	0	0	0	0	0	0
<i>Helicobia</i> sp	<b>23</b>	0	0	1	0	0	0	0

<i>O. diana</i>	25	0	0	1	0	0	0	0
-----------------	----	---	---	---	---	---	---	---

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para as espécies *O. thornax*, *O. culmiforcipes*, *L. crispula*, *Peckia* sp, *O. adminixta*, *Helicobia* sp e *O. diana*: 4 – Rural (fígado) ≤ 8 – Floresta (carne) ≤ 5 - Rural (fígado) ≤ 2 – Urbana (carne) ≤ 7 – Floresta (fígado) ≤ 1 – Urbana (fígado) ≤ 3- Urbana (sardinha).

- \* A{30} = [3 4 17 18 20 23 25 ]; B{30} = [1 2 4 5 6 7 8 ];
- \* A{31} = [3 16 17 21 22 23 25 ]; B{31} = [1 2 3 4 5 7 9 ];
- \* A{32} = [3 16 17 21 22 23 25 ]; B{32} = [1 2 4 5 6 7 9 ];
- \* A{33} = [3 4 13 16 18 20 21 22 ]; B{33} = [1 2 3 4 5 6 7 ];
- \* A{34} = [4 8 9 17 18 20 23 25 ]; B{34} = [1 2 4 5 6 8 9 ];
- \* A{35} = [9 12 16 17 19 21 22 23 25 ]; B{35} = [1 2 4 5 6 7 9 ];
- \* A{36} = [8 9 12 17 18 19 20 23 25 ]; B{36} = [1 2 4 5 6 7 8 ];
- \* A{37} = [5 8 9 13 15 17 18 20 23 25 ]; B{37} = [1 2 4 5 6 8 9 ];
- \* A{38} = [5 13 15 17 18 20 23 25 ]; B{38} = [1 2 4 5 6 7 8 9 ];
- \* A{39} = [13 14 16 17 18 20 21 22 23 25 ]; B{39} = [1 2 3 4 5 6 7 ];
- \* A{40} = [9 14 17 18 20 23 25 ]; B{40} = [1 2 3 4 5 6 8 ];
- \* A{41} = [9 14 16 17 19 21 22 23 25 ]; B{41} = [1 2 3 4 5 6 9 ];
- \* A{42} = [14 16 17 21 22 23 25 ]; B{42} = [1 2 3 4 5 6 7 9 ];
- \* A{43} = [4 13 14 16 18 20 21 22 ]; B{43} = [1 2 3 4 5 6 7 ];
- \* A{44} = [4 14 17 18 20 23 25 ]; B{44} = [1 2 3 4 5 7 8 ];
- \* A{45} = [4 14 17 18 20 23 25 ]; B{45} = [1 2 4 5 6 7 8 ];
- \* A{46} = [7 17 18 20 23 24 25 ]; B{46} = [1 2 4 5 6 7 8 ];
- \* A{47} = [7 16 17 21 22 23 24 25 ]; B{47} = [1 2 4 5 6 7 9 ];
- \* A{48} = [2 9 12 17 18 19 20 23 25 ]; B{48} = [1 2 4 5 6 7 8 ];
- \* A{49} = [2 13 15 16 17 18 20 21 22 23 25 ]; B{49} = [1 2 4 5 6 7 9 ];
- \* A{50} = [2 9 13 15 17 18 20 23 25 ]; B{50} = [1 2 4 5 6 8 9 ];
- \* A{51} = [9 12 17 19 23 24 25 ]; B{51} = [2 4 5 6 7 8 9 ];
- \* A{52} = [2 17 18 20 23 24 25 ]; B{52} = [2 4 5 6 7 8 9 ];
- \* A{53} = [4 17 18 20 23 24 25 ]; B{53} = [2 4 5 6 7 8 9 ];
- \* A{54} = [2 4 8 17 18 20 23 25 ]; B{54} = [1 2 4 5 7 8 9 ];

- \*  $A\{55\} = [8\ 17\ 18\ 20\ 23\ 24\ 25]$ ;  $B\{55\} = [2\ 4\ 5\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{56\} = [5\ 13\ 15\ 17\ 18\ 20\ 23\ 24\ 25]$ ;  $B\{56\} = [2\ 4\ 5\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{57\} = [14\ 17\ 18\ 20\ 23\ 24\ 25]$ ;  $B\{57\} = [2\ 3\ 4\ 5\ 6\ 7\ 8]$ ;
- \*  $A\{58\} = [14\ 16\ 17\ 21\ 22\ 23\ 24\ 25]$ ;  $B\{58\} = [2\ 3\ 4\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{59\} = [6\ 9\ 12\ 13\ 15\ 16\ 18\ 19\ 20\ 21\ 22]$ ;  $B\{59\} = [1\ 2\ 3\ 4\ 5\ 6\ 7]$ ;
- \*  $A\{60\} = [6\ 9\ 12\ 13\ 15\ 16\ 18\ 19\ 20\ 21\ 22]$ ;  $B\{60\} = [1\ 2\ 3\ 4\ 5\ 6\ 8]$ ;
- \*  $A\{61\} = [6\ 9\ 12\ 13\ 15\ 16\ 18\ 19\ 20\ 21\ 22]$ ;  $B\{61\} = [1\ 2\ 3\ 4\ 5\ 6\ 9]$ ;
- \*  $A\{62\} = [4\ 5\ 6\ 13\ 15\ 16\ 18\ 20\ 21\ 22]$ ;  $B\{62\} = [1\ 2\ 3\ 4\ 5\ 7\ 8]$ ;
- \*  $A\{63\} = [4\ 5\ 6\ 8\ 13\ 15\ 16\ 18\ 20\ 21\ 22]$ ;  $B\{63\} = [1\ 2\ 3\ 4\ 5\ 7\ 9]$ ;
- \*  $A\{64\} = [4\ 5\ 6\ 8\ 9\ 13\ 15\ 18\ 20]$ ;  $B\{64\} = [1\ 2\ 3\ 4\ 5\ 8\ 9]$ ;
- \*  $A\{65\} = [6\ 13\ 15\ 16\ 18\ 20\ 21\ 22]$ ;  $B\{65\} = [1\ 2\ 3\ 4\ 5\ 6\ 7\ 8]$ ;
- \*  $A\{66\} = [6\ 13\ 15\ 16\ 18\ 20\ 21\ 22]$ ;  $B\{66\} = [1\ 2\ 3\ 4\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{67\} = [4\ 5\ 6\ 13\ 15\ 18\ 20]$ ;  $B\{67\} = [1\ 2\ 3\ 4\ 5\ 7\ 8\ 9]$ ;
- \*  $A\{68\} = [9\ 10\ 12\ 13\ 15\ 16\ 18\ 19\ 20\ 21\ 22]$ ;  $B\{68\} = [1\ 2\ 3\ 4\ 5\ 6\ 8]$ ;
- \*  $A\{69\} = [9\ 10\ 12\ 13\ 15\ 16\ 18\ 19\ 20\ 21\ 22]$ ;  $B\{69\} = [1\ 2\ 3\ 4\ 5\ 6\ 9]$ ;
- \*  $A\{70\} = [4\ 10\ 13\ 16\ 18\ 20\ 21\ 22]$ ;  $B\{70\} = [1\ 2\ 3\ 4\ 5\ 7\ 8]$ ;
- \*  $A\{71\} = [4\ 10\ 13\ 16\ 18\ 20\ 21\ 22]$ ;  $B\{71\} = [1\ 2\ 3\ 4\ 5\ 7\ 9]$ ;
- \*  $A\{72\} = [10\ 13\ 16\ 18\ 20\ 21\ 22]$ ;  $B\{72\} = [1\ 2\ 3\ 4\ 5\ 6\ 7\ 8]$ ;
- \*  $A\{73\} = [10\ 13\ 16\ 18\ 20\ 21\ 22]$ ;  $B\{73\} = [1\ 2\ 3\ 4\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{74\} = [9\ 10\ 12\ 15\ 16\ 19\ 21\ 22]$ ;  $B\{74\} = [1\ 2\ 3\ 4\ 5\ 6\ 8\ 9]$ ;
- \*  $A\{75\} = [5\ 9\ 10\ 12\ 15\ 16\ 19\ 21\ 22]$ ;  $B\{75\} = [1\ 2\ 3\ 4\ 5\ 8\ 9]$ ;
- \*  $A\{76\} = [6\ 10\ 13\ 15\ 16\ 18\ 20\ 21\ 22]$ ;  $B\{76\} = [1\ 2\ 3\ 4\ 6\ 7\ 8]$ ;
- \*  $A\{77\} = [6\ 10\ 13\ 15\ 16\ 18\ 20\ 21\ 22]$ ;  $B\{77\} = [1\ 2\ 3\ 4\ 6\ 7\ 9]$ ;
- \*  $A\{78\} = [1\ 11\ 13\ 16\ 17\ 18\ 20\ 21\ 22\ 23\ 25]$ ;  $B\{78\} = [1\ 2\ 3\ 4\ 5\ 6\ 7]$ ;
- \*  $A\{79\} = [1\ 11\ 13\ 16\ 17\ 18\ 20\ 21\ 22\ 23\ 25]$ ;  $B\{79\} = [1\ 2\ 4\ 5\ 6\ 7\ 8]$ ;
- \*  $A\{80\} = [1\ 11\ 13\ 16\ 17\ 18\ 20\ 21\ 22\ 23\ 25]$ ;  $B\{80\} = [1\ 2\ 4\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{81\} = [1\ 9\ 17\ 18\ 20\ 23\ 25]$ ;  $B\{81\} = [1\ 2\ 3\ 4\ 5\ 6\ 8]$ ;
- \*  $A\{82\} = [1\ 9\ 16\ 17\ 19\ 21\ 22\ 23\ 25]$ ;  $B\{82\} = [1\ 2\ 3\ 4\ 5\ 6\ 9]$ ;
- \*  $A\{83\} = [1\ 9\ 11\ 12\ 15\ 16\ 17\ 19\ 21\ 22\ 23\ 25]$ ;  $B\{83\} = [1\ 2\ 4\ 5\ 6\ 8\ 9]$ ;
- \*  $A\{84\} = [11\ 13\ 16\ 18\ 20\ 21\ 22]$ ;  $B\{84\} = [1\ 2\ 3\ 4\ 5\ 6\ 7\ 8]$ ;
- \*  $A\{85\} = [11\ 13\ 16\ 18\ 20\ 21\ 22]$ ;  $B\{85\} = [1\ 2\ 3\ 4\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{86\} = [1\ 16\ 17\ 21\ 22\ 23\ 25]$ ;  $B\{86\} = [1\ 2\ 3\ 4\ 5\ 6\ 7\ 9]$ ;

