



**UNIVERSIDADE FEDERAL DE OURO PRETO
DEPARTAMENTO DE COMPUTAÇÃO**



**ANÁLISE DE SÉRIES TEMPORAIS POR MEIO DE REDES
COMPLEXAS**

**RELATÓRIO FINAL DO PROJETO DE INICIAÇÃO CIENTÍFICA
EDITAL 03/2020 PIBIC/CNPQ-2020-2021
VIGÊNCIA: 01/09/2020 A 31/08/2021**

IURI DA SILVA DINIZ

ORIENTADOR: PROF. DR. VANDER LUIS DE SOUZA FREITAS

Ouro Preto - Minas Gerais - Brasil, Agosto de 2021

RESUMO

Séries temporais são sequências de observações de um dado fenômeno, ordenadas no tempo, que capturam tanto eventos isolados quanto características globais do sistema. Uma abordagem para análise destas séries se dá a partir do arcabouço de redes complexas, que são grafos compostos por arestas e vértices (nós), onde os vértices representam as entidades que compõem o sistema dinâmico em estudo e as arestas mediam as interações entre elas. O presente projeto tem por objetivo gerar redes complexas a partir de séries temporais reais, utilizando uma técnica ainda em desenvolvimento, mas que já apresentou resultados promissores na caracterização de séries sintéticas de sistemas não-lineares, capturando comportamentos como periodicidade e dinâmica caótica. Como estudo de caso, explorou-se a formação de uma rede geográfica a partir da análise de séries temporais de radares meteorológicos. Os resultados obtidos se mostraram competitivos com técnicas tradicionais da literatura.

Palavras-chave: séries temporais, redes complexas, grafos.

SUMÁRIO

1	INTRODUÇÃO	4
2	OBJETIVOS	5
3	REVISÃO DA LITERATURA	6
3.1	Séries temporais	6
3.2	Redes complexas e Grafos	6
3.2.1	<i>Métricas de rede</i>	6
3.3	Algoritmos para mapeamento de séries temporais em redes complexas	7
3.3.1	<i>Dynamical Characterization with the Top Integral Function (DCTIF)</i>	7
3.3.2	<i>Dynamical Characterization with Symbolic Dynamics (DCSD)</i>	8
3.3.3	<i>Visibility Graphs (VG)</i>	9
3.4	Medidas de distância	11
3.4.1	<i>Distância Euclidiana (Euclidean Distance)</i>	11
3.4.2	<i>Distância de Manhattan (Manhattan Distance)</i>	11
3.4.3	<i>Distância de Canberra (Canberra Distance)</i>	12
3.4.4	<i>Similaridade por Cossenos (Cosine Distance)</i>	12
3.4.5	<i>Distância de Jensen-Shannon (Jensen-Shannon Distance)</i>	12
4	MATERIAIS E MÉTODOS	14
4.1	Replicação: <i>Complex Networks Approach for Dynamical Characterization of Nonlinear Systems</i>	14
4.2	Nova abordagem: <i>Geographical Complex Networks applied to describe meteorological data</i>	14
5	RESULTADOS E DISCUSSÕES	17
5.1	Distância Euclidiana	17
5.2	Distância de Manhattan	20
5.3	Distância de Canberra	24
5.4	Similaridade por Cosseno	27
5.5	Distância de Jensen-Shannon	31
5.6	Formação da rede geográfica	34
5.6.1	<i>Rede formada por meio da comparação usando similaridade por cosseno</i>	34
5.6.2	<i>Rede formada por meio da comparação usando distância de Jensen-Shannon</i>	36
6	CONCLUSÃO	38

7	PRODUÇÃO CIENTÍFICA	39
---	---------------------------	----

REFERÊNCIAS	40
-------------------	----

1 INTRODUÇÃO

O entendimento deste relatório se concentra, inicialmente, em torno de três conceitos principais: séries temporais, sistemas complexos e grafos. Assim, o objeto de análise, a série temporal, é estudada por meio das redes complexas. Uma série temporal pode ser definida como um conjunto de observações dispostas ao longo do tempo e distribuídas em intervalos contínuos ou discretos ([BROCKWELL; DAVIS, 2016](#)). Os sistemas complexos, por outro lado, possuem um comportamento global definido por interações locais, sem um controlador único e centralizado, fazendo com que a observação da dinâmica dessa classe de sistemas passe por uma observação comportamental de seus componentes ([BOCCARA, 2010](#)). Quando um sistema complexo é representado com o uso de grafos — onde vértices se conectam por meio de arestas a fim de representar algum comportamento —, tem-se então as redes complexas, uma ciência aplicada em diversas áreas e que, recentemente, vem sendo incorporada na análise de séries temporais.

Resultados apresentados por [Freitas, Lacerda e Macau \(2019\)](#) e [Yu, Jia e Jian \(2013\)](#), por exemplo, ilustram a eficiência do uso das redes complexas na caracterização das dinâmicas de séries sintéticas (Mapa logístico e mapa de Henón), fornecendo bases que justificam a continuidade dos estudos nessa área.

2 OBJETIVOS

O objetivo geral deste projeto é gerar uma rede geográfica a partir de séries temporais de um radar meteorológico, a partir de técnicas de análise de séries temporais por meio de redes complexas.

Os objetivos específicos são:

- Realizar uma revisão de literatura sobre os métodos utilizados para resolver o problema abordado.
- Modificar, adaptar e aperfeiçoar algoritmos baseados em redes complexas para análise de séries temporais.
- Preparar as séries temporais a partir dos dados de radar para representação em forma de redes.
- Gerar uma rede geográfica a partir das diversas séries temporais.
- Explorar e comparar os algoritmos usados para conversão de séries temporais em redes complexas.
- Estudar sistemas com base na análise de suas representações como redes complexas.

3 REVISÃO DA LITERATURA

3.1 Séries temporais

As séries temporais são formadas por um conjunto de observações distribuídas ao longo do tempo (em intervalos contínuos ou discretos), estando presente nas mais diversas áreas do conhecimento (BROCKWELL; DAVIS, 2016). A análise dessas séries, segundo Cryer (1986), é baseada em dois propósitos gerais: identificar um modelo capaz de reproduzir a distribuição desses dados; tornar possível a predição de observações ainda não medidas, sendo viável também, segundo Brockwell e Davis (2016) a predição de valores baseado no comportamento de outra série temporal.

3.2 Redes complexas e Grafos

A interação entre certas populações de animais é comumente usada para exemplificar e explicar sistemas complexos. Boccara (2010) e Amaral e Ottino (2004) usam o exemplo de interação entre formigas e gansos, respectivamente, para destacar os pontos que determinam sistemas como complexos. Inerente a esses dois sistemas (revoada de gansos e colônia de formigas) está a adaptabilidade e a falta de um controle centralizado, pois ambos apresentam uma dinâmica que se altera de acordo com o ambiente e são desprovidos de um líder para gerenciar o comportamento global. A presença dessas características é definida por Boccara (2010) como comportamento emergente, onde um evento de impacto global dependente de interações locais entre agentes, que, mesmo simples, são difíceis de prever. Como descrito por Amaral e Ottino (2004), uma das ferramentas usadas na análise de sistemas complexos é a teoria das redes, que por sua vez faz o uso da teoria dos grafos para representação de um dado sistema.

Os grafos foram criados pelo matemático Leonhard Euler como auxílio no entendimento de um problema real. Na representação, Euler usou quatro letras para identificar diferentes porções de terra e as conectou de acordo com a presença de pontes entre essas áreas. A partir daí, toda uma teoria foi desenvolvida, e os grafos se tornaram parte essencial para o desenvolvimento da teoria das redes. É necessário, porém, apontar algumas diferenças conceituais entre redes e grafos. Grafos possuem vértices que se ligam através de arestas a fim de representar um comportamento desejado, mas sem representar sistemas reais. As redes são formadas por grafos que representam sistemas reais, o que introduz outra nomenclatura: vértices são também chamados de nós e as arestas de links (BARABÁSI et al., 2016).

3.2.1 Métricas de rede

O desenvolvimento da metodologia de comparação entre as redes adotada neste trabalho depende de métricas que relacionam e quantificam diversas características de um grafo. Muitas

métricas dependem do entendimento do conceito de caminho. Um caminho entre dois vértices de um grafo é a relação das arestas percorridas para compor esse trajeto; o comprimento desse caminho, portanto, é dado pela quantidade de arestas percorridas. Assim, é possível definir também o caminho mínimo, dado pelo menor conjunto de arestas necessários para conectar dois vértices (BARABÁSI et al., 2016).

O diâmetro é dado pelo comprimento do maior caminho mínimo entre dois vértices, enquanto a densidade do grafo é a relação entre o número de arestas existentes e o número de arestas possíveis ($L_{máx}$), que, para os grafos abordados neste trabalho, é definida como

$$L_{máx} = \frac{N(N - 1)}{2}, \quad (3.1)$$

onde N é o número de vértices.

A medida *Betweenness* quantifica a relação entre o número de caminhos mínimos que passam por uma aresta (ou vértice) e o número total de caminhos mínimos existentes; *Closeness* define a proximidade de um dado vértice a todos os outros do grafo, ao levar em conta o caminho mínimo entre o vértice analisado e todos os outros vértices do grafo. O Grau (*degree*) de um vértice é dado pela quantidade de ligações realizadas por ele, o seu número de arestas. Se um grafo for ponderado, indicando que as arestas possuem pesos numéricos, tem-se a definição da medida *Strength*, que é o somatório dos pesos das arestas que se ligam a um vértice. Outra medida relacionada ao *Degree*, o parâmetro de heterogeneidade, é dado por:

$$\kappa = \frac{\langle k^2 \rangle}{\langle k^2 \rangle^2}, \quad (3.2)$$

onde $\langle k \rangle$ é o grau médio da rede e $\langle k^2 \rangle$ é a média dos quadrados dos graus.

Quanto mais homogênea for a distribuição, mais próximo κ será de 1. Distribuições heterogêneas produzirão valores de κ maiores que 1 (MENCZER; FORTUNATO; DAVIS, 2020).

3.3 Algoritmos para mapeamento de séries temporais em redes complexas

A análise das séries, como o próprio título do projeto já diz, é realizada por meio de redes complexas. Isso se torna possível através do uso de algoritmos capazes de converter uma série em uma rede. Essas redes, porém, precisam herdar determinadas características da série, pois só assim será possível extrair informações passíveis de análise. Os algoritmos utilizados durante o projeto foram: *Dynamical Characterization using Symbolic Dynamics (DCSD)*, *Dynamical Characterization with the Top Integral Function (DCTIF)* e *Visibility Graphs (VG)*.

3.3.1 *Dynamical Characterization with the Top Integral Function (DCTIF)*

O algoritmo DCTIF, apresentado por Freitas, Lacerda e Macau (2019), realiza essa conversão através do mapeamento dos valores de uma série temporal normalizada em uma

sequência de valores inteiros, de tamanho N , que irá corresponder ao conjunto de nós da rede resultante. A relação matemática que define esse mapeamento é dada por

$$Y_k = [N \cdot i_k] = \min\{i \in \mathbb{Z} \mid N \cdot i_k \leq i\}, \quad (3.3)$$

onde:

- Y_k é o k -ésimo elemento resultante do mapeamento;
- N define o tamanho do intervalo onde uma série de entrada será mapeada, e consequentemente o número de nós da rede resultante;
- i_k é o k -ésimo elemento da série a ser mapeada, distribuídos no intervalo $[0,1]$;
- i é um elemento pertencente ao conjunto dos números inteiros.

De posse do novo conjunto, com valores distribuídos de 1 a N , são estabelecidas as ligações entre um dado elemento e seu sucessor. Algumas situações, como a permissão de autoconexões, devem ser avaliadas para cada caso.

3.3.2 *Dynamical Characterization with Symbolic Dynamics (DCSD)*

O próximo algoritmo de caracterização dinâmica, o DCSD, também explorado em Freitas, Lacerda e Macau (2019), realiza o mapeamento de uma série com valores reais em um conjunto binário, que depois será convertido para o sistema decimal usando agrupamentos binários de tamanho N . Esses agrupamentos terão um comportamento deslizante, iniciando no primeiro elemento e se movendo uma posição a cada conversão.

A Figura 1 exemplifica o desenvolvimento do algoritmo DCSD através de 3 etapas básicas. Inicialmente, temos uma distribuição que simula as observações de uma série, podendo ser evidenciado a presença de um segmento contínuo que divide os elementos em dois grupos. Essa linha é traçada a partir de um ponto, p , definido por

$$p = \frac{\max(x) + \min(x)}{2}, \quad (3.4)$$

indicando que a série será agrupada em termos da média aritmética entre a maior e a menor observação. A partir daí, a série binária (simbólica) é gerada tomando como 1 os elementos maiores que p e 0 os elementos menores que p . De posse desse conjunto binário — o tamanho da série binária é o mesmo da série original —, define-se uma janela de tamanho fixo que servirá para realizar a conversão binário/decimal, tal como mostrado na segunda etapa da Figura 1. Ao final teremos uma nova série no sistema numérico decimal, onde cada elemento nessa nova série corresponderá a um nó da rede. Assim como citado para o DCTIF, a definição dos links entre esse conjunto de nós precisa ser avaliada particularmente.

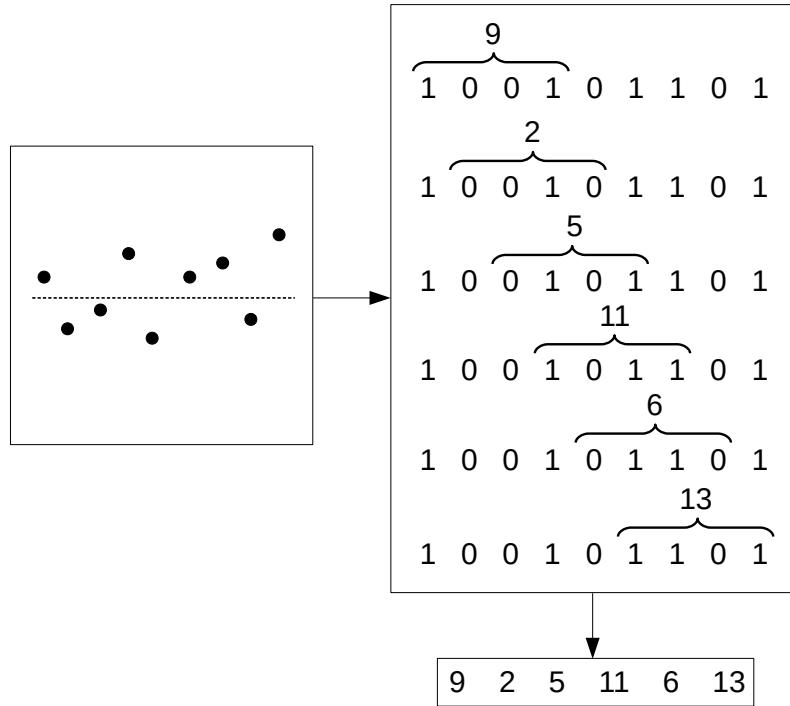


Figura 1 – Funcionamento do algoritmo DCSD.

3.3.3 Visibility Graphs (VG)

Os grafos de visibilidade, apresentados por [Lacasa et al. \(2008\)](#), diferente dos algoritmos de caracterização dinâmica apresentados anteriormente, possuem a característica de manutenção da quantidade de elementos da série a ser convertida. Assim, uma série com N observações será convertida em um grafo com N vértices. As conexões, porém, serão feitas de acordo com um critério específico de visibilidade, tal como mostra a Figura 2. Em termos abstratos, uma conexão entre duas barras verticais (t_a, y_a) e (t_b, y_b) só é feita se for possível “ver” uma barra a partir da outra, ou seja, é possível traçar uma linha reta entre elas. Do ponto de vista quantitativo, essa conexão de visibilidade pode ser dada pela relação

$$y_c < y_b + (y_a - y_b) \left(\frac{t_b - t_c}{t_b - t_a} \right), \quad (3.5)$$

que avalia a visibilidade entre os pontos a e b levando em conta um ponto c , disposto entre eles. Assim, são levantadas todas as relações entre os pares de pontos possíveis em uma determinada série temporal, sendo que a não ocorrência de uma observação (c) entre um par de pontos (a e b) implica em uma relação verdadeira em (3.5) e faz com que um link seja criado entre os nós correspondentes a esses pontos. Havendo um ponto intermediário entre o par analisado, o link só

será criado se a ordenada do ponto y_c for menor que a parcela da direita na relação (3.5).

