

Hiper-redes para análise de relações de coautoria

Matheus H. B. dos Santos¹, Jussara M. de Almeida²,
Carolina R. Xavier¹, Vinícius da F. Vieira¹

¹Departamento de Ciência da Computação – Universidade Federal de São João del-Rei
São João del-Rei, MG, Brasil

²Departamento de Ciência da Computação – Universidade Federal de Minas Gerais
Belo Horizonte, MG, Brasil

vinicius@ufsj.edu.br

Abstract. *Complex networks are a powerful tool for understanding phenomena in the most diverse contexts. However, modeling networks as graphs, as it is centered on pairwise relationships, offers limitations in modeling many-to-many interactions, as is the case with collaboration in scientific articles. This work provides an overview of central concepts for the use of hypernetworks as models for representing social relations, discussing advantages and disadvantages, challenges and opportunities. The comparison of network and hypernetwork models built on CSBCSet, a database of scientific articles published in CSBC, allows exploring the impact of using hypernetworks to study the phenomenon of coauthorship of scientific articles.*

Resumo. *Redes complexas são uma poderosa ferramenta para a compreensão de fenômenos nos mais diversos contextos. Entretanto, a modelagem de redes como grafos, por ser centrada em relações em pares, oferece limitações na modelagem de interações muitos-para-muitos, como é o caso da colaboração em artigos científicos. Este trabalho traz uma visão geral de conceitos centrais para a utilização de hiper-redes como modelos para a representação de relações sociais, discutindo vantagens e desvantagens, desafios e oportunidades. A comparação de modelos de redes e hiper-redes construídos sobre a CSBCSet, uma base de dados de artigos científicos publicados no CSBC, permite explorar o impacto no uso de hiper-redes para o estudo do fenômeno de coautoria de artigos científicos.*

1. Introdução

A representação de sistemas complexos como redes, onde elementos são modelados como vértices e suas interações são modeladas como arestas, tem se mostrado como uma poderosa ferramenta para a compreensão de fenômenos nos mais diversos contextos [Barabási and Pósfai 2016, Newman 2006, Barthélemy et al. 2005]. Particularmente, a modelagem da relação entre pessoas como redes sociais fornece informação valiosa sobre a forma como sociedades se organizam, ideias são propagadas e doenças são espalhadas, apenas para citar alguns exemplos [Barrat et al. 2004, Barabási and Pósfai 2016].

Em uma forma mais simples, as relações em uma rede social podem ser modeladas por grafos nos quais as arestas conectam pares de indivíduos, representados por vértices. Entretanto, esse tipo de representação pode, muitas vezes, falhar em capturar aspectos

importantes dos fenômenos aos quais as redes sociais estão relacionados e, por isso, outras características precisam ser incorporadas ao modelo matemático do grafo correspondente. Às arestas podem ser associadas direções, a fim de distinguir a incidência das relações modeladas, ou pesos, indicando uma função que quantifica a intensidade das interações [Barrat et al. 2004, Barthélemy et al. 2005]. Pode-se, ainda, definir camadas às redes, identificando diferentes classes dos elementos [Kivelä et al. 2014, Boccaletti et al. 2014]. Aspectos temporais podem também ser associados às redes, permitindo a representação dinâmica dos elementos e suas relações [Holme and Saramäki 2012].

Modelos de grafos direcionados, ponderados, multi-camadas e temporais, apesar de muito úteis em um grande conjunto de aplicações, possuem uma importante limitação: são capazes de representar apenas interações diádicas, entre pares de elementos. Por outro lado, em vários sistemas complexos, as relações são intrinsecamente poliádicas, com uma única interação podendo ocorrer entre muitos indivíduos. Em sistemas biológicos, por exemplo, quando diversas proteínas fazem parte de uma mesma reação metabólica, a representação da rede por interações em pares deixa escapar aspectos fundamentais do processo [Battiston et al. 2020, Joslyn et al. 2021]. Em um contexto de colaboração científica, é impossível desconsiderar a natureza intrínseca das relações muitos-para-muitos sem que haja uma perda significativa de informação e a modelagem através de grafos impede que se tenha clareza, por exemplo, se uma clique de tamanho três representa a colaboração em pares dos autores em três trabalhos distintos ou uma colaboração conjunta dos mesmos autores em uma única publicação [Lerner and Hâncean 2023, Patania et al. 2017, Battiston et al. 2020, Antelmi 2021].

Hipergrafos são generalizações de grafos nas quais as *hiperarestas* são capazes de modelar relações poliádicas e, por isso, podem ser utilizados como um modelo mais apropriado para sistemas complexos que apresentam relações muitos-para-muitos. Assim, a generalização possibilitada pelas hiperarestas adiciona um grande poder de representação do ponto de vista de modelagem. Porém, essa mudança de perspectiva ao se representar as relações de alta-ordem demanda que toda a teoria de grafos, assim como algoritmos e ferramentas, tomados como base para a análise de redes complexas, sejam revisitados. Então, se por um lado a representação de redes através de hipergrafos adiciona complexidade e generalidade ao modelo, por outro lado, perde-se a possibilidade de utilização de uma série de algoritmos e ferramentas, tradicionalmente utilizados em redes representadas por grafos, que precisam ser revisados e podem ter sua aplicação inviabilizada em hipergrafos.

Este trabalho tem como objetivo trazer uma visão geral de conceitos centrais para explorar a utilização de hipergrafos como modelos para representação de redes sociais, discutindo vantagens e desvantagens, desafios e oportunidades em um contexto específico: coautoria de trabalhos científicos. Redes de coautoria apresentam muitas propriedades topológicas que podem ser observadas em redes sociais [Newman 2004] e as relações entre autores têm sido exploradas para melhor compreender muitos aspectos da dinâmica de publicações acadêmicas em contextos como física, química e ciência da computação [Newman 2000, Velden et al. 2009, Ullah et al. 2022, Vieira et al. 2024]. De maneira mais específica, o presente trabalho toma como base a CSBCSet [Filho et al. 2023], que apresenta um conjunto de dados de publicações do Congresso da Sociedade Brasileira de Computação (CSBC) no intervalo entre 2013 e 2022 que contempla 10 dos seus eventos.

