

Elielma Santana de Jesus

**REGRAS DE DESCOBERTA DE ASSOCIAÇÃO NA
ELABORAÇÃO DE REDES BAYESIANAS APLICADAS AO SISTEMA
DE NOTIFICAÇÕES DE DOENÇAS VESICULARES EM BOVINOS NO
BRASIL**

**RECIFE-PE
ABRIL DE 2019**



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

REGRAS DE DESCOBERTA DE ASSOCIAÇÃO NA
ELABORAÇÃO DE REDES BAYESIANAS APLICADAS AO SISTEMA DE
NOTIFICAÇÕES DE DOENÇAS VESICULARES EM BOVINOS NO BRASIL

Dissertação apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como exigência parcial à obtenção do título de Mestre.

Área de Concentração: Biometria e Estatística Aplicada

Orientador: Prof. Dr. Kleber Régis Santoro

Recife-PE
Abril de 2019

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas da UFRPE
Biblioteca Central, Recife-PE, Brasil

C376a Jesus, Elielma Santana de.
Regras de descoberta de associação na elaboração de redes Bayesianas aplicadas ao sistema de notificações de doenças vesiculares em bovinos no Brasil / Elielma Santana de Jesus. – Recife, 2019.
55 f.: il.

Orientador(a): Kleber Régis Santoro.
Dissertação (Mestrado) – Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Biometria e Estatística Aplicada, Recife, BR-PE, 2019.
Inclui referências.

1. Febre aftosa 2. Vigilância animal 3. Probabilidade 4. Algoritmos de aprendizagem I. Santoro, Kleber Régis, orient. II. Título

CDD 310

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA**

**REGRAS DE DESCOBERTA DE ASSOCIAÇÃO NA
ELABORAÇÃO DE REDES BAYESIANAS APLICADAS AO SISTEMA DE
NOTIFICAÇÕES DE DOENÇAS VESICULARES EM BOVINOS NO BRASIL**

Elielma Santana de Jesus

Dissertação julgada adequada para obtenção do título de Mestre em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 25/02/2019 pela Banca Examinadora.

Orientador:

**Prof. Dr. Kleber Régis Santoro
Universidade Federal Rural de Pernambuco**

Banca Examinadora:

**Prof. Dr. Jones Oliveira de Albuquerque
Universidade Federal Rural de Pernambuco**

**Prof. Dr. Frank Sinatra Gomes da Silva
Universidade Federal Rural de Pernambuco**

Aos meus pais Maria e Antônio.

AGRADECIMENTOS

À Deus pelo dom da vida, por me encorajar para enfrentar as dificuldades.

A todos os familiares em especial meus pais, Maria e Antônio, que mesmo sem entender o porquê de continuar estudando e ainda por cima longe de casa, me apoiam e desejam o meu sucesso que também é deles.

Ao meu orientador, professor Kleber Régis Santoro, pelo conhecimento transmitido e acolhida em Garanhuns. Além da paciência e compreensão, principalmente na reta final.

Aos irmãos “lobinhos”, pelas contribuições e incentivo ao longo do curso. Aos colegas e amigos da Biometria pela troca de experiências, ajuda nos momentos difíceis e por tornarem meus dias mais alegres. Aos amigos de Sergipe que mesmo distantes sempre se fizeram presentes.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro. Ao Programa de Pós-Graduação em Biometria e Estatística Aplicada (PPGBEA) e também a Unidade Acadêmica de Garanhuns (UAG).

À Agência de Defesa e Fiscalização Agropecuária do Estado de Pernambuco (ADAGRO-PE) por ceder os dados.

Enfim, a todos que direta ou indiretamente contribuíram para que eu pudesse chegar até aqui.

*“A grande descoberta da minha geração é que
os seres humanos podem alterar suas vidas
ao alterar as suas atitudes mentais”*

William James

RESUMO

A febre aftosa se destaca entre as enfermidades vesiculares, recebendo atenção especial da vigilância em saúde animal, devido aos danos causados, tais como queda de produtividade, perda de mercado, implicações sanitárias aos animais e humanos. Utilizando-se de diferentes critérios de descoberta de regras de associação foram analisadas diferentes redes Bayesianas obtidas por algoritmos de aprendizagem estrutural para avaliação de fatores associados às notificações de doenças vesiculares em bovinos no Brasil. As redes propostas pelos algoritmos de pontuação Hill-Climbing e Tabu recuperaram as probabilidades marginais contidas na base de dados. Porém, apresentaram inconsistências nas relações de causalidade, mesmo utilizando-se de conhecimento prévio a respeito do sistema de vigilância. A aplicação da técnica de mineração de regras de associação mostrou-se viável na descoberta de relacionamentos entre as variáveis no conjunto de dados, permitindo a montagem das redes de acordo com os diferentes níveis de combinação.

Palavras-chaves: Febre aftosa; vigilância animal; probabilidade; algoritmos de aprendizagem

ABSTRACT

Foot-and-mouth disease stands out among vesicular diseases, receiving special attention from animal health surveillance, due to the damages caused, such as productivity decrease, market loss, health implications to animals and humans. Using different criteria for the discovery of association rules, different Bayesian networks obtained by structural learning algorithms were analyzed to assess factors associated with notifications of vesicular diseases in bovine in Brazil. The networks proposed by the Hill-Climbing and Tabu scoring algorithms recovered the marginal probabilities contained in the database. However, they presented inconsistencies in the causal relations, even using previous knowledge about the surveillance system. The application of the association rules mining technique proved to be feasible in the discovery of relationships among the variables in the dataset, allowing the assembly of the networks according to the different levels of combination.

Keywords: Foot-and-mouth disease; animal surveillance; probability; learning algorithms

LISTA DE FIGURAS

Figura 1	– Condição sanitária oficial da Febre Aftosa dos membros da OIE na América do Sul.....	5
Figura 2	– Organização dos estados brasileiros em blocos para o processo de transição da condição sanitária.....	6
Figura 3	– Cronograma para o processo de transição da condição sanitária.....	7
Figura 4	– Exemplo de rede Bayesiana.....	12
Figura 5	– Exemplo de <i>Markov blanket</i> da variável X.....	13
Figura 6	– Ilustração do procedimento Hill-Climbing.....	15
Figura 7	– Exemplo da desigualdade de processamento de dados.....	17
Figura 8	– Número de regras de associação conforme <i>support</i> , <i>confidence</i> e <i>lift</i>	23
Figura 9	– Comparação da topologia das redes Bayesianas de HC com as de Tabu, GS e Aracne.....	24
Figura 10	– Redes Bayesianas para 13 variáveis com algoritmos HC e Tabu.....	27
Figura 11	– Redes Bayesianas para 10 variáveis com algoritmos HC e Tabu.....	28
Figura 12	– Redes Bayesianas para 8 variáveis com algoritmos HC e Tabu.....	29
Figura 13	– Redes Bayesianas para as 13 variáveis com <i>whitelist</i> e <i>blacklist</i> , através dos algoritmos HC e Tabu.....	31

LISTA DE TABELAS

Tabela 1 – Comparação das frequências da base de dados e probabilidades marginais das RB com 8, 10 e 13 variáveis com os algoritmos de pontuação HC e Tabu.....	30
---	----

LISTA DE ABREVIATURAS E SIGLAS

Amos	Coleta de Amostra Biológica
BIC	Bayesian Information Criterion
Bov	Bovino
Bub	Bubalino
Cap	Caprino
CF	Confirmação Final
DPD	Desigualdade de Processamento de Dados
DSI	Doenças com Sintomas Iguais
DSPB	Doenças com Sintomas Parecidos Causadas por Bactéria
DSPM	Doenças com Sintomas Parecidos Causadas por Micose
DSPNI	Doenças com Sintomas Parecidos Não Infecciosas
DSPV	Doenças com Sintomas Parecidos Causadas por Vírus
Enf	Classe de Enfermidades
Equ	Equino
Est	Estados Brasileiros
F	Falso
FA	Febre Aftosa
GS	Grow-Shrink
HC	Hill-Climbing
IBGE	Instituto Brasileiro de Geografia e Estatística
IM	Informação Mútua
IP	Imunidade da População
IP. NA	Imunidade da População Não Avaliada
IZL	Implantação de Zonas Livres
KDD	Knowledge Discovery in Database
L.CV	Livre com Vacinação
L.SR	Livre sem Reconhecimento
L.SV	Livre sem Vacinação
LSO	Lesões
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
Not	Origem das notificações

Notn	Número de Notificações
OIE	Organização Mundial de Saúde Animal
Ovi	Ovino
PANAFTOSA	Centro Pan-Americano de Febre Aftosa
PCri	Período Crítico
Per	Período
PEst	Período Estável
PHEFA	Plano Hemisférico de Erradicação da Febre Aftosa
PMud	Período de Mudança
PNEFA	Programa Nacional de Erradicação de Febre Aftosa
Pop	População
PP	Portador com Proteção
Pro	Proprietário
RB	Redes Bayesianas
RCP	Recuperação com Proteção
SHD	Structural Hamming Distance
Sui	Suíno
SVO	Serviço Veterinário Oficial
Ter	Terceiros
TL	Timeliness
V	Verdadeiro
Vig	Vigilância
Z.Inf	Zona Infectada
Z.LRS	Zona Livre com Reconhecimento Suspenso
Z.Prot	Zona de Proteção

SUMÁRIO

1	INTRODUÇÃO.....	1
2	OBJETIVO.....	3
3	REVISÃO DE LITERATURA.....	4
3.1	Sistema de notificações de doenças vesiculares.....	4
3.2	Febre aftosa.....	8
3.3	Descoberta de regras de associação (<i>Rules mining</i>).....	9
3.4	Redes Bayesianas.....	11
3.4.1	Aprendizagem estrutural de redes Bayesianas.....	13
3.4.1.1	Hill-Climbing.....	14
3.4.1.2	Tabu Search.....	15
3.4.1.3	Grow-Shrink.....	16
3.4.1.4	Aracne.....	17
4	MATERIAL E MÉTODOS.....	18
4.1	Material.....	18
4.2	Métodos.....	20
5	RESULTADOS.....	23
6	DISCUSSÃO.....	32
7	CONCLUSÃO.....	34
	REFERÊNCIAS BIBLIOGRÁFICAS.....	35

1 INTRODUÇÃO

O sistema de vigilância epidemiológica em saúde animal representa o conjunto de atividades que tem por objetivo conhecer, detectar ou prever a ocorrência de enfermidades nesta população para que sejam tomadas as medidas necessárias (FOSSAERT; LLOPIS; TIGRE, 1974).

Dentre as doenças que recebe maior atenção por parte do Serviço Veterinário Oficial (SVO) está a febre aftosa (FA), causada por um vírus do gênero *Aftovirus*, da família *Picornaviridae*, existindo sete tipos diferentes (OIE, 2017). Enfermidade vesicular de notificação obrigatória e compulsória, caracteriza-se por ser altamente contagiosa e causar graves perdas econômicas em animais ungulados de casco dividido, em que são suscetíveis os bovinos, bubalinos, ovinos, caprinos, suínos, ruminantes silvestres e outros nos quais a infecção foi demonstrada cientificamente (MAPA, 2007a).

A estimativa quantitativa de risco e de fatores associados à introdução do vírus da FA em um país são pré-requisitos para a elaboração de diferentes políticas para a prevenção e eventual controle de epidemias por este vírus (MARTÍNEZ-LÓPEZ et al., 2008). Vários são os estudos desenvolvidos sobre a probabilidade de introdução do vírus da FA em países da União Europeia e Inglaterra (HALASA et al., 2013; HORST et al., 1999; MCLAWS et al., 2009; VOLKOVA et al., 2011).

A bovinocultura é um dos principais destaques do agronegócio brasileiro no cenário mundial, onde o Brasil possui o maior rebanho do mundo, com mais de 220 milhões de cabeças. A carne bovina representou 3% do total de exportações do agronegócio em 2017, sendo comercializada em mais de 180 países (ABIEC, 2018).

O primeiro caso de FA registrado no Brasil foi em 1895 em Minas Gerais e o último em 2005-2006 nos estados do Rio Grande do Sul e Santa Catarina (MAPA, 2018b). Em 1992, criou-se o Programa Nacional de Erradicação de Febre Aftosa (PNEFA) que a partir de um conjunto de medidas permitiu o avanço da condição sanitária em todo o país. Atualmente o Brasil está certificado pela Organização Mundial da Saúde Animal (OIE) como livre da FA, o que é considerado um marco na história da pecuária (MAPA, 2018a).

A atuação dos SVO se dá a partir da notificação de suspeita de doença vesicular, seja esta realizada pelo proprietário dos animais, por qualquer outra pessoa ou pelo veterinário oficial, devendo ser feita em até 24 horas e o atendimento pelo SVO em até 12 horas (MAPA, 2009).

As notificações são registradas desde 2004 na plataforma web do Sistema Continental de Informação e Vigilância Epidemiológica – SivCont, em que consta desde as informações básicas como hora, data e quem as realizou; localização da propriedade; a espécie desses animais com suspeitas, até o diagnóstico final que pode ser a confirmação ou descarte de enfermidade vesicular, ou ainda o diagnóstico diferencial (MAPA, 2009).

A combinação entre os diferentes níveis classificatórias das variáveis presentes no conjunto de dados de notificações leva a um grande número de combinações possíveis, que devem ser exploradas através da técnica mineração de regras de associação (*association rules mining*). Esta técnica sugere comportamento associativo através de limites mínimos estabelecidos previamente, explora a base de dados com todas as variáveis e utiliza cálculos de probabilidade simples, conjunta e condicional (AGRAWAL, SRIKANT, 1994; AGRAWAL, IMIELINSKI, SWAMI, 1993).

As redes Bayesianas (KOSKI, NOBLE, 2009; NAGARAJAN, SCUTARI, LÈBRE, 2013; SCUTARI, DENIS, 2014) surgem como opção para estudar as associações sugeridas. Em que propõem o cálculo probabilístico das variáveis aleatórias através da união de teoria de grafos e teoria da probabilidade, permitindo explorar os mais diferentes cenários e tipos de associação entre variáveis. Atuam principalmente no campo da incerteza, com fácil interpretação dos relacionamentos entre as variáveis através da visualização dos grafos e ainda permite a interação entre os conhecimentos de especialistas e probabilísticos.

Assim, a partir de diferentes critérios de descoberta de regras de associação, objetiva-se desenvolver redes Bayesianas para avaliação de fatores associados às notificações de doenças vesiculares em bovinos no Brasil.

2 OBJETIVO

Busca-se formular redes Bayesianas para avaliação de fatores associados às notificações de doenças vesiculares em bovinos no Brasil, usando como base diferentes critérios de descoberta de regras de associação.

Para isso, os objetivos específicos são:

- Avaliar critérios de descoberta de regras de associação e suas viabilidades;
- Avaliar a quantidade de regras e variáveis associadas com diferentes critérios de descoberta de regras de associação;
- Avaliar diferentes redes Bayesianas formadas a partir de diferentes valores de critérios de descoberta de regras de associação.

