UNIVERSIDADE FEDERAL DO CEARÁ

CENTRO DE CIÊNCIAS

DEPARTAMENTO DE FÍSICA

CURSO DE GRADUAÇÃO EM FÍSICA

DIEGO SILVA DE FRANÇA

UM ESTUDO DE HIPERGRAFOS EM REDES DE COLABORAÇÃO DE COAUTORES

FORTALEZA

2022

DIEGO SILVA DE FRANÇA

UM ESTUDO DE HIPERGRAFOS EM REDES DE COLABORAÇÃO DE COAUTORES

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Física do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Física.

Orientador: Prof. Dr. César Ivan Nunes Sampaio Filho

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

F881e França, Diego Silva de.
Um estudo de hipergrafos em redes de colaboração de coautores / Diego Silva de França. – 2022.
36 f. : il.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Ciências, Curso de Física, Fortaleza, 2022.
Orientação: Prof. Dr. César Ivan Nunes Sampaio Filho.

1. Física Estatística. 2. Ciência de Redes. 3. Grafos e Hipergrafos. I. Título.

CDD 530

DIEGO SILVA DE FRANÇA

UM ESTUDO DE HIPERGRAFOS EM REDES DE COLABORAÇÃO DE COAUTORES

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Física do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Física.

Aprovada em: 05 de Dezembro de 2022

BANCA EXAMINADORA

Prof. Dr. César Ivan Nunes Sampaio Filho (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Diego Araújo Frota
Instituto Federal do Ceará (IFCE)

Msc. Hermes Alfredo Velazquez Urquijo
Universidade Federal do Ceará (UFC)

À minha família, por sua capacidade de acreditar em mim e investir em mim. Mãe, seu cuidado e dedicação foi que deram, em alguns momentos, a esperança para seguir.

AGRADECIMENTOS

Agradeço à minha mãe Evani, á minha tia Solange e a minha Avó Casiana por todo o apoio incondicional.

Ao meu Orientador César Sampaio, pela exímia orientação, compreensão, paciência e ensinamentos durante esse trabalho de conclusão.

A todos os professores que fazem parte do corpo docente do departamento, em especial ao professor Ramos, pela sua compreensão e ensinamentos

A todos os meus amigos e colegas, em especial ao Antonio Geovane, Mateus Martins, Pedro Deleon e Francisco Rubens.

“Aja como se cada ação fosse a última ação de sua vida.”

(MARCUS AURELIUS, 167 d.C.)

RESUMO

Neste trabalho é estudado medidas de centralidades em redes de colaboração de coautores para físicos teóricos notáveis, Albert-László Barabási, Giorgio Parisi, Frank Wilczek e Donald Truhlar. As medidas foram realizadas utilizando a representação das redes em grafos e em hipergrafos. Ao longo do procedimento, foi utilizado sub-redes para o cálculo das centralidades. Além disso, foi comparado os resultados encontrados usando-se grafos e hipergrafos para as redes de colaborações, o que promoveu a discussão da viabilidade do uso de hipergrafos no estudo de redes.

Palavras-chave: grafos; hipergrafos; centralidades; redes.

ABSTRACT

We investigate the use of graphs and hypergraphs to measure network centralities, the networks used was co-authorship networks of four theorical physicist: Albert-László Barabási, Giorgio Parisi, Frank Wilczek and Donald Truhlar. The measures was made using the graph representation and the hipergraph one, during the procedure we used sub-networks to measure the centralities, because if was analysed the whole network we could face problems during the cleaning, normalization and visualization process. Besides it was compared the results usind graph and hyperhraph which formented the discution of the viability of the use of hypergraphs in networks.

Keywords: graph; hypergraphs; centrilities; networks.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

CI *Centralidade de Intermediação*

CP *Centralidade de Proximidade*

LISTA DE SÍMBOLOS

$G(V,E)$	Grafo com V vértices e E nós
$A^{n\times m}$	Matriz Adjacente
k_i	Grau de Conexões
$P(k)$	Distribuição do grau de conexões
$d(v_i,v_j)$	Distância entre os nós v_i e v_j
C_i	Closseness Centrality
g_i	Betwenness Centrality
$ecc(v)$	Excentricidade do nó v
$H(V,E)$	HiperGrafo com V nós e E hiper-links
I	Matriz incidente
$\Omega(E,X)$	Dual do hiper-grafo $H(V,E)$
$C_s(i)$	s-Closseness Centrality
$g_s(i)$	s-Betwenness Centrality
$ecc_s(u)$	s-Excentricidade

SUMÁRIO

1 INTRODUÇÃO

O estudo de grafos tem-se mostrado uma ferramenta fundamental em diversas áreas do conhecimento como: Matemática, Física, Engenharias, Ciência da Computação etc,. As propriedades topológicas dos grafos nos permitem modelar e analisar estruturas complexas, tanto de forma quantitativa quanto qualitativa. Ademais, a dinâmica de grafos mostra-se fundamental tanto nas ciências exatas quanto nas ciências humanas (DAVERN, 1997).

Porém, com a possibilidade de base de dados cada vez maiores é notório que há sistemas na natureza que a abordagem em ligações somente de pares, ou seja, uma aresta para cada dois nós, não é suficiente para descrever completamente tais sistemas (FATEMI *et al.*, 2019).

As evidências empíricas desse fenômeno aparecem em fenômenos que envolvem um grade número de grupos de colaboração como: interação proteica (KLIMM *et al.*, 2021), redes de coautores (LUNG *et al.*, 2018), teoria da informação (VIGNESWARAN *et al.*, 2020), etc. Dessa forma, com o intuito de abordar apropriadamente tais temas teóricos desenvolveram uma generalização de grafos conhecida como hipergrafo. A título de comparação, podemos entender um hipergrafo como uma generalização natural de um grafo assim como tensores são uma generalização natural de vetores.

Assim como tensores, hipergrafos podem adicionar complexidades ao sistema, mas também adicionam maneiras mais abrangentes de lidar com redes e suas propriedades métricas possibilitando melhores *insights* em como abordar o sistema estudado.

O estudo de hipergrafos tem-se mostrado muito emergente nas últimas décadas, por exemplo, há grupos de físicos teóricos tentando descrever fenômenos fundamentais na natureza como: causalidade, propriedades gravitacionais e dimensionais do universo a partir de simples propriedades encontradas em hipergrafos (GORARD, 2021).

2 REDES E SUAS PROPRIEDADES MÉTRICAS

Formalmente, as estruturas que regem o comportamento e dinâmica de uma rede são os grafos, tais estruturas representam relações em pares entre objetos. Um grafo é um par ordenado $G(V,E)$, onde V é um conjunto de vértices (também conhecido como nó) e E é um conjunto de *links* ou arestas. Tal grafo é representado por uma matriz adjacente A^{nxn} onde $n = |V|$ e $m = |E|$.

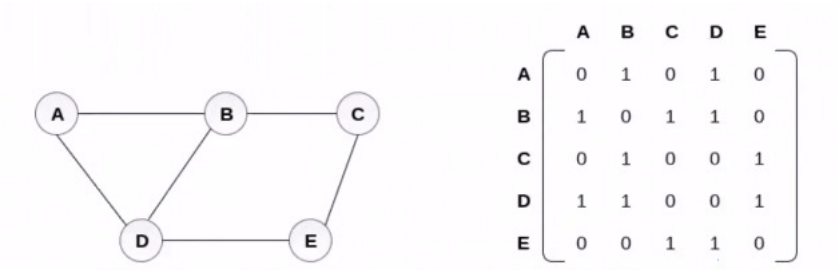

Figura 1 – Grafo indireto e sua respectiva matriz adjacente. Fonte: Criado pelo autor

A figura 1 mostra um grafo indireto e sua representação matricial, um grafo indireto é aquela cuja arestas não tem uma direção precisa, ou seja, a conexão entre dois nós é mútua, por exemplo, se eu conheço meu amigo logo meu amigo também me conhece. Poderá haver redes na qual não há uma conexão mútua como é o caso das redes sociais, onde é possível seguir o perfil de um indivíduo mas tal individuo não o segue de volta. Todas as redes analisadas nesse trabalho de conclusão de curso são do tipo indiretas.

