LLM-Based Summarization of Academic Communities Using Relational Group Characterization

Vinicius V. Santos[1], **Hilário Tomaz Alves de Oliveira**[2], **João E. Ambrósio Gomes**[1]

[1]Instituto Federal do Sertão Pernambucano - Campus Salgueiro

[2]Programa de Pós-graduação em Computação Aplicada (PPComp),
Instituto Federal do Espírito Santo (IFES), Serra, Brasil

vinicius.vasconcelos@aluno.ifsertao-pe.edu.br
joao.ambrosio@ifsertao-pe.edu.br, hilario.oliveira@ifes.edu.br

Abstract. *Community characterization is a fundamental step in social network analysis. Although traditional methods use representative terms to label communities, the understanding of these descriptions can be limited. This work investigates using LLMs to summarize academic communities from terms obtained through relational methods. To this end, communities in an arXiv co-authorship network were analyzed using various characterization methods and centrality measures. The results indicate that different centralities yield distinct semantic representations, and that LLMs generate consistent, informative summaries even from a few terms.*

Resumo. *A caracterização de comunidades é uma etapa fundamental na análise de redes sociais. Embora métodos tradicionais utilizem termos representativos para rotular comunidades, o entendimento dessas descrições pode ser limitado. Este trabalho investiga o uso de LLMs para sumarizar comunidades acadêmicas a partir de termos obtidos por métodos relacionais. Para isso, foram analisadas comunidades detectadas em uma rede de coautoria do repositório arXiv, considerando diferentes métodos de caracterização e medidas de centralidade. Os resultados indicam que distintas centralidades produzem representações semânticas diferentes e que os LLMs geram sumarizações consistentes e informativas mesmo a partir de poucos termos.*

1. Introdução

O crescimento das redes sociais e plataformas digitais ampliou a produção e o compartilhamento de informações, consolidando esses ambientes como importantes fontes de dados. Em outubro de 2025, estimava-se que cerca de 5,66 bilhões de pessoas possuíam identidades ativas em ambientes digitais, evidenciando a relevância dessas plataformas na dinâmica social contemporânea [Allan 2025]. Nesse contexto, a Análise de Redes Sociais (ARS) tornou-se uma importante área de estudo, permitindo investigar padrões de interação, formação de grupos e estruturas de relacionamento entre usuários. Devido à natureza desses dados, sua representação por meio de redes ou grafos é amplamente utilizada para modelar e compreender as conexões existentes [Getoor and Diehl 2005].

Entre os principais problemas investigados na ARS está o estudo de comunidades, entendidas como grupos de indivíduos que apresentam maior probabilidade de estarem

conectados entre si do que com os demais elementos da rede [Fortunato and Hric 2016]. A formação desses grupos está associada ao princípio da homofilia, segundo o qual indivíduos com características semelhantes tendem a estabelecer relações com maior frequência [McPherson et al. 2001]. Como resultado, surgem comunidades formadas por usuários que compartilham interesses, características ou padrões de interação semelhantes [Wasserman and Faust 1994, McPherson et al. 2001]. Devido à sua importância para a compreensão da organização e da dinâmica das redes, a análise de comunidades constitui um dos tópicos mais estudados na ARS [Fortunato and Hric 2016].

Tradicionalmente, a análise de comunidades envolve duas etapas: detecção e caracterização [Gomes 2018]. A detecção identifica agrupamentos de usuários na rede, enquanto a caracterização descreve os temas, interesses ou atributos que distinguem cada comunidade. A caracterização de grupos/comunidades (*Group Profiling*) busca identificar termos representativos de cada comunidade. Métodos como Caracterização de Grupos Baseada em Agregação, Caracterização de Grupos Baseada em Diferenciação (CGBD) e Caracterização de Grupos Baseada em Diferenciação Egocêntrica utilizam diferentes critérios estatísticos para selecionar esses termos/usuários e descrever as características mais relevantes de cada grupo [Tang et al. 2011].

Além das informações textuais, abordagens baseadas em informações relacionais exploram a estrutura da rede para selecionar elementos representativos das comunidades. Nesse contexto, medidas de centralidade como *Degree*, *Closeness*, *Betweenness*, *Eigenvector*, *PageRank* e *Eccentricity* permitem identificar os nós mais relevantes de cada grupo, produzindo diferentes representações das comunidades e reduzindo o custo computacional da caracterização em cenários de larga escala. Essa abordagem foi definida em [Gomes et al. 2018], denominada como Caracterização de Grupos Baseada em Informações Relacionais (CGBIR).