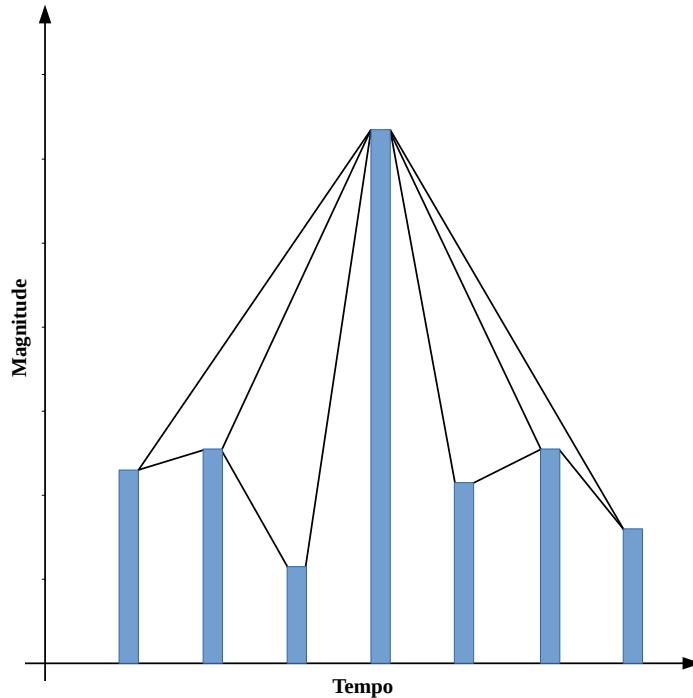


Figura 2 – Demonstração das conexões do algoritmo VG.

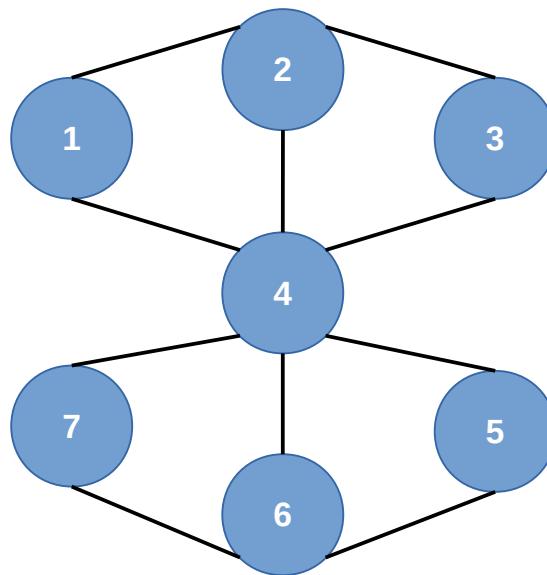


Figura 3 – Grafo formado através do algoritmo VG.

A Figura 3 exibe o grafo resultante da aplicação do VG no exemplo da Figura 2. O *layout* do grafo, ou seja, a disposição dos nós, foi feita com o intuito de destacar uma característica importante: os valores de maior magnitude em uma série corresponderão aos vértices mais

conectados no grafo, ou, em função das terminologias aplicadas aos grafos, vértices com maiores valores de grau.

3.4 Medidas de distância

A comparação entre as redes geradas através das séries temporais estudadas em uma das etapas deste trabalho foi feita mediante o uso de cinco medidas de distâncias já conhecidas na literatura. Foram elas: Euclidiana, Manhattan, Camberra, Cosseno e Jensen-Shannon. Cada uma possui características diferentes, o que possibilitou a realização de diversas comparações.

3.4.1 Distância Euclidiana (*Euclidean Distance*)

A distância Euclidiana entre dois pontos, p e q , no espaço bi-dimensional \mathbb{R}^2 , produz uma formulação amplamente conhecida, dada por:

$$d_{p,q} = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}. \quad (3.6)$$

Podemos ver que, para o caso bi-dimensional, tem-se a aplicação do quadrado da diferença entre as coordenadas (abscissa e ordenada) de cada ponto. Em um caso genérico, considerando a distância entre dois pontos de n dimensões reais (\mathbb{R}^n), reescreve-se a relação na forma:

$$d_{p,q} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}. \quad (3.7)$$

3.4.2 Distância de Manhattan (*Manhattan Distance*)

A distância de Manhattan, também conhecida como *City Block* ou *Taxicab Geometry*, como é abordado em [Thompson e Dray \(2000\)](#), pode ser apresentada de forma semelhante à distância euclidiana. Para um caso bi-dimensional em \mathbb{R}^2 , é dado:

$$d(p, q) = |p_x - q_x| + |p_y - q_y|. \quad (3.8)$$

Generalizando em termos de n dimensões reais (\mathbb{R}^n), tem-se:

$$d_{p,q} = \sum_{i=1}^n |p_i - q_i|. \quad (3.9)$$

Para o caso bi-dimensional, a relação entre a distância euclidiana e a distância de Manhattan é de fácil visualização, tal como mostra a Figura 4. A partir da definição dos dois pontos, p e q , evidencia-se um triângulo retângulo de cuja hipotenusa corresponde à distância euclidiana e a soma dos catetos corresponde à distância de Manhattan.

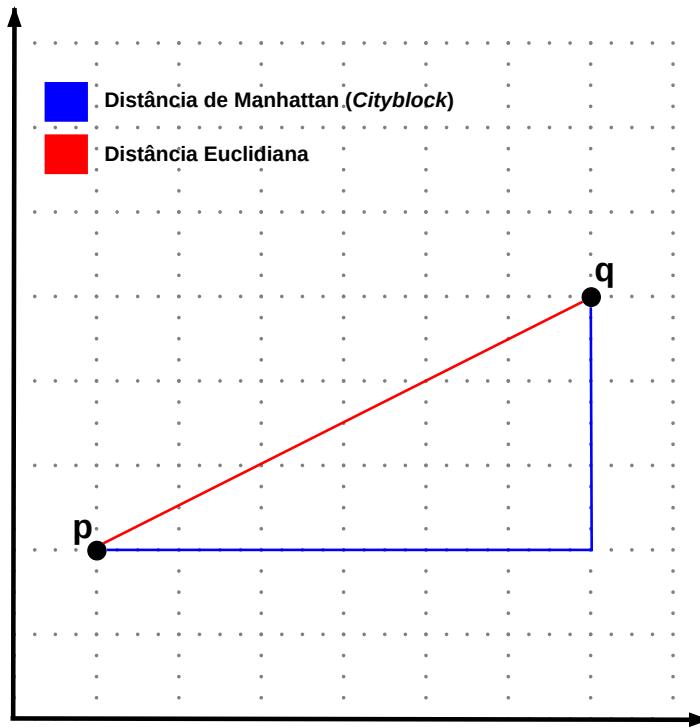


Figura 4 – Relação entre a distância euclidiana e a distância de *Manhattan*.

3.4.3 Distância de Camberra (*Canberra Distance*)

[Jurman et al. \(2009\)](#) definem a distância de *Canberra* entre dois vetores no espaço \mathbb{R}^n como:

$$d_{(p,q)} = \sum_{i=1}^n \frac{|p_i - q_i|}{|p_i| + |q_i|}. \quad (3.10)$$

3.4.4 Similaridade por Cossenos (*Cosine Distance*)

Usada em [George et al. \(2015\)](#) para analisar a similaridade entre amostras de áudio, similaridade por cosseno pode ser extraída a partir do produto escalar entre dois vetores:

$$|p \cdot q| = \|p\| \|q\| \cos \theta, \quad (3.11)$$

$$\cos \theta = \frac{|p \cdot q|}{\|p\| \|q\|}. \quad (3.12)$$

Sabendo que o módulo (norma) de um vetor no \mathbb{R}^n é dada por:

$$\|p\| = \sqrt{\sum_{i=1}^n p_i^2}. \quad (3.13)$$

3.4.5 Distância de Jensen-Shannon (*Jensen-Shannon Distance*)

A divergência de Kullback-Leibler, base para distância de Jensen-Shannon, é apresentada em [Cabral \(2013\)](#) como uma medida comumente utilizada para aproximar duas distribuições de

probabilidade. Através dela é possível verificar quanto de informação se perde ao se usar uma distribuição para explicar outra. A divergência de Kullback-Leibler entre duas distribuições de probabilidade p e q é dada por:

$$D_{KL}(p,q) = \sum_{i=1}^n p_i \log_2 \frac{p_i}{q_i}. \quad (3.14)$$

A divergência de Jensen-Shannon é definida a partir de (3.14), sendo:

$$D_{JS}(p,q) = \frac{D_{KL}(p,m) + D_{KL}(q,m)}{2}, \quad (3.15)$$

onde:

$$m = \frac{p+q}{2}. \quad (3.16)$$

Sendo definida tais questões, pode-se, agora, desenvolver a distância de Jensen-Shanon como a raiz quadrada de (3.15).

4 MATERIAIS E MÉTODOS

O processo de desenvolvimento das atividades passou por fases iniciais de revisões literárias e amadurecimento nos conhecimentos em linguagem de programação Python. As bibliotecas *igraph* e *networkx* foram largamente utilizadas durante as atividades, pois fornecem funções para a criação, obtenção de métricas e exibição de grafos.

4.1 Replicação: *Complex Networks Approach for Dynamical Characterization of Nonlinear Systems*

Após as etapas iniciais de revisão, iniciou-se a reprodução dos resultados do artigo desenvolvido por [Freitas, Lacerda e Macau \(2019\)](#). Esse trabalho explorou o uso de redes complexas na identificação das dinâmicas de sistemas não lineares, especificamente a dinâmica das séries sintéticas geradas pelo mapa logístico e pelo mapa de Henón. O uso de algoritmos que mapeiam as observações de uma série em um conjunto de nós possibilitou a geração de redes que puderam ser analisadas através do uso de métricas já desenvolvidas no estudo de redes. A atividade desenvolvida sobre o artigo supracitado ocorreu de duas formas: reprodução dos mesmos resultados e geração de novos resultados advindos da variação de um parâmetro em um dos algoritmos usados para converter as séries em conjuntos de nós.

Com o acesso aos códigos utilizados no trabalho foi possível replicar e revisar os resultados obtidos, assim como desenvolver uma comparação baseada na alteração do parâmetro de geração das redes no algoritmo DCTIF, já descrito anteriormente. Dois novos valores para o parâmetro N foram testados: 512 e 2048, o que corresponde, respectivamente, à metade e ao dobro do valor originalmente utilizado. Tais modificações tiveram o intuito de averiguar o comportamento dos resultados mediante valores variados de intervalos para mapeamento das observações das séries sintéticas analisadas.

4.2 Nova abordagem: *Geographical Complex Networks applied to describe meteorological data*

As atividades seguintes foram realizadas tendo como base o trabalho desenvolvido por [Jorge, Costa e Santos \(2020\)](#). Esse trabalho analisou os dados de precipitação distribuídos em pontos ao longo do espaço geográfico da bacia de Tamanduateí no estado São Paulo. Cada ponto possuía uma série temporal associada e, através do uso da correlação de Pearson, foi possível estabelecer um link entre as séries cuja correlação estava acima de um limiar determinado. Os resultados mostraram a presença de uma correlação espaço-temporal, onde as séries com maior valor de correlação de Pearson eram as mais próximas geograficamente. Partindo dos mesmos dados e da mesma premissa apresentada nesse trabalho, iniciou-se o desenvolvimento

de uma outra abordagem para o alcance desses resultados. As séries temporais, antes comparadas entre si, foram agora convertidas em redes com o intuito de investigar características que pudessem evidenciar a correlação espaço-temporal encontrada no trabalho base. Alguns métodos de comparação de redes foram explorados.



Figura 5 – Disposição geográfica das séries temporais ao longo da região da bacia de Tamanduateí. Cada ponto representa uma série.

A Figura 5 mostra a região de estudo. São mostrados 587 pontos divididos ao longo da região da bacia, com resolução espacial de 1km. Além da divisão espacial, tem-se a discretização dos dados no tempo, com amostras a cada 10min durante todo o mês de janeiro de 2015. Desse modo, cada ponto corresponde a uma série temporal com mais de 4000 observações.

A partir da preparação das séries temporais, o próximo passo consistiu na definição da metodologia usada para investigação da existência de uma correlação entre elas. Assim como proposto no trabalho original, tendo como intuito uma replicação inicial dos resultados, fez-se o uso da correlação de Pearson entre todos os pares possíveis de séries.

A próxima etapa visou o uso dos algoritmos DCTIF, DCSD e VG para conversão das séries temporais em redes. A partir da conversão, novas estratégias precisaram ser definidas para comparação desses novos elementos, como o uso das medidas de distância descritas na Seção 3.4. A ideia por trás do uso de tais medidas consistiu na aplicação das mesmas durante a comparação de características extraídas de cada rede. Assim, comparava-se algumas características de pares de redes a fim de quantificar e classificar as relações entre todo o conjunto.

Como primeira estratégia de comparação, usou-se um vetor contendo características

que detinham informações sobre aspectos globais da rede e média aritmética de medidas que apresentam informações para cada nó, sendo: grau médio, *betweenness* médio, *closeness* médio, densidade e diâmetro. Esses vetores foram obtidos para cada rede gerada a partir de cada um dos três algoritmos de conversão utilizados. Sendo assim, esses dados foram comparados entre si — apenas há comparação entre redes geradas através do mesmo algoritmo — por meio das 5 medidas de distância propostas. Baseada nessa estratégia apresentada, derivou-se uma segunda, focada na criação de um vetor contendo métricas globais e outro vetor contendo a média de métricas locais. Assim, cada um desses vetores, extraídos de cada uma das redes, também são comparados através das mesmas medidas de distância. Como composição do vetor de métricas globais, temos o parâmetro de heterogeneidade, diâmetro e densidade. Já o vetor contendo as médias das métricas locais ficou constituído por grau médio, *betweenness* médio e *closeness* médio.

Em resultados discutidos por [Lacasa et al. \(2008\)](#), é apontado como o VG é capaz de herdar características das séries, tendo distribuições de graus específicas para determinadas condições. Baseado nessa observação, uma outra abordagem começou a ser explorada, passando a utilizar sequência de métricas (primeiramente grau e posteriormente *betweenness* e *closeness*) como elementos a serem comparados. Tais sequências não são distribuições probabilísticas, e sim listas contendo um valor de métrica (o grau de um nó, por exemplo) para cada um dos nós da rede.

Todas essas abordagens, essas definições de como criar um grupamento de métricas passíveis de comparação via medidas de distância, foram pensadas e aplicadas para os 3 algoritmos responsáveis pela criação das redes. Porém, além dessas abordagens, alterações nos algoritmos DCSD e DCTIF também foram propostas com o intuito de explorar e avaliar novos padrões. Essas alterações consistiram na adição de pesos nas ligações entre os nós, sendo que o peso de uma ligação é definido a partir da observação da quantidade de vezes que essa conexão aparece repetida na lista de adjacências gerada ao fim da execução do algoritmo. De posse dessa informação, os algoritmos DCSD e DCTIF passaram a ser classificados em DCSD ponderado e não-ponderado (wDCSD e uDCSD, respectivamente) e DCTIF ponderado e não-ponderado (wDCTIF e uDCTIF, respectivamente).

5 RESULTADOS E DISCUSSÕES

Em seção anterior foi realizada a apresentação do trabalho desenvolvido por [Jorge, Costa e Santos \(2020\)](#) e a definição das abordagens exploradas a partir dele. Agora, a fim de observar o comportamento dos dados obtidos a partir da comparação entre todas as redes que compõem a região analisada, foram utilizados recursos de visualização que permitiram investigar a relação entre os valores de distância fornecidos pelas métricas e a distância geográfica (em quilômetros) entre cada uma das séries temporais analisadas. As seções a seguir ilustram essas relações levando em conta as 5 métricas de distância utilizadas nas comparações.

Sabendo que o número de séries temporais é 587 e que as comparações feitas entre pares iguais e simétricos foram descartadas, o número de elementos resultantes para cada comparação foi de 171991.

5.1 Distância Euclidiana

Partindo da Figura 6 até a Figura 11, são mostradas as representações em boxplot que relacionam a distância entre um par de séries na região analisada (eixo das abscissas) e o valor da distância euclidiana entre elas.