3 REVISÃO DE LITERATURA

3.1 Sistema de notificações de doenças vesiculares

Com o objetivo de desenvolver e fortalecer programas nacionais de prevenção e erradicação da FA, em 1951 foi criado pela Organização Pan-Americana da Saúde / Organização Mundial da Saúde (OPAS/OMS) o Centro Pan-Americano de Febre Aftosa (PANAFTOSA) com sede no Brasil (PANAFTOSA, 2018).

Posteriormente em 1987 os países da América do Sul assinaram o Plano Hemisférico de Erradicação da Febre Aftosa (PHEFA) que tem como objetivos: manter e expandir zonas livres de FA; aumentar a disponibilidade de carne e leite na região; e ampliar o poder de negociação de produtos agropecuários originados de países afetados no mercado internacional (MELO; SARAIVA; ASTUDILLO, 2002).

No Brasil, em 1992 foi implantado o PNEFA em acordo com PHEFA, promovendo a participação social, regionalização no combate à doença, vacinação sistemática de bovinos e bubalinos, entre outras medidas (MAPA, 2018a). Sendo de grande importância a existência de tal programa para a condição sanitária do país diante da OIE.

Mediante esforços despendidos pelo governo federal, serviços veterinários estaduais e setor privado, além da participação dos próprios criadores para erradicação da FA no Brasil, em 1998 aconteceu o primeiro reconhecimento de livre da FA com vacinação pela OIE, envolvendo os estados do Rio Grande do Sul e Santa Catarina. Porém, em 2000 houve a reintrodução da enfermidade no Rio Grande do Sul e conseqüentemente a suspensão da condição sanitária de livre com vacinação para os dois estados (MAPA, 2018b). No mesmo ano, os estados do Paraná e Distrito Federal e parte dos estados de Goiás, Mato Grosso, Minas Gerais e São Paulo foram reconhecidos como zona livre de FA com vacinação (MAPA, 2018b).

A OIE (2011) determina que somente após dois anos da ocorrência do último surto uma zona pode ser reconhecida como livre. Assim, em 2002 tanto o Rio Grande do Sul como Santa Catarina voltam a zona livre com vacinação. No ano anterior, houve ampliação da zona livre com vacinação com os estados da Bahia, Espírito Santo, Mato Grosso do Sul, Rio de Janeiro, Sergipe, Tocantins e parte de Goiás, Mato Grosso, Minas Gerais e São Paulo. Em 2005, com a reintrodução do vírus no Rio Grande do Sul e no Paraná, muitos estados perderam a condição sanitária já conquistada como livre da doença com vacinação (MAPA, 2018b).

Em 2007, Santa Catarina recebeu a condição de zona livre sem vacinação e desde então a mantém. Já em 2014, quase todo o país era considerado livre com vacinação somente os estados do Amapá, Amazonas e Roraima não eram reconhecidos (MAPA, 2018b).

Os esforços técnicos e financeiros do Ministério da Agricultura, Pecuária e Abastecimento (MAPA) em conjunto com a iniciativa privada continuavam e em 2018 o Brasil recebeu o reconhecimento internacional de livre de FA (26 estados livres com vacinação e um estado livre sem vacinação).

Em 2018 todos os países da América do Sul são considerados livres de FA, com exceção da Colômbia que está com a condição sanitária de livre com vacinação suspensa e a Venezuela que não tem reconhecimento pela OIE (Figura 1).

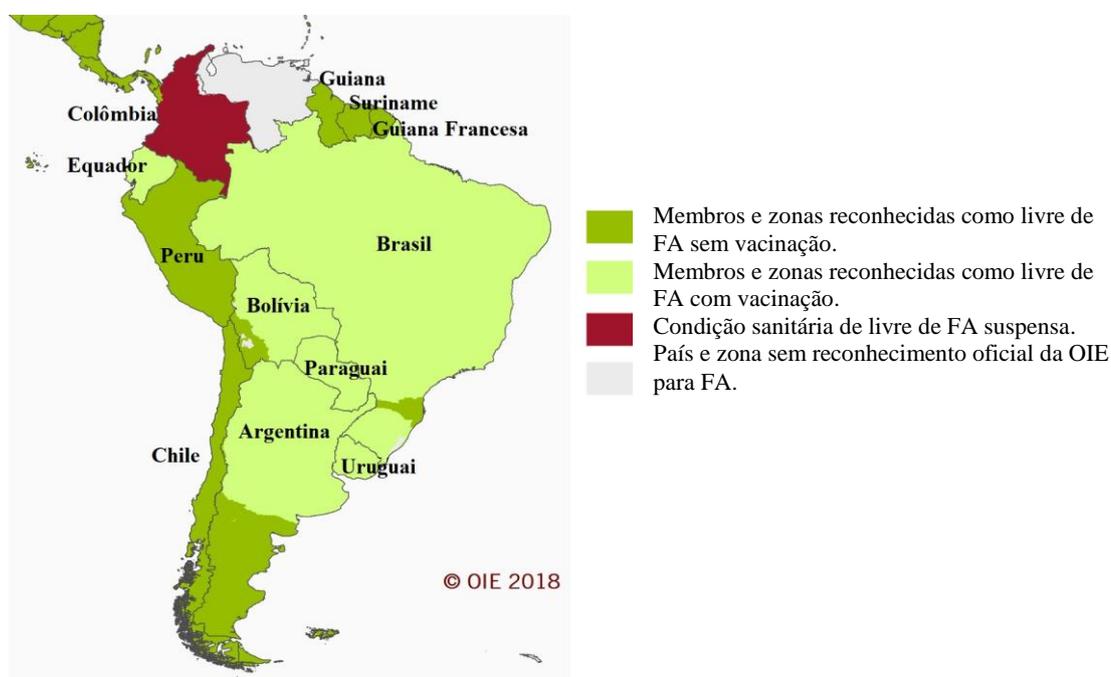


Figura 1: Condição sanitária oficial da Febre Aftosa dos membros da OIE na América do Sul.
Fonte: OIE (2018)

No Brasil os desafios passam a ser outros, agora busca-se alcançar a condição sanitária de livre sem vacinação para todo o país. Em 2017, o MAPA publicou o Plano Estratégico do PNEFA 2017 – 2026 que deve ser executado em 10 anos, tendo como principal objetivo criar e manter as condições sustentáveis para garantir a condição de país livre da FA e ampliar as zonas livres sem vacinação, protegendo o patrimônio pecuário e gerando o máximo de benefícios aos atores envolvidos e à sociedade brasileira (MAPA, 2018a).

Para realização do Plano Estratégico estão previstas operações distribuídas nas seguintes categorias: interação com as partes interessadas no Programa de prevenção da FA; ampliação das capacidades dos Serviços Veterinários; fortalecimento do sistema de vigilância em saúde animal, e realização da transição de zona livre de FA com vacinação para livre sem vacinação no país (MAPA, 2017). A suspensão da vacina será realizada gradativamente, os estados foram organizados em blocos (Figura 2) e irá seguir o cronograma estabelecido no plano (Figura 3).

Organização Geográfica para Zonificação

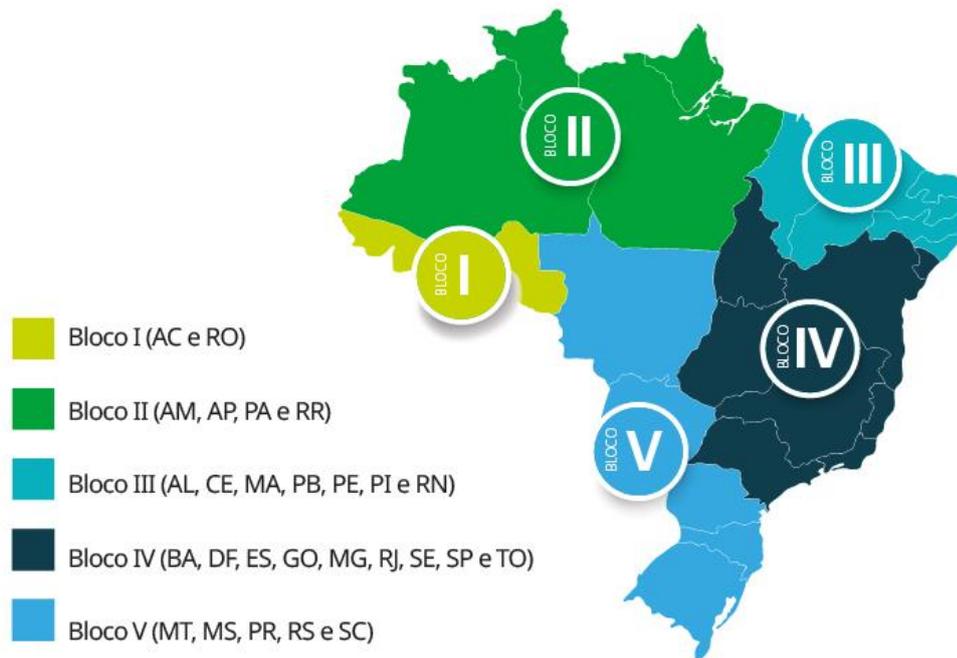


Figura 2: Organização dos estados brasileiros em blocos para o processo de transição da condição sanitária. Fonte: MAPA (2018a)

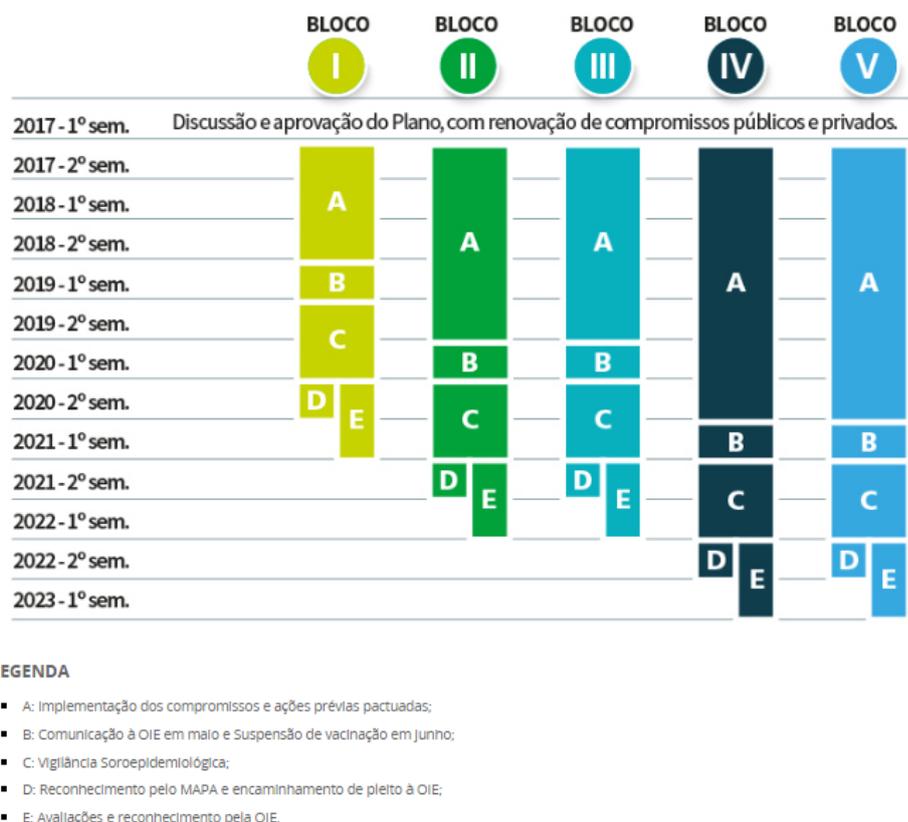


Figura 3: Cronograma para o processo de transição da condição sanitária. Fonte: MAPA (2018a)

Para que os programas de controle, erradicação e prevenção de enfermidades atuem de maneira eficiente é necessário que se tenha informações de qualidade para auxiliar nas medidas a serem tomadas. Ressalta-se que os países membros da OIE tem por obrigação notificar a ocorrência de doenças através dos sistemas atualmente em funcionamento (PANAFTOSA, 2011).

O SivCont é a plataforma web do Sistema Continental de Informação e Vigilância Epidemiológica em que são realizadas em tempo real as comunicações a PANAFTOSA de eventos que exijam atenção do serviço veterinário. Sendo responsabilidade da unidade central de serviço veterinário estadual alimentar a plataforma a partir dos formulários enviados pelas unidades veterinárias locais (MAPA, 2009).

O sistema de vigilância tem sua competência avaliada a partir do número de suspeitas notificadas, além da rapidez e o correto registro das informações em cada atendimento realizado (MAPA, 2009).

As enfermidades vesiculares são de notificação obrigatória e compulsória, devem ser informadas ao serviço veterinário em até 24 horas a suspeita de presença da doença, seja através

do proprietário, terceiros ou até mesmo da própria vigilância (MAPA, 2007a). E, os SVO tem até 12 horas para iniciar o atendimento, visita à propriedade suspeita, seguindo os procedimentos definidos nos manuais e planos de contingência (MAPA, 2007b).

O registro inicial da notificação é realizado no livro de ocorrências com as informações básicas como data, hora, quem realizou a notificação, localização da propriedade. E, durante o atendimento os médicos veterinários devem preencher o Formulário de Investigação de Doenças – Inicial (FORM-IN) e Formulário de Investigação – Complementar (FORM-COM) (MAPA, 2007b). O encerramento da investigação deve ser realizado no FORM-IN quando não houver suspeita de enfermidade vesicular ou outras infecciosas. No caso de confirmação da suspeita de caso provável de enfermidade vesicular, os dois formulários devem ser enviados para a unidade central de serviço veterinário estadual, e deve ser comunicado aos demais níveis do serviço veterinário oficial (MAPA, 2009). O envio de amostras para exame laboratorial também deve ser acompanhado pelos respectivos formulários.

Se confirmada a ocorrência de FA, devem ser realizadas ações e medidas de emergência sanitária com comunicação imediata ao MAPA, OIE, PANAFTOSA, demais estados brasileiros, blocos comerciais e países que mantêm acordos bilaterais de comércio (MAPA, 2007b). Com o descarte de ocorrência de FA, deve-se realizar o diagnóstico diferencial e adotar as devidas ações sanitárias. Independente do diagnóstico o atendimento deve ser registrado no SivCont (MAPA, 2007a).

3.2 Febre aftosa

A FA está entre as doenças definidas pela OIE como sendo de notificação obrigatória aos SVO. É a enfermidade mais contagiosa dos mamíferos e possui um grande potencial para causar graves perdas econômicas nos animais ungulados de casco dividido (LAWSON; ONI-CESCU; ELLERBE, 2011). Das espécies domésticas, o gado bovino, suíno, ovino, caprino e o búfalo aquático (*Bubalus bubalis*) são suscetíveis. Podem infectar-se também muitas espécies selvagens de casco dividido (OIE, 2017).

O vírus causador da FA é do gênero *Aphthovirus*, da família *Picornaviridae*, existindo sete tipos diferentes: A, O, C, SAT1, SAT2, SAT3 e Asia1; a infecção por um dos vírus não concede imunidade contra qualquer outro (OIE, 2017). Somente os três primeiros estiveram presentes no Brasil.

Tal enfermidade se caracteriza pelo aparecimento de febre e de vesículas nos pés, dentro e ao redor da cavidade oral e nas glândulas mamárias das fêmeas, podendo ocorrer também no

interior das narinas e nos pontos de pressão dos membros (OIE, 2017). A mortalidade pode ocorrer com maior frequência em animais jovens devido à miocardite multifocal, já em animais adultos essa taxa é baixa (ALEXANDERSEN et al., 2003).