- \*  $A\{87\} = [1\ 11\ 16\ 17\ 21\ 22\ 23\ 25]$ ;  $B\{87\} = [1\ 2\ 4\ 5\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{88\} = [9\ 11\ 12\ 13\ 15\ 16\ 18\ 19\ 20\ 21\ 22]$ ;  $B\{88\} = [1\ 2\ 3\ 4\ 5\ 6\ 8]$ ;
- \*  $A\{89\} = [9\ 11\ 12\ 13\ 15\ 16\ 18\ 19\ 20\ 21\ 22]$ ;  $B\{89\} = [1\ 2\ 3\ 4\ 5\ 6\ 9]$ ;
- \*  $A\{90\} = [9\ 11\ 12\ 15\ 16\ 19\ 21\ 22]$ ;  $B\{90\} = [1\ 2\ 3\ 4\ 5\ 6\ 8\ 9]$ ;
- \*  $A\{91\} = [4\ 11\ 13\ 16\ 17\ 18\ 20\ 21\ 22\ 23\ 25]$ ;  $B\{91\} = [1\ 2\ 4\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{92\} = [4\ 11\ 13\ 16\ 18\ 20\ 21\ 22]$ ;  $B\{92\} = [1\ 2\ 3\ 4\ 5\ 7\ 8]$ ;
- \*  $A\{93\} = [4\ 11\ 13\ 16\ 18\ 20\ 21\ 22]$ ;  $B\{93\} = [1\ 2\ 3\ 4\ 5\ 7\ 9]$ ;
- \*  $A\{94\} = [5\ 9\ 11\ 12\ 15\ 16\ 19\ 21\ 22]$ ;  $B\{94\} = [1\ 2\ 3\ 4\ 5\ 8\ 9]$ ;
- \*  $A\{95\} = [5\ 11\ 13\ 15\ 16\ 17\ 18\ 20\ 21\ 22\ 23\ 25]$ ;  $B\{95\} = [1\ 2\ 4\ 5\ 6\ 7\ 8]$ ;
- \*  $A\{96\} = [5\ 8\ 11\ 13\ 15\ 16\ 17\ 18\ 20\ 21\ 22\ 23\ 25]$ ;  $B\{96\} = [1\ 2\ 4\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{97\} = [5\ 9\ 11\ 12\ 15\ 16\ 17\ 19\ 21\ 22\ 23\ 25]$ ;  $B\{97\} = [1\ 2\ 4\ 5\ 6\ 8\ 9]$ ;
- \*  $A\{98\} = [5\ 11\ 15\ 16\ 17\ 21\ 22\ 23\ 25]$ ;  $B\{98\} = [1\ 2\ 4\ 5\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{99\} = [1\ 4\ 17\ 18\ 20\ 23\ 25]$ ;  $B\{99\} = [1\ 2\ 3\ 4\ 5\ 7\ 8]$ ;
- \*  $A\{100\} = [1\ 5\ 11\ 15\ 16\ 17\ 21\ 22\ 23\ 25]$ ;  $B\{100\} = [1\ 4\ 5\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{101\} = [6\ 11\ 13\ 15\ 16\ 17\ 18\ 20\ 21\ 22\ 23\ 25]$ ;  $B\{101\} = [1\ 2\ 4\ 5\ 6\ 7\ 8]$ ;
- \*  $A\{102\} = [6\ 11\ 13\ 15\ 16\ 17\ 18\ 20\ 21\ 22\ 23\ 25]$ ;  $B\{102\} = [1\ 2\ 4\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{103\} = [6\ 9\ 13\ 15\ 17\ 18\ 20\ 23\ 25]$ ;  $B\{103\} = [1\ 2\ 4\ 5\ 6\ 8\ 9]$ ;
- \*  $A\{104\} = [6\ 13\ 15\ 17\ 18\ 20\ 23\ 25]$ ;  $B\{104\} = [1\ 2\ 4\ 5\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{105\} = [5\ 6\ 13\ 15\ 17\ 18\ 20\ 23\ 25]$ ;  $B\{105\} = [1\ 4\ 5\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{106\} = [4\ 5\ 6\ 13\ 15\ 17\ 18\ 20\ 23\ 25]$ ;  $B\{106\} = [1\ 2\ 4\ 5\ 7\ 8\ 9]$ ;
- \*  $A\{107\} = [10\ 11\ 13\ 16\ 17\ 18\ 20\ 21\ 22\ 23\ 25]$ ;  $B\{107\} = [1\ 2\ 4\ 5\ 6\ 7\ 8]$ ;
- \*  $A\{108\} = [10\ 11\ 13\ 16\ 17\ 18\ 20\ 21\ 22\ 23\ 25]$ ;  $B\{108\} = [1\ 2\ 4\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{109\} = [9\ 10\ 11\ 12\ 15\ 16\ 17\ 19\ 21\ 22\ 23\ 25]$ ;  $B\{109\} = [1\ 2\ 4\ 5\ 6\ 8\ 9]$ ;
- \*  $A\{110\} = [10\ 11\ 16\ 17\ 21\ 22\ 23\ 25]$ ;  $B\{110\} = [1\ 2\ 4\ 5\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{111\} = [10\ 11\ 15\ 16\ 17\ 21\ 22\ 23\ 25]$ ;  $B\{111\} = [1\ 2\ 4\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{112\} = [1\ 10\ 11\ 16\ 17\ 21\ 22\ 23\ 25]$ ;  $B\{112\} = [1\ 2\ 4\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{113\} = [10\ 13\ 16\ 18\ 20\ 21\ 22\ 24]$ ;  $B\{113\} = [2\ 3\ 4\ 5\ 6\ 7\ 8]$ ;
- \*  $A\{114\} = [10\ 13\ 16\ 18\ 20\ 21\ 22\ 24]$ ;  $B\{114\} = [2\ 3\ 4\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{115\} = [9\ 10\ 12\ 15\ 16\ 19\ 21\ 22\ 24]$ ;  $B\{115\} = [2\ 3\ 4\ 5\ 6\ 8\ 9]$ ;
- \*  $A\{116\} = [6\ 13\ 15\ 16\ 18\ 20\ 21\ 22\ 24]$ ;  $B\{116\} = [2\ 3\ 4\ 5\ 6\ 7\ 8]$ ;
- \*  $A\{117\} = [6\ 13\ 15\ 16\ 18\ 20\ 21\ 22\ 24]$ ;  $B\{117\} = [2\ 3\ 4\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{118\} = [6\ 9\ 13\ 15\ 18\ 20\ 24]$ ;  $B\{118\} = [2\ 3\ 4\ 5\ 6\ 8\ 9]$ ;

- \*  $A\{119\} = [6\ 13\ 15\ 17\ 18\ 20\ 23\ 24\ 25]$ ;  $B\{119\} = [2\ 4\ 5\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{120\} = [4\ 5\ 6\ 13\ 15\ 18\ 20\ 24]$ ;  $B\{120\} = [2\ 3\ 4\ 5\ 7\ 8\ 9]$ ;
- \*  $A\{121\} = [11\ 13\ 16\ 18\ 20\ 21\ 22\ 24]$ ;  $B\{121\} = [2\ 3\ 4\ 5\ 6\ 7\ 8]$ ;
- \*  $A\{122\} = [11\ 13\ 16\ 18\ 20\ 21\ 22\ 24]$ ;  $B\{122\} = [2\ 3\ 4\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{123\} = [9\ 11\ 12\ 15\ 16\ 19\ 21\ 22\ 24]$ ;  $B\{123\} = [2\ 3\ 4\ 5\ 6\ 8\ 9]$ ;
- \*  $A\{124\} = [5\ 11\ 15\ 16\ 17\ 21\ 22\ 23\ 24\ 25]$ ;  $B\{124\} = [2\ 4\ 5\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{125\} = [10\ 11\ 16\ 17\ 21\ 22\ 23\ 24\ 25]$ ;  $B\{125\} = [2\ 4\ 5\ 6\ 7\ 8\ 9]$ ;
- \*  $A\{126\} = [1\ 17\ 18\ 20\ 23\ 24\ 25]$ ;  $B\{126\} = [2\ 3\ 4\ 5\ 6\ 7\ 8]$ ;
- \*  $A\{127\} = [1\ 16\ 17\ 21\ 22\ 23\ 24\ 25]$ ;  $B\{127\} = [2\ 3\ 4\ 5\ 6\ 7\ 9]$ ;
- \*  $A\{128\} = [1\ 11\ 16\ 17\ 21\ 22\ 23\ 24\ 25]$ ;  $B\{128\} = [2\ 4\ 5\ 6\ 7\ 8\ 9]$ ;

**APÊNDICE F – Biclusters do tipo OPSM, obtidos a partir da Tabela 6 (Muscidae).**

Apenas como exemplificação, será demonstrado de forma mais detalhada o bicluster de número 1. As interpretações dos demais será mais direta.

Abaixo, tem-se a demonstração do bicluster de número 1.

$$* A\{1\} = [1\ 2\ 3\ 4\ 5]; B\{1\} = [1\ 2\ 3\ 7];$$

Tabela 35 – Bicluster OPSM de número 1, da análise 3.

		URBANA			FLORESTA
		FÍGADO	CARNE	SARDINHA	FÍGADO
		<b>1</b>	<b>2</b>	<b>3</b>	<b>7</b>
<i>A. orientalis</i>	<b>1</b>	14	225	132	0
<i>S.nudiseta</i>	<b>2</b>	1	17	15	0
<i>M. domestica</i>	<b>3</b>	1	26	5	0
<i>O. chalcogaster</i>	<b>4</b>	0	21	4	0
<i>Muscina sp</i>	<b>5</b>	0	1	0	0

Observando atentamente a Tabela 36, temos a seguinte ordenação nas colunas:

$$\text{Coluna 7} \leq \text{Coluna 1} \leq \text{Coluna 3} \leq \text{Coluna 2}$$

Ou seja, as espécies presentes na Tabela 36, que foram coletadas no ambiente de floresta com a isca fígado, apresentam valores (de indivíduos coletados) menores ou iguais àquelas coletadas na área urbana com a isca de fígado e estes são menores ou iguais àquelas também da área urbana, mas com a isca de sardinha e estes últimas são menores ou iguais do que aquelas coletadas na área urbana com a isca de carne.

$$* A\{2\} = [1\ 2\ 3\ 4\ 5]; B\{2\} = [2\ 3\ 6\ 7];$$

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para todas as espécies: 7 - Floresta (fígado) ≤ 6 - Rural (sardinha) ≤ 3 - Urbana (sardinha) ≤ 2 - Urbana (carne).

$$* A\{3\} = [1\ 2\ 3\ 4\ 5]; B\{3\} = [2\ 3\ 7\ 9];$$

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para todas as espécies: 7 - Floresta (fígado) ≤ 9 - Floresta (sardinha) ≤ 3 - Urbana (sardinha) ≤ 2 - Urbana (carne).

$$* A\{4\} = [1\ 2\ 3\ 4\ 5]; B\{4\} = [2\ 7\ 8\ 9];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para todas as espécies: 7 - Floresta (fígado) ≤ 9 - Floresta (sardinha) ≤ 8 - Floresta (carne) ≤ 2 - Urbana (carne).

$$* A\{5\} = [1\ 2\ 3\ 4\ 5]; B\{5\} = [4\ 5\ 7];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para todas as espécies: 7 - Floresta (fígado) ≤ 4 - Rural (fígado) ≤ 5 - Rural (carne).

$$* A\{6\} = [3\ 4\ 5]; B\{6\} = [1\ 3\ 4\ 5\ 6\ 7\ 9];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *M. domestica*, *O. chalcogaster* e *Muscina* sp : 7 - Floresta (fígado) ≤ 9 - Floresta (sardinha) ≤ 1 - Urbana (fígado) ≤ 6 - Rural (sardinha) ≤ 3 - Urbana (sardinha) ≤ 4 - Rural (fígado) ≤ 5 - Rural (carne).

$$* A\{7\} = [3\ 4\ 5]; B\{7\} = [3\ 4\ 5\ 6\ 7\ 8\ 9];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *M. domestica*, *O. chalcogaster* e *Muscina* sp : 7 - Floresta (fígado) ≤ 9 - Floresta (sardinha) ≤ 8 - Floresta (carne) ≤ 6 - Rural (sardinha) ≤ 3 - Urbana (sardinha) ≤ 4 - Rural (fígado) ≤ 5 - Rural (carne).

$$* A\{8\} = [1\ 2\ 5]; B\{8\} = [1\ 2\ 3\ 4\ 6\ 7];$$



que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *S. nudiseta* e *Muscina* sp: 7 -Floresta (fígado) ≤ 4- Rural (fígado) ≤ 6- Rural (sardinha) ≤ 1 – Urbana (fígado) ≤ 3- Urbana (sardinha) ≤ 2 – Urbana (carne).

$$* A\{9\} = [1\ 2\ 5\ ]; B\{9\} = [2\ 3\ 4\ 5\ 7\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *S. nudiseta* e *Muscina* sp: 7 -Floresta (fígado) ≤ 4- Rural (fígado) ≤ 9- Floresta (sardinha) ≤ 3- Urbana (sardinha) ≤ 2 – Urbana (carne) ≤ 5 – Rural (carne).

$$* A\{10\} = [1\ 2\ 5\ ]; B\{10\} = [2\ 4\ 5\ 7\ 8\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *S. nudiseta* e *Muscina* sp: 7 -Floresta (fígado) ≤ 4- Rural (fígado) ≤ 9- Floresta (sardinha) ≤ 8- Floresta (carne) ≤ 2 – Urbana (carne) ≤ 5 – Rural (carne).

$$* A\{11\} = [1\ 3\ 4\ 5\ ]; B\{11\} = [1\ 5\ 7\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *M. domestica*, *O. chalcogaster* e *Muscina* sp: 7 -Floresta (fígado) ≤ 1- Urbana (fígado) ≤ 5 - Rural (carne).

$$* A\{12\} = [1\ 3\ 4\ 5\ ]; B\{12\} = [5\ 6\ 7\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *M. domestica*, *O. chalcogaster* e *Muscina* sp: 7 -Floresta (fígado) ≤ 6- Rural (sardinha) ≤ 5 - Rural (carne).

$$* A\{13\} = [1\ 3\ 5\ ]; B\{13\} = [1\ 5\ 6\ 7\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *M. domestica* e *Muscina* sp: 7 -Floresta (fígado) ≤ 6- Rural (sardinha) ≤ 1- Urbana (fígado) ≤ 5 - Rural (carne).

$$* A\{14\} = [1\ 4\ 5\ ]; B\{14\} = [1\ 2\ 5\ 7\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *O. chalcogaster* e *Muscina* sp: 7 -Floresta (fígado) ≤ 1- Urbana (fígado) ≤ 2- Urbana (carne) ≤ 5 - Rural (carne).

$$* A\{15\} = [1\ 4\ 5\ ]; B\{15\} = [1\ 2\ 3\ 7\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *O. chalcogaster* e *Muscina* sp: 7 -Floresta (fígado) ≤ 1- Urbana (fígado) ≤ 9- Floresta (sardinha) ≤ 3- Urbana (sardinha) ≤ 2- Urbana (carne).

$$* A\{16\} = [1\ 4\ 5\ ]; B\{16\} = [1\ 2\ 7\ 8\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *O. chalcogaster* e *Muscina* sp: 7 -Floresta (fígado) ≤ 1- Urbana (fígado) ≤ 9- Floresta (sardinha) ≤ 8- Floresta (carne) ≤ 2- Urbana (carne)

$$* A\{17\} = [1\ 4\ 5\ ]; B\{17\} = [2\ 5\ 6\ 7\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *O. chalcogaster* e *Muscina* sp: 7 -Floresta (fígado) ≤ 6- Rural (sardinha) ≤ 2- Urbana (carne) ≤ 5 – Rural (carne).

