RAFAEL ISHIBASHI CIPULLO

Comparação e análise da rede de movimento de
bovinos de propriedades positivas e negativas para
brucelose no Estado de Mato Grosso

São Paulo

2013

RAFAEL ISHIBASHI CIPULLO

Comparação e análise da rede de movimento de
bovinos de propriedades positivas e negativas para
brucelose no Estado de Mato Grosso

Dissertação apresentada ao
Programa de Pós-Graduação em
Epidemiologia Experimental
Aplicada às Zoonoses da
Faculdade de Medicina
Veterinária e Zootecnia da
Universidade de São Paulo, para
obtenção do título de Mestre em
Ciências

Departamento:

Medicina Veterinária Preventiva e
Saúde Animal

Área de Concentração:

Epidemiologia Experimental
Aplicada às Zoonoses

Orientador:

Prof. Dr. Marcos Amaku

De acordo:

Orientador

São Paulo

2013

Obs: A versão original se encontra disponível na Biblioteca da FMVZ/USP

Autorizo a reprodução parcial ou total desta obra, para fins acadêmicos, desde que citada a fonte.

DADOS INTERNACIONAIS DE CATALOGAÇÃO-NA-PUBLICAÇÃO

(Biblioteca Virgínie Buff D'Ápice da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo)

T.2770
FMVZ

Cipullo, Rafael Ishibashi
Comparação e análise da rede de movimento de bovinos de propriedades positivas e negativas para brucelose no Estado de Mato Grosso / Rafael Ishibashi Cipullo. -- 2013.
59 f. : il.

Dissertação (Mestrado) - Universidade de São Paulo. Faculdade de Medicina Veterinária e Zootecnia. Departamento de Medicina Veterinária Preventiva e Saúde Animal, São Paulo, 2013.

Programa de Pós-Graduação: Epidemiologia Experimental Aplicada às Zoonoses.

Área de concentração: Epidemiologia Experimental Aplicada às Zoonoses.

Orientador: Prof. Dr. Marcos Amaku.

1. Bovinos. 2. Epidemiologia. 3. Redes complexas. 4. Mato Grosso. 5. Movimentação Animal. I. Título.

UNIVERSIDADE DE SÃO PAULO

FACULDADE DE MEDICINA VETERINÁRIA E ZOOTECNIA

Comissão de Ética no uso de animais

CERTIFICADO

Certificamos que o Projeto intitulado "Comparação e análise da rede de movimento de bovinos de propriedades positivas e negativas para brucelose no estado do Mato Grosso", protocolado sob o nº 2252/2011, não utilizando animais, sob a responsabilidade do(a) Prof. Dr. Marcos Amaku, está de acordo com os princípios éticos de experimentação animal da "Comissão de Ética no uso de animais" da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo e foi aprovado em reunião de 22/6/2011.

We certify that the Research "Comparison and analysis of the network of cattle movement from holdings positive and negative to brucellosis in the state of Mato Grosso", protocol number 2252/2011, not using animals, under the responsibility Prof. Dr. Marcos Amaku, agree with Ethical Principles in Animal Research adopted by "Ethic Committee in the use of animals" of the School of Veterinary Medicine and Animal Science of University of São Paulo and was approved in the meeting of day 06/22/2011.

São Paulo, 27 de junho de 2011.

Profa. Dra. Denise Tabacchi Fantoni
Presidente

FOLHA DE AVALIAÇÃO

Nome: CIPULLO, Rafael Ishibashi

Título: Comparação e análise da rede de movimento de bovinos de propriedades positivas e negativas para brucelose no Estado de Mato Grosso

Dissertação apresentada ao Programa de Pós-Graduação em Epidemiologia Experimental Aplicada às Zoonoses da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo, para obtenção do título de Mestre em Ciências

Data: ____/____/____

Banca Examinadora

Prof. Dr. _____

Instituição: _____ Julgamento: _____

Prof. Dr. _____

Instituição: _____ Julgamento: _____

Prof. Dr. _____

Instituição: _____ Julgamento: _____

DEDICATÓRIAS

Esta dissertação marca uma fase e uma passagem muito importante na minha vida, a qual ficará registrada para sempre. Por esse motivo gostaria de aproveitar a oportunidade de deixar registrado o meu sentimento por aquelas pessoas que fizeram de mim o que sou hoje. Dedico este momento:

Aos meus pais, William e Marta, pelo esforço e dedicação para me proporcionar a melhor educação possível, pelo apoio incondicional e pelo exemplo de como se viver.

Ao restante da minha família, meu irmão Bruno, meus tios e tias, meus primos e primas pela ajuda e amizade.

À Dea, meu grande amor, pelo companheirismo, amizade e amor em todas as situações, alegres ou tristes, durante não só o mestrado, mas durante toda a nossa união. Juntos, construiremos ainda mais capítulos da nossa vida. Amo você! E gostaria de dedicar e agradecer também à sua família (Marcos, Lúcia e Luciana) a minha segunda casa, obrigado por tudo.

MEUS ETERNOS AGRADECIMENTOS...

AGRADECIMENTOS

Muitas pessoas contribuíram, direta ou indiretamente, com a realização desse trabalho e desde já agradeço a todas.

Em particular gostaria de agradecer:

Ao Prof. Dr. Marcos Amaku, excepcional orientador, pelos ensinamentos, amizade e sua dedicação ao trabalho que inspira todos ao seu redor.

Ao Prof. Dr. Ricardo Augusto Dias, pelos ensinamentos e apoio desde o tempo da iniciação científica.

Aos companheiros do LEB, Fernando Marques, Zé Grisi e Raul pela ajuda na execução deste trabalho, ideias e sugestões. A todos os colegas do LEB, Guilherme Braga, Gina, Oswaldo... pelo convívio de fundamental importância durante estes anos.

À FAPESP, pelo apoio financeiro e o incentivo na área de pesquisa acadêmica.

A todos que de qualquer forma ajudaram e participaram na concretização deste trabalho.

MEU MUITO OBRIGADO...

Este trabalho contou com o apoio financeiro da

Fundação de Amparo à Pesquisa do Estado de São Paulo
Processo n. 2011/03732-7

RESUMO

CIPULLO, R. I. **Comparação e análise da rede de movimento de bovinos de propriedades positivas e negativas para brucelose no Estado de Mato Grosso.** [Comparison and analysis of the network of cattle movement from positive and negative holdings to brucellosis in the State of Mato Grosso]. 2013. 59 f. Dissertação (Mestrado em Ciências) – Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2013.

Uma análise da rede de movimentação de bovinos entre estabelecimentos do Estado de Mato Grosso foi realizada utilizando os dados das guias de trânsito animal (GTA) emitidas durante o ano de 2007. A utilização de parâmetros descritivos da rede de movimentação de bovinos permitiu a caracterização da intensidade de comercialização de bovinos entre os estabelecimentos. Os parâmetros calculados foram: grau (entrada e saída), *betweenness*, *closeness*, coeficiente de aglomeração e *PageRank*. Dados referentes ao status das propriedades (foco ou livre de brucelose) obtidos do banco de dados do levantamento para brucelose bovina do Programa Nacional de Controle e Erradicação da Brucelose e da Tuberculose Animal (PNCEBT) foram utilizados para comparar os parâmetros das propriedades positivas e negativas para brucelose. A comparação foi realizada pelo teste de Mann Whitney e foi encontrada diferença estatisticamente significativa ($P < 0,05$) apenas para os parâmetros grau total e de saída ponderados por número de animais e ponderados por número de lotes movimentados. A análise descritiva da rede de movimentação de bovinos mostra que a distribuição de valores para todos os parâmetros, exceto o *closeness*, seguiu qualitativamente a Lei de potências. As informações obtidas são importantes para auxiliar em medidas de controle e prevenção de doenças em rebanhos bovinos, como a brucelose.

Palavras-chave: Bovinos. Epidemiologia. Redes Complexas. Mato Grosso. Movimentação Animal.

ABSTRACT

CIPULLO, R. I. **Comparison and analysis of the network of cattle movement from positive and negative holdings to brucellosis in the State of Mato Grosso.** [Comparação e análise da rede de movimento de bovinos de propriedades positivas e negativas para brucelose no Estado de Mato Grosso.]. 2013. 59 f. Dissertação (Mestrado em Ciências) – Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2013.

An analysis of the network of cattle movement between farm premises of the state of Mato Grosso was carried out using data from the records of animal transit (GTA) collected during 2007. The use of descriptive parameters of the network of movement allowed the characterization of the intensity of bovine trade between the farm premises. The parameters calculated were: degree (incoming and outgoing), betweenness, closeness, clustering coefficient and *PageRank*. The database concerning the status of the premises (positive or brucellosis free) obtained from the survey for bovine brucellosis of National Programme for Control and Eradication of Brucellosis and Tuberculosis Animal (PNCEBT) was used to compare the parameters of positive and negative premises for brucellosis. The comparison was done by using the Mann Whitney test and found a statistically significant difference ($P < 0.05$) for the parameters total degree and outgoing degree weighted by number of animals and number of batches. The descriptive analysis of the network of bovine movements showed that the distribution of values for all parameters, except the closeness, qualitatively followed a power law. The information obtained is important to assist in control measures and prevention of diseases in cattle herds, such as brucellosis.

Keywords: Bovine. Epidemiology. Complex networks. Mato Grosso. Animal movement.