Quando se analisa uma rede um dos parâmetros mais fundamentais é conhecer como os nós estão conectados, a forma de quantificar essa grandeza é medindo o grau $k_i = \sum_{j=1}^{n} A_{ij}$, que é o número de nós que estão conectados à i, ou seja, k_i mede a quantidade de "amigos"que o indivíduo i tem. Logo, é natural definir a média de k_i para uma rede de n indivíduos e m arestas.

$$\overline{k} = \frac{1}{n}\sum_{i=1}^{n} k_i = \frac{2m}{n} = \frac{2|E|}{|V|}$$

Essa grandeza nos diz quantas conexões em média há em uma rede, por exemplo, quantas pessoas em média um estudante de física da UFC irá conhecer durante sua graduação.

Outra propriedade importante é a distribuição dos graus de conexão k_i essa distribuição se resume a contar quantos nós com grau k_i estão presentes na rede, portanto.

$$P(k) = \frac{n_k}{\sum_k n_k}$$

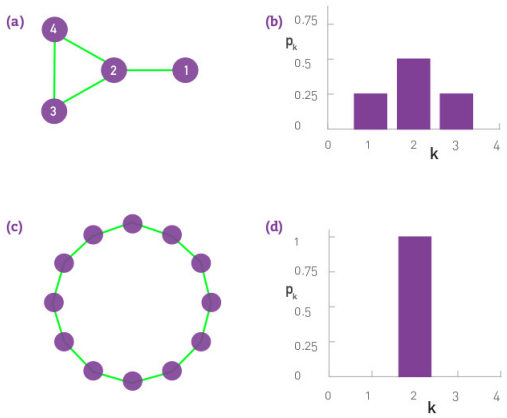

Figura 2 – Representação da distribuição $P(k)$ para dois grafos distintos. Fonte: Retirada do livro (BARABÁSI, 2014)

A partir da figura 2, percebemos que a distribuição $P(k)$ nos diz muito sobre a forma de um grafo, uma vez que $P(k)$ representa a probabilidade de encontra um nó de grau k na rede.

2.1 Caminhos, Transitividade e Coeficiente de Agrupamento

Se estamos lidando com uma rede conectada, ou seja, $\overline{k} \geq 1$ então podemos definir a distância $d(v_i, v_j)$ entre nós, que é o número de arestas que constituem o menor caminho entre o nó v_i à v_j. Assim como $\overline{k}$ queremos definir uma métrica global que expresse a média de $d(v_i, v_j)$, ou seja, queremos um número que nos diga quantos passos em média é preciso dar para ir para de um nó para outro.

$$\overline{L} = \frac{1}{n(n-1)} \sum_{i \neq j} d(v_i, v_j)$$

Onde $n(n-1)$ pode ser pensado como duas vezes o número total de pares de nós conectados na rede. Uma métrica local muito usada é o coeficiente de agrupamento, que mede a razão entre o número N_i de links na vizinhança do nó i e o valor máximo possível para tal nó, ou seja:

$$C_i = \frac{N_i}{k_i(k_i-1)/2}$$

Onde $k_i(k_i-1)/2$ é o número máximo de conexões entre vizinhos, podemos resumir o coeficiente de agrupamento como uma grandeza que mede quantos triângulos existem ao redor de um nó, por exemplo, se todos os meus amigos se conhecem então o meu coeficiente de agrupamento é 100%.

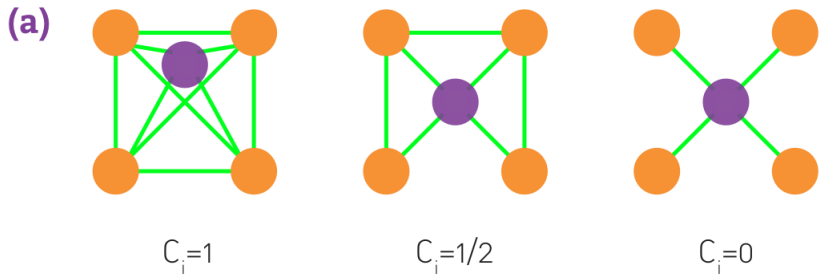

Figura 3 – Representação de diferentes valores para o coeficiente de agrupamento. Fonte: Retirada do livro (BARABÁSI, 2014)

Dessa formar, podemos definir o coeficiente de agrupamento global ou transitividade que basicamente mede o coeficiente de agrupamento da rede como um todo.

$$C_{global} = \frac{3 \times \Delta}{\eta}$$

Onde Δ é o número de triângulos no grafo e η é o número de trios ou triângulos incompletos conectados no grafo, por exemplo, se temos os nós (A,B,C) , onde A está conectado com B e C mas B não está conectado com C, logo está faltando uma aresta em BC para formar um triângulo. Tanto o coeficiente de agrupamento local quanto global são propriedades muito presentes em redes sociais, tal propriedade pode nos dizer muito como seres humanos se organizam em sociedade (SAID *et al.*, 2018).

Outra importante medida em grafos é a centralidade de proximidade, que é uma grandeza que diz quão próximo um nó está em relação à todos os outros nós na rede. Logo, podemos definir tal métricas como:

$$C_i = \frac{n}{\sum_j d(i,j)}$$

Podemos interpretar os nós com maior centralidade de proximidade, como aqueles repensáveis por passar informação de forma eficiente para o resto da rede.

Também podemos definir outra métrica conhecida como centralidade de intermediação que é uma medida de centralidade na rede baseada nos mínimos caminhos na rede , ou seja, dado o nó i:

$$g_i = \sum_{j \neq i \neq t} \frac{\sigma_{jt}(i)}{\sigma_{jt}}$$

Onde σ_{jt} é o número total de distâncias mínimas do nó j para o nó t e $\sigma_{ij}(i)$ é a quantidade de tais distâncias que passam por i. A centralidade de intermediação pode ser pensada como uma formar de medir a quantidade de influência que um nó tem sobre o fluxo de informação na rede (KIM, 2010).

Por fim, podemos definir a excentricidade de um nó v que pertence à rede V , como sendo a maior distância do nó v e todos os demais nós da rede.

$$ecc(v) = max(d(v,u) : u \in V)$$

Logo, a excentricidade é uma grandeza que computa o maior caminho entre o nó v e todos os demais nós, portanto se a excentricidade de um nó é pequena isso significa que todos os outro nós são próximos a ele. O maior valor da excentricidade em uma rede é igual ao valor do diâmetro da rede e o menor valor da excentricidade é igual ao raio da rede.

3 HIPEREDES E SUAS PROPRIEDADES MÉTRICAS

A diferencia fundamental entre um grafo e um hipergrafo é que em grafos as conexões são em pares, ou seja, uma aresta pode se conectar apenas à dois nós. Já em hipergrafos as conexões são em grupo, isto é, as arestas podem abranger mais de dois nós ao mesmo tempo.

Logo, um hipergrafo $H(V,E)$ é um par (V,E) onde V é um conjunto de nós e E é uma família de subconjuntos de V, ou seja, agora as arestas podem se conectar a mais de dois nós ao mesmo tempo. A matriz representativa para hipergrafos é a matriz incidente I de tamanho $n \times m$, onde n é o número de nós e m é o número de arestas. Por exemplo, seja $H(V,E)$ composto por $V = \{A,B,C,D,E\}$ e $E = \{e_1,e_2,e_3\}$.

