

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS - ICEx
DEPARTAMENTO DE ESTATÍSTICA

RELATÓRIO DE MONOGRAFIA

**UMA APLICAÇÃO DO MODELO DE GRAFOS
ALEATÓRIOS EXPONENCIAL PARA DADOS SOBRE
REPRESENTAÇÃO SOCIAL DE ZIKA VÍRUS**

ORIENTADOR: PROF. DR. GILVAN R. GUEDES
CO-ORIENTADORA: PROFA. DRA. DENISE DUARTE

AUTOR: MARINA ALVES AMORIM
MATRÍCULA: 2013034835

BELO HORIZONTE
NOVEMBRO DE 2017

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas - ICEx
Departamento de Estatística

Marina Alves Amorim

Relatório de Monografia

Uma aplicação do modelo exponencial de grafos aleatórios para dados sobre representação social de Zika vírus

Monografia apresentada ao curso de Estatística, como requisito parcial para a obtenção do Título de Bacharel em Estatística, Instituto de Ciências Exatas da Universidade Federal de Minas Gerais

Orientador: Prof. Dr. Gilvan Ramalho Guedes
Departamento de Demografia CEDEPLAR - FACE - UFMG
Co-orientadora: Profa. Dra. Denise Duarte
Departamento de Estatística - ICEx - UFMG

Belo Horizonte
Novembro de 2017

AGRADECIMENTOS

Quero agradecer aos meus pais; Emir e Cleunice. Meu companheiro Mateus e minha família, em especial minha tia Eucy pelo apoio incondicional ao longo desta jornada.

Agradeço também aos meus orientadores, Prof. Dr. Gilvan Ramalho Guedes e Profa. Dra. Denise Duarte, por me auxiliarem na minha vida acadêmica, fazendo com que eu possa cooperar com a comunidade científica. Agradeço também a um amigo muito querido Wesley Pereira , pela contribuição e auxílio para a realização deste trabalho.

“A ciência nunca resolve um
problema sem criar pelo menos
outros dez.”

George Bernard Shaw

RESUMO

Embora as epidemias de Zika já ocorressem na África, sudeste Asiático e ilhas do Pacífico, foi apenas no final de 2015 que a infecção pelo vírus tomou caráter epidêmico nas Américas. Apesar das intensas campanhas promovidas pelas autoridades na área da saúde e pela mídia, pouco se conhece sobre a interpretação que as pessoas dão à doença ou mesmo se elas estão transformando o conhecimento adquirido em práticas preventivas.

Estudos sobre representações sociais sugerem que vários fatores atuam no sentido de promover diferenças na forma como os indivíduos pensam sobre um objeto social. Este pensamento pode sofrer influência do contexto em que o indivíduo se encontra, ou seja, o nível de exposição, gênero e diferenciais socioeconômicos são alguns fatores responsáveis por essa heterogeneidade de pensamento.

Para conseguir captar essas representações sociais aplicamos a Técnica de Associação Livre de Palavras (TALP). Posteriormente, utilizamos o conceito de redes para conseguir caracterizar as representações sociais.. As redes tem estrutura relacional e representam ligações entre objetos denominados vértices. Tais ligações podem representar amizade entre indivíduos, migrações, parceria entre empresas, etc. Uma vez que temos nossa rede construída, nos utilizamos de técnicas que nos auxiliaram no entendimento da formação destas conexões entre o indivíduos. Aplicamos então os modelos de grafos aleatórios exponenciais (ERGM) para analisar a probabilidade de afinidade cognitiva segundo os atributos observados. Os resultados dão indícios de que existem diferenças entre os indivíduos expostos pela doença, bem como existem diferenças na forma como eles dão significado ao Zika Vírus. Também existe uma indicação de diferenças tanto entre sexos quanto por nível de escolaridade. Os resultados encontrados neste trabalho sugerem que políticas preventivas de saúde ou mensagens via mídia devem entender as diferenças cognitivas sobre uma epidemia para melhorar o foco e aderência das pessoas-alvo ao conteúdo pretendido pela campanha.

Palavras-chave: Zika vírus, Representações Sociais, grafos, Análise de Redes Sociais, ERGM, Governador Valadares.

ABSTRACT

Although Zika's epidemics already occurred in Africa, Southeast Asia, and the Pacific Islands, it was not until the end of 2015 that the virus became epidemic in the Americas. Despite intense campaigns promoted by health authorities and the media, little is known about how people interpret the disease or whether they are transforming their knowledge into preventive practices.

Studies of social representations suggest that several factors act to promote differences in how individuals think about a social object. This thought can be influenced by the context in which the individual is, ie the level of exposure, gender and socioeconomic differentials are some factors responsible for this heterogeneity of thought.

In order to be able to capture these social representations, we apply the Free Word Association Technique (TALP). Subsequently, we use the concept of networks to be able to characterize social representations. The networks have a relational structure and represent links between objects called vertices. Such connections can represent friendship between individuals, migrations, partnership between companies, etc. Once we have our network built, we use techniques that have helped us in understanding the formation of these connections between individuals. We then applied the exponential random graph (ERGM) models to analyze the probability of cognitive affinity according to the observed attributes. The results give clues that there are differences between individuals exposed by the disease, as well as there are differences in how they give meaning to the Zika Virus. There is also an indication of differences between both sexes and educational level. The results found in this paper suggest that preventive health policies or messages through the media should understand the cognitive differences about an epidemic to improve the focus and adherence of target people to the content intended by the campaign.

Keywords: Zika Virus, Social Representation, , Social Network Analysis, ERGM, Governador Valadares.

LISTA DE ILUSTRAÇÕES

Figura 1 – Grafo completo de 6 vértices	15
Figura 2 – Padronização com base na ideia original da evocação	16
Figura 3 – União de evocações originais com a mesma ideia	16
Figura 4 – Correção para evocações repetidas via reclassificação	17
Figura 5 – Correção para evocações repetidas via exclusão	17
Figura 6 – Ponderação da ordem por importância	18
Figura 7 – Relativização dos pesos em relação ao maior vetor	19
Figura 8 – Relação da ponderação por posição	20
Figura 9 – Relação da ponderação por distância	20
Figura 10 – Funcionamento da função indicadora	21
Figura 11 – Comparação de vetores com falta de informação mútua e tamanhos distintos	21
Figura 12 – Distribuição dos graus (a) e dos graus ponderados (b)	30
Figura 13 – Distribuição do número de comunidades - Método de Monte Carlo - 100.000 réplicas	32
Figura 14 – Gráfico de radar do pensamento coletivo da rede	34
Figura 15 – Gráfico de radar do pensamento coletivo das comunidades	34
Figura 16 – Bondade de Ajuste	36
Figura 17 – Item do questionário utilizado para análises	40

LISTA DE TABELAS

Tabela 1	–	Evocações únicas	29
Tabela 2	–	Indivíduos menos ativos na rede	30
Tabela 3	–	Indivíduos mais ativos na rede	31
Tabela 4	–	Número de vértices componentes das comunidades	31
Tabela 5	–	Número de indivíduos e densidade das comunidades	33
Tabela 6	–	Divisão em subgrupos das evocações únicas	33
Tabela 7	–	Estimativa e Razão de Chance para os coeficientes do modelo ERGM	35

SUMÁRIO

1	INTRODUÇÃO	9
2	OBJETIVOS	11
2.1	Objetivo geral	11
2.2	Objetivos específicos	11
3	METODOLOGIA	12
3.1	Descrição do banco de dados, questionário e do módulo de associação livre de palavras para Zika Vírus	13
3.2	Representação da rede de indivíduos através da Teoria dos Grafos	13
3.2.1	Algumas medidas topológicas de um grafo	14
3.2.2	Preparação dos dados de evocações sobre Zika Vírus	16
3.2.2.1	Padronização das evocações originais	16
3.2.2.2	Verificação e tratamento de <i>loops</i>	17
3.2.3	Construção da matriz de adjacências	18
3.3	Modularidade, Algoritmo de Louvain e verificação da partição encontrada	22
4	MODELOS DE GRAFOS ALEATÓRIOS EXPONENCIAIS - ERGM	25
4.1	Pacote ERGM	26
4.2	Bondade de Ajuste	27
5	RESULTADOS E DISCUSSÕES	29
5.1	Análise exploratória dos dados	29
5.2	Informações gerais do grafo representativo da rede de pensamento coletivo	30
5.3	Decompondo a rede em comunidades e analisando a qualidade da partição encontrada	31
5.3.1	Verificando a partição encontrada	31
5.4	Comparando os pensamentos coletivos	33
5.5	Ajuste de modelo	35
5.5.1	Bondade de Ajuste	35
6	CONCLUSÃO	37
7	REFERÊNCIAS	38

8	ANEXO	40
---	-----------------	----

1 INTRODUÇÃO

No final de 2015 o vírus da Zika registrava os primeiros sinais de infecção nas Américas. Embora seja transmitido pelo mesmo vetor da Dengue e da Chikungunya, o Zika ganhou rapidamente a atenção da mídia em razão de seu efeito mais alarmante - a microcefalia, um tipo de Síndrome Congênita da Zika. A mídia, nacional e internacional, concentrou sua cobertura nos bebês nascidos com o crânio de circunferência reduzida, com um futuro provavelmente comprometido. O temor de uma consequência de larga escala para toda uma geração levou as autoridades a recomendar a postergação da gravidez para as mulheres. Foi também reforçada a utilização de medidas de eliminação do vetor que já estavam sendo praticadas em razão da Dengue.

A intensa divulgação de imagens mostrando bebês nascidos com microcefalia produziu uma agitação social e um senso de vulnerabilidade coletiva nos países em que o Zika se espalhou. Nos meses iniciais da epidemia, havia ainda muita incerteza sobre as causas, consequências e potencial para difusão da doença, tornando difícil para a sociedade civil entender essa nova e crescente ameaça de saúde pública. A despeito da intensa e dramática cobertura dada ao Zika, ainda se conhece muito pouco sobre como as pessoas interpretam a doença, particularmente se elas a percebem como uma ameaça real, como elas dão sentido à vulnerabilidade em que estão expostas e se estão colocando o conhecimento adquirido em prática.

