



10° EBBC

Encontro Brasileiro de
Bibliometria e Cientometria



O USO DA BIBLIOTECA NETFOG PARA GERAÇÃO E VISUALIZAÇÃO DE REDES CIENTÍFICAS VIA OPENALEX

Tiago Fogolin Ragassi

<https://orcid.org/0009-0005-5338-2916> | ticofogolin@gmail.com

Universidade de Marília (UNIMAR)

Rafael Gutierrez Castanha

<http://orcid.org/0000-0002-3117-1780> | r.castanha@gmail.com

Universidade de Marília (UNIMAR)

Resumo:

Apresenta o desenvolvimento de uma nova funcionalidade na biblioteca NetFog para extração e modelagem automatizada de redes científicas. Esta nova funcionalidade acessa dados abertos da base OpenAlex e constrói grafos de coautoria, coocorrência de palavras-chave e cocitação de documentos e autores. Avaliou-se o desempenho computacional em cenários massivos, revelando alta escalabilidade. O processamento interno estruturou milhões de conexões em poucos segundos, concentrando o custo temporal estritamente na renderização visual. Conclui-se que a solução oferece infraestrutura robusta, rápida e reprodutível para os estudos bibliométricos.

Palavras-Chave: Bibliometria. Redes Científicas. OpenAlex.

EIXO TEMÁTICO:

Técnicas, Ferramentas e Infraestruturas

MODALIDADE:

Resumo expandido

DOI: 10.22477/x.ebbc.1101

COMO CITAR ESTE ARTIGO:

RAGASSI, Tiago Fogolin; CASTANHA, Rafael Gutierrez. O uso da biblioteca NetFog para geração e visualização de redes científicas via OpenAlex. *In: EBBC - Encontro Brasileiro de Bibliometria e Cientometria*. 10. **Anais [...]**. Curitiba, 2026. DOI: 10.22477/x.ebbc.1101. Disponível em: <https://doi.org/10.22477/x.ebbc.1101>.



1 INTRODUÇÃO

Baseada na teoria dos grafos, a modelagem de redes científicas alia matemática e visualização para simplificar dados e revelar padrões complexos invisíveis a métodos tradicionais (Grácio e Oliveira, 2020). Nos Estudos Métricos da Informação (EMI), mapear a disseminação do conhecimento científico requer a construção dessas redes, evidenciando as relações topológicas entre documentos, autores, temas e países.

Sendo centrais para entender a organização e difusão do conhecimento (Rossi; Mena-Chalco, 2018), as redes impulsionam ferramentas nos EMI. Castanha (2022), ao criar o *The Coupler*, elenca diversas soluções bibliométricas (Bibexcel, Bibliometrix, CiteSpace, CoPalRed, IN-SPIRE™, InCites, Metaknowledge, Network Workbench, Publish or Perish, Sci², SCImago, SciMAT, SciVal, VOSviewer). Essa abundância demonstra a consolidação da área, mas também revela a urgência por melhorias contínuas na infraestrutura, especialmente em integração automatizada e escalabilidade.

Softwares baseados em grafos (como VOSviewer, Gephi e Ucinet) operam majoritariamente via interfaces visuais (*point and click*). Apesar de populares, apresentam limitações de desempenho com grandes volumes de dados e restrições de customização, por nem sempre terem código aberto. Nesse cenário, linguagens como Python e R (através de bibliotecas como igraph, NetworkX e graph-tool) surgem como alternativas robustas para análises em larga escala, garantindo automação e reprodutibilidade. Contudo, falta a essas bibliotecas suporte nativo e otimizado para gerar redes bibliométricas específicas (coautoria, cocitação e acoplamento) com integração direta a APIs como a OpenAlex, funcionalidade já presente em algumas ferramentas visuais.

A disponibilização de bases científicas via API, como a OpenAlex, ampliou a coleta automatizada de metadados estruturados, viabilizando a construção de redes reproduzíveis. É neste cenário que se insere a NetFog (<https://pypi.org/project/netfog/>), elemento central desta pesquisa. Sendo uma biblioteca nacional para Python focada na geração e visualização de grafos, a NetFog recebeu uma nova funcionalidade: aproveitando sua versatilidade e o acesso público à OpenAlex, a ferramenta agora recupera dados, gera e visualiza redes científicas a partir dos metadados da API.

Diante o exposto, esta pesquisa busca responder ao seguinte problema de pesquisa: como estruturar e implementar, na biblioteca NetFog, uma funcionalidade capaz de integrar automaticamente a extração de metadados científicos da OpenAlex à geração e visualização de redes científicas? O objetivo deste artigo é apresentar o desenvolvimento e a implementação dessa nova funcionalidade na biblioteca NetFog, demonstrando seu funcionamento na construção de redes científicas a partir de dados coletados via API da OpenAlex.

