Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Departamento de Estatística

Aplicação de métodos inferenciais a redes bipartidas no estudo do fluxo turístico Europa - Brasil

Ana Luiza Barcelos Pereira

Trabalho de Conclusão de Curso

Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Departamento de Estatística

Aplicação de métodos inferenciais a redes bipartidas no estudo do fluxo turístico Europa - Brasil

Ana Luiza Barcelos Pereira
Orientadora: Profª Drª Andressa Cerqueira

Trabalho de Conclusão de Curso apresentado como parte dos requisitos para obtenção do título de Bacharel em Estatística.

São Carlos
Fevereiro de 2025

Este trabalho dedico àqueles que, sob muito sol, me permitiram chegar até aqui pela sombra, e sonharam junto a mim. Em especial, ao meu irmão Hendryck, para quem quero sempre ser um bom exemplo.

Resumo

O presente Trabalho de Conclusão de Curso aborda um estudo aprofundado de métodos inferenciais aplicados a dados de redes bipartidas, com uma aplicação prática no conjunto de dados que registra a chegada de turistas provenientes da Europa aos vinte e seis estados brasileiros, além do Distrito Federal, exclusivamente por via aérea.

Este estudo tem como foco compreender as dinâmicas do fluxo turístico europeu para o Brasil, analisando a chegada de turistas ao longo dos semestres entre os anos de 2019 e 2022. Para isso, serão utilizados métodos inferenciais e ferramentas computacionais que permitem uma análise detalhada dos fatores que influenciam essa movimentação, identificando padrões e tendências ao longo do tempo.

A modelagem será feita a partir da construção de Redes Bipartidas, nas quais os vértices representam os países emissores de turistas e os estados brasileiros de destino, enquanto as arestas indicam a intensidade das conexões entre eles. Posteriormente, a projeção one-mode será aplicada, permitindo que as relações entre países e entre estados sejam analisadas separadamente. Para complementar a investigação, serão utilizados histogramas, um recurso estatístico simples, mas eficaz para visualizar variações e mudanças ao longo do período estudado.

Além disso, será realizada a detecção de comunidades, uma abordagem mais complexa que busca identificar grupos de vértices mais densamente conectados entre si do que com o restante da rede. Essa análise tem o potencial de revelar padrões de interação que não são perceptíveis pelos métodos convencionais, oferecendo uma compreensão mais profunda das relações turísticas entre a Europa e o Brasil.

Dessa forma, este trabalho visa contribuir tanto para a área de análise de redes bipartidas e detecção de comunidades quanto para a compreensão do turismo internacional no Brasil antes, durante e após a pandemia da COVID-19.

Palavras-chave: *fluxo turístico, projeção, redes bipartidas.*

Abstract

The present Undergraduate Thesis explores an in-depth study of inferential methods applied to bipartite network data, with a practical application to a dataset that records the arrival of tourists from Europe to the twenty-six Brazilian states, as well as the Federal District, exclusively by air travel.

This study focuses on understanding the dynamics of European tourist flows to Brazil by analyzing the arrival of tourists over the semesters between the years 2019 and 2022. To achieve this, inferential methods and computational tools will be employed to provide a detailed analysis of the factors influencing this movement, identifying patterns and trends over time.

The modeling will be conducted through the construction of Bipartite Networks, in which the vertices represent the countries sending tourists and the Brazilian states of destination, while the edges indicate the intensity of the connections between them. Subsequently, the one-mode projection will be applied, allowing the relationships between countries and between states to be analyzed separately. To complement the investigation, histograms will be used—a simple yet effective statistical tool for visualizing variations and changes over the studied period.

Additionally, community detection will be performed, a more complex approach aimed at identifying groups of vertices that are more densely connected among themselves than with the rest of the network. This analysis has the potential to reveal interaction patterns that are not perceptible through conventional methods, providing a deeper understanding of the tourism relationships between Europe and Brazil.

Thus, this study aims to contribute both to the field of bipartite network analysis and community detection and to the understanding of international tourism in Brazil before, during, and after the COVID-19 pandemic.

Keywords: *tourist flow, projection, bipartite networks.*

Lista de Figuras

Sumário

Capítulo 1

Introdução

O turismo consiste em um conjunto de atividades realizadas por um indivíduo em locais distintos daquele onde ele reside. Em geral, as atividades turísticas estão associadas ao lazer, mas também podem ocorrer por motivos de negócios. Durante a pandemia global da COVID-19, o setor turístico foi severamente afetado, uma vez que grande parte das cidades ao redor do mundo implementou restrições rígidas de entrada para conter a propagação do vírus. No entanto, à medida que os meses passaram e a vacinação avançou, muitas fronteiras foram reabertas, permitindo a retomada gradual da atividade turística, ainda que sob novas regulamentações e restrições.

Neste contexto, o presente estudo busca compreender os impactos da pandemia no fluxo turístico entre a Europa e o Brasil, considerando exclusivamente turistas europeus que chegaram ao país por via aérea. Para isso, será analisado o fluxo de passageiros ao longo dos semestres entre os anos de 2019 e 2022, permitindo a observação de padrões antes, durante e após a pandemia.

A fim de estruturar essa análise, utilizamos o conceito de redes, que representam sistemas compostos por elementos individuais interconectados, formando um padrão de conexões. Redes são amplamente aplicadas em diversos campos, como ciências sociais, biologia e tecnologia, sendo a internet um dos exemplos mais conhecidos. Estruturalmente, as redes são compostas por vértices (ou nós), que representam os elementos do sistema, e por arestas (ou conexões), que indicam as interações entre eles. Um exemplo clássico de aplicação está na biologia, por meio das redes neurais, onde os vértices correspondem aos neurônios e as arestas representam as sinapses nervosas.

A análise e visualização de redes se tornam essenciais para a compreensão dos padrões de conexão. No entanto, a identificação desses padrões a olho nu se torna inviável quando

o número de vértices é elevado. Para contornar essa limitação, utilizamos medidas de centralidade, que quantificam a importância dos vértices na rede. Uma dessas medidas é o grau de um vértice, definido pelo número de arestas a ele conectadas. Quanto maior o grau de um vértice, maior sua relevância dentro da rede, tornando essa métrica um guia importante para a análise estrutural do sistema.

No presente estudo, a modelagem do fluxo turístico europeu para o Brasil será feita por meio de redes bipartidas. Nessa abordagem, os vértices representarão os estados brasileiros e os países europeus, enquanto as arestas indicarão o número de turistas que viajaram de um determinado país europeu para um estado brasileiro. A análise será realizada considerando a estrutura das conexões entre os dois grupos, além da projeção one-mode, que permitirá estudar separadamente as relações entre os países emissores e entre os estados receptores. Para complementar a investigação, histogramas serão utilizados para visualizar variações e mudanças no fluxo turístico ao longo do período analisado.

A estrutura deste Trabalho de Conclusão de Curso está organizada da seguinte forma: o Capítulo 2 introduzirá o conceito de redes, apresentando a teoria fundamental e exemplos de diferentes formas de grafos, juntamente com suas matrizes de adjacência correspondentes. Além disso, abordará tópicos como redes ponderadas, redes direcionadas e a importância dos graus dos vértices na estrutura das redes. Esse capítulo é essencial para fornecer a base teórica necessária à compreensão dos capítulos seguintes. O Capítulo 3 será dedicado ao estudo das redes bipartidas, explicando sua representação por meio da matriz de incidência e introduzindo o conceito de projeção one-mode. Essa técnica consiste na transformação matemática da rede bipartida em duas redes simples, onde as conexões passam a ser analisadas dentro dos próprios grupos (países europeus e estados brasileiros). Além disso, este capítulo abordará a teoria de detecção de comunidades, um método avançado utilizado para identificar grupos de vértices mais densamente conectados entre si do que com o restante da rede. Como as redes em estudo possuem grande volume de dados, utilizaremos conceitos como a matriz laplaciana para particionar os vértices de forma eficiente. O Capítulo 4 descreverá o conjunto de dados que será analisado, bem como a aplicação prática das teorias abordadas nos capítulos anteriores. Esse estudo será conduzido por meio do software igraph, reconhecido por sua eficiência na análise de redes complexas, permitindo a aplicação de algoritmos avançados para a detecção de comunidades. Esse pacote é detalhado em Csardi e Nepusz (2006) e amplamente utilizado em pesquisas da área. Além disso, outro material de referência será Li *et al.* (2017), que

apresenta explicações sobre pacotes adicionais utilizados no estudo dentro do ambiente RStudio.

Com isso, este trabalho busca não apenas contribuir para o avanço da análise de redes bipartidas e detecção de comunidades, mas também fornecer uma visão detalhada sobre a evolução do turismo internacional no Brasil antes, durante e após a pandemia, auxiliando na compreensão das mudanças no comportamento dos turistas europeus e dos impactos gerados pela crise sanitária no setor.

Capítulo 2

Introdução sobre redes

Neste capítulo apresentaremos uma introdução ao campo das redes. A teoria que será apresentada é fundamentada principalmente no livro (Newman, 2018), que serve como um guia fundamental para o estudo das redes e de como desempenham um papel muito importante em diversas áreas do conhecimento.

2.1 Redes

As redes ou grafos, tema tratado nesta monografia, são estruturas matemáticas que representam a relação entre objetos. Um grafo é um conjunto de vértices ou nós e um conjunto de arestas que conectam esses vértices. Os nós representam objetos e as arestas as conexões entre eles. Um grafo pode ser representado pela Figura 2.1.

O interesse em usar redes surge na necessidade de representação de informações as quais são agrupadas e ligadas entre os diferentes grupos e internamente em cada grupo, isso faz com que seja possibilitado o estudo do todo e também do indivíduo. Há uma terceira vertente de estudo e a mais utilizada nesses casos que é o padrão de conexões no sistemas, nas quais os elementos do sistema são vértices da rede e as conexões são as arestas, tais padrões esses que podem ter grandes efeitos no comportamento do sistema.

Para realizar a análise de uma rede é necessário que se entenda a estrutura, pois caso contrário não será possível que se tenha o entendimento preciso de como tal sistema funciona. A rede é a redução do sistema, no qual só é capturado o básico dos padrões, fazendo com que muitas informações sejam perdidas no processo de redução sistêmica, isso tem suas desvantagens e vantagens. Algumas vantagens são:

- Simplicidade: A representação em redes permite simplificar um sistema complexo,

dividindo-o, isso torna mais fácil entender e visualizar a estrutura e os relacionamentos entre os elementos do sistema.

- Visualização intuitiva: Os grafos oferecem uma maneira intuitiva de visualizar um sistema, permitindo assim que padrões sejam identificados.

- Análise de conectividade: Ao reduzir um sistema, é possível analisar a conectividade entre os elementos, revelando assim a importância de nós.

- Modelagem flexível: Os grafos oferecem uma estrutura flexível que pode ser facilmente adaptada e modificada à medida que o sistema evolui. Elementos podem ser adicionados ou removidos sem afetar a estrutura geral da rede.

E assim tem-se como algumas das desvantagens:

- Perda de detalhes: A simplificação de um sistema complexo para uma representação de rede, pode ocasionar em perda de detalhes importantes, como informações específicas sobre os componentes individuais que podem ser reduzidas ou omitidas, resultando em uma visão mais abstrata e geral do sistema.

- Limitação da estrutura de rede: A representação em redes pode não capturar adequadamente certos aspectos ou relações do sistema. Algumas interações complexas ou informações contextuais podem ser difíceis de representar,levando a uma compreensão limitada deste sistema.

- Dificuldade na representação de atributos: A representação de atributos adicionais associados a nós ou arestas em um grafo pode ser uma tarefa desafiadora. Informações quantitativas ou contextuais podem não ser facilmente integradas à estrutura do grafo, dificultando a análise e a visualização desses atributos.

Apesar das desvantagens listadas acima o uso de redes não é invalidado, porém é importante trazer tais limitações para que o analista possa considerar esses aspectos e utilizar abordagens complementares.

2.2 Matriz de adjacência

Utilizaremos a Figura 2.1 para exemplificar a matriz adjacência. A rede utilizada possui n = 5 vértices rotulados com as cinco primárias letra do alfabeto, cada um desses

rótulos dados aos vértices são únicos. Indicaremos a aresta entre os vértices 1 e 2 por (1,2), assim ao saber o valor de n e a lista de arestas será possível obter a rede completa. Na Figura 2.1, temos como lista de arestas, (A,B),(A,D),(A,E),(B,E),(C,D),(D,E). As vezes, em contextos matemáticos como os abordados neste capítulo, as listas de arestas podem ser consideradas inconvenientes, embora sejam ocasionalmente utilizadas para armazenar a estrutura de redes em computadores. 2.1:

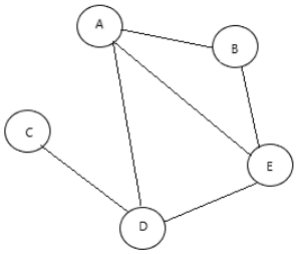


Figura 2.1: Exemplo de grafo não dirigido

Uma forma mais adequada de representar uma rede para se realizar um estudo, é através da matriz de adjacência. A matriz de adjacência da Figura 2.1 será representada por A, tal qual, suas entradas são definidas por:

$$A_{ij} = \begin{cases} 1, & \text{se existir uma aresta entre os vértices i e j.} \\ 0, & \text{caso contrário.} \end{cases}$$

E assim temos que a matriz adjacência referente ao grafo da Figura 2.1 de tamanho 5×5

$$\mathrm{A} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Para o tipo de rede que foi representada na matriz acima temos que a diagonal principal possui todos os elementos iguais a zero, já que a mesma não possui arestas próprias, a matriz também será simétrica, já que existindo um aresta que está entre i e j, logo existe uma aresta entre j e i. Vale ressaltar que para cada tipo de grafo, existe uma representação diferente de matriz de adjacência.

Existem casos em que entre dois vértices podem ter mais de uma aresta conectando-os, essas são referidas como multi arestas, uma rede que possui multi arestas é nomeada de multigrafos, como a representada na Figura 2.2.

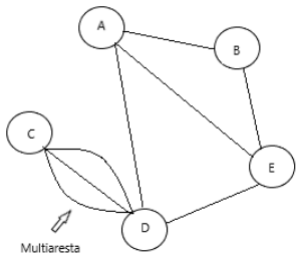


Figura 2.2: Exemplo de rede multiaresta

E assim temos que a matriz adjacência referente ao grafo da Figura 2.2

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 3 & 0 \\ 1 & 0 & 3 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

2.3 Redes ponderadas

Nesta seção trataremos de redes ponderadas, que trazem peso em suas arestas, geralmente esses valores são positivos, porém não há nenhuma razão para que não devam ser negativos. Os pesos, também assim chamado, são utilizados para acrescentar mais informações no grafo e assim tornando sua interpretação mais completa.