Embora os métodos de caracterização de comunidades apresentem estratégias distintas para identificar elementos representativos dos grupos, de modo geral, todas as abordagens supracitadas geram como saída listas contendo os top-n termos mais relevantes. Apesar de úteis para destacar os principais tópicos de uma comunidade, essas listas frequentemente não fornecem contexto suficiente para uma compreensão clara de sua temática, tornando a interpretação dependente do conhecimento prévio do pesquisador e sujeita a diferentes entendimentos. Essa limitação evidencia a necessidade de abordagens capazes de transformar conjuntos reduzidos de termos em descrições mais interpretáveis e semanticamente coerentes. Nesse contexto, os Grandes Modelos de Linguagem (LLMs, do inglês *Large Language Models*) surgem como uma alternativa promissora para gerar descrições que auxiliem na compreensão e caracterização das comunidades analisadas [Rahimzadeh et al. 2025, Wang et al. 2025].

Este trabalho tem como objetivo investigar o uso de LLMs para a geração de sumarizações de comunidades acadêmicas do arXiv a partir dos dez termos mais representativos (top-10) produzidos por métodos de CGBIR. Paralelamente, realizou-se uma análise para verificar se diferentes medidas de centralidade produzem visões distintas de uma mesma comunidade, resultando em interpretações semânticas complementares de sua estrutura e temática. As seguintes questões de pesquisa guiaram este trabalho:

Pergunta de Pesquisa 1 (PP1): *A partir da variação das medidas de centralidade, em uma abordagem CGBIR, é possível obter diferentes visões de uma mesma comunidade?*

Pergunta de Pesquisa 2 (PP2): *É possível gerar um resumo da comunidade usando LLMs tendo como entrada os top-10 termos mais representativos, gerados por um método de Group Profiling?*

Para responder a essas questões, foram analisados os top-10 termos obtidos por diferentes combinações entre métodos de caracterização e medidas de centralidade em comunidades acadêmicas em uma rede de coautoria do repositório de artigos científicos do arXiv. Esses termos foram utilizados como entrada para diferentes LLMs, e as descrições (resumos) geradas foram avaliadas por meio de medidas de avaliação baseadas em similaridade semântica em comparação com resumos de referência produzidos a partir do conteúdo textual completo das comunidades.

O conjunto de dados utilizado neste trabalho está disponível publicamente em um repositório no GitHub[1]. Além disso, os códigos desenvolvidos para a execução dos experimentos, incluindo as implementações em Java e Python, estão disponíveis no repositório do GitHub do artigo[2].

2. Trabalhos Relacionados

Estudos sobre comunidades em redes sociais evoluíram desde análises focadas exclusivamente na estrutura das redes para abordagens capazes de caracterizar semanticamente os grupos identificados. Os fundamentos da ARS foram consolidados por Wasserman e Faust [Wasserman and Faust 1994], que sistematizaram conceitos, métricas e métodos para análise de estruturas relacionais. Posteriormente, Getoor e Diehl [Getoor and Diehl 2005] ampliaram esse cenário ao introduzirem o conceito de *Link Mining*, explorando técnicas de mineração de dados aplicadas a grafos e relações entre entidades.

Com o avanço das pesquisas em redes sociais, a atenção passou a se concentrar também na identificação automática de grupos. Tang e Liu [Tang and Liu 2010] investigaram métodos para detecção de comunidades, permitindo identificar subconjuntos de nós fortemente conectados dentro das redes. Entretanto, embora essas abordagens fossem capazes de detectar comunidades, elas não forneciam mecanismos para compreender ou descrever semanticamente os grupos encontrados. Buscando suprir essa limitação, Tang et al. [Tang et al. 2011] propuseram o conceito de *Group Profiling*, cujo objetivo é caracterizar comunidades por meio da identificação de atributos e termos representativos. Apesar de representar um avanço importante na interpretação dos grupos, a abordagem produz listas de termos ou características, exigindo interpretação humana para compreensão dos temas associados às comunidades.

Em redes de coautoria, trabalhos como os de [Savić et al. 2015] investigaram propriedades estruturais das comunidades científicas e fatores relacionados à sua formação. Posteriormente, Gomes [Gomes 2018] propôs uma abordagem baseada em informações relacionais para seleção de termos representativos de comunidades, enquanto Gomes et al. [Gomes et al. 2018] investigaram o impacto de diferentes medidas de centralidade

[1] `https://github.com/andrecamara/dataset_ngc`

[2] `https://github.com/Viniciusvass/llm-community-summarization`

na construção desses perfis. Embora essas abordagens permitam caracterizar comunidades por meio de conjuntos de termos relevantes, os resultados ainda dependem de interpretação humana para obtenção de descrições semânticas mais completas.