SUMÁRIO

1 INTRODUÇÃO	14
1.1 PARÂMETROS DESCRITIVOS DAS REDES	15
1.1.1 Grau	16
1.1.2 <i>Betweenness</i>	18
1.1.3 <i>Closeness</i>	21
1.1.4 Coeficiente de Aglomeração	23
1.1.5 <i>PageRank</i>	24
1.2 APLICAÇÃO DA METODOLOGIA DE REDES NA VETERINÁRIA.....	30
2 JUSTIFICATIVA E OBJETIVO	31
3 MATERIAIS E MÉTODOS	31
3.1 ANÁLISE DESCRITIVA DO EFETIVO BOVINO	32
3.2 ANÁLISE DOS PARÂMETROS DA REDE.....	32
3.2.1 Grau	33
3.2.2 <i>Betweenness</i>	34
3.2.3 <i>Closeness</i>	34
3.2.4 Coeficiente de Aglomeração	35
3.2.5 <i>PageRank</i>	35
3.3 COMPARAÇÃO DOS PARÂMETROS EM RELAÇÃO AO STATUS DAS PROPRIEDADES.....	36
4 RESULTADOS E DISCUSSÃO	36
4.1 DISTRIBUIÇÃO DO EFETIVO BOVINO	38
4.2 ANÁLISE DESCRITIVA DA REDE.....	41
4.3 COMPARAÇÃO DOS PARÂMETROS EM RELAÇÃO AO STATUS DAS PROPRIEDADES.....	49
5 CONCLUSÕES	53
REFERÊNCIAS.....	54
GLOSSÁRIO.....	57
APÊNDICE A – CÓDIGO ELABORADO EM R PARA O CÁLCULO DO PARÂMETRO COEFICIENTE DE AGLOMERAÇÃO	58
APÊNDICE B – CÓDIGO ELABORADO EM R PARA O CÁLCULO DA MATRIZ <i>H</i> E DO PARÂMETRO <i>PAGERANK</i>.....	59

1 INTRODUÇÃO

As redes complexas estão presentes em muitos aspectos do nosso dia a dia e podem ser representadas por grafos, que são conjuntos de arestas ligando alguns ou todos os nós (Figura 1).

Figura 1 – Ilustração de um grafo direcionado

Fonte: (CIPULLO, 2013).

Um exemplo de grafo é a Internet, na qual os computadores podem ser representados por nós e os cabos entre eles representados por arestas, formando um grafo que expressa a relação entre os diferentes nós dessa rede. Nos casos em que as arestas seguem uma direção específica (i.e. do nó A para o nó B), então as redes são representadas por grafos direcionados (CALDARELLI, 2007). Utilizando conceitos e propriedades similares aos descritos anteriormente, as redes complexas podem ser empregadas em muitas outras situações, inclusive na área da epidemiologia veterinária.

Na epidemiologia veterinária, as redes complexas são importantes para explicar a transmissão das doenças nos rebanhos através do contato entre os animais. O conhecimento do padrão de movimentação dos animais proporciona informações importantes sobre a transmissão das doenças (BIGRAS-POULIN et al., 2006).

A transmissão de doenças em rebanhos bovinos envolve diversos fatores que juntos irão determinar a sua dispersão em maior ou menor grau. O potencial de dispersão de algumas doenças pode ser melhor explicado através da análise da rede de movimentação dos bovinos, que na realidade é apenas um dos

componentes de uma rede maior que inclui a movimentação de veterinários, caminhões de insumos e demais atores da cadeia agroprodutiva.

No caso das redes envolvendo a movimentação de bovinos, há dois níveis de organização que podem ser considerados na transmissão de doenças: movimentações dentro da propriedade e as movimentações entre as propriedades. A utilização da metodologia de redes para explicar a transmissão de doenças dentro da propriedade é uma abordagem interessante, porém não foi tratada nesse estudo.

O conceito de redes aplicado a esse estudo foi definido de modo que os nós (vértices) são representados pelas propriedades de bovinos, abatedouros e demais estabelecimentos enquanto as arestas são representadas pelas movimentações dos bovinos entre os nós.

O número de nós e de arestas são duas características básicas de uma rede (grafo). Para uma análise mais aprofundada, existem outros parâmetros descritivos que foram detalhados a seguir.

1.1 PARÂMETROS DESCRITIVOS DAS REDES

Os grafos podem ser representados por matrizes de adjacência (A) que são simplesmente tabelas de números em que a intersecção de cada coluna e cada linha representa uma aresta do grafo. Os valores de A_{ij} iguais a 1 indicam a existência de uma aresta ligando o nó i -ésimo ao nó j -ésimo, e valores iguais a 0 indicam a inexistência de ligação entre esses nós. A figura 2 mostra um exemplo de matriz de adjacência.

Figura 2 – Exemplo de uma matriz de adjacência e seu grafo direcionado

	A	B	C	D
A	0	0	0	1
B	1	0	0	0
C	0	0	0	0
D	0	1	1	0

Fonte: (CIPULLO, 2013).

A matriz de adjacência A apresenta nas linhas a origem das arestas e nas colunas os destinos das arestas e, por se tratar de uma representação de um grafo direcionado, a matriz não é simétrica ($A_{ij} \neq A_{ji}$). Uma forma de calcular o número máximo de arestas em um grafo direcionado e sem “laço” (i.e. sem arestas que possuem a mesma origem e destino) é simples e pode ser feita através da fórmula: $m_{\text{máx}} = n(n - 1)$ (CALDARELLI, 2007).

Os parâmetros calculados para o presente estudo foram: grau, centralidade por intermediação (*betweenness centrality*)¹, centralidades por proximidade (*closeness centrality*)¹, coeficiente de aglomeração e *PageRank*¹.

1.1.1 Grau

O grau representado por k é um parâmetro simples de ser calculado e é também uma das principais medidas de centralidade da rede. A definição tradicional de grau é o número de arestas conectadas a um determinado nó (CALDARELLI, 2007). Em um grafo direcionado, cada nó apresenta dois graus: um grau de entrada (k_{in}) e um grau de saída (k_{out}) que são, respectivamente, os números de arestas que chegam ao nó e que partem dele.

Para um grafo não-direcionado, o grau k_i de um nó i é calculado somando a linha i ou a coluna i da matriz de adjacência $A(n,n)$, e pode ser definido como (CALDARELLI, 2007):

$$k_i = \sum_{j=1,n} a_{ji}$$

Fonte: (CALDARELLI, 2007).

¹ A referência aos parâmetros *betweenness*, *closeness* e *PageRank* será feita considerando os termos em inglês por preferência do autor.

Para um grafo direcionado, o grau de entrada (k_{in}) e o grau de saída (k_{out}) são calculados somando a coluna i e linha i respectivamente, e podem ser definidos como (CALDARELLI, 2007):

$$k_i^{in} = \sum_{j=1,n} a_{ji} \qquad k_i^{out} = \sum_{j=1,n} a_{ij}$$

Fonte: (CALDARELLI, 2007).

O grau total de um grafo direcionado é a soma do grau de entrada e de saída (COSTA et al., 2007).

Para a análise do grau em uma rede de movimentação de bovinos, o nó pode ser representado por um estabelecimento (abatedouro, fazenda, outros) e as arestas podem ser definidas de 3 formas, resultando em 3 classificações distintas de grau:

- a) Grau ponderado por animais movimentados: a quantidade de bovinos comprados (s_{in}) e vendidos (s_{out}) pelo estabelecimento,
- b) Grau ponderado por lotes movimentados: a quantidade de lotes comprados (s_{in}) e vendidos (s_{out}) pelo estabelecimento,
- c) Grau por vizinhos: o número de estabelecimentos com que comercializou (k_{in} e k_{out}).

O grau ponderado é também chamado de *strength* (força) por alguns autores (COSTA et al., 2007), o que explica a notação s utilizada.

A figura 3 exemplifica uma rede fictícia em que o estabelecimento “A” comercializa com outros 4 estabelecimentos a quantidade de bovinos que está indicada nas setas.

Figura 3 – Rede fictícia de movimentação de bovinos do estabelecimento “A”

Fonte: (CIPULLO, 2013).

A tabela 1 apresenta os valores do grau do estabelecimento A, da rede fictícia da figura 3.

Tabela 1 – O grau do estabelecimento “A” por animais movimentados, por lotes movimentados e por fazendas vizinhas

GRAU DO ESTABELECIMENTO “A”		
Por animais movimentados	Por lotes movimentados	Por estabelecimentos vizinhos
$s = 550$	$s = 5$	$k = 4$
$s_{in} = 450$	$s_{in} = 4$	$k_{in} = 3$
$s_{out} = 100$	$s_{out} = 1$	$k_{out} = 1$

Considerando uma rede de movimentação animal, o grau de uma propriedade irá expressar o quanto ela está conectada com a rede e o quanto de animais e lotes ela movimenta na rede, mostrando, portanto, a força de sua conexão na rede.

1.1.2 *Betweenness*

O *betweenness* de um nó é dado pelo total de geodésicas (i.e distâncias mais curtas entre dois nós na rede) entre pares de outros nós que incluem o nó de referência (FREEMAN, 1977) e indica a importância do nó em ligar duas regiões distintas da rede.

Representando matematicamente o *betweenness*, n_{st}^i do nó i será igual a 1 quando a geodésica entre s e t passar por i , e será igual a zero quando isso não ocorrer. Portanto o *betweenness* é dado por (NEWMAN, 2010):

$$B_i = \sum_{st} n_{st}^i$$

Fonte: (NEWMAN, 2010).

Para exemplificar o cálculo do *betweenness*, considere a rede direcionada fictícia da figura 4.

Figura 4 – Rede direcionada fictícia de 8 nós

Fonte: (CIPULLO, 2013).

Essa rede fictícia possui 8 nós e os únicos que têm *betweenness* diferente de zero são os nós 2, 4 e 7, já que apenas eles estão situados entre geodésicas de outros dois nós. É importante ressaltar que embora o nó 1 esteja situado entre os nós 2 e 3, seu *betweenness* é igual a zero, pois a geodésica entre esses dois nós é uma aresta “direta” ligando o nó 3 ao 2.