A disseminação viral pode ocorrer a partir das secreções e excreções de animais com inflamações intensas, incluindo o ar expirado. A transmissão acontece pelo contato direto entre animais infectados ou objetos infectados com animais suscetíveis, e ainda por consumo de produtos originados de carne ou leite contaminados (OIE, 2017).

Por apresentar sintomas iguais ou parecidos, a FA pode facilmente ser confundida com outras doenças vesiculares como estomatite vesicular, diarreia viral bovina, peste bovina, rinotraqueíte infecciosa bovina, língua azul, mamilite bovina, varíola bovina, ectima contagioso; além de doenças não infecciosas como intoxicações e traumatismos físicos ou químicos (MAPA, 2009). O MAPA deixa claro que o descarte de FA deve estar muito bem fundamentado tecnicamente, e em caso de dúvidas deve-se continuar com a investigação dada a possibilidade de doença vesicular. O principal instrumento de confirmação ou descarte é através de diagnóstico clínico e epidemiológico.

3.3 Descoberta de regras de associação (*Rules mining*)

Diante do enorme conteúdo das bases de dados que se tem atualmente, são necessários avanços tecnológicos, para extração de informações úteis para a tomada de decisões, sendo uma delas o processo de descoberta de conhecimento em bancos de dados (*Knowledge Discovery in Database - KDD*). A principal etapa deste processo é a mineração de dados (*data mining*) que visa encontrar e extrair padrões de interesse utilizando métodos específicos (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

As regras de associação são um tipo de informação obtida com a mineração e representam relacionamentos ou padrões frequentes entre itens de uma base de dados (GONÇALVES, 2005). Uma aplicação comum é a análise de cesta de mercado (*market basket analysis*) que tem por objetivo descobrir quais produtos costumam ser comprados juntos pelos clientes, e assim contribuir para as estratégias de marketing das empresas. Por exemplo, uma regra de associação gerada a partir de uma base de dados que armazena mercadorias adquiridas por consumidores de um supermercado pode ser a seguinte: {feijão, couve} \Rightarrow {linguiça}, que significa que os consumidores que compram feijão e couve, tendem a comprar também linguiça.

A problemática que envolve a mineração de regras de associação foi introduzida por Agrawal, Imielinski e Swami (1993) como segue. Considere a base de dados D formada por

um conjunto de itens $I = \{i_1, i_2, \dots, i_n\}$ e um conjunto de transações $T = \{t_1, t_2, \dots, t_m\}$, sendo que cada transação contém itens pertencentes a I .

Uma regra de associação é definida da forma $X \Rightarrow Y$, em que $X, Y \subset I$ e $X \cap Y = \emptyset$. O lado esquerdo da regra (LHS) representado pelo conjunto de itens X é chamado de antecedente, enquanto que o lado direito da regra (RHS) representado pelo conjunto de itens Y é chamado de conseqüente (AGRAWAL; IMIELINSKI; SWAMI, 1993).

O algoritmo Apriori (AGRAWAL; SRINKANT, 1994) é amplamente empregado tanto para descobrir regras de associação como para encontrar conjuntos de itens frequentes. Este algoritmo realiza a busca em duas fases, na primeira obtêm o conjunto de itens frequentes que satisfaz o *support* mínimo (*minsup*) e em seguida determina as regras que atendem a *confidence* mínima (*minconf*) dentro do conjunto encontrado na primeira fase (ERSEL; GÜNAY, 2012).

Support e *confidence* são medidas objetivas utilizadas para encontrar regras interessantes na base de dados. O *support* indica a frequência em que a regra $X \Rightarrow Y$ ocorre no conjunto T , assim $supp(X \Rightarrow Y) = P(X \cap Y)$. A *confidence* indica a frequência em que os itens em Y ocorrem nas transações que também contêm os itens em X , $conf(X \Rightarrow Y) = P(X \cap Y)/P(X)$ (HAHSLER; BUCHTA; HORNIK, 2008).

Como medida adicional *lift* (SILVERSTEIN; BRIN; MOTWANI, 1998) quantifica o relacionamento entre X e Y , definida como

$$lift(X \Rightarrow Y) = \frac{P(X \cap Y)}{P(X) \times P(Y)}. \quad (1)$$

Se $lift(X \Rightarrow Y) = 1$, X e Y são independentes. Se $lift(X \Rightarrow Y) > 1$, X e Y são positivamente dependentes e se $lift(X \Rightarrow Y) < 1$, então X e Y são negativamente dependentes. Esta medida varia entre 0 e ∞ e é interpretada da seguinte forma: quanto maior o valor de *lift*, mais interessante a regra, visto que X aumentou Y numa taxa maior.

As vantagens por trás do uso da mineração de regras de associação é a descoberta de informações implícitas e a que níveis de *support* e *confidence* os itens de interesse estão ocorrendo junto a outros. Porém, não é possível realizar inferências a respeito dos dados. Além de que esta técnica só é aplicada após o pré-processamento dos dados, ou seja, os dados devem ser consistentes e permitir a transformação em transações para extração de conhecimentos úteis.

3.4 Redes Bayesianas

As redes Bayesianas (RB) (*Bayesian networks*) ou redes de crenças são representações gráficas de uma distribuição de probabilidade sobre um conjunto de variáveis (MARGARITIS, 2003).

De acordo com Neapolitan (2004), as RB surgiram com a necessidade de analisar a influência probabilística não direta de uma variável sobre as demais, dentro de um conjunto com muitas variáveis. Desde então, são aplicadas para solucionar problemas nas diversas áreas que possuem algum grau de incerteza, como diagnóstico médico (ZAGORECKI; ORZECHOWSKI; HOTOWNIA, 2013), previsão de falência (AGHAIE; SAEEDI, 2009), epidemiologia (RODIN; BOERWINKLE, 2005).

Combinam teoria dos grafos e teoria das probabilidades, de modo que cada variável aleatória é representada por um nó (vértice) e a dependência probabilística direta entre elas por arestas (arcos); quando existe um arco de X_i para X_j significa há dependência estatística entre essas variáveis, de modo que um valor tomado por X_j depende do valor tomado por X_i (BENGAL, 2007). Pode-se dizer que X_i é pai de X_j e X_j é filho de X_i , são termos utilizados para se referir aos nós que podem ser alcançados por um caminho direto.

A teoria dos grafos é considerada um ramo da Topologia, mas fortemente ligada à Álgebra e à Teoria de Matrizes, o matemático suíço Leonhard Euler (1707-1783) iniciou os estudos de grafos com o mistério conhecido atualmente como “O Problema das Pontes de Königsberg” e em 1736 escreveu seu primeiro artigo a respeito desta teoria (COSTA, 2011). Sendo atualmente empregada em vários campos, um grafo $G = (V, A)$ consiste em um conjunto não-vazio V de nós ou vértices e um conjunto finito (mas possivelmente vazio) A de pares de vértices chamados arcos, links ou arestas.

A construção de RB ocorre em duas etapas, a identificação do modelo grafo e em seguida a estimação dos parâmetros. A primeira, chamada de aprendizagem estrutural é realizada por meio de algoritmos disponíveis na literatura que consiste em descobrir as relações de dependência entre as variáveis da rede Bayesiana, ou ainda, pode-se construir a topologia da rede manualmente, por meio do conhecimento humano. A segunda etapa é o aprendizado dos parâmetros da distribuição local, ou seja, as probabilidades condicionais e/ou conjuntas de cada nó de acordo com a estrutura (NAGARAJAN; SCUTARI; LÈBRE, 2013).

Ressalta-se que a topologia (estrutura) da rede é representada por grafo acíclico direto, ou seja, um nó não pode ser seu próprio antepassado ou seu próprio descendente; o que permite a fatoração da probabilidade conjunta de uma coleção de nós (XIMENES, 2018).

Na Figura 4 observa-se a estrutura de uma rede Bayesiana, além das distribuições condicionais de probabilidades. Neste exemplo, a variável *Cloudy* é pai das variáveis *Sprinkler* e *Rain* e, essas duas tem como filha a variável *WetGrass*; todas as variáveis são binárias (verdadeiro – T; falso – F). O evento grama molhada representado pela variável *WetGrass* (W) tem duas causas possíveis, o aspersor está ligado que é representado pela variável *Sprinkler* (S) ou está chovendo, representado pela variável *Rain* (R). O evento tempo nublado (*Cloudy* denotado por C) influencia a ocorrência de chuva e se o aspersor será ligado para molhar a grama. A força dos relacionamentos é mostrada nas tabelas de probabilidades (MARQUES; DUTRA, 2002). Por exemplo, a probabilidade da grama está molhada sabendo que choveu e que o aspersor não foi ligado é de 90%. No caso de *Cloudy* (tempo nublado) que não tem nó pai, sua probabilidade condicional é a própria marginal (50%).

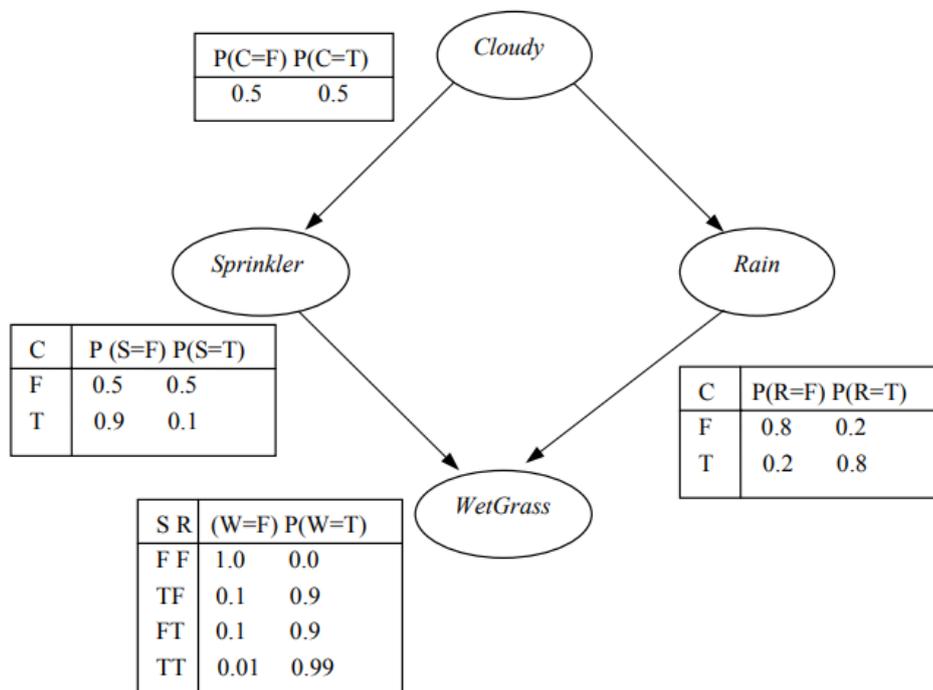


Figura 4: Exemplo de rede Bayesiana. Fonte: Marques e Dutra (2002)

As distribuições locais de probabilidade condicional, juntamente com a estrutura da rede são suficientes para representar a distribuição de probabilidade conjunta do domínio (MARGARITIS, 2003). Definida como

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa_i), \quad (2)$$

em que, Pa_i é o conjunto que contém os pais de X_i . Quer dizer que a distribuição de probabilidade conjunta é a multiplicação das distribuições locais de probabilidade condicional, que envolve um nó e seus pais.

No caso do exemplo, temos a distribuição conjunta dada por

$$P(C, S, R, W) = P(C)P(S|C)P(R|C)P(W|S, R). \quad (3)$$

Importante definição em RB é *Markov blanket*, uma variável qualquer $X \in \mathcal{U}$, *Markov blanket* $BL(X) \subseteq \mathcal{U}$ é qualquer conjunto de variáveis tais que para qualquer $Y \in \mathcal{U} - BL(X) - \{X\}$, $X \perp Y | BL(X)$ (MARGARITIS, 2003). Em outras palavras, é o conjunto formado pelos nós pais, filhos e pais de filhos da variável de interesse, como mostra o exemplo da Figura 5.

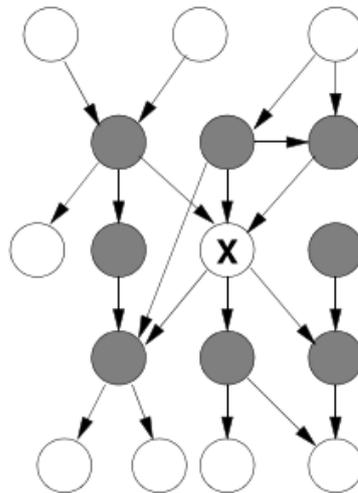


Figura 5: Exemplo de *Markov blanket* da variável X . Os nós sombreados são os membros do *Markov blanket*. Fonte: Margaritis (2003)

3.4.1 Aprendizagem estrutural de redes Bayesianas

Na aprendizagem da estrutura de RB a partir dos dados são utilizados algoritmos baseados em pontuação, restrição e híbridos. Os que são baseados em pontuação utilizam um escore

que mede quão bem a rede proposta se ajusta ao conjunto de dados e é realizada a busca pela rede que maximiza esta pontuação (MARGARITIS, 2003).

Tomando \mathcal{G} como a estrutura e \mathcal{D} o conjunto de dados, sua pontuação é dada por

$$Score(\mathcal{G}, \mathcal{D}) = P(\mathcal{G}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{G})P(\mathcal{G})}{P(\mathcal{D})}. \quad (4)$$

Entre as pontuações normalmente empregadas está o Critério de Informação Bayesiano - BIC (*Bayesian Information Criterion*) proposto por Schwarz (1978), sendo

$$BIC = \sum_{i=1}^n \log P_{X_i}(X_i|\Pi_{X_i}) - \frac{d}{2} \log n, \quad (5)$$

em que d é o número de parâmetros da distribuição global.

Nos algoritmos baseados em restrição são utilizados testes de independência condicional para calcular a estrutura da rede Bayesiana, assim os arcos são tratados de forma individual (SCUTARI; DENIS, 2014). Com estes métodos não é possível determinar a direção para todos os arcos, como mencionado por Margaritis (2003) podem existir outras redes não direcionadas que fazem parte da mesma classe de equivalência quanto a independência condicional entre as variáveis, contida nas estruturas. Estes algoritmos fazem uso da definição de *Markov blanket*, conjunto de nós que d-separa um nó de interesse dos demais.

Por fim, os algoritmos híbridos são combinações dos dois anteriores. Ou seja, a estrutura de RB é encontrada de acordo com uma pontuação que a maximiza e com a restrição de que haja independência condicional entre variáveis não conectadas por um arco (NAGARAJAN; SCUTARI; LÈBRE, 2013). O processo de construção é realizado em duas etapas, ocorrendo primeiro a fase de restrição e em segundo a busca pela maximização da pontuação.

3.4.1.1 Hill-Climbing

Algoritmo de pontuação, Hill-Climbing (HC) aprende a estrutura da rede de modo a maximizar sua pontuação, explorando o espaço de busca iniciado a partir de uma estrutura de rede, geralmente vazia, e adicionando, excluindo ou invertendo um arco de cada vez até que a pontuação não possa mais ser melhorada (SCUTARI; DENIS, 2014).