$$* A\{18\} = [1\ 4\ 5\ ]; B\{18\} = [2\ 6\ 7\ 8\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *O. chalcogaster* e *Muscina* sp: 7 -Floresta (fígado) ≤ 6- Rural (sardinha) ≤ 8- Floresta (carne) ≤ 2- Urbana (carne).

$$* A\{19\} = [2\ 3\ 4\ 5\ ]; B\{19\} = [1\ 2\ 3\ 6\ 7\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *S. nudiseta*, *M. domestica*, *O. chalcogaster* e *Muscina* sp: 7 -Floresta (fígado) ≤ 9 - Floresta (sardinha) ≤ 1- Urbana (fígado) ≤ 6- Rural (sardinha) ≤ 3- Urbana (sardinha) ≤ 2- Urbana (carne).

$$* A\{20\} = [2\ 3\ 4\ 5\ ]; B\{20\} = [2\ 3\ 6\ 7\ 8\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *S. nudiseta*, *M. domestica*, *O. chalcogaster* e *Muscina* sp: 7 -Floresta (fígado)  $\leq$  9 - Floresta (sardinha)  $\leq$  8- Floresta (carne)  $\leq$  6- Rural (sardinha)  $\leq$  3- Urbana (sardinha)  $\leq$  2- Urbana (carne).

$$* A\{21\} = [2\ 3\ 4\ 5\ ]; B\{21\} = [4\ 5\ 7\ 8\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *S. nudiseta*, *M. domestica*, *O. chalcogaster* e *Muscina* sp: 7 -Floresta (fígado)  $\leq$  9 - Floresta (sardinha)  $\leq$  8- Floresta (carne)  $\leq$  4- Rural (fígado)  $\leq$  5- Rural (carne).

$$* A\{22\} = [2\ 3\ 5\ ]; B\{22\} = [1\ 2\ 3\ 6\ 7\ 8\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *S. nudiseta*, *M. domestica* e *Muscina* sp: 7 -Floresta (fígado)  $\leq$  9 - Floresta (sardinha)  $\leq$  8- Floresta (carne)  $\leq$  6- Rural (sardinha)  $\leq$  1- Urbana (fígado)  $\leq$  3- Urbana (sardinha)  $\leq$  2 - Urbana (carne).

$$* A\{23\} = [2\ 4\ 5\ ]; B\{23\} = [2\ 4\ 5\ 7\ 8\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *S. nudiseta*, *O. chalcogaster* e *Muscina* sp: 7 -Floresta (fígado)  $\leq$  9 - Floresta (sardinha)  $\leq$  8- Floresta (carne)  $\leq$  4 -Rural (fígado)  $\leq$  5- Rural (carne)  $\leq$  2 Urbana (carne).

$$* A\{24\} = [1\ 2\ 3\ 5\ ]; B\{24\} = [1\ 2\ 3\ 6\ 7\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *S. nudiseta*, *M. domestica* e *Muscina* sp: 7 -Floresta (fígado)  $\leq$  6 - Rural (sardinha)  $\leq$  1 Urbana (fígado)  $\leq$  3- Urbana (sardinha)  $\leq$  2 Urbana (carne).

$$* A\{25\} = [1\ 2\ 4\ 5\ ]; B\{25\} = [2\ 4\ 5\ 7\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *A. orientalis*, *S. nudiseta*, *O. chalcogaster* e *Muscina* sp: 7 -Floresta (fígado)  $\leq$  4 -Rural (fígado)  $\leq$  2 Urbana (carne)  $\leq$  5- Rural (carne).

**APÊNDICE G – Biclusters do tipo OPSM, obtidos a partir da Tabela 7 (Calliphoridae).**

$$* A\{1\} = [1\ 2\ 3\ 4\ 5]; B\{1\} = [1\ 4\ 7];$$

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para todas as espécies: 4 - Rural (fígado) ≤ 1- Urbana (fígado) ≤ 7 - Floresta (fígado).

$$* A\{2\} = [1\ 2\ 3\ 4]; B\{2\} = [1\ 2\ 4\ 7];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. segmentaria*, *H. semidiphana* e *C. megacephala*: 2- Urbana (carne) ≤ 4- Rural (fígado) ≤ 1 - Urbana (fígado) ≤ 7 - Floresta (fígado).

$$* A\{3\} = [1\ 2\ 3\ 4]; B\{3\} = [1\ 3\ 4\ 7];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. segmentaria*, *H. semidiphana* e *C. megacephala*: 3 - Urbana (sardinha) ≤ 4- Rural (fígado) ≤ 1 - Urbana (fígado) ≤ 7 - Floresta (fígado).

$$* A\{4\} = [1\ 2\ 3\ 4]; B\{4\} = [1\ 5\ 7];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. segmentaria*, *H. semidiphana* e *C. megacephala*: 5- Rural (carne) ≤ 1 - Urbana (fígado) ≤ 7 - Floresta (fígado).

$$* A\{5\} = [1\ 2\ 3\ 4]; B\{5\} = [6\ 7\ 9];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. segmentaria*, *H. semidiphana* e *C. megacephala*: 6 - Rural (sardinha) ≤ 9 – Floresta (sardinha) ≤ 7 - Floresta (fígado).

$$* A\{6\} = [1\ 2\ 3\ 4]; B\{6\} = [2\ 7\ 9];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. segmentaria*, *H. semidiphana* e *C. megacephala*: 2- Urbana (carne) ≤ 9 – Floresta (sardinha) ≤ 7 - Floresta (fígado).

$$* A\{7\} = [1\ 2\ 3\ 4]; B\{7\} = [5\ 7\ 9];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. segmentaria*, *H. semidiphana* e *C. megacephala*: 5 – Rural (carne) ≤ 9 – Floresta (sardinha) ≤ 7 - Floresta (fígado).

$$* A\{8\} = [1\ 2\ 3\ 4]; B\{8\} = [2\ 4\ 8];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. segmentaria*, *H. semidiphana* e *C. megacephala*: 2- Urbana (carne) ≤ 4- Rural (fígado) ≤ 8 - Floresta (carne).

$$* A\{9\} = [1\ 2\ 3\ 4]; B\{9\} = [3\ 4\ 8];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. segmentaria*, *H. semidiphana* e *C. megacephala*: 3- Urbana (sardinha) ≤ 4- Rural (fígado) ≤ 8 - Floresta (carne).

$$* A\{10\} = [1\ 3\ 4]; B\{10\} = [1\ 2\ 4\ 5\ 6\ 7];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. semidiphana* e *C. megacephala*: 6- Rural (sardinha) ≤ 5- Rural (carne) ≤ 2- Urbana (carne) ≤ 4- Rural (fígado) ≤ 1 - Urbana (fígado) ≤ 7 - Floresta (fígado).

$$* A\{11\} = [1\ 3\ 4]; B\{11\} = [2\ 4\ 5\ 6\ 8];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. semidiphana* e *C. megacephala*: 6- Rural (sardinha) ≤ 5- Rural (carne) ≤ 2- Urbana (carne) ≤ 4- Rural (fígado) ≤ 8 – Floresta (carne).

$$* A\{12\} = [1\ 3\ 4]; B\{12\} = [2\ 5\ 6\ 7\ 9];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. semidiphana* e *C. megacephala*: 6- Rural (sardinha) ≤ 5- Rural (carne) ≤ 2- Urbana (carne) ≤ 9 – Floresta (sardinha) ≤ 7 - Floresta (fígado).

$$* A\{13\} = [1\ 3\ 4]; B\{13\} = [1\ 3\ 4\ 6\ 7];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. semidiphana* e *C. megacephala*: 6- Rural (sardinha) ≤ 3- Urbana (sardinha) ≤ 4- Rural (fígado) ≤ 1 - Urbana (fígado) ≤ 7 - Floresta (fígado).

$$* A\{14\} = [1\ 3\ 4]; B\{14\} = [3\ 4\ 6\ 8];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. semidiphana* e *C. megacephala*: 6- Rural (sardinha) ≤ 3 - Urbana (sardinha) ≤ 4- Rural (fígado) ≤ 8 - Floresta (carne).

$$* A\{15\} = [2\ 3\ 4]; B\{15\} = [1\ 2\ 3\ 4\ 7];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *H. segmentaria*, *H. semidiphana* e *C. megacephala*: 2- Urbana (carne) ≤ 3- Urbana (sardinha) ≤ 4- Rural (fígado) ≤ 1 - Urbana (fígado) ≤ 7 - Floresta (fígado).

$$* A\{16\} = [2\ 3\ 4]; B\{16\} = [1\ 2\ 3\ 4\ 8];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *H. segmentaria*, *H. semidiphana* e *C. megacephala*: 2- Urbana (carne) ≤ 3- Urbana (sardinha) ≤ 4- Rural (fígado) ≤ 1 - Urbana (fígado) ≤ 8- Floresta (carne).

$$* A\{17\} = [2\ 3\ 4]; B\{17\} = [1\ 5\ 8];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *H. segmentaria*, *H. semidiphana* e *C. megacephala*: 5 - Rural (carne) ≤ 1 - Urbana (fígado) ≤ 8 - Floresta (carne).

$$* A\{18\} = [1\ 2\ 3]; B\{18\} = [1\ 2\ 3\ 4\ 7];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. segmentaria* e *H. semidiphana* : 3- Urbana (sardinha) ≤ 2- Urbana (carne) ≤ 4- Rural (fígado) ≤ 1 - Urbana (fígado) ≤ 7 - Floresta (fígado).

$$* A\{19\} = [1\ 2\ 3]; B\{19\} = [2\ 3\ 4\ 7\ 8];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *Lucilia eximia*, *H. segmentaria* e *H. semidiphana* : 3- Urbana (sardinha) ≤ 2- Urbana (carne) ≤ 4- Rural (fígado) ≤ 8 – Floresta (carne) ≤ 7 - Floresta (fígado).

$$* A\{20\} = [1\ 2\ 3\ ]; B\{20\} = [2\ 3\ 7\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *Lucilia eximia*, *H. segmentaria* e *H. semidiphana* : 3- Urbana (sardinha) ≤ 2- Urbana (carne) ≤ 9- Floresta (sardinha) ≤ 7 - Floresta (fígado).

$$* A\{21\} = [1\ 2\ 3\ ]; B\{21\} = [1\ 3\ 5\ 7\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *Lucilia eximia*, *H. segmentaria* e *H. semidiphana* : 3- Urbana (sardinha) ≤ 5 – Rural (carne) ≤ 1 - Urbana (fígado) ≤ 7 - Floresta (fígado).

$$* A\{22\} = [1\ 2\ 3\ ]; B\{22\} = [3\ 5\ 7\ 8\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *Lucilia eximia*, *H. segmentaria* e *H. semidiphana* : 3- Urbana (sardinha) ≤ 5 - Rural (carne) ≤ 8 - Floresta (carne) ≤ 7 - Floresta (fígado).

$$* A\{23\} = [1\ 2\ 3\ ]; B\{23\} = [3\ 5\ 7\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *Lucilia eximia*, *H. segmentaria* e *H. semidiphana* : 3- Urbana (sardinha) ≤ 5 - Rural (carne) ≤ 9 - Floresta (sardinha) ≤ 7 - Floresta (fígado).

$$* A\{24\} = [2\ 3\ 5\ ]; B\{24\} = [1\ 4\ 5\ 7\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *H. segmentaria*, *H. semidiphana* e *C. albiceps* : 4- Rural (fígado) ≤ 1- Urbana (fígado) ≤ 5 - Rural (carne) ≤ 7 - Floresta (fígado).

$$* A\{25\} = [2\ 3\ 5\ ]; B\{25\} = [1\ 4\ 5\ 8\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *H. segmentaria*, *H. semidiphana* e *C. albiceps* : 4- Rural (fígado) ≤ 1- Urbana (fígado) ≤ 5 - Rural (carne) ≤ 8 - Floresta (carne).

$$* A\{26\} = [2\ 3\ 5\ ]; B\{26\} = [1\ 4\ 5\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *H. segmentaria*, *H. semidiphana* e *C. albiceps* : 4- Rural (fígado) ≤ 1- Urbana (fígado) ≤ 6 - Rural (sardinha) ≤ 9 - Floresta (sardinha).

$$* A\{27\} = [2\ 3\ 5\ ]; B\{27\} = [1\ 4\ 6\ 7\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *H. segmentaria*, *H. semidiphana* e *C. albiceps* : 4- Rural (fígado) ≤ 1- Urbana (fígado) ≤ 6 - Rural (sardinha) ≤ 7 - Floresta (fígado).

$$* A\{28\} = [2\ 3\ 5\ ]; B\{28\} = [1\ 4\ 6\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *H. segmentaria*, *H. semidiphana* e *C. albiceps* : 4- Rural (fígado) ≤ 1- Urbana (fígado) ≤ 6 - Rural (sardinha) ≤ 9 - Floresta (sardinha).

$$* A\{29\} = [2\ 3\ 5\ ]; B\{29\} = [2\ 3\ 4\ 8\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *H. segmentaria*, *H. semidiphana* e *C. albiceps* : 4- Rural (fígado) ≤ 2- Urbana (carne) ≤ 3 - Urbana (sardinha) ≤ 8- Floresta (carne).

$$* A\{30\} = [2\ 3\ 5\ ]; B\{30\} = [2\ 3\ 4\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *H. segmentaria*, *H. semidiphana* e *C. albiceps* : 4- Rural (fígado) ≤ 2- Urbana (carne) ≤ 3 - Urbana (sardinha) ≤ 9- Floresta (sardinha).