Na figura 5, é possível visualizar passo a passo a maneira como o *betweenness* é calculado e todas as geodésicas (com distâncias maiores que 1) possíveis da rede fictícia apresentada na figura 4.

Figura 5 – As onze geodésicas possíveis compostas por suas arestas (setas em vermelho) e a variação da cor dos nós representando o valor do *betweenness* (de 0 a 8)

Fonte: (CIPULLO, 2013).

Em uma rede de movimentação animal, o *betweenness* de uma propriedade indica a sua importância como “ponte” de ligação entre duas outras propriedades, ou entre dois grupos distintos de propriedades. Uma propriedade com alto *betweenness* tem uma grande influência no controle do fluxo de movimentação animal da rede, pois é muito provável que uma grande quantidade de movimentações passe por essa propriedade.

1.1.3 Closeness

O *closeness* mede o quanto um determinado nó está próximo dos demais nós da rede. Segundo Freeman (1979), a forma mais simples de cálculo dessa medida de centralidade é definida matematicamente como o inverso da somatória das geodésicas entre o nó de referência e os demais da rede e foi proposta de forma que o *closeness* assuma valores mais altos para aqueles nós que possuam as geodésicas mais curtas entre os demais nós da rede.

$$C_i = \frac{1}{\sum_{i \neq j} d_{ij}}$$

Fonte: (FREEMAN, 1979).

onde C_i é a medida *closeness* do nó i e d_{ij} representa a distância da geodésica do nó i ao j . Caso não exista a geodésica do nó i ao j , pode-se utilizar o número total de nós da rede (n) em vez do valor da geodésica (d) (CSARDI; NEPUSZ, 2006).

Imagine uma rede baseada em comunicação na qual os nós representam as pessoas e as arestas são os caminhos percorridos pelas mensagens enviadas. As pessoas com os maiores valores de *closeness* serão as mais eficientes em sua comunicação, pois não possuem muitos “intermediários” em suas conexões fazendo com que tenham menor custo ou tempo para se comunicar com outras pessoas.

O valor do *closeness* é dependente do número de nós presentes na rede, portanto a comparação entre redes diferentes só poderá ser realizada caso seja utilizada a padronização sugerida por Beauchamp (1965), que se resume em multiplicar a equação tradicional do *closeness* por $n-1$.

$$C'_i = \frac{n-1}{\sum_{i \neq j} d_{ij}}$$

Fonte: (BEAUCHAMP, 1965).

Admitindo que um nó i conectado diretamente a todos os outros nós o valor da somatória das geodésicas será o menor possível e igual a $n-1$, após a padronização o valor de C'_i será igual a 1 e todos os valores dos demais nós estarão contidos em um intervalo de zero a 1.

Para exemplificar o cálculo do *closeness*, considere a rede direcionada fictícia da figura 6 e sua matriz de geodésicas da tabela 2.

Figura 6 – Rede direcionada fictícia de 11 nós

Fonte: (CIPULLO, 2013).

Tabela 2 – Matriz das geodésicas entre os nós, a somatória das geodésicas e os valores do *closeness* (C) e *closeness* padronizado (C').

Matriz das geodésicas entre os nós													$\sum$ geodésicas	C	C'
	1	2	3	4	5	6	7	8	9	10	11				
1	-	1	1	2	3	4	5	0	0	0	0		16	0,016666667	0,16666667
2	1	-	1	1	2	3	4	0	0	0	0		12	0,017857143	0,17857143
3	1	1	-	2	3	4	5	0	0	0	0		16	0,016666667	0,16666667
4	4	4	3	-	1	2	3	0	0	0	0		17	0,016393443	0,16393443
5	3	3	2	4	-	1	2	0	0	0	0		15	0,016949153	0,16949153
6	2	2	1	3	4	-	1	0	0	0	0		13	0,017543860	0,17543860
7	0	0	0	0	0	0	-	0	0	0	0		0	0,009090909	0,09090909
8	0	0	0	0	0	0	0	-	0	0	0		0	0,009090909	0,09090909
9	0	0	0	0	0	0	0	0	-	1	2		3	0,010989011	0,10989011
10	0	0	0	0	0	0	0	0	0	-	1		1	0,010000000	0,10000000
11	0	0	0	0	0	0	0	0	0	0	-		0	0,009090909	0,09090909

Fonte: (CIPULLO, 2013).

Em uma rede de movimentação animal, o *closeness* de uma propriedade indica a oportunidade de comercializar com as outras propriedades da rede.

1.1.4 Coeficiente de Aglomeração

Para redes não-direcionadas a definição do coeficiente de aglomeração (CA) se refere ao número de triângulos formados na rede (COHEN; HAVLIN, 2010) ou seja é a medida de quanto os nós de uma rede estão unidos em grupos de três.

Segundo Caldarelli (2007) o cálculo do coeficiente de aglomeração é dado por:

$$CA_i = \frac{1}{(k_i)(k_i - 1)/2} \sum_{j,k} a_{ij} a_{ik} a_{jk}$$

Fonte: (CALDARELLI, 2007).

O valor do coeficiente de aglomeração é alto quando dois nós que compartilham um vizinho em comum possuem alta probabilidade de estarem conectados entre si (COHEN; HAVLIN, 2010).

Em redes direcionadas, à semelhança do grau, podem ser calculados o coeficiente de aglomeração de entrada (CA_{in}) e o coeficiente de aglomeração de saída (CA_{out}). Esse cálculo é feito utilizando o grau por vizinhos e a quantidade de triângulos formados. O valor do coeficiente de aglomeração varia de zero a 1, sendo que, quando atinge o valor 1, todos os vizinhos desse nó possuem conexões entre si (em ambos os sentidos).

Para um nó i com grau k_i maior que 1, o coeficiente de aglomeração entrada (CA_{in}) e o coeficiente de aglomeração de saída (CA_{out}) são definidos respectivamente por (CALDARELLI, 2007):

$$CA_i^{in} = \frac{1}{(k_i^{in})(k_i^{in} - 1)/2} \sum_{j,k} a_{ij} a_{ki} \frac{a_{jk} + a_{kj}}{2}$$

Fonte: (CALDARELLI, 2007).

$$CA_i^{out} = \frac{1}{(k_i^{out})(k_i^{out} - 1)/2} \sum_{j,k} a_{ij} a_{ki} \frac{a_{jk} + a_{kj}}{2}$$

Fonte: (CALDARELLI, 2007).

Considere a figura 7 que representa uma rede direcionada fictícia de 9 nós e a tabela 3 com os valores do coeficiente de aglomeração.

Figura 7 – Rede direcionada fictícia de 9 nós

Fonte: (CIPULLO, 2013).

Tabela 3 – Valores do coeficiente de aglomeração de entrada (CA_{in}) e de saída (CA_{out}) da rede fictícia da figura 6.

Nós	CA_{in}	CA_{out}
1	0	0
2	0	$2/6 = 0,33$
3	0	0
4	$3/12 = 0,25$	0
5	0	0
6	0	$1/2 = 0,5$
7	$1/6 = 0,167$	0
8	0	0
9	0	0

Sendo o mais simplista possível, o cálculo do coeficiente de aglomeração tanto de entrada (CA_{in}) como de saída (CA_{out}) pode ser feito como uma relação de triângulos formados sobre os triângulos possíveis que esse nó poderia formar. Desta maneira, considerando o cálculo do coeficiente de aglomeração de saída (CA_{out}) do nó 2 da rede da figura 7 é $2/6$, pois ele formou 2 triângulos com arestas que saiam dele (1º triângulo: $2 \rightarrow 3$, $2 \rightarrow 4$, $3 \rightarrow 4$ e 2º triângulo: $2 \rightarrow 4$, $2 \rightarrow 7$, $7 \rightarrow 4$) e poderia formar até 6 triângulos com as arestas que saem dele (1º triângulo: $2 \rightarrow 3$, $2 \rightarrow 4$, $3 \rightarrow 4$; 2º triângulo: $2 \rightarrow 3$, $2 \rightarrow 4$, $4 \rightarrow 3$; 3º triângulo: $2 \rightarrow 4$, $2 \rightarrow 7$, $7 \rightarrow 4$; 4º triângulo: $2 \rightarrow 4$, $2 \rightarrow 7$, $4 \rightarrow 7$; 5º triângulo: $2 \rightarrow 3$, $2 \rightarrow 7$, $3 \rightarrow 7$ e 6º triângulo: $2 \rightarrow 3$, $2 \rightarrow 7$, $7 \rightarrow 3$).

1.1.5 PageRank

O *PageRank* é um algoritmo de análise de conexões em redes utilizado no mecanismo de busca da Internet do *Google* (BRIN; PAGE, 1998). Esse algoritmo atribui um valor da importância relativa de cada nó da rede (valor de *PageRank*), considerando a quantidade de arestas que determinado nó recebe (grau de entrada)

e também a quantidade de arestas que seus vizinhos recebem. Além disso, é uma medida propagada pela rede que se baseia na premissa de que nós importantes recebem ligações de nós importantes, como mostra a figura 8.

Figura 8 – Rede fictícia exemplificando o cálculo do *PageRank*

Legenda: Os nós A, B e C recebem o mesmo número de ligações (i.e. grau de entrada igual a 1). No entanto, o nó C possui um valor de *PageRank* maior que A e B, pois a ligação que C recebe provém de um nó importante (nó D, com maior grau de entrada), elevando assim seu *PageRank*.

O *PageRank* é dado por (NEWMAN, 2010):

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j^{out}} + \beta$$

Fonte: (NEWMAN, 2010).

onde α e β são constantes positivas e x representa o valor de *PageRank*.