Para escapar de um máximo local são utilizadas técnicas de reinício e perturbação. O primeiro é o número de vezes que o algoritmo vai reiniciar o processo na busca da estrutura que melhor represente a base de dados de acordo com a medida de pontuação desejada. Já o segundo, trata-se do número de perturbações que serão realizadas em cada reinício ao encontrar um máximo local, estas perturbações são as operações de adição, remoção e inversão de arcos entre as variáveis (CHICKERING; GEIGER; HECKERMAN, 1995).

A Figura 6 mostra operações realizadas pelo algoritmo para encontrar a RB que maximize a pontuação e não seja um ótimo local.

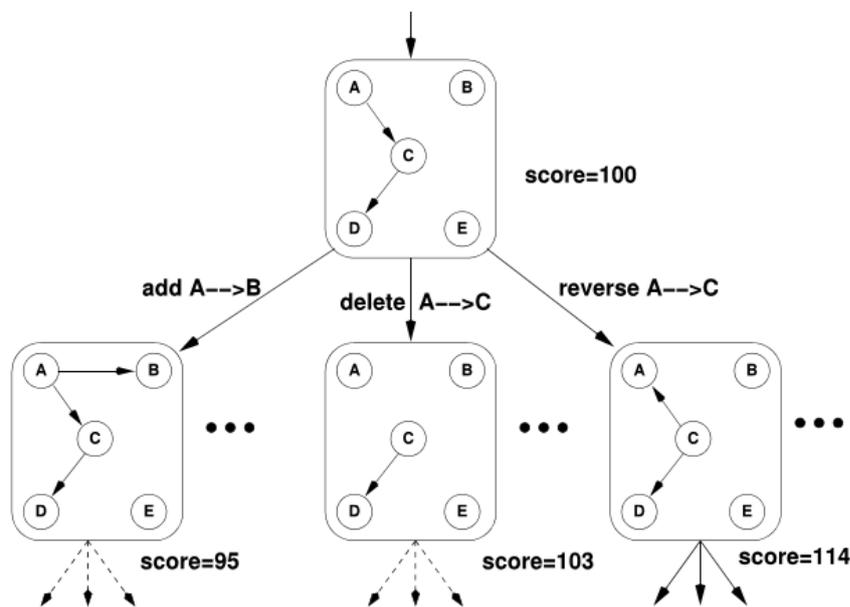


Figura 6: Ilustração do procedimento Hill-Climbing. Fonte: Margaritis (2003)

3.4.1.2 Search Tabu

É um algoritmo de pontuação e como o próprio nome sugere tabu está ligado a “proibição de algo”, que neste caso, é a estacionariedade a uma solução ótima local. Como abordagem meta-heurística tem a capacidade de resolver problemas de otimização como descrito por Glover (1989).

Uma “lista tabu” é incorporada ao algoritmo, que representa a quantidade de movimentos considerados proibidos, isso impede o retorno ao ótimo local já visitado e evita o ciclismo (GENDREAU; POTVIN, 2010).

Assim, na rede Bayesiana as arestas são adicionadas, removidas e tem a direção alterada respeitando a lista que faz uso de memória adaptativa e exploração responsiva. Através da memória adaptativa é possível realizar a busca de forma econômica e efetiva no espaço de soluções. E, a exploração responsiva visa descobrir boas soluções ao explorar novas áreas promissoras (KUMAR; VENKATESWARLU, 2014).

De acordo com Glover e Laguna (1997), as estruturas de memória fazem referência a quatro dimensões: recência, frequência, qualidade e influência. As duas primeiras são complementares, recência diz respeito a memória de curto prazo que rastreia as soluções que mudaram recentemente durante a pesquisa e frequência verifica os elementos que apareceram nas soluções anteriores.

A qualidade refere-se à capacidade de diferenciar o mérito das soluções visitadas durante a pesquisa, e influência considera o impacto das escolhas feitas durante a pesquisa, não apenas na qualidade, mas também na estrutura (GLOVER; LAGUNA, 1997).

A pesquisa Tabu gera inicialmente soluções aleatórias e então, busca a vizinhança para cada uma dessas soluções. Quando o melhor vizinho não supera a solução atual, este entra para a lista tabu, caso contrário passa a ser a solução “inicial” para a próxima iteração. O critério de parada do algoritmo é o número fixo de iterações ou o esforço computacional, definido previamente (ESWARI; KAVYA, 2016).

3.4.1.3 Grow-Shrink

Grow-Shrink (GS) é baseado em restrições e foi o primeiro algoritmo a utilizar o conceito de *Markov blanket*, o que facilita a aprendizagem estrutural da rede Bayesiana (NAGARAJAN; SCUTARI, 2013).

Como descrito em Margaritis (2003), inicialmente identifica-se o *Markov blanket*, vizinhos locais, de todas variáveis e em seguida determina-se a estrutura da rede. Ou seja, verifica-se quais nós do passo anterior realmente são vizinhos diretos (pais e filhos), através de testes de independência condicional como informação mútua (IM) (COVER; THOMAS, 2006) em que pode-se utilizar a distribuição assintótica $\chi^2_{(R-1)(C-1)L}$ ou a abordagem de permutação de Monte Carlo (NAGARAJAN, SCUTARI, LÈBRE, 2013). Posteriormente, busca-se determinar a direção das arestas e caso ocorra ciclos, tais arestas tem a direção alterada. Até que todas as variáveis estejam conectadas por arcos direcionados ou não-direcionados, mantendo a aciclicidade.

3.4.1.4 Aracne

ARACNE – Algoritmo para a Reconstrução de Redes Celulares Precisas (*Algorithm for the Reconstruction of Accurate Cellular Networks*) apresentado por Margolin et al. (2006) é um algoritmo de busca local baseado em restrição. Em que, utiliza a IM de pares de variáveis para identificar a relação entre elas, sem considerar quaisquer suposições. Posteriormente, através da desigualdade de processamento de dados (DPD), propriedade teórica da informação, elimina-se as possíveis interações indiretas (MARGOLIN et al., 2006).

Em Cover e Thomas (1991), a DPD assegura que se o relacionamento existente entre as variáveis X_i e X_z ocorre somente através da variável X_j , então

$$IM(X_i, X_z) \leq \min[IM(X_i, X_j); IM(X_j, X_z)]. \quad (6)$$

Assim, as variáveis são analisadas em trio independente de ordem, eliminando-se a aresta existente entre o par de variáveis que apresentar menor valor de IM. Resultando em um grafo não direcionado.

Na Figura 7, um exemplo da desigualdade de processamento de dados, em que são removidos os arcos tracejados por representarem interações consideradas indiretas e preservadas as linhas sólidas, descrevendo assim a reconstrução de uma rede de genes.

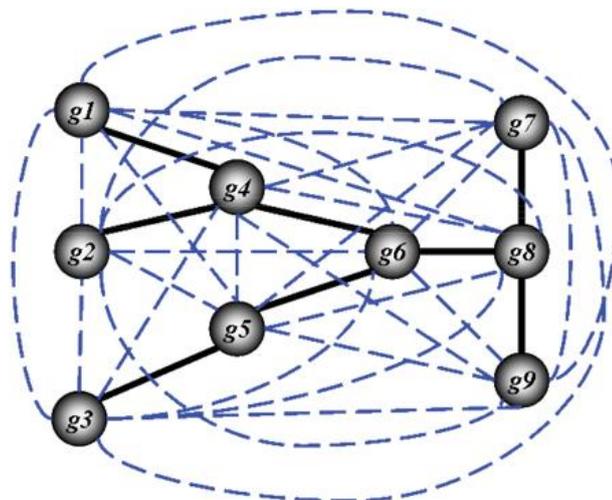


Figura 7: Exemplo da desigualdade de processamento de dados. Fonte: Margolin et al. (2006)

4 MATERIAL E MÉTODOS

4.1 Material

A base de dados utilizada refere-se às notificações de suspeitas de doenças vesiculares registradas na plataforma Web do Sistema Continental de Vigilância Epidemiológica (SivCont), datadas de 2004 a 2017 relativas a todos os estados do Brasil. Foram criadas variáveis a partir das já existentes na plataforma e ainda se acrescentou de outras bases de dados, totalizando 13 variáveis, 79 níveis e 4.494 observações.

- **Variáveis existentes no SivCont**

As Unidades da Federação Brasileira representada por Est. A origem da notificação (Not) com três níveis (proprietário – Pro, terceiros – Ter, vigilância – Vig). Além da variável espécie (Esp) com seis níveis (Bovina – Bov, Suína – Sui, Ovina – Ovi, Caprina – Cap, Equina – Equ, Bubalina – Bub). Se houve coleta de amostra biológica (Amos) para exame laboratorial e se a suspeita da enfermidade foi confirmada (CF), para ambas as variáveis são dois níveis, verdadeiro (V) e falso (F).

- **Variáveis criadas**

A variável Notn com cinco níveis (≤ 10 , ≤ 15 , ≤ 20 , ≤ 25 e > 50) corresponde ao número de notificações realizadas por estado, ano e origem.

As enfermidades vesiculares notificadas (Enf) ao SivCont foram classificadas segundo a similaridade dos sintomas com os da FA e com sua causa, seja por vírus, bactéria, micose e as não infecciosas. Assim, com seis grupos:

- i. Febre Aftosa (FA): Notificações de suspeitas de FA;
- ii. Doenças com sintomas iguais (DSI): Estomatite Vesicular (EV);
- iii. Doenças com sintomas parecidos causadas por vírus (DSPV): Língua Azul, Ectima Contagioso, Diarreia Viral Bovina, Rinotraqueites Viral Bovina, Febre Catarral Maligna, Varíola Bovina, Mamilites Bovina;
- iv. Doenças com sintomas parecidos causadas por bactéria (DSPB): Actinobaciloses, Podo Dermatite Bovina;

- v. Doenças com sintomas parecidos causadas por micose (DSPM): Actinomicoses;
- vi. Doenças com sintomas parecidos não infecciosas (DSPNI): Corpo Estranho e Lesões traumáticas, Mal do Eucalipto, Fotossensibilidade, Irritantes Químicos e Queimaduras, Intoxicações.

A partir dos anos estudados criou-se a variável período (Per) com três níveis, considerando as análises realizadas pelo MAPA em relação à imunidade da população vacinada e a circulação viral (MAPA, 2007a; MAPA, 2011; MAPA, 2016), assim como a implantação de zonas livres da FA (MAPA, 2018b). Em que de 2004 a 2006 chama-se período crítico (PCri), momento em que houve o último surto de FA no Brasil, nos estados do Paraná e Mato Grosso do Sul. O período de mudança (PMud) vai desde 2007 a 2013, fase em que ocorreu a perda e mais tarde a recuperação da condição livre com vacinação para os estados afetados pelo surto no período anterior, também foram reconhecidos como livres com vacinação estados do norte e nordeste, e livre sem vacinação o estado de Santa Catarina. Já no período de estabilidade (PEst), de 2014 a 2017 a condição de livre com vacinação foi estendido para quase todo o país, com o Amapá, Amazonas e Roraima sem reconhecimento até 2017.

O tempo entre o aparecimento dos sinais de enfermidade vesicular até a conclusão final por parte dos SVO foi dividido em dois intervalos – *timeliness* (HOINVILLE et al., 2013). O *timeliness-1* (TL.1) representa o tempo entre o provável início dos sintomas até a notificação realizada pelo proprietário, terceiro ou pela própria vigilância veterinária. De acordo com a evolução teórica da FA (MAPA, 2009), esta variável foi dividida em três níveis:

- i. Lesões (LSO): de zero a seis dias, quando aparecem os sintomas. Esta é a melhor época para isolamento viral;
- ii. Recuperação com proteção (RCP): mais de seis dias, quando começa a recuperação e é a melhor época para colheita de soro;
- iii. Portador com proteção (PP): mais de 16 dias.

O *timeliness-2* (TL.2) dividido em cinco níveis (≤ 1 dia, ≤ 3 dias, ≤ 14 dias, ≤ 30 dias e > 30 dias) representa a duração de tempo desde a primeira visita dos SVO à propriedade com suspeita de animais infectados após recebida a notificação até a confirmação final.

- **Variáveis de outras fontes**

A variável imunidade da população vacinada (IP) com cinco níveis (IP > 90%, IP 80-89%, IP < 80%, IP 0% e IP.NA) para cada período e por estado, de acordo com a descrição realizada pelo MAPA em suas pesquisas (MAPA, 2007a; MAPA 2011; MAPA, 2015), na qual se observou a imunidade dos rebanhos bovino e bubalino vacinados para os vírus A, O e C. Os dois últimos níveis desta variável correspondem ao estado de Santa Catarina (livre sem vacinação) e aos estados não avaliados nas investigações sorológicas, respectivamente.

Observando-se também o processo de implantação de zonas livres (IZL) da FA para cada período (MAPA, 2018b), com seis níveis:

- i. Livre com vacinação (L.CV): condição sanitária concedida pela OIE;
- ii. Livre sem vacinação (L.SV): condição sanitária concedida pela OIE;
- iii. Livre sem reconhecimento (L.SR): não reconhecida pela OIE, mas aceita pelo MAPA;
- iv. Zona de proteção (Z.Prot): condição sanitária utilizada pelo MAPA;
- v. Zona infectada (Z.Inf);
- vi. Zona livre com reconhecimento suspenso (Z.LRS): suspensão realizada pela OIE.

Acrescentou-se ainda ao banco de dados, informações da população (Pop) das espécies notificadas por estado e ano, com seis níveis (≤ 1 milhão, ≤ 7 milhões, ≤ 13 milhões, ≤ 19 milhões, ≤ 25 milhões e > 25 milhões) obtidas no Instituto Brasileiro de Geografia Estatística (IBGE).

4.2 Métodos

Para verificar a existência de associação entre as variáveis por meio de seus níveis aplicou-se a técnica de mineração de regras de associação (*association rules mining*) (AGRAWAL; IMIELINKI; SWAMI, 1993), através do algoritmo Apriori (AGRAWAL; SRIKANT, 1994) com combinações dos valores mínimos de *support* (30%, 35%, 40%, 45%, 50% e 55%) e *confidence* (50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% e 95%), escolhidos de forma arbitrária, com aumentos de 5%.

Foram eliminadas as regras redundantes, isto é, aquelas que são sub regras de outras mais gerais, além das não-significativas que foram avaliadas através do Teste Exato de Fisher

(HAHSLER; HORNIK, 2007) para 5% de significância, resultando em regras e variáveis associadas para cada combinação de *support* e *confidence*, para então realizar a aprendizagem estrutural das redes Bayesianas.

Na fase de aprendizagem foram utilizados os algoritmos de pontuação Hill-Climbing (NAGARAJAN; SCUTARI; LÈBRE, 2013) e Tabu Search (GLOVER; LAGUNA, 1998), de restrição Grow-Shrink (MARGARITIS, 2003) e de descoberta local também com restrição por Aracne (MARGOLIN et al., 2006).