Mais recentemente, o avanço da aplicação de LLMs impulsionou pesquisas voltadas à geração automática de perfis e descrições textuais. Trabalhos como [Rahimzadeh et al. 2025] utilizaram LLMs para perfilamento de usuários em redes sociais, enquanto Yin et al. [Yin et al. 2026] combinaram LLMs e validação de especialistas para gerar personas comportamentais a partir de milhões de publicações em redes sociais. Além disso, Wang et al. [Wang et al. 2025] exploraram a geração de temas e descrições semânticas a partir de dados relacionados a redes sociais. Entretanto, essas abordagens não utilizam técnicas de caracterização relacional de comunidades nem investigam o impacto de diferentes medidas de centralidade sobre as representações produzidas.

A análise dos trabalhos relacionados evidencia avanços importantes em diferentes etapas do processo de estudo de comunidades em redes sociais. Os trabalhos de ARS e detecção de comunidades consolidam os fundamentos para identificação de grupos, enquanto as abordagens de *Group Profiling* e CGBIR permitem selecionar termos representativos de forma eficiente e considerando aspectos estruturais da rede. Paralelamente, estudos recentes demonstram o potencial dos LLMs na interpretação de grandes volumes de dados e na geração de descrições semanticamente ricas.

Entretanto, observa-se uma lacuna na literatura quanto à integração entre métodos tradicionais de caracterização de comunidades e modelos generativos de linguagem. Embora existam abordagens capazes de identificar os termos mais representativos de uma comunidade, ainda são escassos os estudos que investigam a utilização desses termos como única fonte de informação para a geração de descrições textuais interpretáveis. Além disso, permanece pouco explorado o impacto que diferentes medidas de centralidade exercem sobre as representações semânticas produzidas para uma mesma comunidade.

3. Metodologia

A Figura 1 apresenta uma visão geral do fluxo experimental, abrangendo as etapas de obtenção dos dados, geração dos gabaritos e descrições das comunidades, bem como os procedimentos de avaliação adotados. Inicialmente, são construídas redes acadêmicas de coautoria a partir de artigos coletados no repositório do arXiv, sobre as quais são aplicados algoritmos de detecção de comunidades. Em seguida, são utilizados diferentes métodos de representação de termos, combinados com medidas de centralidade, para a seleção dos dez termos mais representativos de cada comunidade.

Na segunda etapa, todas as informações da comunidade são consolidadas e utilizadas como entrada para o modelo GPT-OSS de 120 bilhões de parâmetros, responsável pela geração de um gabarito textual que serve como referência. Na terceira etapa, são geradas sumarizações a partir dos top-10 termos utilizando diferentes LLMs executados localmente, e os resultados são armazenados para análise posterior. Na quarta etapa, realiza-se a análise de similaridade entre os resumos produzidos pelos diferentes LLMs, com base em medidas de similaridade textual. Por fim, na quinta etapa, os resumos são avaliados em relação ao gabarito utilizando medidas de avaliação semântica, como BERTScore [Zhang et al. 2020] e a similaridade do cosseno, além da aplicação de testes estatísticos para análise de significância entre os modelos.

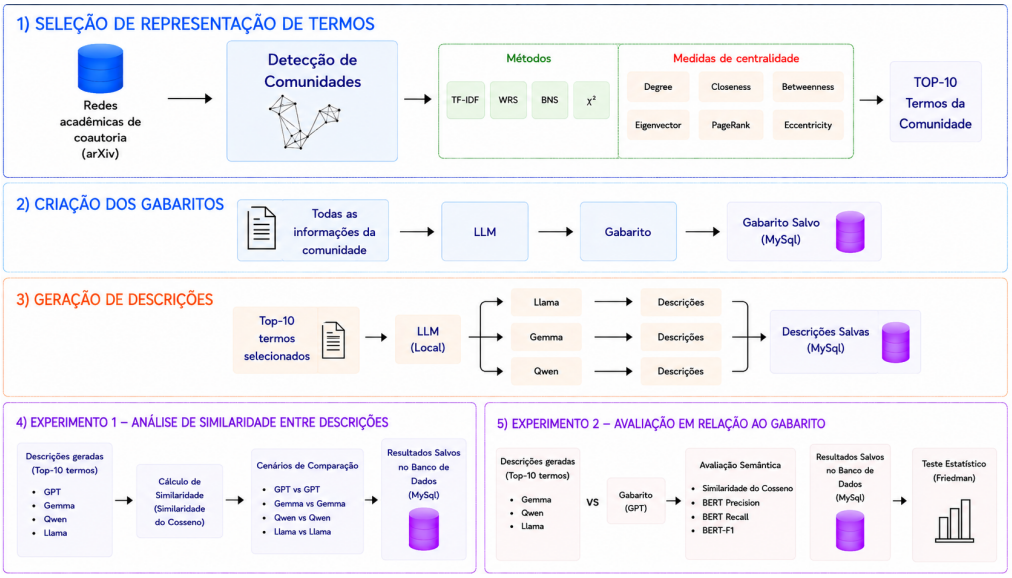


Figura 1. Fluxo metodológico proposto para geração e avaliação de sumarizações de comunidades acadêmicas.