Didaticamente, podemos imaginar o cálculo do *PageRank* como a dinâmica de um volume fixo de fluido que circula pela rede passando pelos nós através de suas arestas e acumulando em nós de maior importância da rede (EASLEY; KLEINBERG, 2010), de modo que esse fluido não é criado e nem destruído ao longo de seu cálculo.

O cálculo do *PageRank* é realizado em uma série de passos sendo que o valor inicial que cada nó recebe é igual a $1/n$ (admitindo uma rede com n nós) e o valor da soma do valor de *PageRank* de todos os nós sempre será igual a 1.

Considere a rede da figura 9 e a tabela 4 com seus respectivos passos para o cálculo do *PageRank*.

Figura 9 – Rede direcionada fictícia de 8 nós

Fonte: (figura adaptada de EASLEY; KLEINBERG, 2010).

Tabela 4 – Valores do *PageRank* a cada passo do cálculo da rede fictícia da figura 9

Nº passos	A	B	C	D	E	F	G	H
-	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
1º	1/2	1/16	1/16	1/16	1/16	1/16	1/16	1/8
2º	5/16	1/4	1/4	1/32	1/32	1/32	1/32	1/16
⋮								
60º	207/673	2/13	2/13	1/13	1/13	1/13	1/13	1/13
61º	4/13	2/13	2/13	1/13	1/13	1/13	1/13	1/13
⋮								
*kº	4/13	2/13	2/13	1/13	1/13	1/13	1/13	1/13

Legenda: *infinitos passos.

Fonte: (tabela adaptada de EASLEY; KLEINBERG, 2010).

Inicialmente, todos os nós da rede possuem o mesmo valor de *PageRank*: 1/8 (sendo 8 o número de nós da rede). Após o primeiro passo, o nó A assumiu o valor de 1/2, pois ele recebeu dos nós F, G e H o total do valor do *PageRank* de cada um e recebeu a metade dos valores do *PageRank* dos nós D e E. Já os nós B e C receberam, cada um, a metade do valor do *PageRank* do nó A. Portanto o valor do *PageRank* do nó A ao final do 1º passo foi calculado da seguinte maneira:

$$\frac{1}{8} (\text{do nó } F) + \frac{1}{8} (\text{do nó } G) + \frac{1}{8} (\text{do nó } H) + \frac{1}{16} (\text{do nó } D) + \frac{1}{16} (\text{do nó } E) = \frac{1}{2}$$

Para cada passo do cálculo do *PageRank* é utilizada a matriz H (*row normalized hyperlink matrix*) que representa a porção do valor do *PageRank* que cada nó passa aos seus nós vizinhos. Essa matriz é construída a partir do grau de saída de cada nó, desta maneira:

Os elementos da matriz H são definidos da seguinte forma:

- quando o nó i não possuir aresta de saída com o nó $j \rightarrow H_{i,j} = 0$;
- se o nó i possuir aresta de saída com o nó $j \rightarrow H_{i,j} = \frac{1}{k_{out}}$.

A matriz H da rede da figura 9 está representada na figura 10.

Figura 10 – Matriz H usada para cada passo do cálculo do *PageRank* da rede da figura 9

	A	B	C	D	E	F	G	H
A	0	0,5	0,5	0	0	0	0	0
B	0	0	0	0,5	0,5	0	0	0
C	0	0	0	0	0	0,5	0,5	0
D	0,5	0	0	0	0	0	0	0,5
E	0,5	0	0	0	0	0	0	0,5
F	1	0	0	0	0	0	0	0
G	1	0	0	0	0	0	0	0
H	1	0	0	0	0	0	0	0

Fonte: (CIPULLO, 2013).

Ao final do cálculo do *PageRank*, os nós B e C apresentaram um valor de *PageRank* superior ao nó H que possui um grau de entrada (igual a 2) maior que os nós B e C , mostrando que, pelo fato deles receberem arestas do nó A (o mais importante da rede), beneficiaram-se e tiveram o valor de seu *PageRank* aumentado.

Pode-se notar pela tabela 4 que os valores do *PageRank* de cada nó atingiram um equilíbrio a partir do 62º passo, e se repetiram até o k -ésimo passo. Portanto, admite-se que o cálculo do valor do *PageRank* converge a um valor limite

quando nos passos seguintes os valores de todos os nós não se alteram mais. Em redes muito grandes, o valor final do *PageRank* é atingido quando o número de passo tende ao infinito.

Em redes com nós isolados em componentes que só recebem arestas e não enviam arestas de volta aos demais nós da rede, ocorre um problema com a definição básica do *PageRank* dada anteriormente. Por exemplo, se a rede analisada tivesse algumas diferenças em relação à rede da figura 9, em que os nós *F* e *G* ligassem um ao outro em vez de conectarem-se ao nó *A*. O resultado dessa nova rede está na figura 11.

Figura 11 – Rede direcionada fictícia de 8 nós com características que causariam problema no cálculo da definição básica do *PageRank*

Fonte: (figura adaptada de EASLEY; KLEINBERG, 2010).

Nessa nova rede, o nó *A* deixa de ser tão importante quanto era anteriormente. Porém, além disso, o *PageRank* que passa do nó *C* para os nós *F* e *G* não circulará novamente na rede, fazendo com que todo o *PageRank* da rede seja dividido entre os nós *F* e *G* (1/2 para cada um).

Para contornar esse problema, é utilizado um parâmetro α na fórmula do algoritmo do *PageRank* que fornece a todos os nós da rede um valor de *PageRank* “residual”, não permitindo que todo o *PageRank* da rede acumule-se em determinados nós (e.g. nós *F* e *G* da rede da Figura 11).

Utilizando a analogia feita por Easley e Kleinberg (2010), podemos admitir, de forma simplista, que toda a água da Terra não escoar pelos desníveis e acumula nos pontos de menor altitude por causa de um processo de equilíbrio de evaporação e chuva, fazendo com que a água retorne aos pontos mais altos novamente. Esse processo de equilíbrio no cálculo do *PageRank* é feito pelo parâmetro α que pode variar de zero a 1. O mecanismo de busca do *Google* utiliza um valor de α de 0,85 em seu cálculo, apesar de não estar claro de que haja uma justificativa analítica para a escolha desse determinado valor, aparentemente é a alternativa que mostrou uma melhor otimização do cálculo (NEWMAN, 2010).

O cálculo utilizando esse parâmetro multiplica por α os valores de *PageRank* de todos os nós, portanto a somatória do *PageRank* da rede passa a ser α e não mais 1. Após isso, é adicionado a cada nó um valor de *PageRank* igual a $\frac{1-\alpha}{n}$ (sendo n o número de nós da rede) fazendo com que a somatória do *PageRank* da rede volte a ser igual a 1. A tabela 5 mostra os passos para o cálculo dos valores de *PageRank* da rede mostrada na figura 11.

Tabela 5 – Valores do *PageRank* a cada passo do cálculo da rede fictícia da figura 11 utilizando um α de 0,85

Nº passos	A	B	C	D	E	F	G	H
-	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
1º	71/160	23/320	23/320	23/320	23/320	23/320	23/320	1/8
2º	201/652	96/463	96/463	7/142	7/142	7/142	7/142	61/764
⋮								
*kº	3/28	9/140	9/140	17/369	17/369	168/547	168/547	41/708

Legenda: *infinitos passos.

Fonte: (tabela adaptada de EASLEY; KLEINBERG, 2010).

O valor do *PageRank* do nó *B*, por exemplo, ao final do 1º passo foi calculado da seguinte maneira:

$$\left[\frac{1}{16} \text{ (do nó A)} \times 0,85 \right] + \frac{1 - 0,85}{8} = \frac{23}{320}$$

1.2 APLICAÇÃO DA METODOLOGIA DE REDES NA VETERINÁRIA

A metodologia de análise de redes já é usada há muito tempo nas áreas da Sociologia, Psicologia, Antropologia e Biologia (MARTINEZ-LOPEZ; PEREZ; SANCHEZ-VIZCAINO, 2008), porém na Medicina Veterinária a sua aplicação é relativamente recente, sendo os primeiros estudos feitos por Corner, Pfeiffer e Morris (2003) e Christley et al. (2003). Atualmente a metodologia de análise de redes aplicada à Epidemiologia Veterinária vai além da simples descrição do perfil de movimentação de animais, sendo usada principalmente em modelagem matemática para verificar o espalhamento de doenças na rede, como subsídio para a análise baseada em risco, na detecção de circuitos agropecuários (GRISI-FILHO et al., 2012) e na compartimentação de zonas agropecuárias com diferentes status sanitários.

Com o aumento da preocupação em relação à sanidade dos animais de produção, novas aplicabilidades da metodologia de análise de redes vêm sendo desenvolvidas tanto para a prevenção das doenças como também para ações no controle de focos de doenças. Com isso o uso da metodologia de redes na Epidemiologia Veterinária passa a ser cada vez mais importante e essencial.

As redes formadas por movimentações animais, em especial bovinos, apresentam algumas características próprias, entre elas: são direcionadas, dinâmicas, apresentam alto grau de heterogeneidade (BIGRAS-POULIN et al., 2006, 2007; NATALE et al., 2009) e podem ser descritas pela Lei de potências (NEGREIROS, 2010). Portanto em estudos realizados com redes de movimentação animal deve-se ter o cuidado de considerar essas características nas análises.

2 JUSTIFICATIVA E OBJETIVO

A análise descritiva dos diversos parâmetros da rede do Mato Grosso (MT) é importante para conhecer o perfil de movimentação dos bovinos e entender a relação entre os estabelecimentos de produção deste Estado. O objetivo do presente trabalho consiste em realizar a caracterização da intensidade de comercialização de bovinos entre os estabelecimentos e realizar uma comparação dos parâmetros da rede entre propriedades positivas e negativas para brucelose bovina, já que a movimentação de bovinos é um fator preponderante no espalhamento da doença (BEER, 1999). Deste modo, essas análises podem auxiliar tanto em estratégias para prevenção e controle da brucelose como também em estudos futuros envolvendo a brucelose no Mato Grosso.

