

Caracterização de Comunidades de I.A. no Brasil Utilizando Análise Bibliométrica e Descoberta de Subgrupos

1st Gabriel Fadoul

Departamento de Ciência da Computação
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil
gabriel.fadoul@dcc.ufmg.br

Orient. Renato Vimieiro

Departamento de Ciência da Computação
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil
rvimieiro@dcc.ufmg.br

Abstract—Redes Científicas são o centro de desenvolvimento do conhecimento humano. É através da pesquisa aprofundada e colaborativa que surgem novas possibilidades de expansão da fronteira das ciências. Pesquisadores da mesma subárea tendem a formar comunidades dentro de tais redes. Porém, a atividade de caracterizar tais comunidades não é tão simples. Esse projeto se propõe a aplicar métodos de *Subgroup Discovery*, especificamente a técnica de *Local Subgroup Discovery* em redes de coautoria e utilizar dados bibliométricos dos autores para analisar e caracterizar tais comunidades, com foco específico nos autores que publicaram na Conferência Brasileira de Sistemas Inteligentes (BRACIS). Foi possível encontrar comunidades de autores por regiões geográficas e por instituições que possuem maior relevância em seus artigos publicados, além de também identificar autores dentro dessas comunidades que possuem um desempenho abaixo do esperado, permitindo realizar uma análise qualitativa dos seus resultados.

Index Terms—Group Profiling, Subgroup Discovery, Local Subgroup Discovery, Co-author Network

I. INTRODUÇÃO

Redes de pesquisa científicas são o cerne de desenvolvimento do conhecimento. São nessas redes em que novas pesquisas são realizadas, subáreas do conhecimento se interceptam e a fronteira do conhecimento humano é expandida. Especialmente, a área de pesquisa acadêmica que têm se desenvolvido mais rapidamente é a de Inteligência Artificial (I.A.). Tais avanços resultam em incontáveis impactos sociais e econômicos, a tal ponto que há uma relação próxima entre o nível de pesquisa do país nessa área e o desenvolvimento econômico do mesmo [7]. No Brasil, apesar da adoção da I.A. no ambiente de mercado estar em par com os países mais desenvolvidos, o campo de pesquisa científica ainda tem uma baixa produção em comparação com esses mesmos países [8]

Decorre daí, que a compreensão do funcionamento dessas redes, da relação entre os atores, é de alta estima, tendo em vista que uma análise de tais conexões pode trazer benefícios de maior produtividade acadêmica para os membros das mesmas. Já existem estudos que buscam realizar essa análise através do processo de descoberta de comunidades científicas. [1], [2]

Mesmo com estudos que tratam exatamente dessa tarefa de análise das estruturas dessas redes, ainda há desafios concernentes à descoberta de comunidades e, mais especificamente, maneiras de caracterizá-las. [3] O processo de caracterização de comunidades necessita que seja possível extrair descrições de cada comunidade com base em atributos multidimensionais que são extraídos do processamento dos textos autores. Para tal desafio, surge a oportunidade da contribuição de outras áreas da computação para auxiliar nesse processo.

Nesse projeto, iremos utilizar a técnica de Mineração de Dados Supervisionada conhecida como Subgroup Discovery (SD), que permite com que possamos identificar descrições interessantes com base em uma variável alvo específica dentro de um conjunto de dados [4], [5]. A técnica de SD que será implementada terá como entrada um conjunto de dados de autores e suas características particulares estruturado como uma Rede de Coautoria. Foi escolhido especialmente autores que publicaram na Conferência Brasileira de Sistemas Inteligentes (BRACIS), principal conferência de I.A. no Brasil. O escopo dessa etapa do projeto será especificamente o processo de construção desta Rede.

O projeto está organizado da seguinte maneira. Na segunda seção trataremos dos conceitos básicos e os trabalhos relacionados. A terceira seção detalha a metodologia aplicada para o tratamento dos dados e os resultados obtidos com os atributos da Rede de Coautoria, por fim a seção 5 conclui o projeto com as possíveis próximas etapas.