Este trabalho procura contribuir com a literatura sobre aspectos sociais e cognitivos relacionado ao Zika ao examinar como os indivíduos coletivamente significam a epidemia. A partir da evidência empírica de outros estudos que o envolvimento pessoal influencia e molda a representação da doença, este trabalho estabelece duas hipóteses. (1) Os sentidos (significados) que as pessoas dão ao que vêem ou ouvem durante a explosão de uma epidemia estão sujeitos as suas experiências pessoais com o vírus. Indivíduos infectados por Zika, Dengue ou Chikungunya podem ver a doença de forma diferente quando comparados aqueles para os quais o contato com esses vírus ocorreu apenas por intermédio das notícias ou campanhas de saúde. Nesse sentido, pessoas que compartilham a mesma experiência de exposição ao Zika teriam uma maior afinidade cognitiva. (2) Essa afinidade cognitiva também depende de fatores adicionais, como o contato com pessoas que também foram expostas ao vírus e outros atributos sociodemográficos (como sexo, renda e escolaridade). Diferenciais de gênero podem emergir pela forma como as campanhas de divulgação apresentam sempre o foco sobre a gravidez (vista pela ótica das mulheres como agentes responsáveis quase que exclusivas pela prevenção). Os diferenciais socioeconômicos podem emergir das diferentes capacidades de auto-proteção e acesso a informações técnicas, bem como as diferenças na capacidade de negociação de gênero para

implementação dos mecanismos de proteção quando classes socioeconômicas distintas são comparadas

Para conseguir representar como os indivíduos significam a epidemia, utilizamos o conceito de redes, que pode ser representada de várias formas, uma delas é como um grafo, que consiste em nós conectados por arestas e vértices. É importante ressaltar que ao decorrer do trabalho utilizaremos o termo *redes* como sinônimo de *grafo*. A teoria dos grafos nos possibilita uma lista de particularidades que podem ser utilizadas para reproduzir muitas propriedades das representações sociais, a teoria também disponibiliza operações matemáticas que podem quantificar e medir estas propriedades.

Estas redes tem estrutura relacional e representam ligações entre os vértices, tais como amizade entre indivíduos, migrações, parceria entre empresas, etc. Existe uma dependência entre os vértices como também uma dependência entre as ligações (arestas). Temos vários modelos estatísticos que são baseados em suposições de independência, mas a natureza complexa de dependência nas redes tem retardado o desenvolvimento desses modelos para estruturas de rede. Temos inúmeros tipos de dependência e compreendê-las é importante, pois os diferentes modelos estatísticos que existem na literatura divergem na forma de representá-la (Snijders 2011).

Com isso, utilizaremos os modelos de grafos aleatórios exponenciais para analisar a probabilidade de afinidade cognitiva segundo atributos epidemiológicos (exposição pessoal ao Zika e contato com outra pessoa segundo exposição ao vírus) e sociodemográficos (sexo e escolaridade). Este modelo estatístico é mais utilizado no contexto de redes sociais, além disso também são utilizados em outras áreas como demografia, biologia e outras. Para representar o pensamento coletivo da rede, introduziremos o modelo no contexto de representações sociais, bem como no livro de Kolaczyk e Csárdi (2014).

2 OBJETIVOS

2.1 OBJETIVO GERAL

Identificar e compreender como se representa o pensamento coletivo sobre Zika vírus em Governador Valadares-MG, suas heterogeneidades em subgrupos e quais variáveis explicam os padrões de afinidade cognitiva formados pela rede.

2.2 OBJETIVOS ESPECÍFICOS

1. Analisar a estrutura das representações sociais sobre Zika Vírus através da Análise de redes sociais;
2. Identificar possíveis diferenças nas representações sociais sobre Zika vírus através da decomposição da rede coletiva;
3. Ajuste dos modelos ERGM para entender a formação de conexões na rede de acordo com os atributos observados.

3 METODOLOGIA

Para a implementação dos objetivos propostos para neste trabalho, vamos adotar uma abordagem para extrair dos indivíduos observados informações sobre a forma como dão significado ao Zika vírus. Essa abordagem deve ser eficaz para representar este conjunto de significados, tanto no nível individual, quanto no nível coletivo.

Utilizaremos a Teoria das Representações Sociais (TRS) proposta por Moscovici (1961), no intuito de explicar os fenômenos associados a uma determinada sociedade a partir de uma perspectiva coletiva dos indivíduos. Para a representação de objetos sociais a TRS considera vários fatores como: cultura, comunicação e memória coletiva. Estes fatores sociais exercem influência sobre o indivíduo na representação do objeto social. Além de tudo, a TRS preserva a individualidade pessoal, considerando o fator cognição como mais um componente da representação individual. Esta teoria tem mecanismos capazes de capturar o pensamento leigo de cada indivíduo e representar, baseado na amostra coletada, o pensamento coletivo na sociedade a qual esses indivíduos pertencem. Portanto, vários elementos de atribuição de significados individual expressam, na verdade, um processo socialmente compartilhado, e pode revelar a estrutura comum de pensamento dessa comunidade, sejam elas realísticas do ponto de vista científico ou não.

Foi proposto por Abric em 1994 e 2001, a Técnica de Associação Livre de Palavras (TALP) é um dos instrumentos mais comumente utilizados para coletar dados sobre as Representações Sociais. Essa técnica consiste na proposição de um determinado termo indutor (representando o objeto social) ao entrevistado, a partir do qual este relata de forma livre e instantânea as k primeiras palavras ou expressões (denominadas evocações) que lhe surgem à mente. Por se tratar de uma coleta de evocações espontâneas, o elemento cognitivo é colocado em perspectiva, dando espaço para que o pensamento leigo expresse o seu papel na construção mental daquele objeto social (Pereira, 2017). Em seguida, é solicitado ao entrevistado que classifique, por ordem de relevância pessoal, as evocações por ele citadas. A cada indivíduo é atribuído um vetor com até k evocações fornecidas e, a partir desse conjunto de vetores, reconstrói-se a estrutura de pensamento coletivo da sociedade à qual o indivíduo pertence.

3.1 DESCRIÇÃO DO BANCO DE DADOS, QUESTIONÁRIO E DO MÓDULO DE ASSOCIAÇÃO LIVRE DE PALAVRAS PARA ZIKA VÍRUS

O questionário utilizado para a coleta dos dados amostrais possui 34 itens a serem respondidos pelos entrevistados. Dentre as informações coletadas temos, conhecimento do domicílio, identificação pessoal, classificação social e exposição a doença (Zika, Dengue ou Chikungunya).

Utilizaremos a questão de número 8 do questionário, relacionada a Evocação Livre de Palavras, no qual o termo indutor é "Zika vírus". Cada entrevistado foi convidado a expressar até cinco palavras ou expressões que lhe ocorreram instantaneamente em relação a este termo indutor. Já o item 9 representa a ordenação dessas expressões por ordem de importância atribuída pelo entrevistado, variando de 1 a 5, sendo, 1 a expressão mais importante e 5 a de menor importância.

Temos um total de 150 entrevistados, 15 indivíduos não evocou qualquer palavra ou expressão, logo temos 135 indivíduos contribuindo com pelo menos 1 evocação. Como cada indivíduo pode fornecer no máximo 5 evocações, a amostra válida conterá no máximo 675 evocações.

Os dados utilizados são fruto do projeto "Demografia da Exceção: intenções reprodutivas e migração em um contexto de Zika Vírus", financiado pela Rede Clima, CNPq e pela FAPEMIG, e coordenado pelos professores Dr. Alisson Barbieri e Dr. Gilvan Ramalho Guedes, da Universidade Federal de Minas Gerais (Processo: 431872/2016-3) e FAPEMIG (Processo n.º. CSA - APQ-01553-16) - UFMG.

3.2 REPRESENTAÇÃO DA REDE DE INDIVÍDUOS ATRAVÉS DA TEORIA DOS GRAFOS

Nesta análise, cada indivíduo contribui com até cinco evocações para a estrutura coletiva de significados. Existem técnicas propostas na literatura para este tipo de dados (Abric (1994), Sa (1996), Hora (2013)). No entanto, essas técnicas não quantificam as características topológicas da rede de indivíduos, como a sua centralidade e sua intermediação para a formação da rede de significados. Uma alternativa à essas técnicas é a representação da rede através de grafos (Bachrach, 2014). Na Teoria dos Grafos, um grafo $G = (V, E)$ é definido como um conjunto de objetos de duas naturezas: o vértice (V), a unidade fundamental sob a qual o grafo é construído e as arestas (E), que representam a ligação entre pares de vértices. A essas ligações podem ser atribuídas muitas características, como direção, distância e peso. Existem aplicações onde uma aresta conecta

o vértice a ele próprio, denominado *loop*. Matematicamente, esses conjuntos podem ser representados conjuntamente através de matrizes de adjacência, onde linhas e colunas representam os vértices e as células representam a ligação entre cada par linha-coluna, ou mesmo através de listas de conexões, onde cada linha da lista se refere a uma aresta ligando os dois vértices listados na mesma.

Representar a rede de indivíduos por intermédio de um grafo possibilita evidenciar indivíduos "populares", os quais possuem um grande número de conexões, resumindo as ideias mais difundidas em sua comunidade. O estudo destes indivíduos de alta popularidade pode ser utilizado em campanhas, experimentando a receptividade dos mesmos a ações, serviços e/ou produtos relacionados à temática da rede, em busca de estimação da receptividade geral da comunidade (Pereira, 2017).

Outra propriedade desta abordagem consiste na possibilidade de encontrar a difusão espacial do pensamento coletivo, sob a condição de se conhecer o georreferenciamento dos indivíduos. Analisando a distribuição do padrão espacial do grupo de indivíduos com a mesma ideia, verifica-se se existe concentração destes indivíduos em uma determinada região ou dispersão dos mesmos ao longo do território. Dessa forma, procura-se delimitar a área de propagação do pensamento, verificando se existe nele algum indício de influência geoespacial. Além disso, pode-se realizar a validação das partições encontradas, bem como a verificação sobre indícios de processos latentes responsáveis pelo regimento comportamental da rede.