3 MATERIAIS E MÉTODOS

As informações sobre o status das propriedades (foco ou livre da brucelose) foram obtidas do banco de dados do levantamento para brucelose bovina do Programa Nacional de Controle e Erradicação da Brucelose e da Tuberculose Animal (PNCEBT) realizado no ano de 2003.

Os dados das movimentações de bovinos utilizados no presente trabalho foram obtidos a partir das Guias de Trânsito Animal (GTA) emitidas ou recebidas pelo Instituto de Defesa Agropecuária do Estado de Mato Grosso (INDEA-MT) durante o ano de 2007.

As informações sobre a quantidade de bovinos em cada propriedade foram obtidas da análise de banco de dados de efetivo populacional cedido pelo INDEA-MT referentes aos exercícios de 2003, 2004, 2005, 2006 e 2007 (KURODA, 2012).

Para o início das análises no banco de dados da rede de movimentação de bovinos do Estado de Mato Grosso (MT), foi necessária a sua preparação extraíndo somente o trânsito de bovinos e excluindo as movimentações com origem e destino na mesma propriedade (i. e. laços). A rede apresentou um total de 88.862 nós (estabelecimentos) e 539.494 arestas entre os nós.

No levantamento epidemiológico realizado no Estado do MT pelo PNCEBT no ano de 2003, foi estimada uma prevalência de propriedades soropositivas para brucelose de 41,2% de um total de 1.115 propriedades consideradas na amostragem (NEGREIROS, 2006). A partir da identificação dessas propriedades na rede de movimentação animal do MT, foi possível obter informações sobre a comercialização de bovinos dessas propriedades no ano de 2007.

3.1 ANÁLISE DESCRITIVA DO EFETIVO BOVINO

Como um dos objetivos do trabalho era analisar as movimentações de bovinos realizadas durante o ano de 2007 pelas propriedades positivas e negativas para brucelose, realizou-se uma análise descritiva do efetivo bovino por propriedades ao longo dos anos de 2003, 2004, 2005, 2006 e 2007 com o intuito de observar se durante esses anos a distribuição do efetivo bovino por propriedades apresentaria um perfil semelhante, caracterizando um padrão na comercialização de bovinos ao longo dos anos.

A análise da distribuição do efetivo bovino foi realizada utilizando o programa Minitab[®] Statistical Software.

A seguir analisou-se a evolução do efetivo bovino das propriedades em que foi possível recuperar as informações da quantidade de bovinos nos anos de 2003 e 2007 (783 propriedades), observando se houve alteração na quantidade do efetivo bovino entre esses dois anos. Utilizando os quartis (1º quartil, mediana e 3º quartil) da distribuição dos efetivos nos anos 2003 e 2007 como pontos de cortes, foram definidas 4 categorias de propriedades para cada ano. Assim, calculou-se a diferença entre as categorias das propriedades no ano de 2003 e no ano de 2007.

3.2 ANÁLISE DOS PARÂMETROS DA REDE

Optou-se por utilizar o programa R (R Development Core Team, 2012) na etapa de análise dos parâmetros da rede, pois esta ferramenta é disponibilizada

gratuitamente pela *web* e possui código aberto, possibilitando a implementação de algoritmos específicos para a caracterização de redes. Além disso, já existem os pacotes *igraph* (CSARDI; NEPUSZ, 2006) e *statnet* (HANDCOOK et al., 2003) desenvolvidos para a análise de redes.

A análise descritiva dos parâmetros da rede foi realizada considerando todos os 88.862 estabelecimentos (nós) e suas 539.494 movimentações (arestas).

Além da análise descritiva dos parâmetros da rede, foi realizada uma outra análise para verificar se as distribuições dos parâmetros *betweenness*, *closeness*, coeficiente de aglomeração e *PageRank* da rede do MT seguem a Lei de Pareto. A Lei de Pareto ocorre quando 20% dos estabelecimentos com os maiores valores dos parâmetros analisados realizaram 80% das movimentações de bovinos. Já foi demonstrado que o parâmetro grau pôde ser ajustado pela Lei de Potências (NEGREIROS, 2010) e consequentemente seguiu a Lei de Pareto, que é uma de suas características.

Com a análise da Lei de Pareto, conhecida também como razão 80/20, foi possível demonstrar a presença de nós espalhadores de doença na rede do MT e quais os parâmetros mais adequados para visualizar esses nós.

3.2.1 Grau

O parâmetro grau, conforme mencionado anteriormente, pode ser classificado de três maneiras distintas: grau ponderado por animais movimentados, grau ponderado por lotes movimentados e grau por vizinhos. Para a análise desse parâmetro foram consideradas as três formas distintas de grau e suas variantes (grau total, grau de entrada e grau de saída).

O cálculo do grau por animais movimentados e o grau por lotes movimentados foi realizado utilizando a função *degree* do pacote *sna* (CARTER, 2010).

Os códigos utilizados no cálculo do grau por animais movimentados considerando o grau total, grau de entrada e grau de saída foram, respectivamente:

```
ktotal <- degree(dat, cmode="freeman")
```

```
kin <- degree(dat, cmode="indegree")
kout <- degree(dat, cmode="outdegree")
```

Os códigos utilizados no cálculo do grau por lotes movimentados considerando o grau total, grau de entrada e grau de saída foram, respectivamente:

```
ktotal_lote <- degree(dat, cmode="freeman", ignore.eval=TRUE)
kin_lote <- degree(dat, cmode="indegree", ignore.eval=TRUE)
kout_lote <- degree(dat, cmode="outdegree", ignore.eval=TRUE)
```

Sendo o objeto *dat* uma lista de adjacência criada nos padrões do pacote *sna*.

O cálculo do grau por vizinhos foi realizado utilizando a função *degree* do pacote *igraph* (CSARDI; NEPUSZ, 2006).

Os códigos utilizados no cálculo do grau por vizinhos considerando o grau total, grau de entrada e grau de saída foram, respectivamente:

```
ktotal_viz <- degree(graph, mode = "all")
kin_viz <- degree(graph, mode = "in")
kout_viz <- degree(graph, mode = "out")
```

Sendo o objeto *graph* um grafo criado nos padrões do pacote *igraph*.

3.2.2 Betweenness

O cálculo do parâmetro *betweenness* foi realizado utilizando a função *betweenness_w* do pacote *tnet* (OPSAHL, 2009).

O código utilizado no cálculo do *betweenness* foi:

```
bet <- betweenness_w(net, alpha = 0.5)
```

Sendo o objeto *net* uma lista de adjacência criada no formato de matriz.

3.2.3 Closeness

O cálculo do parâmetro *closeness* foi realizado utilizando a função *closeness* do pacote *igraph* (CSARDI; NEPUSZ, 2006).

O código utilizado no cálculo do *closeness* foi:

```
clos <- closeness(graph, mode = "out")
```

Sendo o objeto *graph* um grafo criado nos padrões do pacote *igraph*.

3.2.4 Coeficiente de Aglomeração

Como não foram encontrados pacotes com funções que permitissem o cálculo do coeficiente de aglomeração de uma rede direcionada, foi desenvolvida uma função no programa R que utilizou como base para o cálculo a matriz de adjacência da rede do MT. O *script* com a função utilizada no cálculo está apresentado no Apêndice A.

Os códigos utilizados no cálculo do coeficiente de aglomeração de entrada e saída foram, respectivamente:

```
cin <- clustering("in", matriz)
cout <- clustering("out", matriz)
```

Sendo o objeto *matriz* uma matriz de adjacência em que as linhas indicam origem e as colunas indicam destino das movimentações. Nessa matriz os valores das células devem ser 1 quando há aresta entre os dois nós e zero caso não haja aresta.

3.2.5 PageRank

Para o cálculo dos valores de *Pagerank* foi criado um script que utiliza a Matriz *H* (*row normalized hiperlink matrix*) apresentada no item 1.1.5.

O código com a função *rnhm* utilizado para calcular a matriz *H* foi:

```
H <- rnhm(matriz)
```

Sendo o objeto *matriz* uma matriz *n* por *n* preenchida por zeros e uns, em que as linhas representam a origem e as colunas representam o destino das movimentações.

O código com a função *pagerank* utilizado para calcular os valores de *PageRank* foi:

```
pr <- pagerank(H)
```

O *script* com as funções *rnhm* e *pagerank* está apresentado no Apêndice B.

3.3 COMPARAÇÃO DOS PARÂMETROS EM RELAÇÃO AO STATUS DAS PROPRIEDADES

A comparação dos parâmetros da rede do MT em relação ao status da propriedade (positiva ou negativa para brucelose) foi feita pelo teste não paramétrico Mann-Whitney para um nível de significância de 5% ($\alpha=0,05$).

Deve ser salientado que devido à necessidade de relacionar dois bancos de dados distintos (levantamento epidemiológico e movimentação animal), algumas informações sobre as propriedades amostradas não puderam ser utilizadas, sendo possível relacionar os dados de 641 propriedades (262 positivas e 379 negativas) das 1.115 amostradas no levantamento epidemiológico do PNCEBT.

4 RESULTADOS E DISCUSSÃO

O Estado do MT apresentou um fluxo muito intenso de movimentação com uma grande quantidade de bovinos movimentados durante o ano de 2007. Além

disso, detém o maior rebanho bovino do Brasil com cerca de 12,9% do efetivo total de bovinos do país (IBGE, 2007).