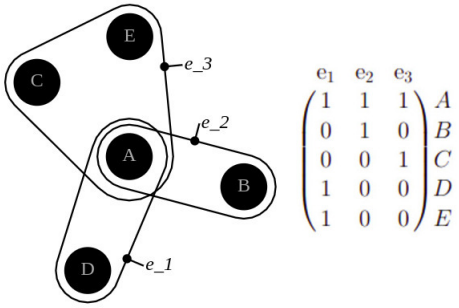

Figura 4 – Hipergrafo e sua respectiva matriz incidente. Fonte: Criado pelo autor

Perceba pela figura 4 que diferentemente da matriz adjacente a matriz incidente não será necessariamente simétrica. Veja que as arestas e_2 e e_3 estão conectadas a apenas dois nós, logo tais links contém a mesma complexidade de links de um grafo usual já e_3 contém 3 nós. Essa possibilidade de expressar conexões em pares e em grupos no mesmo grafo torna o estudo de hipergrafos bastante promissor. A partir da matriz incidente I podemos fazer uma transformação que vá de $H(V,E)$ para $G(V,E)$, ou seja, transformar I em A, i.e.

$$A = I^T I - D$$

Onde D é a uma matriz diagonal cujo elementos da diagonal são os graus k_i dos nós do hipergrafo onde agora definimos $k_i = \sum_{j=1}^{n} I_{ij}$. Devido à possibilidade de uma aresta se conectar a mais de dois nós ao mesmo tempo é interessante definir o grau $k_j = \sum_{i=1}^{n} I_{ij}$ como o grau ou tamanho de uma aresta, por exemplo, pela figura 4 vemos que $k_{e_3} = 3$, pois ela contém 3 nós.

Por fim, uma última propriedade peculiar em hipergrafos é uma transformação chamada dual de um hipergrafo, que formalmente se resume a: dado um hipergrafo com E nós e

V arestas, o dual de $H(V,E)$ será $\Omega(E,X)$ onde X é a família de conjuntos de nós E contidos em V.

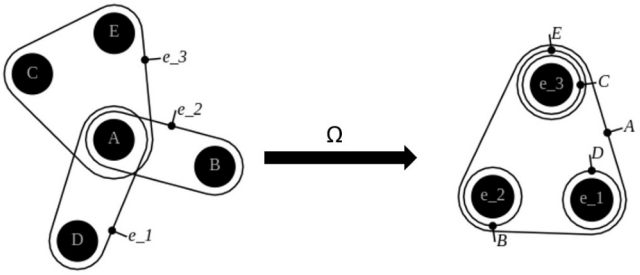

Figura 5 – Hipergrafo e seu respectivo dual. Fonte: Criado pelo autor.

Dessa forma, estamos transformando nós em arestas e arestas em nós.

3.1 Propriedades métricas de Hiperedes

Assim como grafos, hipergrafos contém componentes métricas, tais componentes podem ser pensadas como generalizações das componentes vistas em grafos. Em grafos se existe uma sequência de nós conectadas por uma sequência única de arestas denominados essa sequência de um caminho. Já em hipergrafos, existe uma redundância em definir caminhos, pois geralmente não há uma sequência única de arestas, uma vez que pode haver arestas conectando mais de dois nós. Pela figura 6 vemos que se definirmos no grafo a sequência de nós (E,C,D,A)

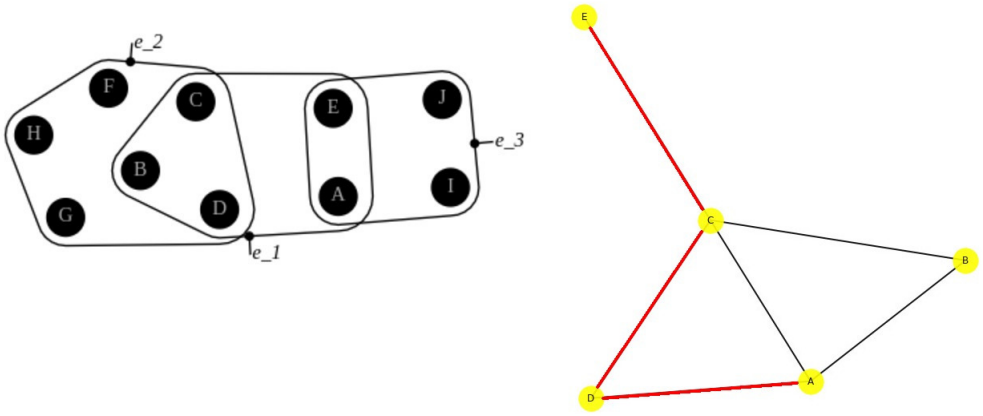

Figura 6 – Comparação entre caminhos em grafo e hipergrafos. Fonte: Criado pelo autor.

, teremos um caminho unívoco que liga tal sequência. Já se fizermos tal sequência no hipergrafo

teremos uma redundância no caminho, pois há intersecções entre as arestas, tais intersecções não permitem criar um caminho unívoco em um hipergrafo, ou seja, temos uma espécie de degenerescência ao definirmos caminhos em hipergrafos usando essa abordagem. Logo, ao tratarmos hipergrafos torna-se mais claro definirmos caminhos entre arestas e não entre nós.

Em hipergrafos, uma caminhada de tipo s é a sequência de arestas $(e_1, e_2, ...)$ tal que $e_i \cap e_{i+1} \geq s$, por exemplo, na figura 6 o caminhado é de tipo s=2, pois a quantidade de nós entre as intersectações é pelo menos 2. Logo, o valor de s nos permite aferir a forca da conectividade entre arestas, em outras palavras, quanto maior o valor de s mais robusta será a conexão e consequentemente a hiperede.

Como discutido anteriormente, certas generalizações em hipergrafos tornam-se um desafio, como é o caso do coeficiente de agrupamento, pois dependendo do grau s, já discutido, teremos diferentes triângulos. Logo, é necessário fixar o grau s e analisar no hipergrafo apenas os elementos com essa propriedade. Logo, podemos definir o coeficiente de agrupamento do tipo s para o nó f como:

$$C_s(f) = \frac{2\sum_{v,w \in E_s} I_{E_s}(v,w,f)}{N(f)[N(f)-1]}$$

Onde, $E_s(v,w,f)$ é diferente de zero se há um "hipertriangulo"entre os nós v,w,f. Similarmente podemos definir a centralidade de fechamento para hipergrafos como.

$$C_s(i) = \frac{E_s - 1}{\sum_{f \in E_s} d_s(i,f)}$$

Onde, $E_s = \{i \in E : i \geq s\}$, ou seja, estamos restringindo nossa análise para arestas de tipo pelo menos s , devido à redundância discutida anteriormente. Outra generalização que podemos fazer é sobre a centralidade de proximidade.

$$g_s(i) = \sum_{j \neq i \neq f \in E_s} \frac{\sigma^s_{jt}(i)}{\sigma^s_{jt}}$$

Onde, a única diferença para a definição em grafos é que agora estamos introduzindo o vínculo E_s, já discutido anteriormente. Por fim, a redefinição da excentricidade em hipergrafos é bastante direta.

$$ecc_s(u) = max(d(u,v); v \in V$$

Onde a única diferencia da definição utilizada em grafos, é que em hipergrafos é necessário definir o grau s da conexão.

3.2 Amostragem, Matriz de Correlação, Fator de Impacto e Grafo Bipartido

3.2.1 Amostragem

Em estatística Amostragem (sampling) , é um subconjunto dos dados estudados com a caraterística de ser representativo à todo o conjunto de dados, ou seja, esse pequeno subconjunto tem caraterísticas similares ao conjunto como um todo.

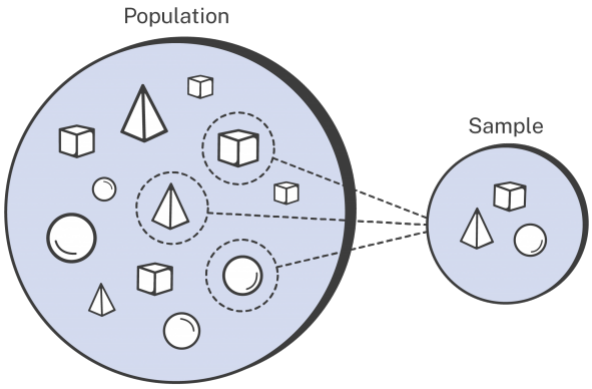

Figura 7 – Ilustração do uso da amostragem. Fonte:<https://supermetrics.com/blog/google-analytics-sampling> , acesso: 26 de outubro de 2022.

Em um mundo com dados cada vez maiores o uso de amostragem tem-se tornando uma ferramenta essencial para qualquer inferência ou estudo estatístico. A grande questão que emerge ao tentar criar uma amostragem é quais critérios os indivíduos da população devem ter para serem escolhidos para compor a amostragem, normalmente aplica-se uma abordagem aleatória e variações dessa abordagem são propostas na literatura (ALTMANN, 1974).