3.2.1 ALGUMAS MEDIDAS TOPOLÓGICAS DE UM GRAFO

Baseado na representação de redes através de grafos, utilizam-se algumas medidas descritivas topológicas para interpretar as características da rede. Seja u_i o i -ésimo vértice da rede u . Algumas de suas medidas descritivas topológicas são:

- **Medidas sobre os vértices:**

- *Grau*: Representa o número de arestas incidentes sobre u_i .
- *Grau ponderado*: Representa a soma dos pesos das arestas incidentes em u_i ponderadas por valores atribuídos segundo algum critério. Vértices com alto grau ponderado são caracterizados por uma combinação entre conexões de peso elevado e alta frequência de conexões. Será utilizada neste trabalho para mensurar a atividade de um vértice na rede.
- *Proximidade*: É definida como o inverso do afastamento de u_i em relação aos demais vértices. O afastamento é definido como sendo a soma das distâncias entre u_i e os demais vértices.
- *Intermediação*: Quantifica o número de vezes em que u_i age como ponte ao longo do caminho mais curto entre outros dois vértices.

• **Medidas sobre o grafo:**

- *Ordem*: A ordem de um grafo é definida como o número de vértices que o formam.
- *Tamanho*: O tamanho de um grafo é definido como sendo a soma entre o número de vértices e o número de arestas que o formam.
- *Homofilia*: Tendência dos nós de se relacionarem com outros nós que apresentem características similares (cor, raça, sexo, etc).
- *Transitividade* : É a proporção de conjuntos de três vértices na rede que formam triângulos.
- *Densidade*: Mensura a proporção de arestas observadas em um subgrafo H para as possíveis arestas do grafo G . Seja o subgrafo $H = (V_H, E_H)$, sua densidade é:

$$den(H) = \frac{|E_H|}{|V_H|(|V_H| - 1)/2}$$

Assim, altas densidades são alcançadas quando todas as possíveis conexões são realizadas. Num grafo completo, a densidade é igual a 1, tal como na Figura (1). Espera-se que vértices com alta centralidade formem um subgrafo mais denso com seus vizinhos.

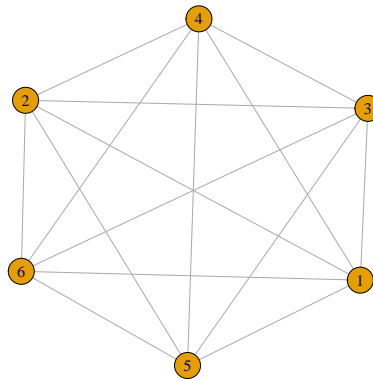


Figura 1 – Grafo completo de 6 vértices

Para este trabalho, a rede consistirá em um grafo não-direcionado, onde cada indivíduo será considerado um vértice que possui conexão com outro indivíduo quando seus vetores de evocações tiverem ao menos uma evocação em comum. A relação de arestas estará alocada em uma matriz de adjacências ponderada pela força das conexões, calculada com base no número de evocações em comum, sua ordem de relevância, no distanciamento

entre as ordens de relevância atribuídas pelos indivíduos para as evocações em comum e na quantidade de evocações expressadas pelos indivíduos.

3.2.2 PREPARAÇÃO DOS DADOS DE EVOCAÇÕES SOBRE ZIKA VÍRUS

A princípio, as evocações poderiam ser compostas por uma única palavra até orações completas. A coleta em nível individual resultou em uma enorme diversidade de palavras e expressões. Em termos de análises estatísticas, é de interesse haver variabilidade, desde que associado a frequências satisfatórias para cada nível do conjunto de dados. Dessa forma, foram adotadas transformações nos dados originais de forma a garantir frequências satisfatórias, mantendo ao mesmo tempo heterogeneidade dos dados e a ideia subjacente as expressões originais.

3.2.2.1 Padronização das evocações originais

A primeira etapa do tratamento do banco de dados é identificar a ideia por trás de cada evocação. Desta forma, pode-se reduzir frases inteiras a uma expressão mais simples, contendo apenas a ideia principal expressa pelo entrevistado. A Figura 2 apresenta um exemplo desta etapa.

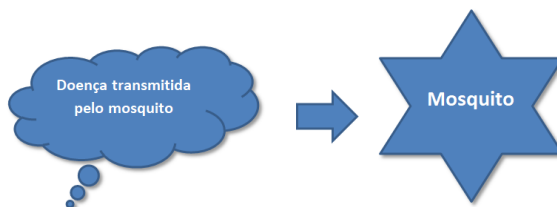


Figura 2 – Padronização com base na ideia original da evocação

A segunda etapa consiste em uma reavaliação das novas evocações, buscando unir em um mesmo nível evocações com sentidos similares. Nessa etapa, o número de níveis distintos que o banco de dados possuirá estará definido. A Figura 3 apresenta um exemplo desta etapa.

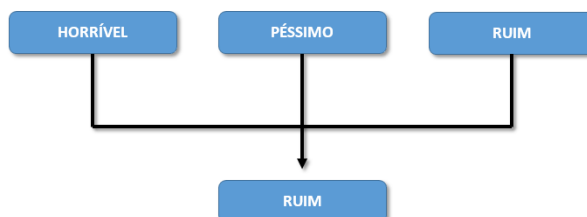


Figura 3 – União de evocações originais com a mesma ideia

O processo de padronização descrito acima resultou em 106 evocações únicas, expressando sentimentos negativos e danos provocados pelo Zika vírus. Deve-se levar em

consideração que, mesmo que as evocações possuam sentidos similares, deve-se evitar unir grupos com frequências observadas muito altas, sob ônus de criar um grupo muito grande que pode desprezar a possível heterogeneidade entre as ideias difundidas na sociedade.

3.2.2.2 Verificação e tratamento de *loops*

Uma consequência de padronizar as evocações coletadas é o risco de se obter um vetor contendo duas ou mais evocações padronizadas idênticas. No caso da utilização destes vetores padronizados em análise de redes, a repetição de expressões idênticas representaria um *loop*. No entanto, este não é um padrão natural, mas criado artificialmente pelo processo de padronização. Neste trabalho, utilizaram-se dois passos para corrigir essa situação:

1. Ao ser detectado um *loop*, realiza-se a reavaliação da classificação das evocações com o auxílio das evocações reduzidas ou mesmo as evocações originais (Figura 2). Sempre que possível, a evocação padronizada com menos similaridade à ideia original é classificada em uma nova categoria. A Figura 4 apresenta um exemplo de execução deste passo.

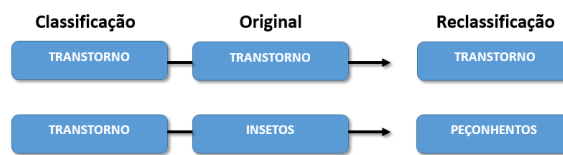


Figura 4 – Correção para evocações repetidas via reclassificação

2. Na eventualidade de o Passo 1 não ser efetivo, tanto por ideias realmente similares ou por falta de classes que melhor represente alguma das evocações repetidas, opta-se pela exclusão da evocação duplicada com ordem de menor importância. Caso existam evocações com ordem menor que evocação excluída, essas evocações herdam sucessivamente as posições do vetor, de forma a deixarem as células vazias nas últimas posições do vetor. A Figura 5 apresenta um exemplo de execução deste passo.

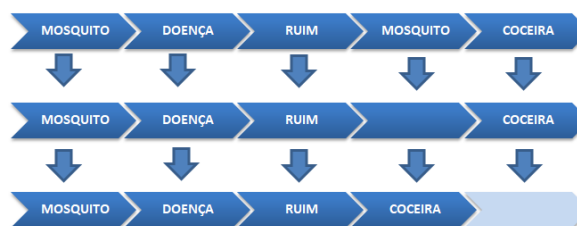


Figura 5 – Correção para evocações repetidas via exclusão

Após esta etapa, conclui-se o processo de tratamento do banco de dados. Do total de 675 evocações possíveis, o processo de padronização descrito gerou 643 evocações padronizadas, distribuídas em 106 níveis diferentes (evocações únicas), expressadas por 135 entrevistados.

3.2.3 CONSTRUÇÃO DA MATRIZ DE ADJACÊNCIAS

Para a construir a matriz de adjacências, considerou-se que duas pessoas estão ligadas uma entre si através do nível de afinidade entre um par de vetores de evocações. Como explicado anteriormente, cada vetor de evocações é ordenado de acordo com o grau de importância daquela expressão para o indivíduo. Foram atribuídos pesos variando entre 1 e k , onde k representa o tamanho do vetor de evocações do indivíduo. São atribuídos pesos para as posições em ordem decrescente, sendo k para a evocação com a ordem mais importante e 1 para a evocação com a ordem menos importante. A Figura 6 apresenta um exemplo deste critério.

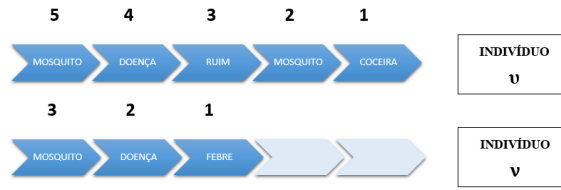


Figura 6 – Ponderação da ordem por importância

Foi proposto por Pereira (2017), um coeficiente de afinidade generalizado entre dois indivíduos que leva em consideração a informação fornecida pelas evocações: o conteúdo da evocação, a ordem de importância para o indivíduo e a distância relativa entre a posição das evocações nos dois vetores. Sejam u e v os vetores de evocação de dois indivíduos comparáveis entre si, e n_u e n_v são os tamanhos dos vetores u e v , respectivamente. O coeficiente de afinidade é definido por

$$\alpha(u, v) = \underbrace{\left\{ \frac{\sum_{i=1}^n \sum_{j=1}^n [\theta(i, j, n) * \rho(i, j, n) * 1_{u_i=v_j}]}{n^2 * (n + 1)} \right\}}_{\beta(u, v)} * \underbrace{\left\{ 1 - \frac{(N - n) * (N - n + 1)}{N * (N + 1)} \right\}}_{\omega(n)}, \quad (3.1)$$

onde:

- $n = \max(n_u, n_v)$. Para este coeficiente, se os vetores têm tamanho diferente, o vetor menor de tamanho n_{min} será considerado como tendo tamanho n , atribuindo a ele posições vazias nas $n - n_{min}$ últimas posições. Exemplificando, se u tem comprimento $n_u = 4$ e v tem comprimento $n_v = 3$, $n = \max(4, 3) = 4$. Então v

passa a ter comprimento $n = 4$, sendo as $4 - 3 = 1$ últimas posições do vetor v posições vazias, como mostrado na Figura 7.