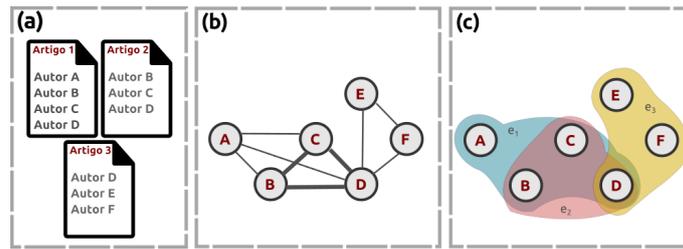


Figura 1. Exemplo da representação de autoria de artigos científicos (a) modelada como: um grafo clássico; (b) um hipergrafo (c).

Dentro desse cenário, o presente trabalho é guiado pelas seguintes questões de pesquisa: *QPI) Como as relações muitos-para-muitos representadas na CSBCSet podem ser caracterizadas sob o ponto de vista de hiper-redes? QPII) Qual o impacto do uso de modelos de hipergrafos para a compreensão de fenômenos relacionados à coautoria de artigos científicos em relação a modelos baseados em grafos clássicos?* Para que essas questões sejam respondidas, a representação de coautoria a partir da CSBCSet será analisada, de maneira comparativa, tomando como base redes modeladas pelas suas relações em pares (grafos) e redes modeladas pelas suas relações muitos-a-muitos (hipergrafos).

O restante deste trabalho está dividido da seguinte forma: a Seção 2 apresenta uma série de conceitos relacionados a hipergrafos, essenciais para o desenvolvimento e a compreensão do trabalho; a Seção 3 apresenta uma breve visão dos dados considerados e dos métodos utilizados durante o trabalho; a Seção 4 apresenta os experimentos conduzidos e discute seus resultados; por fim, as conclusões obtidas a partir do presente trabalho, assim como algumas direções de trabalhos futuros são apresentados na Seção 5.

2. Hipergrafos

Antes de apresentar algumas definições e propriedades relacionadas a hipergrafos, o que será feito ao longo desta seção, a Figura 1 apresenta um pequeno exemplo que associa três artigos científicos a seis autores distintos (Fig. 1(a)). A Fig. 1(b) apresenta a representação desse exemplo através de um grafo clássico, no qual cada vértice representa um autor e cada aresta representa a coautoria entre um par de autores. Já a Fig. 1(c) apresenta a representação do mesmo exemplo através de um hipergrafo, no qual cada vértice também representa um autor, mas cada hiperaresta representa a colaboração de todos os coautores de um artigo. É possível perceber que a representação baseada no grafo clássico (Fig. 1(a)) não permite distinguir, por exemplo, se o triângulo formado pelos vértices D, E e F indica um único artigo dos autores D, E e F, ou se indica a existência de três artigos dos três autores tomados em pares. Por outro lado, essa distinção é possível pelo modelo de hipergrafo, que representa cada relação, originada pela colaboração em um artigo científico como uma hiperaresta, ilustrada por uma região de uma cor. Nesse exemplo, a representação da colaboração de um conjunto de autores como uma hiperaresta contorna também a necessidade de associar pesos às arestas, o que no caso da Fig. 1(b) é ilustrado por larguras diferentes das arestas, evitando ambiguidades na interpretação.

O termo “*hiper-rede*” será utilizado para permitir a distinção entre sistemas modelados como hipergrafos e sistemas modelados como grafos, aos quais associa-se o termo “*rede*”. De forma direta, *hipergrafos* são generalizações de grafos cujas arestas podem relacionar mais de dois vértices. De maneira mais formal, um hipergrafo $\mathcal{H} = (\mathcal{V}, \mathcal{E})$

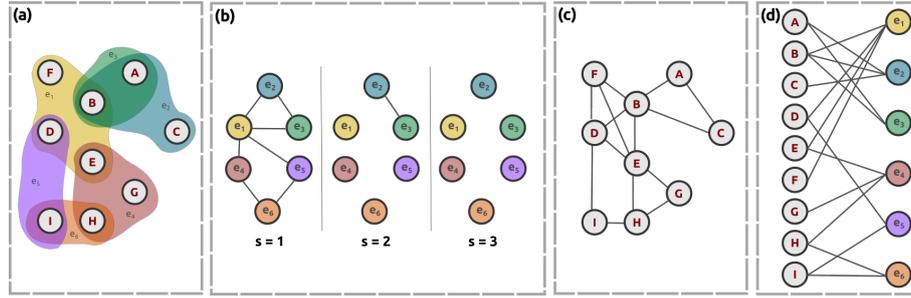


Figura 2. Projeções de um hipergrafo (a) em grafos s -linha (b), grafo clique (c), grafo bipartido (d).

corresponde a um conjunto de vértices $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ e uma família de hiperarestas $\mathcal{E} = \{e_1, e_2, \dots, e_m\}$ na qual cada elemento é um conjunto $e_i \subseteq \mathcal{V}$, $i = 1, \dots, m$.

Algumas definições aplicadas a hipergrafos são análogas às aplicadas a grafos. Algumas delas são apresentadas a seguir, tomando como base as definições propostas por Bretto [Bretto 2013]. Um hipergrafo \mathcal{H} tem ordem n e tamanho m , correspondentes às cardinalidades dos seus conjuntos \mathcal{V} e \mathcal{E} , respectivamente. Um hipergrafo é trivial se $\mathcal{V} \neq \emptyset$ e $\mathcal{E} = \emptyset$, e vazio se $\mathcal{V} = \emptyset$ e $\mathcal{E} = \emptyset$. Dois vértices $v_i \in \mathcal{V}$ e $v_j \in \mathcal{V}$ são adjacentes se existe uma hiperaresta $e_k \in \mathcal{E}$ tal que $v_i, v_j \in e_k$. Duas hiperarestas são adjacentes se sua interseção é não vazia e uma hiperaresta $e_k \in \mathcal{E}$ é incidente a um vértice $v_i \in \mathcal{V}$ se $v_i \in e_k$.

Um caminho P em um hipergrafo \mathcal{H} entre os vértices v_i e v_j é uma seqüência alternada entre vértices e hiperarestas da forma $v_i = v_1, e_1, v_2, e_2, \dots, v_s, e_s, v_{s+1} = v_j$ tal que v_1, v_2, \dots, v_{s+1} são vértices distintos e e_1, e_2, \dots, e_s são arestas distintas. s é o comprimento do caminho P . Assim como em grafos clássicos, a distância $l(v_i, v_j)$ entre os vértices v_i e v_j é o comprimento do menor caminho entre v_i e v_j .