$$* A\{31\} = [2\ 3\ 5\ ]; B\{31\} = [2\ 4\ 7\ ];$$



que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *H. segmentaria*, *H. semidiphana* e *C. albiceps* : 4- Rural (fígado) ≤ 2- Urbana (carne) ≤ 7- Floresta (fígado).

$$* A\{32\} = [1\ 3\ 4\ 5\ ]; B\{32\} = [1\ 4\ 6\ 7\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. semidiphana*, *C. megacephala* e *C. albiceps* : 6 – Rural (sardinha) ≤ 4- Rural (fígado) ≤ 1- Urbana (fígado) ≤ 7- Floresta (fígado).

$$* A\{33\} = [1\ 3\ 4\ 5\ ]; B\{33\} = [2\ 5\ 6\ 7\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. semidiphana*, *C. megacephala* e *C. albiceps* : 6 – Rural (sardinha) ≤ 5- Rural (carne) ≤ 2 - Urbana (carne) ≤ 7- Floresta (fígado).

$$* A\{34\} = [1\ 3\ 4\ 5\ ]; B\{34\} = [2\ 5\ 6\ 8\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. semidiphana*, *C. megacephala* e *C. albiceps* : 6 – Rural (sardinha) ≤ 5- Rural (carne) ≤ 2 - Urbana (carne) ≤ 9- Floresta (sardinha) ≤ 8 - Floresta (carne).

$$* A\{35\} = [1\ 3\ 4\ 5\ ]; B\{35\} = [3\ 6\ 8\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. semidiphana*, *C. megacephala* e *C. albiceps* : 6 – Rural (sardinha) ≤ 3 – Urbana (sardinha) ≤ 8 - Floresta (carne).

$$* A\{36\} = [1\ 3\ 4\ 5\ ]; B\{36\} = [4\ 6\ 8\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. semidiphana*, *C. megacephala* e *C. albiceps* : 6 – Rural (sardinha) ≤ 4 – Rural (fígado) ≤ 8 - Floresta (carne).

$$* A\{37\} = [3\ 4\ 5\ ]; B\{37\} = [1\ 4\ 6\ 8\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *H. semidiphana*, *C. megacephala* e *C. albiceps* : 6 - Rural (sardinha)  $\leq$  4 – Rural (fígado)  $\leq$  1 – Urbana (fígado)  $\leq$  8 - Floresta (carne).

$$* A\{38\} = [3\ 4\ 5\ ]; B\{38\} = [2\ 3\ 5\ 6\ 8\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *H. semidiphana*, *C. megacephala* e *C. albiceps* : 6 - Rural (sardinha)  $\leq$  5 - Rural (carne)  $\leq$  2 - Urbana (carne)  $\leq$  3 – Urbana (sardinha)  $\leq$  8 - Floresta (carne).

$$* A\{39\} = [1\ 3\ 5\ ]; B\{39\} = [3\ 6\ 8\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *L. eximia*, *H. semidiphana* e *C. albiceps* : 6 - Rural (sardinha)  $\leq$  3 – Urbana (sardinha)  $\leq$  9 - Floresta (sardinha)  $\leq$  8 – Floresta (carne).

$$* A\{40\} = [2\ 3\ 4\ 5\ ]; B\{40\} = [2\ 3\ 8\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *H. segmentaria*, *H. semidiphana*, *C. megacephala* e *C. albiceps* : 2 – Urbana (carne)  $\leq$  3 – Urbana (sardinha)  $\leq$  8 – Floresta (carne).

$$* A\{41\} = [2\ 3\ 4\ 5\ ]; B\{41\} = [1\ 4\ 8\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *H. segmentaria*, *H. semidiphana*, *C. megacephala* e *C. albiceps* : 2 – Urbana (carne)  $\leq$  3 – Urbana (sardinha)  $\leq$  8 – Floresta (carne).

**APÊNDICE H – Biclusters do tipo OPSM, obtidos a partir da Tabela 8 (Sarcophagidae).**

$$* A\{1\} = [1\ 2\ 3\ 4\ 5]; B\{1\} = [2\ 4\ 5];$$

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para todas as espécies: 4 – Rural (fígado) ≤ 5- Rural (carne) ≤ 2- Urbana (carne).

$$* A\{2\} = [1\ 2\ 3\ 4\ 5]; B\{2\} = [3\ 4\ 5];$$

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para todas as espécies: 4 – Rural (fígado) ≤ 5- Rural (carne) ≤ 3- Urbana (sardinha).

$$* A\{3\} = [1\ 2\ 3\ 4\ 5]; B\{3\} = [4\ 5\ 6];$$

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para todas as espécies: 4 – Rural (fígado) ≤ 5- Rural (carne) ≤ 6 – Rural (sardinha).

$$* A\{4\} = [1\ 2\ 3\ 4\ 5]; B\{4\} = [4\ 5\ 7];$$

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para todas as espécies: 4 – Rural (fígado) ≤ 5- Rural (carne) ≤ 7- Floresta (fígado).

$$* A\{5\} = [1\ 2\ 3\ 4\ 5]; B\{5\} = [4\ 5\ 9];$$

que refere-se à seguinte ordenação das colunas referente ao número de indivíduos coletados para todas as espécies: 4 – Rural (fígado) ≤ 5- Rural (carne) ≤ 9- Floresta (sardinha).

$$* A\{6\} = [1\ 3\ 4\ 5]; B\{6\} = [1\ 2\ 4\ 5\ 8];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (S) lambens*, *P. (Euboetcheria) sp*, *P. (E) collusor*, *P. (P) intermutans*: 1 – Urbana (fígado) ≤ 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 2- Urbana (carne) ≤ 8- Floresta (carne).

$$* A\{7\} = [1\ 3\ 4\ 5]; B\{7\} = [1\ 2\ 4\ 5\ 9];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (S) lambens*, *P (Euboetcheria) sp*, *P. (E) collusor*, *P. (P) intermutans*: 1 – Urbana (fígado) ≤ 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 2- Urbana (carne) ≤ 9- Floresta (sardinha).

$$* A\{8\} = [1\ 3\ 4\ 5\ ]; B\{8\} = [1\ 3\ 4\ 5\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (S) lambens*, *P (Euboetcheria) sp*, *P. (E) collusor*, *P. (P) intermutans*: 1 – Urbana (fígado) ≤ 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 3- Urbana (sardinha).

$$* A\{9\} = [1\ 3\ 4\ 5\ ]; B\{9\} = [1\ 4\ 5\ 6\ 7\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (S) lambens*, *P (Euboetcheria) sp*, *P. (E) collusor*, *P. (P) intermutans*: 1 – Urbana (fígado) ≤ 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 6- Rural (sardinha) ≤ 7- Floresta (fígado).

$$* A\{10\} = [1\ 3\ 4\ 5\ ]; B\{10\} = [1\ 4\ 5\ 6\ 8\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (S) lambens*, *P (Euboetcheria) sp*, *P. (E) collusor*, *P. (P) intermutans*: 1 – Urbana (fígado) ≤ 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 6- Rural (sardinha) ≤ 8- Floresta (carne).

$$* A\{11\} = [1\ 3\ 4\ 5\ ]; B\{11\} = [1\ 4\ 5\ 6\ 9\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (S) lambens*, *P (Euboetcheria) sp*, *P. (E) collusor*, *P. (P) intermutans*: 1 – Urbana (fígado) ≤ 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 6- Rural (sardinha) ≤ 9- Floresta (sardinha).

$$* A\{12\} = [1\ 3\ 5\ ]; B\{12\} = [1\ 2\ 3\ 4\ 5\ ];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (S) lambens*, *P (Euboetcheria) sp*, *P. (P) intermutans*: 1 – Urbana (fígado) ≤ 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 2- Urbana (carne) ≤ 3- Urbana (sardinha).

$$* A\{13\} = [1\ 3\ 5]; B\{13\} = [1\ 2\ 4\ 5\ 8\ 9];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (S) lambens*, *P. (Euboetcheria) sp*, *P. (P) intermutans*: 1 – Urbana (fígado) ≤ 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 2- Urbana (carne) ≤ 9- Floresta (sardinha) ≤ 8 – Floresta (carne).

$$* A\{14\} = [1\ 3\ 5]; B\{14\} = [1\ 4\ 5\ 6\ 8\ 9];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (S) lambens*, *P. (Euboetcheria) sp*, *P. (P) intermutans*: 1 – Urbana (fígado) ≤ 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 6- Rural (sardinha) ≤ 9- Floresta (sardinha) ≤ 8 – Floresta (carne).

$$* A\{15\} = [3\ 4\ 5]; B\{15\} = [1\ 2\ 3\ 4\ 5\ 7];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (Euboetcheria) sp*, *P. (E) collusor* e *P. (P) intermutans*: 1 – Urbana (fígado) ≤ 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 3- Urbana (sardinha) ≤ 2- Urbana (carne) ≤ 7- Floresta (fígado).

$$* A\{16\} = [3\ 4\ 5]; B\{16\} = [1\ 2\ 3\ 4\ 5\ 8];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (Euboetcheria) sp*, *P. (E) collusor* e *P. (P) intermutans*: 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 3- Urbana (sardinha) ≤ 1 – Urbana (fígado) ≤ 2- Urbana (carne) ≤ 8- Floresta (carne).

$$* A\{17\} = [3\ 4\ 5]; B\{17\} = [1\ 2\ 3\ 4\ 5\ 9];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (Euboetcheria) sp*, *P. (E) collusor* e *P. (P) intermutans*: 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 3- Urbana (sardinha) ≤ 2- Urbana (carne) ≤ 1 - Urbana (fígado) ≤ 9- Floresta (sardinha).

$$* A\{18\} = [1\ 4\ 5]; B\{18\} = [1\ 2\ 4\ 5\ 6\ 8];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (S) lambens*, *P. (E) collusor* e *P. (P) intermutans*: 1 - Urbana (fígado) ≤ 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 6- Rural (sardinha) ≤ 2- Urbana (carne) ≤ 8- Floresta (carne).

$$* A\{19\} = [1\ 4\ 5]; B\{19\} = [1\ 2\ 4\ 5\ 6\ 9];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (S) lambens*, *P. (E) collusor* e *P. (P) intermutans*: 1 - Urbana (fígado) ≤ 9 - Floresta (sardinha) ≤ 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 6- Rural (sardinha) ≤ 2- Urbana (carne).

$$* A\{20\} = [1\ 4\ 5]; B\{20\} = [1\ 3\ 4\ 5\ 6];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (S) lambens*, *P. (E) collusor* e *P. (P) intermutans*: 1 - Urbana (fígado) ≤ 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 6- Rural (sardinha) ≤ 3- Urbana (sardinha).

$$* A\{21\} = [1\ 3\ 4]; B\{21\} = [1\ 4\ 5\ 6\ 7\ 8];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (S) lambens*, *P. (Euboetcheria) sp* e *P. (E) collusor* : 1 - Urbana (fígado) ≤ 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 6- Rural (sardinha) ≤ 7- Floresta (fígado) ≤ 8- Floresta (carne).

$$* A\{22\} = [1\ 3\ 4]; B\{22\} = [1\ 4\ 5\ 6\ 7\ 9];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (S) lambens*, *P. (Euboetcheria) sp* e *P. (E) collusor* : 1 - Urbana (fígado) ≤ 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 6- Rural (sardinha) ≤ 7- Floresta (fígado) ≤ 9- Floresta (sardinha).

$$* A\{23\} = [2\ 3\ 5]; B\{23\} = [1\ 2\ 3\ 4\ 5\ 6];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *O. thornax*, *P. (Euboetcheria) sp* e *P. (P) intermutans* : 4- Rural (fígado) ≤ 5-

Rural (carne) ≤ 2- Urbana (carne) ≤ 1 - Urbana (fígado) ≤ 3- Urbana (sardinha) ≤ 6- Rural (sardinha).

$$* A\{24\} = [2\ 3\ 4\ 5]; B\{24\} = [1\ 3\ 4\ 5];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *O. thornax*, *P (Euboetcheria) sp*, *P. (E) collusor* e *P. (P) intermutans* : 5- Rural (carne) ≤ 4- Rural (fígado) ≤ 1 - Urbana (fígado) ≤ 3- Urbana (sardinha).

$$* A\{25\} = [2\ 3\ 4\ 5]; B\{25\} = [1\ 4\ 5\ 6];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *O. thornax*, *P (Euboetcheria) sp*, *P. (E) collusor* e *P. (P) intermutans* : 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 1 - Urbana (fígado) ≤ 6- Rural (sardinha).

$$* A\{26\} = [2\ 3\ 4\ 5]; B\{26\} = [2\ 4\ 5\ 7];$$

que refere-se a seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *O. thornax*, *P (Euboetcheria) sp*, *P. (E) collusor* e *P. (P) intermutans* : 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 2 – Urbana (carne) ≤ 7- Floresta (fígado).

$$* A\{27\} = [2\ 3\ 4]; B\{27\} = [4\ 8\ 9];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *O. thornax*, *P (Euboetcheria) sp* e *P. (E) collusor* : 4- Rural (fígado) ≤ 8- Floresta (carne) ≤ 9 - Floresta (sardinha).

$$* A\{28\} = [1\ 2\ 3\ 5]; B\{28\} = [2\ 3\ 4\ 5];$$

que refere-se à seguinte ordenação referente ao número de indivíduos coletados nas colunas para as espécies *P. (S) lambens*, *O. thornax*, *P (Euboetcheria) sp* e *P. (P) intermutans* : 4- Rural (fígado) ≤ 5- Rural (carne) ≤ 2 – Urbana (carne) ≤ 3- Urbana (sardinha).