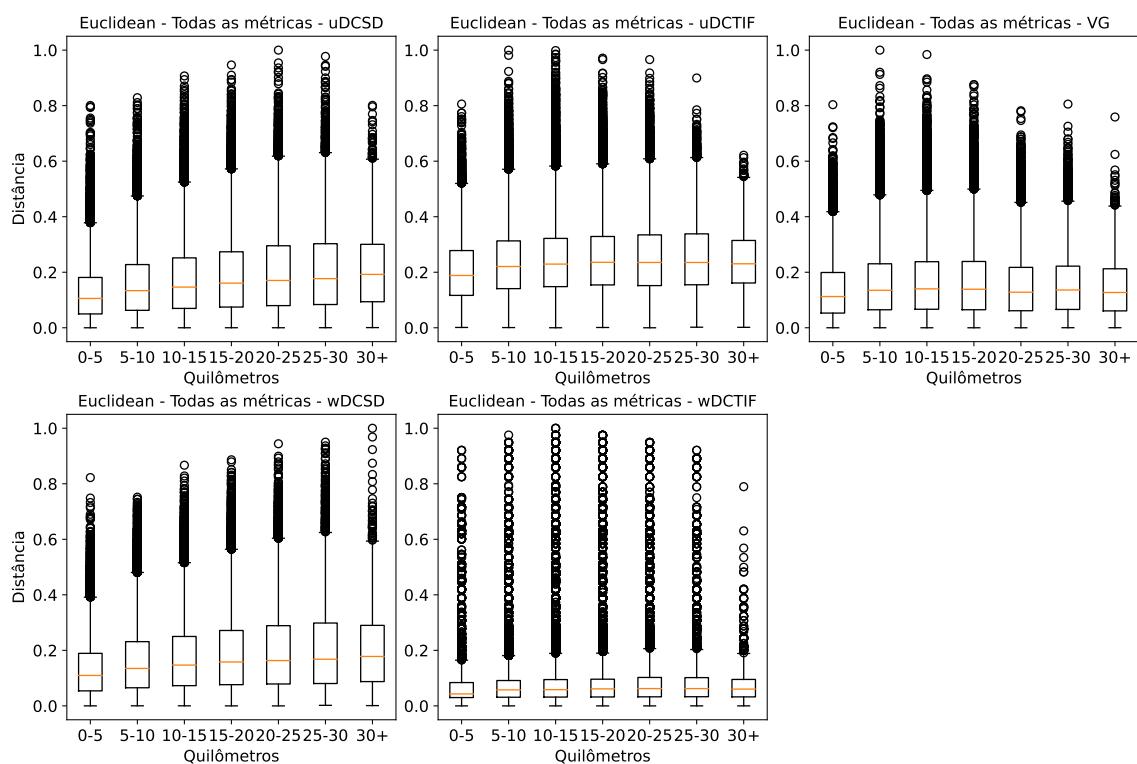


Figura 6 – Relação entre a distância geográfica e a distância Euclidiana para comparação usando vetor de várias métricas.

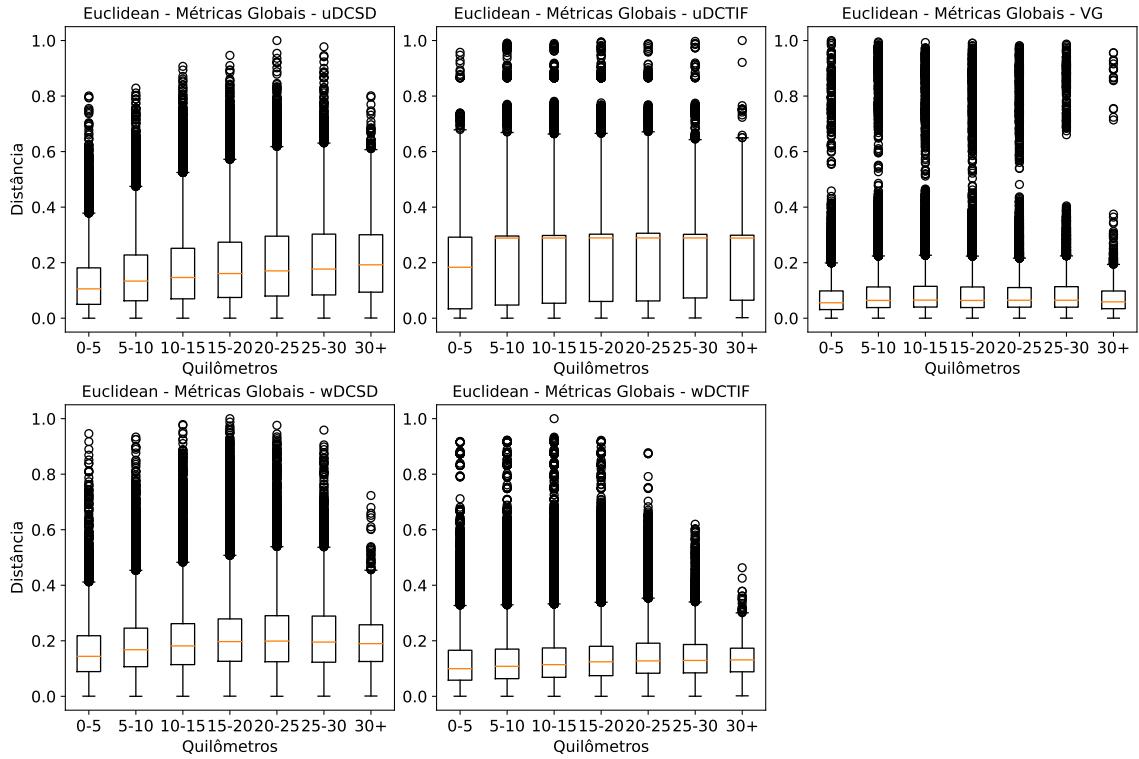


Figura 7 – Relação entre a distância geográfica e a distância Euclidiana para comparação usando vetor de métricas globais.

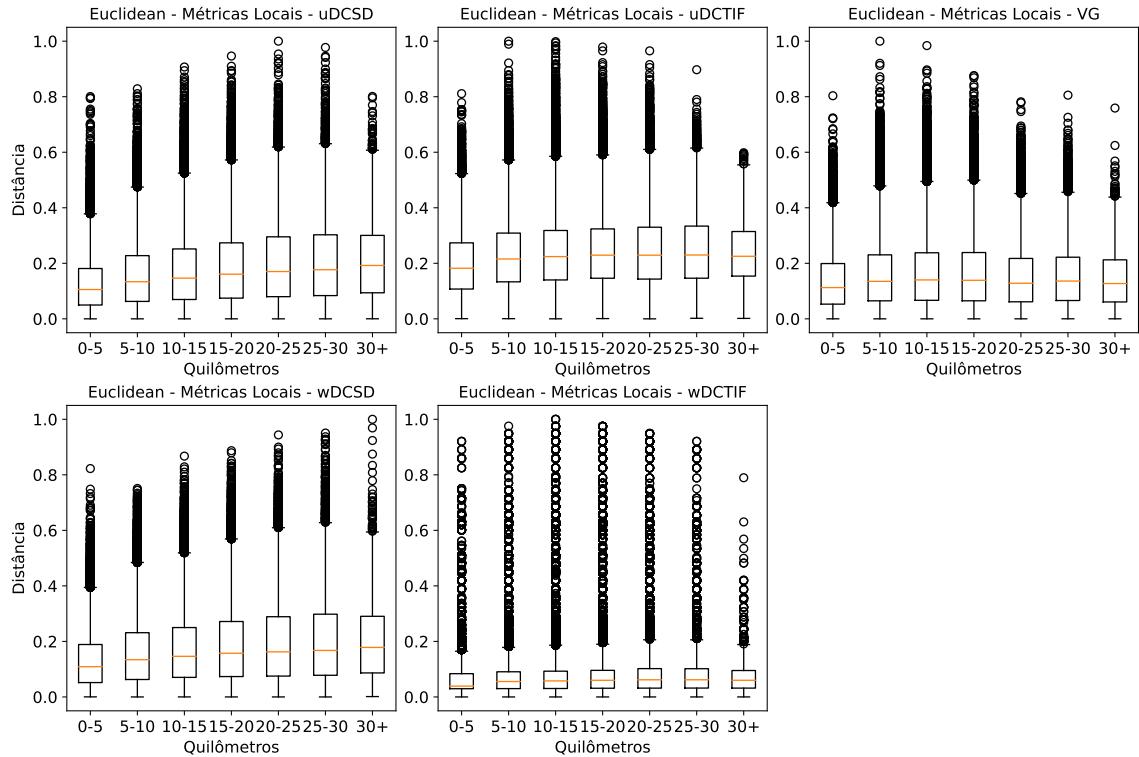


Figura 8 – Relação entre a distância geográfica e a distância Euclidiana para comparação usando vetor de métricas locais.

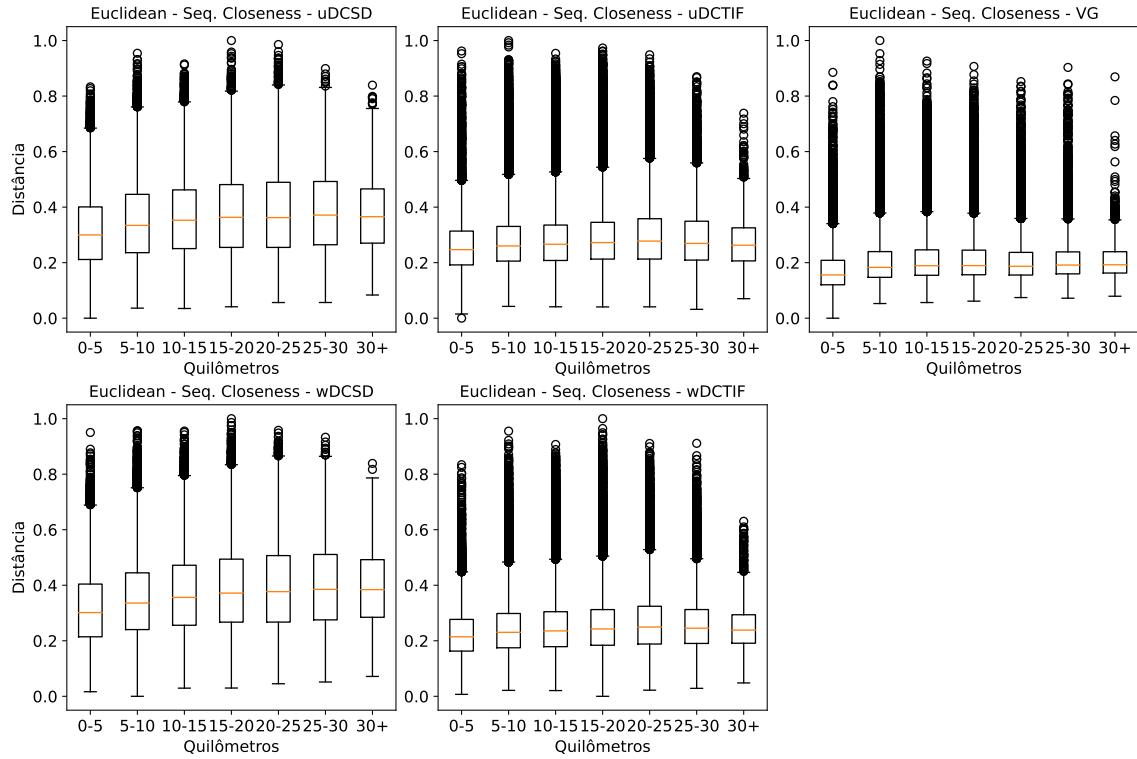


Figura 9 – Relação entre a distância geográfica e a distância Euclidiana para comparação usando sequência de *closeness*.

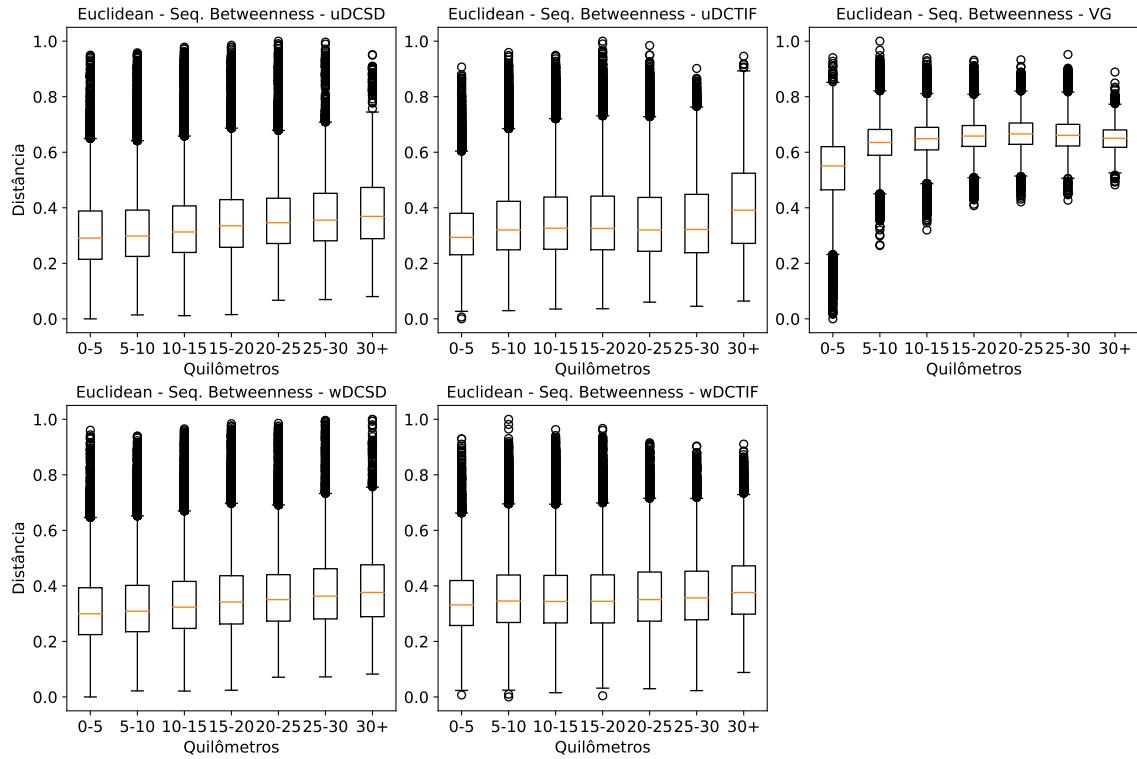


Figura 10 – Relação entre a distância geográfica e a distância Euclidiana para comparação usando sequência de *betweenness*.

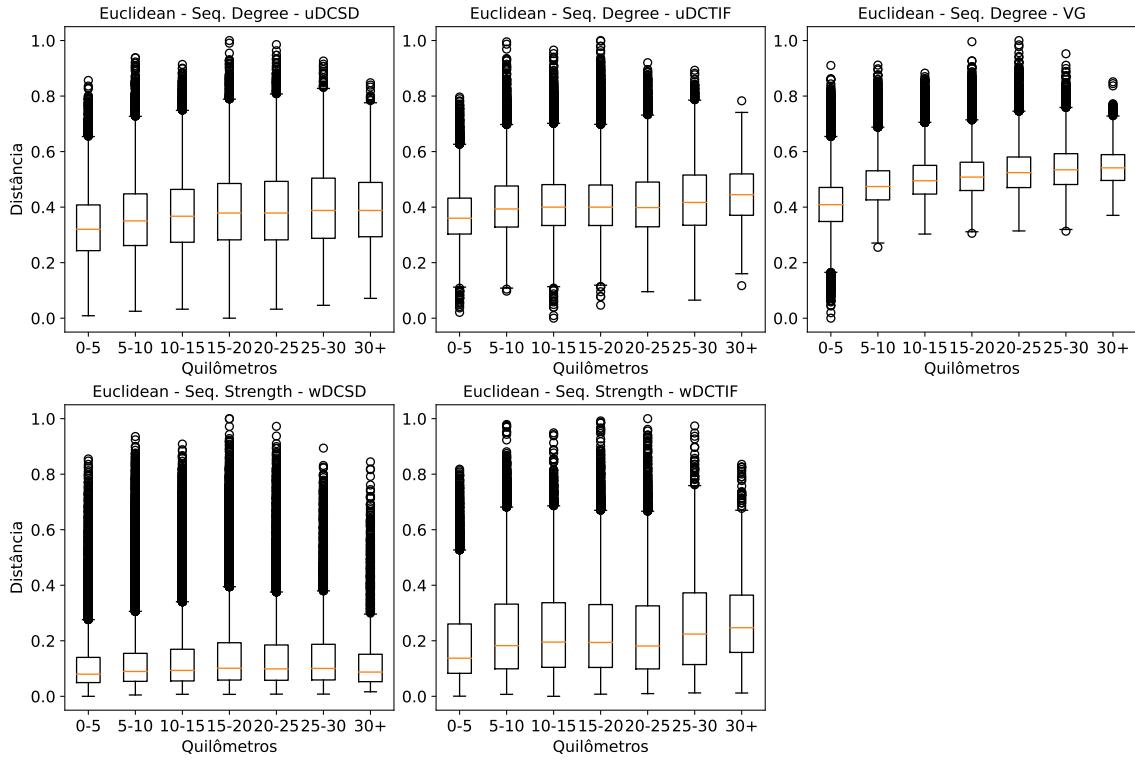


Figura 11 – Relação entre a distância geográfica e a distância Euclidiana para comparação usando sequência de *degree* e *strength*.