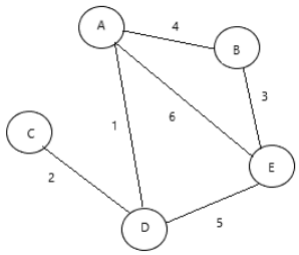


Figura 2.3: Exemplo de grafo ponderado

A rede da Figura 2.3 também podem ser representada por uma matriz adjacente, na qual os pesos são os valores de entrada da matriz.

$$A = \begin{bmatrix} 0 & 4 & 0 & 1 & 6 \\ 4 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 2 & 0 \\ 1 & 0 & 2 & 0 & 5 \\ 6 & 3 & 0 & 5 & 0 \end{bmatrix}$$

2.4 Redes direcionadas

A rede direcionada ou grafo direcionado, apresenta um direcionamento em cada aresta, indicando se as arestas estão saindo em direção a outro vértice ou entrando, as arestas que fazem essas conexões apresentam setas que indicam a direção, como exemplificado na Figura 2.4.

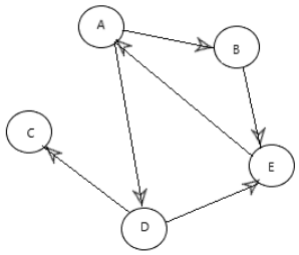


Figura 2.4: Exemplo de grafo direcionado

Para grafos deste tipo também é possível obter uma matriz de adjacência correspondente, porém assimétrica e preenchida de uma maneira não intuitiva, porém foi a maneira apresentada pelo autor do livro referência (Newman, 2018) utilizado. As entradas de tal matriz são definidas por.

$$A_{ij} = \begin{cases} 1, & \text{se existe uma aresta indo do vértice j para o vértice i.} \\ 0, & \text{caso contrário.} \end{cases}$$

Assim teremos o grafo direcionado da Figura 2.4 representado em forma de matriz.

$$\mathrm{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

2.5 Graus

Anteriormente muito se falou das arestas do grafo, porém agora, será abordado um tema relacionado aos vértices, o grau, que representa o número de arestas conectadas a um mesmo vértice. Nos casos de grafos não direcionados o grau será escrito em termos da matriz de adjacência.

$$k_i = \sum_{j=1}^{n} A_{ij} \tag{2.1}$$

Já sabemos que nos casos de grafos não direcionados cada aresta possui duas extremidades, sendo m o número de arestas, logo 2 será a quantidade de extremidades, podendo também essa quantidade ser representada como a soma dos graus dos vértices do grafo, dado por

$$2m = \sum_{i=1}^{n} k_i \tag{2.2}$$

ou

$$m = \frac{1}{2}\sum_{i=1}^{n} k_i = \frac{1}{2}\sum_{i,j=1} A_{ij} \tag{2.3}$$

Em capítulos adiante será utilizado o grau médio c de um vértice em um grafo não direcionado, representado pela expressão

$$c = \frac{1}{n}\sum_{i=1}^{n} k_i \tag{2.4}$$

a qual pode ser combinada com a Equação (2.2) e se obter

$$c = \frac{2m}{n} \tag{2.5}$$

Capítulo 3

Redes bipartidas

Neste capítulo será explicado o conceito de redes bipartida, o tema principal deste trabalho e detecção de comunidades uma técnica utilizada para que seja possível analisar de maneira mais detalhada os dados em estudo. Anteriormente trouxemos o conceito de redes para que ao chegar aqui o leitor já tenha o conhecimento básico e consiga compreender os novos conceitos trazidos aqui.

3.1 Redes bipartidas

As redes bipartidas são redes utilizadas para representar a relação entre dois grupos distintos em estudo, como por exemplo os países europeus e os estados brasileiros que serão tratados neste trabalho. Há dois tipos de vértices, o que representa os vértices originais e os vértices que secundários aos quais os originais pertencem. Nesse tipo de rede as arestas ligam apenas os diferentes tipos de vértices. No estudo desse tipo de redes o equivalente a matriz adjacência visto anteriormente, é a matriz de incidência. Para exemplificar melhor, será apresentado a matriz incidência da Figura 3.1, na qual os vértices rosas representados por $g = 3$ e os vértices azuis são $n = 4$.

Para exemplificar, suponha que os vértices rosas são grupos de estudos e os vértices azuis estudantes do curso de Estatística, assim teremos uma aresta entre o vértice azul e o vértice rosa, se um aluno (representado pelo vértice azul) fizer parte de um grupo de estudo (representado pelo vértice rosa).

Assim a matriz incidência é uma matriz $g \times n$, em que seus elementos serão 1 se o

vértice azul pertencer ao vértice rosa.

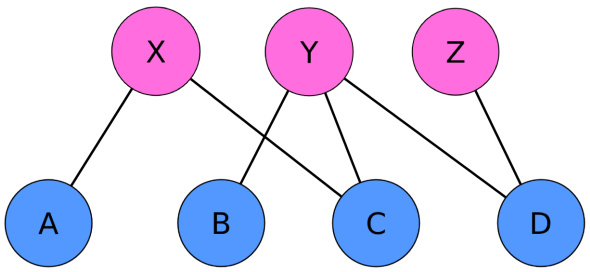


Figura 3.1: Exemplo de rede bipartida

Assim as entradas da matriz de incidência são definidas por:

$$B_{ij} = \begin{cases} 1, & \text{se o vértice j pertence ao grupo i.} \\ 0, & \text{caso contrário.} \end{cases}$$

E no caso da rede bipartida apresentada acima teremos então uma matriz de incidência de tamanho 3×4 é dada por

$$\mathrm{B} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

A redes bipartidas também são conhecidas como redes *two-mode* e a partir desse conceito apresentado anteriormente, pode ser realizado a projeção de *one-mode*, que consiste na transformação da rede bipartida em duas redes simples, a teoria desse método foi estudada no material Opsahl (2009). Para exemplificar utilizaremos a rede da Figura 3.1, e faremos as projeções *one-mode* a partir dela.

Começando pela projeção dos vértices azuis, criaremos uma rede *n-vertex*, apresentada na Figura 3.2, na qual esses vértices serão interligados quando estiverem ambos conectados a um mesmo vértice rosa. Essa projeção revela conexões diretas entre os vértices azuis com base em sua associação com os vértices rosas, voltando ao nosso exemplo, um vértice azul só estará conectado a outro vértice azul, se esses alunos, fizerem parte de um mesmo grupo de estudo.

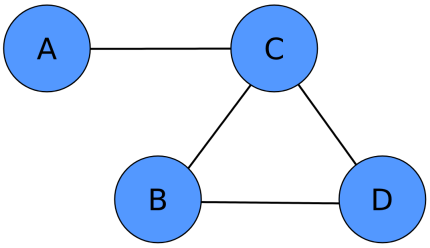


Figura 3.2: Projeção *one-mode* dos vértices azul da Figura 3.1

Em seguida, realizaremos a projeção *one-mode* correspondente aos vértices rosas que resultará em uma rede *g-vertex*, apresentada na Figura 3.3. Nessa rede, os nós rosas estarão interligados somente se compartilharem um nó azul comum. Isso nos permite explorar as conexões entre os vértices rosas com base na sua relação com os vértices azuis, voltando ao exemplo, um vértice rosa só estará conectado a outro vértice rosa se um mesmo aluno fizer parte de ambos grupos de estudo.

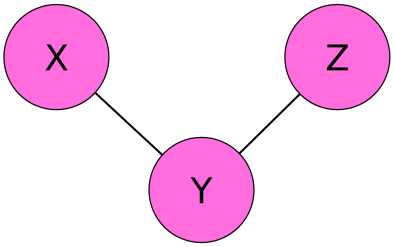


Figura 3.3: Projeção *one-mode* dos vértices rosa da Figura 3.1

Através dessa abordagem de projeção, podemos simplificar e analisar a estrutura da rede bipartida original, destacando as relações relevantes entre os vértices azuis e rosas de forma mais clara e interpretável.

As projeções *one-mode* apresentadas nas Figuras 3.2 e 3.3, são uteis e muito utilizadas, porém há uma grande perda de informações da rede bipartida original, com isso esse método se torna menos poderoso na representação dos dados. Para que não haja uma perda tão grande das informações, utiliza-se a projeção ponderada, no qual se adiciona peso em cada aresta, o valor desse peso é igual ao número de grupos que compartilham o mesmo vértice. A matriz de adjacência da projeção ponderada *one-mode* é calculada como $P = B^T B$, assim resultando em uma matriz $n \times n$. Assim podemos representar a rede bipartida da Figura 3.1 pela matriz de projeção ponderada a seguir

$$\mathrm{P} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix}$$

Como a rede bipartida possui dois conjuntos de vértices, haverá então mais uma representação *one-mode* por matriz de projeção ponderada, essa então será calculada como $P' = BB^T$ e de dimensão $g \times g$, sendo apresentada a seguir

$$\mathrm{P'} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

Mesmo utilizando a projeção ponderada existe a perda de informação da rede original, perda essa que se refere ao não registro do número de associações exatas de cada grupo. Tratando agora da parte matemática da projeção das redes bipartidas, temos que ela pode ser escrita em termos da matriz incidência B. A equação matemática que representa a projeção ponderada apresentada na matriz P será:

$$P_{ij} = \sum_{k=1}^{g} B_{ki} B_{kj} = \sum_{k=1}^{g} B_{ik}^T B_{jk}$$

e assim considerando que B assumirá os valores zero ou um, temos o número total P_{ij}, de grupos aos quais i e j pertencem.

Como exemplo, vamos utilizar a expressão acima para calcular 3 valores da matriz P, para exemplificar o uso da fórmula.

$$P_{12} = B_{11}B_{12} + B_{21}B_{22} + B_{31}B_{32} = 1 \cdot 0 + 0 \cdot 1 + 0 \cdot 0 = 0$$

$$P_{23} = B_{12}B_{13} + B_{22}B_{23} + B_{32}B_{33} = 0 \cdot 1 + 1 \cdot 1 + 0 \cdot 0 = 1$$

$$P_{44} = B_{14}B_{14} + B_{24}B_{24} + B_{34}B_{34} = 0 \cdot 0 + 1 \cdot 1 + 1 \cdot 1 = 2$$

Ao verificar na matriz P, temos exatamente esses valores nas posições P_{12}, P_{23} e P_{44}.

3.2 Detecção de comunidades

Nas seções anteriores foram apresentadas teorias sobre grafos direcionados, tanto com peso, quanto simples de maneira introdutória para o assunto que será apresentado nessa seção.

A detecção de comunidades é uma técnica usada para entender como um grupo de indivíduos se relaciona. Essa relação é geralmente representada por grafos simples, ou seja, grafos sem peso e direção nas conexões. No entanto, se for necessário representar essas relações de forma mais detalhada, é possível usar grafos com pesos e direções.

Uma comunidade é organizada de forma que os grupos de vértices estejam conectados dentro de uma mesma comunidade, de modo que os vértices tenham mais conexões entre si, enquanto entre comunidades diferentes, as conexões ocorrem com menos frequência. O principal objetivo de dividir uma rede em comunidades é identificar grupos que tenham mais conexões internas, ou seja, encontrar vértices que estejam mais relacionados entre si.

Os problemas em torno da detecção de comunidades são semelhantes aos encontrados no clustering, onde o objetivo é agrupar elementos com base em similaridades. No clustering, os elementos são agrupados com base em características como proximidade em um espaço de dados. Já na detecção de comunidades, o foco está nas conexões, buscando-se agrupar vértices que possuem mais ligações entre si do que com o restante da rede. Apesar das diferenças, ambos compartilham a ideia central de identificar padrões de agrupamento.

A detecção de comunidades é o problema que será solucionado por meio de diferentes abordagens, incluindo a modelagem via o Modelo Estocástico de Blocos (SBM) e o Modelo Estocástico de Blocos com Grau Corrigido (DCSBM). Esses métodos serão apresentados a seguir, destacando suas características e aplicações na análise de redes complexas.

3.2.1 Modelo Estocástico de Blocos (SBM)

Para criar uma rede com estrutura de comunidade, usamos o Modelo Estocástico de Blocos (SBM), proposto por Holland *et al.* (1983), assumimos que a rede tem K comunidades e definimos n variáveis aleatórias $Z_1, \ldots, Z_n$, que podem assumir valores de 1 a K. Cada variável Z_i indica a comunidade à qual o vértice i pertence, isto é,

$$Z_i = a$$

significa que o vértice i pertence à comunidade a. As comunidades $1, \ldots, K$ são atribuídas aos n vértices da rede de forma aleatória, de modo que:

$$\mathbb{P}(Z_i = a) = \pi_a.$$

O parâmetro π_a representa a proporção esperada de vértices que pertencem à comunidade a, e seus valores devem satisfazer $0 \leq \pi_a \leq 1$ e $\sum_{a=1}^{K} \pi_a = 1$.

A partir do vetor de comunidades $\mathbf{Z}_n = (Z_1, \ldots, Z_n)$, onde as entradas são independentes e identicamente distribuídas (i.i.d.), definimos como as conexões (arestas) do grafo são criadas. Para um par de vértices i, j, com $1 \leq i < j \leq n$, a variável aleatória $A_{ij} \in \{0, 1\}$ mostra se há uma conexão entre i e j, sendo gerada por:

$$A_{ij} \mid Z_i = a, Z_j = b \sim \text{Ber}(P_{ab}),$$

onde $\text{Ber}(\cdot)$ é a distribuição Bernoulli, P_{ab} é a probabilidade de existir uma aresta entre um vértice da comunidade a e um vértice da comunidade b. Esses valores formam uma matriz de probabilidade $P = [P_{ab}]$, onde $P_{ab} \in [0, 1]$. Essa matriz define as intensidades das conexões entre e dentro das comunidades, sendo que P_{aa} normalmente representa a densidade de conexões dentro de uma mesma comunidade a, enquanto P_{ab} (para $a \neq b$) determina a força das conexões entre comunidades diferentes.