3.1. Base de Dados e Seleção dos Dados de Entrada

Como base de dados para os experimentos, foi utilizada uma rede de coautoria extraída da plataforma arXiv, aplicando-se um filtro para seleção de trabalhos da área de Inteligência Artificial (cs.IA), composta por publicações realizadas entre 2012 e 2014. Em uma rede de coautoria, os nós representam autores e as arestas representam relações de coautoria entre pesquisadores. A Tabela 1 apresenta as principais características estruturais das comunidades identificadas, incluindo tamanho, grau médio, densidade e coesão.

Tabela 1. Estatísticas das comunidades analisadas.

Comunidade	Tamanho	Grau Médio	Densidade	Coesão
6	41	2,829	7,1%	0,248
80	28	2,929	10,8%	0,384
104	68	3,735	5,6%	0,307
116	28	3,929	14,6%	0,363
134	60	3,167	5,4%	0,299
145	14	3,429	26,4%	0,377
151	18	3,333	19,6%	0,346
153	53	3,321	6,4%	0,337
156	29	3,724	13,3%	0,352
256	33	3,273	10,2%	0,316

Observa-se a existência de comunidades de diferentes dimensões e níveis de conectividade, evidenciando grupos de pesquisadores com padrões distintos de colaboração. Ao final da etapa de caracterização, foram obtidos os top-10 termos mais representativos de cada comunidade para as diferentes combinações entre métodos estatísticos e medidas de centralidade, os quais foram utilizados como entrada para a geração das sumarizações pelos modelos de linguagem. Os detalhes para a obtenção da rede, detecção das comunidades e de seus termos caracterizadores podem ser acessados em [Gomes et al. 2018].

Diversos fatores podem contribuir para a formação de comunidades em redes de coautoria [Savić et al. 2015]. Nesse contexto, descrever e compreender as características que justificam a formação dessas comunidades é fundamental para o avanço do entendimento das estruturas acadêmicas. A investigação de suas características e peculiaridades pode contribuir para o desenvolvimento de novas abordagens capazes de fortalecer e ampliar as redes de colaboração científica.

3.2. Criação dos Gabaritos

A segunda etapa da metodologia consiste na geração dos resumos de referência (gabaritos) utilizados na avaliação das sumarizações automáticas produzidas ao longo do estudo. Diferentemente dos resumos gerados a partir dos top-10 termos representativos de cada comunidade, os gabaritos foram produzidos com base em todas as informações textuais disponíveis para cada agrupamento. Essas informações foram fornecidas ao modelo GPT-OSS de 120B de parâmetros, executado por meio da plataforma DeepInfra[3], sendo responsável pela geração de descrições completas das comunidades analisadas. Os parâmetros de temperatura e número máximo de *tokens* não foram definidos manualmente durante a geração dos gabaritos, sendo mantidas as configurações padrão disponibilizadas pelo modelo.

Os gabaritos gerados foram armazenados em um banco de dados e utilizados como referência durante o processo de avaliação. O objetivo dessa etapa foi construir representações textuais com maior riqueza semântica e contextual sobre cada comunidade, permitindo comparar a capacidade dos LLMs em produzir descrições representativas a partir de informações condensadas, como os top-10 termos obtidos pelos métodos de *Group Profiling*.

3.3. Criação das Descrições das Comunidades

Na terceira etapa da metodologia, foram gerados resumos utilizando exclusivamente os top-10 termos representativos de cada comunidade, obtidos a partir dos experimentos realizados por [Gomes 2018]. Esses termos foram utilizados como entrada para os modelos do Llama 3.2 3B, Qwen 3.5 2B e Gemma 3 1B, responsáveis pela geração automática das descrições textuais das comunidades acadêmicas analisadas. O modelo Qwen 3.5 2B também foi executado por meio da plataforma DeepInfra, enquanto os modelos Gemma 3 1B e Llama 3.2 3B foram executados localmente. Os parâmetros de temperatura e número máximo de tokens seguiram as configurações padrão de cada modelo em suas respectivas plataformas de execução.