5.2 Distância de Manhattan

As representações em boxplot a seguir, tendo início na Figura 12 e terminando na Figura 17, demonstram a distribuição dos dados usando a distância de *Manhattan*.

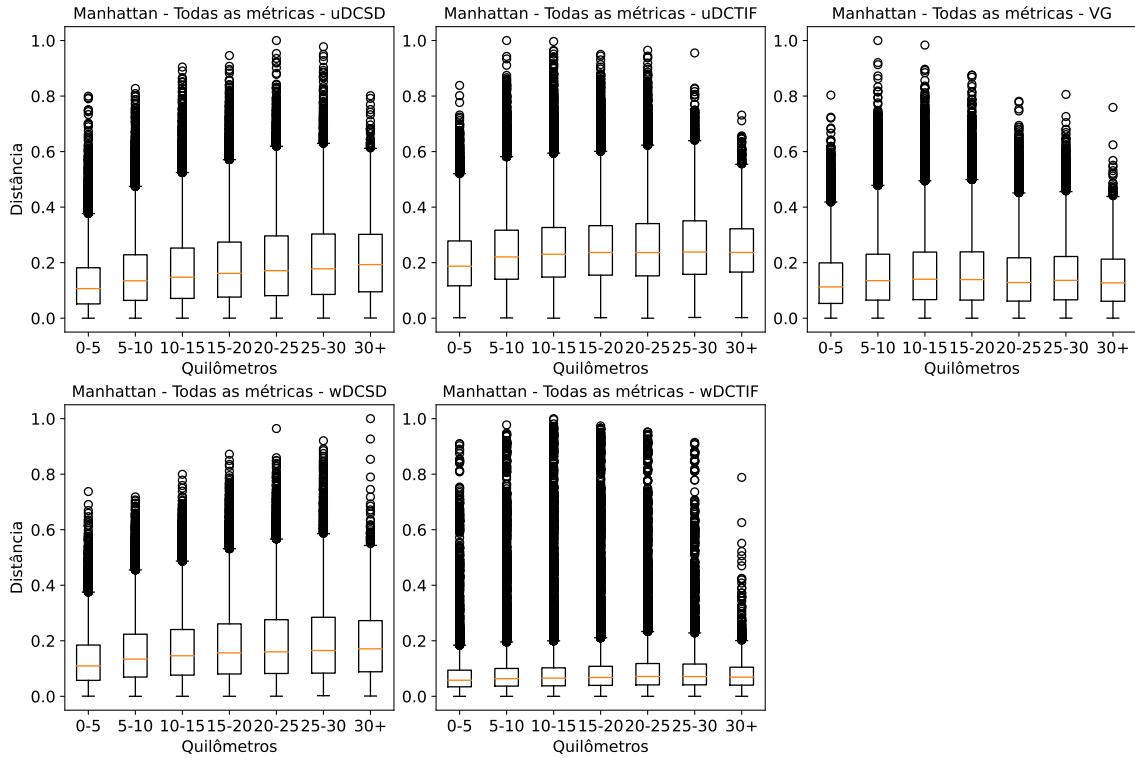


Figura 12 – Relação entre a distância geográfica e a distância de Manhattan para comparação usando vetor de várias métricas.

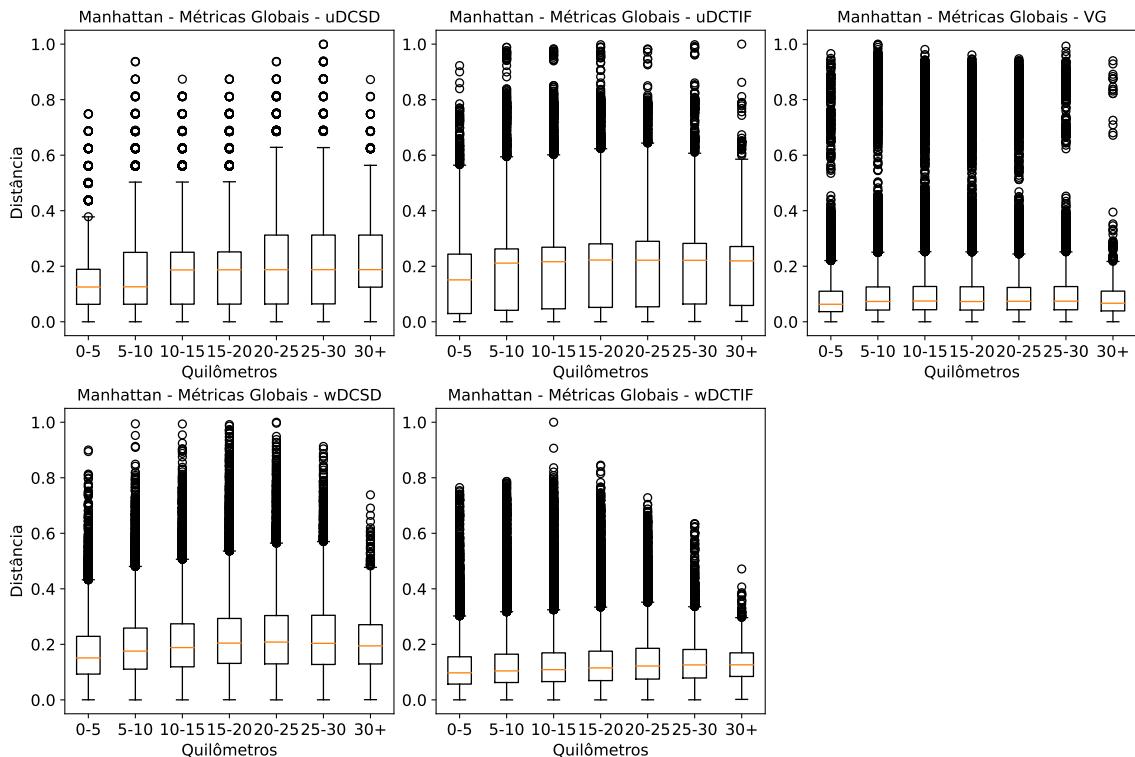


Figura 13 – Relação entre a distância geográfica e a distância de Manhattan para comparação usando vetor de métricas globais.

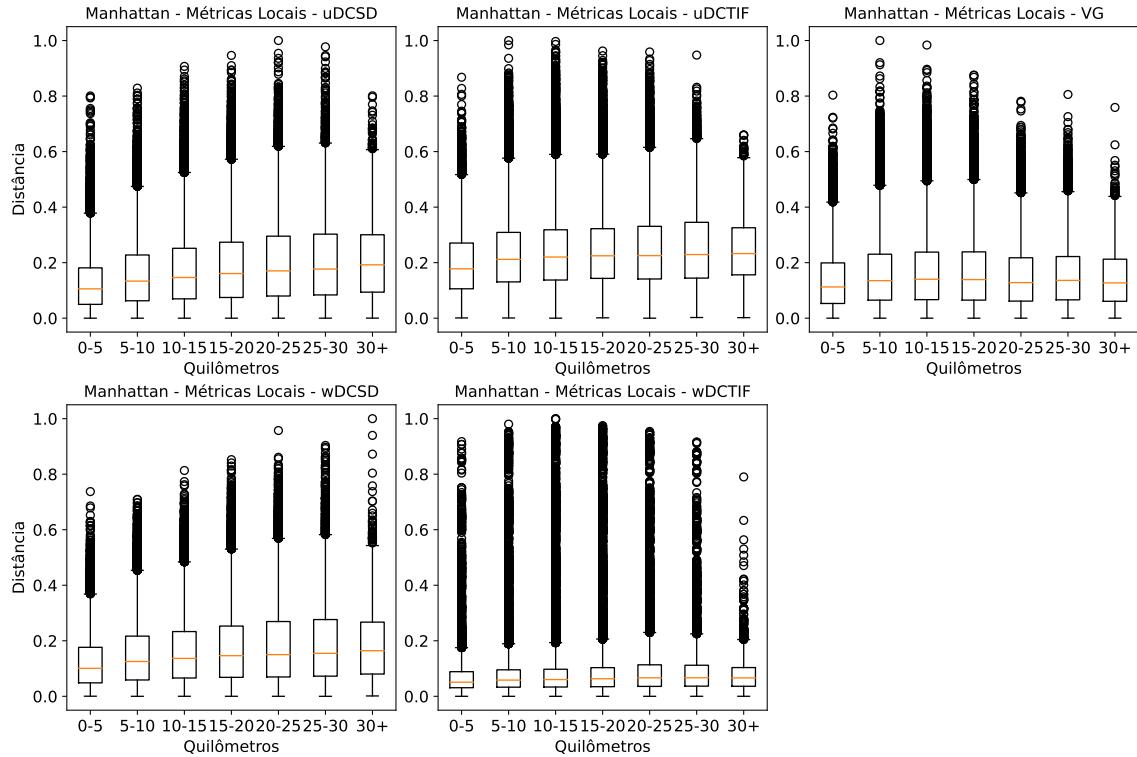


Figura 14 – Relação entre a distância geográfica e a distância de Manhattan para comparação usando vetor de métricas locais.

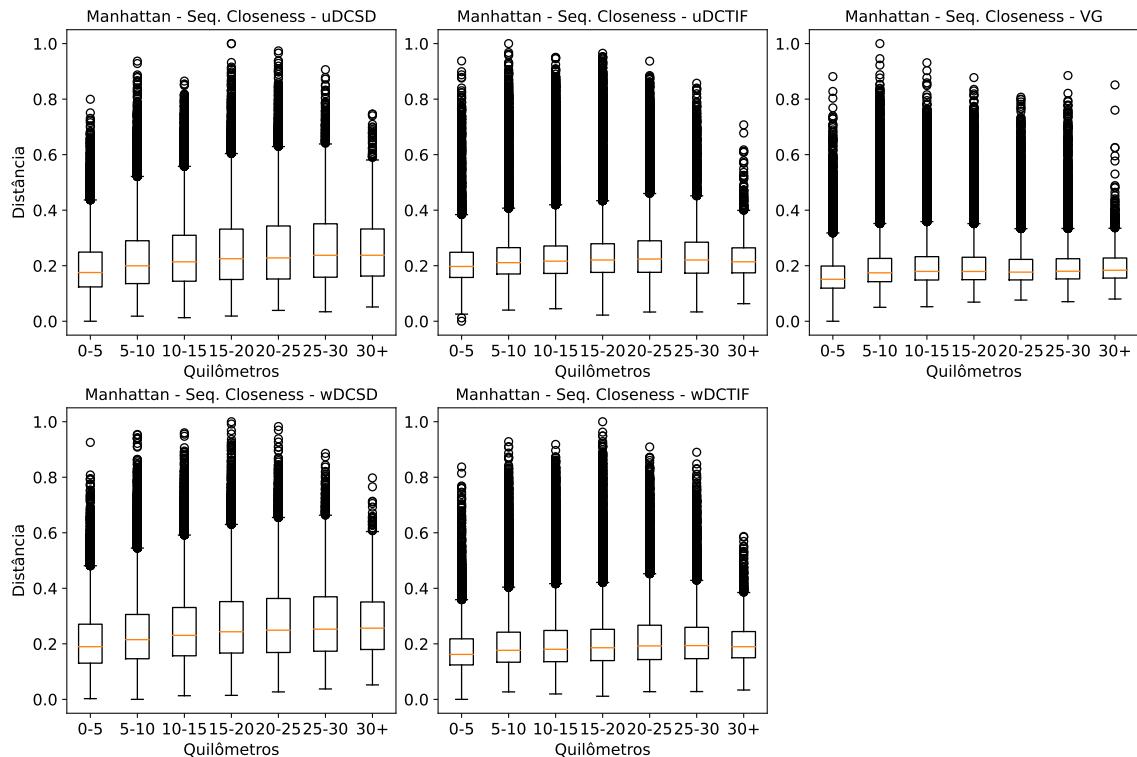


Figura 15 – Relação entre a distância geográfica e a distância de Manhattan para comparação usando sequência de *closeness*.

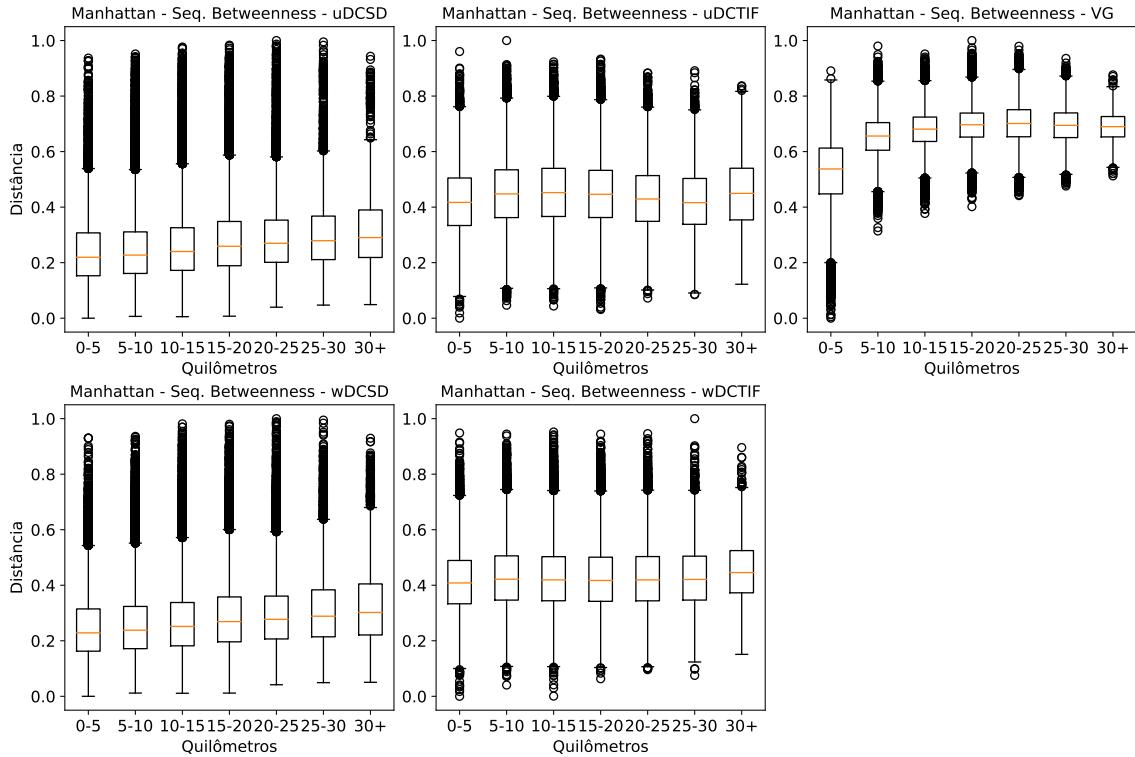


Figura 16 – Relação entre a distância geográfica e a distância de Manhattan para comparação usando sequência de *betweenness*.

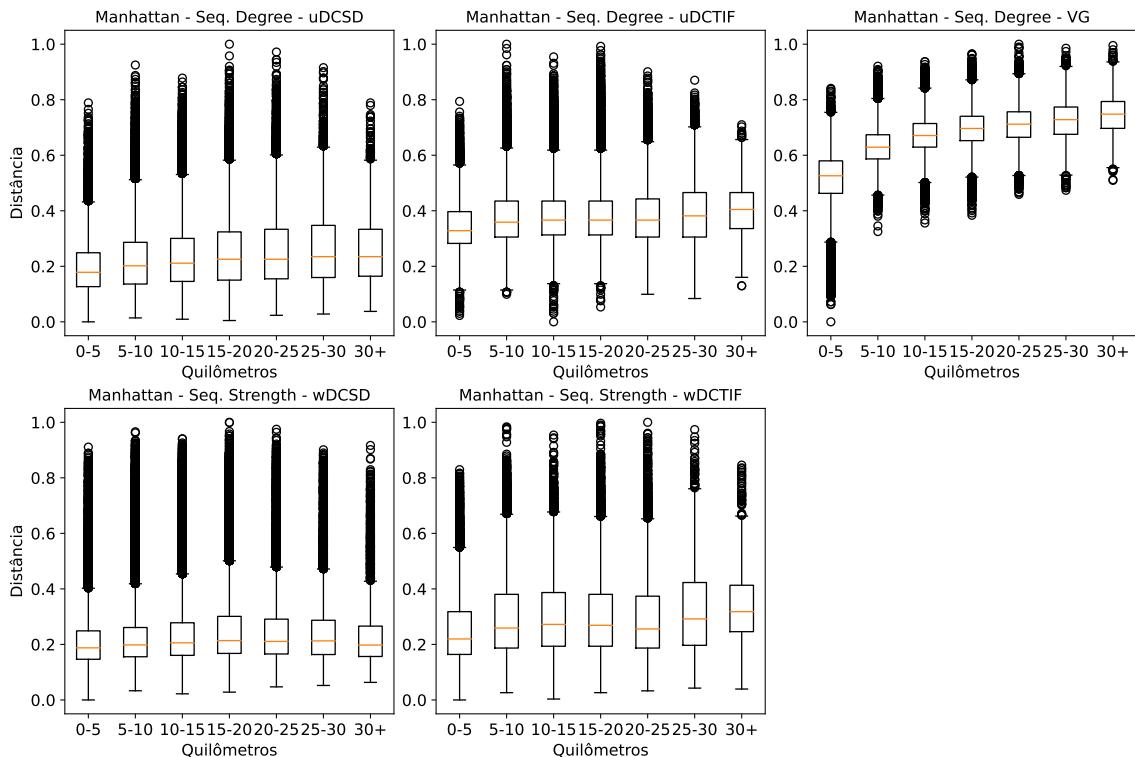


Figura 17 – Relação entre a distância geográfica e a distância de Manhattan para comparação usando sequência de *degree* e *strength*.