APÊNDICE I – Árvores de decisão das espécies de dípteros necrófagos obtidas durante o presente trabalho.

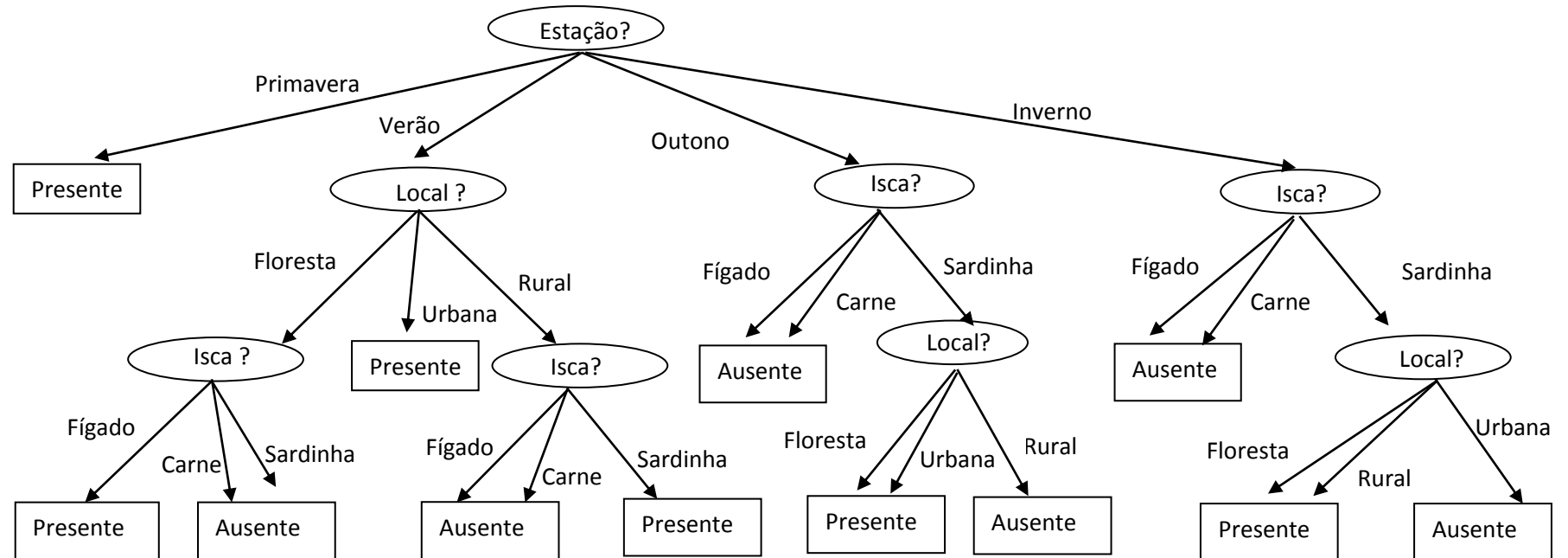


Figura 19 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Oxysarcodexia thornax*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.



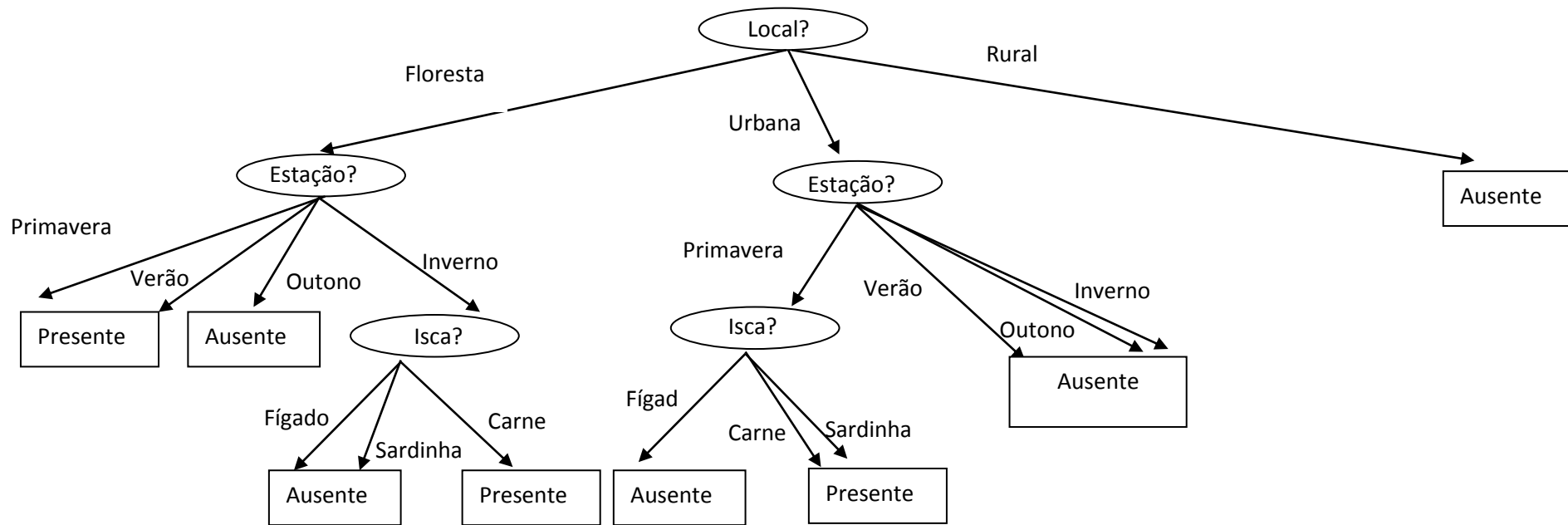


Figura 20 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Peckia (Euboettcheria) collusor*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.

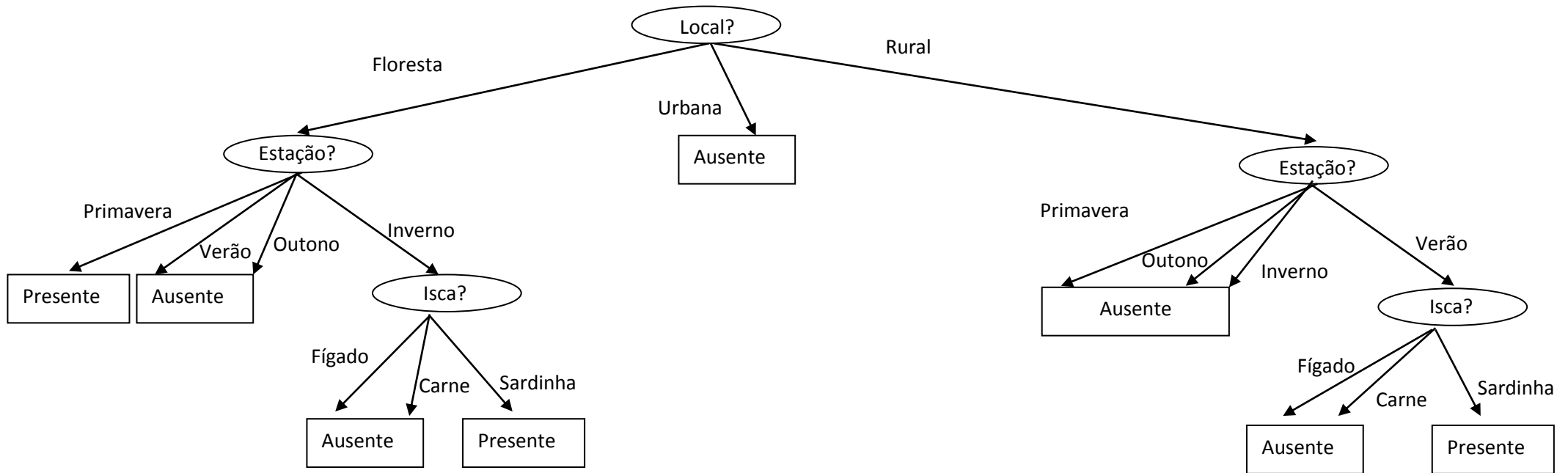


Figura 21 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Oxysarcodexia avuncula*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.

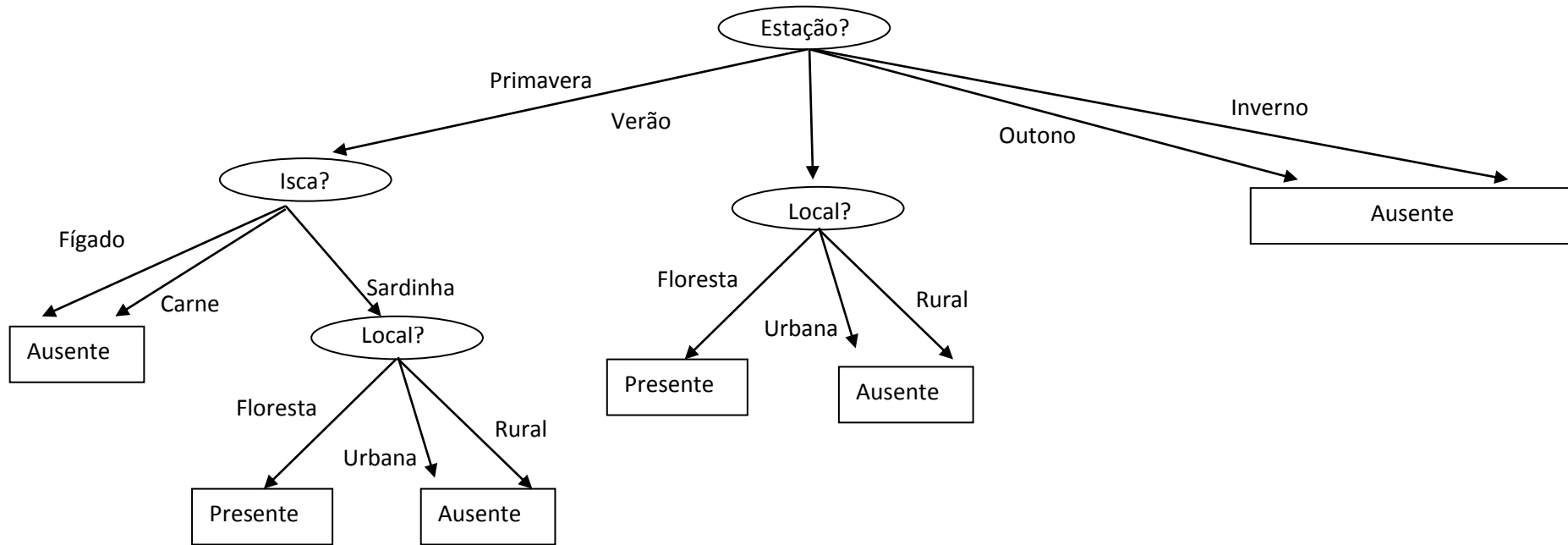


Figura 22 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Peckia (Squamatoles) ingens*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.

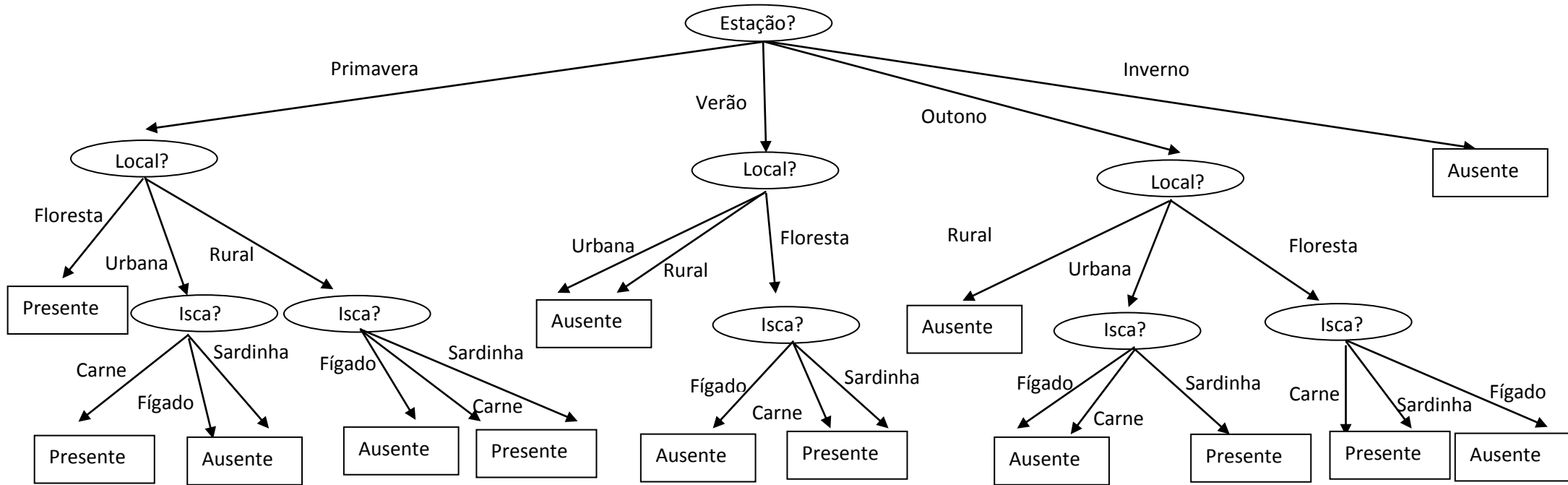


Figura 23 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Peckia (Euboettcheria) anguilla*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.