Com Hill-Climbing iniciou-se a aprendizagem sem nenhum valor para reinício e perturbação, em seguida foram realizadas combinações destes parâmetros para os valores 10, 20, 50 e 100. Com Tabu manteve-se a lista fixa em 10 e o número máximo de interações variando de 10 a 100, com intervalo de 10. A rede que apresentou menor *Bayesian Information Criterion* (BIC) (SCHAWARZ, 1978) em cada método de pontuação foi selecionada como sendo a que melhor representa os dados. Em Grow-Shrink a restrição de independência condicional foi através da informação mútua com 10 permutações Monte Carlo. E, em Aracne a restrição foi por informação mútua em pares.

As estruturas das redes aprendidas pelos diferentes métodos foram comparadas através de suas topologias, em que se observou a presença e/ou ausência de arcos não direcionados entre as variáveis. Tomando uma das redes como verdadeira, a métrica Distância de Hamming Estrutural – SHD (*Structural Hamming Distance*) indicou o número de arcos que as diferenciam, ou seja, a quantidade de operações necessárias seja adição e/ou remoção de arcos, para que as estruturas fossem equivalentes (TSAMARDINOS; BROWN; ALIFERIS, 2006).

Realizou-se o ajuste das maiores redes (8, 10 e 13 variáveis) escolhidas como as melhores com os algoritmos HC e Tabu que são totalmente direcionadas, analisou-se o quanto cada uma recuperaria das informações contidas na base de dados, através das probabilidades marginais.

O conhecimento prévio quanto ao relacionamento entre as variáveis pode ser incorporado aos algoritmos através de lista de conexões obrigatórias (*whitelist*) e lista de conexões proibidas (*blacklist*).

Posteriormente, a aprendizagem foi realizada com os algoritmos de pontuação para as 13 variáveis, utilizando-se:

- i. *whitelist* (Est \rightarrow Esp; Esp \rightarrow TL.1; Esp \rightarrow Enf; Enf \rightarrow Not; TL.1 \rightarrow TL.2; TL.2 \rightarrow CF; Not \rightarrow TL.2);
- ii. *blacklist* (CF \rightarrow TL.2).

Para a geração das regras de associação utilizou-se o pacote `arules` (versão 1.6-1) (HAHSLER; GRÜN; HORNIK, 2005), na aprendizagem e ajuste das redes Bayesianas `bnlearn` (versão 4.3) (SCUTARI, 2010) e `gRain` (versão 1.3-0) (HØJSGAARD, 2012), respectivamente, e os pacotes `Rgraphviz` (versão 2.24.0) e `scatterplot3d` (versão 0.3-41) para os gráficos, do software R (versão 3.5.1).

5 RESULTADOS

Para os diversos valores mínimos de *support* e *confidence* foram obtidos os valores de *lift* e as regras de associação. Após eliminadas as regras redundantes e não significativas, observou-se que variando a *confidence* não houve mudança quanto ao total de variáveis associadas, assim para *support* de 30% as 13 variáveis estão presentes nas regras, para 35% são 10 variáveis, 40% são 8 variáveis, 45% são 7 variáveis, 50% são 5 variáveis e para 55% são 4 variáveis.

Na Figura 8 é possível observar o comportamento do número de regras de acordo com os critérios de descoberta de associação, depois de eliminadas as regras redundantes e não significativas. Com as 13 variáveis o número de regras variou entre 377 e 922, já com o mínimo de variáveis (quatro) manteve-se constante em quatro regras, com exceção de três regras obtidas quando o *support* foi 55% e *confidence* 95%.

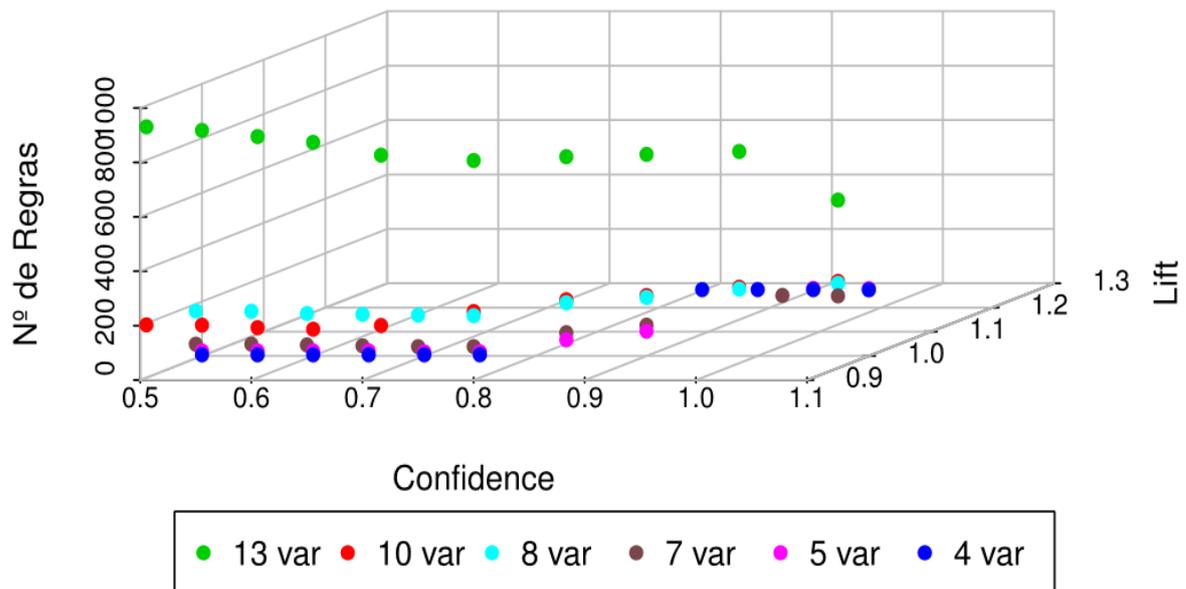
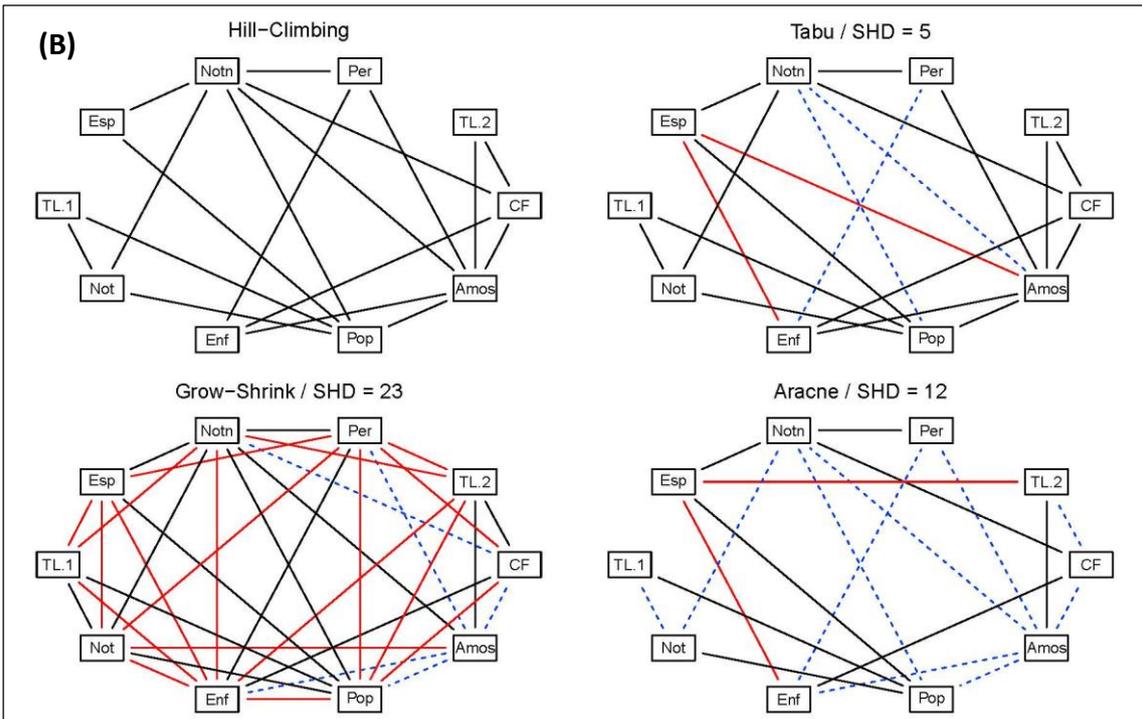
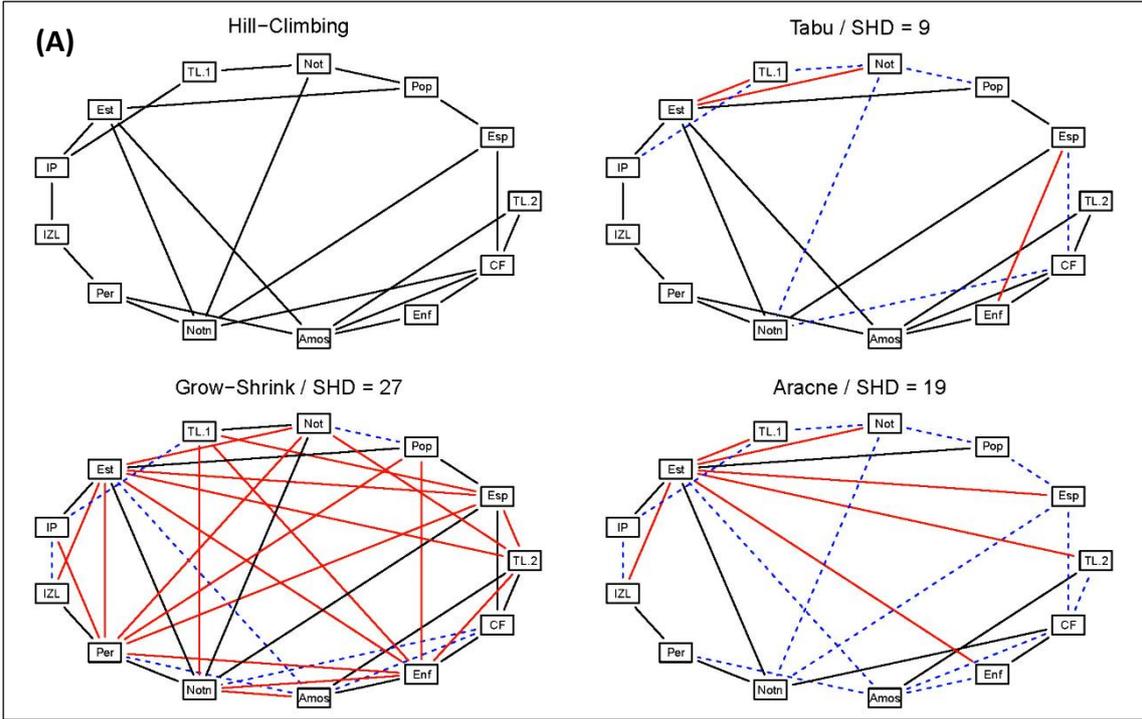
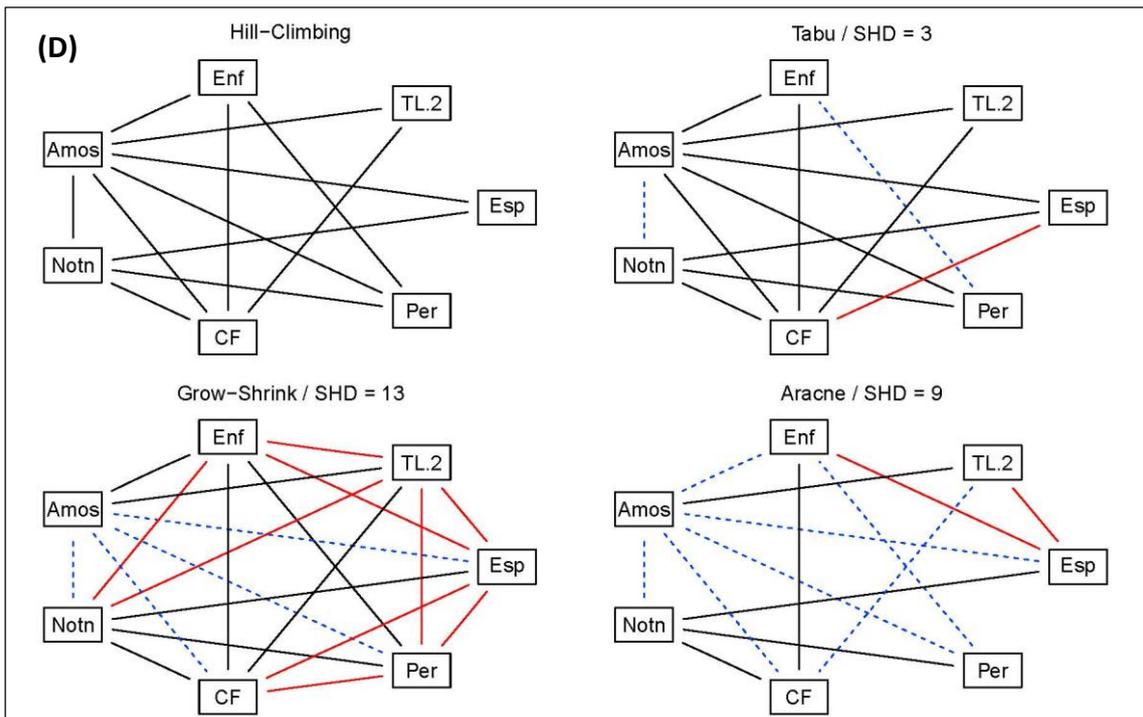
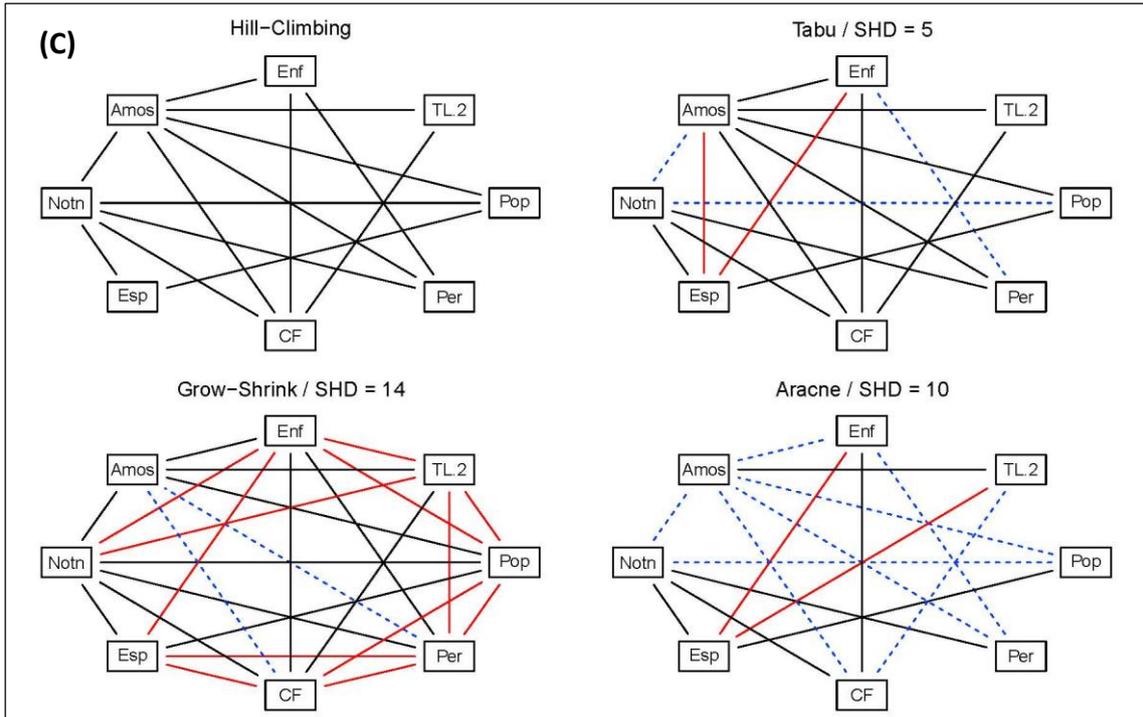


Figura 8: Número de regras de associação conforme *support*, *confidence* e *lift*.