É possível estabelecer métodos para a projeção de hipergrafos em grafos sob diferentes perspectivas. Em um *grafo clique* projetado a partir de um hipergrafo $\mathcal{H} = (\mathcal{V}, \mathcal{E})$, cada vértice v_i, v_j está ligado se $v_i, v_j \in e_k$, tal que $e_k \in \mathcal{E}$. Um *grafo bipartido* projetado a partir de \mathcal{H} possui dois conjuntos de vértices, um deles contendo elementos correspondentes aos elementos de \mathcal{V} e outro com elementos correspondentes aos elementos de \mathcal{E} . Uma aresta no grafo bipartido conecta os elementos v_i às suas relações e_k correspondentes. Um hipergrafo \mathcal{H} pode também ser projetado como um *grafo linha*, onde cada vértice e_k corresponde a uma hiperaresta de \mathcal{H} e um par de vértices e_k, e_l está conectado se suas hiperarestas correspondentes em \mathcal{E} compartilham ao menos um vértice. Como observado por Aksoy *et al.* ([Aksoy et al. 2020]), grafos linha são frequentemente utilizados como uma alternativa para a análise de grafos, desempenhando um papel central na ciência de hiper-redes. Aksoy *et al.*, então, apresentam a definição de grafos s -linha como uma generalização de grafos linha na qual cada par de vértices e_k, e_l está conectado se suas arestas correspondentes em \mathcal{E} compartilham ao menos s vértices. A Figura 2 apresenta um exemplo de projeção de um hipergrafo (Fig. 2(a)) em três tipos de grafos: grafos s -linha, para $s = 1$, $s = 2$ e $s = 3$ (Fig. 2(b)), grafo clique (Fig. 2(c)) e grafo bipartido (Fig. 2(d)). Embora a projeção de hipergrafos em grafos seja muito útil para a aplicação de um grande conjunto de técnicas clássicas utilizadas em análise de redes, descritas e aplicadas de maneira vasta na literatura, sua utilização resulta, inevitavelmente,

em modelos com uma perda de informação. A projeção de um hipergrafo em grafos linha deixa escapar as informações sobre a composição de suas hiperarestas, representando nelas apenas a existência de uma interseção de vértices. Ao se projetar um hipergrafo como grafo clique, toda hiper-relação é substituída por relações em pares, fazendo com que o modelo perca a natureza das relações de alta-ordem, assim como a noção de grupos de vértices representada no hipergrafo. Em uma projeção de hipergrafo em grafo bipartido, vértices originados a partir do conjunto $\mathcal{V} \in \mathcal{H}$ têm suas interações mediadas por vértices originados a partir do conjunto $\mathcal{E} \in \mathcal{H}$ e sua estrutura intrinsecamente bipartida faz com que todas as ferramentas de análise de redes, desde métricas, até algoritmos e modelos de difusão, precisem ser revistas para que sejam potencialmente aplicadas.

Assim, há um enorme desafio em generalizar técnicas de ciência de redes baseadas em grafos para um contexto de hiper-redes, em que as relações de alta-ordem sejam respeitadas. Askoy *et al.* [Aksoy et al. 2020] fazem uma revisão de definições relacionadas a hipergrafos e propõem uma série de conceitos baseados em passeios que estão na base de métodos para a definição de coeficientes e centralidades aplicados a hiper-redes que sejam análogos àqueles tradicionalmente aplicados na análise de redes.

Uma delas é a definição de *s-caminho*, que conecta um par de vértices v_i e v_j em um hipergrafo utilizando apenas hiperarestas que compartilham ao menos s vértices e permite definir a *s-distância* $s - l(v_i, v_j)$, que caracteriza o tamanho do menor *s-caminho* entre v_i e v_j em um hipergrafo. Dessa forma, as noções de centralidade de proximidade (*closeness centrality*) e centralidade de intermediação (*betweenness centrality*), tradicionalmente utilizadas em grafos clássicos, podem ser definidas para hipergrafos como *s-closeness centrality* e *s-betweenness centrality*, respectivamente¹.

A *s-closeness centrality* de um vértice v_i , mede sua proximidade topológica a outros vértices do hipergrafo utilizando *s-caminhos*. A *s-closeness centrality* normalizada C_{s-c} de um vértice v_i pode ser definida como

$$C_{s-c}(v_i) = \frac{n - 1}{\sum_{v_j \in \mathcal{V} \setminus \{v_i\}} s - l(v_i, v_j)}. \quad (1)$$

A *s-betweenness centrality* mede a capacidade de um vértice v_i em intermediar informação que está sendo propagada pelo hipergrafo utilizando *s-caminhos*. A *s-betweenness centrality* C_{s-b} de um vértice v_i pode ser definida como

$$C_{s-b}(v_i) = \frac{\sum_{v_k, v_j \in \mathcal{V} \setminus \{v_i\}, k \neq j} \sigma_{k,j}(v_i)}{\sigma_{k,j}}, \quad (2)$$

onde $\sigma_{k,j}(v_i)$ é o número de *s-caminhos* mínimos entre v_k e v_j que passam por v_i e $\sigma_{k,j}$ é o número total de *s-caminhos* mínimos entre v_k e v_j .

A definição de centralidade de grau (*s-degree centrality*) de um vértice v_i em um hipergrafo também pode ser facilmente estendida daquela aplicada a grafos e pode ser calculada proporcionalmente ao número de hiperarestas que possuem ao menos s vértices às quais o vértice v_i pertence. A *s-degree centrality* pode ser definida como

$$C_{s-d}(v_i) = \frac{k_i}{n}, \quad (3)$$

¹Neste trabalho, será utilizada a terminologia *s-closeness centrality* e *s-betweenness centrality* no idioma inglês, como no trabalho de Askoy *et al.* [Aksoy et al. 2020].

onde k_i é o grau do vértice v_i , o número de hiperarestas com ao menos s vértices aos quais v_i pertence.

