



DISSERTAÇÃO DE MESTRADO PROFISSIONAL

**Rastreado a inovação industrial: análise de redes sociais
de artigos e patentes de SAF assistida por LLM**

Matheus Noschang de Oliveira

Programa de Pós-Graduação Profissional em Engenharia Elétrica
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
FACULDADE DE TECNOLOGIA
UNIVERSIDADE DE BRASÍLIA

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

**Tracking Industrial Innovation: Social Network Analysis of SAF Articles and Patents aided by
LLM**

Matheus Noschang de Oliveira

Orientador: Prof. Dr. André Luiz Marques Serrano, PPEE/UnB

PUBLICAÇÃO: PPEE.MP.089

BRASÍLIA-DF

UNIVERSIDADE DE BRASÍLIA
Faculdade de Tecnologia

DISSERTAÇÃO DE MESTRADO PROFISSIONAL

**Rastreando a inovação industrial: análise de redes sociais
de artigos e patentes de SAF assistida por LLM**

Matheus Noschang de Oliveira

*Dissertação de Mestrado Profissional submetida ao Departamento de Engenharia
Elétrica como requisito parcial para obtenção
do grau de Mestre em Engenharia Elétrica*

Banca Examinadora

Prof. Dr. André Luiz Marques Serrano, _____
EPR/FT/UnB
Orientador

Prof. Dr. Vinícius Pereira Gonçalves, ENE/FT/UnB _____
Examinador interno

Prof. Dr. Rodrigo Bonacin, Centro Universitário _____
Campo Limpo Paulista (UNIFACCAMP)
Examinador externo

Prof. Dr. Clóvis Neumann, EPR/FT/UnB _____
Suplente

FICHA CATALOGRÁFICA

DE OLIVEIRA, MATHEUS NOSCHANG

Rastreando a inovação industrial: análise de redes sociais de artigos e patentes de SAF assistida por LLM [Distrito Federal] 2025.

xvi, 54 p., 210 x 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2025).

Dissertação de Mestrado Profissional - Universidade de Brasília, Faculdade de Tecnologia.

Departamento de Engenharia Elétrica

- | | |
|----------------|------------------------|
| 1. LLM | 2. Social Networks |
| 3. SAF Patents | 4. SAF Research Papers |
| I. ENE/FT/UnB | II. Título (série) |

REFERÊNCIA BIBLIOGRÁFICA

DE OLIVEIRA, M.N. (2025). *Rastreando a inovação industrial: análise de redes sociais de artigos e patentes de SAF assistida por LLM*. Dissertação de Mestrado Profissional, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 54 p.

CESSÃO DE DIREITOS

AUTOR: Matheus Noschang de Oliveira

TÍTULO: Rastreando a inovação industrial: análise de redes sociais de artigos e patentes de SAF assistida por LLM.

GRAU: Mestre em Engenharia Elétrica ANO: 2025

É concedida à Universidade de Brasília permissão para reproduzir cópias desta Dissertação de Mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. Do mesmo modo, a Universidade de Brasília tem permissão para divulgar este documento em biblioteca virtual, em formato que permita o acesso via redes de comunicação e a reprodução de cópias, desde que protegida a integridade do conteúdo dessas cópias e proibido o acesso a partes isoladas desse conteúdo. O autor reserva outros direitos de publicação e nenhuma parte deste documento pode ser reproduzida sem a autorização por escrito do autor.

Matheus Noschang de Oliveira
Depto. de Engenharia Elétrica (ENE) - FT
Universidade de Brasília (UnB)
Campus Darcy Ribeiro
CEP 70919-970 - Brasília - DF - Brasil

DEDICATÓRIA

A todas as pessoas de boa fé que acreditam ainda haver tempo para o combate à crise ambiental.

“Se vi mais longe, foi por estar sobre os ombros de gigantes.”

Isaac Newton

AGRADECIMENTOS

Agradeço a Deus, por me ter concedido saúde, disposição e discernimento para enfrentar os mais diversos desafios que surgiram ao longo desta trajetória acadêmica.