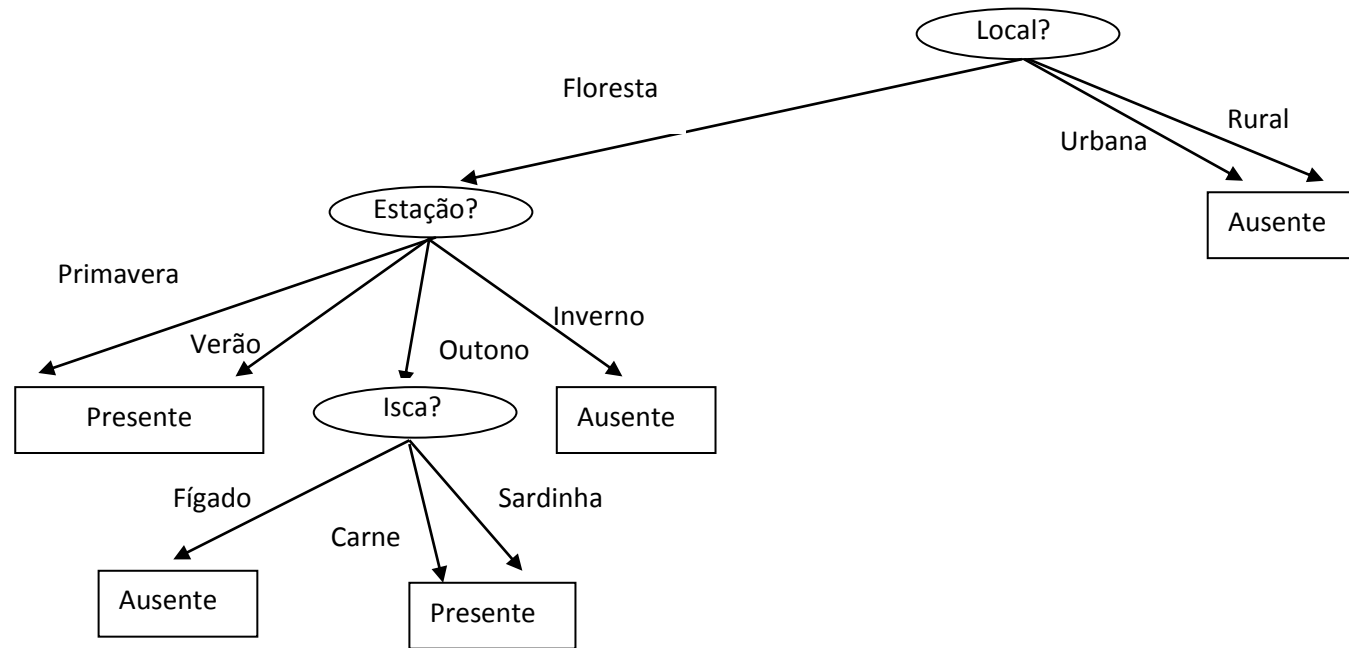


Figura 24 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Peckia (Patonella) intermutans*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.

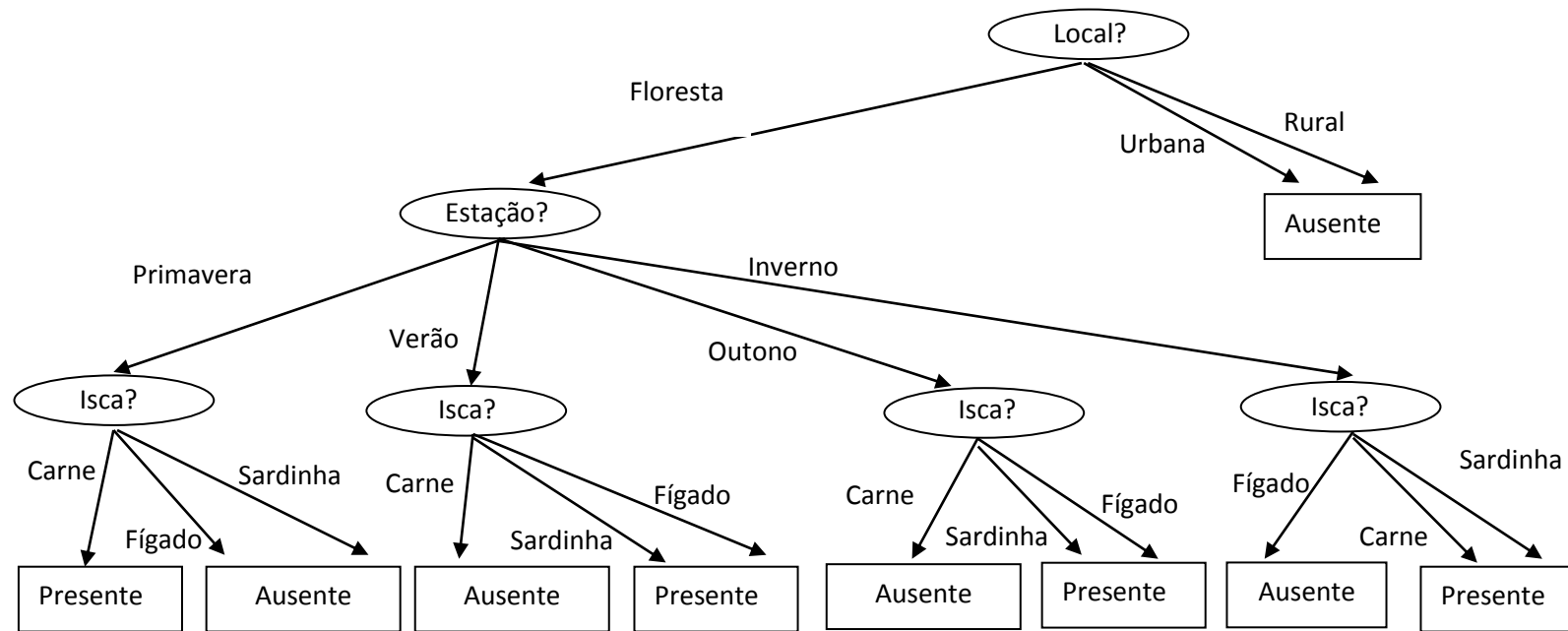


Figura 25 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Peckia (Euboettcheria) florencioi*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.

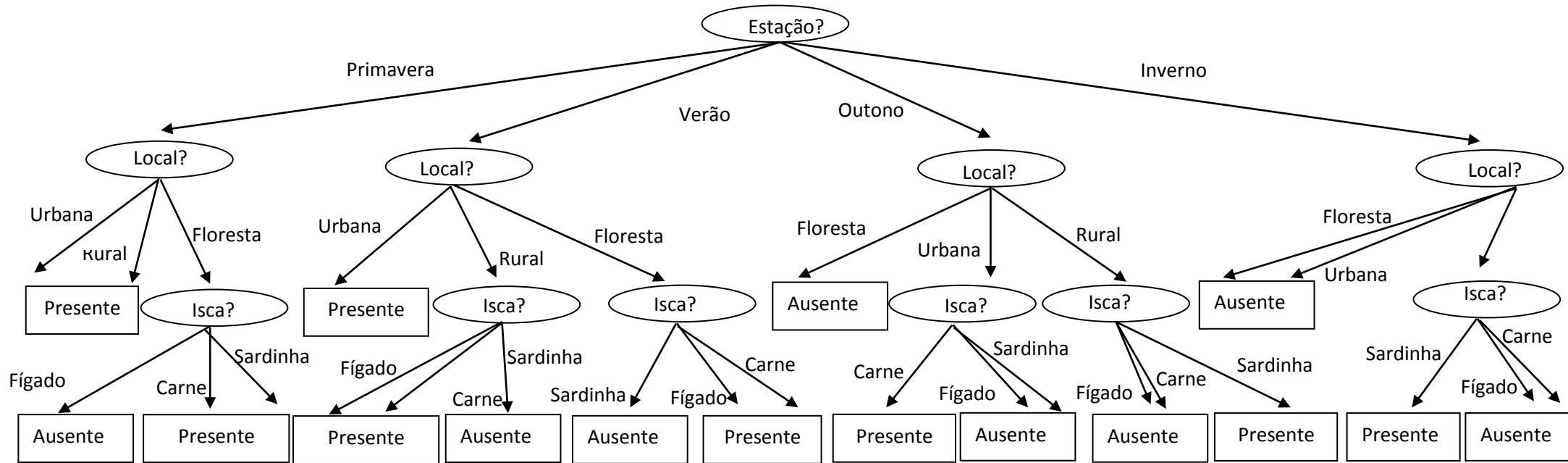


Figura 26 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Atherigona orientalis*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.

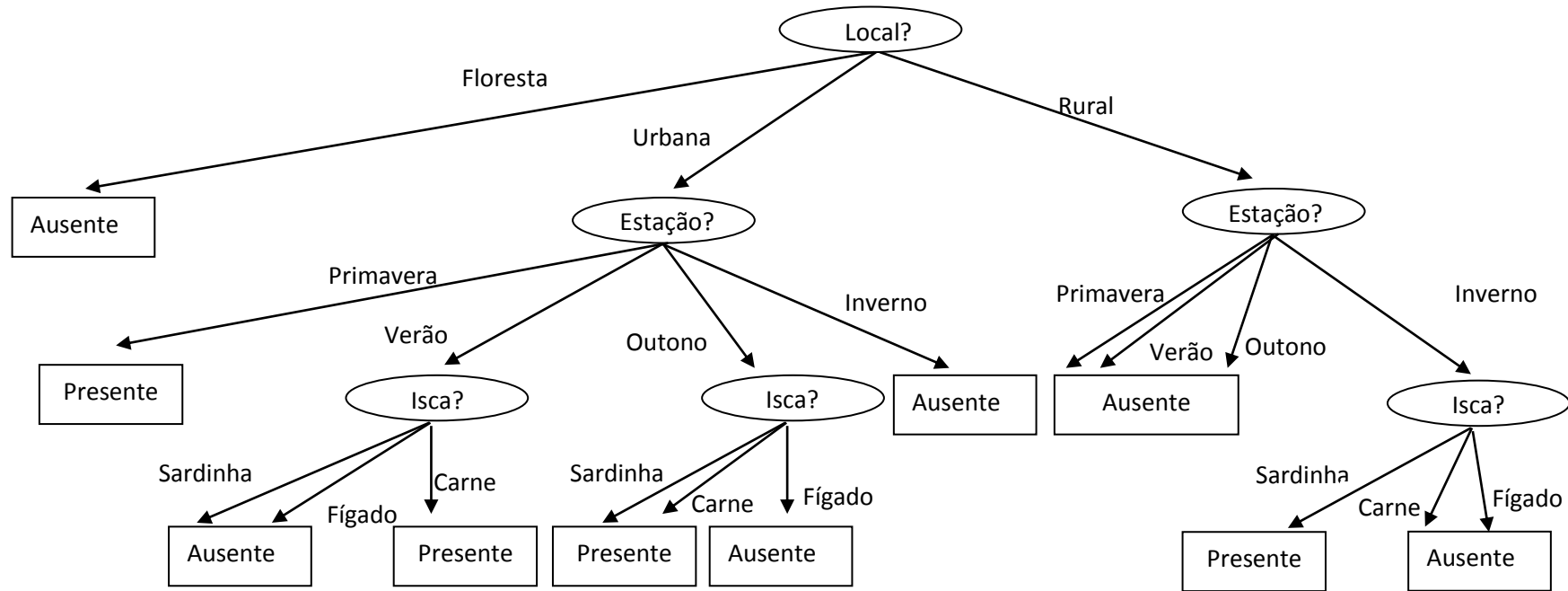


Figura 27 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Synthesiomyia nudiseta*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.