2 DESENVOLVIMENTO

A fim de demonstrar o uso da biblioteca NetFog para geração e visualização de redes científicas, via OpenAlex, o fluxo metodológico deste trabalho compreende a extração automatizada de metadados advindos da base e sua transformação em estruturas matemáticas de grafos. Para viabilizar esse fluxo, a principal etapa técnica desta pesquisa consistiu no desenvolvimento de um novo módulo nativo para a biblioteca NetFog capaz de realizar consultas em tempo real na OpenAlex via API pública da base de dados. A NetFog é uma biblioteca de manipulação e análise de grafos para a linguagem de programação *Python*. A biblioteca é capaz de oferecer um conjunto robusto de funcionalidades para auxiliar na modelagem

de redes complexas, abrangendo visualização, cálculo de métricas e execução de algoritmos. Embora a NetFog seja utilizada na linguagem de programação *Python*, todas as funcionalidades da biblioteca foram desenvolvidas em linguagem *Rust* a fim de garantir maior eficiência de processamento e segurança de memória.

Nesse sentido, foi implementada uma nova funcionalidade, que permite a construção automatizada de redes com base em registros extraídos a partir do *endpoint* base <https://api.openalex.org/works>, que retorna todos os registros disponíveis da plataforma paginados. A extração é executada de maneira iterativa por meio de um sistema de paginação baseado em cursor (*cursor-based pagination*)¹. Assim, foi implementado um algoritmo à biblioteca NetFog capaz de realizar requisições contínuas à API, atualizando o cursor dinamicamente a cada resposta, até que todos os dados sejam recuperados ou até que o teto amostral definido pelo usuário seja atingido, momento em que a estrutura de dados é truncada para garantir a precisão do limite. Para adequar a extração aos diferentes objetivos de pesquisa, a funcionalidade criada na biblioteca é parametrizável segundo diversos filtros, detalhados no Quadro 1.

Quadro 1 - Parâmetros implementados

PARÂMETRO	DESCRIÇÃO
<i>search</i>	Busca geral nos artigos pelo termo informado.
<i>author</i>	Permite buscar trabalhos com base no nome de um autor.
<i>author_id</i>	Busca específica e precisa utilizando o identificador único do autor na plataforma OpenAlex.
<i>author_orcid</i>	Busca específica utilizando o ORCID do autor.
<i>keyword</i>	Busca publicações associadas a uma palavra-chave.
<i>limit</i>	Quantidade máxima de resultados que serão retornados pela api e utilizados para a criação da rede.
<i>min_weight</i>	Intensidade mínima da aresta para que a relação seja incluída na rede.

¹ Documentação disponível em: <https://docs.openalex.org/how-to-use-the-api/get-lists-of-entities/paging>

PARÂMETRO	DESCRIÇÃO
<i>graph_type</i>	Indica o tipo de rede que será gerada (Coautoria, Coocorrência de Palavras-Chave ou Cocitação).
<i>api_key</i>	Chave de autenticação da OpenAlex, utilizada para garantir limites maiores de requisição.

Fonte: Elaboração dos autores

Para garantir a flexibilidade na coleta de dados, a funcionalidade oferece diferentes escopos de busca. Os parâmetros *search* e *keyword* permitem mapear áreas do conhecimento ou tópicos específicos de forma ampla. Enquanto o *search* realiza uma varredura geral em títulos, resumos e no texto completo das publicações, conforme mencionado na documentação da API da base² o parâmetro *keyword* restringe os resultados a trabalhos formalmente associados a uma palavra específica, sendo ideal para construir redes de coocorrência de palavras-chave ou explorar a estrutura da literatura de um nicho de pesquisa.

Quando o foco da análise recai sobre a produção de pesquisadores específicos, os filtros de autoria (*author*, *author_id* e *author_orcid*) assumem um papel fundamental. O parâmetro *author* realiza uma busca textual, priorizando o primeiro perfil retornado, o que pode ser útil para consultas rápidas. No entanto, para pesquisas mais específicas e rigorosas, recomenda-se o uso de *author_id* ou *author_orcid*, que se utilizam de identificadores persistentes alfanuméricos, em que o primeiro refere-se ao identificador da própria base e o segundo, ao número de ORCID do pesquisador.