Na Tabela 2 é apresentado o *prompt template* usado tanto para a geração dos gabaritos quanto para a produção dos resumos das comunidades identificadas a partir do top-10. O *prompt* foi projetado para orientar os LLMs a produzirem descrições acadêmicas em inglês exclusivamente com base no conjunto de palavras-chave fornecido. Para reduzir a ocorrência de informações não suportadas pelos dados de entrada, foram definidas restrições explícitas que limitam a geração ao conteúdo diretamente inferível dos termos selecionados, impedindo a inclusão de contextos, aplicações ou áreas de pesquisa não representadas pelas palavras-chave. Além disso, o modelo é instruído a produzir um texto contínuo, objetivo e coerente, sem títulos, listas ou conclusões, garantindo maior

[3]`https://deepinfra.com/`

padronização entre as descrições geradas e permitindo comparações mais consistentes nas etapas de avaliação.

Tabela 2. *Prompt template* utilizado para criação dos gabaritos e resumos.

Instruções de Sistema
`You are an academic researcher.`
Generate an academic description in English about a research community using exclusively the provided keywords.
RULES:
The text must be written entirely in ENGLISH.
Use only information that can be directly inferred from the keywords.
Do not invent contexts, applications, domains, or research areas that are not represented by the terms.
Be objective, coherent, and descriptive.
Do not use bullet points or lists.
Do not use titles.
Do not use subtitles.
Do not use expressions such as “Academic Summary”, “Conclusions”, “Preliminary Conclusions”, “Final Considerations”, “In summary”, “To conclude”, “Finally”, or equivalent phrases.
Do not make recommendations or judgments about the quality of the research.
Do not present final conclusions.
Return only the community description as continuous text, organized into two or three paragraphs.
KEYWORDS:
{keywords}

3.4. Metodologia Experimental

Para responder às perguntas de pesquisa (**PP1** e **PP2**) que nortearam este trabalho, foram conduzidos dois experimentos. O primeiro teve como foco analisar se diferentes medidas de centralidade são capazes de produzir diferentes interpretações de uma mesma comunidade a partir dos termos selecionados para sua representação e se elas se complementam. O segundo experimento buscou avaliar a qualidade dos resumos das comunidades geradas por diferentes LLMs, comparando-os com os gabaritos de referência produzidos.

Para responder à **PP1**, foi realizado um primeiro experimento visando analisar a similaridade entre os resumos gerados a partir dos top-10 termos obtidos pela abordagem de CGBIR. As comparações foram realizadas entre resumos pertencentes à mesma comunidade, mantendo-se os métodos utilizados para seleção dos termos, sendo eles o *Term Frequency-Inverse Document Frequency* (TF-IDF), *Bi-Normal Separation* (BNS), *Wilcoxon Rank Sum* (WRS) ou *Qui-quadrado* (X^2), e variando apenas a medida de centralidade empregada. Para cada resumo, foi gerada uma representação vetorial usando o modelo *Paraphrase Multilingual MiniLM-L12-v2* da biblioteca *Sentence Transformers* [Reimers and Gurevych 2019]. A similaridade do cosseno foi utilizada para comparar os vetores, permitindo avaliar a similaridade entre as descrições produzidas para uma mesma comunidade. A partir dos valores de similaridade, foi obtida a dissimilaridade entre os pares de resumos das comunidades, permitindo verificar o grau de convergência ou divergência entre as descrições produzidas para uma mesma comunidade.

Para responder à **PP2**, foi realizado um segundo experimento de avaliação dos resumos das comunidades gerados por diferentes LLMs, utilizando exclusivamente os top-10 termos produzidos pela abordagem de CGBIR. Os resumos produzidos pelos modelos

Gemma, Qwen e Llama foram comparados aos gabaritos de referência gerados pelo modelo GPT-OSS 120B. A avaliação foi conduzida utilizando a medida de similaridade do cosseno a partir da representação gerada pelo modelo Multilingual MiniLM e a medida de F1-escore computada a partir da medida de similaridade semântica do BERTScore. Adicionalmente, foi aplicado o teste não paramétrico de Friedman sobre os valores de F1-escore computado do BERTScore obtidos para cada comunidade, permitindo verificar a existência de diferenças estatisticamente significativas entre os modelos.

4. Resultados e Discussão

4.1. Resultados do Experimento 1

As descrições textuais geradas por todos os LLMs avaliados, considerando as variações dos conjuntos de top-10 termos selecionados por diferentes combinações de métodos de caracterização e medidas de centralidade em uma abordagem CGBIR, estão disponibilizadas no repositório do trabalho. Com base nessas descrições, foram realizadas análises de dissimilaridade semântica entre os perfis gerados. Na Tabela 3 é apresentado, para cada comunidade analisada, o par de perfis que obteve a maior dissimilaridade. Observa-se que as maiores divergências concentram-se principalmente nos métodos BNS e WRS, reforçando a influência da combinação entre método e medida de centralidade na produção de diferentes representações semânticas de uma mesma comunidade.

Tabela 3. Perfis com maior dissimilaridade observados para cada comunidade com o modelo GPT-OSS 120B.