No ano de 2007, a maior parte das movimentações de bovinos realizadas foram com a finalidade de abate e engorda (Figura 12), mostrando a aptidão característica do Estado na produção de gado de corte.

Figura 12 – Gráfico de barras com a distribuição da movimentação de bovinos para cada Finalidade de Movimentação

Fonte: (CIPULLO, 2013).

4.1 DISTRIBUIÇÃO DO EFETIVO BOVINO

Os sumários gráficos da distribuição do efetivo bovino por propriedades ao longo dos anos de 2003, 2004, 2005, 2006 e 2007, respectivamente, apresentados nas figuras 13, 14, 15, 16 e 17, mostram que há um perfil semelhante entre os anos de modo que a média e os quartis pouco se alteraram durante o período analisado.

Figura 13 – Sumário gráfico da distribuição do efetivo bovino por propriedades no ano de 2003

Figura 14 – Sumário gráfico da distribuição do efetivo bovino por propriedades no ano de 2004

Figura 15 – Sumário gráfico da distribuição do efetivo bovino por propriedades no ano de 2005

Figura 16 – Sumário gráfico da distribuição do efetivo bovino por propriedades no ano de 2006.

Figura 17 – Sumário gráfico da distribuição do efetivo bovino por propriedades no ano de 2007.

Analizou-se também a evolução da quantidade de bovinos nos anos de 2003 e 2007 observando as mesmas propriedades nesses dois anos. Utilizando a média dos quartis (1º quartil, mediana e 3º quartil) das distribuições dos anos de 2003 a 2007 como pontos de cortes, foram definidas 4 categorias de propriedades para cada ano (Tabela 6).

Tabela 6 – Os pontos de cortes da distribuição dos efetivos nos anos 2003 e 2007 utilizados para definir as 4 categorias de propriedades

CATEGORIAS			
1	2	3	4
$\leq 62,9$	$> 62,9 \leq 163,6$	$> 163,6 \leq 522,2$	$> 522,2$

Fonte: (CIPULLO, 2013).

Calculou-se então a diferença entre as categorias das propriedades no ano de 2003 e no ano de 2007. Foi observado que 54,4% das propriedades (do 1º quartil ao 3º quartil na Figura 18 A) mantiveram a mesma classificação de categoria (ou seja, diferença igual a zero) e que menos de 10% (9,6%) das propriedades alteraram sua classificação em pelo menos duas categorias para mais ou para menos (ou seja,

diferença ≥ 2 ou ≤ -2 respectivamente). A figura 18 B mostra a classificação das propriedades nos anos de 2003 e 2007.

Figura 18 – A: Sumário gráfico da diferença das categorias das propriedades no ano de 2003 e 2007. B: Classificação das propriedades nos anos de 2003 e 2007

Fonte: (CIPULLO, 2013).

Portanto, a partir das análises da distribuição e evolução do efetivo bovino de 2003 a 2007, é possível observar que a quantidade de bovinos na maior parte das propriedades é mantido, caracterizando um perfil potencialmente estável da comercialização de bovinos ao longo desses anos.

4.2 ANÁLISE DESCRITIVA DA REDE

A análise descritiva da rede é essencial para identificar e caracterizar o padrão de movimentação de bovinos entre todos os estabelecimentos da rede do MT, permitindo a avaliação do número de animais e lotes movimentados (grau), da existência de estabelecimentos de ligação que intermedeiam o fluxo de animais (*betweenness*), da distância geodésica entre os estabelecimentos (*closeness*), da formação de aglomerados de estabelecimentos (coeficiente de aglomeração) e da importância de determinados estabelecimentos (*PageRank*).

A análise descritiva da rede também permite identificar as características da estrutura de uma rede de movimentação animal, podendo ser definida como aleatória, mundo pequeno (*small world*) ou como livre de escala (*free-scale networks*), entre outras possibilidades. Essa identificação é importante, pois permite inferir comportamentos epidemiológicos feitos a partir de estudos teóricos realizados na rede.

Nas redes livres de escala, existe um padrão em que poucos nós possuem muitas conexões e a maior parte dos nós possui poucas conexões. Essas redes podem ser descritas pela Lei de potência (*power law*). Algumas redes de movimentação animal descritas em trabalhos anteriores seguiram a Lei de potência (BIGRAS-POULIN et al., 2006; NATALE et al., 2009; NEGREIROS, 2010). Como pode ser observado nas figuras 19 a 28, a rede de movimentação de bovinos do MT no ano de 2007 apresentou visualmente características da Lei de potência em todos os parâmetros, exceto o *closeness* que apresentou uma distribuição dos valores diferente da encontrada nos demais parâmetros.

As figuras 19 e 20 mostram, respectivamente, os histogramas da distribuição dos valores do grau de entrada e de saída por animais movimentados.

Figura 19 – Distribuição dos valores do grau de entrada por número de animais movimentados (bovinos comprados) sem limitar os eixos X e Y (gráfico maior) e limitando os eixos X e Y (gráfico menor)

Grau por animais movimentados

Fonte: (CIPULLO, 2013).

Figura 20 – Distribuição dos valores do grau de saída por número de animais movimentados (bovinos vendidos) sem limitar os eixos X e Y (gráfico maior) e limitando os eixos X e Y (gráfico menor)

Grau por animais movimentados

Fonte: (CIPULLO, 2013).

As figuras 21 e 22 mostram, respectivamente, os histogramas da distribuição dos valores do grau de entrada e de saída por lotes movimentados.

Figura 21 – Distribuição dos valores do grau de entrada por número de lotes movimentados (lotes comprados) sem limitar os eixos X e Y (gráfico maior) e limitando os eixos X e Y (gráfico menor)

Figura 22 – Distribuição dos valores do grau de entrada por número de lotes movimentados (lotes comprados) sem limitar os eixos X e Y (gráfico maior) e limitando os eixos X e Y (gráfico menor)

As figuras 23 e 24 mostram, respectivamente, os histogramas da distribuição dos valores do grau de entrada e de saída por vizinhos.

Figura 23 – Distribuição dos valores do grau de entrada por estabelecimentos vizinhos sem limitar os eixos X e Y (gráfico maior) e limitando os eixos X e Y (gráfico menor)

Grau por estabelecimentos "vizinhos"

Fonte: (CIPULLO, 2013).

Figura 24 – Distribuição dos valores do grau de saída por estabelecimentos vizinhos sem limitar os eixos X e Y (gráfico maior) e limitando os eixos X e Y (gráfico menor)

Grau por estabelecimentos "vizinhos"

Fonte: (CIPULLO, 2013).

As figuras 25 e 26 mostram, respectivamente, os histogramas da distribuição dos valores do *betweenness* e do *closeness*.

Figura 25 – Distribuição dos valores do *betweenness* sem limitar os eixos X e Y (gráfico maior) e limitando os eixos X e Y (gráfico menor)

Figura 26 – Distribuição dos valores normalizados do *closeness* sem limitar os eixos X e Y (gráfico maior), limitando os eixo X e Y para visualização dos valores em torno de 1,00 (gráfico menor à esquerda) e para visualização dos valores em torno de 1,36 (gráfico menor à direita)

As figuras 27 e 28 mostram, respectivamente, os histogramas da distribuição dos valores do coeficiente de aglomeração e do *PageRank*.

Figura 27 – Distribuição dos valores do coeficiente de aglomeração de entrada (gráfico esquerda) e de saída (gráfico direita)

Coeficiente de Aglomeração

Fonte: (CIPULLO, 2013).

Figura 28 – Distribuição dos valores do *PageRank* sem limitar os eixos X e Y (gráfico maior) e limitando os eixos X e Y (gráfico menor)

PageRank

Fonte: (CIPULLO, 2013).

É possível notar pela análise das figuras 19 a 28 que apenas o parâmetro *closeness* não apresentou visualmente características da Lei de potência, demonstrando no gráfico uma aparente predominância de 2 valores (1,00 e 1,36). O fato de o parâmetro *closeness* apresentar uma aparente predominância de 2 valores (1,00 e 1,36) ocorre devido à sua definição matemática e ao modo como foi calculado. Os nós que fazem parte do Grande Componente (*Giant Component* – GC) apresentaram valores próximos a 1,36 já os nós que não fazem parte do GC e aqueles que só possuem arestas de entrada (i. e. com grau de saída igual a zero) apresentaram valores próximos a 1,00 (Figura 26). Para facilitar a apresentação dos valores do *closeness* foi feita uma normalização, multiplicando os valores originários da fórmula do item 1.1.3 por n (sendo n o número de nós da rede).

Na tabela 7 é apresentado o sumário dos parâmetros da rede analisados.

Tabela 7 – Sumário dos parâmetros da rede

Sumário				
Parâmetros	Mínimo	Mediana	Média	Máximo
Grau por animais movimentados	0	53,0	359,7	380.313,0
Grau de entrada por animais movimentados	0	4,0	179,8	380.313,0
Grau de saída por animais movimentados	0	29,0	179,8	53.951,0
Grau por lotes movimentados	0	3,0	12,1	21.672,0
Grau de entrada por lotes movimentados	0	1,0	6,0	21.672,0
Grau de saída por lotes movimentados	0	2,0	6,0	1.585,0
Grau por "vizinhos"	0	2,0	4,6	1.522,0
Grau de entrada por "vizinhos"	0	1,0	2,3	1.522,0
Grau de saída por "vizinhos"	0	1,0	2,3	183,0
<i>Betweenness</i>	0	0	140.300	237.300.000,0
<i>Closeness</i>	1,0	1,361	1,201	1,364
Coeficiente de aglomeração de entrada	0	0	0,009	0,5
Coeficiente de aglomeração de saída	0	0	0,038	0,5
<i>PageRank</i>	$4,634 \times 10^{-6}$	$5,422 \times 10^{-6}$	$1,125 \times 10^{-5}$	$5,491 \times 10^{-3}$

Fonte: (CIPULLO, 2013).