3.2.2 Matriz de Correlação

Vivemos em um mundo rodeado por correlações, por exemplo a altura dos pais de uma criança e a altura da criança ou a qualidade dos serviços de um município e seu PIB. Os dois exemplos apresentados anteriormente representam correlações, pois quando uma grandeza cresce ou descresse a outra tende a segui-la de forma proporcional, em algumas situações, pode haver anti-correlações que é quando uma grandeza varia inversamente à outra, por exemplo a quantidade de óbitos por COVID-19 e o número de vacinados. Formalmente, sejam as variáveis aleatórias X e Y e μ_X μ_Y suas respectivas medias, a correlação $corr(X,Y)$ é dada por:

$$corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

Onde $\sigma(i)$ é o desvio padrão da variável i e $E(i)$ é o operador média de i, ou simplesmente média de i. Onde se $corr(X,Y) = 0$ X e Y não estão correlacionados se $corr(X,Y) = 1$ X e Y são completamente correlacionados e se $corr(X,Y) = -1$ então X e Y são completamente anti-correlacionados.

3.2.3 Fator de impacto

O fator de impacto de uma revista científica avalia a relativa importância da revista em seu "nicho"científico, o fator de impacto F_y é uma medida anual e relaciona o número de citações e publicações que uma dada revista fez durante um determinado ano.

$$F_y = \frac{\beta y}{\alpha_{y-1} - \alpha_{y-2}}$$

Onde β_y é o número de citações que a revista obteve durante o ano y , e α_y é o número de publicações que a revista fez durante o ano y. O fator de impacto não é uma ferramenta totalmente completa para medir a qualidade dos artigos, porém ele é uma boa aproximação para tal. Ademais, as revistas com maior fator de impacto geralmente são aquelas mais difíceis de admitirem artigos, tal fato mostra a relevância do fator de impacto à comunidade cientifica.

3.2.4 Grafo Bipartido

Um grafo bipartido é um grafo cujo vértices podem ser divididos em dois conjuntos distintos U e V frequentemente se escreve $G = (U,V,E)$ para denotar um grafo bipartido cuja partição tem as partes U e V. Se $|U| = |V|$, ou seja, se os dois subconjuntos tem o mesmo número de elementos.

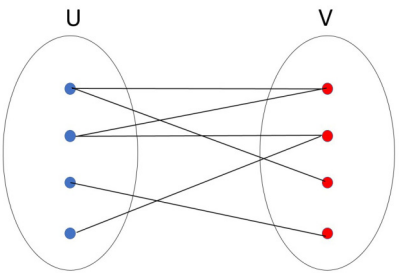

Figura 8 – Ilustração de um grafo bipartido com conjuntos U e V. Fonte: Criado pelo autor.

Grafos bipartidos são usados extensivamente para modelagem de relações entre duas classes diferentes de objetos. Por exemplo, um grafo de jogadores e clubes de futebol, com uma aresta entre um jogador e um clube caso o jogador tenha jogado por aquele clube. Ademais grafos bipartidos podem ser utilizados para a modelagem das redes de colaboração de coautores estudadas neste trabalho,

4 METODOLOGIA

Os dados dos pesquisadores foram obtidos na plataforma Google Scholar por meio do pacote scholary da linguagem R, os dados foram analisados e tratados utilizando-se tanto R quanto pacotes no Python. Foi criado um conjunto de dados para cada pesquisador cuja linhas são os coautores que compõem a rede de colaboração e as colunas são as revistas na qual o pesquisador publicou junto com os co-autores, por exemplo, digamos que estamos analisando um pesquisador com a seguinte rede.

	Nature	Science	Physical Review Letters
João	1	0	4
José	0	1	0
Maria	2	0	1

Tabela 1 – Exemplo de uma matriz que representa a rede de colaboração para o pesquisador estudado.

Pela tabela 1, vemos que o pesquisador estudado publicou junto com João uma vez na Nature e quatro vezes na Physical Review Latters, dessa forma, podemos interpretara a tabela 1 como a matriz incidente I do pesquisador estudado.

Com intuito de melhorar a visualização e diminuir vieses foi estudado sub-hipergrafos para cada pesquisador, ou seja, foi escolhido apenas certas revistas (colunas). O critério para a escolha dessas colunas foi baseado no fator de impacto da reavista e pela matriz do correlação das revistas escolhidas. Tais critérios não caracterizam uma amostragem, pois as escolhas das revistas não foram aleatórias mas o objetivo final é obter um subconjunto que seja característico à todo o hipergrafo.

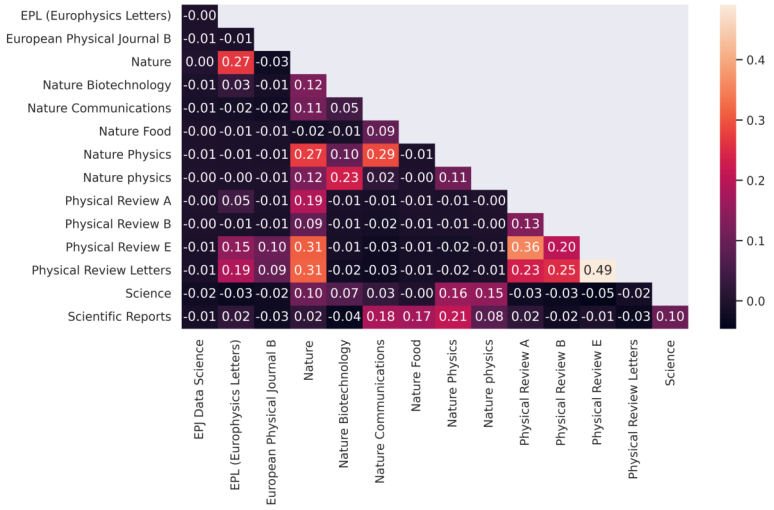

Figura 9 – Representação da matriz de correlação para o sub-hipergrafo do perquisidor Barabási. Fonte: Criado pelo autor.

Por exemplo, a partir da matriz de correlação da figura 8 vemos que a revista EPJ Data Sciencee é a revista que menos se correlaciona com as demais, logo ela não seria adicionada ao sub-hipergrafo. Os critérios usados para criar os sub-hipergrafos são critérios heurísticos e uma outra forma mais precisa e não enviesada para a escolha dos sub-hipergrafos seria o uso da técnica dos motifs (LOTITO *et al.*, 2022), porém tal técnica foge do escopo desse trabalho de conclusão de curso .

Um outro método heurístico empregado para verificar se o sub-hipergrafo é representativo foi a contagem da quantidade de palavras usadas nos títulos das publicações usando todo a população e comparar com o a amostra.

A figura 10 mostra a frequência de palavras dos títulos das publicações do pesquisador Barabási para todo o hiper-grafo, ou seja, para a população completa, quanto maior o tamanho da palavra maior a frequência sua frequência.