Uma das tarefas mais importantes na ciência de redes é a identificação de comunidades que, de maneira consensual pode ser descrita como a busca por grupos de vértices com alta densidade interna quando comparada ao volume de ligações externas [Newman 2006]. A generalização das relações em pares para relações muitos-para-muitos faz com que problema de identificação de grupos, embora análogo, seja acrescido de uma enorme complexidade. Alguns autores buscaram enfrentar esse problema, principalmente em hipergrafos uniformes, i.e, hipergrafos cujas hiperarestas possuem um número uniforme de vértices [Zhou et al. 2006, Bulò and Pelillo 2009]. Em um esforço de identificar comunidades em hiper-redes, Kumar *et al.* ([Kumar et al. 2018]) definem um método baseado no tradicional método de Louvain [Blondel et al. 2008], frequentemente utilizado para a caracterização de comunidades em grafos e que visa resolver o problema de maximização da modularidade

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij}P_{ij})\delta(c_i, c_j), \quad (4)$$

onde A é a matriz de adjacência da rede, c_i representa a comunidade do vértice v_i e P representa o modelo-nulo que preserva a distribuição de graus na rede, ou seja, P_{ij} representa a expectativa de ligação entre v_i e v_j , preservando seus graus k_i e k_j , respectivamente. Kumar *et al.* apontam que uma estratégia simples para generalização do problema de identificação de comunidades em hiper-redes poderia ser a projeção do hipergrafo em um grafo clássico e a aplicação de algoritmos tradicionais, mas que essa abordagem faria com que informações essenciais das hiperarestas fossem perdidas. Tomando como exemplo o vértice **B** da rede ilustrada na Fig. 2(a), temos que $k_B = 3$, já que participa de três hiperarestas (e_1, e_2 e e_3). Porém, quando o hipergrafo é projetado no grafo clique (Fig. 2(c)), $k_B = 5$, já que, nesta projeção, suas relações com todos os vértices pertencentes a e_1, e_2 e e_3 são representadas. Assim, o método para identificação de comunidades baseado no método de Louvain definido por Kumar *et al.*, além de considerar o cálculo dos graus de maneira apropriada para hipergrafos em P , também adapta a matriz de adjacência A para o contexto de hipergrafos.

3. Materiais e métodos

3.1. Visão geral da base de dados

Os experimentos neste estudo são conduzidos tomando como base a CSBCSet [Filho et al. 2023], um conjunto de dados sobre publicações acadêmico-científicas nos eventos mais longevos realizados no Congresso da Sociedade Brasileira de Computação (CSBC) ². A CSBCSet apresenta metadados de todas as publicações disponíveis entre 2013 e 2022 de 97 edições de dez eventos: BraSNAM (*Brazilian Workshop on Social Network Analysis and Mining*); SemISH (Seminário Integrado de Software e Hardware); WIT (*Women in Information Technology*); WEI (Workshop sobre Educação em Computação); BreSCI (*Brazilian e-Science Workshop*); WCAMA (Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais); WPerformance

²<https://csbc.sbc.org.br/>

	# de nós	# de (hiper)-arestas
Rede	4960	11972
Hiper-rede	4960	1996
Hiper-rede $s = 1$	4940	11972
Hiper-rede $s = 2$	804	1174
Hiper-rede $s = 3$	266	296

Tabela 1. Características básicas das (Hiper-)redes utilizadas nos experimentos.

(WPerf. – Workshop em Desempenho de Sistemas Computacionais e de Comunicação); SBCUP (Simpósio Brasileiro de Computação Ubíqua e Pervasiva); CTIC (Concurso de Trabalhos de Iniciação Científica); e CTD (Concurso de Teses e Dissertações). Há na CSBCSet um total de 4961 autores, já desambiguados pelos na base de dados, e 1997 publicações distintas.

3.2. Modelos de (hiper-)redes

Os modelos de (hiper-)redes explorados neste trabalho visam representar as (hiper-)relações de colaboração entre os autores em artigos científicos publicados no CSBC, de acordo com os meta-dados disponibilizados na CSBCSet, como descrito na Subseção 3.1. Assim, nosso modelo é construído a partir de um conjunto de \mathcal{A} autores e um conjunto de suas \mathcal{P} publicações. Na base de dados, há a identificação do ano de publicação de cada trabalho que, neste trabalho, foi ignorada.

São construídas uma rede e uma hiper-rede, representadas, respectivamente, por um grafo $\mathcal{G} = (\mathcal{V}, \mathcal{L})$ e por um hipergrafo $\mathcal{H} = (\mathcal{V}, \mathcal{E})$. Cada autor é representado por um vértice em \mathcal{G} e \mathcal{H} . Cada publicação $p \in \mathcal{P}$ é representada por uma clique em \mathcal{G} , cujas arestas $l_k \in \mathcal{L}$ relacionam os pares de vértices (v_i, v_j) correspondentes a p e por uma hiperaresta $e_k \in \mathcal{E}$ que relaciona o conjunto de autores v_i correspondentes a p .

Considerando a definição de Aksoy *et al.* [Aksoy et al. 2020], são também construídas redes baseadas em grafos s -linha para $s = 1$, $s = 2$ e $s = 3$, que serão utilizados na investigação sobre os autores mais centrais, tomando como base as definições de *s-degree centrality*, *s-closeness centrality* e *s-betweenness centrality* apresentadas na Seção 2. Para o restante do trabalho, as projeções da hiper-rede em grafos s -linha, para $s = 1$, $s = 2$ e $s = 3$, serão chamadas, respectivamente, de *Hiper-rede $s = 1$* , *Hiper-rede $s = 2$* e *Hiper-rede $s = 3$* .

4. Experimentos e discussão

Considerando a metodologia proposta (Seção 3), esta seção apresenta os experimentos conduzidos com o objetivo de responder às questões de pesquisa (Seção 1): *QPI) Como as relações muitos-para-muitos representadas na CSBCSet podem ser caracterizadas sob o ponto de vista de hiper-redes? QPII) Qual o impacto do uso de modelos de hipergrafos para a compreensão de fenômenos relacionados à coautoria de artigos científicos em relação a modelos baseados em grafos clássicos?* Inicialmente, a Tabela 1 apresenta as características básicas das (Hiper-)redes.