Em estudos anteriores, foi demonstrado que a distribuição do parâmetro grau para a rede do MT pôde ser ajustada, em sua maior parte, pela Lei de potências (NEGREIROS, 2010) e consequentemente seguiu a razão 80/20, que é uma de suas características. Também conhecida como Lei de Pareto (BARABASI, 2003), a razão 80/20 em redes de movimentação animal pode ser evidenciada quando 80% das movimentações são realizadas por 20% dos estabelecimentos da rede.

No presente estudo, foi realizada uma análise para verificar se os demais parâmetros (*betweenness*, *closeness*, coeficiente de aglomeração e *PageRank*) da rede do MT seguem a Lei de Pareto, ou seja, se 20% dos estabelecimentos com os maiores valores dos parâmetros analisados realizaram 80% das movimentações de bovinos. Na tabela 8, está apresentado o resultado dessa análise.

Tabela 8 – Porcentagem da quantidade de animais movimentados (Grau total, Grau de entrada e Grau de saída) por 20% dos estabelecimentos com os maiores valores dos parâmetros analisados: *betweenness*, *closeness*, coeficiente de aglomeração (entrada e saída) e *PageRank*

Porcentagem de animais movimentados por 20% dos estabelecimentos com os maiores valores dos parâmetros			
Parâmetros	Grau total	Grau de entrada	Grau de saída
<i>Betweenness</i>	47,0	39,0	54,0
<i>Closeness</i>	12,0	4,0	19,0
<i>Coeficiente de Aglomeração de entrada</i>	58,0	74,0	* 0,0
<i>Coeficiente de Aglomeração de saída</i>	38,0	* 0,0	51,0
<i>PageRank</i>	66,0	83,0	49,0

* Os parâmetros coeficientes de aglomeração de entrada e saída não consideram respectivamente a saída e a entrada de animais para o cálculo Fonte: (CIPULLO, 2013).

É possível afirmar que o parâmetro *PageRank*, quando analisado em relação a movimentação de entrada de bovinos, foi o único que seguiu a Lei de Pareto, já que 83% das movimentações de entrada de bovinos foram realizadas por 20% dos estabelecimentos com os maiores valores de *PageRank*. Portanto, o parâmetro *PageRank* identifica não só a importância de cada nó na rede baseada em suas conexões, mas também indica a importância baseada na quantidade de animais adquiridos por cada estabelecimento (grau de entrada).

4.3 COMPARAÇÃO DOS PARÂMETROS EM RELAÇÃO AO STATUS DAS PROPRIEDADES

Os parâmetros das propriedades positivas e negativas para brucelose foram comparados pelo teste não-paramétrico de Mann-Whitney, já que como mostrado anteriormente as distribuições dos parâmetros apresentaram características da Lei de potência ou distribuições bem peculiares, como no caso do *closeness*.

A tabela 9 mostra o valor de *P* da análise pelo teste de Mann-Whitney, o 1º quartil, a mediana e o 3º quartil do parâmetro grau.

Tabela 9 – Valor de *P*, 1º quartil, mediana e 3º quartil dos valores do grau das propriedades positivas e negativas

Comparação do grau entre propriedades positivas (P) e negativas (N)					
Parâmetros		1º Quartil	Mediana	3º Quartil	Valor de <i>P</i>
Grau por animais movimentados	P	37,0	160,5	587,8	0,0016 ^D
	N	33,0	97,0	314,5	
Grau de entrada por animais movimentados	P	0	8,0	100,0	0,0604
	N	0	1,0	60,5	
Grau de saída por animais movimentados	P	29,2	112,5	414,0	0,0054 ^D
	N	20,5	76,0	213,5	
Grau por lotes movimentados	P	2,0	6,0	16,7	0,0262 ^D
	N	2,0	5,0	11,0	
Grau de entrada por lotes movimentados	P	0	1,0	3,0	0,1587
	N	0	1,0	2,0	
Grau de saída por lotes movimentados	P	2,0	5,0	12,7	0,0404 ^D
	N	1,5	4,0	9,0	
Grau por vizinhos	P	2,0	4,0	8,0	0,0619
	N	2,0	3,0	6,0	
Grau de entrada por vizinhos	P	0	1,0	2,0	0,2283
	N	0	1,0	2,0	
Grau de saída por vizinhos	P	1,0	3,0	6,0	0,0829
	N	1,0	2,0	4,0	

^D valores que apresentaram diferença estatística ($\alpha = 0,05$)

A tabela 10 mostra o valor de P da análise pelo Mann-Whitney, 1º quartil, a mediana e o 3º quartil do parâmetro *betweenness*.

Tabela 10 – Valor de P , 1º quartil, mediana e 3º quartil dos valores do *betweenness* das propriedades positivas e negativas

Comparação do <i>betweenness</i> entre propriedades positivas (P) e negativas (N)					
Parâmetro		1º Quartil	Mediana	3º Quartil	Valor de P^*
<i>Betweenness</i>	P	0	0	24.360	0,3364
	N	0	0	23.590	

* $\alpha = 0,05$

A tabela 11 mostra o valor de P da análise pelo Mann-Whitney, 1º quartil, a mediana e o 3º quartil do parâmetro *closeness*.

Tabela 11 – Valor de P , 1º quartil, mediana e 3º quartil dos valores do *closeness* das propriedades positivas e negativas

Comparação do <i>closeness</i> entre propriedades positivas (P) e negativas (N)					
Parâmetro		1º Quartil	Mediana	3º Quartil	Valor de P^*
<i>Closeness</i>	P	1,0	1,361	1,361	0,1871
	N	1,0	1,361	1,361	

* $\alpha = 0,05$

A tabela 12 mostra o valor de P da análise pelo Mann-Whitney, 1º quartil, a mediana e o 3º quartil do parâmetro coeficiente de aglomeração.

Tabela 12 – Valor de P , 1º quartil, mediana e 3º quartil dos valores do coeficiente de aglomeração das propriedades positivas e negativas

Comparação do coeficiente de aglomeração entre propriedades positivas (P) e negativas (N)					
Parâmetro		1º Quartil	Mediana	3º Quartil	Valor de P^*
Coeficiente de aglomeração de entrada	P	0	0	0	0,4085
	N	0	0	0	
Coeficiente de aglomeração de saída	P	0	0	0,055	0,1608
	N	0	0	0,044	

* $\alpha = 0,05$

A tabela 13 mostra o valor de P da análise pelo Mann-Whitney, 1º quartil, a mediana e o 3º quartil do parâmetro *PageRank*.

Tabela 13 – Valor de P , 1º quartil, mediana e 3º quartil dos valores do *PageRank* das propriedades positivas e negativas

Comparação do <i>PageRank</i> entre propriedades positivas (P) e negativas (N)					
Parâmetro		1º Quartil	Mediana	3º Quartil	Valor de P^*
<i>PageRank</i>	P	$4,634 \times 10^{-6}$	$5,462 \times 10^{-6}$	$9,359 \times 10^{-6}$	0,4186
	N	$4,634 \times 10^{-6}$	$4,728 \times 10^{-6}$	$9,204 \times 10^{-6}$	

* $\alpha = 0,05$

As comparações dos parâmetros entre as propriedades positivas e negativas utilizando o teste de Mann Whitney (Tabelas 9, 10, 11, 12 e 13) indicaram que houve diferença estatisticamente significativa apenas para o grau total e de saída ponderados por número de animais e por número de lotes movimentados. Além disso, é possível perceber pelos quartis (1º quartil, mediana, 3º quartil) da tabela 9 que os valores dos parâmetros que apresentaram diferença estatística significativa são maiores para as propriedades positivas que para as propriedades negativas.

As informações obtidas pela análise do grau sugerem que as propriedades positivas possuem uma movimentação de animais e lotes de animais mais intensa que as propriedades negativas, sendo a movimentação de venda (saída de animais e lotes) a principal responsável pela diferença encontrada.

A análise da movimentação de compra (entrada de animais ou lotes) não demonstrou diferença estatisticamente significativa entre as propriedades positivas e negativas, porém era esperado que as propriedades positivas apresentassem um grau de entrada maior que as negativas, já que um dos fatores de risco para que a propriedade se torne positiva é a entrada de grande quantidade de animais (alto grau de entrada), que consequentemente aumenta a probabilidade de adquirir bovinos doentes.

A análise dos parâmetros *betweenness*, *closeness*, coeficiente de aglomeração e *Pagerank* demonstrou que não houve diferença estatisticamente significativa entre as propriedades positivas e negativas. Portanto, a partir dos resultados desses parâmetros, é possível inferir que as propriedades positivas e negativas não apresentam diferenças, respectivamente: quanto à existência de propriedades de ligação que intermedeiam o fluxo de animais, quanto à distância

geodésica entre as propriedades, quanto à formação de aglomerados de propriedades e quanto à importância das propriedades na rede de movimentação.

Portanto, um importante achado deste estudo é que o padrão de distribuição do efetivo bovino nas propriedades foi mantido entre os anos de 2003 e 2007. Outra observação importante é que as propriedades que foram detectadas como positivas para brucelose no levantamento de 2003 venderam uma quantidade de lotes e animais superior à quantidade vendida pelas propriedades negativas ($p < 0,05$). Caso o status das propriedades positivas tenha se mantido ao longo desses anos, essa observação pode implicar em um potencial risco de espalhamento de brucelose no Estado do MT.