Na Figura 9 são comparadas as topologias das redes aprendidas por HC com os dos demais métodos de acordo com as variáveis presentes nas regras de associação, e SHD mostra em quantos arcos ocorrem as diferenças. As redes obtidas por algoritmos de pontuação são semelhantes e no caso de quatro e cinco variáveis não houve diferença.





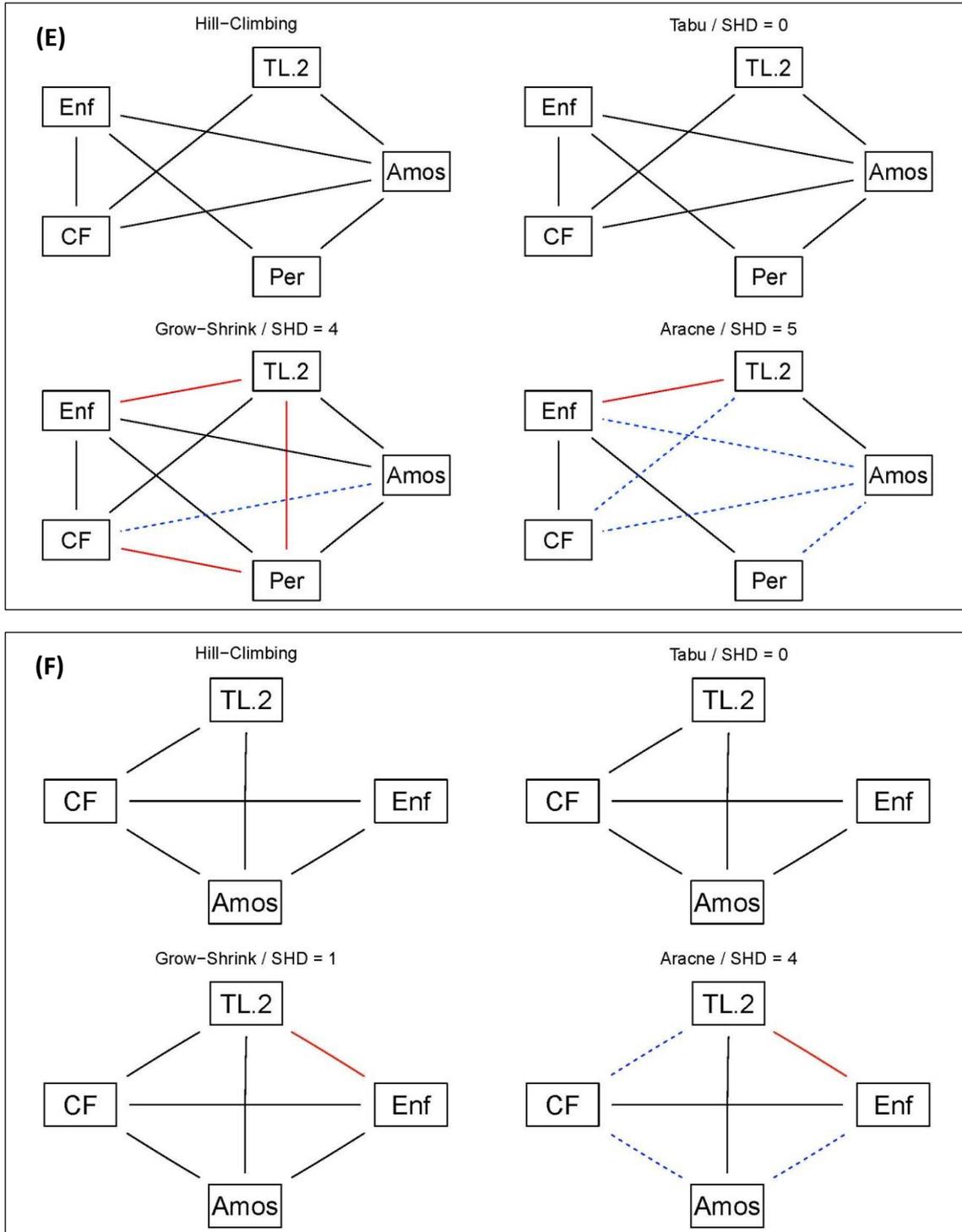


Figura 9: Comparação da topologia das redes Bayesianas de HC com as de Tabu, GS e Aracne. SHD mostra o número de arcos que as diferencia, (—) comum, (---) somente em HC, (—) somente na outra rede. (A) 13 variáveis, (B) 10 variáveis, (C) 8 variáveis, (D) 7 variáveis, (E) 5 variáveis e (F) 4 variáveis.

As redes aprendidas propostas com 13 (Figura 10) e 8 (Figura 12) variáveis apresentaram os menores BIC com o algoritmo HC, e a menor rede é a que se destaca entre todas como

a de melhor ajuste. A maior distância entre os critérios de ajuste para os diferentes métodos foi observada quando todas as variáveis estavam presentes na rede (471,31). Com 10 variáveis (Figura 11) o método Tabu mostrou-se com o melhor ajuste e com diferença de ajuste próxima a encontrada com 8 variáveis (77,79).

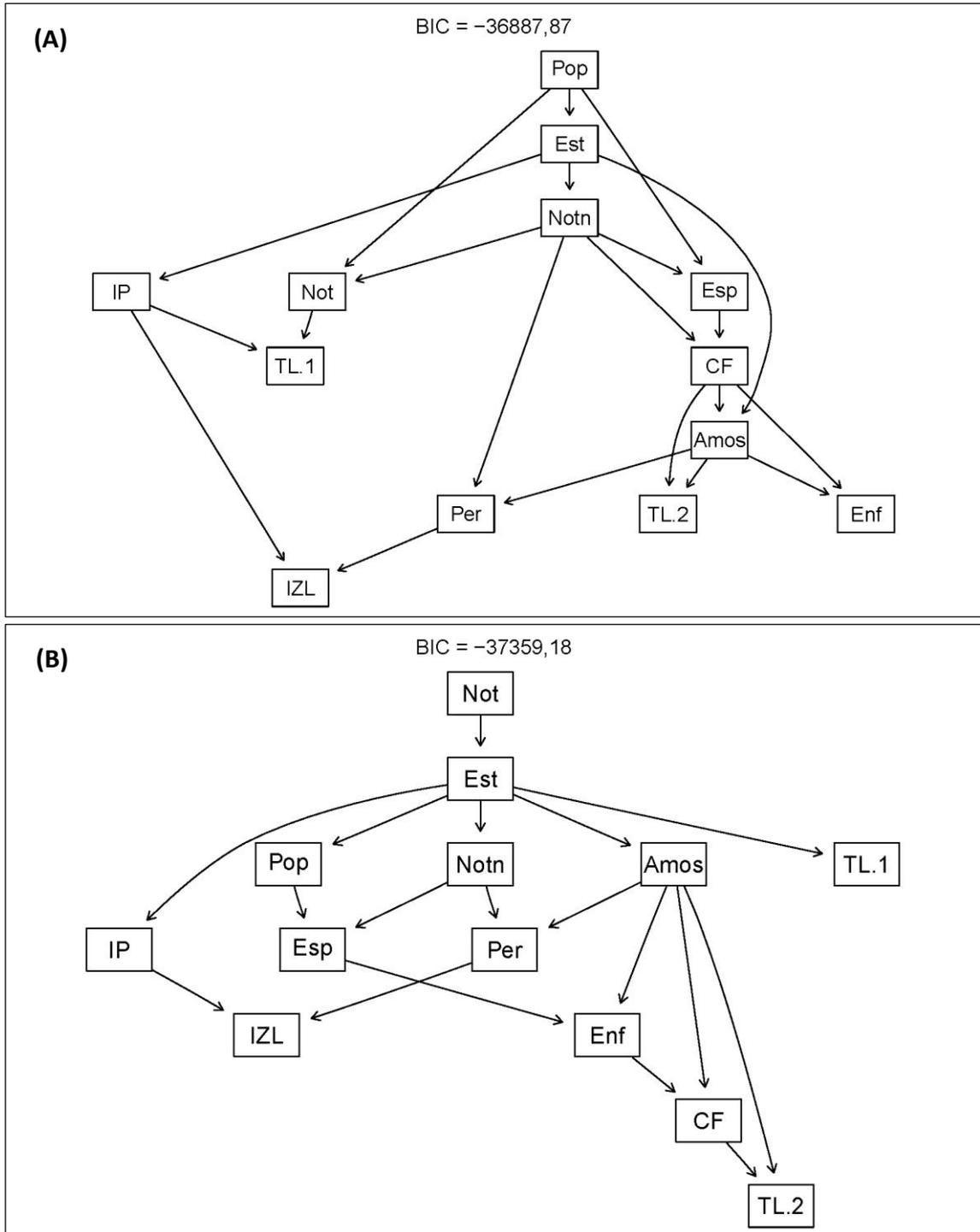


Figura 10: Rede Bayesiana para 13 variáveis com algoritmo HC (A) e Tabu (B).

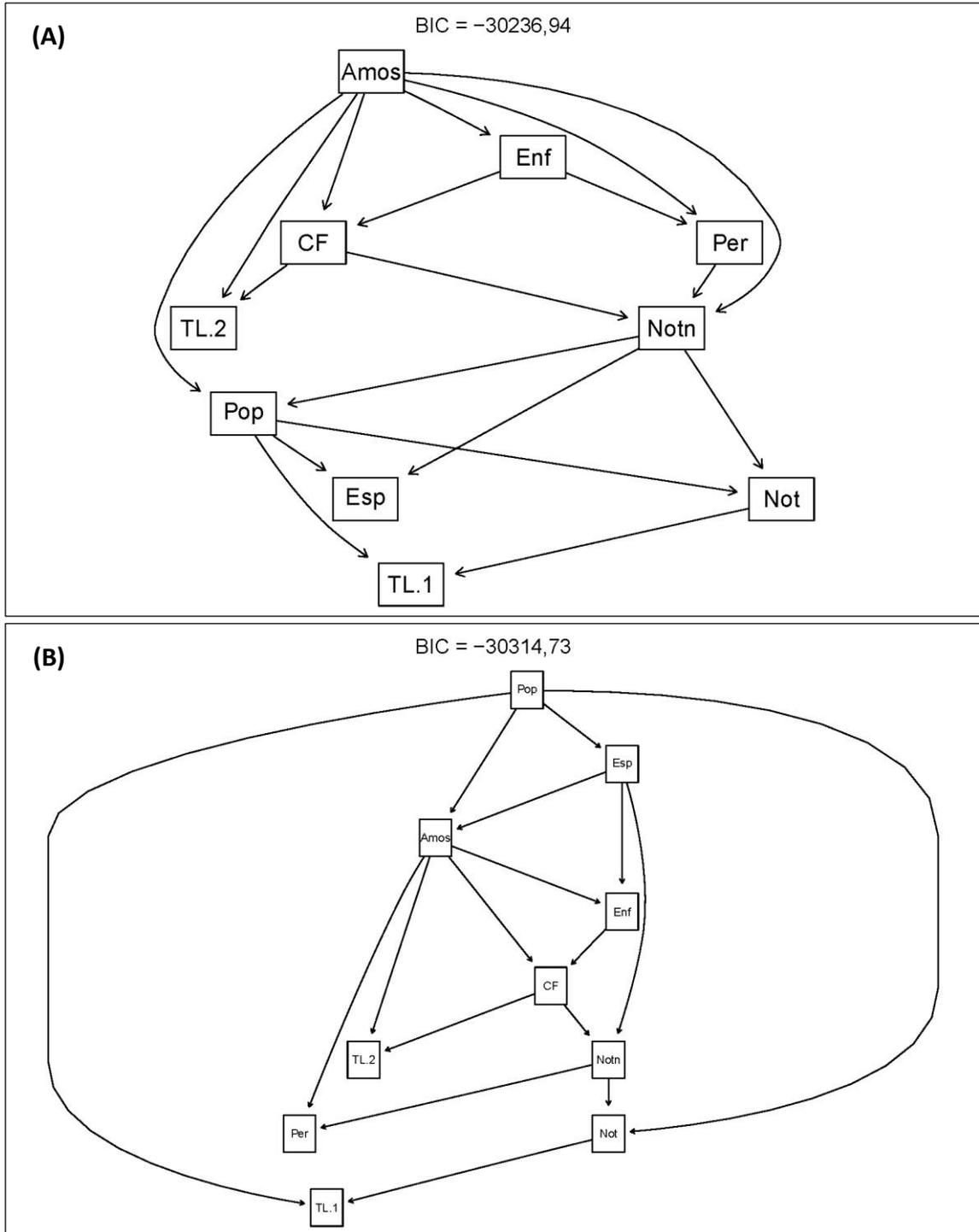


Figura 11: Rede Bayesiana para 10 variáveis com algoritmo HC (A) e Tabu (B).