A Figura 3 apresenta uma representação visual dos modelos de Rede (Figs. 3a e 3b) e Hiper-rede (Figs. 3c e 3d). É importante destacar que a Hiper-rede de coautorias resultou em uma visualização bastante poluída e pouco atrativa e, por isso,

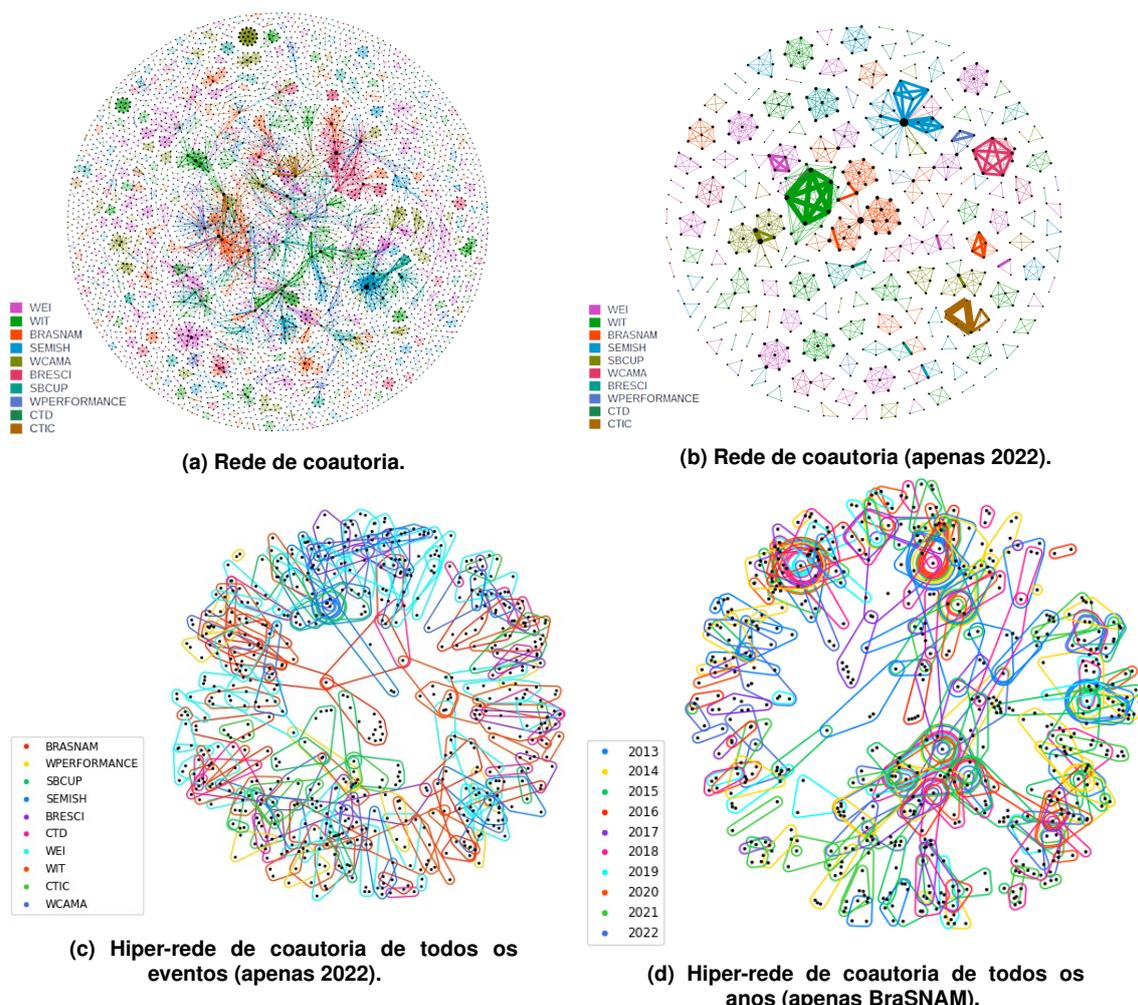


Figura 3. Representação visual da hiper-rede e da rede de coautoria da CSBCSet.

optou-se por aplicar: um filtro de todos os trabalho, mas apenas do ano de 2022 (Fig. 3c) e um filtro de todos os anos, mas apenas de trabalhos do BraSNAM (Figs. 3d). Primeiramente, ao observar a Figura 3 é importante notar que a representação visual das relações de coautorias na CSBCSet possui limitações impostas pelo seu número de vértices, tanto no modelo de Rede, como no modelo de Hiper-rede. Na Hiper-rede, especificamente, há uma limitação adicional dada pela ilustração das linhas que delimitam as hiper-relações (o que, inclusive, acabou por inviabilizar a visualização da Hiper-rede completa). Porém, nota-se que, de fato, como discutido na Seção 2, a representação através de hiper-redes permite algumas distinções na autoria que não são possíveis no modelo de redes. Por exemplo, há na Hiper-rede da Fig. 3c diversas hiper-arestas que referem-se a um único trabalho, mas que envolvem diversos autores, gerando cliques na Rede da Fig. 3b, que não permitem essa distinção, o que reforça a profundidade de riqueza na representação trazida pelos hiper-grafos à ciência de redes.

Continuando a análise das relações de coautorias de trabalhos da CSBCSet, foi realizado um estudo sobre os autores mais centrais, que podem representar pessoas com maior importância dentro do contexto explorado. Para comparar os autores mais centrais identificados nos modelos de (Hiper-)redes estudados, foram calculados os *ranks* de

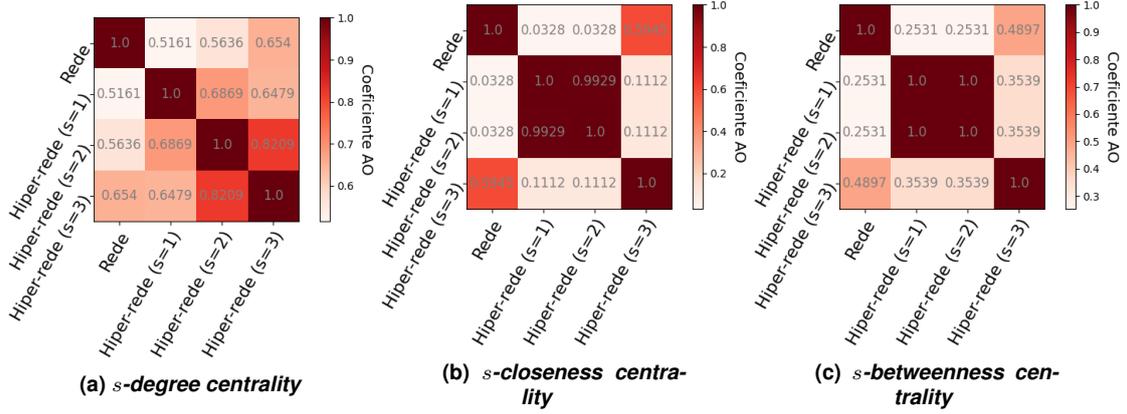


Figura 4. Coeficientes AO dos *ranks* obtidos em cada modelo de (Hiper-)rede para cada centralidade estudada.