Figura 7 – Relativização dos pesos em relação ao maior vetor

A justificativa para este procedimento é o fato de que supõe-se que os indivíduos sejam igualmente importantes para o estudo. Desta forma, garantimos que para as comparações entre as pessoas, elas sejam comparadas em igualdade de pesos. Assim, assume-se as posições vazias como alguma evocação implícita, que por algum motivo não foi expressada e que possui menos relevância para o indivíduo do que as evocações mencionadas, dada a construção do instrumento de Associação Livre de Palavras, possui menos relevância para o indivíduo do que as evocações mencionadas. Ademais, pelo desconhecimento de tal evocação implícita, ela será considerada na comparação como discordante de qualquer evocação expressada.

- $\theta(i, j, n)$ pondera a informação da ordem das evocações dos indivíduos. é definido por

$$\theta(i, j, n) = 2 * (n + 1) - (i + j) \quad (3.2)$$

A Equação 3.2 é baseada na soma dos pesos atribuídos às posições u_i e v_j . Para $n = 5$, $i = 1$ e $j = 4$,

$$k_1 + k_4 = 5 + 2 = 7 = 2 * (5 + 1) - (1 + 4) = \theta(1, 4, 5). \quad (3.3)$$

O exemplo da Equação 3.3 é ilustrado pela Figura 8.

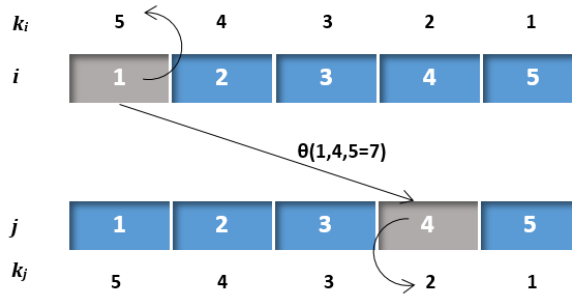


Figura 8 – Relação da ponderação por posição

- $\rho(i, j, n)$ pondera a informação da distância entre as posições das evocações dos indivíduos. é definido por

$$\rho(i, j, n) = n - |i - j|. \quad (3.4)$$

A Equação 3.4 é baseada no quão distantes estão as evocações u_i e v_j . Supõe-se que dois indivíduos estão mais afins se as evocações por eles expressadas estiverem em posições próximas de relevância. No entanto, se os indivíduos expressam a mesma evocação, mas com relevâncias muito distintas, esses indivíduos serão estar menos afins. A Figura 9 exemplifica a Equação 3.4.

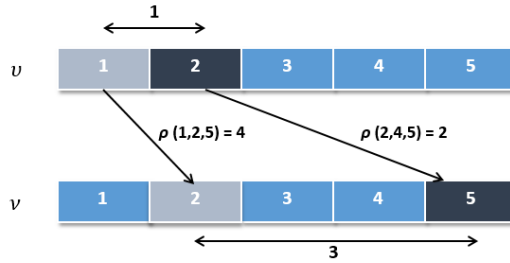


Figura 9 – Relação da ponderação por distância

- A função indicadora

$$1_{u_i=v_j} = \begin{cases} 1, & u_i = v_j \\ 0, & u_i \neq v_j \end{cases} \quad (3.5)$$

é responsável pela comparação das evocações, zerando o termo aditivo da equação sempre que as evocações u_i e v_j são diferentes. A Figura 10 apresenta um exemplo deste critério.

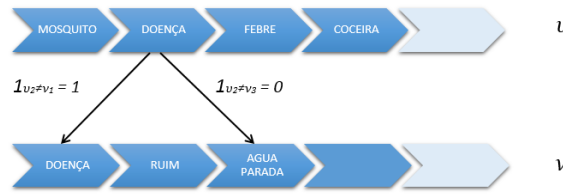


Figura 10 – Funcionamento da função indicadora

- O denominador $n^2 * (n + 1)$ representa o valor máximo que $\beta(u, v)$ pode assumir, garantindo que

$$0 \leq \beta(u, v) \leq 1.$$

- N é o comprimento do vetor de maior comprimento da amostra. Para este trabalho $N = 5$, pois pelo menos um entrevistado contribuiu com 5 evocações padronizadas.
- O fator $\omega(n) = 1 - \frac{(N-n)*(N-n+1)}{N*(N+1)}$, se responsabiliza por penalizar comparações onde ambos os vetores possuem menos evocações do que o máximo permitido N . Esta restrição implica em

$$0 \leq \alpha(u, v) \leq \omega(n).$$

Assim, ao calcular a afinidade entre dois indivíduos, será levado em consideração a quantidade de informação fornecida pelos entrevistados. O fator $\omega(n)$ aplica tal penalidade baseado na soma dos pesos das posições faltantes. Ademais, está intrínseca ao coeficiente proposto na Equação 3.1 a penalidade para comparações onde os vetores têm tamanho distinto, uma vez que a fórmula considera células vazias como diferentes de células contendo evocações. A Figura 11 apresenta um exemplo deste critério, onde poderia alcançar no máximo $1 - \frac{2+1}{5+4+3+2+1} = \frac{12}{15} = 0,8$.

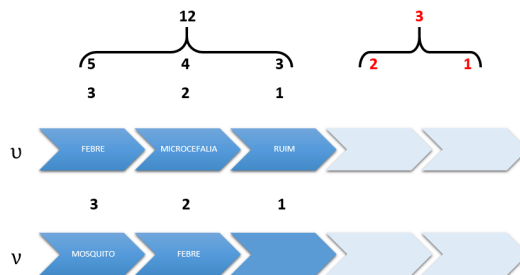


Figura 11 – Comparação de vetores com falta de informação mútua e tamanhos distintos

Para compreender melhor o funcionamento do coeficiente de afinidade apresentado em 3.1, utilizou-se do exemplo da Figura 11. Note que

$$1_{u_i=v_j} = 0, \forall (i, j) \neq (1, 2).$$

Logo,

$$\begin{cases} \theta(1, 2, 3) = 2 * (n + 1) - (i + j) = 2 * (3 + 1) - (1 + 2) = 5 \\ \rho(1, 2, 3) = n - |i - j| = 3 - |1 - 2| = 2 \\ n^2 * (n + 1) = 3^2 + (3 + 1) = 36. \end{cases} \quad (3.6)$$

Desta forma, $\beta(u, v)$ foi calculado a seguir.

$$\frac{\sum_{i=1}^3 \sum_{j=1}^3 [\theta(i, j, n) * \rho(i, j, n) * 1_{u_i=v_j}]}{n^2 * (n + 1)} = \frac{5 * 2}{36} = 0,2778. \quad (3.7)$$

Por fim, multiplicou-se a Equação 3.7 por $\omega(3)$.

$$\alpha(u, v) = 0,2778 * \left[1 - \frac{(5 - 3) * (5 - 3 + 1)}{5 * (5 + 1)} \right] = 0,2778 * \frac{12}{15} = 0,2222. \quad (3.8)$$

O coeficiente viabiliza a construção da matriz de adjacências ponderada, onde a força da ligação entre dois indivíduos é definida como a afinidade entre os mesmos. A matriz é quadrada, simétrica, possui diagonal nula e, para este trabalho, possui dimensão [135 x 135]. A implementação do coeficiente de afinidade e da construção da matriz de adjacência está implementada no *software R*, na biblioteca . Será utilizada a biblioteca *Wordevok* (Pereira, 2017) e *igraph*, implementada no *software R* para a construção da rede de indivíduos.

3.3 MODULARIDADE, ALGORÍTMO DE LOUVAIN E VERIFICAÇÃO DA PARTIÇÃO ENCONTRADA

A modularidade é um coeficiente utilizado na detecção de comunidades em uma rede. Assumindo valores entre -1 e 1 , a modularidade é a medida da densidade das arestas que interligam atores dentro de um mesmo grupo em relação à arestas que interligam atores de grupos distintos.

A fórmula para calcular a modularidade pode ser parametrizada em (Newman, 2006):

$$Q = \frac{1}{2m} \sum_{i,j} [A_{i,j} - P_{i,j}] \delta(c_i, c_j) \quad (3.9)$$

onde $A_{i,j}$ representa o peso da aresta que liga os vértices i e j ; $m = \frac{1}{2} \sum_{i,j} A_{i,j}$ representa a soma dos pesos de todas as arestas contidas no grafo; $P_{i,j}$ é o valor esperado para o peso da aresta que liga os vértices i e j . Usualmente, $P_{i,j} = \frac{k_i k_j}{2m}$, onde k_i representa o grau ponderado do vértice i ; c_i é comunidade da partição ao qual pertence o vértice i ; $\delta(c_i, c_j)$ refere-se ao delta de Kroneker, que tem funcionamento muito similar ao de uma função indicadora, onde

$$\delta(c_i, c_j) = \begin{cases} 1, & c_i = c_j \\ 0, & c_i \neq c_j, \end{cases} \quad (3.10)$$

de forma que não sejam consideradas para a modularidade comparações entre vértices pertencentes à comunidades diferentes. Neste trabalho, foi utilizado o algoritmo Louvain (Clauset, Newman e Moore, 2004). Este algoritmo possui um funcionamento simples divididos em dois passos e alegadamente possui vantagens computacionais para redes muito grandes.