Dado o vetor $\mathbf{Z}_n$, as variáveis $\{A_{ij}\}_{1 \leq i < j \leq n}$ são independentes. Sabendo o número de comunidades K, podemos estimar as comunidades dos vértices $\mathbf{Z}_n = (Z_1, \ldots, Z_n)$ usando métodos baseados no espectro da rede.

Como exemplo simples de aplicação desse método podemos supor que temos uma rede social com 6 pessoas divididas em dois grupos:

- Grupo 1 (A, B, C)
- Grupo 2 (D, E, F)

Definimos a matriz de probabilidade de conexão entre os grupos como:

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.6 \end{pmatrix}$$

sendo assim,

- A probabilidade de conexão dentro do Grupo 1 é 0.8.
- A probabilidade de conexão dentro do Grupo 2 é 0.6.
- A probabilidade de conexão entre um membro do Grupo 1 e um do Grupo 2 é 0.2.

Para gerar uma rede com base nesse modelo, para cada par de nós (i, j) sorteamos uma aresta com a probabilidade correspondente à matriz P de acordo com os grupos de (i, j).

3.2.2 Espectro de um Grafo

O espectro de um grafo é o conjunto de autovalores da matriz de adjacência A. Um autovalor λ e o respectivo autovetor v satisfazem a equação:

$$Av = \lambda v.$$

Nessa equação:

- A é a matriz de adjacência do grafo, que descreve as conexões entre os vértices.
- v é o autovetor associado ao autovalor λ, representando uma direção no espaço que permanece inalterada (exceto pelo fator de escala λ) após a multiplicação por A.
- λ é o autovalor, que indica a magnitude da transformação na direção do autovetor v.

A decomposição espectral da matriz A é dada por:

$$A = UDU^T,$$

no qual:

- D é uma matriz diagonal cujos elementos na diagonal são os autovalores $\text{diag}(D) = (\lambda_1, \ldots, \lambda_n)$.
- U é a matriz cujas colunas são os autovetores correspondentes.

Como exemplo, considere um grafo simples com 6 vértices e as conexões representadas pela matriz de adjacência A:

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

Os autovalores da matriz de adjacência são:

$$\lambda_1 = 2.66, \quad \lambda_2 = 1.00, \quad \lambda_3 = 1.00, \quad \lambda_4 = 0.00, \quad \lambda_5 = -2.14, \quad \lambda_6 = -2.14.$$

Esses autovalores refletem propriedades estruturais do grafo. Por exemplo, autovalores positivos maiores indicam maior conectividade entre os vértices, enquanto autovalores negativos sugerem padrões de desconexão. No caso de $\lambda_4 = 0.00$, isso pode estar relacionado à existência de componentes desconexas ou redundâncias na estrutura da rede.

3.2.3 Modelo Estocástico de Blocos com Grau Corrigido (DCSBM)

O Modelo Estocástico de Blocos com Grau Corrigido (DCSBM) é uma versão melhorada do SBM, que permite ajustar a variação dos graus dos vértices na rede. Isso é feito com um parâmetro extra, w_i, que ajusta o grau de cada vértice, esse ajuste foi proposto por Karrer e Newman (2011).

Definição do Modelo DCSBM

1. **Variáveis Latentes**:

   - O vetor de variáveis latentes é $Z_n = (Z_1, \ldots, Z_n)$, onde cada Z_i indica a comunidade do vértice i, com $\mathbb{P}(Z_i = a) = \pi_a$.

2. **Parâmetro de Grau**:

   - Cada vértice i tem um parâmetro w_i, que ajusta a probabilidade de se conectar a outros vértices.

3. **Variáveis Observadas**:

   - A rede é representada pela matriz de adjacência A de tamanho $n \times n$, onde a distribuição condicional de A_{ij}, dada as comunidades, é:

$$A_{ij} \mid Z_i = a, Z_j = b \sim \text{Ber}(w_i w_j P_{ab}),$$

     onde P_{ab} é a probabilidade de conexão entre as comunidades a e b.

O modelo de rede DCSBM é importante porque permite capturar a heterogeneidade dos graus dos vértices, algo frequentemente observado em redes reais, como redes sociais, biológicas ou de comunicação. Enquanto o SBM clássico assume graus aproximadamente homogêneos dentro das comunidades, o DCSBM incorpora essa variação ao ajustar a probabilidade de conexão de cada vértice, oferecendo uma descrição mais realista da estrutura da rede.

Este modelo é amplamente utilizado em aplicações práticas para detectar comunidades de forma mais precisa, especialmente em redes onde existem vértices altamente conectados (hubs). Ele também auxilia na análise de padrões complexos de conectividade, permitindo uma melhor compreensão das interações dentro e entre comunidades em redes reais.

Utilizando o mesmo exemplo apresentado no Modelo Estocástico de Blocos, temos agora que cada vértice i recebe um fator θ_i que representará sua atividade na rede, no nosso exemplo redes sociais. Suponha que

$$\theta = [0.5, 1.2, 0.8, 1.0, 0.7, 1.5]$$

A probabilidade de conexão entre dois nós i e j passa a ser:

$$P_{ij} = \theta_i \theta_j P_{g(i),g(j)}$$

onde $g(i)$ e $g(j)$ representam os grupos dos vértices i e j.

Isso significa que, dentro do mesmo grupo, um nó com $\theta = 1.5$ terá mais conexões do que um nó com $\theta = 0.5$ mesmo que ambos pertencendo ao mesmo bloco. Esse modelo melhora a flexibilidade, permitindo que alguns nós sejam hubs enquanto outros tenham menos conexões.

3.2.4 Grau Esperado no SBM e DCSBM

O grau esperado de conexão entre dois vértices i e j é calculado de maneiras diferentes nos modelos SBM e DCSBM.

No cálculo do grau esperado no modelo SBM, a fórmula que descreve o grau esperado de conexão entre dois vértices i e j no Modelo Estocástico de Blocos (SBM) é:

$$\mathbb{E}[A_{ij}] = P(A_{ij} = 1) = \sum_{a=1}^{K} \sum_{b=1}^{K} P(Z_i = a, Z_j = b) P_{ab}$$

- Na fórmula $P(Z_i = a, Z_j = b)$ representa a probabilidade de que o vértice i pertença à comunidade a e o vértice j pertença à comunidade b. Como as comunidades são definidas de forma aleatória, temos:

  $$P(Z_i = a, Z_j = b) = \pi_a \pi_b$$

  onde π_a é a probabilidade de o vértice i estar na comunidade a, e π_b é a probabilidade de o vértice j estar na comunidade b.

O grau esperado é dado pela soma das probabilidades de que os vértices i e j pertençam a todas as combinações possíveis de comunidades, multiplicada pela probabilidade de conexão entre essas comunidades.

No Modelo Estocástico de Blocos com Grau Corrigido (DCSBM), a fórmula para o grau esperado inclui um fator adicional para ajustar a variação dos graus dos vértices. A fórmula é:

$$\mathbb{E}[A_{ij}] = P(A_{ij} = 1) = \sum_{a=1}^{K} \sum_{b=1}^{K} P(Z_i = a, Z_j = b) w_i w_j P_{ab}$$

- Já neste caso $P(Z_i = a, Z_j = b)$ representa a probabilidade de que i pertença à comunidade a e j à comunidade b, e como no SBM, temos:

  $$P(Z_i = a, Z_j = b) = \pi_a \pi_b$$

- Além disso w_i e w_j representam os parâmetros adicionais do DCSBM, representando

o grau esperado de conexão do vértice i e do vértice j. Esses parâmetros ajustam a probabilidade de conexão levando em consideração o número de conexões que cada vértice tem na rede. O parâmetro w_i ajusta a probabilidade de i se conectar com qualquer outro vértice, e w_j faz o mesmo para j.

Além de considerar a probabilidade de pertencimento às comunidades e a probabilidade de conexão entre essas comunidades, o modelo DCSBM leva em conta o grau individual de cada vértice i e j, ajustado pelos parâmetros w_i e w_j. Esses parâmetros adicionais permitem que o modelo reflita a heterogeneidade dos graus dos vértices, o que é importante em redes reais onde nem todos os vértices têm o mesmo número de conexões.

3.2.5 Laplaciano Normalizado de um Grafo

Quando falamos sobre a variação dos graus dos vértices, estamos nos referindo ao fato de que, em muitas redes reais, nem todos os vértices têm o mesmo número de conexões (ou grau).

Para lidar com essa variação de graus e evitar que vértices com muitos graus dominem os cálculos, utilizamos o Laplaciano Normalizado, que ajuda a suavizar a influência de vértices com graus muito altos e a tornar as conexões mais equilibradas, as fórmulas e teorias expplicadas a frente foram estudadas a partir dos materiais Shigehalli e Shettar (2011) e Lei e Rinaldo (2015).

A fórmula do Laplaciano Normalizado é:

$$L = D^{-1/2} A D^{-1/2}$$

onde:

- A é a matriz de adjacência (que indica quais vértices estão conectados)
- D é a matriz diagonal de graus, ou seja, D_{ii} é o grau do vértice i.
- $D^{-1/2}$ é a inversa da raiz quadrada de D, o que ajuda a normalizar a influência de vértices de graus diferentes.

Os autovalores do Laplaciano variam no intervalo $[-1, 1]$, o que permite representar as relações estruturais da rede de maneira mais robusta e eficiente, considerando a variação nos graus.

O Agrupamento Espectral Esférico algoritmo para DCSBM é usado para identificar comunidades em redes, especialmente em redes modeladas pelo DCSBM. ele usa o conceito de autovalores e autovetores da matriz de adjacência (ou do Laplaciano Normalizado) para agrupar os vértices com base nas suas relações estruturais.

1. Calcule o Laplaciano Normalizado L: O primeiro passo é calcular o Laplaciano Normalizado, que ajuda a considerar as diferenças de grau entre os vértices e a normalizar as conexões. Como explicado anteriormente, ele é calculado com a fórmula $L = D^{-1/2}AD^{-1/2}$, usando a matriz de adjacência A e a matriz diagonal de graus D.

2. Ordene os autovalores de L e escolha os K maiores: Após calcular o Laplaciano, ordenamos seus autovalores do maior para o menor. A ideia é que os K maiores autovalores podem nos dar as informações necessárias para dividir a rede em K comunidades.

3. Crie a matriz U com os autovetores correspondentes: Cada autovalor tem um autovetor associado. Esses autovetores representam as relações dos vértices com as comunidades. Então, criamos uma matriz U cujas colunas são os autovetores correspondentes aos K maiores autovalores.

4. Normalize as linhas de U para obter $\tilde{U}$: Para que os autovetores sejam comparáveis e não sejam dominados por valores grandes, normalizamos as linhas da matriz U. A normalização é feita dividindo cada linha pelo seu comprimento (ou norma), fazendo com que as linhas da matriz resultante $\tilde{U}$ tenham norma igual a 1.

5. Aplique o algoritmo de K-médias em $\tilde{U}$ para identificar as comunidades: Depois de normalizar as linhas de U, usamos o algoritmo de K-médias em $\tilde{U}$ para agrupar os vértices em K comunidades. O algoritmo de K-médias tenta agrupar os dados (neste caso, os vértices) de modo que os vértices dentro de uma comunidade sejam mais semelhantes entre si do que com vértices de outras comunidades.

O algoritmo de agrupamento espectral retorna as comunidades que foram identificadas com base na estrutura da rede, dividindo a rede em K grupos de vértices, onde os vértices dentro de uma comunidade estão mais fortemente conectados entre si do que com vértices de outras comunidades.

Esse algoritmo é útil principalmente para redes em que as conexões entre os vértices podem ser difíceis de entender à primeira vista, especialmente em redes grandes e complexas. Ele pode ser utilizado em redes com e sem pesos.

Em resumo, o algoritmo de agrupamento espectral é uma técnica importantissima para a detecção de comunidades em redes, sendo especialmente útil quando há variação no grau de conectividade entre os vértices e quando os pesos das conexões influenciam a estrutura da rede.

Como exemplo de aplicação do método, vamos considerar um grafo com três nós e as seguintes arestas:

$$V = \{1, 2, 3\}, \quad E = \{(1, 2), (2, 3)\}$$

A representação gráfica deste grafo é:

$$1 \longleftrightarrow 2 \longleftrightarrow 3$$

A matriz de adjacência A é construída com base nas arestas do grafo:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

A matriz de graus D é uma matriz diagonal onde cada entrada D_{ii} representa o grau do vértice i:

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

A matriz laplaciana L é dada por:

$$L = D - A$$

Portanto, temos:

$$L = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}$$

O Laplaciano Normalizado é calculado como:

$$L_{\text{norm}} = D^{-1/2} L D^{-1/2}$$

Onde:

$$D^{-1/2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Agora, calculamos o produto:

$$L_{\text{norm}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Após as multiplicações, obtemos:

$$L_{\text{norm}} = \begin{pmatrix} 1 & -\frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 1 & -\frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & 1 \end{pmatrix}$$

Assim temos que a matriz Laplaciana Normalizada do grafo dado é:

$$L_{\text{norm}} = \begin{pmatrix} 1 & -\frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 1 & -\frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & 1 \end{pmatrix}$$

Esta matriz pode ser utilizada para estudar as propriedades espectrais e de conectividade do grafo.

Capítulo 4

Análise de dados

Neste capítulo, serão apresentadas as informações sobre o banco de dados utilizado no estudo e o pré-processamento realizado a ele antes da aplicação das técnicas apresentadas anteriormente. Iniciaremos analisando as redes bipartidas, que nos permite visualizar as relações entre os nós. Em seguida realizaremos a análise descritiva dos dados com a apresentação dos histogramas do grau dos vértices, para entendermos a distribuição das conexões entre os vértices de uma rede. A técnica de *projeção one-mode* será apresentada também como forma de facilitar a análise. Por fim, estudaremos a detecção de comunidades, que torna possível a identificação de grupos de vértices com forte conexão.

4.1 Banco de dados

O conjunto de dados analisado é disponibilizado pelo Ministério do Turismo[1] e consiste em registros administrativos de migração tratados estatisticamente conforme as Recomendações Internacionais de Estatística de Turismo, estabelecidas em 2008 pela Organização Mundial de Turismo (OMT). Esse tratamento inclui um filtro para capturar exclusivamente as entradas de indivíduos no país que se alinham com a definição de turismo. Ou seja, são consideradas apenas as pessoas que chegam ao Brasil com o objetivo de conhecer e explorar a cultura local e que possuem pernoite inclusa na viagem.