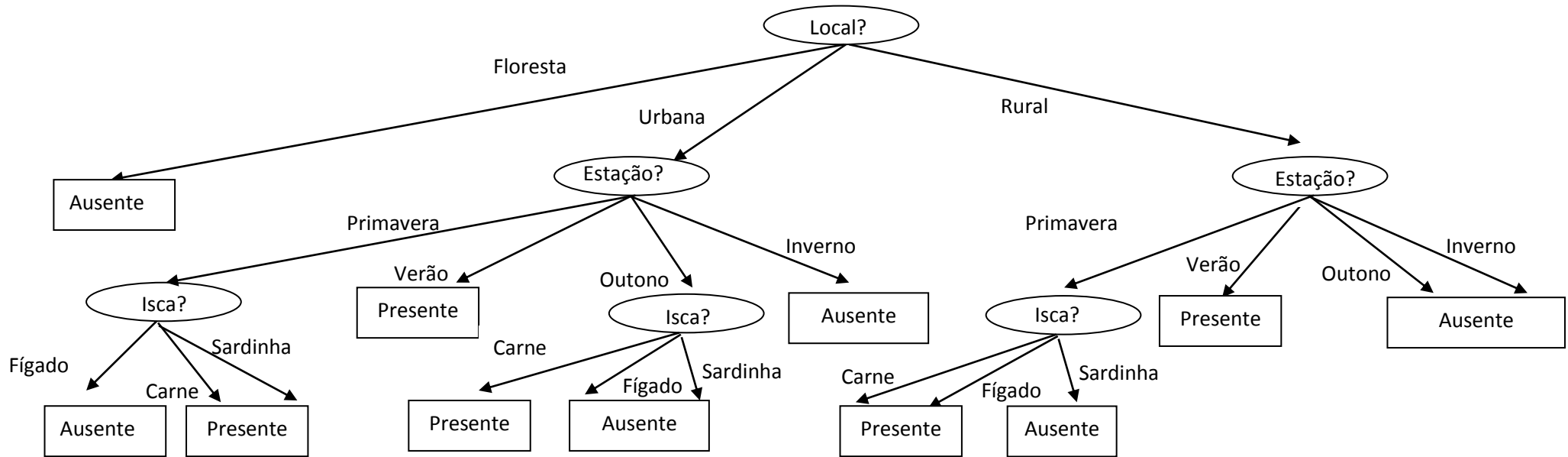


Figura 28 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Musca domestica*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.

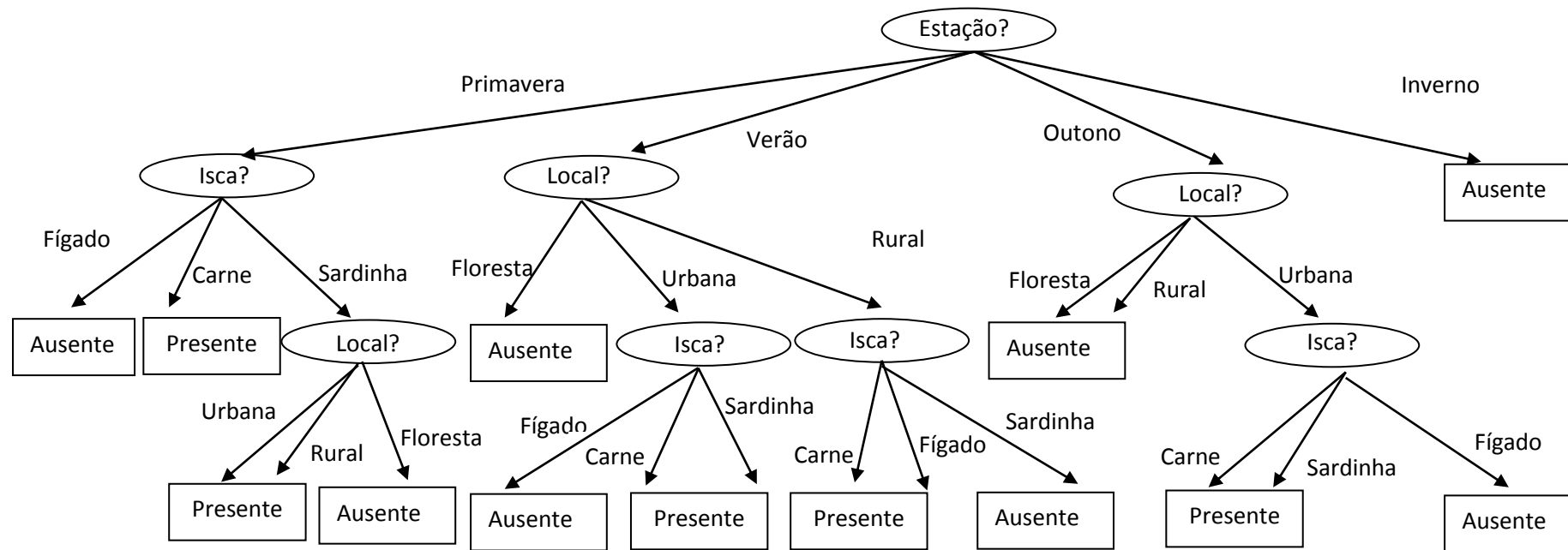


Figura 29 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Ophyra calchogasther*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.

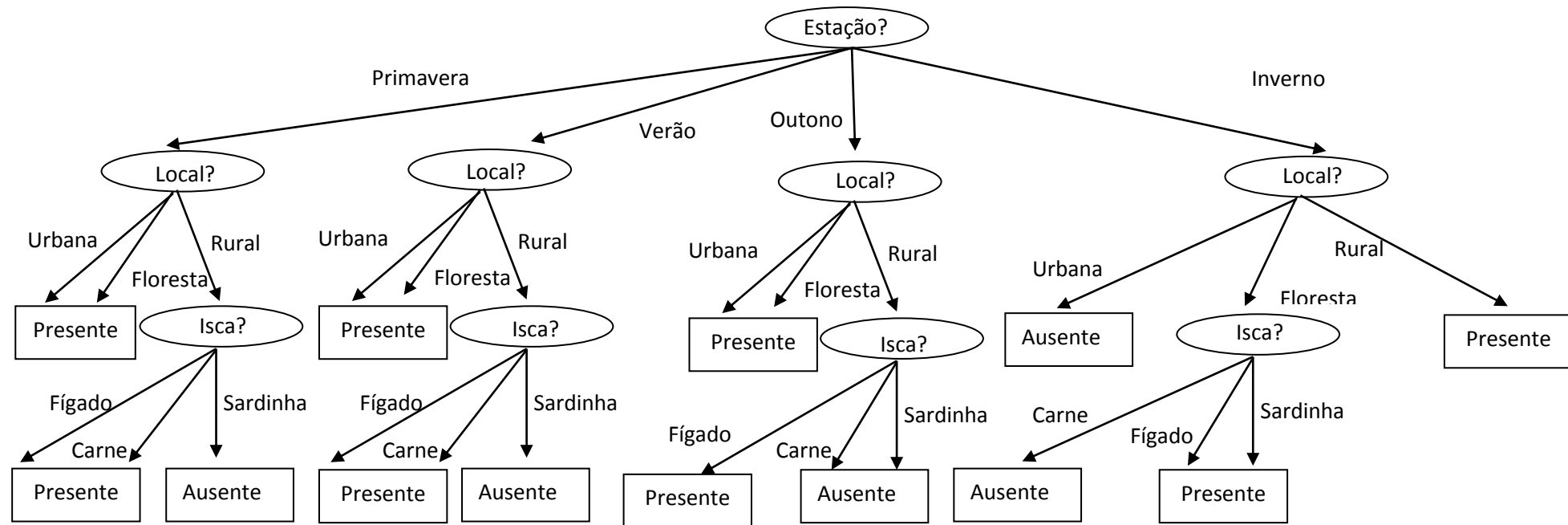


Figura 30 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Lucilia eximia*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.

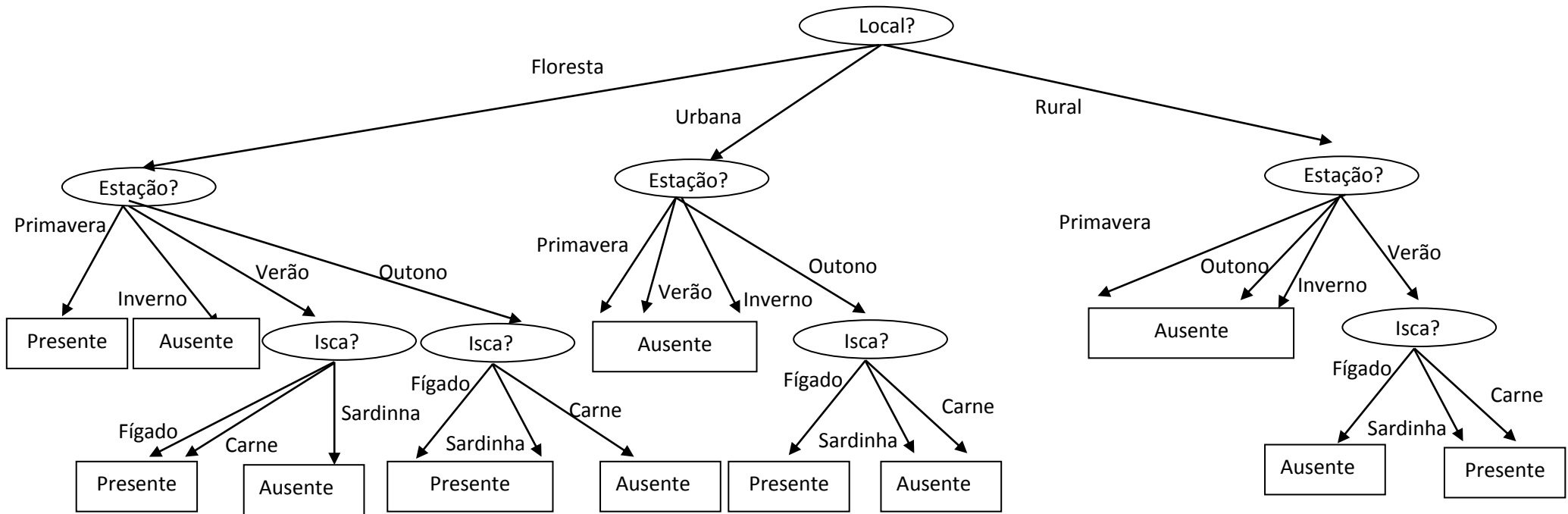


Figura 31 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Hemilucilia segmentaria*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.

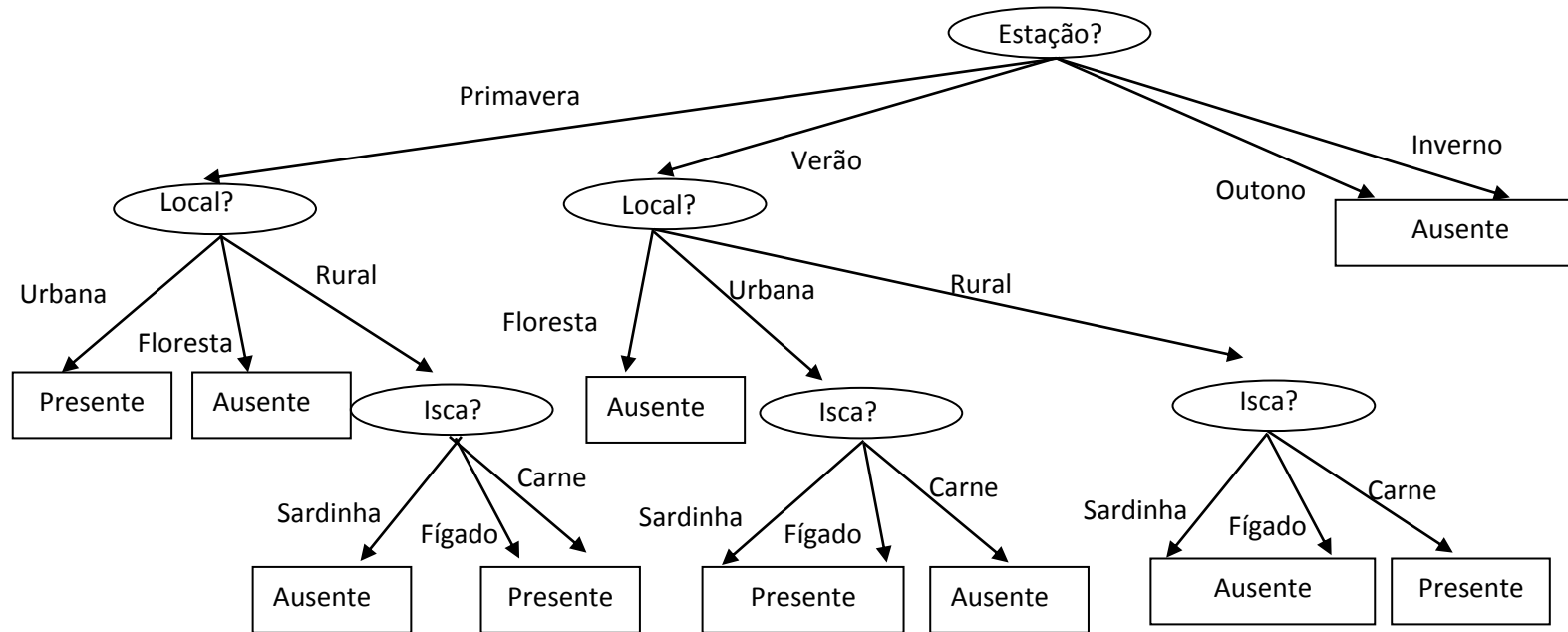


Figura 32 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Lucilia cuprina*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.