Comunidade	Método	Centralidade 1	Centralidade 2	Dissimilaridade
6	WRS	*Eccentricity*	*Degree*	0,55221
80	BNS	*Betweenness*	*Closeness*	0,44094
104	WRS	*Eccentricity*	*Degree*	0,58969
116	BNS	*Degree*	*Closeness*	0,46883
134	BNS	*Eigenvector*	*Betweenness*	0,72215
145	WRS	*Eigenvector*	*Eccentricity*	0,47722
151	BNS	*PageRank*	*Closeness*	0,52658
153	WRS	*Eccentricity*	*PageRank*	0,58624
156	WRS	*Eccentricity*	*Degree*	0,59204
256	BNS	*Eccentricity*	*Closeness*	0,47755

A análise das características estruturais das comunidades também ajuda a explicar os resultados observados. A comunidade 134 apresentou o maior valor de dissimilaridade e possui uma das menores densidades da amostra, sugerindo uma estrutura mais heterogênea. Situação semelhante foi observada nas comunidades 104, 153 e 156, que também figuram entre os casos de maior divergência. Nesses cenários, diferentes medidas de centralidade tendem a destacar autores posicionados em regiões distintas da rede, produzindo conjuntos de termos menos sobrepostos e descrições semanticamente mais divergentes. Em contraste, comunidades mais densas e coesas tendem a favorecer a seleção de autores semelhantes entre as centralidades, resultando em representações mais estáveis. Esses resultados sugerem que comunidades menos densas favorecem a obtenção de perspectivas complementares sobre seus temas de pesquisa.

É possível observar ainda uma predominância dos métodos BNS e WRS entre os pares com maior dissimilaridade. Diferentemente do TF-IDF, esses métodos possuem

maior sensibilidade às diferenças de distribuição dos atributos entre grupos, favorecendo a seleção de termos mais específicos ou discriminativos. Quando combinados com diferentes medidas de centralidade, tendem a produzir conjuntos de termos mais heterogêneos, enfatizando aspectos distintos da comunidade e resultando em descrições semanticamente mais distantes entre si.

Observa-se também um padrão recorrente nas combinações de centralidades associadas às maiores dissimilaridades. Em quatro das dez comunidades analisadas (6, 104, 153 e 156), a centralidade *Eccentricity* aparece associada a medidas como *Degree* ou *PageRank*. Esse resultado é coerente com as diferenças conceituais entre essas métricas, uma vez que *Eccentricity* tende a selecionar autores localizados em regiões mais periféricas da rede, enquanto *Degree* e *PageRank* priorizam autores mais centrais e influentes. Consequentemente, os conjuntos de autores selecionados apresentam menor sobreposição, produzindo termos representativos e sumarizações semanticamente distintas.

De forma semelhante, as combinações envolvendo *Betweenness*, *Closeness* e *Eigenvector*, observadas nas comunidades 80 e 134, refletem diferentes perspectivas estruturais da comunidade. Enquanto *Betweenness* privilegia autores que atuam como pontes entre subgrupos, *Closeness* destaca autores localizados em posições centrais da rede e *Eigenvector* favorece autores conectados a outros autores influentes. Essas diferenças na seleção dos autores mais representativos contribuem para a geração de descrições que enfatizam aspectos temáticos distintos de uma mesma comunidade.

A análise dos pares mais dissimilares mostrou-se consistente com as propriedades estruturais das medidas de centralidade. *Degree* destacou tópicos amplamente compartilhados pelos autores mais conectados, enquanto *Closeness* enfatizou temas de regiões centrais da rede. *Betweenness* evidenciou conceitos relacionados à interseção entre subtemas, refletindo o papel de intermediação dos autores. *Eigenvector* e *PageRank* destacaram tópicos associados a autores influentes, enquanto *Eccentricity* revelou perspectivas mais específicas de autores situados em regiões periféricas da rede.

Em diversos casos, os perfis gerados apresentaram diferenças semânticas relevantes, porém não contraditórias, indicando que diferentes medidas de centralidade capturam aspectos distintos e complementares de uma mesma comunidade de coautoria. Esses resultados confirmam a **PP1**, evidenciando que a escolha da centralidade influencia os termos selecionados e as descrições geradas pelos LLMs. Os cenários de maior dissimilaridade reforçam essa complementaridade, contribuindo para uma caracterização mais abrangente das comunidades analisadas.