A Tabela 4.1 representa a maneira a qual os dados são fornecidos pelo Ministério do Turismo, sendo eles organizados em planilhas de *Excel*, divididas em doze colunas: Continente, cod continente, País, cod país, UF, cod UF, Via, cod via, ano, mês, cod mes e Chegadas.

[1] http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html

Continente	cod continente	País	cod país	UF	cod UF	Via	cod via	Ano	Mês	cod mês	Chegadas
África	1	África do Sul	2	Acre	1	Aérea	1	2019	janeiro	1	0
...	...	...	...	...	...	...	...	...	...	...	...
América do Sul	4	Argentina	11	Paraná	16	Terrestre	2	2019	abril	4	20417
Oceania	7	Austrália	14	Rio de Janeiro	19	Aérea	1	2019	dezembro	12	1255
...	...	...	...	...	...	...	...	...	...	...	...

Tabela 4.1: Tabela de demostração de como o banco de dados é apresentado pelo Ministério do Turismo, essa em especifico para o ano de 2019.

4.2 Pré-processamento dos dados

Como mencionado na seção anterior, o banco de dados contém uma grande quantidade de informações, mas para este estudo não foi necessário a utilização de todas. Portanto, foi realizado um pré-processamento dos dados para incluir apenas as informações relevantes para o estudo. O primeiro filtro foi aplicado à variável Via, selecionando apenas a modalidade Aérea. Em seguida, foi filtrada a variável cod continente, escolhendo-se o valor 6, que é referente ao continente Europeu. Por fim, foram excluídos os países que não tiveram nenhum residente vindo ao Brasil com intenção de turismo, assim como foram retirados os estados que não receberam nenhum turista vindo dos países da Europa. Além disso os grafos são semestrais, sendo possível observar com mais detalhes a interação entre os países da Europa e os estados do Brasil.

Após realizar o pré-processamento, os dados se encontram organizados como descritos na 4.2.

Continente	cod continente	País	cod país	UF	cod UF	Via	cod via	Ano	Mês	cod mês	Chegadas
Europa	6	Alemanha	4	Acre	1	Aérea	1	2019	janeiro	1	0
...	...	...	...	...	...	...	...	...	...	...	...
Europa	6	Dinamarca	57	Santa Catarina	24	Aérea	1	2019	abril	4	1
Europa	6	Suíça	216	São Paulo	25	Aérea	1	2019	junho	6	1134
...	...	...	...	...	...	...	...	...	...	...	...

Tabela 4.2: Tabela que apresenta o banco de dados do primeiro semestre de 2019 após o pré-processamento de dados.

Depois de realizar o pré-processamento dos dados, os mesmo estão prontos para serem representados por grafos, nos quais os vértices serão representados pelos países e os estados em estudo e as arestas a relação entre eles. No caso das redes bipartidas as arestas irão representar a quantidade de turistas que se deslocam de um país europeu em direção a um estado brasileiro, cada aresta representa a quantidade de turistas. Já nas projeções e detecção de comunidade as arestas que conectam os estados entre si representa que os estados que compartilham essa ligação, receberam turistas vindos do mesmo País, já no caso das arestas que conectam os países, essas representam que o turista vindo de um país foi para os estados conectados entre si, sendo possível assim observamos e analisarmos se

há relação entre os turistas de vindos de um mesmo país com os estados visitados pelos mesmos.

4.3 Rede Bipartida

A partir dessa seção serão apresentadas interpretações e análise do tema central deste trabalho, as redes bipartidas e suas projeções.

A Tabela 4.3 apresenta em ordem alfabética os estados brasileiros e os países europeus que estão presentes no banco de dados e compõem as redes bipartidas.

Tabela 4.3: Unidades Federativas do Brasil e Países da Europa presentes nas Redes Bipartidas

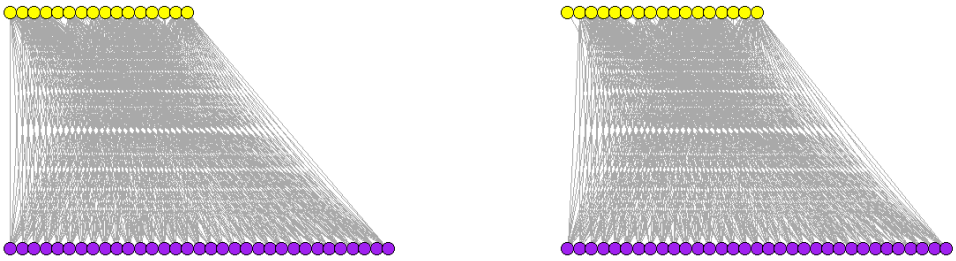
Acre	Alemanha	Amapá	Amazonas	Áustria
Bahia	Bélgica	Bulgária	Ceará	Croácia
Dinamarca	Distrito Federal	Eslováquia	Eslovênia	Espanha
Estônia	Finlândia	França	Grécia	Holanda
Hungria	Irlanda	Itália	Letônia	Lituânia
Luxemburgo	Mato Grosso do Sul	Minas Gerais	Noruega	Outras Unidades da Federação
Outros países	Pará	Paraná	Pernambuco	Polônia
Portugal	Reino Unido	República Tcheca	Rio de Janeiro	Rio Grande do Norte
Rio Grande do Sul	Romênia	Roraima	Rússia	Santa Catarina
São Paulo	Sérvia	Suécia	Suíça	Turquia
Ucrânia				

As redes bipartidas mostradas nas Figuras 4.1,4.2,4.3 e 4.4 dão uma visão geral do fluxo turístico entre a Europa e o Brasil. Ao ligar os estados brasileiros (vértices amarelos) aos países de origem dos turistas (vértices roxos), essas redes ajudam a entender como se dão as conexões e as hierarquias no turismo.

O padrão visual visto nos períodos analisados indica que o turismo no Brasil se mantém mais ou menos estável. Porém, algumas pequenas mudanças nas conexões sugerem que pode haver alterações nos destinos mais procurados ou nos países que mais enviam turistas.

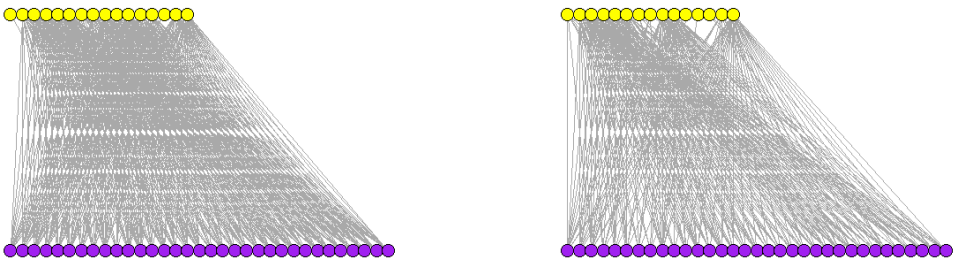
Ao analisar as linhas e colunas da matriz de adjacência, observamos que em todos os anos o vértice referente ao estado do Acre não possuía nenhuma aresta conectada. Além disso, o estado do Mato Grosso do Sul não recebeu nenhum turista no ano de 2020 e no primeiro semestre de 2021. O estado de Roraima também apresentou ausência de turistas no primeiro semestre de 2021 e no segundo semestre de 2022. Já o vértice que representa o Amapá ficou desconectado apenas no segundo semestre de 2020. Devido à ausência de conexões ou registros turísticos nos períodos mencionados, os estados citados foram

desconsiderados na representação da rede bipartida e nos estudos realizados sobre ela. Em contra partida, temos que em todos os anos os estados que mais receberem turistas foram São Paulo, Rio de Janeiro e Ceará, respectivamente, já os países europeus com mais vinda de turista para o Brasil, foram Alemanha, Portugal e França, esses alternaram suas posições ao longo dos anos.



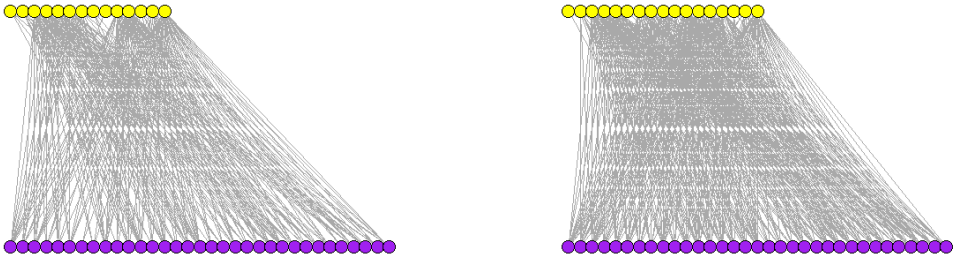
(a) 1º semestre de 2019 (b) 2º semestre de 2019

Figura 4.1: Rede Bipartida que representa a relação entre as UFs brasileiras e os estados europeus no ano de 2019



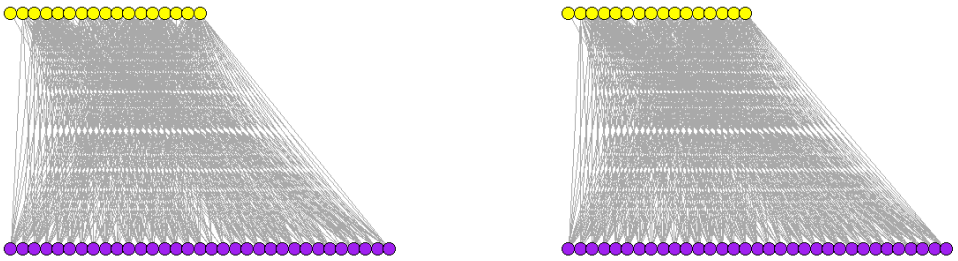
(a) 1º semestre de 2020 (b) 2º semestre de 2020

Figura 4.2: Rede Bipartida que representa a relação entre as UFs brasileiras e os estados europeus no ano de 2020



(a) 1º semestre de 2021 (b) 2º semestre de 2021

Figura 4.3: Rede Bipartida que representa a relação entre as UFs brasileiras e os estados europeus no ano de 2021



(a) 1º semestre de 2022 (b) 2º semestre de 2022

Figura 4.4: Rede Bipartida que representa a relação entre as UFs brasileiras e os estados europeus no ano de 2022

4.3.1 Análise descritiva dos dados

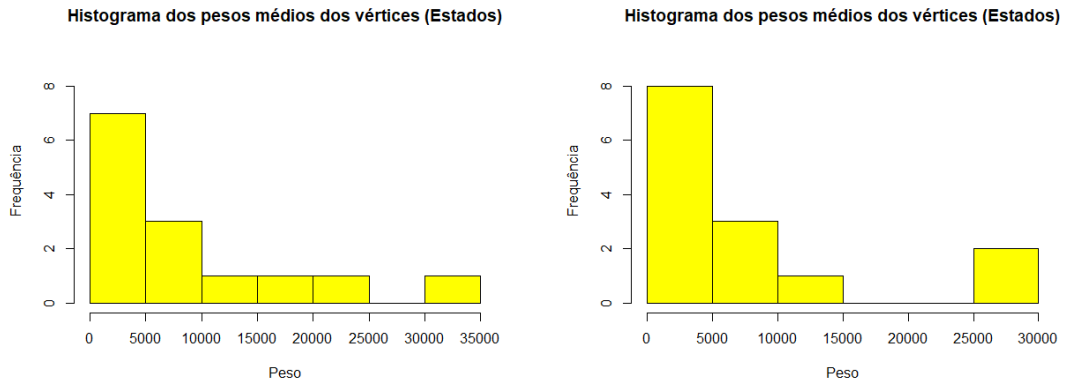
Nesta subseção será realizada a análise descritiva dos dados, por meio de histogramas e tabelas com as medidas resumo, separados por semestre em cada ano. Vale ressaltar que os histogramas a seguir apresentam apenas valores de grau menores que 50.000, pois há muitos valores entre 0 e 50.000 e poucos valores maiores que 50.000 mas que influenciam na apresentação gráfica dessas informações, porém os gráficos com todos os dados podem ser encontrados no apêndice desse relatório. Para complementar a analise dos histogramas utilizaremos as medidas resumos dos pesos dos vértices tanto para os países quanto para

os estados.

A seguir, será apresentada a análise dos dados referentes aos pesos dos vértices, que representam as Unidades Federativas (UFs) brasileiras, extraídos de uma rede bipartida entre as UFs do Brasil e os países europeus. Os dados são organizados em tabelas de medidas resumo e histogramas, separados por semestres dos anos de 2019, 2020, 2021 e 2022. Essas tabelas e histogramas visam ilustrar a distribuição e as variações nos pesos dos vértices ao longo do tempo, proporcionando uma visão detalhada do comportamento das interações entre as UFs e os países europeus, com foco nas mudanças ocorridas durante este período.

Tabela 4.4: Medidas resumo do peso médio dos vértices das Unidades Federativas do Brasil do ano de 2019

Semestre de 2019	**Min.**	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**
1º	1	1235	4634	7929	10616	30535
2º	2	156.5	2913	7164.6	9169.5	28898



(a) 1º semestre de 2019

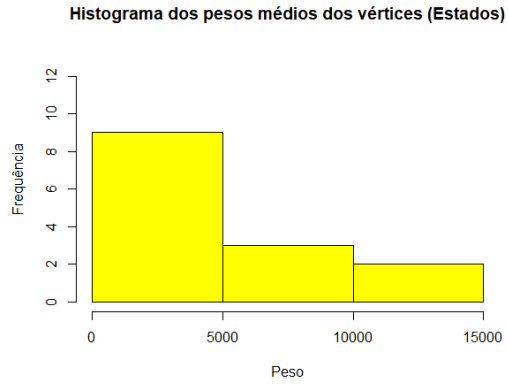
(b) 2º semestre de 2019

Figura 4.5: Histogramas que apresentam o peso médio dos vértices das Unidades Federativas do Brasil no ano de 2019

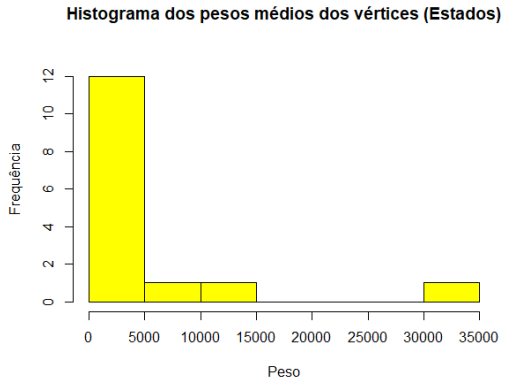
Fazendo a interpretação dos dois semestres do ano de 2019 observamos uma grande amplitude entre o valor máximo e o valor mínimo, é observado também que a média é maior que a mediana, de uma maneira significativa, o que indica uma distribuição assimétrica positiva em ambos os casos, refletindo que a maioria dos valores se concentram na faixa mais baixa, o que podemos confirmar ao observar os histogramas referentes a esses dados.