autores considerando as centralidades estudadas para os modelos de Grafo, Hipergrafo ($s = 1$), Hipergrafo ($s = 2$) e Hipergrafo ($s = 3$) descritos na Subseção 3.2. Após definir um *rank* dos autores mais centrais, concentramos a análise nas *top-k* posições, com $k = 100$, de forma a minimizar o impacto de divergências que posições inferiores (e menos interessantes) pudessem trazer ao resultado. Assim, é necessário correlacionar *ranks* com elementos potencialmente diferentes e, para isso, utilizamos a métrica *Average Overlap (AO)* [Webber et al. 2010], que mede a similaridade entre *ranks* de itens possivelmente diferentes. Dados dois *ranks* S e T de tamanhos iguais, o coeficiente AO é definido como

$$AO(S, T, k) = \frac{1}{k} \sum_{d=1}^k A(S, T, d), \quad (5)$$

onde k é a profundidade da avaliação. O cálculo de $A(S, T, d)$, que quantifica a sobreposição S e T nas primeiras d posições é dado por

$$A(S, T, d) = \frac{\|S_{1:d} \cap T_{1:d}\|}{d}. \quad (6)$$

A Figura 4 apresenta o resultado dos coeficientes AO, em forma de mapa de calor, comparando os *ranks* obtidos em cada modelo de (Hiper-)rede para a *s-degree centrality* (Fig. 4a), *s-closeness centrality* (Fig. 4b) e *s-betweenness centrality* (Fig. 4c).

Padrões bastante distintos de correlação para a *s-degree centrality* podem ser observados na Figura 4, quando comparada às *s-closeness centrality* e *s-betweenness centrality*. Para a *s-degree centrality* (Fig. 4a), há uma correlação apenas moderada entre os *ranks* gerados para a Rede e a Hiper-rede ($s = 1$), o que está ligado à diferença na própria definição de grau para redes e hiper-redes. Enquanto o grau em uma rede de coautoria está relacionado à quantidade de coautores aos quais um indivíduo está relacionado, em uma hiper-rede de coautoria o grau está relacionado à quantidade de trabalhos dos quais ela participa. À medida em que o valor s aumenta para a construção do modelo de Hiper-rede, autores que participam em artigos com menos coautores são desconsiderados, e o *rank* prioriza, em suas posições mais altas, autores que colaboram em trabalhos com mais colaboradores. Dessa forma, aumenta também a correlação entre os *ranks* gerados para a Rede e as Hiper-redes ($s = 2$) e ($s = 3$). Nos *ranks* gerados para *s-closeness centrality* e *s-betweenness centrality*, observa-se que há uma baixa correlação entre os *ranks*

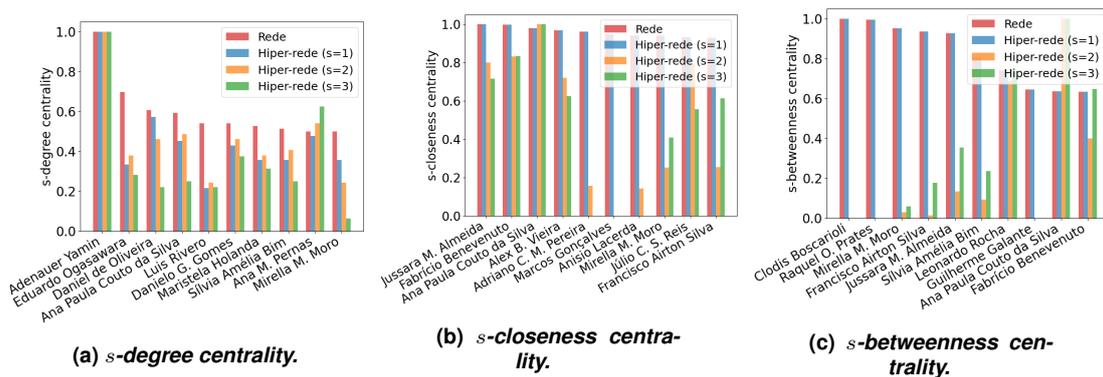


Figura 5. Comparação de valores de centralidade das dez pessoas autoras mais centrais tomando como referência o modelo de Rede.

dos autores das Redes e Hiper-redes ($s = 1$) e ($s = 2$) e apenas a Hiper-rede ($s = 3$) apresenta uma correlação significativa com a Rede. Novamente, ao desconsiderar autores com participação em trabalhos com poucos autores, o modelo de Hiper-rede ($s = 3$) coloca como mais centrais as pessoas com um maior volume de autorias, assim como o modelo de Rede. O resultado revelado pela Fig. 4 indica que o modelo de grafo clique, tradicionalmente utilizado para representar redes de coautoria, ao priorizar como autores mais centrais aqueles que possuem um volume maior de colaboração, pode estar subestimando o potencial de autores com um menor número de colaborações, mas que mantêm colaborações mais consistentes em servir como ponte intermediadora entre relações.