Aos meus pais, que me deram todo o suporte para que eu pudesse me concentrar no desenvolvimento desta pesquisa.

Ao meu filho, Yuri, que foi a força-motriz para a conclusão deste curso.

Ao meu orientador, André, que me conduziu com bom grado para muito além de sua alçada, me habilitou e me proporcionou insumos para que esta pesquisa atingisse seu objetivo.

Aos meus colegas pesquisadores do grupo *Projectum*, que me ajudaram imensamente na produção dos artigos publicados para este trabalho e me ensinaram muito sobre o mundo acadêmico.

Aos colegas de classe, que agraciaram meus dias e me transmitiram muitas lições de vida. A academia é verdadeiramente um ambiente fantástico para criar conexões e trocar experiências.

À Confederação Nacional da Indústria (CNI), pelo apoio na condução desta pesquisa.

RESUMO

Esta dissertação investiga a dinâmica de inovação dos Combustíveis Sustentáveis de Aviação (SAF) por meio de uma abordagem integrada, que combina análise de redes de patentes, teoria de redes sociais e técnicas de aprendizado de máquina. Ao unir perspectivas tecnológicas e institucionais, o estudo revela barreiras críticas que dificultam a conversão da expertise brasileira consolidada em biocombustíveis em uma posição de liderança em inovação em SAF.

A análise das redes globais de patentes em SAF identifica três principais agrupamentos tecnológicos — hidrocessamento de lipídios, conversão termoquímica via gaseificação e síntese de Fischer–Tropsch, e rotas eletroquímicas emergentes — nos quais os atores brasileiros permanecem sub-representados. Análises complementares de redes de co-invenção e coautoria evidenciam uma desconexão estrutural entre a pesquisa acadêmica e o desenvolvimento industrial no Brasil, em contraste com as colaborações universidade–indústria integradas observadas nos Estados Unidos e na Europa.

Para enfrentar essas lacunas, desenvolveu-se um sistema de pareamento entre patentes e artigos, baseado em arquiteturas de transformadores e em um grafo de conhecimento específico do domínio, alcançando 89% de acurácia na identificação de domínios de inovação compartilhados. Esse sistema foi integrado a uma ferramenta de apoio à decisão, baseada em modelos de linguagem de grande porte (LLM), permitindo a identificação de trajetórias de colaboração de alto potencial — como o redirecionamento da expertise brasileira em etanol para rotas de álcool-para-querosene (ATJ).

As descobertas fundamentam recomendações políticas-chave, incluindo a necessidade de marcos regulatórios híbridos, incentivos direcionados à inovação e estratégias de colaboração internacional. O estudo contribui metodologicamente ao demonstrar o valor da combinação entre análise de redes multicamadas e aprendizado de máquina para o mapeamento de ecossistemas de inovação sustentável.

Trabalhos futuros devem expandir o arcabouço para redes dinâmicas e multicamadas, incorporar modelagem tecnoeconômica e ampliar o escopo geográfico da análise. Em última instância, a pesquisa oferece insights teóricos e práticos para acelerar a transição para combustíveis de aviação de baixo carbono, enfatizando que soluções tecnológicas só são eficazes quando inseridas em redes institucionais e colaborativas robustas.

Palavras-chave: LLM; Redes Sociais; Patentes de SAF; Artigos de SAF

ABSTRACT

This dissertation investigates the innovation dynamics of Sustainable Aviation Fuels (SAF) through an integrated approach that combines patent network analysis, social network theory, and machine learning techniques. By bridging technological and institutional perspectives, the study reveals critical barriers that hinder the transformation of Brazil’s established biofuel expertise into a leadership position in SAF innovation.

The analysis of global SAF patent networks identifies three main technological clusters — hydroprocessing of lipids, thermochemical conversion via gasification and Fischer–Tropsch synthesis, and emerging electrochemical pathways — in which Brazilian actors remain underrepresented. Complementary analyses of co-inventorship and co-authorship networks highlight a structural disconnect between academic research and industrial development in Brazil, in contrast to the integrated university–industry collaborations observed in the United States and Europe.