Tabela 4.5: Medidas resumo do peso médio dos vértices das Unidades Federativas do Brasil do ano de 2020

Semestre de 2020	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1º	1	367.2	2420.5	4040.7	5897.2	14239
2º	1	8.5	433	3653.4	2071.5	32221



(a) 1º semestre de 2020



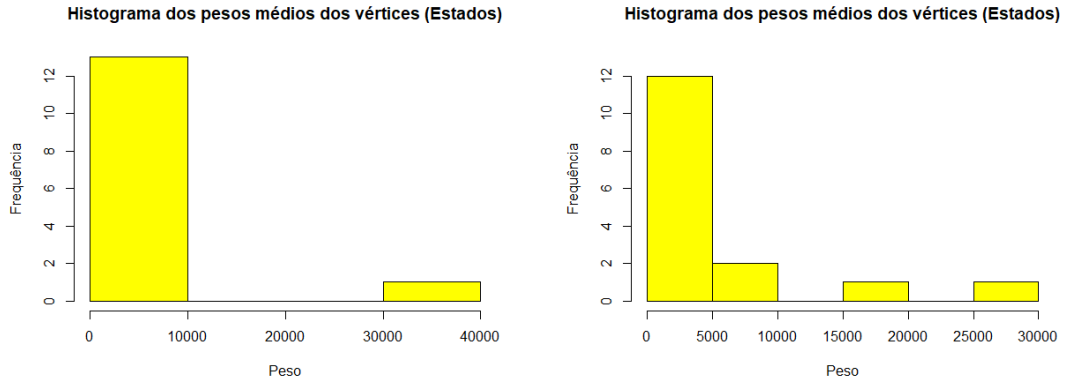
(b) 2º semestre de 2020

Figura 4.6: Histogramas que apresentam o peso médio dos vértices das Unidades Federativas do Brasil no ano de 2020

Já no ano de 2020 o primeiro semestre apresenta a assimetria positiva, porém no segundo semestre desse ano é observado um valor máximo muito alto influenciando assim a média, sugerindo a presença de *outliers*, já que a mediana é baixa. Os histogramas deste ano em estudo também são assimétricos positivos, porém com escalas diferentes em relação ao ano anterior, mostrando a redução geral dos valores. O *outlier* é representado no segundo semestre como uma barra isolada à direita.

Tabela 4.6: Medidas resumo do peso médio dos vértices das Unidades Federativas do Brasil do ano de 2021

Semestre de 2021	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1º	1	12	383	3476	1215	36078
2º	1	32.25	532	3973.81	3371.25	28143



(a) 1º semestre de 2021

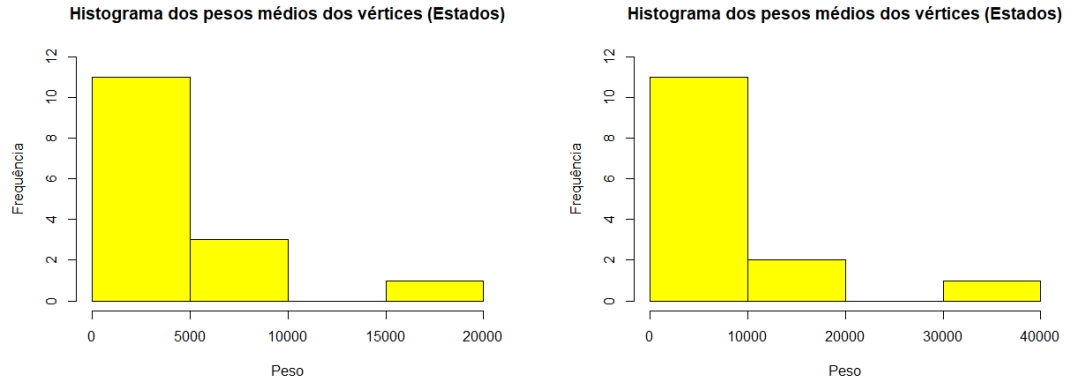
(b) 2º semestre de 2021

Figura 4.7: Histogramas que apresentam o peso médio dos vértices das Unidades Federativas do Brasil no ano de 2021

Ao observar as medidas resumo do ano de 2021 é possível perceber nos dois semestres uma grande diferença entre o valor máximo e o valor mínimo, com as médias muito discrepantes das medianas, indicando assim a uma assimetria positiva e alguns valores de *outliers*, essa hipótese é confirmada ao se observar os histogramas, nos quais são apresentados caudas longas a direita.

Tabela 4.7: Medidas resumo do peso médio dos vértices das Unidades Federativas do Brasil do ano de 2022

Semestre de 2022	**Min.**	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**
1º	3	142.5	1204	3638.5	5553	15946
2º	1	404	3294	7089	7404	35909



(a) 1º semestre de 2022

(b) 2º semestre de 2022

Figura 4.8: Histogramas que apresentam o peso médio dos vértices das Unidades Federativas do Brasil no ano de 2022

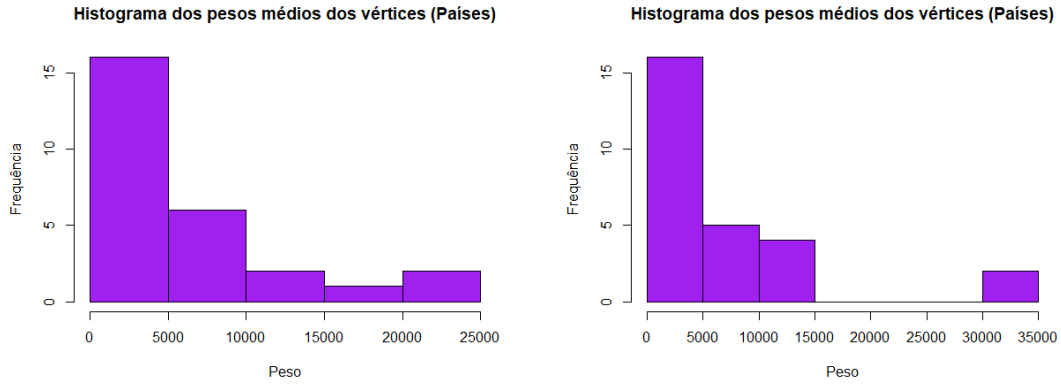
No ano de 2022 a assimetria positiva ainda é refletida nas medidas resumo e confirmada ao se analisar os histogramas. A assimetria positiva é apresentada em todos os anos em estudo, significando uma concentração de valores menor e alguns valores muito maiores que influenciam e 'puxam' a média, foi observado também que os valores máximos variam bastante entre os semestres e anos, o que impacta na amplitude e na média.

Ao comparar os anos de 2019, 2020, 2021 e 2022, é observado que, em todos os anos, a distribuição dos pesos é assimétrica positiva, ou seja, a maioria dos valores está concentrada em uma faixa mais baixa, mas com alguns valores muito altos que influenciam a média, porém em 2019, houve uma amplitude muito grande, especialmente no primeiro semestre, com valores extremos que puxaram a média para cima, enquanto em 2020, apesar de uma redução geral nos valores, ainda houve uma presença de outliers, especialmente no segundo semestre. Em 2021, a amplitude continuou alta, mas a média foi um pouco menor que em 2019, refletindo uma distribuição mais equilibrada, embora ainda assimétrica. Já em 2022, a amplitude foi mais moderada, com uma concentração maior de valores na faixa mais baixa em comparação com 2021, embora ainda houvesse valores atípicos que afetaram a média.

Agora, será apresentada a análise dos dados referentes aos pesos dos vértices, que representam os países europeus.

Tabela 4.8: Medidas resumo do peso médio dos vértices dos Países Europeus do ano de 2019

Semestre de 2019	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1º	550	1120.2	3096	7929	10616	30535
2º	503	1101	3037	6540	8688	32006



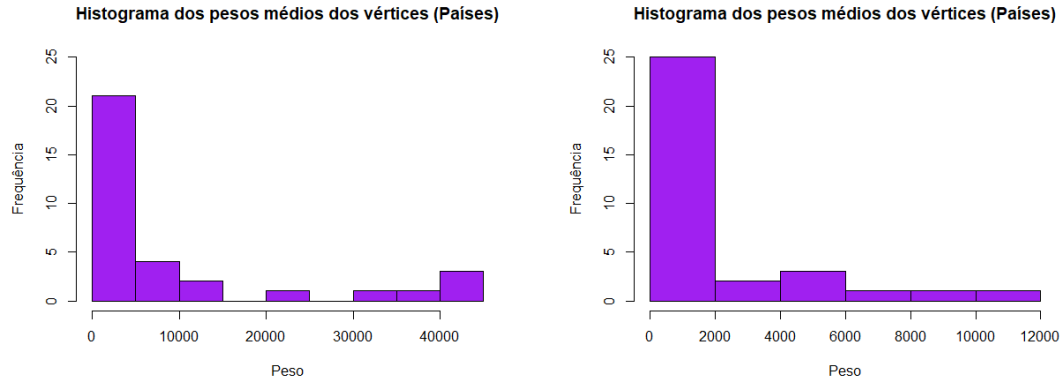
(a) 1° semestre do ano de 2019 (b) 2° semestre do ano de 2019

Figura 4.9: Histogramas que apresentam o peso médio dos vértices dos Países Europeus no ano de 2019

No primeiro semestre de 2019, a amplitude dos dados foi grande, com a média de 7.929 superando a mediana de 3.096, isso indicou uma distribuição assimétrica positiva, com a maioria dos valores concentrada em números menores, mas com alguns valores muito altos. No segundo semestre, o mesmo se repete e a média 6.540 continuou sendo maior que a mediana 3.037 confirmando a assimetria positiva, de forma semelhante ao primeiro semestre, o que também é possível ser observado e confirmado nos histogramas.

Tabela 4.9: Medidas resumo do peso médio dos vértices dos Países Europeus do ano de 2020

Semestre de 2020	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1º	296	822	2823	9238	8881	44934
2º	82	158	444	1661	1329	10353



(a) 1° semestre do ano de 2020 (b) 2° semestre do ano de 2020

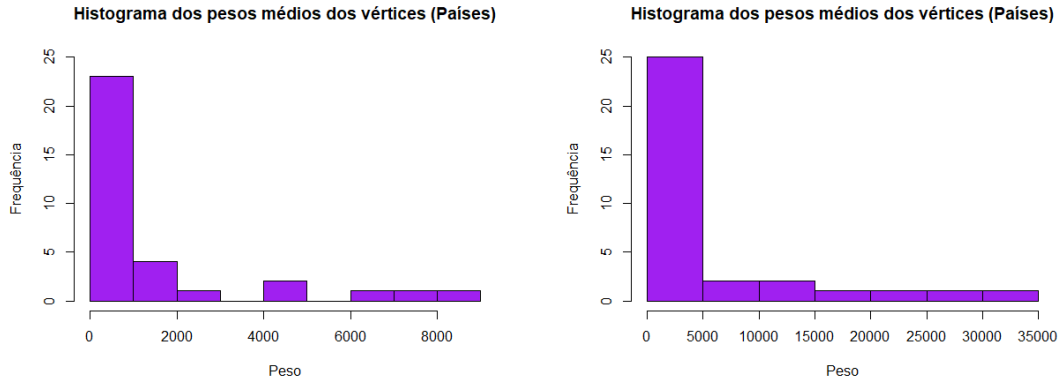
Figura 4.10: Histogramas que apresentam o peso médio dos vértices dos Países Europeus no ano de 2020

Em 2020, o primeiro semestre teve uma amplitude ainda maior que 2019 porém ainda sim é indicado uma forte assimetria positiva. No segundo semestre, houve uma mudança significativa, com a amplitude reduzida, embora a média ainda fosse maior que a mediana, a diferença foi menor, refletindo uma distribuição mais equilibrada em comparação com o primeiro semestre. A assimetria positiva ainda estava presente, mas de forma mais suave.

Tabela 4.10: Medidas resumo do peso médio dos vértices dos Países Europeus do ano de 2021

Semestre de 2021	**Min.**	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**
1º	52	169	430	1475	1507	8317
2º	187	417	1299	5097	4618	30269

Tabela 4.11: Histograma que apresenta o peso dos vértices dos Países Europeus no 1° semestre do ano de 2021



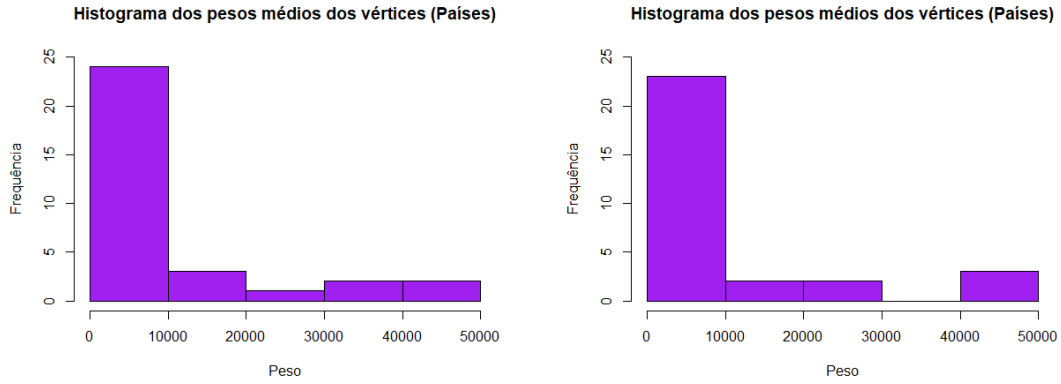
(a) 1° semestre do ano de 2021 (b) 2° semestre do ano de 2021

Figura 4.11: Histogramas que apresentam o peso médio dos vértices dos Países Europeus no ano de 2021

No ano de 2021 nos dois semestres é possível observar que ambas as médias foram inferiores aos terceiros quartis, mas ainda superior à mediana, indicando uma assimetria positiva menos acentuada que nos anos anteriores, e assim consigo observar nos histogramas caudas longas devido ao valor máximo.