5.3 Distância de Canberra

As representações em boxplot a seguir, tendo início na Figura 18 e terminando na Figura 23, demonstram a distribuição dos dados usando a distância de *Canberra*.

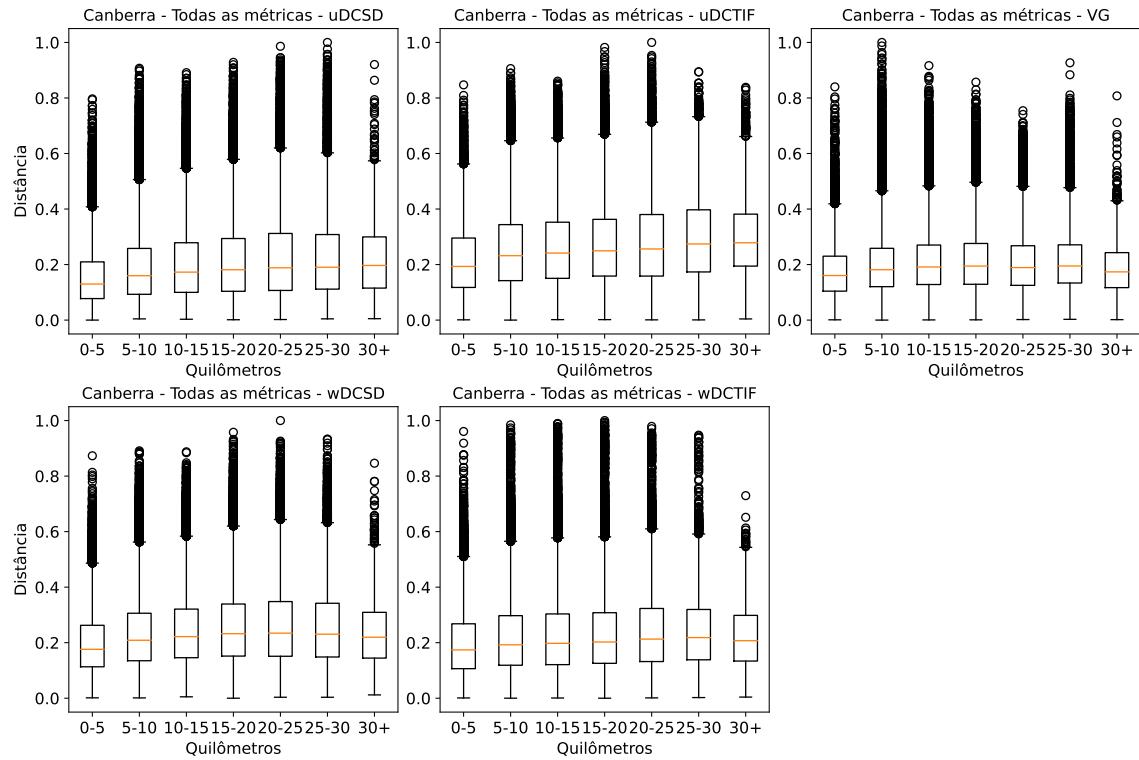


Figura 18 – Relação entre a distância geográfica e a distância de Canberra para comparação usando vetor de várias métricas.

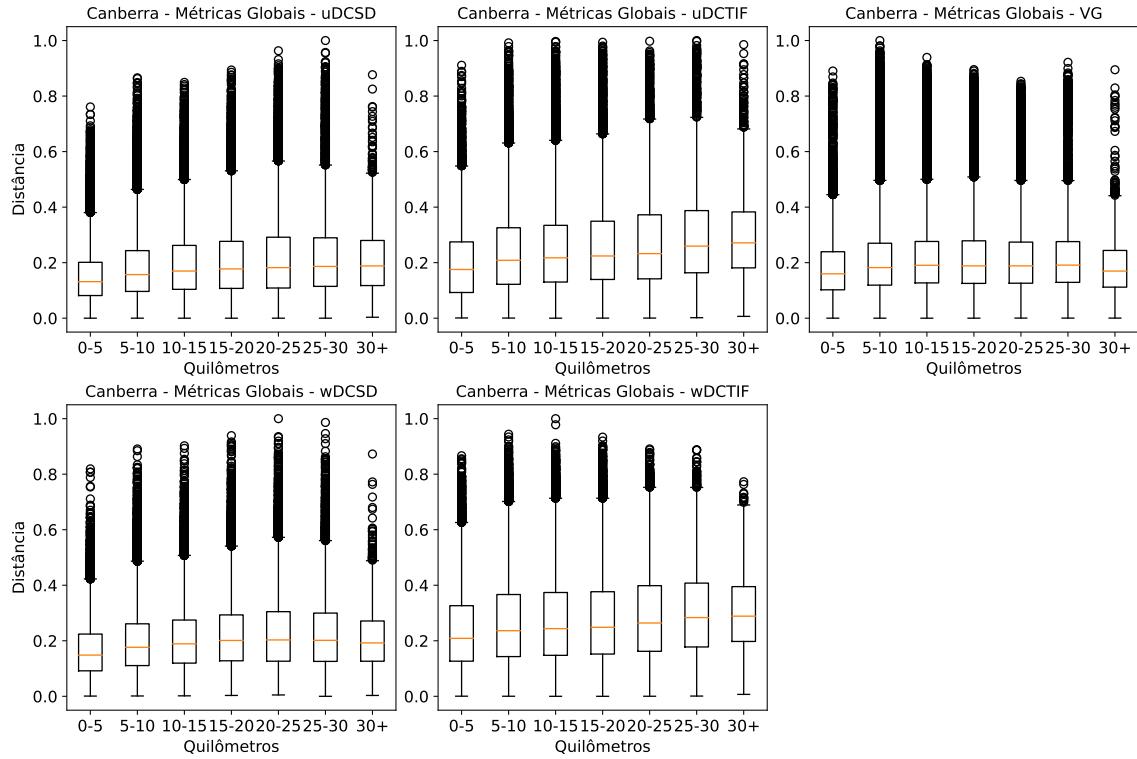


Figura 19 – Relação entre a distância geográfica e a distância de Canberra para comparação usando vetor de métricas globais.

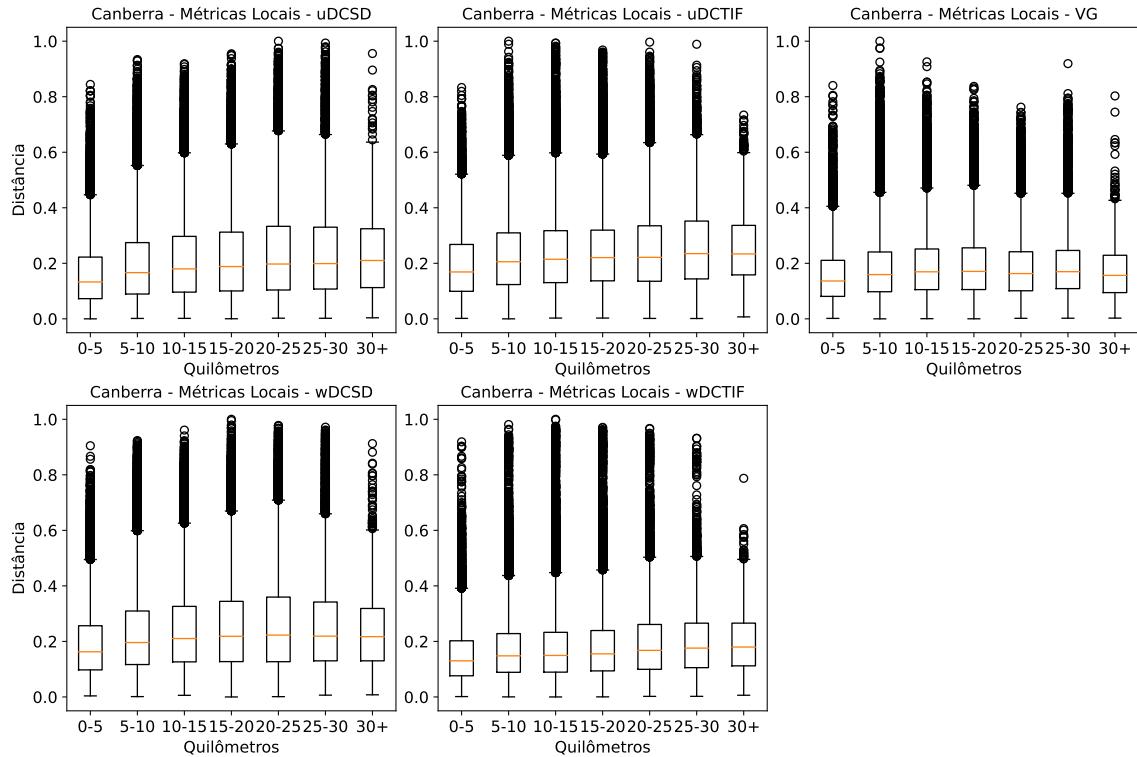


Figura 20 – Relação entre a distância geográfica e a distância de Canberra para comparação usando vetor de métricas locais.

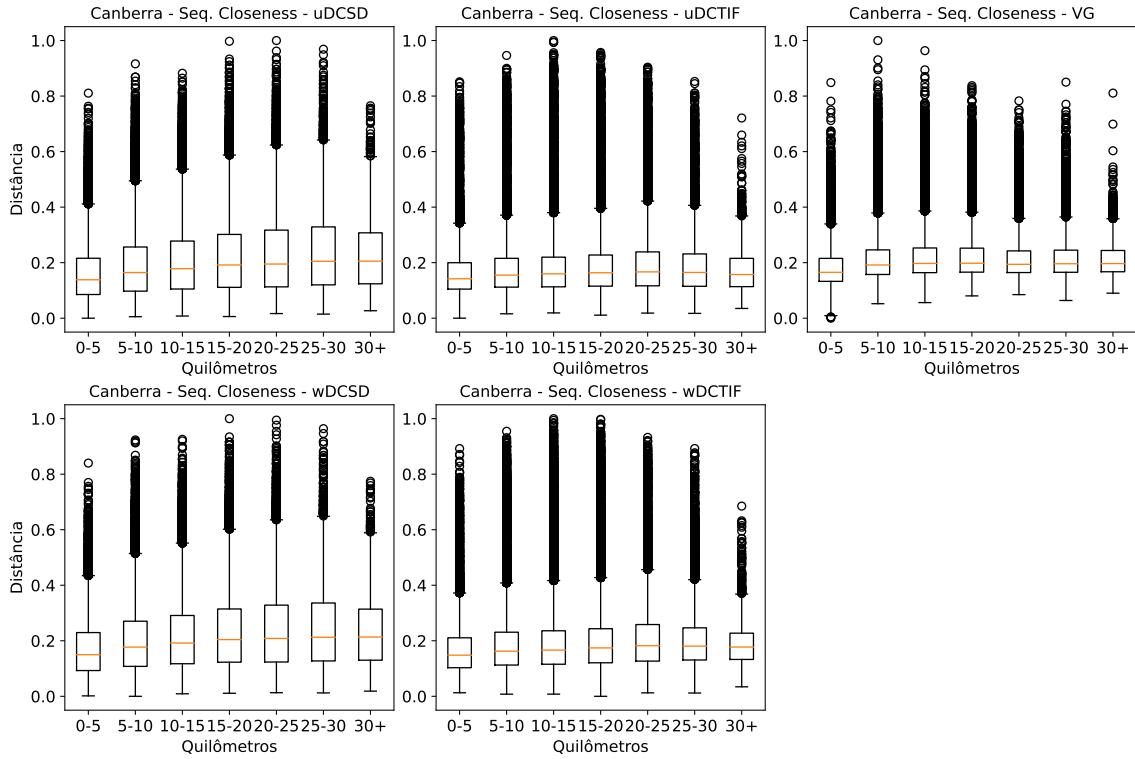


Figura 21 – Relação entre a distância geográfica e a distância de Canberra para comparação usando sequência de *closeness*.

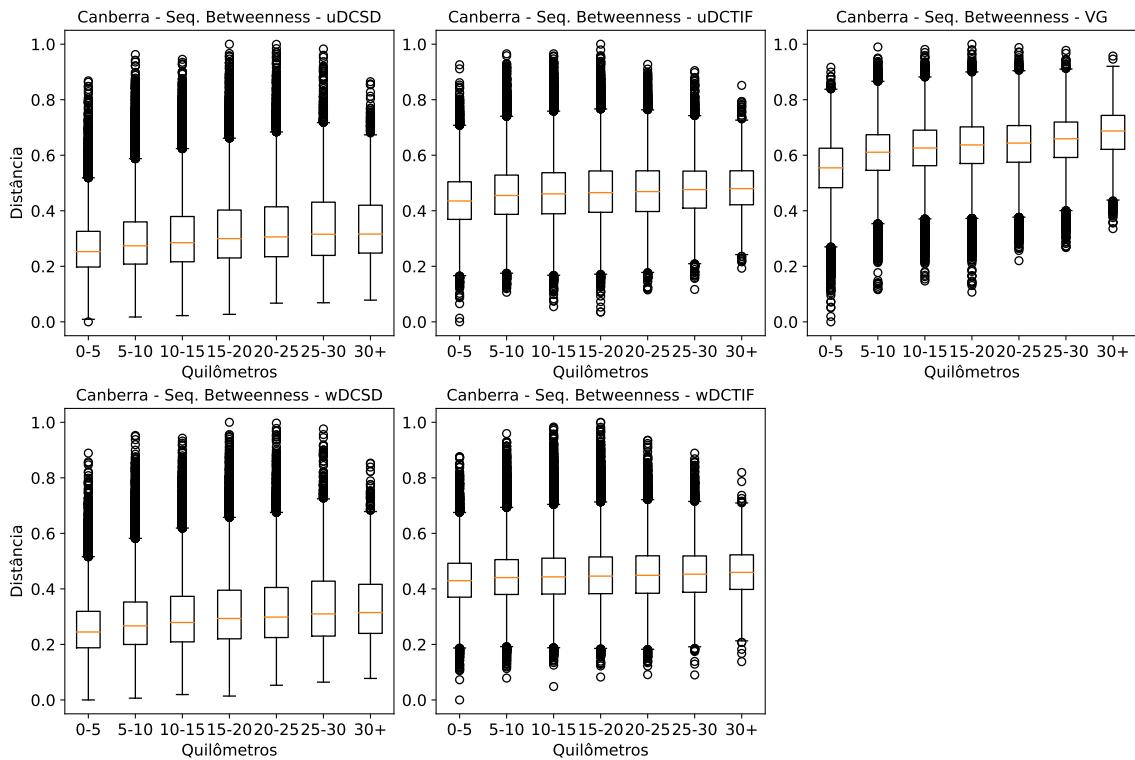


Figura 22 – Relação entre a distância geográfica e a distância de Canberra para comparação usando sequência de *betweenness*.