II. REFERENCIAL TEÓRICO E TRABALHOS CORRELATOS

A. Comunidades Científicas e Redes de Coautoria

A investigação de estruturas de redes de pesquisa científicas já é um tópico bem abordado na literatura, com inúmeras técnicas de estruturá-las. [6] Em muitas das investigações, uma das melhores maneiras de estruturar elas é através de uma *Rede de Coautoria*. Nessa Rede, estão representados os autores juntamente com as relações que possuem entre si. Essa relação, neste tipo de Rede, se dá quando há ao menos uma atividade de pesquisa realizada entre os autores.

Formalmente, uma *Rede de Coautoria* pode ser definida como um grafo $G = (V, E)$, com vértices $V = \{v_1, \dots, v_n\}$ e arestas $E \subseteq V \times V$, tal que:

$$\forall v \in V, \forall u \in V (u \neq v \rightarrow ((v, u) \leftrightarrow (u, v))) \quad (1)$$

$$\forall v \in V, (v, v) \notin E \quad (2)$$

Ou seja, é um grafo não direcionado (1) e sem arestas entre um vértice e ele mesmo (2).

Ao estruturar a Rede, é comum haver agrupamento de indivíduos que atuam entre si, e também possíveis relações com indivíduos de outros grupos não relacionados a princípio. Esses agrupamentos são chamados de *comunidades científicas*, onde é possível investigar sua formação e estrutura.

B. Group Profiling

Para compreender essas comunidades, é preciso encontrar uma maneira de descrevê-las, associando à cada uma um conjunto de atributos, chamados perfis descritivos. A tarefa de gerar e encontrar estes perfis descritivos, ou *group profiling*, é um problema que pode ser definido formalmente a partir da seguinte configuração. Dado uma rede representada como um grafo $G = (V, E)$ com vértices $V = \{v_1, \dots, v_n\}$ e arestas $E \subseteq V \times V$. Cada vértice é associado a um vetor de atributos de dimensão d , $\mathbf{a} \in A^d$, $\mathbf{a} = (a_1, \dots, a_d)$, onde $\text{dom}(a_j)$ é definido como o domínio dos atributos: $\{0, 1\}$. Já as comunidades podem ser definidas como um subgrafo $P_i = (V_{P_i}, E_{P_i})$, tal que $V_{P_i} \subseteq V$, $E_{P_i} \subseteq V_{P_i} \times V_{P_i}$, $E_{P_i} \subseteq E$.

A finalidade deste problema é selecionar os melhores k atributos que descrevem cada comunidade. É necessário definir uma medida de qualidade $f(a_j, P_i)$ para poder associar uma importância a cada atributo, e poder recuperar os k melhores.

C. Subgroup Discovery

Subgroup Discovery (SD) é uma atividade de mineração de dados que tem como objetivo identificar subgrupos em uma base de dados, onde a presença de uma certa variável alvo é destoante dos demais grupos. Cada subgrupo pode ser definido a partir de uma linguagem de descrição consistindo de pares atributo-valor na forma normal conjuntiva ou disjuntiva. A SD é uma técnica de extração de padrões que se encontra como meio termo entre indução descritiva e preditiva, seu objetivo é gerar, de maneira interpretável, subgrupos que descrevam relações entre as variáveis independentes e um certo valor de variável alvo.

De maneira formal, uma base de dados pode ser definida como $D = (I, A)$, sendo I um conjunto de indivíduos e A um conjunto de atributos. Para atributos nominais, um seletor ($a_i = v_j$) é uma função Booleana $I \rightarrow \{0, 1\}$ que assume o valor de verdade *verdadeiro* se o valor do atributo $a_i \in A$ for igual à v_j . Já para atributos numéricos a_{num} , os seletores ($a_{num} \in [\min_j; \max_j]$) podem ser definidos de maneira análoga para cada intervalo $[\min_j; \max_j]$ no domínio de a_{num} . A função Booleana assume valor de verdade *verdadeiro* se o valor do atributo a_{num} está no respectivo intervalo. Existem vários algoritmos de SD que aplicam inúmeras

técnicas para essa descoberta, porém neste trabalho iremos utilizar uma técnica SD Local, que está sendo explorada recentemente.