Passo 1: Retira-se o primeiro vértice da sua comunidade. Este será alocado em cada uma das outras $n - 1$ comunidades. Calcula-se a variação da modularidade

$$\Delta Q = \left[\frac{\sum_{in} + k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \quad (3.11)$$

onde:

- \sum_{in} é a soma dos pesos das arestas dentro da comunidade C_j para a qual estamos movendo o vértice i ;
- $k_{i,in}$ é a soma dos pesos das arestas que ligam o vértice i à todos os membros da referida comunidade C_j ;
- \sum_{tot} é a soma dos pesos das arestas incidentes sobre todos os vértices da referida comunidade C_j ;
- k_i é grau ponderado do vértice i ;
- m é a soma dos pesos de todas as arestas da rede.

Calcula-se o ΔQ alcançado na mudança do primeiro vértice de sua comunidade original para cada outra comunidade. A comunidade que alcançar o ΔQ mais positivo aloca o primeiro vértice. Caso $\max(\Delta Q) = 0$, então o primeiro vértice permanece em sua comunidade. Após alocar o primeiro vértice, realiza-se este mesmo procedimento para todos os demais vértices.

Após a primeira execução, o Passo 1 é repetido começando do primeiro vértice novamente, porém agora respeitando as alocações feitas no procedimento anterior. A cada repetição do Passo 1, o número de comunidades diminui ou se mantém igual. A repetição termina quando nenhuma mudança individual de comunidade resulta em $\max(\Delta Q) > 0$.

Passo 2: Uma vez definidas as comunidades do Passo 1, um novo grafo é gerado. Cada comunidade é representada no novo grafo por um vértice. As arestas que conectam vértices de duas comunidades diferentes são agora representadas por uma única aresta a qual atribui-se peso igual à soma dos pesos de todas as arestas que conectam as duas comunidades. Além disso, para não alterar os graus dos vértices, as arestas que ligam indivíduos dentro da mesma comunidade são representadas por um *auto loop* ao qual atribui-se peso igual à soma dos pesos de todas as arestas dentro da comunidade. Após este procedimento, reinicia-se o Passo 1. O algoritmo é finalizado quando a partição resultante do Passo 2 é igual à partição resultante do Passo 1 realizado imediatamente a seguir.

Após encontrar a partição ótima, representa-se cada comunidade através dos respectivos subgrafos da rede geral. Desta forma, será possível descrever o significado coletivo sobre o Zika vírus, a partir da observação das evocações mais representativas do pensamento coletivo em cada uma das comunidades.

A validação das comunidades encontradas será realizada obtendo-se da distribuição dos graus da rede. Geram-se grafos com arestas distribuídas aleatoriamente e condicionalmente à distribuição dos graus, que se mantém fixa. Cada um desses grafos simulados é uma réplica de Monte Carlo do grafo original, e deve ser particionado pelo mesmo critério aplicado ao grafo original observado. As proporções esperadas do número de comunidades para o conjunto de réplicas geradas é comparada com o número de comunidades da partição da rede original. Haverá indícios de que um processo latente não-aleatório é responsável pela partição do grafo se a proporção esperada em grafos aleatórios para o número de comunidades observado na partição do grafo original for pequena (Pereira, 2017).

4 MODELOS DE GRAFOS ALEATÓRIOS EXPONENCIAIS - ERGM

Temos que um grafo é uma representação matemática de uma rede binária. Como notação assumimos o número de nós como n , a variável aleatória como $Y_{i,j}$ que indica se há uma ligação entre os nós i e j ou caso contrário $Y_{i,j} = 0$. O grafo aleatório é dado por uma matriz de adjacência $Y_{i,j}$. Os grafos por definição são não direcionados, ou seja $Y_{i,j} = Y_{j,i}$ para todo i e j .

Os modelos de grafos exponenciais (ERGM), modela as dependências estruturais entre as arestas, que são consideradas variáveis aleatórias, no qual hipóteses são formuladas em termos de parâmetros que apresentam dependências. Estes modelos, muitas vezes chamados de p^* , que pertence a uma família de modelos estatísticos para redes que nos permite inferir sobre os padrões proeminentes nos dados dado a presença de outras estruturas contida na rede.

Os modelos de grafos foram inicialmente estudados por Erdos-Renyi (1959), que propôs um modelo Bernoulli no qual as arestas são consideradas independentes e são de origem aleatória de acordo com uma probabilidade p fixa. Posteriormente Frank e Strauss em 1986 propuzeram os modelos de grafos aleatórios Markovianos onde cada par de arestas é condicionalmente dependente quando eles tem uma ligação em comum e estes modelos são a base do modelo ERGM atual.

Em uma generalização do grafo de Markov proposta por Frank, 1991 e Wasserman e Patisson, 1996, existe um novo modelo proposto que aceita cováriaveis observadas, ou seja :

$$P(Y = y|\theta) = \frac{\exp\{\theta_\kappa S_\kappa(y)\}}{\kappa(\theta)} \quad (4.1)$$

onde θ é um vetor de parâmetros do modelo associado a $s(y)$ e $\kappa(\theta)$ é uma constante de normalização. Podemos ainda incluir atributos ao modelo, com objetivo de melhorar a análise, da seguinte forma :

$$P(Y = y|\theta) = \frac{\exp(\theta^T s(y, X))}{\kappa(\theta)}, \quad (4.2)$$

onde a estatística $s(y, X)$ depende agora de variáveis exógenas à rede, representadas pelo vetor de atributos X . Estas variáveis podem ser usadas para representar as indagações da pesquisa e para obter um bom ajuste do modelo. Este modelo pode ser usado para representar qualquer distribuição no espaço de grafos, desde que atribua uma probabilidade

positiva para cada grafo. O modelo foi chamado p^* por Wasserman e Pattinson (1996) e mais recentemente denominado *Grafos Aleatórios Exponenciais* (ERGM).

O modelo ERGM tem como característica importante que as arestas dependem umas das outras, assim, a existência de uma aresta pode afetar a existência de outras. O modelo também é aberto, o que possibilita ao pesquisador avaliar várias teorias de acordo com o interesse pessoal, uma vez que o mesmo consiga traduzir essas teorias em termos quantitativos relacionais.

Conforme descrito anteriormente, um grafo aleatório não direcionado pode ser representado por uma matriz de adjacência \mathbf{Y} da seguinte forma : n nós e um conjunto de arestas $Y_{i,j} : i = 1, 2, 3, \dots, n; j = 1, 2, 3, \dots, n$, no qual $Y_{i,j} = 1$ se o par de nós (i, j) está conectado e $Y_{i,j} = 0$, se não estão conectados. *Loops* não são permitidos ($Y_{i,i} = 0$). O modelo ERGM representa a distribuição conjunta de todas as arestas do grafo e a distribuição de probabilidade de \mathbf{Y} com a função de verossimilhança :

$$\pi(\mathbf{y}|\theta) = \frac{q_{\theta}(\mathbf{y})}{z(\theta)} = \frac{\exp\{\theta^t s(\mathbf{y})\}}{z(\theta)} \quad (4.3)$$

onde, $s(\mathbf{y})$ é um vetor conhecido de estatísticas suficientes, como número de arestas, $\theta \in \Theta$ é um vetor de parâmetros do modelo, no qual um valor positivo para θ_i resulta em uma maior probabilidade de se observar a estatística $s(\mathbf{y})$ nos grafos, por fim, $z(\theta) = \sum_{\mathbf{y} \in Y} \exp\{\theta^t s(\mathbf{y})\}$ a constante de normalização.

4.1 PACOTE ERGM

Para explicar e entender melhor os métodos e técnicas utilizadas neste trabalho, vamos utilizar um pacote impletando no *software R* o pacote *ergm*, da equite *statnet*, que fornece ferramentas para modelar redes com base nos modelos de grafos aleatórios exponenciais. O pacote possui ferramentas para poder realizar três importantes tarefas envolvendo os modelos de grafos aleatórios exponenciais, são elas estimativa, simulação e bondade de ajuste (Hunter et al 2008). Para ser mais exato, o pacote tem a capacidade de aproximar um estimador de máxima verossimilhança para um modelo dado o conjunto de dados de uma rede simulando um novo conjunto de dados de uma rede através de cadeias de Markov via Monte Carlo e avaliando o quão bem o modelo ajustado abrange o conjunto de características reais da rede observada.

Um modelo ERGM é construído seguindo a seguinte nomenclatura:

$$y \sim \langle term1 \rangle + \langle term2 \rangle + \dots$$

No qual y é um objeto de rede ou uma matriz que pode ser interpretada como uma rede, $\langle term1 \rangle$, $\langle term2 \rangle$ e etc são termos escolhidos, como por exemplo arestas, e atributos da rede.

- **Efeito dos atributos :**

- **nodefactor (attrname, base=1):** Efeito principal de um atributo que está na classe fator : o argumento *attrname* é um atributo categórico na lista de características dos vértices da rede, este termo adiciona várias estatísticas da rede ao modelo, para cada um dos subconjuntos do atributo *attrname*. O argumento *base* conta quais os valores (numerados em ordem de acordo com a função de classificação) deve ser omitido, significa que a categoria base, ou definida como padrão.
- **nodematch(attrname, diff=FALSE, keep=NULL):** Efeito de homofilia uniforme e homofilia diferencial : O argumento *attrname* é um atributo na lista de características dos vértices da rede. Quando *diff=FALSE*, este termo adiciona uma estatística de rede ao modelo, que conta o número de arestas (i, j) para qual o atributo escolhido $attrname(i) == attrname(j)$, quando temos *diff=TRUE*, p estatísticas de rede são acrescentadas ao modelo, onde p é o número de valores únicos do atributo.
- **nodemain(attrname) ou nodecov:** Efeito principal de uma covariável : O argumento *attrname* é um atributo quantitativo (não categórico) na lista de características dos vértices da rede, este termo adiciona uma única estatística de rede ao modelo igualando a soma de $attrname(i)$ e $attrname(j)$ para todas as arestas (i, j) .