Com o objetivo de aprofundar o estudo sobre os autores mais centrais encontrados em cada um dos modelos de (Hiper-)redes, foi realizada uma investigação mais próxima dos nomes presentes nas primeiras posições dos *ranks*. O resultado desse experimento é apresentado na Figura 5, onde cada uma das subfiguras (Fig. 5a, 5b e 5c) apresenta a comparação para cada uma das centralidades estudadas. Para isso, o modelo de rede foi utilizado como referência e suas dez pessoas autoras mais centrais foram encontradas. Em cada uma das subfiguras da Fig. 5, o eixo- x apresenta os nomes das pessoas nas dez primeiras posições do respectivo *rank* no modelo de Rede. O eixo- y apresenta a medida de centralidade observada para essas pessoas no modelo de Rede, mas também a medida observada no modelos de Hiper-redes $s = 1$, $s = 2$ e $s = 3$, possibilitando, assim, uma avaliação da estabilidade dos *ranks* de centralidade nos diferentes modelos. É importante destacar que o valor de centralidade foi normalizado entre 0 e 1 para cada medida em cada modelo de (Hiper-)rede. Primeiramente, nota-se que há pouca concordância nas primeiras posições dos *ranks* encontrados para o modelo de Rede e os modelos de Hiper-redes, o que pode ser visto pelo fato que há um comportamento monotônico para as barras referentes ao modelo de Rede – naturalmente, já que é esse o modelo usado como referência –, mas que não se repete para os modelos de Hiper-redes. Uma notória exceção ocorre na primeira posição observada para a *s-degree centrality* (Fig. 5a), onde há um alinhamento na primeira posição dos *ranks* para todos os modelos. Para a *s-closeness centrality* (Fig. 5b) e a *s-betweenness centrality* (Fig. 5c) nota-se que há algumas pessoas para as quais nem há valor de centralidade associado nos modelos de Hiper-redes, indicando que essas foram desconsideradas por esses modelos. Esse fato é ainda mais evidente nas primeiras posições do gráfico da Fig. 5c, o que mostra que pessoas que colaboram em

artigos com poucos coautores têm papel de destaque como intermediadores de relações no modelo representado por Redes, corroborando o resultado apresentado pela Figura 4.

Os resultados das Figuras 4 e 5 evidenciam a importância da modelagem de coautorias como Hiper-redes como ferramenta para análise do fenômeno de colaboração em artigos científicos sob uma ótica alternativa àquela fornecida por modelos de Redes para a investigação local. Uma investigação da organização topológica das (Hiper-)redes foi também realizada para que as coautorias na CSBCSet pudessem ser exploradas sob uma perspectiva global. Tomando como base o método de Louvain [Blondel et al. 2008] e o método de Kumar *et al.* [Kumar et al. 2018], como descrito na Seção 2, foram identificadas as comunidades considerando os modelos de Redes e Hiper-redes, respectivamente. Para a aplicação do algoritmo de Louvain, foi utilizada a implementação da biblioteca CDLib³, já o algoritmo de Kumar *et al.* foi aplicado pela implementação disponibilizada pela biblioteca HiperNetX. A modularidade da partição obtida pelo método de Louvain para o modelo de Rede foi de $Q_G = 0.96$, enquanto um valor $Q_H = 0.97$ foi obtido para a partição encontrada pelo método de Kumar *et al.* Esses resultados mostram que, quando são consideradas as coautorias de trabalhos no CSBC no período investigado, há uma clara divisão topológica da (Hiper-)rede de autores, implicando em um alto valor de modularidade para ambos modelos considerados.

A extensão da sobreposição da partição obtida nos modelos de (Hiper-)redes, foi avaliada através do coeficiente *Overlapping Normalized Mutual Information (ONMI)* [McDaid et al. 2011], que retorna um valor 0 quando não há informação mútua entre as partições e 1 quando há uma perfeita correlação. Para isso, foi aplicada a implementação fornecida pela biblioteca CDLib. Quando todas as comunidades são consideradas, obtém-se um valor de $ONMI = 0.91$, indicando uma grande sobreposição. Com o objetivo de verificar se o coeficiente ONMI está dominado por comunidades de tamanhos muito pequenos, um filtro foi aplicado para que apenas comunidades com mais que dez vértices fossem avaliadas. Após a aplicação desse filtro, foi obtido um valor de $ONMI = 0.83$, indicando que há, sim, uma sobreposição concentrada nas comunidades menores, mas, mesmo sem considerá-las, a sobreposição entre as partições é bastante significativa.

A comparação entre a organização das comunidades pode ser melhor compreendida através da Figura 6, que apresenta a Função de Distribuição Cumulativa Complementar dos tamanhos das comunidades da Rede e da Hiper-rede de coautoria, na qual, o eixo- x representa os tamanhos das comunidades e o eixo- y representa a probabilidade de que uma comunidade seja maior que o valor no eixo- x . Nessa figura, é possível perceber que as distribuições de comunidades de um determinado tamanho são coincidentes até o tamanho 11 para a Rede e a Hiper-rede. A partir desse ponto, percebe-se uma distinção das curvas, mesmo que pequena. Nota-se, ainda, a existência de comunidades maiores na Rede de coautoria, mesmo que em probabilidade muito baixa.

5. Conclusão e direções futuras

Este trabalho investiga o uso de hiper-redes para a modelagem de relações sociais de alta-ordem, i.e., relações que envolvem um conjunto maior que dois indivíduos, de forma a generalizar relações entre pares de indivíduos, representadas por redes tradicionais. Após apresentar uma série de definições e ferramentas computacionais para a representação e

³<https://cdlib.readthedocs.io/>

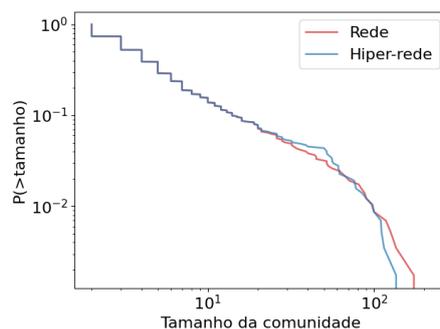


Figura 6. Função de Distribuição Cumulativa Complementar dos tamanhos das comunidades identificadas para a rede (vermelho) e hiper-rede (azul).

análise de hiper-grafos como modelo matemático de hiper-redes, é realizado um estudo em uma hiper-rede de coautoria construída a partir da CSBCSet, uma base de dados que contém todas as publicações no Congresso da Sociedade Brasileira da Computação em um intervalo de dez anos.