D. Local Subgroup Discovery

A técnica de Local Subgroup Discovery (LSD) [14] funciona como uma análise da rede, combinando tanto os atributos dos nós quanto as informações estruturais para identificar excepcionalidades em subgrupos locais. O interesse nessa técnica é encontrar subgrupos que possuem divergência não com a população global, mas com um grupo de referência que possui indivíduos com características similares. Isso possibilita a descoberta de subgrupos que possuem traços distintos do comportamento esperado do seu grupo de referência.

Para formalizar o funcionamento do algoritmo, precisamos acrescentar algumas informações à rede. Para cada nó $v \in V$ da Rede de Coautoria, estão associados os atributos $a(v) = \{a_1(v), \dots, a_h(v)\}$ juntamente com uma variável objetivo binária $t(v) \in \{0, 1\}$. Um protótipo é qualquer nó $x \in V$ que serve como um centro para definir um subgrupo S . A partir do protótipo, são definidos dois subgrupos R_ρ e S_σ , onde $\rho, \sigma \in \mathbb{N}$ e $\rho > \sigma$. O subgrupo R_ρ , chamado de *subgrupo referência*, consiste nos ρ vizinhos mais próximos a partir do protótipo de acordo com uma métrica de distância d . Analogamente, o subgrupo S_σ , chamado de *subgrupo local*, consiste nos σ vizinhos mais próximos a partir do protótipo de acordo com uma métrica de distância d . Por construção, $S_\sigma \subset R_\rho$. O objetivo é encontrar subgrupos S_σ e R_ρ tais que a distribuição da variável alvo t dentro destes subgrupos revele um padrão local.

Na proposta original da técnica, havia a possibilidade de utilizar o algoritmo de Dijkstra como a maneira de medir os nós mais próximo, mas no caso deste projeto o grafo que mapeia a rede de coautoria não é ponderado, restando sempre aplicar a Distância de Gower. Como esse método é possível avaliar a similaridade entre o protótipo x e qualquer nó v utilizando o conjunto de atributos $a(v)$ associado a cada um deles. A Distância de Gower pode ser definida como:

$$d(u, v) = \frac{1}{h} \sum_{k=1}^h \delta_k(u, v) \quad (3)$$

onde $u, v \in V$ são nós quaisquer, h é o chamado atributo de dimensionalidade e $\delta_k(u, v)$ é a distância parcial entre os nós u e v com base no atributo k . Para o nosso caso de atributos numéricos, essa distância parcial é definida como:

$$\delta_k(u, v) = \frac{|a_k(u) - a_k(v)|}{\max_{x \in V}(a_k(x)) - \min_{x \in V}(a_k(x))} \quad (4)$$

E. Função de Qualidade

Quanto à função de qualidade adotada para classificar os subgrupos encontrados, foi adotada a divergência de Kullback-Leibler Ponderada (Weighted Kullback-Leibler [WKL]). A WKL é uma adaptação da divergência de Kullback-Leibler utilizada no campo da teoria da informação que mede como uma distribuição de probabilidade diverge de uma segunda,

porém agora ponderando pelo tamanho do subgrupo. Especificamente, essa função quantifica o quanto a presença da variável alvo difere entre dois subgrupos:

$$q(S, R) = \frac{|S|}{|R|} \sum_{y \in \{0,1\}} P_S(y) \log\left(\frac{P_S(y)}{P_R(y)}\right) \quad (5)$$

onde $P_A(y) = \mathbb{P}(t(x) = y | x \in A)$

De maneira prática, quanto maior o valor de $q(S, R)$, maior é a diferença entre as distribuições, o que indica um comportamento discrepante dos seus vizinhos. No artigo original do LSD [14], os autores relatam que empiricamente valores entre 0.20 e 0.40 representam subgrupos locais excepcionais e de alta relevância, o que nos capacita a uma correta interpretação dos nossos resultados.