5 CONCLUSÕES

Face ao exposto neste presente trabalho, conclui-se que:

- A utilização da metodologia de redes permite compreender melhor o fluxo de animais em uma determinada região, proporcionando uma importante ferramenta no controle e prevenção de doenças transmissíveis entre animais, como a brucelose.
- A distribuição dos valores de alguns dos parâmetros estudados (*betweenness*, *closeness*, coeficiente de aglomeração e *PageRank*) apresentaram características que seguem a Lei de potências, qualitativamente, porém apenas o *PageRank* quando considerado o grau de entrada apresentou a razão 80/20.
- As propriedades positivas e negativas para brucelose apresentaram padrões de movimentação de bovinos semelhantes, diferenciando-se apenas quanto ao grau total e de saída ponderados por número de animais e por número de lotes movimentados.

REFERÊNCIAS

- BARABÁSI, A. L. **Linked**. New York: Plume, 2003.
- BEAUCHAMP, M. A. An improved index of centrality. **Behavioral Science**, v. 10, p. 161-163, 1965.
- BEER, J. **Doenças infecciosas em animais domésticos**. São Paulo: Roca, 1999. p. 168.
- BIGRAS-POULIN, M.; THOMPSON, R. A; CHRIEL, M.; MORTENSEN, S.; GREINER, M. Network analysis of Danish cattle industry trade patterns as an evaluation of risk potential for disease spread. **Preventive Veterinary Medicine**, v. 76, n. 1-2, p. 11-39, 2006.
- BIGRAS-POULIN, M.; BARFOD, K.; MORTENSEN, S.; GREINER, M. Relationship of trade patterns of the Danish swine industry animal movements network to potential disease spread. **Preventive Veterinary Medicine**, v. 80, n. 2-3, p.143-165, 2007.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Departamento de Defesa Animal. **Programa Nacional de Controle e Erradicação da Brucelose e Tuberculose**. Brasília, 2004. (manual técnico).
- BRIN, S; PAGE, L. The anatomy of a large-scale Hypertextual web search engine. **Computer networks**, v. 30, p. 107-117, 1998.
- CALDARELLI, G. **Scale-Free Networks**. Oxford: Oxford University Press, 2007.
- CARTER, T. B. **sna: Tools for Social Network Analysis**. R package version 2.2-0, 2010. Disponível em: <<http://CRAN.R-project.org/package=sna>>. Acesso em: 16 de outubro de 2012.
- CHRISTLEY, R. M.; PINCHBECK, G. L.; BOWERS, R. G.; CLANCY, D.; FRENCH, N.P.; BENNETT, R.; TURNER, J. Infection in social networks: using network analysis to identify high-risk individuals. **American Journal of Epidemiology**, v. 162, p. 1024–1031, 2003.
- COHEN, R.; HAVLIN, S. **Complex networks structure, robustness and function**. New York: Cambridge University Press, 2010.
- CORNER, L. A. L.; PFEIFFER, D. U.; MORRIS, R. S. Social-network analysis of Mycobacterium bovis transmission among captive brushtail possums (Trichosurus vulpecula). **Preventive Veterinary Medicine**, v. 59, p. 147–167, 2003.
- COSTA, L. D. F.; RODRIGUES, F. A.; TRAVIESO, G.; VILLAS BOAS, P. R. Characterization of complex networks: a survey of measurements. **Advances in Physics**, v. 56, p. 167–242, 2007.

CSARDI, G.; NEPUSZ, T. The igraph software package for complex network research. **InterJournal**, v. Complex Systems, p.1695, 2006. Disponível em <<http://igraph.sf.net>>. Acesso em: 16 de outubro de 2012.

EASLEY, D.; KLEINBERG, J. **Networks, crowds, and markets**. New York: Cambridge University Press, 2010.

FREEMAN, L. C. A set of measures of centrality based upon betweenness. **Sociometry**, v. 40, p. 35-41, 1977.

FREEMAN, L. C. Centrality in social networks conceptual clarification. **Social Networks**, v. 1, p. 215-239, 1979.

GRISI-FILHO, J. H. H. **Caracterização de circuitos pecuários com base em redes de movimentação de animais**. 2012. 30 p. Tese (Doutorado). Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2012.

HANDCOOK, M.; HUNTER, D.; BUTTS, C.; GOODREAU, S.; KRIVITSKY, P.; MORRIS, M. Statnet software tools for the statistical modeling of network data, 2003. Disponível em: <http://CRAN.R-project.org/package=statnet>. Acesso em: 16 de outubro de 2012.

IBGE. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Produção de pecuária municipal, 2007**. Rio de Janeiro: IBGE, 2007. Disponível em: <http://www.ibge.gov.br/home/presidencia/noticias/noticia_visualiza.php?id_noticia=1269&id_pagina=1>. Acesso em: 17 fev. 2011.

KURODA, R. B. dos S. **Avaliação da dinâmica do efetivo bovino no Estado do Mato Grosso e seu impacto no controle da brucelose bovino**. 2012. 66f. Tese (Doutorado). Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2012.

MARTÍNEZ-LÓPEZ, B.; PEREZ, A. M.; SÁNCHEZ-VIZCAÍNO, J. M. Social network analysis. Review of general concepts and use in preventive veterinary medicine. **Transboundary and Emerging Diseases**, v. 56, n. 4, p. 109-120, 2009.

NATALE, F.; GIOVANNINI, A.; SAVINI, L.; PALMA, D.; POSSENTI, L.; FIORE, G.; CALISTRI, P. Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. **Preventive Veterinary Medicine**, v. 92, n. 4, p. 341-350, 2009.

NEGREIROS, R. L. **Caracterização da brucelose bovina no Estado de Mato Grosso**. 2006. 104 p. Dissertação (Mestrado). Faculdade de Medicina Veterinária e Zootecnia. Universidade de São Paulo, São Paulo, 2006.

NEGREIROS, R. L. **Caracterização e análise da rede de movimento de bovinos no Estado de Mato Grosso**. 2010. 121 p. Tese (Doutorado). Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2010.

NEWMAN, M. E. J. **Networks**: an introduction. Oxford: Oxford University Press, 2010.

OPSAHL, T. **Structure and evolution of weighted networks**. London, UK: University of London, 2009, p. 104-122.

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2012. Disponível em: <<http://www.R-project.org/>>. Acesso em: 16 de outubro de 2012.

GLOSSÁRIO

Grande Componente: ou Giant Component (GC) é um conjunto de nós em que todos os nós estão ligados entre si, independente da direção das arestas.

Lista de adjacência: Tabela composta pela indicação da origem e do destino de cada movimento.

Matriz de adjacência: é a matriz que representa as movimentações de uma rede, na qual as linhas representam a origem das arestas e as colunas os destinos das arestas.

Nós vizinhos: são os nós conectados diretamente ao nó de referência, ainda existem os nós vizinhos primários e secundários. Geralmente quando não é feita a menção do grau da distância dos vizinhos refere-se aos vizinhos primários.

APÊNDICE A – Código elaborado em R para o cálculo do parâmetro coeficiente de aglomeração

```

clustering <- function(tipo,a)
  # a: matriz de adjacencia
  # tipo:'in','out'
{
  kin <- colSums (a)
  kout <- rowSums(a)
  k <- kin + kout
  n <- length(k)
  c <- matrix(0,n,1)
  denin <- denout <- matrix(0,n,1)

  # Calculando os denominadores
  # Para k[i]=1, den[i]=1 para evitar indefinicao 0/0

  for (i in 1:n){
    if (kin[i] > 1) denin[i] <- (kin[i]*(kin[i]-1))
    else denin[i] <- 1
  }
  for (i in 1:n){
    if (kout[i] > 1) denout[i] <- (kout[i]*(kout[i]-1))
    else denout[i] <- 1
  }

  if(tipo=='in')
  {
    nos.calculaveis <- which(kin>1)
    c[-nos.calculaveis] <- 0
    for (i in nos.calculaveis)
    {
      somain <- 0
      ci <- a[,i]
      submatriz <- a[which(ci>0),which(ci>0)]
      submatriz <- submatriz + t(submatriz)
      submatriz[submatriz > 1] <- 1
      somain <- sum(submatriz)/2
      c[i] <- (1/denin[i])*somain
      print(i)
    }
  }

  if(tipo=='out')
  {
    nos.calculaveis <- which(kout>1)
    c[-nos.calculaveis] <- 0
    for (i in 1:n)
    {
      somaout <- 0
      li <- a[i,]
      submatriz <- a[which(li>0),which(li>0)]
      submatriz <- submatriz + t(submatriz)
      submatriz[submatriz > 1] <- 1
      somaout <- sum(submatriz)/2
      c[i] <- (1/denout[i])*somaout
      print(i)
    }
  }
  else
  {warning('erro nos dados de entrada')}

  # list(c,somai)
  # return(somai)
  return(c)
}

```


APÊNDICE B – Código elaborado em R para o cálculo da matriz H e do parâmetro *PageRank*

```

rnhm <- function(A)
{
  Srow <- rowSums(A)
  n <- dim(A)[1]
  H <- Matrix(0,n,n)
  for (i in 1:n)
  {
    print(i)
    if (Srow[i] != 0)
      H[i,] <- A[i,]/Srow[i]
  }
  return(H)
}

```

```

pagerank <- function(H,alpha=0.85,epsilon=1e-8,pr0)
{
  n <- dim(H)[1]
  if (missing(pr0))
    pr0 <- rep(1/n,n)
  rowsumvector <- rowSums(H) == 0
  a <- rowsumvector*1
  residual <- 1
  pr <- pr0
  while (residual >= epsilon)
  {
    prevpr <- pr
    pr <- as.vector((alpha*pr)%*%H) + (alpha*(pr%*%a)+1-alpha)*rep(1/n,n)
    residual <- dist(rbind(pr,prevpr),method='manhattan')
  }
  return(pr)
}

```