4.2 BONDADDE DE AJUSTE

Em qualquer tipo de problema de modelagem, o melhor ajuste é aquele que melhor se adequa os dados, porém, pode acontecer de a própria classe de modelos não contiver um conjunto de modelos suficientemente rico para escolher. O conceito de bondade de ajuste é, importante. Mas, embora este conceito seja bem desenvolvido em contextos de modelagem padrão, como a modelagem linear, é indiscutivelmente ainda nova (primitiva) para a modelagem de grafos.

Para ERGMs, atualmente na avaliação de bondade de ajuste consiste em primeiro simular inúmeros grafos aleatórios do modelo ajustado e, em seguida, comparar as características desses grafos com as do grafo originalmente observado. Exemplos de tais características temos várias medidas de resumo referente a estrutura da rede, como grau, centralidade e distância geodésica.

Se as características do grafo simulado de acordo com a rede observada tem uma correspondência muito fraca com os valores típicos decorrentes das realizações do modelo de grafos aleatórios ajustado, então há indícios, portanto, de falta de qualidade, indicando que o modelo não está bem ajustado. Para avaliar a qualidade do ajuste do nosso modelo,

conforme ajustado pelo *ergm*, a função *gof* executa a simulação *Monte Carlo* necessária e calcula as comparações com o grafo sobre a rede original em termos de distribuição de graus, comprimento geodésico e parceiros compartilhados por arestas (ou seja, o número de vizinhos compartilhados por um par de vértices que definem uma aresta) (Kolaczyk e Csárdi, 2014).

5 RESULTADOS E DISCUÇÕES

5.1 ANÁLISE EXPLORATÓRIA DOS DADOS

Após feita a padronização das evocações, como descrito na 3.2.2, temos 106 (cento e seis) evocações únicas padronizadas, dentre as expressões manifestadas temos sentimentos negativos, sintomas, prevenções e danos provocados pela doença. Do total de 675 evocações possíveis, o processo gerou 673 evocações padronizadas, expressadas por 135 entrevistados. A Tabela 1 apresenta as evocações padronizadas e as frequências em que foram expressadas de acordo com o banco de dados utilizado neste trabalho.

Tabela 1 – Evocações únicas

Frequência	Evocações únicas padronizadas		
$N = 1$	Absence of symptoms	Avoid Pregnancy	Bitter taste
	Care during pregnancy	Chikungunya	Cough
	Cure	Desperate Families	Destruction
	Devastation	Dirty Water	Female partner
	Fight the epidemics	Flower vase	Government
	Hard	Health Consequences	Hospital cot
	Hospital Stay	House	Inept Doctors
	Junk Yard	Lack of Appetite	Lack of Prevention
	Liters	Low imunity	Medical Leave
	Miss Work	Mosquito Bite	Mouth Sores
	News	Population's fault	Proliferation
	Protection	Public health	Qualified Health Professionals
	Sadness	Social Mobilization	Social Problem
	Sore Throat	Spray	Sweat
	Tolerable	Treatment	Vaccine
	Virus	Waiting Line	Water
$N = 2$	Bed	Blood	Heat
	Lack of Committment	Physician	Problem
	Repellent	Rest	Sanitation
$N = 3$	Fight the Epidemics	Mosquito Breeding Site	Public Health Clinic
	Shivery	SUS	Trash
	Vomit		
$4 \leq N < 10$	Discarted Tire	Eye Pain or Infection	Health
	Public Health	Something Bad	Suffering
	Intravenous Rehydration	Pregnant Woman	Caos
	Contagious	Cost	Epidemics
	Fear	Symptoms	Awareness
	Bloated	Medicine	Danger
	Dengue	Joint Pain	Headache
$N \geq 10$	Hospital	Malaise	Prevention
	Weakness	Itch	Rash
	Microcephaly	Standing Water	Fever
	Disease	Mosquito	Pain

Note que existem muitas evocações com baixa frequência, isto se dá porque estas expressões são evocações muito específicas, não se encaixando em nenhuma categoria de

evocações padronizadas ou são evocações que constantemente geram *loops*.

5.2 INFORMAÇÕES GERAIS DO GRAFO REPRESENTATIVO DA REDE DE PENSAMENTO COLETIVO

Como discutido na seção Metodologia, o grafo representativo da rede de pensamento coletivo foi construído abordando cada indivíduo como um vértice, que se conecta a outro indivíduo ao compartilhar com ele pelo menos uma evocação. O peso dessa evocação é definido pelo coeficiente de afinidade apresentado na Equação 3.1. A Figura 12 apresenta o histograma da distribuições de graus e graus ponderados dos vértices, respectivamente.

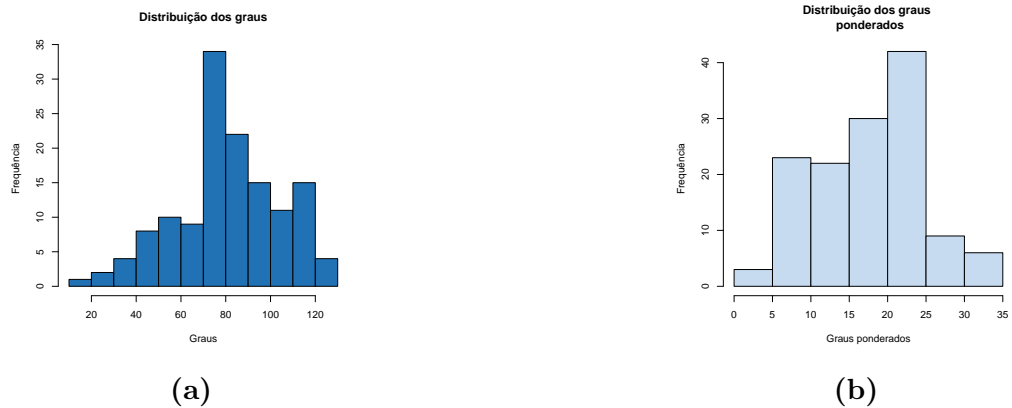


Figura 12 – Distribuição dos graus (a) e dos graus ponderados (b)

A Figura 12 mostra de forma explícita a diferença entre as distribuições dos graus e dos graus ponderados. Para a distribuição dos graus mostrada em 12a, observou-se maiores frequências entre 80 e 90. O histograma sugere que a distribuição dos graus é assimétrica positiva. Por outro lado, para a distribuição dos graus ponderados explicitada em 12b, observou-se maiores frequências entre 20 e 25. O histograma sugere que a distribuição dos graus ponderados é assimétrica negativa. Aparentemente, os critérios pré-estabelecidos impactam a distribuição da atividade dos indivíduos na rede.

Para melhor entendimento da atuação do coeficiente de afinidade, a Tabela 2 apresenta os cinco vértices que apresentaram os menores graus ponderados na rede de pensamento coletivo, bem como as evocações expressadas pelos indivíduos correspondentes.

Tabela 2 – Indivíduos menos ativos na rede

Indivíduo	Grau P.	PALA01	PALA02	PALA03	PALA04	PALA05
35	2,477	Fear	Caos	Cost	—	—
141	4,749	Cough	Join Pain	Shevery	Weakness	Malaise
21	4,977	Epidemics	Government	Hospital	Public Health	Health Profes.
77	5,257	Standing Water	Open Tanks	Discarted tire	Trash	Liters
37	5,803	Weakness	Join Pain	Something Bad	Itch	Tolerable

A Tabela 2 mostra indivíduos que, em geral, expressaram evocações de baixa frequência, como pode ser visto na Tabela 1. Este fator justifica a baixa atividade destes indivíduos na rede.

A Tabela 3 apresenta os cinco vértices que apresentaram os maiores graus ponderados na rede de pensamento coletivo, bem como as evocações expressadas pelos indivíduos correspondentes.

Tabela 3 – Indivíduos mais ativos na rede

Indivíduo	Grau Ponderado	PALA01	PALA02	PALA03	PALA04	PALA05
128	30,73	Standing Water	Mosquito	Desiase	Pain	Hospital
147	30,97	Pain	Mosquito	Itch	Fever	Medical Leave
92	31,59	Pain	Desiase	Mosquito	Mosquito Bite	Fever
140	31,80	Standing Water	Pain	Mosquito	Fever	Malaise
40	31,91	Fever	Pain	Malaise	Mosquito	Desiase

A Tabela 3 mostra indivíduos que, em geral, expressaram evocações de alta frequência, como pode ser visto na Tabela 1. Ademais, os indivíduos tem evocações em posições similares e prioritárias como Mosquito e alguns sintomas como dor e febre, que aparece em todos os cinco indivíduos listados.

A densidade do gráfico representativo da rede do pensamento coletivo é 0.7054. Isto indica que 70,54% de todos os possíveis pares de vértices se conectam, mostrando coesão do pensamento coletivo.

5.3 DECOMPONDO A REDE EM COMUNIDADES E ANALISANDO A QUALIDADE DA PARTIÇÃO ENCONTRADA

Como discutido anteriormente na subseção 3.3, o algoritmo proposto por Blondel (2008) geraram uma partição composta por 3 comunidades. Tal partição otimizou a modularidade em 0,24, indicando que os grupos gerados não tem uma grande propensão a formarem aqueles grupos. A Tabela 4 apresenta o número de vértices que compõem cada uma das 3 comunidades.

Tabela 4 – Número de vértices componentes das comunidades

Comunidade	1	2	3	Geral
Tamanho	25	53	57	135

5.3.1 VERIFICANDO A PARTIÇÃO ENCONTRADA

Conforme discutido na subseção 3.3, a partir do grafo representativo gerado para a rede de pensamentos coletivos e segundo a escolha do método de agrupamento da rede

em comunidades, no intuito de verificar a aleatoriedade da rede realizou-se a geração de réplicas de Monte Carlo do grafo original, onde em uma primeira simulação fixou-se tanto a distribuição (sequência) de graus ponderados e em uma segunda simulação fixou-se o tamanho do grafo. Em seguida, distribuiu-se as arestas aleatoriamente, respeitando as condições pré-fixadas. Foram geradas e 100.000 réplicas para cada simulação. Por fim, cada réplica foi decomposta segundo o mesmo método de agrupamento pelo qual se decompôs a rede geral e observou-se o número de comunidades resultante de cada decomposição, explicitados no gráfico da Figura 13.