Ademais, o dimensionamento e a densidade da rede são controlados pelos parâmetros *limit* e *min_weight*. O primeiro define o teto amostral de artigos, impactando a escala do grafo e o tempo de coleta. Já o *min_weight* atua como mecanismo de poda topológica: ao exigir um peso mínimo (ex: três colaborações) para formar

² Disponível em: <https://docs.openalex.org/api-entities/works/search-works>

arestas, preservam-se as relações mais fortes e eliminam-se ruídos. Computacionalmente, esse controle é vital, pois a geração de conexões possui complexidade combinatória $O(n^2)$ por artigo. Assim, o aumento do limite de registros implica em um crescimento quadrático no custo de processamento e alocação de memória.

Após a extração dos dados da API, a etapa seguinte consiste em mapear os dados retornados para a representação matemática de um grafo, denotada por $G = (V, E)$, em que V representa o conjunto de vértices do grafo, e E o conjunto de arestas. A NetFog realiza esse processo de instanciação dinamicamente e suporta três modelos de redes: Coautoria, Coocorrência de palavras-chave e Cocitação. No caso das redes de coautoria, o conjunto de vértices é composto pelos autores dos trabalhos recuperados. Uma aresta não direcionada é estabelecida entre dois vértices caso dois autores participem de uma mesma publicação, em que o peso da aresta deles é incrementado, refletindo a intensidade da colaboração dos autores.

As redes de coocorrência de palavras-chave são voltadas, sobretudo, para a análise temática dos artigos, em que as palavras-chaves dos artigos serão os vértices do grafo. Uma conexão é criada quando dois conceitos aparecem juntos nos metadados de um trabalho, em que o peso da aresta varia de acordo com a quantidade de coocorrências.

As redes de cocitação evidenciam a proximidade intelectual e teórica entre publicações, tendo como vértices os trabalhos citados conjuntamente nas referências analisadas. Devido à estrutura da API da OpenAlex, a geração dessa rede exige duas etapas. Inicialmente, extraem-se apenas os identificadores únicos (ex: W000000000) do campo *referenced_works*. Para obter os metadados completos, a NetFog isola esses IDs, elimina duplicatas e realiza requisições secundárias em lote. Após essa resolução, formam-se as arestas entre os trabalhos cocitados pela

mesma publicação de origem. Destaca-se que a ferramenta suporta duas abordagens: cocitação de documentos (vértices são títulos) e de autores (vértices são os respectivos nomes). A fim de demonstrar a nova funcionalidade da NetFog, será realizado o seguinte teste de uso, ilustrado na Figura 1, que explora o método criado em quatro cenários de extração, todos com limite amostral de 200 registros.

Figura 1 - Cenários de extração e modelagem na biblioteca NetFog

```
from netfog import Graph, OpenAlexGraphType

# 1. Grafo de coautoria baseado em um autor específico
grafo_coautoria = Graph.from_openalex(
    api_key="SUA_CHAVE_API_AQUI", graph_type=OpenAlexGraphType.Coauthorship,
    author="Jesus Mena-Chalco", limit=200)

# 2. Grafo de coocorrência de palavras-chave (intensidade mínima: 3)
grafo_coocorrencia = Graph.from_openalex(
    api_key="SUA_CHAVE_API_AQUI", graph_type=OpenAlexGraphType.KeywordCooccurrence,
    keyword="Bibliometrics", limit=200, min_weight=3)

# 3. Grafo de cocitação de documentos baseado em um termo de busca geral
grafo_cocitacao_documentos = Graph.from_openalex(
    api_key="SUA_CHAVE_API_AQUI", graph_type=OpenAlexGraphType.WorkCocitation,
    search="Bibliometrics", limit=200)

# 4. Grafo de cocitação de autores baseado em um termo de busca geral
grafo_cocitacao_autores = Graph.from_openalex(
    api_key="SUA_CHAVE_API_AQUI", graph_type=OpenAlexGraphType.AuthorCocitation,
    search="Bibliometrics", limit=200)
```

Fonte: Elaboração dos Autores

Inicialmente, gera-se uma rede de coautoria a partir de um autor específico parâmetro *author*. Neste caso, utilizou-se *author* = “Jesus Mena-Chalco”. Em seguida, estabeleceu-se uma busca exata em palavra-chave pelo termo “Bibliometrics” (parâmetro *keyword*=“Bibliometrics”) e constrói-se um grafo de coocorrência de palavras-chaves aplicando-se um filtro de peso mínimo (*min_weight* = 3) para preservar apenas os conceitos com intensidade de conexão igual à 3. Por fim, exploram-se os modelos de cocitação, instanciando-se separadamente tanto a rede de cocitação entre os documentos referenciados quanto a rede de cocitação entre os autores

citados na amostra, em ambos casos estabeleceu-se uma busca geral por “Bibliometrics” a partir do parâmetro *search = “Bibliometrics”*, diferenciando-se apenas a função para cocitação entre documentos (*graph_type = OpenAlexGraphType.WorkCocitation*) ou autores (*graph_type = OpenAlexGraphType.AuthorCocitation*).