4.2. Resultados do Experimento 2

A Tabela 4 apresenta, para cada comunidade, a combinação entre modelo, método de caracterização e medida de centralidade que obteve o maior valor na medida de F1-escore do BERTScore (BERT-F1). É possível observar que tanto o modelo utilizado quanto a combinação entre método e centralidade influenciam a qualidade das descrições produzidas. Os resultados deste experimento demonstram que os modelos avaliados foram capazes de gerar descrições semanticamente próximas aos gabaritos de referência, mesmo utilizando apenas os top-10 termos representativos de cada comunidade. Esse resultado sugere que os termos selecionados pelos métodos de caracterização preservam informações suficientes para representar os temas centrais das comunidades analisadas.

O modelo Llama 3.2 3B foi responsável por seis dos dez melhores resultados por comunidade, enquanto o Qwen 3.5 2B obteve os quatro resultados restantes. O maior valor de BERT-F1 foi obtido pelo modelo Llama 3.2 3B na comunidade 6, utilizando CHI e *Betweenness*, com valor de 0,83978. Resultados elevados também foram observados para as comunidades 104, 116, 151, 153, 156 e 256, todas com valores superiores a 0,80, indicando alta proximidade semântica entre as descrições geradas e os respectivos gabaritos de referência.

Tabela 4. Melhores resultados por comunidade com base no valor do BERT-F1.

LLM	Comunidade	Método	Centralidade	Similaridade	BERT-F1
Llama 3.2 3B	6	CHI	*Betweenness*	0,57331	0,83978
Llama 3.2 3B	80	CHI	*Betweenness*	0,58423	0,79846
Qwen 3.5 2B	104	TF-IDF	*PageRank*	0,50177	0,82402
Qwen 3.5 2B	116	TF-IDF	*Degree*	0,56799	0,83527
Llama 3.2 3B	134	CHI	*PageRank*	0,44433	0,79922
Qwen 3.5 2B	145	WRS	*Eccentricity*	0,52378	0,77858
Qwen 3.5 2B	151	TF-IDF	*Closeness*	0,59603	0,83583
Llama 3.2 3B	153	CHI	*Betweenness*	0,66999	0,83414
Llama 3.2 3B	156	TF-IDF	*Eccentricity*	0,64960	0,82688
Llama 3.2 3B	256	CHI	*PageRank*	0,45965	0,83832

Em relação aos métodos de caracterização, o TF-IDF e o CHI apresentaram maior frequência entre os melhores resultados, com o TF-IDF aparecendo em quatro comunidades e CHI em cinco comunidades. O método WRS esteve presente em uma comunidade, enquanto BNS não figurou entre os melhores desempenhos observados. Esse resultado sugere que TF-IDF e CHI foram mais eficazes na seleção de termos capazes de representar os principais temas das comunidades, favorecendo a geração de descrições semanticamente mais próximas dos gabaritos. Enquanto o TF-IDF tende a destacar termos característicos e recorrentes da comunidade, o CHI prioriza termos com maior capacidade de diferenciar uma comunidade das demais, fornecendo informações mais representativas para os LLMs.

A análise das medidas de centralidade revelou uma distribuição relativamente equilibrada entre os melhores resultados. *Betweenness* e *PageRank* foram as mais frequentes, presentes em três comunidades, seguidas por *Eccentricity*, *Degree* e *Closeness*. A ausência de uma centralidade dominante indica que a representação mais adequada depende das características estruturais de cada comunidade. A frequência de *Betweenness* e *PageRank* sugere a relevância de autores intermediários e estruturalmente influentes, enquanto as demais medidas também fornecem perspectivas complementares para a caracterização das comunidades.

Esse resultado é consistente com os resultados discutidos no Experimento 1, ou seja, diferentes medidas de centralidade tendem a capturar perspectivas complementares da estrutura temática das comunidades. Enquanto algumas centralidades favorecem autores mais influentes ou centrais, outras destacam autores que atuam como intermediários ou pertencem a regiões mais periféricas da rede, produzindo conjuntos distintos de termos representativos. Consequentemente, diferentes combinações entre métodos de caracterização e centralidades podem resultar em sumarizações de melhor qualidade, dependendo da comunidade analisada.

Para verificar se as diferenças observadas entre os modelos eram estatisticamente significativas, foi aplicado o teste não paramétrico de Friedman sobre os valores do BERT-F1. Os resultados indicaram uma estatística de Friedman de 61,7249 e um p-valor de aproximadamente $3{,}95 \times 10^{-14}$, permitindo rejeitar a hipótese nula de equivalência entre os modelos avaliados. O Qwen 3.5 2B apresentou o melhor desempenho médio (1,7625), seguido pelo Llama 3.2 3B (1,825) e pelo Gemma 3 1B (2,4125). Embora o Llama 3.2 3B tenha obtido o maior número de melhores resultados por comunidade, o ranking médio indica que o Qwen 3.5 2B apresentou desempenho mais consistente ao longo de todos os cenários avaliados.