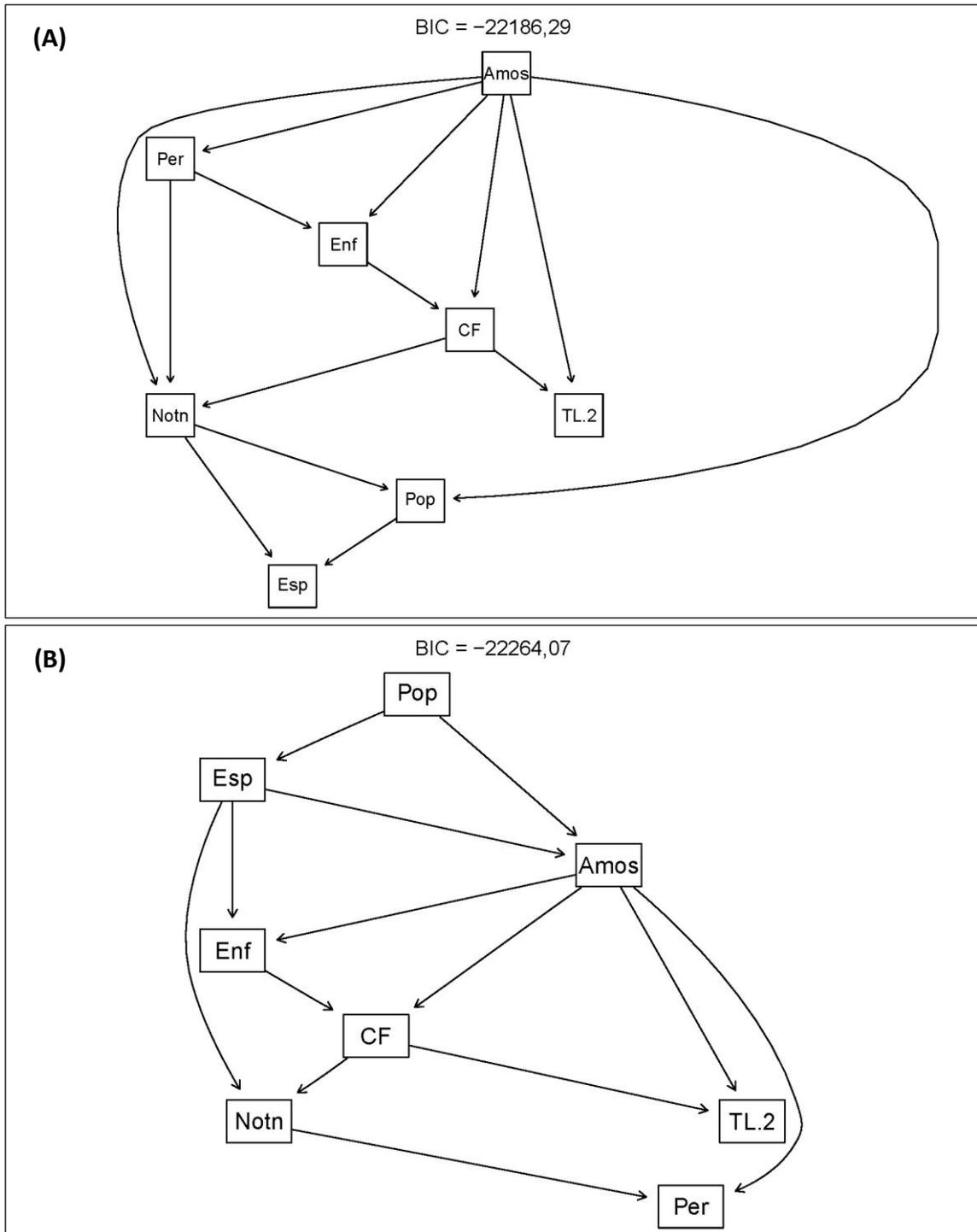


Figura 12: Rede Bayesiana para 8 variáveis com algoritmo HC (A) e Tabu (B).

Com diferentes números de variáveis as redes formadas conseguiram recuperar as informações contidas na base de dados, apresentando pequena variação nas probabilidades marginais em ambos algoritmos de pontuação.

A seguir está a tabela de probabilidades marginais para as oito variáveis (Per, Pop, Est, Notn, Amos, Enf, TL.2, CF) presentes nas três maiores RB (Figura 10, Figura 11 e Figura 12) e

são comparadas com as frequências da base de dados, em vermelho estão as probabilidades das redes que mais se diferenciaram em relação a base de dados.

Tabela 1: Comparação das frequências da base de dados e probabilidades marginais das RB com 8, 10 e 13 variáveis com os algoritmos de pontuação HC e Tabu.

Variáveis	Níveis	Dados	HC			Tabu		
			8v	10v	13v	8v	10v	13v
Per	PCri	0,0407	0,0408	0,0408	0,0429	0,0420	0,0420	0,0431
	PMud	0,3393	0,3393	0,3393	0,3379	0,3395	0,3395	0,3377
	PEst	0,6199	0,6199	0,6199	0,6192	0,6185	0,6185	0,6192
Esp	Bov	0,4528	0,4531	0,4531	0,4615	0,4528	0,4528	0,4615
	Bub	0,0038	0,0039	0,0039	0,0044	0,0038	0,0038	0,0044
	Cap	0,0087	0,0088	0,0088	0,0095	0,0087	0,0087	0,0095
	Equ	0,0263	0,0265	0,0265	0,0248	0,0263	0,0263	0,0248
	Ovi	0,0307	0,0309	0,0309	0,0307	0,0307	0,0307	0,0307
	Sui	0,4777	0,4768	0,4768	0,4692	0,4777	0,4777	0,4692
	TL.2	<=1d	0,7646	0,7644	0,7644	0,7597	0,7644	0,7644
	<=3d	0,0085	0,0085	0,0085	0,0085	0,0085	0,0085	0,0086
	<=14d	0,0490	0,0490	0,0490	0,0498	0,0490	0,0490	0,0492
	<=30d	0,0605	0,0606	0,0606	0,0613	0,0606	0,0606	0,0608
	> 30d	0,1175	0,1175	0,1175	0,1207	0,1175	0,1175	0,1173
Pop	<=1	0,0748	0,0754	0,0754	0,0748	0,0748	0,0748	0,0748
	<=7	0,4784	0,4783	0,4783	0,4783	0,4783	0,4783	0,4783
	<=13	0,2259	0,2259	0,2259	0,2258	0,2258	0,2258	0,2258
	<=19	0,0656	0,0657	0,0657	0,0657	0,0657	0,0657	0,0657
	<=25	0,0995	0,1000	0,1000	0,0995	0,0995	0,0995	0,0995
	> 25	0,0559	0,0546	0,0546	0,0559	0,0559	0,0559	0,0559
Enf	DSI	0,6155	0,6154	0,6154	0,6118	0,6154	0,6154	0,6039
	DSPB	0,0661	0,0661	0,0661	0,0673	0,0661	0,0661	0,0653
	DSPM	0,0011	0,0011	0,0011	0,0012	0,0011	0,0011	0,0012
	DSPNI	0,1500	0,1500	0,1500	0,1526	0,1500	0,1500	0,1599
	DSPV	0,1547	0,1547	0,1547	0,1545	0,1547	0,1547	0,1561
	FA	0,0127	0,0127	0,0127	0,0126	0,0127	0,0127	0,0136
Notn	<=10	0,3147	0,3147	0,3147	0,3131	0,3150	0,3150	0,3131
	<=15	0,0757	0,0757	0,0757	0,0759	0,0759	0,0759	0,0759
	<=20	0,0615	0,0615	0,0615	0,0608	0,0602	0,0602	0,0608
	<=25	0,0767	0,0767	0,0767	0,0764	0,0757	0,0757	0,0764
	> 50	0,4715	0,4715	0,4715	0,4739	0,4732	0,4732	0,4739
Amos	V	0,2181	0,2181	0,2181	0,2225	0,2181	0,2181	0,2181
	F	0,7819	0,7819	0,7819	0,7775	0,7819	0,7819	0,7819
CF	V	0,4143	0,4143	0,4143	0,4169	0,4143	0,4143	0,4248
	F	0,5857	0,5857	0,5857	0,5831	0,5857	0,5857	0,5752

Os resultados (Figura 13) por HC demonstraram o menor BIC (-38886,24) comparativamente a Tabu, porém maior do que o obtido sem restrição; as probabilidades retornadas são idênticas as da base de dados para as variáveis população (Pop), espécie (Esp), coleta de amostra (Amos), confirmação final (CF), imunidade da população (IP) e estado (Est). Para as demais, as probabilidades ficaram próximas, com a maior diferença (1%) para o nível Zona infestada da variável IZL (Introdução de Zonas Livres).

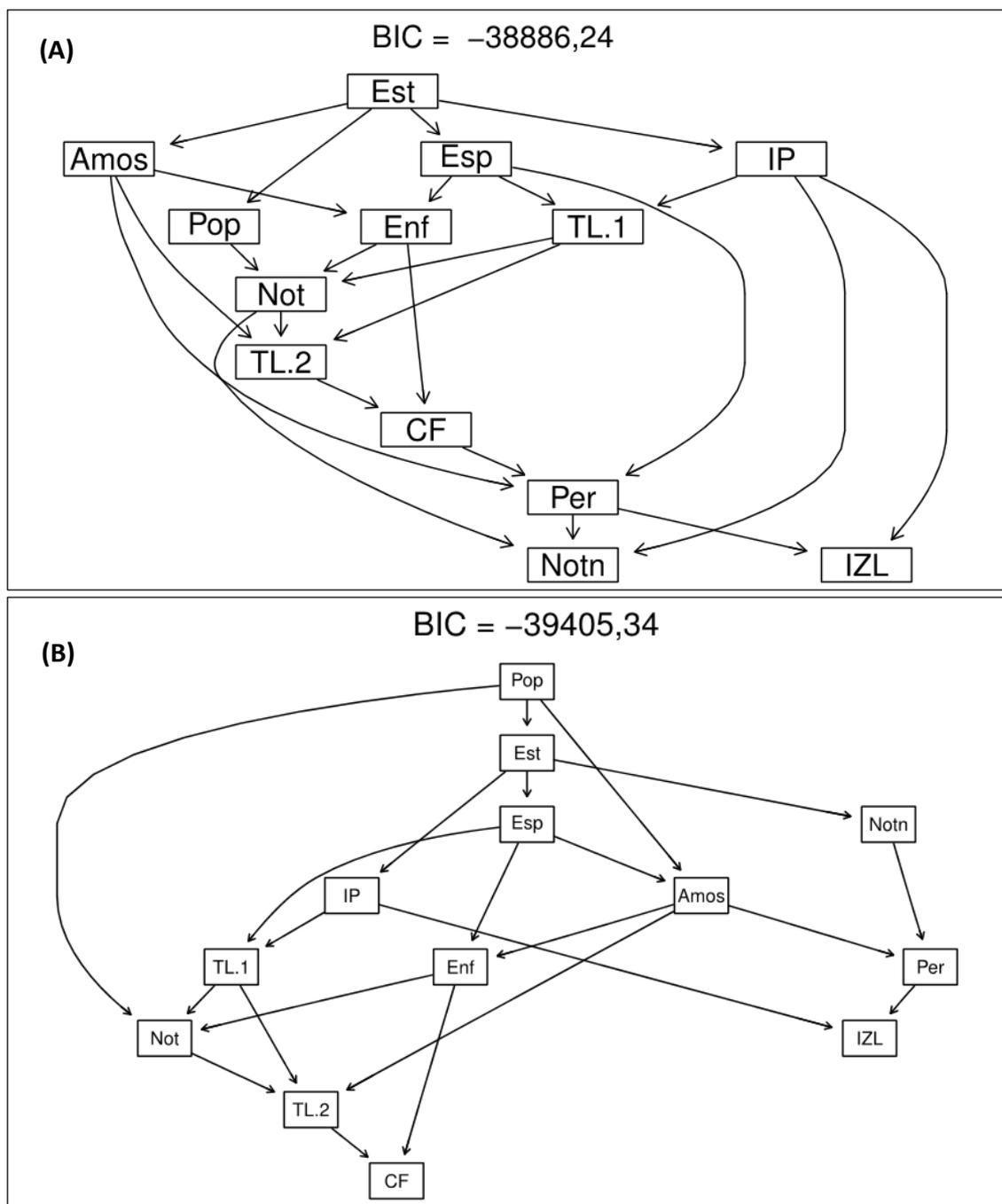


Figura 13: Redes Bayesianas para as 13 variáveis com *whitelist* e *blacklist*, através dos algoritmos HC (A) e Tabu (B).

6 DISCUSSÃO

Como esperado, quanto maior a exigência nos limites mínimos dos parâmetros para a geração das regras de associação, menor o número de regras obtidas e conseqüentemente a quantidade de variáveis associadas também diminui (Figura 8). A queda mais acentuada ocorre quando o *support* passa de 30% para 35% como também se perde três variáveis (Est, IZL e IP) para a construção da rede, devido à baixa probabilidade de ocorrência dos níveis destas variáveis. Torna-se praticável avaliar as regras por diferentes critérios para melhor compreensão da associação entre as variáveis. A combinação de *support* mínimo de 30% independentemente do valor mínimo da *confidence* foi a que manteve todas as variáveis analisadas presentes nas regras.

A frequência (*support*) das regras foi determinante para a seleção das variáveis que iriam participar da aprendizagem das redes Bayesianas, que se justifica devido à natureza do fenômeno em estudo. Com estrutura pré-definida, de modo que na mesma regra não ocorre a presença de mais de um nível da mesma variável.

A aplicação desta mesma abordagem em outros bancos de dados, quer dizer em fenômenos com diferentes características pode resultar em comportamento diferenciado tanto em relação a quantidade de variáveis como a importância dos parâmetros de interesse, ou seja, *support* ou *confidence* pode ser mais importante ou ainda, ambos tenham relevância na seleção das variáveis.

As RB com poucas variáveis dificultaram o entendimento sobre o completo funcionamento do sistema de vigilância e os fatores que envolvem as notificações. Ainda assim, realizou-se aprendizagem para os diferentes números de variáveis e foram comparadas pelos quatro métodos de aprendizagem. Devido à grande quantidade de testes de independência condicional realizados pelo algoritmo de restrição muitos arcos são recuperados, conseqüentemente, são estes os que mais se diferenciam dos esqueletos de HC. Com Aracne tem-se a menor quantidade de arcos, pois são mantidos apenas aqueles que apresentam maiores valores de IM entre pares de variáveis. Embora as estruturas sejam fortemente influenciadas pela escolha de critérios estatísticos específicos e não somente pelas propriedades dos próprios algoritmos (SCUTARI; GRAAFLAND; GUTIÉRREZ, 2018).

Redes obtidas por meio de aprendizagem podem resultar em relações que não condizem com a realidade do fenômeno analisado. Como pode ser visto nas Figuras 10, 11 e 12, a confirmação final (CF) ao atendimento de suspeita de doença vesicular condicionando o tempo que

o SVO leva entre o atendimento inicial e o diagnóstico final (TL.2), esperava-se o contrário nesta relação. Ainda, todas as redes aprendidas pelos algoritmos de pontuação para as três maiores quantidades de variáveis (8, 10 e 13) Enf (classe de enfermidades vesiculares notificadas) recebe influência da coleta de amostra (Amos) o que não pode ser admitido, pois a amostragem só é realizada diante do atendimento de suspeita de alguma enfermidade.

Quando não se conhece o relacionamento entre as variáveis envolvidas no estudo, deve-se aceitar a rede resultante como verdadeira, sendo que conhecendo parcialmente como as variáveis se relacionam utiliza-se deste conhecimento em forma de *blacklist* e *whitelist* para descobrir as demais relações por meio da aprendizagem. No caso do pesquisador ter total conhecimento da topologia da rede, esta seria definida previamente sem realização de aprendizagem, impossibilitando a descoberta de novas e interessantes relações causais ou de dependência.

As probabilidades da base de dados foram recuperadas principalmente pelas redes com 8 e 10 variáveis para os algoritmos de pontuação HC e Tabu que possuem também o melhor ajuste BIC, mesmo apresentando relações de causalidade que não deveriam ocorrer, quando não foram empregadas *whitelist* e *blacklist*. A doença mais notificada foi a estomatite vesicular, caracterizada por apresentar sintomas iguais aos da FA e mais da metade de todos os casos notificados não são confirmados.

A RB com melhor ajuste BIC, utilizando restrição por *whitelist* e *blacklist* (Figura 13 A) ainda não conseguiu recuperar toda a topologia da rede, contudo as probabilidades recuperadas ficaram próximas as da base de dados, a maior diferença foi de 1%.