F. Trabalhos Correlatos

O problema de *Group Profiling* utilizando a abordagem de SD já foi utilizado previamente na literatura, como no Trabalho Acadêmico de Dias [11] que se utiliza do algoritmo SSDP+, que é um algoritmo criado propriamente para esse problema [12], porém, em seu trabalho, ele utiliza uma base de dados mais abrangente do arXiv.

Há também aplicações da mesma técnica, mas dentro do contexto de redes sociais [13], no qual é gerado um grafo de interações entre usuários e buscam-se comunidades formadas dentro das redes.

III. METODOLOGIA

Na primeira etapa do projeto, houve um enfoque na captura, tratamento e estruturação dos dados em uma *Rede Coautoria* para a ingestão dos algoritmos de SD. Já na segunda etapa, o foco foi no enriquecimento desses dados. A Figura 1 apresenta o esquema da metodologia proposta. Na etapa 1, iniciamos o processo coletando os dados dos artigos publicados através da API da *dblp computer science bibliography* DBLP, com suas respectivas informações de autores, ano de publicação etc. Na etapa 2, são filtradas as colunas que de fato serão utilizadas no processo de estruturação da Rede. Na etapa 3, é feito o preenchimento dos dados vazios, especificamente, vários artigos da base de dados recuperados pela API, não possuem a informação do nome dos autores, então utilizamos o DOI de cada artigo na API Crossref, para recuperar os nomes dos autores dos artigos.

Dado que há uma diversidade de maneiras em que o mesmo autor pode ser nomeado, surge a necessidade de realizar uma desambiguação nessa nomenclatura. Na etapa 4, é realizada essa desambiguação através da tentativa de recuperação de um identificador único, para isso utilizamos o *Open Researcher & Contributor ID* (ORCID) que é um identificador que foi criado justamente para esse propósito de resolver o problema da ambiguidade entre o nome de autores [9]. Foi utilizada uma combinação entre a API Crossref e a API da própria plataforma do ORCID, para capturar o ORCID de cada autor, porém, como é uma plataforma de registro voluntário, é possível que tenham autores que não possuam um ORCID.

Na etapa 5, buscamos agrupar autores que possuem nomes similares para tentar contornar os autores que não possuem ORCID, ou até mesmo o mesmo autor, porém, em uma instância específica foi encontrado o ORCID e em outra não foi. A métrica de similaridade utilizada foi a distância de Levenshtein [10] entre o nome dos autores. Definimos uma similaridade mínima de 90 para agruparmos os nomes dos autores. Porém, como alguns autores podem ter nomes muito similares, tivemos que fazer as outras desambiguações manualmente. Por fim, na etapa 6, utilizamos os autores já desambiguados e todas as publicações para montarmos a *Rede de Coautoria*, onde cada vértice representa um autor, com atributos como os artigos publicados por esse autor juntamente com o resumo de cada artigo. As arestas estão definidas sob os autores que tiveram ao menos uma colaboração em artigos.