Para avaliar o desempenho computacional, os quatro cenários foram executados três vezes, adotando-se a média aritmética para garantir a precisão dos dados. Durante estas execuções, foram mensuradas três métricas temporais distintas: o tempo de requisição à API (latência de rede), o tempo de processamento intrínseco da biblioteca e o tempo necessário para a geração da visualização do grafo. Os tempos médios resultantes dessa avaliação, segmentados conforme as etapas descritas, estão sumarizados no Quadro 2 a seguir.

Quadro 2 – Tempos médios de execução por cenário de rede

CENÁRIO	TEMPO REQ. API (S)	PROC. INTERNO (S)	VISUALIZAÇÃO (S)
Coautoria	3,669	0,001	0,12
Coocorrência	3,19	0,01	0,06
Cocitação de documentos	305,62	1,76	580,28
Cocitação de Autores	312,49	8,08	1732,97

Fonte: Elaboração dos Autores

Os dados observados no Quadro 2 revelam a alta escalabilidade da biblioteca. Em topologias menores, o processamento interno foi quase instantâneo (em torno de 0,01s). Já em cenários massivos, como a Cocitação de Autores com até 5,3 milhões de arestas, o gargalo concentrou-se inteiramente no cálculo de layout espacial, exigindo cerca de 1733 segundos. Sob essa mesma carga, o algoritmo da NetFog precisou de apenas 8 segundos, provando que o custo computacional reside primariamente na etapa visual. Todos os arquivos interativos contendo as visualizações geradas para todos os cenários testados foram disponibilizados publica-

mente e podem ser acessados via repositório digital Zenodo³. Adicionalmente, para fins de demonstração interativa, o grafo resultante do cenário de coautoria também se encontra acessível para exploração na plataforma RPubs⁴.

CONSIDERAÇÕES FINAIS

A pesquisa apresentou uma nova funcionalidade bibliométrica da biblioteca de grafos NetFog, voltada para extração, geração e visualização grafos científicos a partir de dados oriundos da base dados OpenAlex. Os testes validaram a implementação de um fluxo metodológico eficiente, operado em *Rust*, assegurando robustez computacional na transformação de dados brutos em estruturas topológicas complexas.

Os resultados evidenciaram que a NetFog supera limitações comuns de escalabilidade presentes em ferramentas tradicionais do campo dos EMI. A análise de desempenho demonstrou que o núcleo algorítmico da biblioteca processa topologias massivas rapidamente, isolando o custo computacional na etapa de renderização de layout. Conclui-se que a NetFog se consolida como uma infraestrutura promissora e escalável, garantindo maior flexibilidade, reprodutibilidade e automação na construção e análise de redes de coautoria, coocorrência de palavras-chave e cocitação a partir da OpenAlex, maior base dados do mundo de acesso aberto.

REFERÊNCIAS

GRÁCIO, Maria Cláudia Cabrini; OLIVEIRA, Ely Francina Tannuri. Análise de redes sociais para visualização do comportamento científico. In: GRÁCIO, M. C. C.; MARTÍNEZ-ÁVILA, D.; OLIVEIRA, E. F. T.; ROSAS, F. S. (org.). **Tópicos da bibliometria para bibliotecas universitárias**. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2020. p. 164-191. DOI: <https://doi.org/10.36311/2020.978-65-86546-91-0.p164-191>. Acesso em 4 mar. 2026

³ Disponível em: <https://zenodo.org/records/18986283>

⁴ Disponível em: <https://rpubs.com/ticofogolin/1427410>



ROSSI, Luciano; MENA-CHALCO, Jesus Pascual. Criação de grafos de tópicos do conhecimento baseada em genealogia acadêmica. **Encontro Brasileiro de Bibliometria e Cientometria**, v. 6, p. 353-363, 2018. DOI: <https://doi.org/10.22477.045656> Acesso em 4 mar. 2026.

CASTANHA, Rafael Gutierrez. The Coupler: uma nova ferramenta bibliométrica para análises relacionais de citação, acoplamento bibliográfico e cocitação. **RDBCI**, Campinas. v. 20, p. e022029, 2022. DOI: <https://doi.org/10.20396/rdbci.v20i00.8671208>. Acesso em 4 mar. 2026