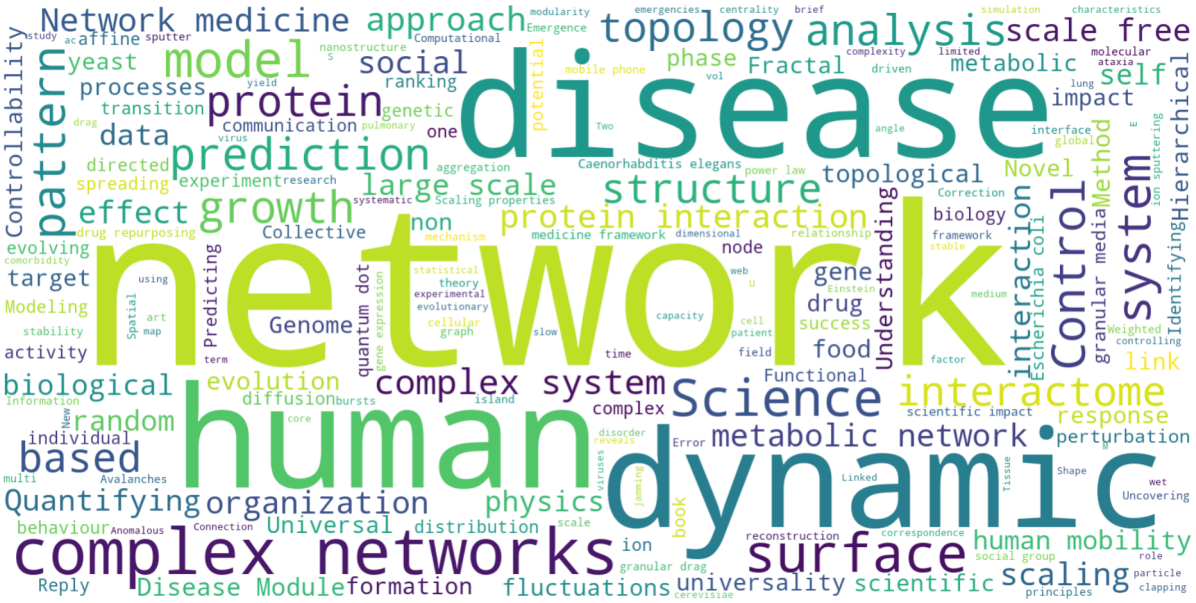

Figura 10 – Contagem da frequência de palavras utilizadas nos títulos das publicações para toda a rede de colaboração do pesquisado Barabási, 27105 palavras utilizadas. Fonte: Criado pelo autor.

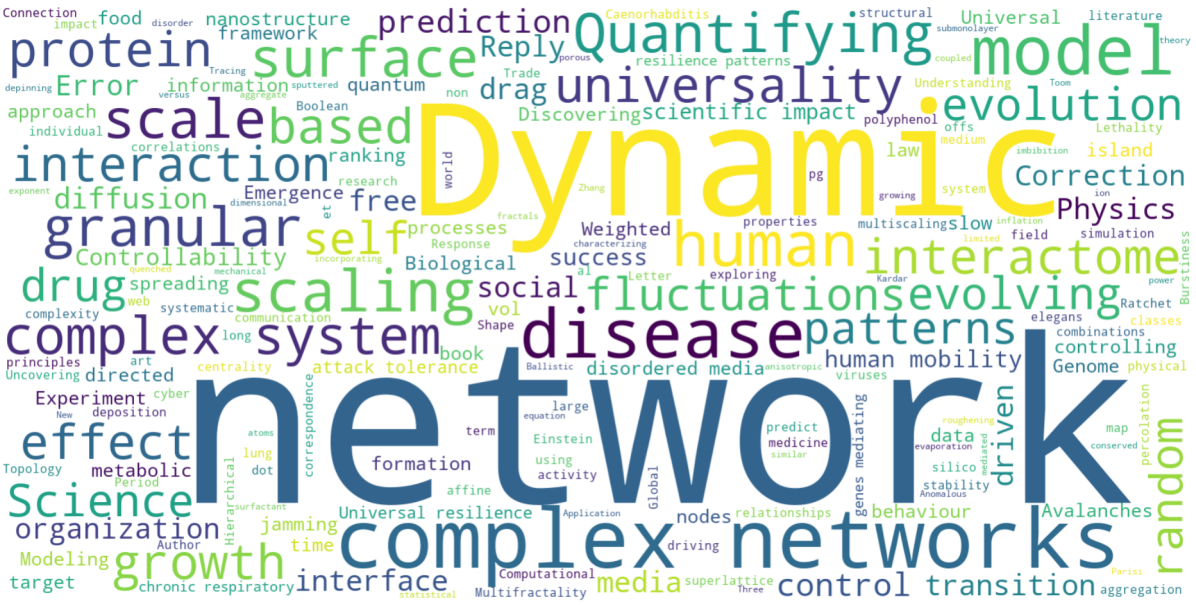

Figura 11 – Contagem da frequência de palavras utilizadas nos títulos das publicações para um sub-conjunto da rede de colaboração do pesquisado Barabási, 9969 palavras utilizadas. Fonte: Criado pelo autor.

A figura 11 mostra a frequência de palavras utilizando um sub-hipergrafo do pesquisador Barabasi, podemos notar que a figura 10 nos dá um insight de que a amostra escolhida é representativa à todo o hipergrafo, pois as mesmas palavras aparecem aproximadamente com a mesma frequência.

5 RESULTADOS

O primeiro resultado encontrado a ser comparado foi a comparação visual entre o hipergrafo e o seu equivalente grafo, tal comparação pode nos ajudar na caracterização da rede e suas propriedades.

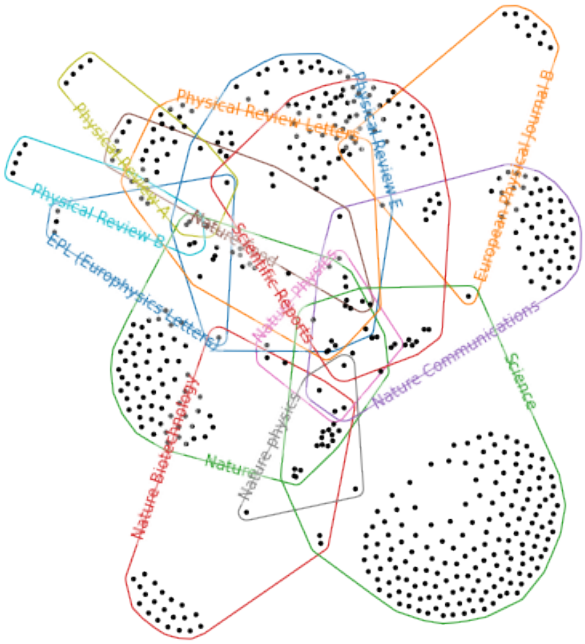

Figura 12 – Sub-hipergrafo da rede de colaboração do pesquisador Barabási. Fonte: Criado pelo autor.

A figura 12 nos informa como os grupos de pesquisadores estão segmentados a partir das arestas, porém ainda podemos observar dificuldades na visualização do hipergrafo e na observação dos relacionamentos entre os nós, tal problema se agrava quanto maior for o hipergrafo.

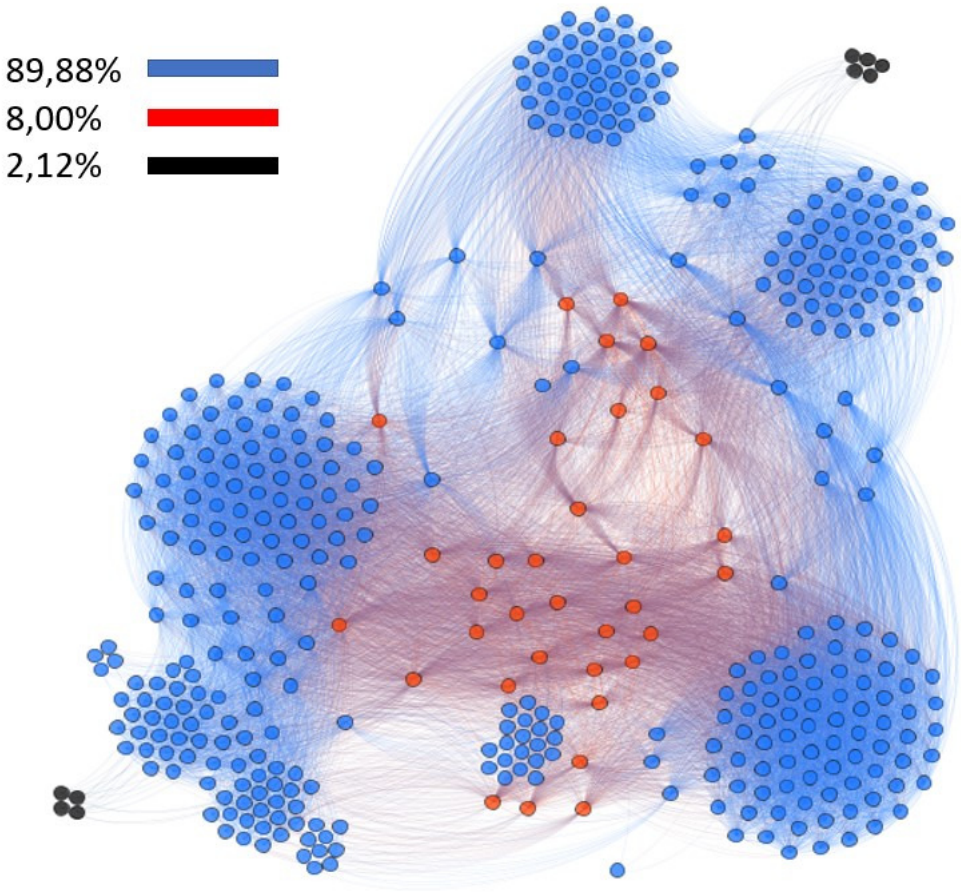

Figura 13 – Grafo da rede de colaboração do pesquisador Barabási, onde os nós estão segmentados pelo valor da sua excentricidade, preto 4, azul 3 e vermelho 2 . Fonte: Criado pelo autor.