Uma série de experimentos são realizados com o objetivo de responder a duas questões de pesquisa: *QPI) Como as relações muitos-para-muitos representadas na CSBCSet podem ser caracterizadas sob o ponto de vista de hiper-redes?* e *QPII) Qual o impacto do uso de modelos de hipergrafos para a compreensão de fenômenos relacionados à coautoria de artigos científicos em relação a modelos baseados em grafos clássicos?*. Em relação à QPI, pode-se notar que ainda há uma certa limitação nas ferramentas computacionais para modelagem e análise de hiper-redes quando compara-se com aquelas voltadas a redes tradicionais. Porém, a análise de hiper-redes, apesar de não ser uma área de pesquisa nova, tem recebido, apenas recentemente, uma atenção mais dedicada, ao contrário da análise de redes tradicionais, que já encontra uma área bastante consolidada. Mesmo assim, há uma grande quantidade de métodos relatados na literatura para investigação de hiper-relações, tanto de maneira local, como de maneira global, o que permite que se tire proveito da riqueza na representação de relações de alta-ordem trazida pelas hiper-redes, mas fazendo análises que sejam coerentes com aquelas realizadas com ferramentas oriundas da ciência de redes. Assim, é possível avançar também na resposta à QPII, verificando que a aplicação de modelos de hipergrafos para a compreensão da coautoria de artigos científicos pode revelar aspectos importantes sobre, não apenas o número de interações realizadas por cada indivíduo, mas a maneira como se dão essas interações. Assim, pode-se compreender com mais clareza o papel de indivíduos com diferentes padrões de colaboração na intermediação de relações e, conseqüentemente, na difusão de ideias e conhecimento através da hiper-rede.

Assim, o presente trabalho pode apontar direções importantes a serem exploradas em futuros estudos. Nesse sentido, é importante destacar que a área ainda é incipiente e oferece um grande potencial de pesquisa, principalmente considerando a baixa quantidade de trabalhos sobre hiper-redes em língua portuguesa (nenhum trabalho em língua portuguesa foi encontrado em assunto relacionado ao aqui explorado, por exemplo). Como possíveis direções futuras, então, pode-se buscar a verificação da generalidade das observações aqui realizadas em outras bases de dados de interações sociais, inclusive de colaboração e, mais especificamente, de coautoria, mas em outros contextos. Além

disso, diversos outros métodos de análise de hiper-redes podem ser empregados, como forma de confrontar os resultados aqui obtidos sob outras perspectivas.

Referências

- Aksoy, S., Joslyn, C., Ortiz Marrero, C., Praggastis, B., and Purvine, E. (2020). Hypernetwork science via high-order hypergraph walks. *EPJ Data Science*, 9.
- Antelmi, A. (2021). *Beyond Pairwise Relationships: Modeling Real-world Dynamics Via High-order Networks*. Phd thesis, Università degli Studi di Salerno, Salerno, Italy.
- Barabási, A.-L. and Pósfai, M. (2016). *Network science*. Cambridge University Press, Cambridge.
- Barrat, A., Barthélemy, M., Pastor-Satorras, R., and Vespignani, A. (2004). The architecture of complex weighted networks. *PNAS*, 101(11):3747–3752.
- Barthélemy, M., Barrat, A., Pastor-Satorras, R., and Vespignani, A. (2005). Characterization and modeling of weighted networks. *Physica A: Statistical Mechanics and its Applications*, 346(1):34–43.
- Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.-G., and Petri, G. (2020). Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 874:1–92.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.
- Boccaletti, S., Bianconi, G., Criado, R., del Genio, C., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., and Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122.
- Bretto, A. (2013). *Hypergraph Theory: An Introduction*. Springer Publishing Company.
- Bulò, S. and Pelillo, M. (2009). A game-theoretic approach to hypergraph clustering. In Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., and Culotta, A., editors, *Advances in Neural Information Processing Systems*, volume 22. Curran.
- Filho, S. L., Carvalho, L., Suzano, J., Brandão, M., Oliveira, J., and Santoro, F. (2023). Csbcset: Um conjunto de dados para uma década de csbc, seus eventos e publicações. In *Anais do XII Brazilian Workshop on Social Network Analysis and Mining*, pages 240–245, Porto Alegre, RS, Brasil. SBC.
- Holme, P. and Saramäki, J. (2012). Temporal networks. *Physics Reports*, 519(3):97–125. Temporal Networks.
- Joslyn, C. A., Aksoy, S. G., Callahan, T. J., Hunter, L. E., Jefferson, B., Praggastis, B., Purvine, E., and Tripodi, I. J. (2021). Hypernetwork science: From multidimensional networks to computational topology. In *Unifying Themes in Complex Systems X*, pages 377–392, Cham. Springer International Publishing.
- Kivelä, M., Arenas, A., Barthélemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, 2(3):203–271.

- Kumar, T., Vaidyanathan, S., Ananthapadmanabhan, H., Parthasarathy, S., and Ravindran, B. (2018). Hypergraph clustering: A modularity maximization approach.
- Lerner, J. and Hâncean, M.-G. (2023). Micro-level network dynamics of scientific collaboration and impact: Relational hyperevent models for the analysis of coauthor networks. *Network Science*, 11(1):5–35.
- McDaid, A., Greene, D., and Hurley, N. (2011). Normalized mutual information to evaluate overlapping community finding algorithms. *CoRR*.
- Newman, M. (2000). Who is the best connected scientist? A study of scientific co-authorship networks. *Santa Fe Institute, Working Papers*, 650.
- Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(suppl_1):5200–5205.
- Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582.
- Patania, A., Petri, G., and Vaccarino, F. (2017). The shape of collaborations. *EPJ Data Science*, 6:18.
- Ullah, M., Shahid, A., Din, I., A., M., Assam, M., Fayaz, M., Ghadi, Y., and Aljuaid, H. (2022). Analyzing interdisciplinary research using co-authorship networks. *Complexity*.
- Velden, T., Haque, A.-u., and Lagoze, C. (2009). A new approach to analyzing patterns of collaboration in co-authorship networks - mesoscopic analysis and interpretation. *Scientometrics*, 85.
- Vieira, V. d. F., Ferreira, C. H. G., Almeida, J. M., Moreira, E., Laender, A. H. F., Meira, W., and Gonçalves, M. A. (2024). A network-driven study of hyperprolific authors in computer science. *Scientometrics*.
- Webber, W., Moffat, A., and Zobel, J. (2010). A similarity measure for indefinite rankings. *ACM Trans. Inf. Syst.*, 28(4).
- Zhou, D., Huang, J., and Schölkopf, B. (2006). Learning with hypergraphs: Clustering, classification, and embedding. In Schölkopf, B., Platt, J., and Hoffman, T., editors, *Advances in Neural Information Processing Systems*, volume 19. MIT Press.