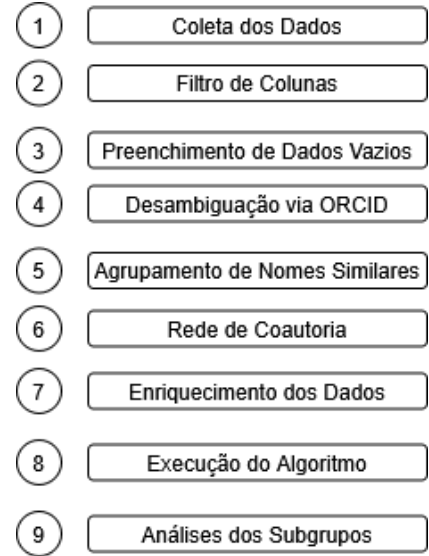


Fig. 1. Esquema da metodologia para montagem da *Rede de Coautoria*

No segundo momento deste projeto, tivemos a etapa 7,8 e 9. Na 7, temos o enriquecimento dos dados de cada autor. Foi necessário capturar novos dados que caracterizassem cada um dos autores com informações como a quantidade de artigos publicados, o H-Index desse autor, as instituições nas quais ele já esteve vinculado e também os tópicos que seus artigos abordam. Utilizamos a OpenAlex API para esses dados mais gerais de cada autor. Essas informações possibilitam caracterizar os nós para a utilização do algoritmo LSD, em que utilizamos esses tópicos como o conjunto de atributos para o cálculo da distância entre os nós. Foi feita uma pivotização dos tópicos e criado um vetor de atributos, em que cada coluna corresponde à um tópico específico, e seus valores são a razão entre a quantidade de artigos publicados no respectivo tópico e a quantidade total de artigos publicados. Com isso, temos para cada autor, a proporção de artigos publicados por tópico. Além disso, foi necessário também definir a variável alvo de maneira empírica, ou seja, definimos quais as possíveis métricas que retornariam subgrupos interessantes.

Tendo essas informações, adentramos na etapa 8, onde

foi possível executar o algoritmo de LSD, que gerava os subgrupos referência e os subgrupos locais, e os ordenava segundo a função de qualidade supracitada. Por fim, na etapa 9, buscamos fazer uma análise qualitativa (com algumas poucas quantitativas), buscando compreender qual o padrão formado pelos subgrupos encontrados, com um foco específico nas instituições dos autores (buscando um padrão geográfico) ou métricas derivadas da variável alvo.

IV. APLICAÇÃO DA METODOLOGIA PROPOSTA

O corpus utilizado foram os dados do BRACIS através da API DBLP, que nos trazem todos os artigos publicados na conferência desde o período de 2013 até 2025. Na etapa 1 e 2, foram filtrados os editoriais da conferência, tendo assim captados 1.034 artigos publicados no período supracitado. De todos eles, 16 não possuíam valores associados a sua coluna de autor. Logo, na etapa 3, esses 16 registros foram preenchidos utilizando da API Crossref.

Na etapa 4, foi criada uma rotina para poder verificar para cada artigo se seria possível capturar os ORCID dos autores, primeiramente via API Crossref e em sequência na API ORCID, após a rotina tivemos 3.674 autores, dos quais 2781 foram possíveis de capturar o ORCID e 893 não. Na etapa 5, foram gerados 81 grupos tal que cada grupo contém os autores com nomes similares. Nesses agrupamentos, foi possível notar que houve falsos positivos, onde autores possuíam nomes realmente similares, mas não significava que eram o mesmo. Nos dados observados, a maior parte desses casos se deu por uma utilização frequente de certos sobrenomes comuns como *Santos* e *Araújo*. Nestes casos, foi necessário fazer a desambiguação de maneira manual.

Na etapa 6, foi montada a *Rede de Coautoria* na base de dados desambiguada e agrupada, gerando 2.395 autores, onde 1.907 autores possuem ORCID e 488 não. Cada vértice também foi associado aos artigos publicados por esse autor e seus respectivos resumos.

Após ter a Rede de Coautoria completa, entramos na etapa 7 na qual enriquecemos os dados dos nós com as informações como a quantidade de artigos publicados, os tópicos abordados em seus artigos e outras informações como colocadas na seção de metodologia. Foi utilizada a OpenAlex API, na qual foi possível coletar 1573 tópicos e utilizados para montar o vetor de atributos de cada nó.