Pela figura 13 vemos que a rede na representação de grafo nos dá informações sobre como os nós se relacionam um com o outro, já do ponto de vista de detecção de comunidades devemos fazer algum cálculo prévio como no caso acima onde foi calculado a excentricidade, o que pode ser visto como uma desvantagem ao compararmos com hipergrafos.

Para obter as medidas das redes usando hipergrafos foi utilizado o dual dos hipergrafos de cada pesquisador, pois se medíssemos os hipergrafos estaríamos usando as revistas como nós. Ademais, todas as grandezas obtidas no hipergrafo foram usando o grau s=2, se quiséssemos encontrar os mesmo resultados tanto em grafos quanto em hipergrafos deveríamos usar a condição s=1, logo a condição s=2 irá nos dar Informação extras que grafos não iriam nos fornecer.

5.1 Resultados Usando Grafos

Os resultados da centralidade de proximidade foram representados por meio de histogramas, onde foi ressaltado os 3 nós com maior e menor *Centralidade de Proximidade* (CP).

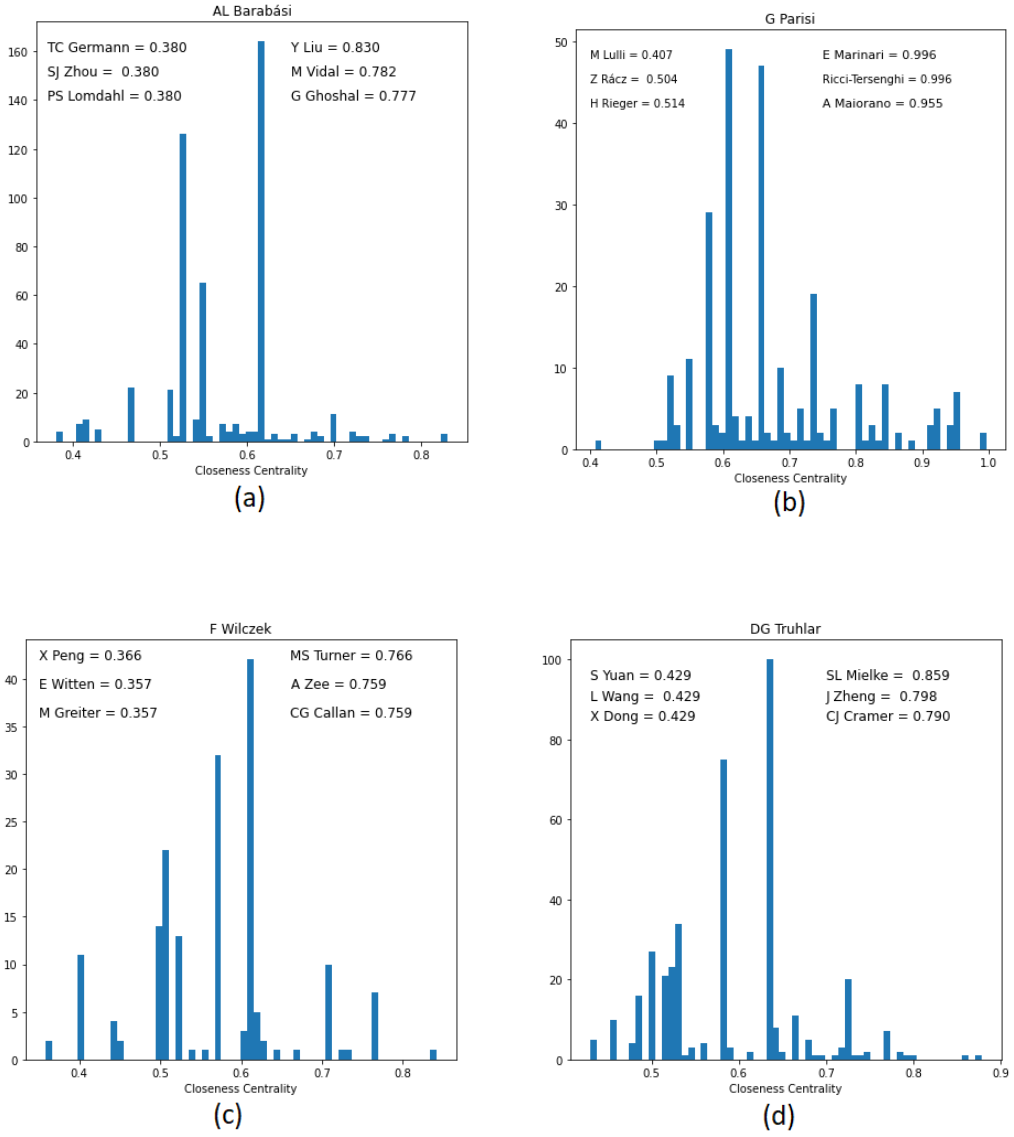

Figura 15 – Distribuição da CP para os 4 pesquisadores estudados, valores obtidos por grafos.
Fonte: Criado pelo autor.

Durante a obtenção dos resultados, observou-se que revistas com uma alta correlação e alto fator de impacto eram capazes de alterar a distribuição da CP drasticamente, talvez esse fato seja porque tais revistas apareciam com uma maior frequência comparada com as outras.

O segundo resultado encontrado foi a centralidade de intermediação e com o intuito de melhor mostrar os resultados optou-se em mostrar os resultados na forma da distribuição

tipo violino, tal método de visualização é muto similar ao Box Plot com a diferença de que ele é capaz de mostrar a distribuição de probabilidade de cada ponto. Ademais, como antes, foi adicionado os três nós com maior valor de *Centralidade de Intermediação* (CI).