Tabela 4.12: Medidas resumo do peso médio dos vértices dos Países Europeus do ano de 2022

Semestre de 2022	**Min.**	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**
1º	305	989	2522	8602	7693	46671
2º	358	792.5	2545	8927.3	8141.8	49494



(a) 1° semestre do ano de 2022 (b) 2° semestre do ano de 2022

Figura 4.12: Histogramas que apresentam o peso médio dos vértices dos Países Europeus no ano de 2022

Em 2022, é observado uma amplitude grande entre os mínimos e os máximos, em ambos os semestres, as médias continuaram menores que os terceiros quartis porém maior

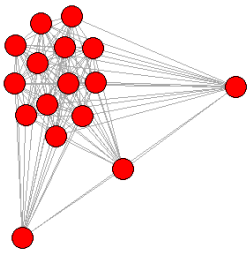
que as medianas mantendo a tendência de assimetria positiva e uma grande amplitude, representada nos histogramas.

Interpretando com uma visão geral os valores apresentados acima, é observado que em 2019 e 2022, a distribuição dos dados foi predominantemente assimétrica positiva, com a presença de valores máximos muito altos em todos os anos, o que puxou as médias para cima. Em 2020, a assimetria foi mais pronunciada no primeiro semestre, mas houve uma mudança no segundo semestre com uma distribuição mais equilibrada. Já em 2021 e 2022, embora a assimetria positiva tenha persistido, os valores máximos aumentaram consideravelmente, refletindo-se nas médias mais altas e em maior dispersão dos dados. O comportamento dos dados ao longo dos anos mostra uma tendência de aumento das amplitudes e dos valores máximos, embora a assimetria positiva tenha sido uma constante.

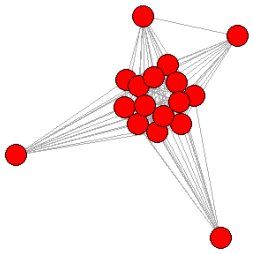
4.3.2 Projeção das redes bipartida de cada ano em estudo

Nesta subseção são apresentadas as projeções das redes bipartidas de cada ano em estudo, são apresentadas as projeções dos estados e dos países, respectivamente.

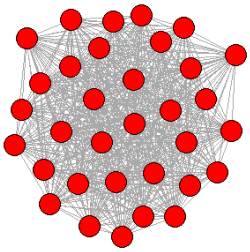
No caso das projeções dos estados, é observado uma rede menos densa, com alguns clusters visíveis, sugerindo grupos de estados com interações mais próximas. Já as redes de países foram mais densas, refletindo maior interconexão entre os países. Quando é feita a comparação entre os dois semestres de 2019, vemos que a estrutura das redes se modificou, com algumas conexões aparecendo ou desaparecendo, indicando que as relações entre as entidades não são estáticas.



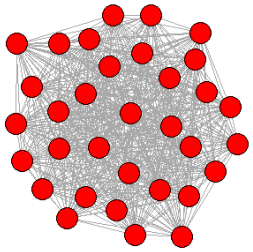
(a) Projeção *one-mode* das Unidades Federativas do Brasil em 2019.1



(b) Projeção *one-mode* das Unidades Federativas do Brasil em 2019.2



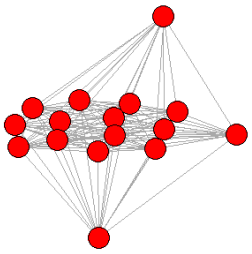
(c) Projeção *one-mode* dos Países da Europa em 2019.1



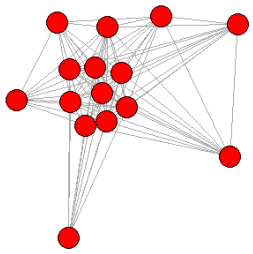
(d) Projeção *one-mode* dos Países da Europa em 2019.2

Figura 4.13: Projeção *one-mode* da rede bipartida do ano de 2019

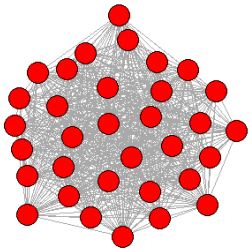
Em 2020, a estrutura das redes de estados e países apresentou mudanças, mas de maneira similar ao que vimos em 2019. As redes de países continuaram sendo mais densas, com uma interação mais forte entre os países. No entanto, as redes de estados continuaram a mostrar uma densidade menor e alguns clusters distintos. Ao comparar os semestres, também foi possível notar alterações nas conexões e nos clusters, o que indicou uma dinâmica nas relações ao longo do ano.



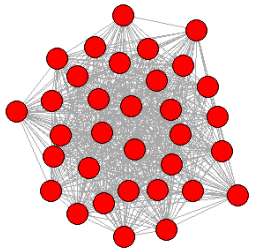
(a) Projeção *one-mode* das Unidades Federativas do Brasil em 2020.1



(b) Projeção *one-mode* das Unidades Federativas do Brasil em 2020.2



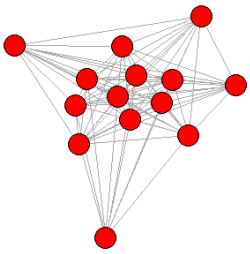
(c) Projeção *one-mode* dos Países da Europa em 2020.1



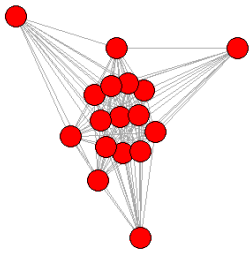
(d) Projeção *one-mode* dos Países da Europa em 2020.2

Figura 4.14: Projeção *one-mode* da rede bipartida do ano de 2020

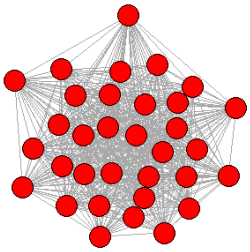
As projeções de 2021 revelaram redes de estados e países com características próprias. As redes de estados continuaram com menor densidade e alguns clusters mais definidos, sugerindo interações mais restritas a grupos específicos. As redes de países, por outro lado, mantiveram uma estrutura mais densa e uniforme, com interações mais generalizadas entre os países. Comparando os semestres de 2021, notamos algumas mudanças nas conexões, com alguns nós mais isolados, especialmente no primeiro semestre, o que indica uma certa fragmentação nas relações entre os estados.



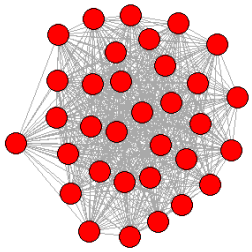
(a) Projeção *one-mode* das Unidades Federativas do Brasil em 2021.1



(b) Projeção *one-mode* das Unidades Federativas do Brasil em 2021.2



(c) Projeção *one-mode* dos Países da Europa em 2021.1

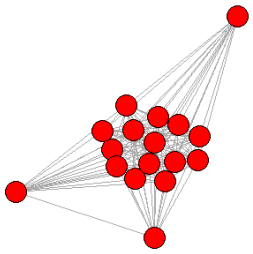


(d) Projeção *one-mode* dos Países da Europa em 2021.2

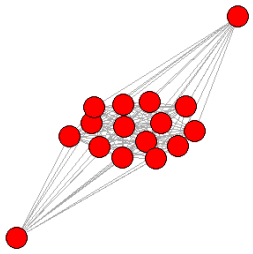
Figura 4.15: Projeção *one-mode* da rede bipartida do ano de 2021

Em 2022, as projeções mostraram um aumento na densidade das redes de países, indicando um nível maior de interação entre eles. As redes de estados, no entanto, se tornaram mais dispersas, especialmente no segundo semestre, sugerindo mudanças nas relações entre os estados. A comparação entre os semestres deste ano mostrou um aumento na densidade das redes de países, o que pode indicar uma intensificação das relações globais. Já as redes de estados continuaram a apresentar clusters, mas com uma distribuição mais ampla, refletindo um cenário de maior dispersão nas interações.

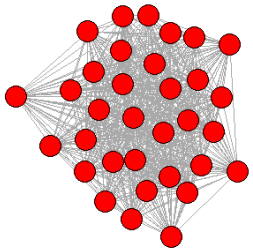
É possível observar que ao longo de 2019 a 2022, as redes de países mostraram uma tendência crescente de densidade, sugerindo um aumento nas interações globais. Por outro lado, as redes de estados mantiveram uma estrutura mais fragmentada, com a formação de clusters mais visíveis, especialmente em 2019 e 2021. As mudanças observadas entre os semestres e os anos indicam que as relações entre as entidades, tanto entre estados quanto entre países, não são estáticas e variam ao longo do tempo. Para uma análise mais precisa,



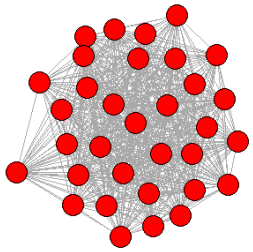
(a) Projeção *one-mode* das Unidades Federativas do Brasil em 2022.1



(b) Projeção *one-mode* das Unidades Federativas do Brasil em 2022.2



(c) Projeção *one-mode* dos Países da Europa em 2022.1



(d) Projeção *one-mode* dos Países da Europa em 2022.2

Figura 4.16: Projeção *one-mode* da rede bipartida do ano de 2022

seria necessário entender o contexto da rede bipartite original e o que realmente representa uma conexão entre as entidades.

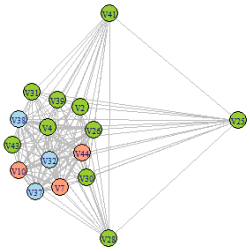
4.4 Detecção de comunidades

Por fim, nesta seção teremos a apresentação dos resultados obtidos após a aplicação dos métodos de detecção de comunidades, que apresentou o fenômeno do pequeno mundo, pois foram observadas distribuições de grau heterogêneas, alguns nós com grau muito alto conectando partes da rede.

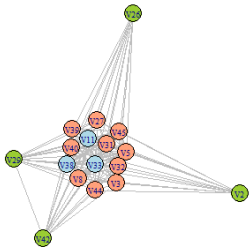
Esse fenômeno caracteriza redes em que os nós estão organizados de forma a combinar dois aspectos importantes: uma alta densidade de conexões locais, ou seja, os nós tendem a formar grupos bem conectados, e caminhos curtos que ligam qualquer par de nós, graças à presença de hubs. Esses hubs são nós altamente conectados que atuam como pontos

centrais e ajudam a conectar diferentes partes da rede.

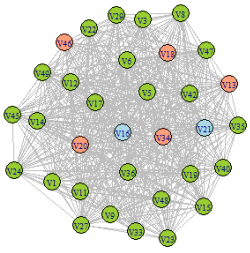
Apesar de ser eficiente para espalhar informações, essa estrutura cria desafios na identificação de comunidades em redes ponderadas. Isso ocorre porque os hubs, ao conectarem diferentes grupos, criam "atalhos"que dificultam separar comunidades claramente. Além disso, os pesos das conexões podem tornar mais confuso distinguir quais ligações são realmente relevantes para cada grupo. Nesse contexto, após testar diferentes opções, optamos por utilizar $k = 3$ comunidades, essa foi a solução mais adequada, equilibrando a complexidade da rede e os padrões de conexão observados, o $k = 3$ foi escolhido pelo método de tentativa e erro, no qual foi testado k de 1 a 5, e o $k = 3$ foi o que melhor representou o problema.



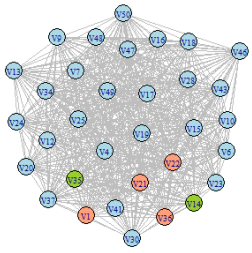
(a) Detecção de comunidades das UF´s do Brasil no 1º semestre de 2019 2019.1



(b) Detecção de comunidades das UF´s do Brasil no 2º semestre de 2019



(c) Detecção de comunidades dos Países Europeus no 1º semestre de 2019



(d) Detecção de comunidades dos Países Europeus no 2º semestre de 2019

Figura 4.17: Detecção de Comunidades da Rede Bipartida do ano de 2019

A análise de detecção de comunidades revelou a formação de grupos distintos dentro da rede bipartida composta por estados brasileiros e países da Europa, considerando $k = 3$ comunidades, representadas pelas cores azul, rosa e verde. Foi detalhado a composição

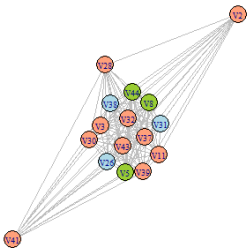
de cada grupo por semestres.

- Primeiro semestre de 2019 (Estados): O grupo azul foi formado pelos estados de Pernambuco, Rio de Janeiro e Rio Grande do Norte. No grupo rosa, tivemos o Ceará, o Distrito Federal e São Paulo. Os demais estados foram agrupados no grupo verde.

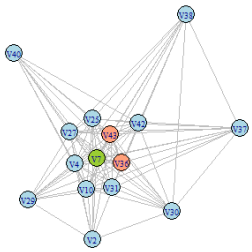
- Segundo semestre de 2019 (Estados): O grupo azul incluiu o Distrito Federal, Rio de Janeiro e Pernambuco. Já o grupo verde foi composto por Amapá, Mato Grosso do Sul, Outras Unidades da Federação e Roraima. Os demais estados permaneceram no grupo rosa.

- Primeiro semestre de 2019 (Países) : O grupo azul incluiu França e Itália, enquanto o grupo rosa foi composto por Espanha, Holanda, Irlanda, Portugal e Suécia. Os demais países ficaram no grupo verde.

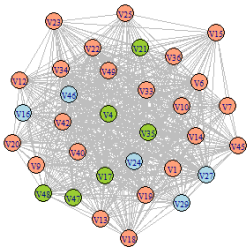
- Segundo semestre de 2019 (Países):O grupo rosa reuniu Alemanha, Irlanda, Itália e Reino Unido, enquanto o grupo verde incluiu Espanha e Portugal. Os demais países permaneceram no grupo azul.



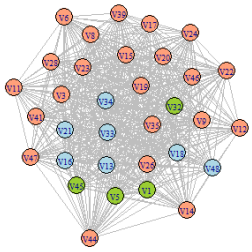
(a) Detecção de comunidades das UF´s do Brasil no 1º semestre de 2020



(b) Detecção de comunidades das UF´s do Brasil no 2º semestre de 2020



(c) Detecção de comunidades dos Países Europeus no 1º semestre de 2020



(d) Detecção de comunidades dos Países Europeus no 2º semestre de 2020

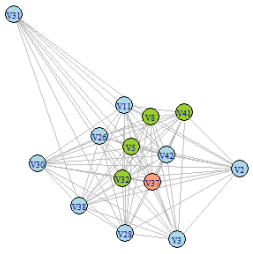
Figura 4.18: Detecção de Comunidades da Rede Bipartida do ano de 2020

- Primeiro semestre de 2020 (Estados): Os vértices do grupo azul foram Minas Gerais, Paraná e Rio Grande do Norte. O grupo verde reuniu Bahia, Ceará e São Paulo. Os outros estados ficaram no grupo rosa.