To address these gaps, the study develops a patent–article matching system based on transformer architectures and a domain-specific knowledge graph, achieving 89% accuracy in identifying shared innovation domains. This system is integrated into a decision support tool based on large language models (LLMs), enabling the identification of high-potential collaboration pathways — such as redirecting Brazil’s ethanol expertise toward alcohol-to-jet (ATJ) routes.

The findings underpin key policy recommendations, including the need for hybrid regulatory frameworks, targeted innovation incentives, and international collaboration strategies. Methodologically, the study demonstrates the value of combining multilayer network analysis with machine learning to map sustainable innovation ecosystems.

Future work should extend the framework to dynamic and multilayer networks, incorporate techno-economic modeling, and broaden the geographical scope of analysis. Ultimately, the research offers both theoretical and practical insights for accelerating the transition to low-carbon aviation fuels, emphasizing that technological solutions are only effective when embedded in robust institutional and collaborative networks.

Keywords: LLM; Social Networks; SAF Patents; SAF Research Papers

SUMÁRIO

1	INTRODUÇÃO	1
1.1	CONTEXTO E MOTIVAÇÃO DA PESQUISA	2
1.2	OBJETIVOS	3
1.3	CONTRIBUIÇÕES	4
1.4	ESTRUTURA DA DISSERTAÇÃO	5
1.5	REVISÃO DE LITERATURA	7
2	RASTREANDO A INOVAÇÃO EM BIOCOMBUSTÍVEIS: UMA ANÁLISE BASEADA EM GRAFOS DE PATENTES DE COMBUSTÍVEL SUSTENTÁVEL PARA AVIAÇÃO	10
2.1	COLETA DE DADOS	10
2.2	EXTRAÇÃO E PRÉ-PROCESSAMENTO DOS DADOS	11
2.3	GERAÇÃO DE GRAFOS	11
2.3.1	REDE IPC-ASSIGNEE	11
2.3.2	REDE IPC-PATENTE-ASSIGNEE	11
2.3.3	REDE IPC-ASSIGNEE-WRITTEN OPINION	13
2.4	BIBLIOTECAS E FERRAMENTAS PYTHON PARA ANÁLISE DE GRAFOS	13
2.5	ANÁLISE E RESULTADOS	13
2.5.1	CONTEXTUALIZAÇÃO DOS DADOS DE PATENTES DE SAF	13
2.5.2	ANÁLISE DE REDES COMPLEXAS DE SAF	18
3	O CAMINHO DO BRASIL PARA CÉUS VERDES: ANÁLISE DE COMBUSTÍVEIS SUSTENTÁVEIS PARA AVIAÇÃO POR MEIO DE PATENTES E ARTIGOS UTILIZANDO TEORIA DE REDES	26
3.1	COLETA DE DADOS	27
3.1.1	RECUPERAÇÃO E PRÉ-PROCESSAMENTO DOS DADOS	28
3.2	ANÁLISE DE GRAFOS	29
3.2.1	REDES BIPARTIDAS	29
3.2.2	REDE DE AFILIAÇÕES–DEPOSITANTES–ÁREAS DE PESQUISA	29
3.2.3	REDE DE ORGANIZAÇÕES FINANCIADORAS–DEPOSITANTES–ÁREAS DE PESQUISA	30
3.2.4	REDE DE ISICs DE ORGANIZAÇÕES FINANCIADORAS BRASILEIRAS–DEPOSITANTES–ÁREAS DE PESQUISA	30
3.2.5	REDES MULTICAMADAS	30
3.2.6	BIBLIOTECAS PYTHON E FERRAMENTAS DE VISUALIZAÇÃO	30
3.3	CORRESPONDÊNCIA PATENTE-ARTIGO	32
3.4	ESTRUTURA DA REDE SOCIAL	32
3.5	OPORTUNIDADES POTENCIAIS NO MERCADO BRASILEIRO	35
4	DO DADO À DECISÃO: ANÁLISE DE GRAFOS COM MODELOS DE LINGUAGEM	39