Na etapa 8, nos preparamos para rodar o algoritmo em cima desses dados. Para isso, também foi necessário definir uma variável alvo para nossos testes. Algumas foram escolhidas, como por exemplo, autores que publicaram ao menos um artigo no tópico de Mineração de Dados recebia o alvo como *verdadeiro*, analogamente um outro teste foi com o tópico de Processamento de Linguagem Natural e por fim o teste que retornou um resultado mais interessante foi ao criar uma métrica usando como base o H-Index dos autores, a métrica é simples: os autores que possuem um H-Index acima da média geral da Rede recebem o valor *verdadeiro* na variável alvo.

Para a etapa 9, irei focar nos resultados da execução usando o H-Index como métrica.

# node	# rho	# sigma	# q
1928	13	4	0.3840
874	8	3	0.3827
317	6	3	0.3535
155	6	3	0.3535
1635	6	3	0.3535
881	6	3	0.3535
871	15	4	0.3442
1122	10	3	0.3286
92	18	6	0.3207
152	5	3	0.3098

TABLE I

SUBGRUPOS ENCONTRADOS COM SEUS RESPECTIVOS NÓS PROTÓTIPOS

Tivemos vários subgrupos locais com uma qualidade q considerada boa, com uma média de 0.3484, assim como todos os 10 melhores grupos com uma qualidade acima de 0.30, vide a tabela I. Ao nos aprofundarmos no melhor subgrupo encontrado, temos um visão geral da organização desse grupo de maneira visual na Figura 2, onde os nós de coloração azul são os do subgrupo referência, os nós verdes são os do subgrupo local e o nó vermelho o protótipo.

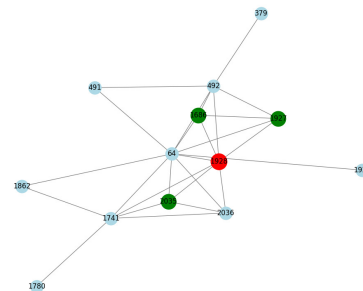


Fig. 2. Exemplo com a variável alvo sendo o H-Index de cada autor

Ao olharmos as características dos membros desse grupo de referência, foi possível notar que todos estavam de alguma maneira relacionados à uma Universidade Federal específica, e que possuíam um H-index excepcional em relação à Rede completa, mas também foi possível identificar um seletor subgrupo que possui um H-Index muito abaixo do normal, ou seja, uma anomalia local que pode ser observada na Figura 3.

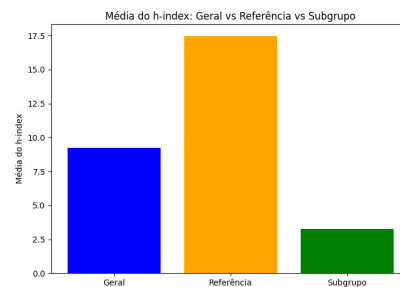


Fig. 3. Distribuição do H-Index por Subgrupo

V. CONCLUSÃO

Esse foi um projeto voltado à caracterização de comunidades científicas brasileiras na área de Inteligência Artificial (I.A.), com foco em autores que publicaram na Conferência Brasileira de Sistemas Inteligentes (BRACIS). Dada a estruturação de uma Rede de Coautoria, organizada com base em dados obtidos das APIs DBLP, Crossref, OpenAlex e ORCID, foi possível criar um conjunto de dados, com 2.395 autores e 1.034 artigos, dentro do período de 2013 a 2025. E além da montagem destes dados foi possível a exploração da estrutura das redes de coautoria através da técnica de mineração de dados chamada Local Subgroup Discovery.

A técnica de Local Subgroup Discovery(LSD) permitiu a identificação de subgrupos estruturais e de uma análise das suas propriedades. Alguns dos resultados mostram padrões institucionais além de demonstrar possíveis diferenças de desempenho acadêmico dentro das regiões de referência.