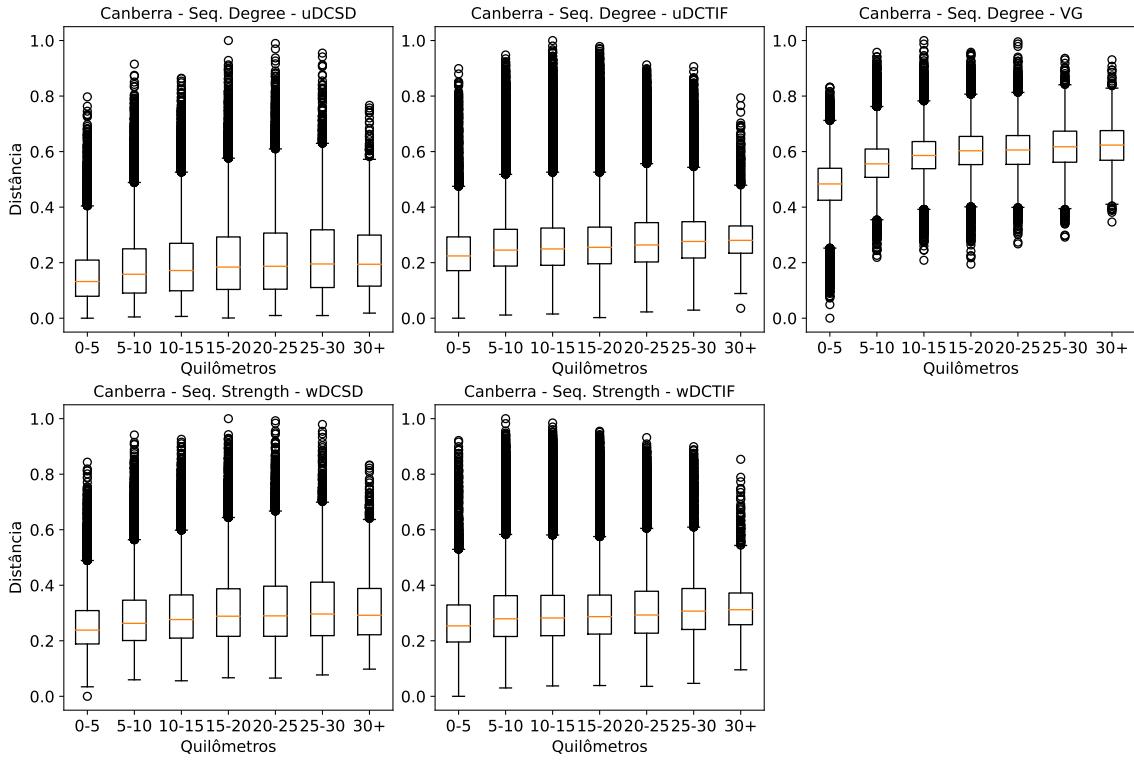


Figura 23 – Relação entre a distância geográfica e a distância de Canberra para comparação usando sequência de *degree* e *strength*.

5.4 Similaridade por Cosseno

As representações em boxplot a seguir, tendo início na Figura 24 e terminando na Figura 29, demonstram a distribuição dos dados usando a similaridade por cosseno.

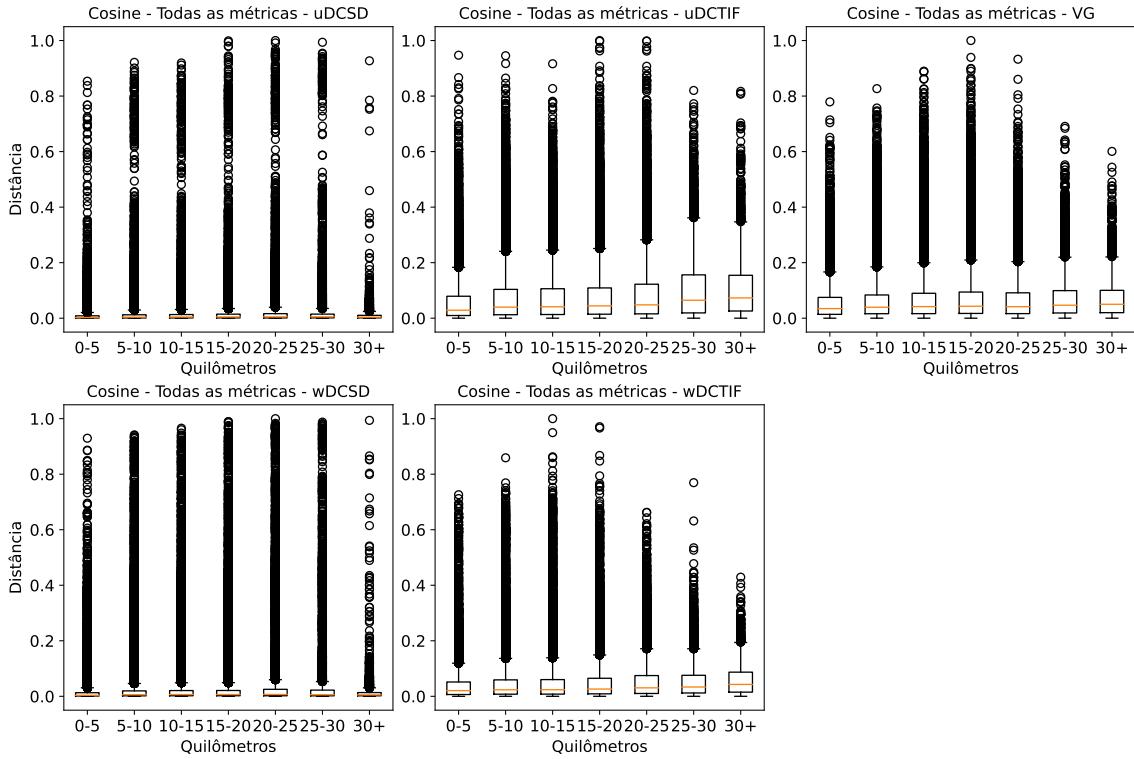


Figura 24 – Relação entre a distância geográfica e a similaridade por cosseno para comparação usando vetor de várias métricas.

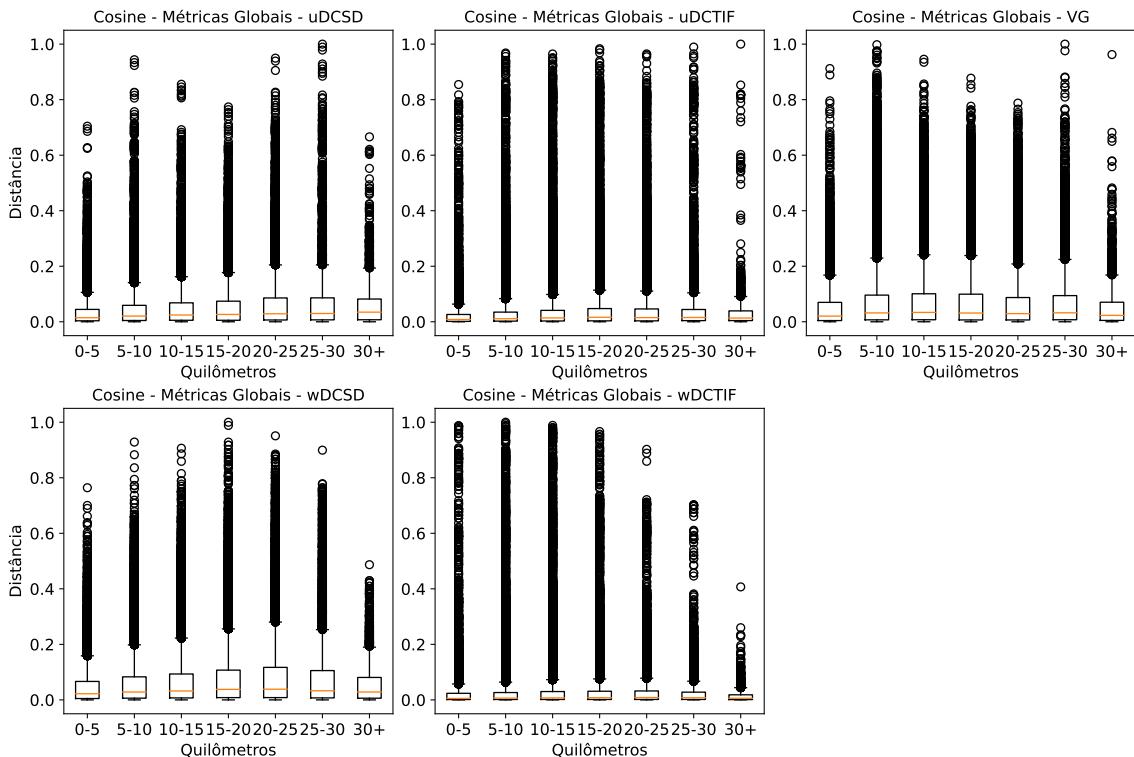


Figura 25 – Relação entre a distância geográfica e a similaridade por cosseno para comparação usando vetor de métricas globais.

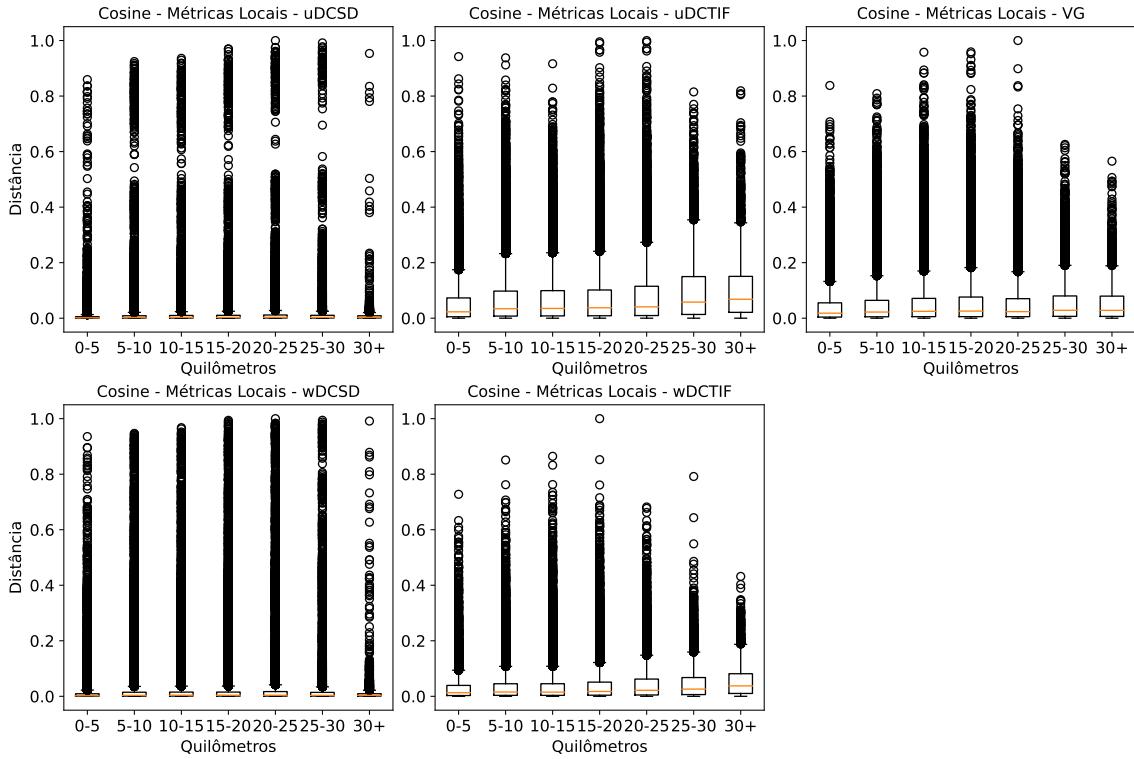


Figura 26 – Relação entre a distância geográfica e a similaridade por cosseno para comparação usando vetor de métricas locais.

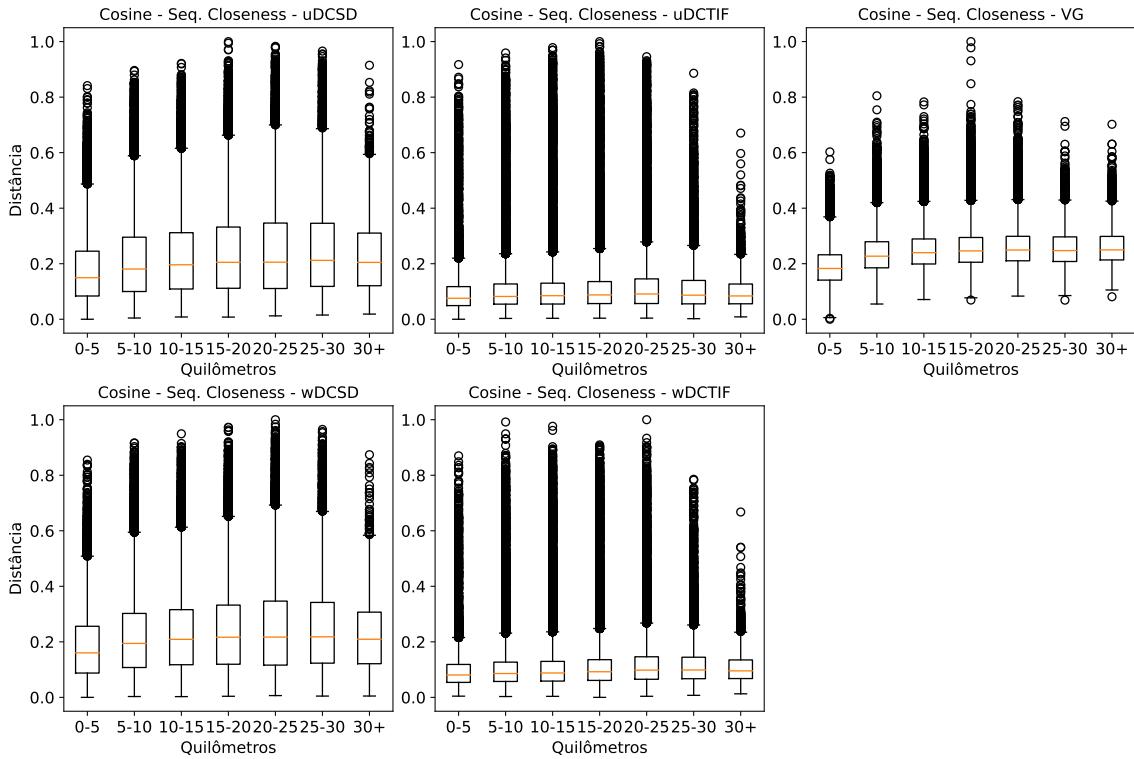


Figura 27 – Relação entre a distância geográfica e a similaridade por cosseno para comparação usando sequência de *closeness*.

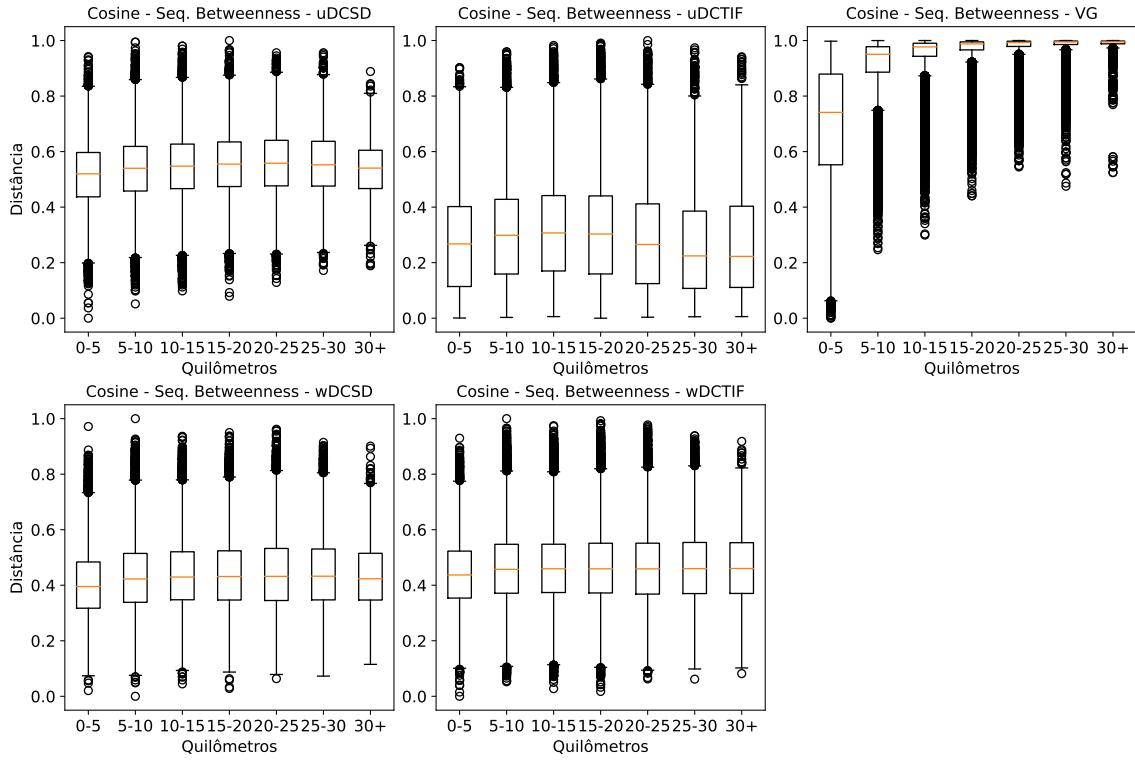


Figura 28 – Relação entre a distância geográfica e a similaridade por cosseno para comparação usando sequência de *betweenness*.