Os resultados obtidos demonstram que os LLMs avaliados são capazes de gerar descrições semanticamente representativas para as comunidades, utilizando apenas os top-10 termos produzidos pelos métodos de caracterização de grupos. Além disso, evidenciam que a qualidade das descrições depende tanto do modelo utilizado quanto da combinação entre método de caracterização e medida de centralidade empregada na seleção dos termos. O teste de Friedman confirmou a existência de diferenças estatisticamente significativas entre os modelos avaliados, com destaque para o Qwen 3.5 2B em desempenho médio e para o Llama 3.2 3B em número de melhores resultados por comunidade. Dessa forma, os resultados permitem responder positivamente à **PP2**, evidenciando que LLMs são capazes de gerar descrições semanticamente coerentes e próximas aos gabaritos de referência, mesmo a partir de representações condensadas das comunidades.

5. Considerações Finais e Trabalhos Futuros

Este trabalho investigou a utilização de LLMs para gerar descrições de comunidades acadêmicas a partir dos 10 termos mais representativos gerados por métodos de CGBIR. Os resultados demonstraram que diferentes medidas de centralidade produzem representações semânticas distintas para uma mesma comunidade, frequentemente evidenciando perspectivas complementares sobre sua estrutura temática. Além disso, verificou-se que os LLMs avaliados foram capazes de gerar descrições semanticamente representativas utilizando apenas informações condensadas das comunidades. Dessa forma, foi possível responder positivamente às duas questões de pesquisa propostas, demonstrando tanto a influência das medidas de centralidade na caracterização das comunidades quanto a viabilidade da utilização de LLMs para geração automática de descrições a partir dos top-10 termos mais descritivos das comunidades criadas.

Como trabalhos futuros, pretende-se avaliar a abordagem em outros tipos de redes sociais e domínios de aplicação, investigando sua capacidade de generalização para diferentes contextos. Também será analisado o impacto da quantidade de termos representativos utilizados como entrada para LLMs, considerando diferentes configurações além do top-10. Outra direção consiste em incorporar informações adicionais das comunidades, como métricas estruturais e metadados, com o objetivo de enriquecer as descrições produzidas. Além disso, pretende-se investigar a aplicação da abordagem em comunidades dinâmicas, de modo a possibilitar o acompanhamento de sua evolução ao longo do tempo, identificando mudanças temáticas, padrões de crescimento, fragmentação e convergência entre comunidades. Por fim, poderão ser desenvolvidas métricas específicas para a avaliação da qualidade das descrições das comunidades, complementando as medidas de similaridade semântica utilizadas neste trabalho.

Referências

Allan, S. (2025). Estatísticas das redes sociais em 2026: últimas tendências e dados. https://digiexe.com/pt/blog/social-media-statistics/. Acesso em: 23 maio 2026.

Fortunato, S. and Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, 659:1–44.

Getoor, L. and Diehl, C. P. (2005). Link mining: a survey. *SIGKDD Exploration Newsletter*, 7(2):3–12.

Gomes, J. a. E. A., Prudêncio, R. B. C., and Nascimento, A. C. A. (2018). Centrality-based group profiling: A comparative study in co-authorship networks. *New Gen. Comput.*, 36(1):59–89.

Gomes, J. E. A. (2018). *Caracterização de Grupos Baseada em Informações Relacionais*. Tese de doutorado, Universidade Federal de Pernambuco, Recife.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444.

Rahimzadeh, V., Hamzehpour, A., Shakery, A., and Asadpour, M. (2025). From millions of tweets to actionable insights: Leveraging llms for user profiling.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Savić, M., Ivanović, M., Radovanović, M., Ognjanović, Z., Pejović, A., and Jakšić Krüger, T. (2015). *ICT Innovations 2014: World of Data*, chapter Exploratory Analysis of Communities in Co-authorship Networks: A Case Study, pages 55–64. Springer International Publishing, Cham.

Tang, L. and Liu, H. (2010). *Community Detection and Mining in Social Media*. Morgan and Claypool, New York, USA.

Tang, L., Wang, X., and Liu, H. (2011). Group profiling for understanding social structures. *ACM Trans. Intell. Syst. Technol.*, 3(1).

Wang, Q., Erqsous, M., Barner, K. E., and Mauriello, M. L. (2025). Lata: A pilot study on llm-assisted thematic analysis of online social network data generation experiences. *Proc. ACM Hum.-Comput. Interact.*, 9(2).

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: methods and applications (structural analysis in the social sciences)*. Cambridge University Press, New York, USA.

Yin, M., Liu, H., Lian, B., and Cai, R. (2026). Copersona: Leveraging llms and expert collaboration to understand user personas through social media data analysis. *Design for Augmented Humanity*, 1(1):49–66.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., and Artzi, Y. (2020). Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.