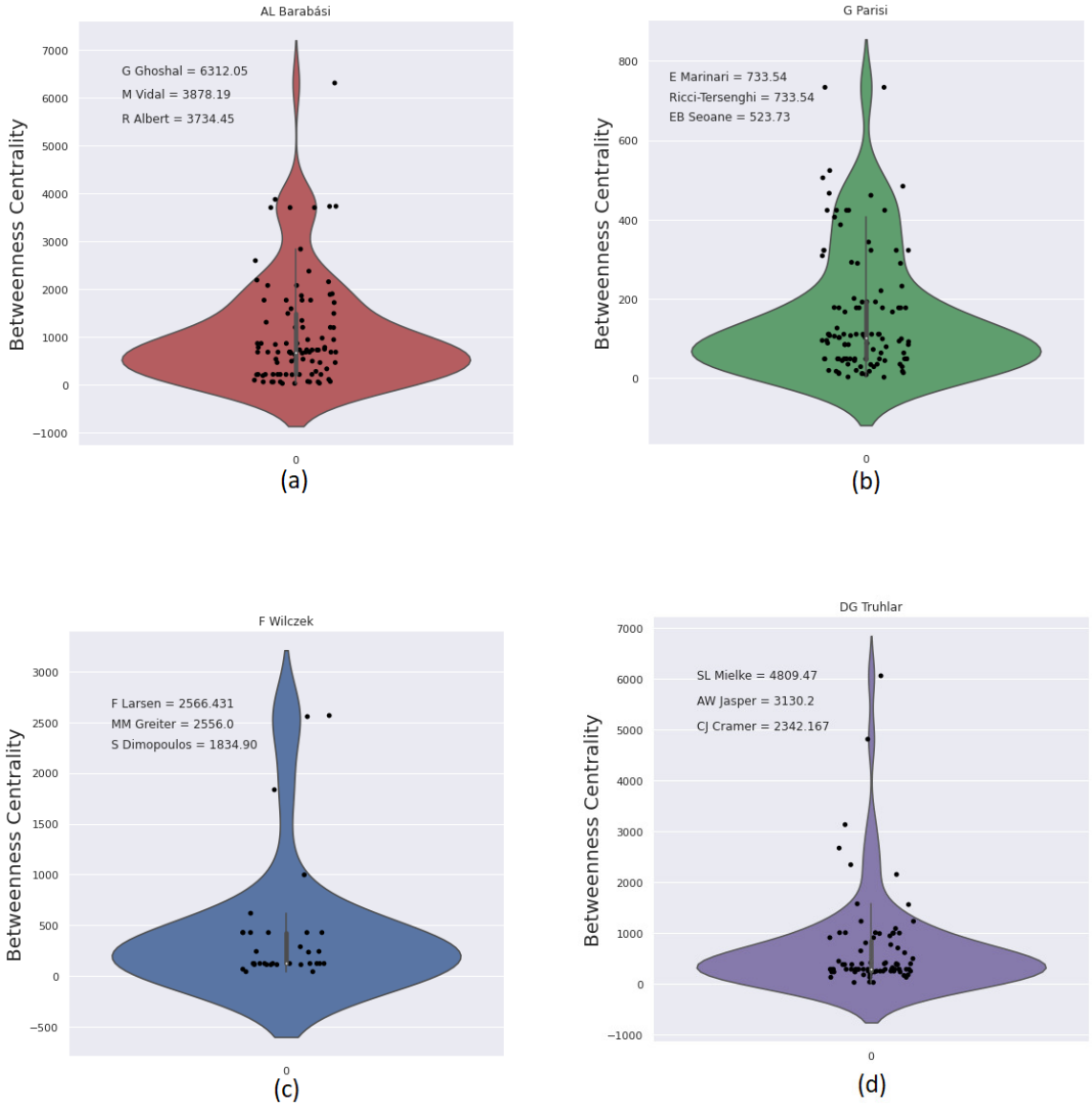

Figura 17 – Distribuição da CI para os 4 pesquisadores estudados, valores obtidos por grafos. Fonte: Criado pelo autor.

Pela figura 17 nota-se que para as quatro redes de colaboração estudadas, as distribuições estão em torno do zero e que há poucos indivíduos com grande CI, ademais os indivíduos com maior CI também são aquelas com maior CP.

Por fim, foi calculado a excentricidade das redes estudadas, o que nos permite obter o diâmetro e raio da rede e observar a relação diâmetro e raio por meio da excentricidade media.

	Diâmetro	Raio	Excentricidade Média
AL Barabási	4	2	2,95
G Parisi	3	2	2,46
F Wilczek	4	2	3.10
DG Truhlar	3	2	2.37

Tabela 2 – Diâmetro e raio das 4 redes estudadas.

5.2 Resultados Usando Hipergrafos

Como dito anteriormente, os resultados usando hipergrafos foram obtidos a parir do dual do hipergrafo por meio do pacote hypernetx, ademais foi utilizado o vínculo s=2, arestas com 2 ou mais nós entre as conexões, tais características devem resultar em novos resultados ao compararmos com grafos.

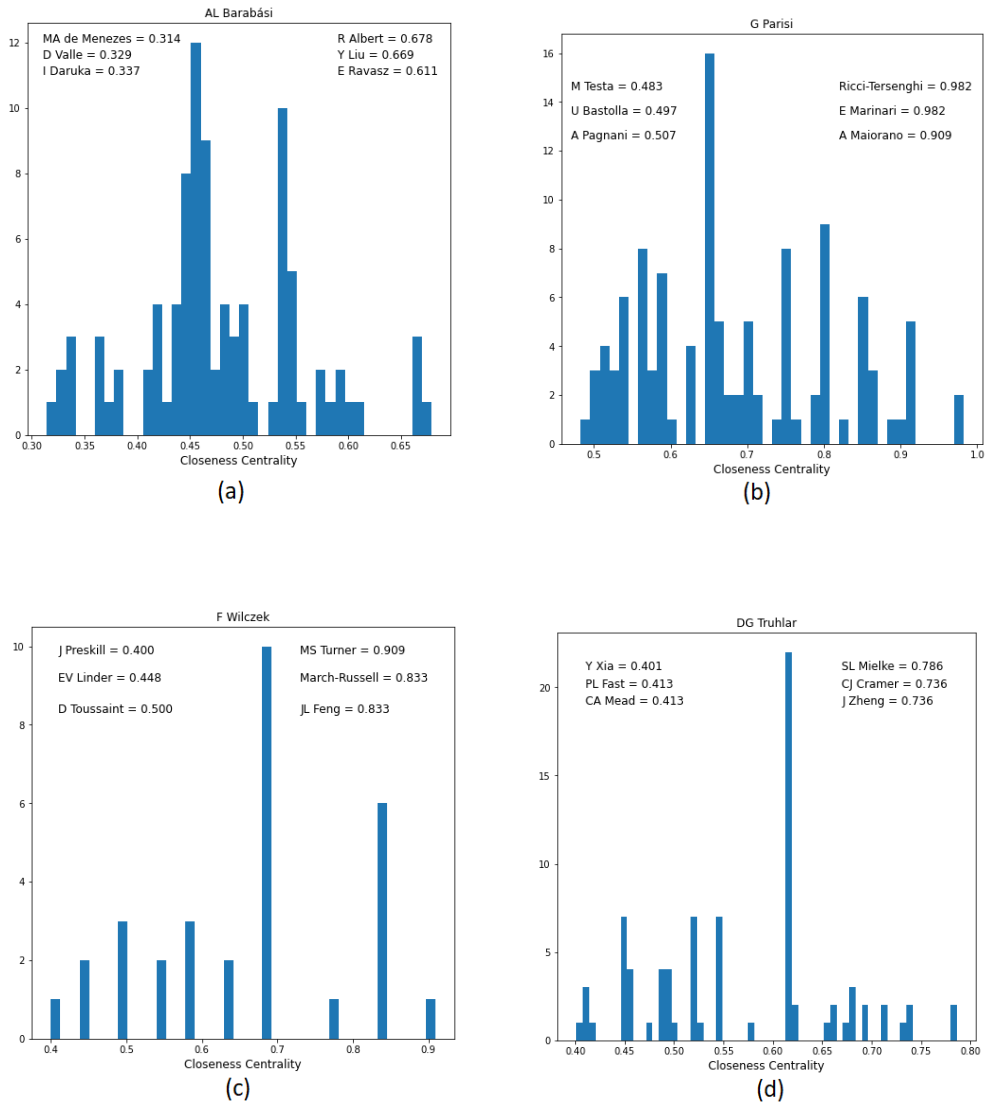

Figura 19 – Distribuição da CP para os 4 pesquisadores estudados, valores obtidos por hipergrafos com s = 2. Fonte: Criado pelo autor

Pela figura 19, nota-se que o eixo horizontal não começa no zero, isso se deve ao fato de que os nós que não satisfazem a condição s=2 irão ter a CP igual a zero, logo tais nós não foram adicionados à distribuição. Ademais, ao compararmos os resultados da CP em grafos e hipergrafos notamos que certos nós com maior CP em grafos também aparecem em hipergrafos , por exemplo, para o AL Barabasi o nó G Ghoshal aparece como um dos maiores valores tanto em grafos quanto hipergrafos, dessa forma tal nó cumpre um papel central à rede.

Assim como discutido anteriormente, os valores da CI foram calculados com s=2 e foram destacados os três indivíduos com maior valor.