7 CONCLUSÃO

Observou-se que a mudança nos valores mínimos para *confidence* não alterou as variáveis presentes nas regras de associação, o que foi alterado ao variar o *support*. Conclui-se que é viável aplicar diferentes critérios de medidas de interesse para melhor compreensão de associação no conjunto de variáveis.

As redes com os três maiores números de variáveis sem restrições de listas de conexões recuperaram as probabilidades contidas na base de dados, mas não a topologia completa da rede. O mesmo aconteceu para a RB com todas as variáveis e com as restrições impostas pelo conhecimento prévio, mesmo assim esta última com aprendizagem HC se configura como a melhor.

Sendo importante a opinião de um especialista na área para que mudanças nas direções dos arcos, ou até mesmo remoção, pudessem ser feitas de modo a representar o sistema de vigilância e assim como os fatores que envolvem o processo de notificações. E, posteriormente realizar inferências de modo a contribuir para a melhoria das estratégias e planos que visam o avanço das zonas livres sem vacinação.

REFERÊNCIAS BIBLIOGRÁFICAS

ABIEC. Associação Brasileira das Indústrias Exportadoras de Carne. **Perfil da pecuária no Brasil. Relatório Anual 2018**. Associação Brasileira das Indústrias Exportadoras de Carne, São Paulo, 48 p., 2018.

AGHAIE, A.; SAEEDI, A. Using bayesian networks for bankruptcy prediction: empirical evidence from Iranian Companies. In: **International Conference on Information Management and Engineering**, Kuala Lumpur, p. 450-455, 2009.

AGRAWAL, R.; IMIELISNKI, T.; SWAMI, A. Mining association rules between sets of items in large databases. In: **Proceedings of the 1993 ACM SIGMOD Conference**, washington DC, USA, p. 10, 1993.

AGRAWAL, R; SRIKANT, R. Fast algorithms for mining association rules. In: **Proceedings of the 20th International Conference on Very Large Data Bases Conference**. Santiago, Chile, p. 487-499, 1994.

ALEXANDERSEN, S. ZHANG, Z.; DONALDSON, A. I.; GARLAND, A. J. The pathogenesis and diagnosis of foot-and-mouth disease. **Journal of Comparative. Pathology**, v. 129, n. 1 p. 1-36, 2003.

BEN-GAL, I. Bayesian networks. **Encyclopedia of statistics in quality and reliability**, Wiley Online Library, 2007.

CHICKERING D. M.; GEIGER, D.; HECKERMAN, D. Learning Bayesian Networks: Search Methods and Experimental Results. In: **Proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics**, p. 112-128, 1995.

COSTA, P. P. **Teoria de Grafos e suas Aplicações**. Dissertação (Mestrado Profissional em Matemática Universitária) – Universidade Estadual Paulista, Instituto de Geociências e Ciências Exatas, Rio Claro, 2011.

COVER, T. M.; THOMAS, J. A. **Elements of Information Theory**. New York: John Wiley & Sons; 1991.

COVER, T. M.; THOMAS, J. A. **Elements of Information Theory**. 2nd, New York: John Wiley & Sons; 2006.

ERSEL, D.; GÜNAY, S. Bayesian networks and association analysis in knowledge discovery process. **İstatistikçiler Dergisi**, v. 5, p. 51-64, 2012.

ESWARI, J. S.; KAVYA, K. Optimal feed profile for the Rhamnolipid kinet models by usig tabu search: metabolic view point. **AMB Express**, v. 6, n. 116, 2016.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge discovery in databases **AI Magazine AAAI**, v. 17, n. 3, p. 37-54, 1996.

FOSSAERT, H.; LLOPIS, A.; TIGRE, C.H. Sistemas de vigilância epidemiológica. **Boletim de la Oficina Sanitaria Panamericana**, p. 512-528, 1974.

GENDREAU, M.; POTVIN, J.Y. Handbook of metaheuristics, New York, Springer, 2nd edition, v.146, p. 41 - 59, 2010.

GLOVER, F. Tabu Search - parte I. **ORSA Journal on Computing**, v.1, n. 3, 1989.

GLOVER, F.; LAGUNA, M. **Tabu Search**. [S. l.]: Kluwer Academic Publishers, 1997.

GLOVER, F.; LAGUNA, M. Tabu search. **Kluwer Academic Publishers**, Colorado, Boulder, CO, 137 p., 1998.

GONÇALVES, E. C. Regras de associação e suas medidas de interesse objetivas e subjetivas. **Universidade Federal Fluminense**, Instituto de Computação, Niterói, 2005.

HAHSLER, M.; GRÜN, B.; HORNIK, K. arules – A Computational Environment for Mining Association Rules and Frequent Item Sets. **Journal of Statistical Software**, v. 14, 25 p., 2005.

HAHSLER, M.; HORNIK, K. New Probabilistic Interest Measures for Association Rules. **Intelligent Data Analysis**, v 11, n. 5, p. 437-455, 2007.

HAHSLER, M.; BUCHTA, C.; HORNIK, K. Selective association rule generation. **Computational Statistics**, v.23, p. 303-315, 2008.

HALASA, T; WILLEBERG, P.; CHRISTIANSEN, L. E.; BOKLUND, A.; ALKHAMIS, M.; PEREZ, A.; ENØE, C. Decisions on control of foot-and-mouth disease informed using model predictions. **Preventive Veterinary Medicine**, v. 112, n. 3-4, p. 194-202, 2013.

HOINVILLE, L. J.; ALBANB, L.; DREWEC, J. A.; GIBBENS, J. C.; GUSTAFSOND, L.; HÄSLER, B.; SAEGERMANE, C.; SALMAN, M.; STÄRK, K. D. C. Proposed terms and concepts for describing and evaluating animal-health surveillance systems. **Preventive Veterinary Medicine**, Elsevier, v.112, p. 1-12, 2013.

HØJSGAARD, S. Graphical independence networks with the gRain package for R. **Journal of Statistical Software**, v. 46, n. 10, p. 1-26, 2012. Disponível em: <<http://www.jstatsoft.or/v46/i10/>>.

HORST, H. S.; DIJKHUIZEN, A. A.; HUIRNE, R. B. M.; MEUWISSEN, M. P. M. Monte Carlo simulation of virus introduction into the Netherlands. **Preventive Veterinary Medicine**, v. 41, n. 2-3, p. 209-229, 1999.

KOSKI, T.; NOBLE, J. M. **Bayesian networks: an introduction**. John Wiley & Sons: Sussex, 2009.

KUMAR, B. S.; VENKATESWARLU, CH. Inverse modeling approach for evaluation of kinetic parameters of a biofilm reactor using tabu search. **Water Environment Research**, v. 86, n. 8, p.675-686, 2014.

LAWSON, A. B.; ONICESCU, G.; ELLERBE, C. Foot and mouth disease revisited: Re-analysis using Bayesian spatial susceptible-infectious-removed models. **Spatial and Spatiotemporal Epidemiology**, v. 2, n. 3, p. 185-194, 2011.

MAPA. Ministério da Agricultura, Pecuária e Abastecimento. **Estudo epidemiológico para avaliação de transmissão viral na zona livre de febre aftosa com vacinação**. Secretaria de Defesa Agropecuária. Departamento de saúde Animal. Brasília, DF, 17 p., 2016.

MAPA. Ministério da Agricultura, Pecuária e Abastecimento. **Monitoramento sorológico para avaliação da eficiência da vacinação contra a febre aftosa na zona livre**. Secretaria de Defesa Agropecuária. Departamento de saúde Animal. Brasília, DF, 16 p., 2011.

MAPA. Ministério da Agricultura, Pecuária e Abastecimento. **Monitoramento sorológico para avaliação da eficiência da vacinação contra a febre aftosa na zona livre**. Secretaria de Defesa Agropecuária. Departamento de saúde Animal. Brasília, DF, 8 p., 2015.

MAPA. Ministério da Agricultura, Pecuária e Abastecimento. **Plano de ação contra a febre aftosa. Atendimento a notificação de suspeita de doença vesicular**. Secretaria de Defesa Agropecuária. Departamento de saúde Animal. Brasília, DF, 96 p., 2009.

MAPA. Ministério da Agricultura, Pecuária e Abastecimento. **Programa Nacional de Erradicação e Prevenção de Febre Aftosa. Plano Estratégico 2017 – 2023**. Secretaria de Defesa Agropecuária. Departamento de saúde Animal. Brasília, DF, 83 p., 2017.

MAPA. Ministério da Agricultura, Pecuária e Abastecimento. **Programa Nacional de Erradicação e Prevenção de Febre Aftosa. Avaliação da imunidade populacional resultante das campanhas de vacinação contra a febre aftosa**. Secretaria de Defesa Agropecuária. Departamento de saúde Animal. Brasília, DF, 68 p., 2007a.

MAPA. Ministério da Agricultura, Pecuária e Abastecimento. **Programa Nacional de Erradicação e Prevenção de Febre Aftosa. Vigilância veterinária de doenças vesiculares**. Secretaria de Defesa Agropecuária. Departamento de Saúde Animal. Brasília, DF, 49 p., 2007b.

MAPA. Ministério de Agricultura, Pecuária e Abastecimento. **Brasil Livre da Febre Aftosa**. Ministério da Agricultura, Pecuária e Abastecimento, 2018a. Disponível em: <<http://www.agricultura.gov.br/assuntos/sanidade-animal-e-vegetal/saude-animal/programas-de-saude-animal/febre-aftosa/brasil-livre-da-aftosa>>. Acesso em: 09 out. 2018.

MAPA. Ministério da Agricultura, Pecuária e Abastecimento. **Evolução das zonas livres de febre aftosa no Brasil**. Divisão de Febre Aftosa. Departamento de saúde Animal. Brasília, DF, 2018b. Disponível em: <http://www.agricultura.gov.br/assuntos/sanidade-animal-e-vegetal/saude-animal/programas-de-saude-animal/febre-aftosa/documentos-febre-aftosa/EvoluodaszonaslivresdeFA_2018.pdf>. Acesso em: 20 dez. 2018.

MARGARITIS, D. **Learning Bayesian Network Model Structure** Thesis (Philosophy) - School of Computer Science, Carnegie-Mellon University, Pittsburgh, PA, 127 p., 2003.

MARGOLIN, A. A.; NEMENMAN, I.; BASSO, K.; WIGGINS, C.; STOLOVITZKY, G.; FAVERA, R. D.; CALIFANO, A. ARACNE: An algorithm for the reconstruction of gene

regulatory networks in a mammalian cellular context. **BMC Bioinformatics**, v. 7 Suppl 1, S7, 2006.

MARQUES, R. L.; DUTRA, I. Redes bayesianas: o que são, para que servem, algoritmos e exemplos de aplicação. Coppe – Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil, 2002.

MARTÍNEZ-LÓPEZ, B.; PEREZ, A. M.; TORRE, A.; RODRIGUEZ, J. M. Quantitative risk assessment of foot-and-mouth disease introduction into Spain via importation of live animals. **Preventive Veterinary Medicine**, v. 86, n. 1-2, p. 43-56, 2008.

MCLAWS, M.; RIBBLE, C.; MARTIN, W.; WILESMITH, J. Factors associated with the early detection of foot-and-mouth disease during the 2001 epidemic in the United Kingdom. **The Canadian Veterinary Journal**, v. 50, n. 1, p. 53, 2009.

MELO, E. C.; SARAIVA, V.; ASTUDILLO, V. Review of the status of foot and mouth disease in country of South America and approaches to control and eradication. **Revue Scientifique et Technique (International Office of Epizootics)**, v. 21, n. 3, p. 429-436, 2002.

NAGARAJAN, R.; SCUTARI, M. Impact of Noise on Molecular Network Inference. **Plos One**, v. 8, n. 8, 2013.

NAGARAJAN, R.; SCUTARI, M.; LÈBRE, S. **Bayesian networks in R with applications in systems biology**. New York: Springer, v. 48, 2013.

NEAPOLITAN, R. E. **Learning Bayesian Networks**. Upper Saddle River: Pearson, 2004.

OIE. World Organisation for Animal Health. Chapter 2.1.8: **Foot and Mouth Disease. (Infection with foot and mouth disease virus)**, 2017. Disponível em: <http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/2.01.08_FMD.pdf>. Acesso em: 15 ago. 2017.

OIE. World Organization for Animal Health. Chapter 8.5: **Foot and mouth disease. In: Terrestrial manual code**, 2011. Disponível em: <http://www.oie.int/eng/A_FMD2012/docs/en_chapitre_1.8.5.pdf>. Acesso em: 10 ago. 2017.

OIE. World Organization for Animal Health. **Map of FMD official status**. World Organization for Animal Health, 2018. Disponível em: <http://www.oie.int/fileadmin/Home/js/images/fmd/FMD_SouthAmerica_ENG.png>. Acesso em: 20 nov. 2018.

PANAFTOSA. Centro Pan-Americano de Febre Aftosa. **Panaftosa compromisso com a erradicação da febre aftosa nas Américas desde 1951**. Centro Pan-Americano de Febre Aftosa. 2018.

PANAFTOSA. Centro Pan-Americano de Febre. **Programa hemisférico de erradicação da febre aftosa. Plano de Ação 2011–2020**. Organização Pan-Americana de saúde, 53 p., 2011.

RODIN, A. S.; BOERWINKLE, E. Mining genetic epidemiology data with Bayesian networks I: Bayesian networks and example application (plasma apoE levels). **Bioinformatics**, Oxford University Press, v. 21, n. 15, p. 3273-3278, 2005.

SCHAWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, v. 6, n. 2 p. 461-464, 1978.

SCUTARI, M. bnlearn: Bayesian network structure learning. **R package**, 2010.

SCUTARI, M.; DENIS, J.B. **Bayesian networks with examples in R**. [S.l.]: CRC Press, 2014.

SCUTARI, M.; GRAAFLAND, C. E.; GUTIÉRREZ, J. M. Who learns better bayesian network structures: constraint-based, score-based or hybrid algorithms? In: **Proceedings of Machine Learning Research**, v. 72, p. 416-127, 2018.

SILVERSTEIN, C.; BRIN, S.; MOTWANI, R. Beyond Market Baskets: Generalizing Association Rules to Dependence Rules. **Data Mining and Knowledge Discovery**, v. 2, p. 39-68, 1998.

TSAMARDINOS, I.; BWOWN, L. E.; ALIFERIS, C. F. The max-min hill-climbing Bayesian network structure learning algorithm. **Machine Learning**, v.65, n.1, p. 31-78, 2006.

VOLKOVA, V. V.; BESSELL, P. R.; WOOLHOUSE, M. E. J.; SAVIL, N. J. Evaluation of risks of foot-and-mouth disease in Scotland to assist with decision making during the 2007 outbreak in the UK. **Veterinary Record**, v. 169, n. 5, p. 124-124, 2011.

XIMENES, P. S. M. P. **Sensibilidade de alarme em um sistema de vigilância sindrômica epidemiológica**. Dissertação (Mestrado em Biometria e Estatística Aplicada) – Universidade Federal Rural de Pernambuco, Recife, 2018.

ZAGORECKI, A.; ORZECHOWSKI, P.; HOLOWNIA, K. A system for automated general medical diagnosis using bayesian networks. **Studies in health technology and informatics**, v. 192, n. 1, p. 461-465, 2013.