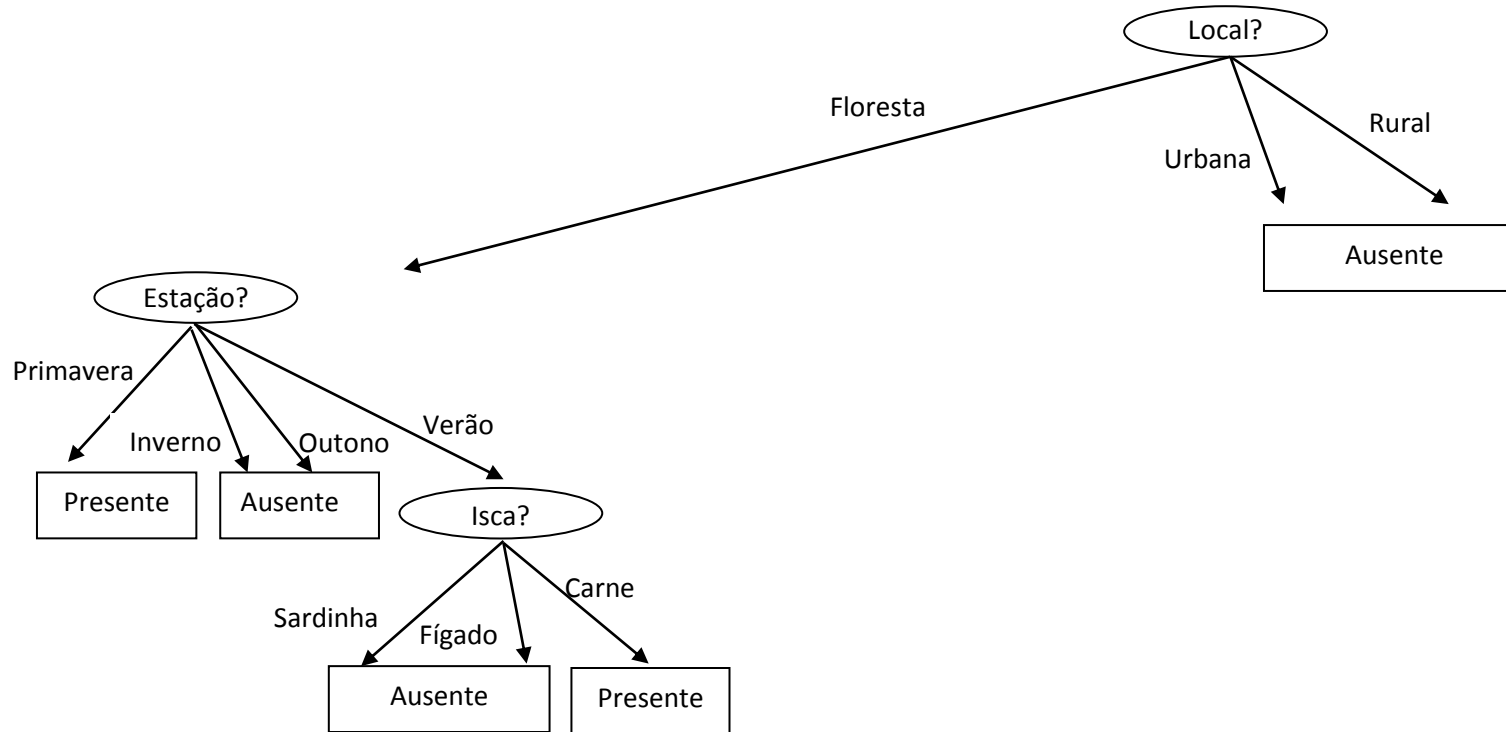


Figura 33 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Hemilucilia semidiaphana*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.

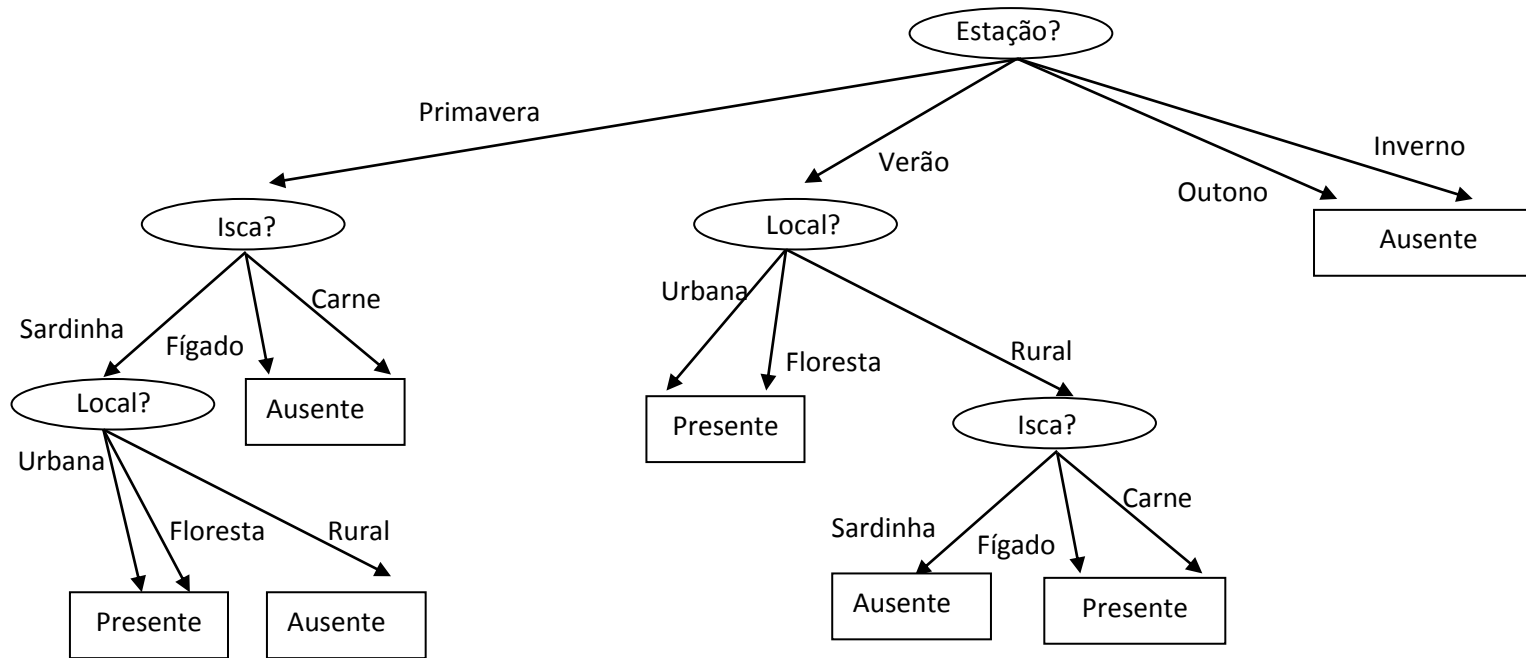


Figura 34 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Chrysomya megacephala*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.

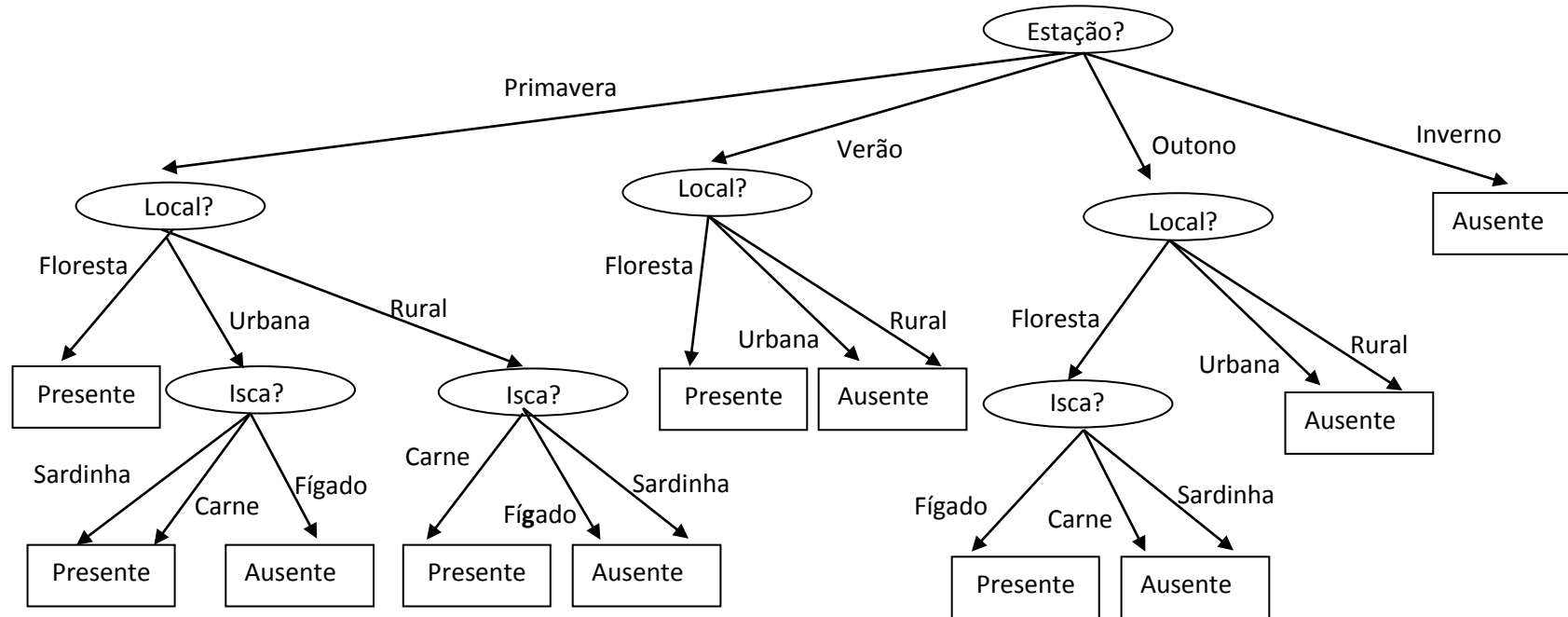


Figura 35 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Chrysomya albiceps*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.



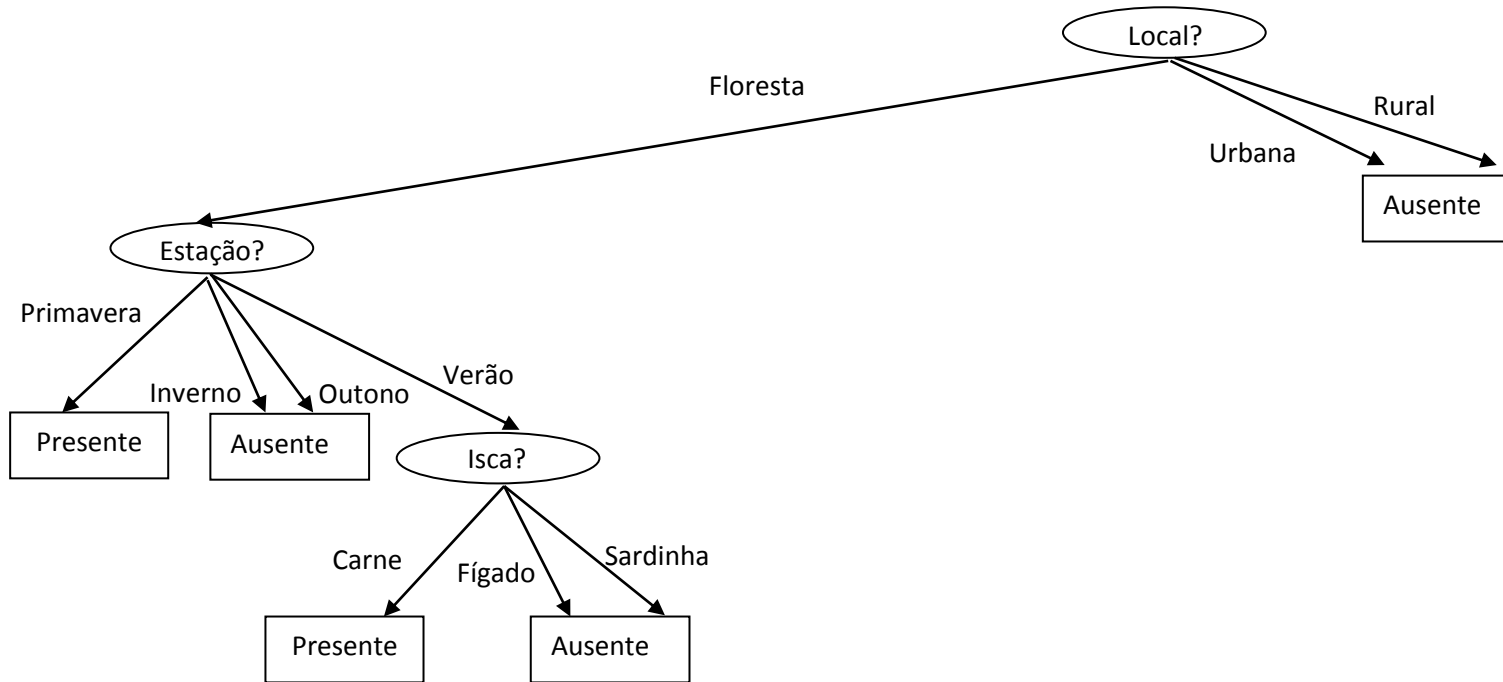


Figura 36 - Modelo da árvore de decisão classificativa em relação à presença ou ausência de *Mesembrinella peregrina*. Os objetos elipsoidais possuem uma variável de entrada (nó interno) sendo portanto um ponto de divisão/escolha. Em cada aresta está discriminada a decisão tomada. Os objetos quadrados são as classes. Uma previsão é conseguida por uma combinação de uma série de critérios definidos pelas variáveis e as decisões tomadas.