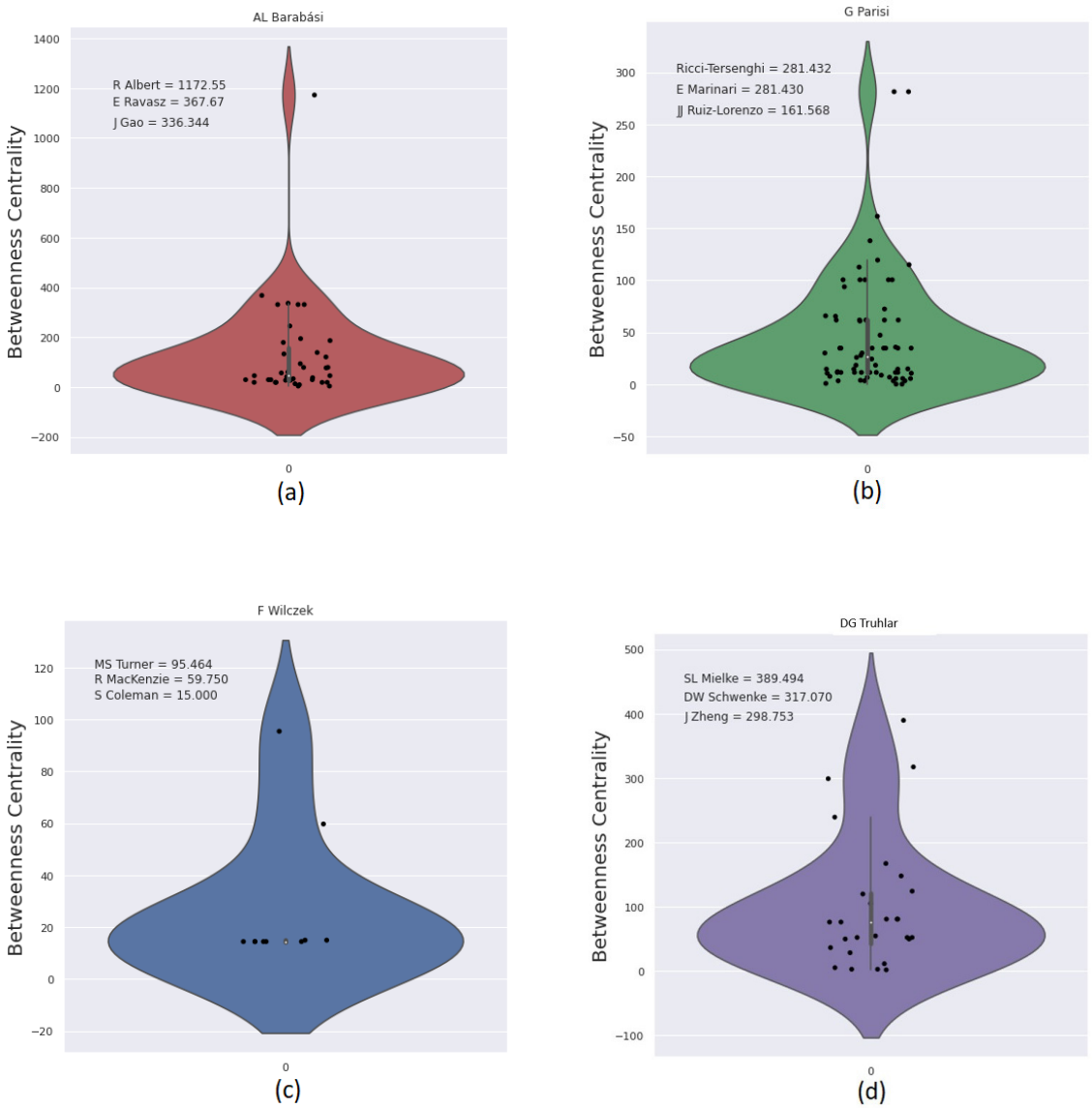

Figura 21 – Distribuição da CI para os 4 pesquisadores estudados, valores obtidos em hipergrafos com s=2. Fonte: Criado pelo autor.

Pela figura 21, percebemos o mesmo comportamento observando na CP, por exemplo, para o pesquisador G Parisi vemos que o nó Ricci-Tersenghi aparece com uma das maiores CI

tanto no grafo quanto no hipergrafo.

Por fim, ao calcular a excentricidade do hipergrafo com s=2 foi considerado apenas os nós com excentricidade diferente de zero, pois como discutido anteriormente indivíduos com o valor igual a zero não satisfazem o vínculo s=2 e não pertencem à rede.

	Diâmetro	Raio	Excentricidade Média
AL Barabási	5	3	3,75
G Parisi	3	2	2,14
F Wilczek	3	2	2,67
DG Truhlar	4	2	2.91

Tabela 3 – Diâmetro e raio das 4 redes estudadas, utilizando hipergrafos com s=2

Notemos, pela tabela 3 que o diâmetro no hipergrafo é maior do que seu equivalente grafo, isso se deve pois em hipergrafos estamos tratando apenas os nós com s=2, ou seja, cada nó já publicou em duas ou mais revistas na rede de colaboração, logo o número de caminhos possíveis é maior comparado com grafos.

6 CONCLUSÕES E TRABALHOS FUTUROS

A partir dos fatos discutidos anteriormente, concluímos que o uso de hipergrafos no estudo de redes mostra-se muito útil para o entendimento das caraterísticas encontradas em redes. Como mostrado, o uso de hipergrafos pode nos dar um complemento e aprofundamento de grandezas encontradas tradicionalmente usando-se grafos. Contudo como já mencionado, ainda existem limitações no que tange o caráter visual de hipergrafos, mais especificamente, ao tratarmos grandes redes hipergrafos não são capazes de expressar claramente os aspectos visuais da rede.

Não obstante, hipergrafos podem ressaltar nós importantes à rede, ademais hipergrafos mostram-se uma excelente forma de expressar redes bipartidas. Por fim, ressaltamos a necessidade de trabalhos futuros referentes à dinâmica e modelos de hipergrafos, tais trabalhos irão fomentar os argumentos utilizados neste trabalho.

REFERÊNCIAS

ALTMANN, J. Observational study of behavior: sampling methods. **Behaviour**, Brill, v. 49, n. 3-4, p. 227–266, 1974.

BARABÁSI, A.-L. Network science book. **Network Science**, Cambridge University Press Cambridge, v. 625, 2014.

DAVERN, M. Social networks and economic sociology: a proposed research agenda for a more complete social science. **American journal of Economics and Sociology**, Wiley Online Library, v. 56, n. 3, p. 287–302, 1997.

FATEMI, B.; TASLAKIAN, P.; VAZQUEZ, D.; POOLE, D. Knowledge hypergraphs: prediction beyond binary relations. **arXiv preprint arXiv:1906.00137**, 2019.

GORARD, J. Hypergraph discretization of the cauchy problem in general relativity via wolfram model evolution. **arXiv preprint arXiv:2102.09363**, 2021.

KIM, J.-Y. Information diffusion and δ-closeness centrality. **Sociological Theory and Methods**, Japanese Association For Mathematical Sociology, v. 25, n. 1, p. 95–106, 2010.

KLIMM, F.; DEANE, C. M.; REINERT, G. Hypergraphs for predicting essential genes using multiprotein complex data. **Journal of Complex Networks**, Oxford University Press, v. 9, n. 2, p. cnaa028, 2021.

LOTITO, Q. F.; MUSCIOTTO, F.; MONTRESOR, A.; BATTISTON, F. Higher-order motif analysis in hypergraphs. **Communications Physics**, Nature Publishing Group, v. 5, n. 1, p. 1–8, 2022.

LUNG, R. I.; GASKÓ, N.; SUCIU, M. A. A hypergraph model for representing scientific output. **Scientometrics**, Springer, v. 117, n. 3, p. 1361–1379, 2018.

SAID, A.; ABBASI, R. A.; MAQBOOL, O.; DAUD, A.; ALJOHANI, N. R. Cc-ga: a clustering coefficient based genetic algorithm for detecting communities in social networks. **Applied Soft Computing**, Elsevier, v. 63, p. 59–70, 2018.

VIGNESWARAN, C.; VS, S. S. *et al.* Unsupervised bin-wise pre-training: a fusion of information theory and hypergraph. **Knowledge-Based Systems**, Elsevier, v. 195, p. 105650, 2020.