4.1	LEIAUTE DA SOLUÇÃO	39
4.2	INTEGRAÇÃO ENTRE VISUALIZAÇÃO DE REDES E MODELOS DE LINGUAGEM	40
5	CONCLUSÃO	47
5.1	TRABALHOS FUTUROS	48
	REFERÊNCIAS	50

LISTA DE FIGURAS

2.1	Framework Metodológico	14
2.2	Evolução das Patentes Solicitadas.....	14
2.3	Patentes por País.....	15
2.4	Perfil tecnológico por país.....	15
2.5	Patentes solicitadas por país por ano.....	16
2.6	Porcentagem das áreas tecnológicas segundo a IPC	17
2.7	Evolução das áreas tecnológicas da IPC nas patentes solicitadas.....	17
2.8	Visão geral do grafo de rede social IPC/Cessionário	18
2.9	Domínio da Rolls-Royce e nós folha de universidades (recorte/ampliação)	19
2.10	IPC mais visada em SAF (recortada/ampliada)	20
2.11	Família de patentes mais relevante do C10G destacada.....	21
2.12	Cluster da Rolls-Royce.....	23
2.13	Rede IPC-Assignee-Opinião Escrita.....	24
2.14	Pedidos próximos da expiração	25
2.15	Colaboração entre universidade e empresa em múltiplas IPCs.....	25
3.1	Framework Metodológico	26
3.2	Histograma das áreas de pesquisa relacionadas aos artigos consultados.....	31
3.3	Áreas de pesquisa com maior ocorrência utilizadas para o treinamento dos modelos	32
3.4	Matrizes de Confusão para todos os modelos treinados.....	33
3.5	Visão geral da rede de responsáveis por patentes e afiliações de artigos.....	34
3.6	Rede em camadas de organizações financiadoras brasileiras e responsáveis por patentes.....	35
3.7	Rede de organizações financiadoras brasileiras e responsáveis por patentes.....	36
3.8	Rede de organizações financiadoras brasileiras e responsáveis por patentes com foco em RORO-C	37
3.9	Rede de responsáveis por patentes e afiliações de artigos filtrada pelos ISICs correspondentes das organizações financiadoras brasileiras	37
4.1	Framework do aplicativo web	40
4.2	Dashboard de análise de redes assistida por LLM.....	43

LISTA DE TABELAS

1.1	Dimensões Integradas da Inovação em SAF	8
2.1	Descrição dos códigos Derwent.....	12
2.2	Top 5 medidas de centralidade de grau para nós de IPC	20
2.3	Top 5 medidas de centralidade de grau para nós de depositários	21
2.4	5 Maiores Valores de Centralidade de Grau para Nós de Patente.....	22
3.1	Três maiores medidas de centralidade de grau para nós de Área de Pesquisa na rede de responsáveis por patentes e afiliações de artigos.....	34
3.2	Top 3 medidas de centralidade de grau para nós de Área de Pesquisa da rede de organizações financiadoras brasileiras e responsáveis por patentes	38
4.1	Descrição das variáveis utilizadas no prompt de análise de redes científicas	42

1 INTRODUÇÃO

O imperativo global de descarbonizar os sistemas energéticos colocou as tecnologias sustentáveis na vanguarda da inovação, sendo a aviação um dos setores mais desafiadores a ser transformado devido aos seus rigorosos requisitos de combustível e limitações técnicas. Esta tese desenvolve uma metodologia inovadora para analisar e acelerar o desenvolvimento tecnológico através do exame sistemático de redes de inovação, utilizando os Combustíveis Sustentáveis de Aviação (Sustainable Aviation Fuel - SAF) como um estudo de caso crítico que exemplifica tanto o potencial quanto os desafios das transições energéticas sustentáveis (Akpan and Olanrewaju, 2023; Vardon et al., 2022). Ao integrar análise de patentes, classificação por aprendizado de máquina e teoria de redes sociais, criamos uma estrutura abrangente de apoio à decisão capaz de identificar tecnologias emergentes, mapear fluxos de conhecimento e revelar oportunidades estratégicas de parceria entre setores e regiões geográficas (Soofi and Abdi, 2020; Nicola et al., 2023).