A análise utilizando a métrica derivada do H-Index se mostrou muito eficaz em capturar informações de contraste entre autores altamente produtivos e outros membros em que as métricas apresentam valores inferiores ao que seria esperado do seu grupo de referência. O que deixa explícito a capacidade que o LSD tem de reconhecer anomalias locais e permite interpretar a coesão dos grupos, tanto no aspecto estrutural quanto também na qualidade dos mesmos.

Resumidamente, o projeto permitiu verificar, de maneira empírica, o potencial que o algoritmo de Local Subgroup Discovery possui de reconhecer estruturas, padrões e anomalias que, somente através de métodos tradicionais de redes, não seria possível detectar. Para trabalhos futuros, a captura de uma maior quantidade de dados dos autores ou a escolha de diferentes variáveis alvo podem permitir um aprofundamento nas interpretações dos resultados e revelando ainda mais características nos padrões de colaboração em redes científicas.

REFERENCES

- [1] Glänzel, Wolfgang, and András Schubert. "Analysing scientific networks through co-authorship." *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems*. Dordrecht: Springer Netherlands, 2004. 257-276.
- [2] Fonseca, Bruna de Paula Fonseca E., et al. "Co-authorship network analysis in health research: method and potential use." *Health research policy and systems* 14 (2016): 1-10.
- [3] Gomes, J. E. A., Prudêncio, R. B., & Nascimento, A. C. (2016, October). A comparative study of group profiling techniques in co-authorship networks. In 2016 5th Brazilian Conference on Intelligent Systems (BRACIS) (pp. 373-378). IEEE.
- [4] Herrera, F., Carmona, C. J., González, P., & Del Jesus, M. J. (2011). An overview on subgroup discovery: foundations and applications. *Knowledge and information systems*, 29, 495-525.
- [5] Atzmueller, M. (2015). Subgroup discovery. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(1), 35-49.
- [6] A. Clauset, M. E. J. Newman, and C. Moore, 'Finding community structure in very large networks', *Phys. Rev. E*, vol. 70, p. 066111, Dec. 2004.
- [7] Qin, Yong, et al. "Artificial intelligence and economic development: An evolutionary investigation and systematic review." *Journal of the Knowledge Economy* 15.1 (2024): 1736-1770.
- [8] Benedeti Rosa, Maurício, and Luis Claudio Kubota. "Artificial intelligence: where does Brazil stand in global scientific production and what are the main technical determinants of adoption by Brazilian companies." *EconomiA* (2025).

- [9] Haak, Laurel L., et al. "ORCID: a system to uniquely identify researchers." *Learned publishing* 25.4 (2012): 259-264.
- [10] Yujian, Li, and Liu Bo. "A normalized Levenshtein distance metric." *IEEE transactions on pattern analysis and machine intelligence* 29.6 (2007): 1091-1095.
- [11] Dias, Ângelo de Sant'Ana Santos. *Caracterização de Comunidades Científicas usando Subgroup Discovery*. Diss. Universidade Federal de Pernambuco, 2018.
- [12] Lucas, T., Gomes, J., Vimieiro, R., Prudêncio, R., & Ludermit, T. (2019, October). A multivariate method for group profiling using subgroup discovery. In 2019 8th Brazilian Conference on Intelligent Systems (BRACIS) (pp. 371-376). IEEE.
- [13] Atzmueller, Martin, Stephan Doerfel, and Folke Mitzlaff. "Description-oriented community detection using exhaustive subgroup discovery." *Information Sciences* 329 (2016): 965-984.
- [14] Heinrich, Carl Vico, Tommie Lombarts, Jules Mallens, Luc Tortike, David Wolf, and Wouter Duivesteijn. "Local Subgroup Discovery on Attributed Network Graphs." *Advances in Intelligent Data Analysis XXIII: 23rd International Symposium on Intelligent Data Analysis, IDA 2025, Konstanz, Germany, May 7-9, 2025, Proceedings*. Berlin, Heidelberg: Springer-Verlag, 2025. 195-208.