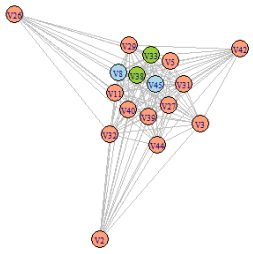
- Segundo semestre de 2020 (Estados): O grupo rosa foi composto por Rio de Janeiro e São Paulo, enquanto o grupo verde teve apenas o estado do Ceará. Os demais estados integraram o grupo azul.

- Primeiro semestre de 2020 (Países): O grupo azul contou com Finlândia, Lituânia, Noruega, Outros Países e Sérvia. Já o grupo verde incluiu Áustria, França, Irlanda, Reino Unido, Suíça e Turquia. Os demais países ficaram no grupo rosa.

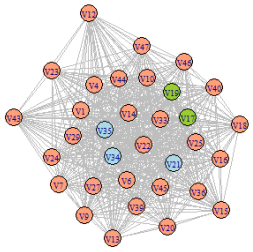
- Segundo semestre de 2020 (Países): O grupo azul foi formado por Espanha, França, Itália, Holanda, Portugal, Reino Unido e Ucrânia. O grupo verde incluiu Alemanha, Bélgica, Polônia e Suécia. Os outros países foram agrupados no grupo rosa.



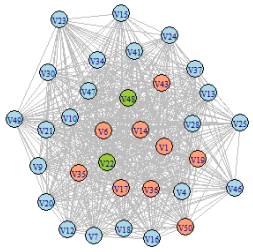
(a) Detecção de comunidades das UF´s do Brasil no 1º semestre de 2021 2021.1



(b) Detecção de comunidades das UF´s do Brasil no 2º semestre de 2021



(c) Detecção de comunidades dos Países Europeus no 1º semestre de 2021



(d) Detecção de comunidades dos Países Europeus no 2º semestre de 2021

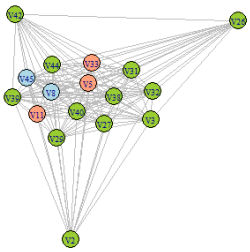
Figura 4.19: Detecção de Comunidades da Rede Bipartida do ano de 2021

- Primeiro semestre de 2021 (Estados): O grupo verde incluiu Bahia, Ceará, Pernambuco e Santa Catarina. Já o grupo rosa contou apenas com o estado do Rio de Janeiro. Os outros estados ficaram no grupo azul.

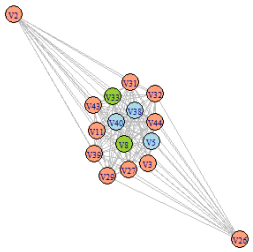
- Segundo semestre de 2021 (Estados): O grupo azul foi formado por Ceará e São Paulo, enquanto Pernambuco e Rio de Janeiro foram agrupados no grupo verde. Os demais estados ficaram no grupo rosa.

- Primeiro semestre de 2021 (Países): O grupo azul teve Irlanda, Portugal e Reino Unido, enquanto o grupo verde contou com França e Holanda. Os demais países ficaram no grupo rosa.

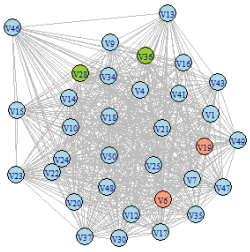
- Segundo semestre de 2021 (Países): O grupo rosa foi formado por Alemanha, Bélgica, França, Holanda, Portugal, Reino Unido, Rússia e Ucrânia. O grupo verde foi composto por Itália e Suíça. Os demais países integraram o grupo azul.



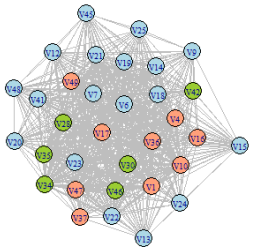
(a) Detecção de comunidades das UF´s do Brasil no 1º semestre de 2022 2022.1



(b) Detecção de comunidades das UF´s do Brasil no 2º semestre de 2022



(c) Detecção de comunidades dos Países Europeus no 1º semestre de 2022



(d) Detecção de comunidades dos Países Europeus no 2º semestre de 2022

Figura 4.20: Detecção de Comunidades da Rede Bipartida do ano de 2022

- Primeiro semestre de 2022 (Estados): O grupo azul foi composto por Ceará e São Paulo, enquanto o grupo rosa reuniu Bahia, Ceará e Distrito Federal. Os outros estados ficaram no grupo verde.

- Segundo semestre de 2022 (Estados): O grupo azul incluiu Bahia, Rio de Janeiro e Rio Grande do Norte, enquanto o grupo verde contou com Ceará e Pernambuco. Os demais estados permaneceram no grupo rosa.

- Primeiro semestre de 2022 (Países): O grupo rosa incluiu Bélgica e Holanda, enquanto o grupo verde foi composto por Noruega e Reino Unido. Os outros países ficaram no grupo azul.

- Segundo semestre de 2022 (Países): O grupo rosa reuniu Alemanha, Áustria, Dinamarca, Finlândia, França, Reino Unido, República Tcheca, Suíça e Ucrânia. O grupo verde incluiu Polônia, Portugal, Noruega, Rússia, Outros Países e Sérvia. Os

demais países ficaram no grupo azul.

Analisando de maneira geral temos que a divisão em três comunidades revela padrões, o grupo azul frequentemente reúne estados geograficamente ou economicamente influentes em sua interação com os países europeus. O grupo rosa, por sua vez, apresenta uma combinação de estados diversificados, incluindo capitais e grandes centros urbanos, que provavelmente desempenham um papel relevante nas conexões internacionais. Já o grupo verde tende a englobar estados menos conectados diretamente ou com menor intensidade nas relações com os países europeus. Destacando a interação entre estados brasileiros e países da Europa, sugerindo que há uma concentração de conexões em determinados estados em função de características específicas, como localização geográfica, infraestrutura ou relevância econômica. Além disso, a formação consistente de pequenos grupos, com mudanças semestrais e anuais, reflete variações no comportamento das interações ao longo do tempo.

A formação das comunidades revela padrões sobre a interação dos países europeus com os estados brasileiros. O grupo azul frequentemente inclui países que, embora geograficamente dispersos, possuem um perfil de interação consistente ao longo do tempo, como França, Reino Unido e Itália. O grupo rosa, por sua vez, apresenta uma concentração de países com forte relevância econômica e geopolítica, enquanto o grupo verde tende a agrupar países mais diversificados, incluindo tanto grandes economias quanto países menos conectados diretamente.

Essas divisões podem refletir diferenças no fluxo de turistas, trocas culturais ou mesmo relações econômicas entre os países europeus e os estados brasileiros. A mudança na composição das comunidades ao longo do tempo sugere que essas interações não são fixas, mas evoluem em resposta a fatores externos, como políticas, eventos internacionais ou mudanças nas conexões regionais.

Capítulo 5

Considerações Finais

O estudo realizado teve como objetivo compreender o fluxo turístico entre a Europa e o Brasil, com ênfase nos períodos antes, durante e após a pandemia da COVID-19. Através da aplicação de uma teoria estatística não abordada ao longo da graduação, foi possível extrair conclusões significativas sobre esse fluxo. Observou-se que, embora o Brasil tenha continuado a receber turistas durante a pandemia, houve uma queda expressiva no número de visitantes europeus nos anos de 2020 e 2021. Esse declínio pode ser atribuído às restrições de viagem e ao fechamento de fronteiras em diversos países, apesar de o Brasil ter mantido suas fronteiras abertas. Com a vacinação em massa e a adaptação ao "novo normal"em 2022, observou-se um aumento no número de turistas, indicando uma recuperação gradual do setor.

Além disso, o estudo enfrentou desafios na detecção de comunidades dentro do fluxo turístico devido à distribuição desigual de conexões entre os vértices analisados. Muitos vértices apresentavam poucas conexões, enquanto um pequeno número de vértices concentrava grande parte das conexões, dificultando a identificação de padrões claros de agrupamento. No entanto, a análise permitiu identificar que a divisão em três comunidades revela tendências significativas: o grupo azul frequentemente agrupa estados geograficamente ou economicamente influentes em sua interação com os países europeus; o grupo rosa apresenta uma combinação de estados diversificados, incluindo capitais e grandes centros urbanos, que desempenham um papel relevante nas conexões internacionais; e o grupo verde tende a englobar estados menos conectados diretamente ou com menor intensidade nas relações com os países europeus.

Essa divisão traz evidências de que há uma concentração de conexões em determinados estados, influenciada por fatores como localização geográfica, infraestrutura e relevância

econômica. Além disso, a formação consistente de pequenos grupos, com variações semestrais e anuais, reflete mudanças no comportamento das interações ao longo do tempo. Da mesma forma, a formação das comunidades revela padrões sobre a interação dos países europeus com os estados brasileiros. O grupo azul inclui países que, embora geograficamente dispersos, mantêm um perfil de interação consistente ao longo do tempo, como França, Reino Unido e Itália. O grupo rosa concentra países com forte relevância econômica e geopolítica, enquanto o grupo verde agrupa nações mais diversificadas, abrangendo tanto grandes economias quanto países com menor volume de conexões diretas.

Essas divisões podem refletir diferenças no fluxo de turistas, trocas culturais ou até mesmo relações econômicas entre os países europeus e os estados brasileiros. A mudança na composição das comunidades ao longo do tempo sugere que essas interações não são fixas, mas evoluem em resposta a fatores externos, como políticas governamentais, eventos internacionais ou transformações nas conexões regionais.

Para atrair um maior número de turistas europeus e impulsionar a economia nacional, sugere-se a implementação de estratégias de marketing direcionadas, destacando a diversidade cultural e natural do Brasil. O investimento em infraestrutura turística e em campanhas de promoção internacional pode aumentar a atratividade do país como destino. Especificamente para estados como Acre, Amapá e Mato Grosso do Sul, que registraram um número reduzido de visitantes, essa baixa demanda pode estar associada a fatores como infraestrutura limitada e o meio de transporte analisado no estudo. Como a pesquisa considerou apenas turistas vindos por via aérea e originários da Europa, isso pode ter influenciado a distribuição do fluxo turístico, uma vez que diferentes regiões do mundo possuem características culturais e atrativos distintos.

Diante disso, recomenda-se o desenvolvimento de pacotes turísticos que explorem as riquezas naturais e culturais dessas regiões. Além disso, parcerias com agências de turismo europeias para a criação de roteiros personalizados, bem como a promoção de eventos culturais e ecológicos, podem contribuir significativamente para o aumento do interesse por esses destinos.

Referências Bibliográficas

Csardi, G. e Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, **Complex Systems**, 1695.

Holland, P. W., Laskey, K. B. e Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, **5**(2), 109–137.

Karrer, B. e Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, **83**(1), 016107.

Lei, J. e Rinaldo, A. (2015). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, páginas 215–237.

Li, T., Levina, E. e Zhu, J. (2017). *randnet: Random Network Model Selection and Parameter Tuning*. R package version 0.1.

Newman, M. (2018). *Networks*. Oxford university press.

Opsahl, T. (2009). *Projecting two-mode networks onto weighted one-mode networks*. Available at: https://toreopsahl.com/2009/05/01/projecting-two-mode-networks-onto-weighted-one-mode-networks/.

Shigehalli, V. S. e Shettar, V. M. (2011). Spectral techniques using normalized adjacency matrices for graph matching.