A Figura 13 mostra que, em geral são esperados em grafos aleatórios com as respectivas restrições impostas 3 comunidades. O número de comunidades encontrado na rede geral demonstra uma alta probabilidade de ser observado em grafos aleatórios. Logo, o gráfico sugere indícios de existência de um processo latente aleatório para o agrupamento dos vértices observado na decomposição da rede geral.

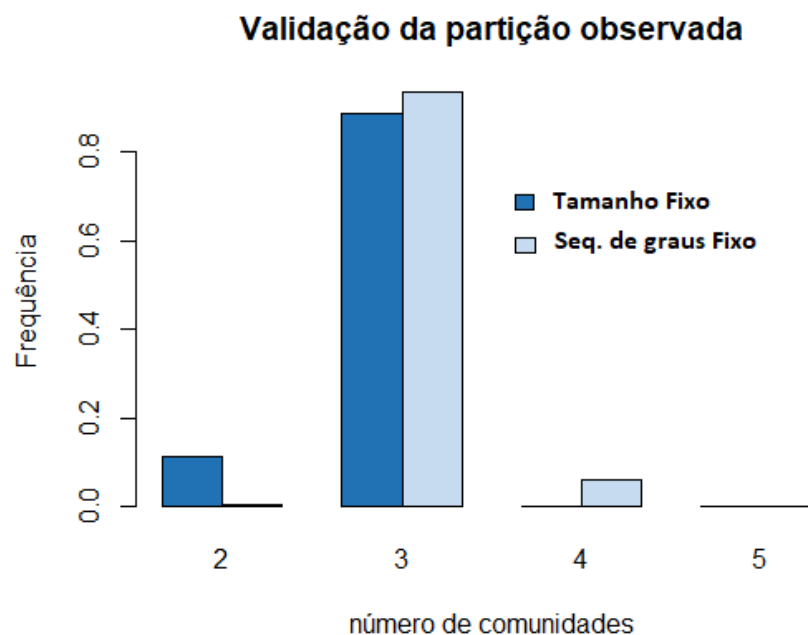


Figura 13 – Distribuição do número de comunidades - Método de Monte Carlo - 100.000 réplicas

O ideal seria que a partição encontrada pela nossa rede seja diferente da encontrada pela rede aleatória, o que não é observado. Temos no geral 3 comunidades para a rede observada e 3 comunidades para as redes aleatórias, isto sugere que a partição encontrada pode ter ocorrido de forma aleatória, mas não podemos tirar conclusões precipitadas.

5.4 COMPARANDO OS PENSAMENTOS COLETIVOS

Após encontrar a partição que melhor representa as heterogeneidades da rede como visto na seção 3.3, encontramos 3 comunidades para a rede observada, são elas

Comunidade	Tamanho	Densidade
1 Tratamento	25	0.7966
2 Prevenção	53	0.9064
3 Sintomas	57	0.9041

Tabela 5 – Número de indivíduos e densidade das comunidades

Temos que, todas as comunidades encontradas apresentam alta densidade, ou seja, o pensamento dentro das comunidades é bem difundido. Para comunidade 1 referente a tratamento, temos a menor densidade dentre as 3 encontradas, com 79,66% de coersão.

Para representarmos graficamente o pensamento geral da rede e das comunidades de forma simples, temos na tabela 6 as evocações únicas agrupadas de forma coerente com seus respectivos sentidos. Os subgrupos foram formados com a finalidade de reduzir o número de evocações e facilitar a visualização.

Tabela 6 – Divisão em subgrupos das evocações únicas

Grupos	Evocações únicas padronizadas		
Symptoms(SYM)	Vomit	Headache	Tolerable
	Absence of symptoms	Bed	Sweat
	Bitter taste	Bloated	Cough
	Eye Pain or Infection	Fever	Weakness
	Health Consequences	Itch	Joint Pain
	Lack of Appetite	Malaise	Mouth Sores
	Pain	Rash	Rest
	Shivery	Sore Throat	Symptoms
Etiology (ETI)	Water	Virus	Mosquito Bite
	Chikungunya	Mosquito	Dengue
	Heat	Disease	Blood
	Proliferation	Contagious	Low immunity
Government & social causes (GSC)	Samarco disaster	Government	Lack of Commitment
	Destruction	Sanitation	Social Problem
	Protect the Environment		
Prevention (PRV)	Repellent	Prevention	Protection
	Spray	Care during pregnancy	
	Health	Public Health	Vaccine
	Awareness	Social Mobilization	Fight the epidemics
Chaos (CHA)	Caos	Danger	Desperate Families
	Epidemics	Fear	Hard
	News	Sadness	Something Bad
	Problem	Suffering	Devastation
Mosquito Breeding site (MBS)	Ceiling	Dirty Water	Discarded Tire
	Flower vase	House	
	Junk Yard	Liters	Mosquito Breeding Site
	Open Water Tanks	Standing Water	Trash
Treatment (TRT)	Qualified Health Professionals	Medicine	Cure
	Hospital	Inept Doctors	SUS
	Intravenous Rehydration	Treatment	Hospital Stay
	Physician	Waiting Line	Doctor Visit
Population's Fault (PPF)	Population's fault	Lack of Prevention	
Vulnerable Groups (VLG)	Pregnant Woman	Female partner	Baby
	Child		

O agrupamento é feito para que possamos montar uma fácil visualização do pensamento coletivo da rede, que neste caso, será feito através do gráfico de radar. O gráfico de radar é uma forma de representar o pensamento da rede. Ele é gerado levando em conta o número médio de evocações e suas frequências. Para montar os gráficos utilizamos o pacote Wordevok de Pereira (2017). O gráfico de radar consegue mostrar de forma clara e simples onde está concentrado o pensamento coletivo dos indivíduos observados. Quanto mais próximo o pensamento está da borda, mais aquele grupo pensa naquele determinado termo. Para a figura 14 temos que a rede geral comumente pensa em Sintomas, Etiologia e Consequências.

Radar do pensamento coletivo da rede geral

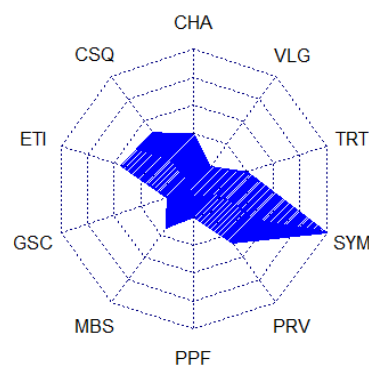


Figura 14 – Gráfico de radar do pensamento coletivo da rede

O pensamento dentro das comunidades se diferem do pensamento geral da rede. Então, temos que para a comunidade 1, os indivíduos usualmente pensam em Tratamento, Prevenção e Etiologia, logo chamamos esta comunidade de Tratamento; a comunidade 2, concentra o pensamento em termos como Prevenção, Etiologia, consequência e Focos do Mosquito, logo, nós a chamamos de Prevenção e por fim, a comunidade 3 em geral pensa em Sintomas e Etiologia e é chamada de Sintomas. Os nomes dados às comunidades é para que possamos identificá-las mais facilmente.

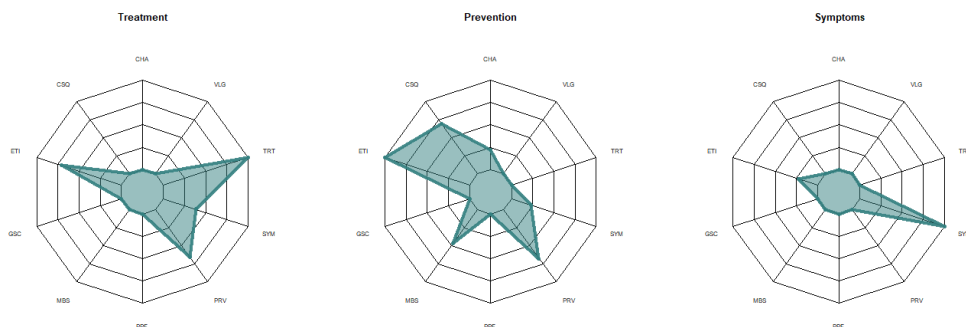


Figura 15 – Gráfico de radar do pensamento coletivo das comunidades

5.5 AJUSTE DE MODELO

Para analisar e explicar a probabilidade de ocorrência das conexões por afinidade cognitiva entre os indivíduos, usamos atributos epidemiológicos (exposição pessoal ao Zika (Ego) e contato com outra pessoa segundo exposição ao vírus (Alter)) e sociodemográficos (sexo e escolaridade), utilizamos os modelos de grafos aleatórios exponenciais (ERGM). Para o modelo ajustado temos os resultados na tabela 7, que nos fornece os coeficientes e suas respectivas estimativas. Para uma fácil interpretação também apresentamos a razão de chance dos coeficientes.

Coeficiente	Estimativa	Razão de Chance
Edge	-0.4114 ***	1.52
Comunidade Prevention	-0.3216 ***	1.38
Comunidade Symptons	-0.1154*	1.12
Escolaridade Média	0.6162 ***	1.85
Escolaridade Alta	0.38938 ***	1.47
Sexo	0.1481 ***	1.16
Alter(Nenhuma Doença)	0.5219***	1.68
Alter(Dengue ou Chikungunya)	-0.39756*	1.48
Alter(Zika)	0.5354***	1.71
Ego	0.39742***	1.48

Tabela 7 – Estimativa e Razão de Chance para os coeficientes do modelo ERGM

Os resultados do modelo ERGM de Bernoulli sugerem uma elevada densidade na rede observada, em que 52% das relações possíveis entre as pessoas por afinidade cognitiva foram realizadas. Exposição ao vírus também se mostrou importante para explicar a afinidade cognitiva; duas pessoas com o mesmo nível de exposição ao ZIKV têm mais chance de pensarem parecido do que pessoas com diferentes níveis de exposição ($RC = 1.49$). O efeito de exposição de pessoas conhecidas sobre afinidade cognitiva foi significativo; quem teve a experiência com o vírus apenas pela mídia ($RC = 1.69$) ou tendo sido infectado pelo ZIKV ($RC = 1.71$) têm maior chance de pensarem parecido entre si, ao passo que quem teve DNG ou CKG ($RC = 1.49$) tem menos chance de afinidade cognitiva com outros com o mesmo nível de exposição. Isso reforça o papel homogeneizador da mídia e do contato direto com a doença. Por fim, pessoas de mesmo sexo têm maior chance de pensarem parecido ($RC = 1.16$), bem como as pessoas de maior escolaridade ($RC = 1.48$).