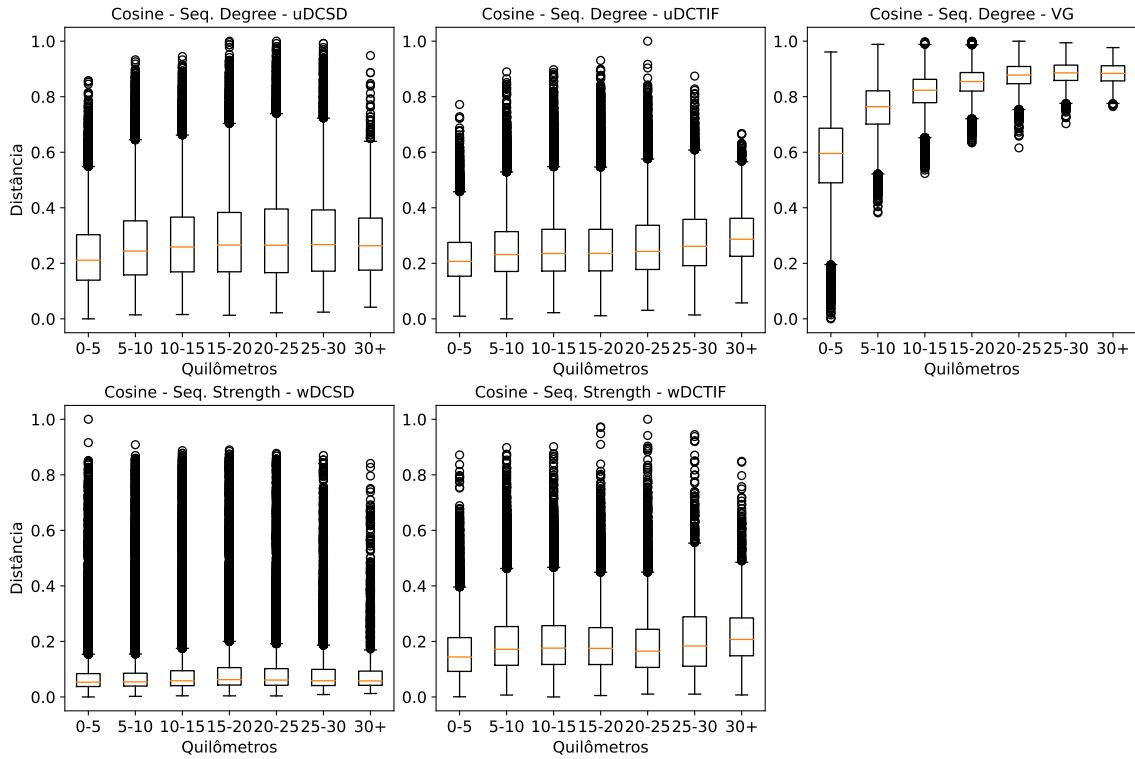


Figura 29 – Relação entre a distância geográfica e a similaridade por cosseno para comparação usando sequência de *degree* e *strength*.

5.5 Distância de Jensen-Shannon

As representações em boxplot a seguir, tendo início na Figura 30 e terminando na Figura 35, demonstram a distribuição dos dados usando a distância de *Jensen-Shannon*.

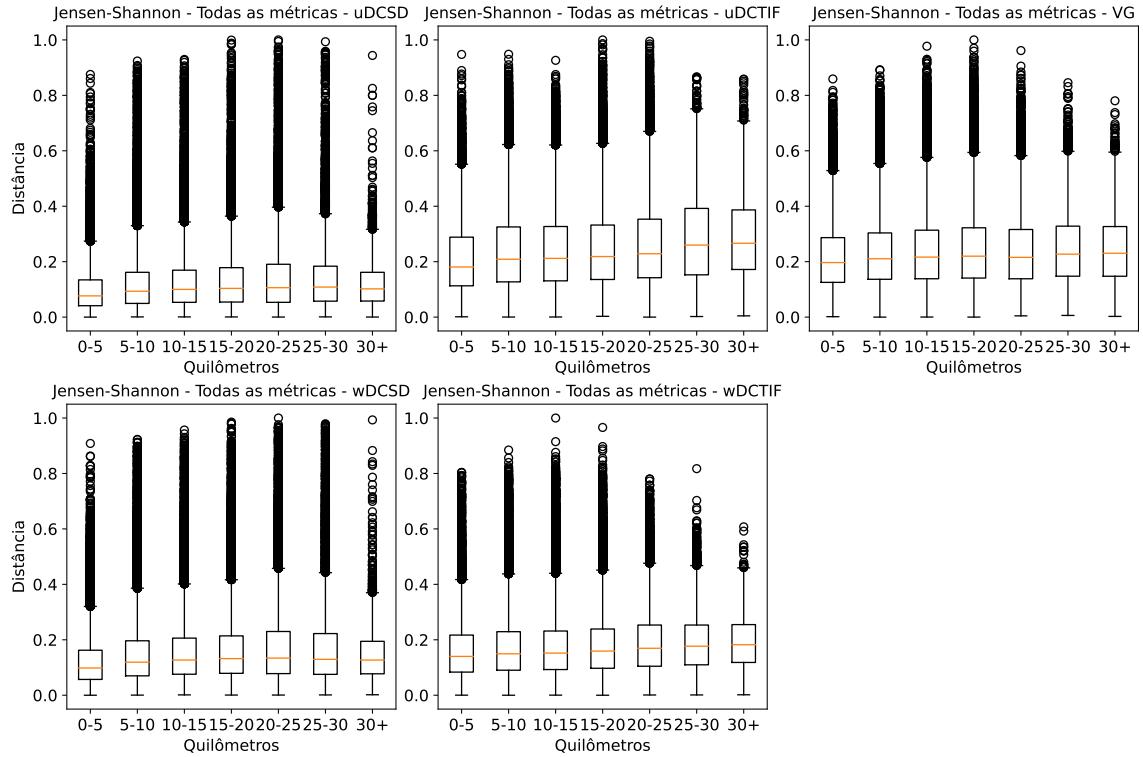


Figura 30 – Relação entre a distância geográfica e a distância de Jensen-Shannon para comparação usando vetor de várias métricas.

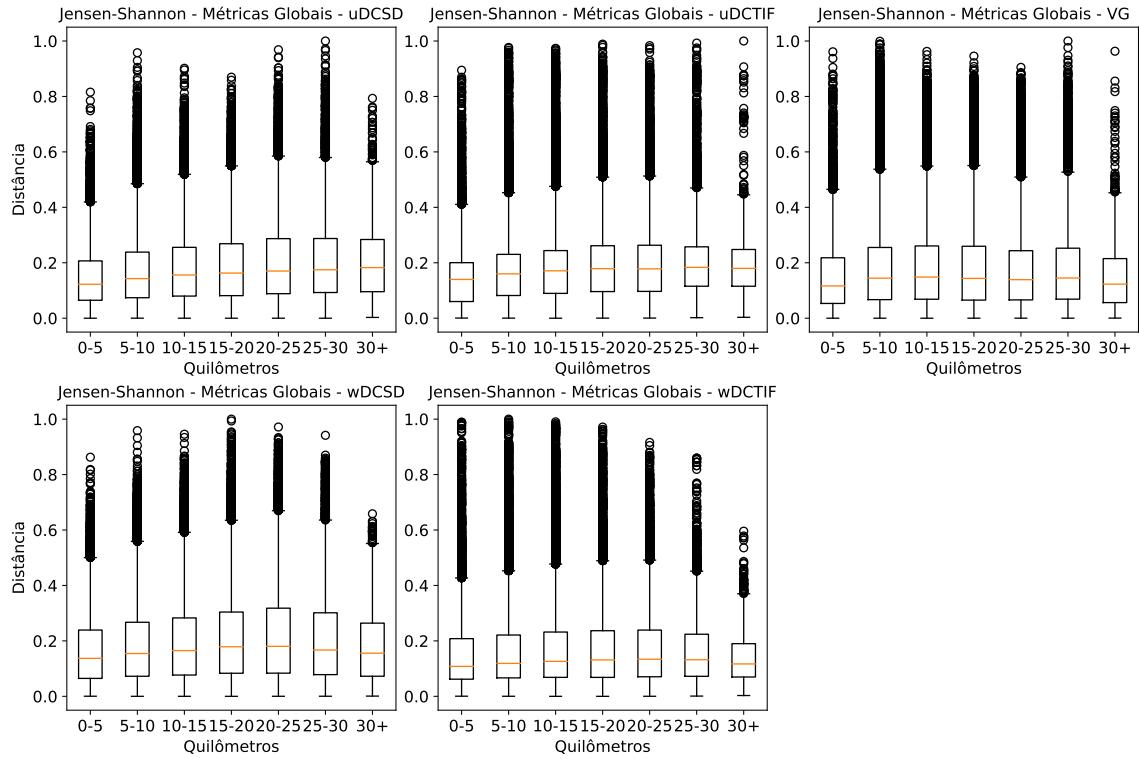


Figura 31 – Relação entre a distância geográfica e a distância de Jensen-Shannon para comparação usando vetor de métricas globais.

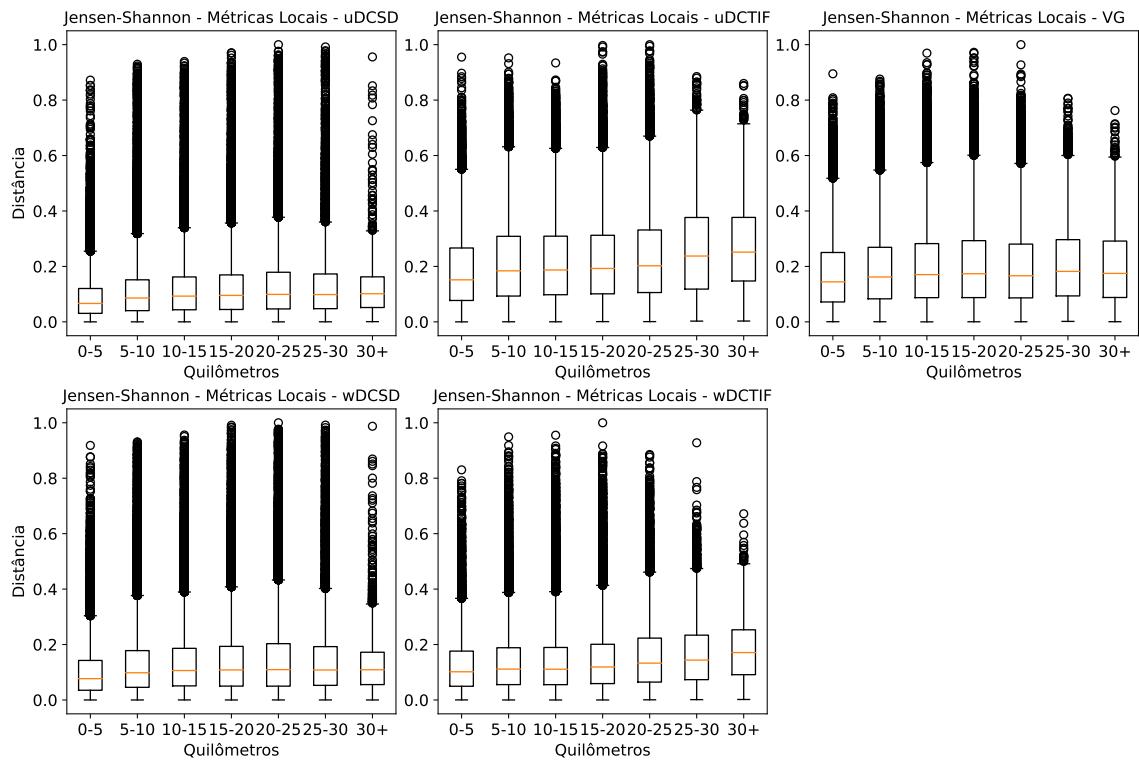


Figura 32 – Relação entre a distância geográfica e a distância de Jensen-Shannon para comparação usando vetor de métricas locais.

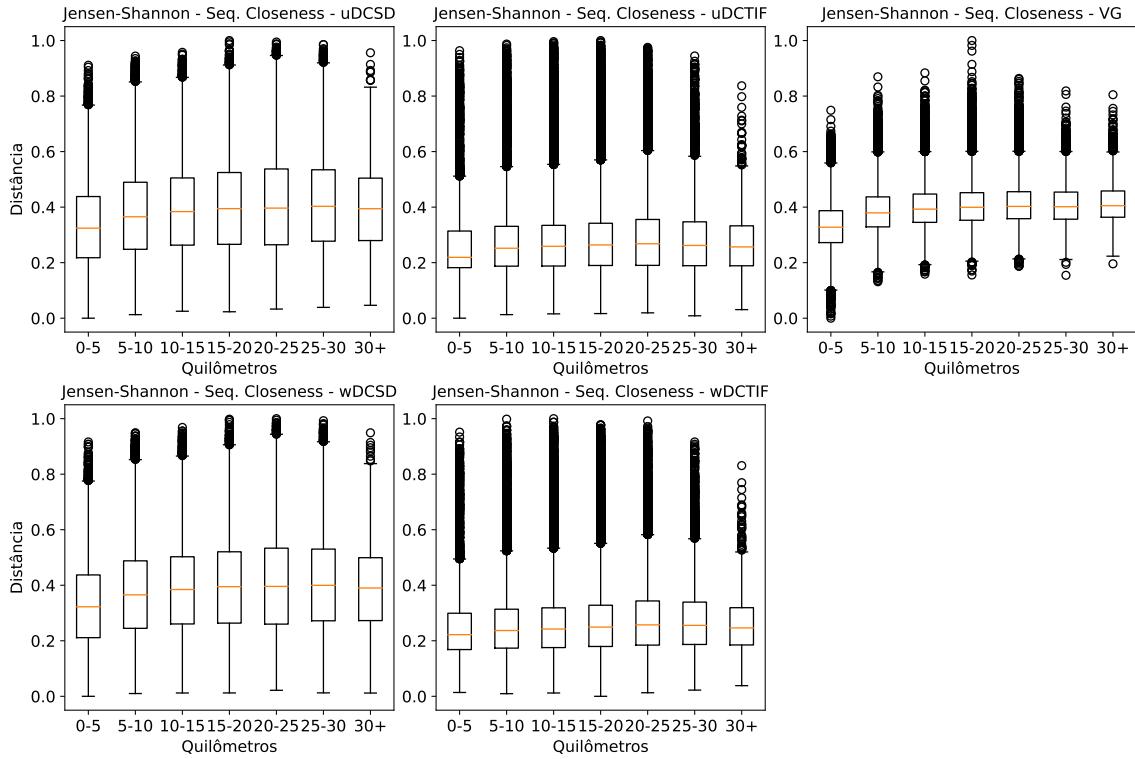


Figura 33 – Relação entre a distância geográfica e a distância de Jensen-Shannon para comparação usando sequência de *closeness*.