Apêndice A

Código utilizado

A seguir é apresentado o código utilizado neste trabalho, foram utilizados oito código como este, mudando apenas o excel o qual os dados eram puxados e o filtro de mês que varia de acordo com o semestre em estudo.

```
rm(list=ls())
library(readxl)
library(readr)
library(stats)
library(base)
library(dplyr)
library(xtable)
library(igraph)
library(randnet)

dados = read_excel("D:/MEUS DOMUMENTOS/Desktop/Faculdade/TG/2019_atualizado.xlsx")

# Nomes das variaveis no banco
str(dados)

# Aqui é feita a seleção do continente europeu e do 1º semestre de 2019
Dados <- dados %>%
  filter(`cod continente` == "6", `cod mes` < 7 )
```

```
# Dados <- subset(dados,`cod mes`==1)
print(Dados)

attach(Dados)

# Obtendo os nomes de todos países e UFs
# Que aparecem no ano em estudo e no continente selecionado
list_nomes = names(table( c(names(table(País)),
                            names(table(UF)))))
print(list_nomes)

# Número de países e UFs
length(list_nomes)



# Retorna: Matriz de adj com pesos sendo o n. de Chegadas no UF
# Dados: matriz com os dados
# list_nomes: nomes dos países e UFs
criar_rede <- function(Dados, list_nomes) {
  adj.mat <- matrix(0, ncol = length(list_nomes), nrow = length(list_nomes))
  # Vamos percorrer as linhas do banco de dados
  # número de linhas é dim(Dados)[1]
  for (i in 1:dim(Dados)[1]) {
    # Selecionar o nome do país de origem da linha i
    k <- which(list_nomes == Dados$País[i])
    # Selecionar o nome do UF de destino da linha i
    l <- which(list_nomes == Dados$UF[i])
    # Selecionar o número de visitas/Chegadas da linha i
    # Salvar as chegadas na entrada (k, l) da matriz de adj
    adj.mat[k, l] = adj.mat[k, l] + Dados$Chegadas[i]
    print(paste("Linha", i, "de", dim(Dados)[1]))
    diag(adj.mat) <- 0
  }
```

```
  return(adj.mat)
}

# Criando a matriz de adj a partir da função criada
mat_adj <- criar_rede(Dados, list_nomes)

# Verificar se a matriz é simétrica
isSymmetric(mat_adj)
# Transformar a matriz em matriz simétrica
mat_adj = mat_adj + t(mat_adj)
isSymmetric(mat_adj)

# Verificar quais linhas (ou colunas) têm soma igual a 0
# Quando isso ocorre não temos nenhuma aresta conectada a esse vértice
linhas <- which(apply(mat_adj, 2, sum) == 0)
linhas
list_nomes[linhas]


# Retirar os vértices que não se conectam com ninguém
mat_adj <- mat_adj[-linhas, -linhas]
list_nomes <- list_nomes[-linhas]
print(list_nomes)
write.table(mat_adj, file= "adj_matrix_Jan_pesos.txt")

# Salvando o nome dos países e UFs
write.table(list_nomes, file = "Paises_UFs.txt", row.names = FALSE, col.names = FAL
print(list_nomes)



# Transformar a matriz de adj em binária apenas para a figura
# 0 são os paises e 1 os estados, essa matriz mostra as conexoes entre les, tem dia
```

```
#=0
mat_adj[which(mat_adj > 0, arr.ind = TRUE)] <- 1

# Salvando a matriz de adjacência binária
write.table(mat_adj, file= "adj_matrix_Jan_binaria.txt")

# Transformar a matriz de adj em objeto igraph
G <- graph_from_adjacency_matrix(mat_adj, mode=c("undirected"))

# Grau dos vértices(quantas conexões cada vértices tem )
degree(G)

# Verifica se é um grafo bipartido
# No caso de ser bipartido retorna qual grupo o nó pertence
#false = paises / true = estados
bipartite_mapping(G)

# Atribuir uma coordenada do eixo y, 1(paises) ou 2(estados), dependendo do seu grupo
V(G)$y <- (bipartite_mapping(G)$type*1) + 1
V(G)$y
# n: número de vértices
# n1: número de vértices do grupo 1
# n0: número de vértices do grupo 0
n <- gorder(G)
n1 <- length(which((bipartite_mapping(G)$type*1) == 1))
n0 <- length(which((bipartite_mapping(G)$type*1) == 0))
c(n1, n0)

# Atribuir uma coordenada do eixo x a cada vértice do grupo 1 e 2
dist <- 5
V(G)$x <- rep(0, n)
V(G)$x[which((bipartite_mapping(G)$type*1) == 1)] <- seq(0, dist*(n1-1), dist)
V(G)$x[which((bipartite_mapping(G)$type*1) == 0)] <- seq(0, dist*(n0-1), dist)
```

```
# Atribuir cores dependendo do grupo
V(G)$color <- rep('gray', n)
V(G)$color[which((bipartite_mapping(G)$type*1) == 1)] <- "yellow" #estados
V(G)$color[which((bipartite_mapping(G)$type*1) == 0)] <- "purple" #países
V(G)$label.cex <- 0.7

par(mfrow = c(1, 1))
plot(G, label.color = "black", vertex.size = 7, vertex.label = NA, asp = 0)

# Usar a matriz de adj COM pesos
#rm(list=ls()) #limpar memória

mat_adj <- read.table("adj_matrix_Jan_pesos.txt")
destinos <- unlist(read.table("Paises_UFs.txt"))
destinos <- as.vector(destinos)
G <- graph_from_adjacency_matrix(as.matrix(mat_adj),
                                 mode=c("undirected"),
                                 weighted = TRUE)
V(G)$type <- bipartite_mapping(G)$type

# Pesos das arestas
E(G)$weight

# Peso dos vértices
strength(G)


#filtrando vertices<50.000
vertices_estado <- strength(G)[V(G)$type == TRUE]
vertices_estado_filtrados <- vertices_estado[vertices_estado < 50000]
hist(vertices_estado_filtrados, col = "yellow", main = "Histograma dos pesos médios
summary(vertices_estado)
```

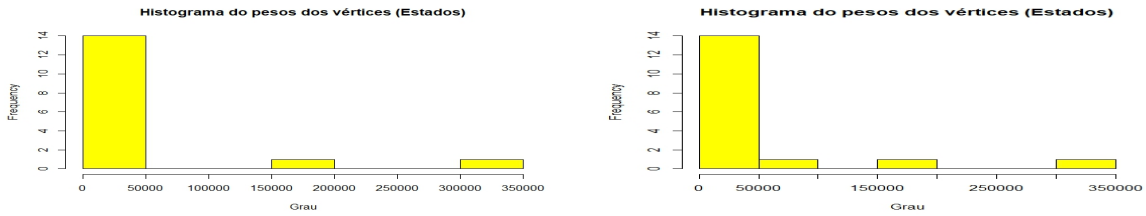
```
vertices_estado <- strength(G)[V(G)$type == TRUE]
vertices_estado_filtrados <- vertices_estado[vertices_estado < 50000]



#filtrando vertices<50.000
vertices_paises <- strength(G)[V(G)$type == FALSE]
vertices_paises_filtrados <- vertices_paises[vertices_paises < 50000]
hist(vertices_paises_filtrados, col = "purple", main = "Histograma dos pesos médios dos
summary(vertices_paises)

# Selecionar os nomes do tipo 1 e 2
#lista dos nomes separados e na ordem
estados<- destinos[V(G)$type == TRUE] #estados
print(estados)
paises <- destinos[V(G)$type == FALSE] #países
print (paises)
# Nós de cada tipo
n1 <- length(estados)
n2 <- length(paises)
c(n1, n2)



# Gerar a projeção do grafo bipartido
G.proj <- bipartite_projection(G)
names(G.proj)

# Figura dos grafos da projeção
plot(G.proj$proj1) #projeção dos países

# Ajuste o tamanho dos nós ao plotar o grafo
plot(G.proj$proj1,vertex.color = "red",vertex.size = 20, vertex.label = NA ) #projeções
```

```
plot(G.proj$proj2,vertex.color = "red",vertex.size = 20, vertex.label = NA) #projeç

# Grau de cada grafo da projeção
degree(G.proj$proj1) #projeção dos países
degree(G.proj$proj2) #projeções dos estados

# Pesos das arestas na projeção
E(G.proj$proj1)$weight #países
E(G.proj$proj2)$weight #estados

# Pesos dos vértices na projeção
strength(G.proj$proj1) #países
strength(G.proj$proj2) #estados


# Transformando o objeto igraph em uma matriz de adjacência
A <- as_adjacency_matrix(G.proj$proj2, type = c("both"), attr = "weight", sparse =


#---------------------------- K = 3 -------------------------------#
# Estimar as comunidades usando Spectral Clustering na projeção dos estados
set.seed(1000)
z.hat.SC <- reg.SSP(A, K = 3)$cluster

# Atribuir cores para as comunidades detectadas
col <- vector()
col[which(z.hat.SC == 1)] <- "olivedrab3"
col[which(z.hat.SC == 2)] <- "lightsalmon1"
col[which(z.hat.SC == 3)] <- "lightblue"
plot(G.proj$proj2,
     vertex.size = 16,
     vertex.color = col,
     vertex.label.cex = 0.55,
```

```
     edge.width = E(G.proj$proj2)$weight / max(E(G.proj$proj2)$weight) * 2,
     edge.color = "grey")

# Transformando o objeto igraph em uma matriz de adjacência
B <- as_adjacency_matrix(G.proj$proj1, type = c("both"), attr = "weight", sparse = FALSE

# Estimar as comunidades usando Spectral Clustering na projeção dos paises
set.seed(1000)
z.hat.SC <- reg.SSP(B, K = 3)$cluster

# Atribuir cores para as comunidades detectadas
col <- vector()
col[which(z.hat.SC == 1)] <- "olivedrab3"
col[which(z.hat.SC == 2)] <- "lightsalmon1"
col[which(z.hat.SC == 3)] <- "lightblue"
plot(G.proj$proj1,
     vertex.size = 16,
     vertex.color = col,
     vertex.label.cex = 0.55,
     edge.width = E(G.proj$proj1)$weight / max(E(G.proj$proj1)$weight) * 2,
     edge.color = "grey")

#nome dos paises e estaos em ordem, retirando os vertices sem conexão
print(list_nomes)
print(summary(vertices_estado_filtrados))
print(summary(vertices_paises_filtrados))
```

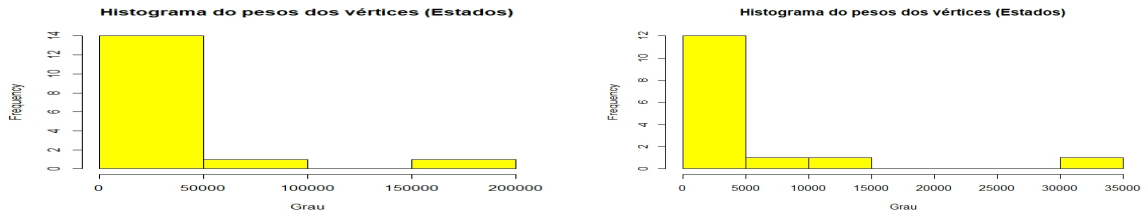
Apêndice B

Histogramas do peso dos vértices dos Estados



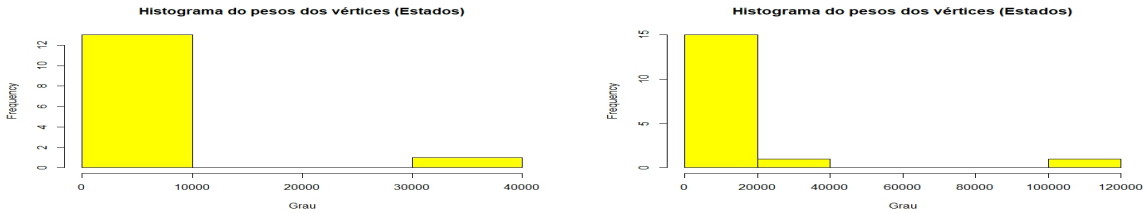
(a) 2019.1 (b) 2019.2

Figura B.1: Histograma do peso dos vértices (Estados)



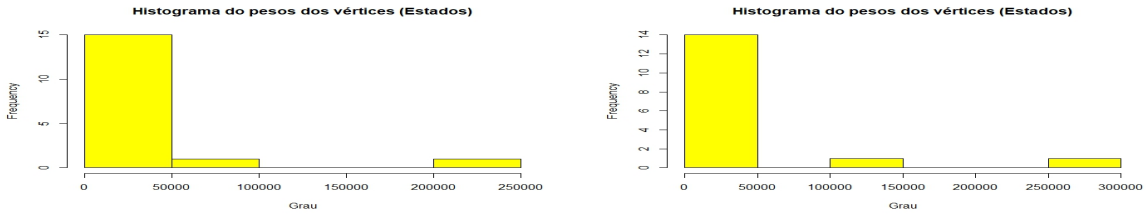
(a) 2020.1 (b) 2020.2

Figura B.2: Histograma do peso dos vértices (Estados)



(a) 2021.1 (b) 2021.2

Figura B.3: Histograma do peso dos vértices (Estados)

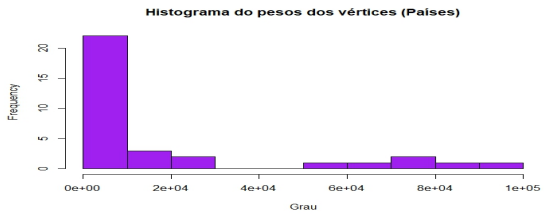


(a) 2022.1 (b) 2022.2

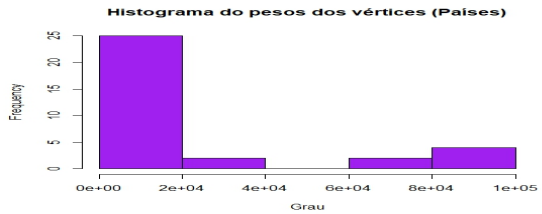
Figura B.4: Histograma do peso dos vértices (Estados)

Apêndice C

Histogramas dos pesos dos vértices dos Países

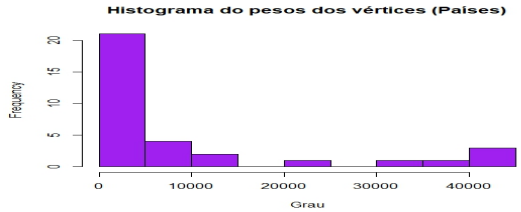


(a) 2019.1

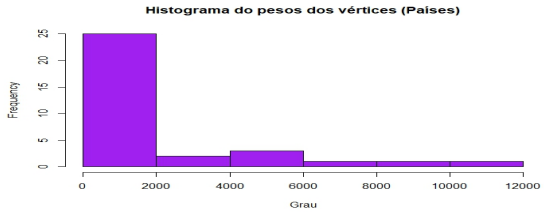


(b) 2019.2

Figura C.1: Histograma do peso dos vértices (Países)

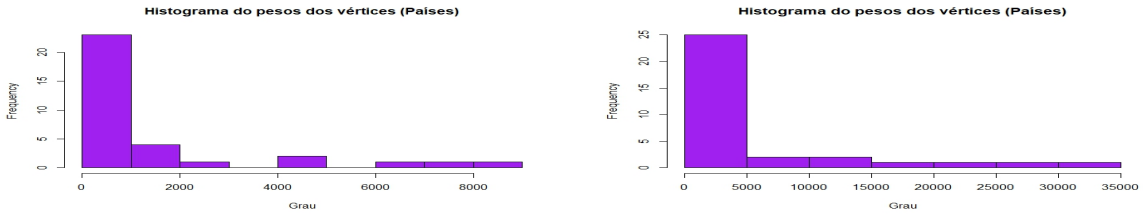


(a) 2020.1



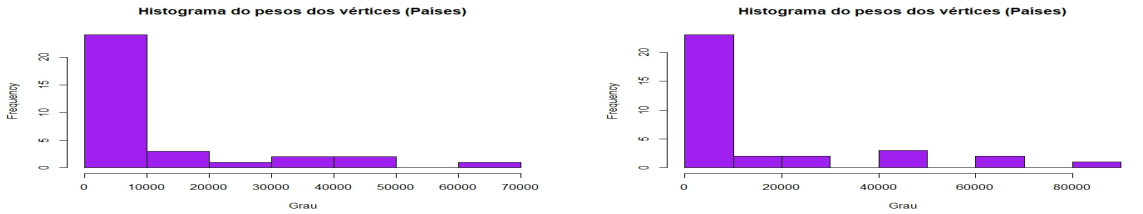
(b) 2020.2

Figura C.2: Histograma do peso dos vértices (Países)



(a) 2021.1 (b) 2021.2

Figura C.3: Histograma do peso dos vértices (Países)



(a) 2022.1 (b) 2022.2

Figura C.4: Histograma do peso dos vértices (Países)