5.5.1 BONDADDE DE AJUSTE

A bondade do ajuste como definido na seção 4.2 é feita por simulação de Monte Carlo via Cadeia de Markov (MCMC), no qual fixamos características do grafo observado para gerar o grafo aleatório. De acordo com o modelo ajustado, conseguiremos comparar

como se comporta o ajuste para um grafo aleatório e o ajuste para o grafo observado. A figura 16 apresenta o bondade de ajuste para o modelo

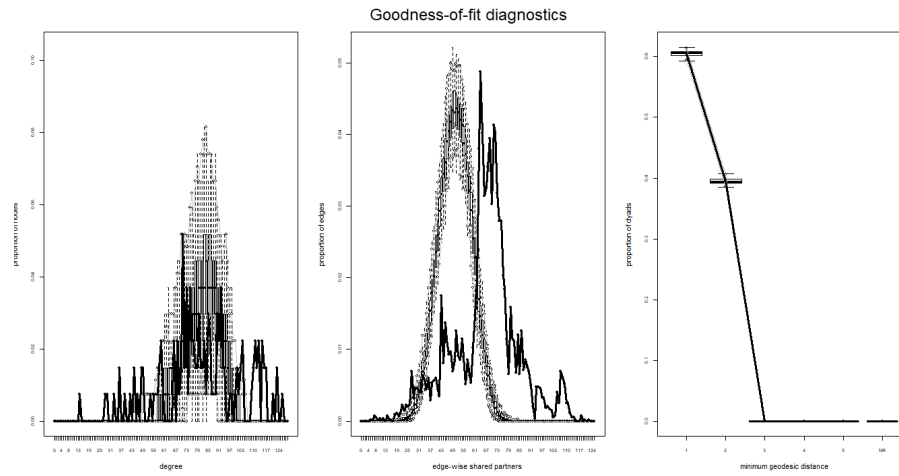


Figura 16 – Bondade de Ajuste

Na figura 16, temos 3 gráficos que nos auxiliam na conclusão da qualidade do ajuste do modelo. No primeiro gráfico apresentado, temos a bondade de ajuste para os graus médios da rede, podemos perceber que o modelo em média consegue capturar bem o grau dos vértices, mas para vértices com graus muito altos ou muito baixos o modelo não consegue captar bem o comportamento, isto pode ser porque temos uma rede muito densa. Já para o segundo gráfico, temos os parceiros em comum ou parceiros compartilhados, nele podemos perceber em média quantos parceiros geralmente são compartilhado na rede, podemos inferir que em média a rede observada compartilha 65 parceiros e a rede aleatória em média 55. Ou seja, a rede é bem mais conectada que o esperado, como já informamos anteriormente. Por fim, temos o gráfico que nos apresenta a distância geodésica, neste gráfico podemos perceber que temos grandes concentrações de distâncias pequenas, isto é, para transmitir informação temos que percorrer caminhos curtos. É utilizado a distância geodésica pois o grafo não é uma estrutura plana.

Claramente podemos perceber que temos problemas com o ajuste do modelo, já que ele não consegue captar bem o comportamento da rede observada. Isto se dá, devido as limitações do modelo ERGM, que tem vários problemas de convergência quando incluímos algumas estatísticas endógenas da rede no modelo. Um exemplo destas estatísticas são os triângulos e estrelas de ordem k . A falta dessas estatísticas para o modelo pode ocasionar a falta de ajuste.

6 CONCLUSÃO

No decorrer deste trabalho foi discutida a união de várias metodologias difundidas no estudo das representações de objetos sociais com uma abordagem mais recente neste campo de estudo, a representação através de redes, grafos. Como mostrado, a união dessas técnicas concede às representações sociais possibilidades muito interessantes, principalmente no que se refere a encontrar heterogeneidades no pensamento coletivo da sociedade e entender a formação do pensamento coletivo.

Através dos resultados obtidos, foi possível perceber algumas diferenças nos focos de pensamentos entre as 3 comunidades encontradas, isto é, conseguimos captar as heterogeneidades de pensamento dentre elas, pois, foram encontradas diferenças nas prioridades que cada comunidade têm ao dar significado para o Zika vírus, embora a modularidade tenha se mostrado baixa.

Os resultados encontrados neste estudo através do modelo ERGM ajustado, nos ajuda a quantificar o quanto as cováriaveis influenciam a densidade da afinidade cognitiva da rede observada e na probabilidade de conexão.

Por fim, o estudo leva a conclusão de que conseguimos perceber que existe diferenças evidentes na perspectiva que as pessoas têm ao significar Zika vírus. Exemplificando, existem alguns subgrupos da sociedade que buscam identificar os sintomas da doença enquanto outros subgrupos se preocupam com a prevenção e tratamento. A união das perspectivas de cada um desses subgrupos formam o pensamento coletivo da sociedade.

7 REFERÊNCIAS

- PEREIRA, WESLEY, Representação da Estrutura do Pensamento Coletivo sobre as Enchentes do Rio Doce: Conectando indivíduos e afins através da teoria dos grafos, 2017.
- DAVID HUNTER, et. al. ERGM: A package of fit, Simulate and diagnose Exponential Family Models for Networks, *Journal of statistical software*, vol.24, 2008
- ABRIC, J.-C. Jeux, conits et representations sociales. Tese (Doutorado), 1976.
- ABRIC, J. C. Pratiques sociales et representations. In: . Paris: Presses Universitaires de France, 1994. cap. Las representations sociales: aspects theoriques.
- ABRIC, J. C. Representations of the social: Bridging theoretical traditions. In: . Malden, UK: Blackwell Publishing, 2001. cap. A structural approach to social representations, p. 4247.
- BACHRACH, C. A. Culture and demography: From reluctant bedfellows to committed partners. *Demography*, Springer, v. 51, n. 1, p. 325, 2014.
- BLONDEL, V. D. et al. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, IOP Publishing, v. 2008, n. 10, p. P10008, 2008.
- CSARDI, G.; NEPUSZ, T. The igraph software package for complex network research. *InterJournal, Complex Systems*, p. 1695, 2006. Disponível em: <<http://igraph.org>>.
- HORA, A. M. Representacoes sociais da comunidade de Tumiritinga sobre o Rio Doce e a Prainha do Jao - Dissertacao de Mestrado. [S.l.]: Governador Valadares: Universidade Vale do Rio Doce, 2013.
- KOLACZYK, E. D. Statistical analysis of network data - methods and models. [S.l.]: Springer, 2009.
- MOSCOVICI, S. La psychanalyse, son image et son public. Paris: Presses Universitaires de France, 1961.
- NAKAZAWA, M. fmsb: Functions for Medical Statistics Book with some Demographic Data. [S.l.], 2015. R package version 0.5.2. Disponível em: <<https://CRAN.R-project.org/package=fmsb>>.
- NEWMAN, M. E. Finding community structure in networks using the eigenvectors of matrices. *Physical review E, APS*, v. 74, n. 3, p. 036104, 2006.
- NEWMAN, M. E.; GIRVAN, M. Finding and evaluating community structure in networks. *Physical review E, APS*, v. 69, n. 2, p. 026113, 2004.
- RStudio Team. RStudio: Integrated Development Environment for R. Boston, MA, 2015. Disponível em: <<http://www.rstudio.com/>>.
- SA, C. P. d. Representações sociais: teoria e pesquisa do núcleo central. *Temas em Psi-*

cologia, Sociedade Brasileira de Psicologia, v. 4, n. 3, p. 1933, 1996.

8 ANEXO

ENTREVISTA SEMIESTRUTURADA	
8) Coleta de evocações: Agora, você deverá dizer as cinco primeiras coisas que vêm à sua cabeça quando você ouve a expressão Zika Vírus:	
<div style="background-color: black; color: white; padding: 5px; margin-bottom: 10px; font-weight: bold;">ZIKA VÍRUS</div> <div style="border: 1px solid black; height: 100px; margin-top: 10px; position: relative;"> <div style="position: absolute; bottom: 10px; right: 10px;"> <div style="border-bottom: 1px solid black; width: 80%;"></div> <div style="border-bottom: 1px solid black; width: 80%;"></div> <div style="border-bottom: 1px solid black; width: 80%;"></div> <div style="border-bottom: 1px solid black; width: 80%;"></div> <div style="border-bottom: 1px solid black; width: 80%;"></div> </div> </div>	<div style="margin-bottom: 5px;">()</div> <div style="margin-bottom: 5px;">()</div> <div style="margin-bottom: 5px;">()</div> <div style="margin-bottom: 5px;">()</div> <div style="margin-bottom: 5px;">()</div>
99[] Nunca ouviu falar em Zika Virus → (Pule para Q11)	
9) Palavras mais importantes Agora, você deverá ordenar as palavras ou expressões que você falou pela ordem de importância. [<i>ENTREVISTADOR: marque a ordem de importância preenchendo os parênteses com a numeração de 1 (a mais importante) até 5 (a menos importante).</i>]	
10) Significados Agora, você deverá indicar o significado que a palavra/expressão que você elegeu como nº 1, tem para você e porque você elegeu a palavra / expressão (nº 1) como a mais importante?	
<div style="border-bottom: 1px solid black; height: 20px; width: 100%;"></div> <div style="border-bottom: 1px solid black; height: 20px; width: 100%;"></div>	

Figura 17 – Item do questionário utilizado para análises