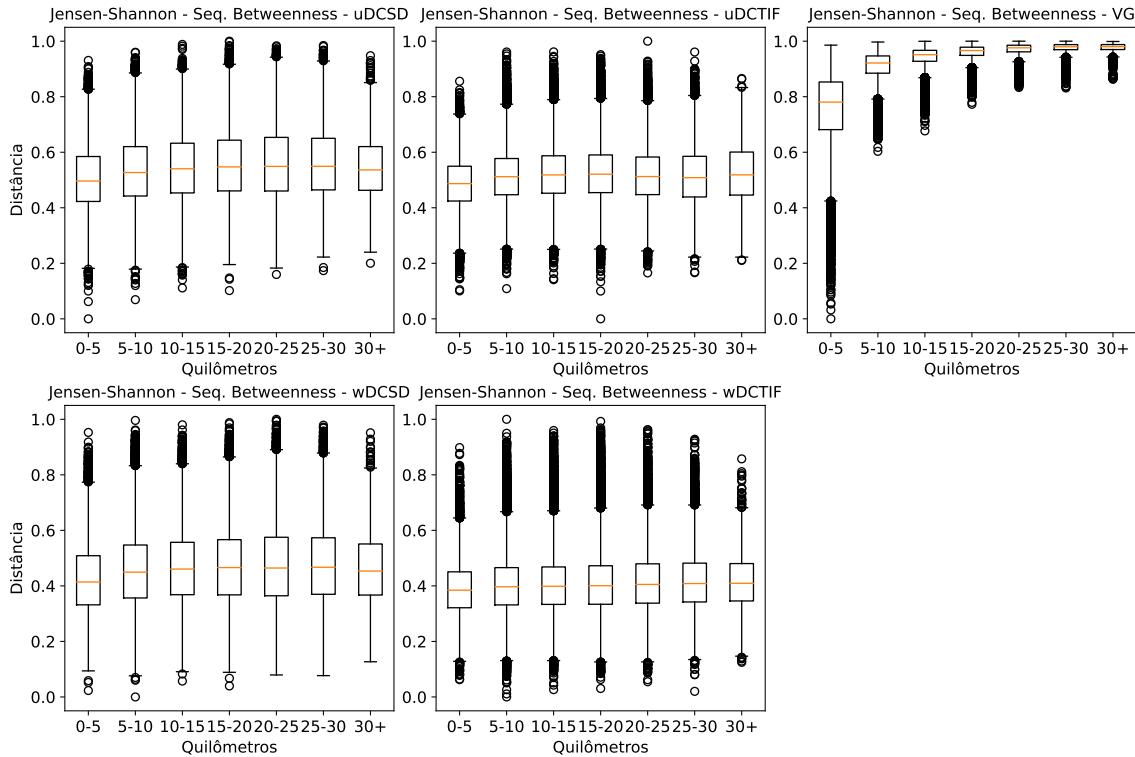


Figura 34 – Relação entre a distância geográfica e a distância de Jensen-Shannon para comparação usando sequência de *betweenness*.

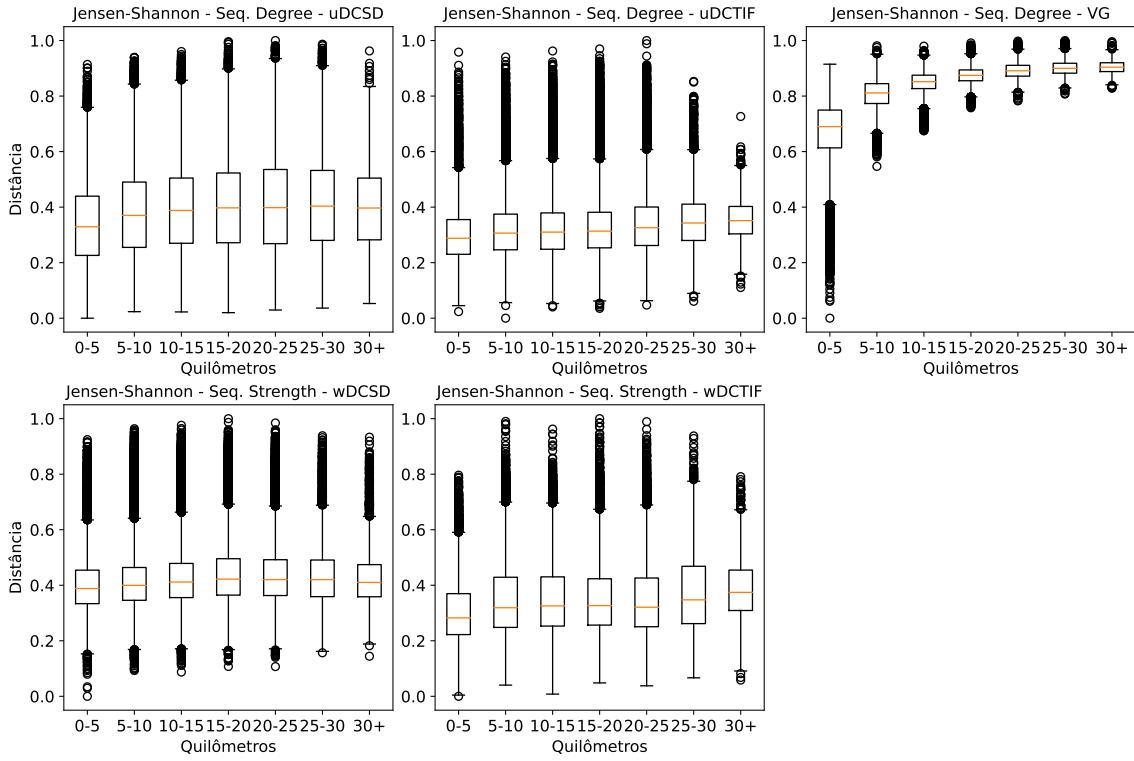


Figura 35 – Relação entre a distância geográfica e a distância de Jensen-Shannon para comparação usando sequência de *degree* e *strength*.

5.6 Formação da rede geográfica

A distribuição das comparações evidenciou alguns padrões compatíveis com o esperado, mas apenas quando foram comparadas as sequências de características de redes geradas pelo VG — as características levadas em consideração eram responsáveis pela descrição de cada nó da rede. O mesmo comportamento não foi observado para o caso dos algoritmos DCSD e DCTIF, mesmo com a modificação para geração de redes ponderadas. Nas Figuras 29 e 35, considerando as distribuições para o VG, pode ser vista a forma com que a amplitude de variação dos dados diminui, e a mediana se eleva, de acordo com o aumento da distância entre as redes no espaço. Assim, pode-se concluir, para os casos mencionados, que quanto menor o resultado da comparação entre um par de redes usando métricas de distância — todas as métricas demonstraram o mesmo comportamento, mas uns foram mais distintos —, menor é a distância entre essas duas redes dentro do espaço geográfico analisado. Essa conclusão permite a definição de um limiar crítico para a seleção de pares que integrarão a rede geográfica.

5.6.1 Rede formada por meio da comparação usando similaridade por cosseno

A Figura 36 ilustra a formação da rede geográfica quando um limiar de 0,3497 é traçado na distribuição do VG mostrado na Figura 29. Foram conectados os pares de rede cujo valor de comparação foi menor que o limiar pré-definido para obtenção de uma rede geográfica com 1270 arestas, mesmo número obtido no estudo apresentado por [Jorge, Costa e Santos \(2020\)](#).



Figura 36 – Rede da região considerando as conexões formadas através da comparação da medida *Degree* usando similaridade por cosseno.

A análise da rede formada levou ao resultado apresentado na Figura 37. O gráfico foi construído levando em conta a distância topológica (caminho mínimo) e espacial entre os pares possível de pontos. É possível observar que a distribuição tem um caráter altamente linear e se assemelha aos resultados obtidos por [Jorge, Costa e Santos \(2020\)](#), conseguindo capturar o mesmo comportamento mesmo fazendo o uso de uma metodologia completamente distinta para a construção da rede geográfica.

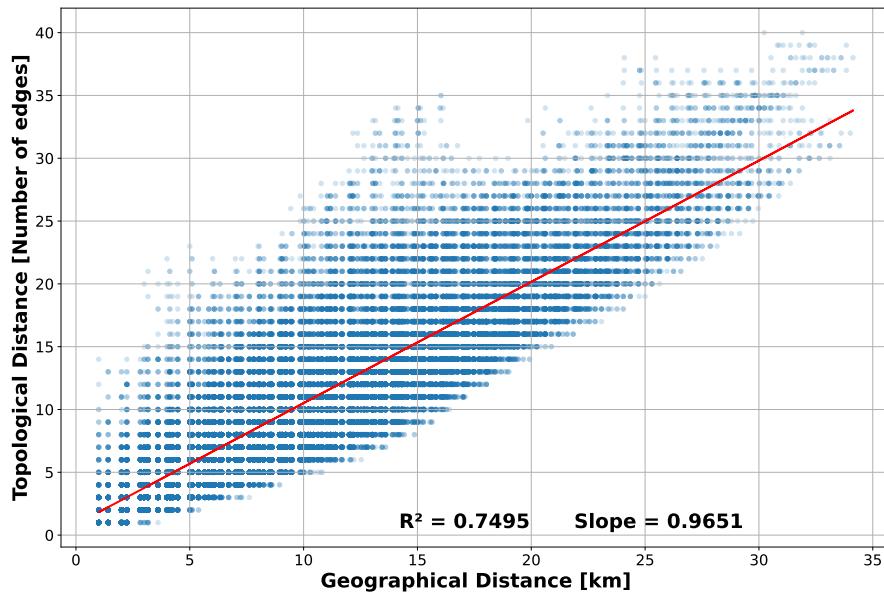


Figura 37 – Relação entre as distâncias topológicas e geográficas.

5.6.2 Rede formada por meio da comparação usando distância de Jensen-Shannon

A rede mostrada na Figura 38 foi obtida a partir da definição de um limiar de 0,4931, aplicado na distribuição para o VG ilustrada na Figura 35. Pode ser observado que a rede e a distribuição das distâncias, Figura 39, apresentam comportamento similar ao observado para o caso anterior, sendo possível notar, inclusive, a manutenção das áreas com maior densidade de links na rede.

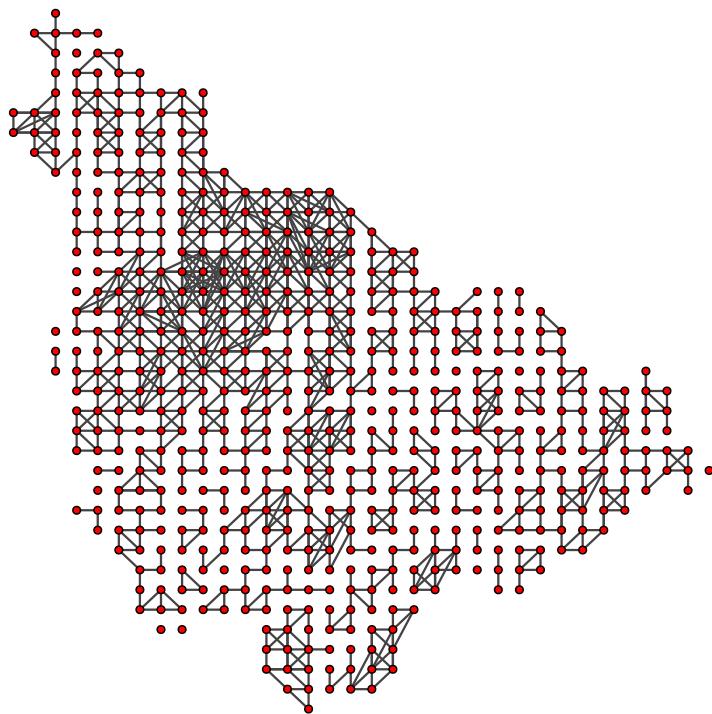


Figura 38 – Rede da região considerando as conexões formadas através da comparação da medida *Degree* usando a distância de Jensen-Shannon.

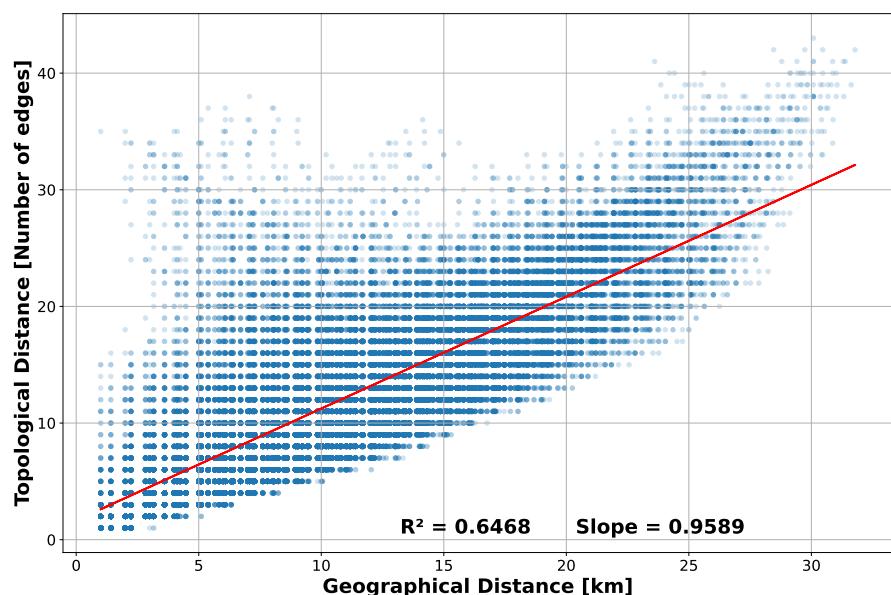


Figura 39 – Relação entre as distâncias topológicas e geográficas.

6 CONCLUSÃO

Ao longo do desenvolvimento da metodologia, ficou claro que o problema proposto, o uso das redes complexas para análise de séries temporais, ainda possui um vasto potencial de exploração. Partindo da base fornecida por trabalhos como o de [Freitas, Lacerda e Macau \(2019\)](#), buscou-se a associação de diversas estratégias para o desenvolvimento de um trabalho com dados reais.

O trabalho desenvolvido por [Jorge, Costa e Santos \(2020\)](#) forneceu uma base de comparação para verificar se os mesmos comportamentos observados usando uma metodologia estatística bem estabelecida também seriam capturados com o uso das estratégias adotadas nesse projeto.

Com os resultados obtidos, pôde-se atestar a capacidade da metodologia proposta em capturar as mesmas dinâmicas vistas no trabalho base, além de ter sido possível perceber como os algoritmos DCSD, DCTIF e VG atuaram na conversão das séries, reforçando a necessidade de critérios que definam o mais apropriado para cada problema.

7 PRODUÇÃO CIENTÍFICA

Submissão de resumo estendido intitulado “Análise de redes de precipitação baseadas em séries temporais de um radar meteorológico” para o Congresso Nacional de Matemática Aplicada e Computação (CNMAC), edição 2021.

Artigo intitulado “*Global-threshold and backbone high-resolution weather radar networks are significantly complementary in a watershed*” em preparação para submissão na revista *Transactions in GIS*.

REFERÊNCIAS

- AMARAL, L. A.; OTTINO, J. M. Complex networks. *The European Physical Journal B*, 2004. Springer, v. 38, n. 2, p. 147–162, 2004. Citado na página [6](#).
- BARABÁSI, A.-L. et al. *Network science*. [S.l.]: Cambridge university press, 2016. Citado 2 vezes nas páginas [6](#) e [7](#).
- BOCCARA, N. *Modeling complex systems*. [S.l.]: Springer Science & Business Media, 2010. Citado 2 vezes nas páginas [4](#) e [6](#).
- BROCKWELL, P. J.; DAVIS, R. A. *Introduction to time series and forecasting*. [S.l.]: springer, 2016. Citado 2 vezes nas páginas [4](#) e [6](#).
- CABRAL, R. da S. Estudo da variabilidade de medidas em redes complexas. 2013. Universidade Federal de Minas Gerais, 2013. Citado na página [12](#).
- CRYER, J. D. *Time series analysis*. [S.l.]: Springer, 1986. Citado na página [6](#).
- FREITAS, V. L.; LACERDA, J. C.; MACAU, E. E. Complex networks approach for dynamical characterization of nonlinear systems. *International Journal of Bifurcation and Chaos*, 2019. World Scientific, v. 29, n. 13, p. 1950188, 2019. Citado 5 vezes nas páginas [4](#), [7](#), [8](#), [14](#) e [38](#).
- GEORGE, K. K. et al. Cosine distance features for improved speaker verification. *Electronics Letters*, 2015. IET, v. 51, n. 12, p. 939–941, 2015. Citado na página [12](#).
- JORGE, A. A.; COSTA, I. C.; SANTOS, L. B. Geographical complex networks applied to describe meteorological data. 2020. GEOINFO, v. 21, p. 258–263, 2020. Citado 5 vezes nas páginas [14](#), [17](#), [34](#), [35](#) e [38](#).
- JURMAN, G. et al. Canberra distance on ranked lists. In: CITESEER. *Proceedings of advances in ranking NIPS 09 workshop*. [S.l.], 2009. p. 22–27. Citado na página [12](#).
- LACASA, L. et al. From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences*, 2008. National Acad Sciences, v. 105, n. 13, p. 4972–4975, 2008. Citado 2 vezes nas páginas [9](#) e [16](#).
- MENCZER, F.; FORTUNATO, S.; DAVIS, C. A. *A first course in network science*. [S.l.]: Cambridge University Press, 2020. Citado na página [7](#).
- THOMPSON, K.; DRAY, T. Taxicab angles and trigonometry. *Pi Mu Epsilon Journal*, 2000. JSTOR, p. 87–96, 2000. Citado na página [11](#).
- YU, X.; JIA, Z.; JIAN, X. Logistic mapping-based complex network modeling. *Applied Mathematics*, 2013. Scientific Research Publishing, v. 4, n. 11, p. 1558, 2013. Citado na página [4](#).