No cerne desta pesquisa reside um desafio fundamental na gestão tecnológica contemporânea: a crescente complexidade dos ecossistemas de inovação, onde avanços ocorrem rotineiramente na intersecção de múltiplas disciplinas científicas e setores industriais (Wang et al., 2021). Os métodos tradicionais de prospecção tecnológica, que frequentemente examinam inovações de forma isolada, mostram-se inadequados para capturar essas relações multidimensionais, resultando frequentemente em decisões de investimento subótimas e na adoção tardia de soluções sustentáveis (Stergiou et al., 2023). Nosso framework metodológico aborda essas limitações através de um processo analítico trifásico que inicia com o mapeamento de redes de patentes utilizando o Derwent Innovation Index (DII) como fonte primária de dados. Ao construir redes temporais de patentes em janelas de três anos e aplicar algoritmos de teoria dos grafos, identificamos titulares dominantes e suas relações colaborativas, revelando a estrutura subjacente dos ecossistemas de inovação de maneira que transcende meras contagens de patentes (Ang et al., 2024). Esta abordagem desvenda padrões cruciais no desenvolvimento tecnológico, incluindo clusters emergentes de titulares com capacidades complementares, lacunas persistentes entre pesquisa fundamental e tecnologias aplicadas, e tendências geográficas significativas na difusão de inovações.

A segunda fase de nossa metodologia concentra-se na integração de conhecimento entre domínios, empregando classificadores de aprendizado de máquina treinados com metadados da literatura acadêmica para estabelecer vínculos bidirecionais entre documentos de patentes e áreas de pesquisa científica (Tsan-kov, 2021). Esse processo gera uma tabela de correspondência robusta, que classifica sistematicamente as patentes em domínios de pesquisa específicos — como catálise e processamento de matérias-primas —, identifica instituições acadêmicas que atuam em desafios técnicos relacionados e quantifica lacunas críticas na tradução do conhecimento científico em aplicações tecnológicas. A síntese resultante das redes de patentes e publicações acadêmicas proporciona uma visibilidade inédita sobre os processos, muitas vezes opacos, que regem a difusão da inovação.

Para sua implementação prática, desenvolvemos um sistema interativo de apoio à decisão que processa o grafo de conhecimento sintetizado por meio de uma interface baseada em Modelo de Linguagem de Grande Escala (Large Language Model - LLM) (Molina et al., 2020). Essa ferramenta inovadora permite que diferentes partes interessadas consultem a rede utilizando linguagem natural — por exemplo, para

identificar instituições brasileiras especializadas em catalisadores HEFA (Hydroprocessed Esters and Fatty Acids) —, além de possibilitar a visualização de potenciais caminhos de colaboração e a simulação de diversos cenários de investimento.

Quando aplicada às tecnologias de Sustainable Aviation Fuel (SAF), essa estrutura integrada revela percepções marcantes sobre os padrões globais de inovação, com destaque para a posição paradoxal do Brasil: apesar de ser um líder mundial na produção de biocombustíveis, o país não conseguiu estabelecer presença significativa no desenvolvimento de SAF, mesmo dispondo de abundantes recursos renováveis e expertise técnica relevante (Caldeira et al., 2016; Mosquera et al., 2024). Nossa análise identifica gargalos específicos no sistema nacional de inovação, incluindo a fraca articulação entre indústria e academia na pesquisa de biocombustíveis avançados, o desalinhamento entre as estratégias de patenteamento e as capacidades tecnológicas domésticas, e redes de colaboração pouco desenvolvidas com atores globais do setor de SAF (Bailis and Baka, 2010).

O desafio da descarbonização do setor de aviação exemplifica a complexa interação entre viabilidade tecnológica, viabilidade econômica e apoio institucional, característica das transições contemporâneas rumo à sustentabilidade (Abrantes et al., 2021). Por ser uma solução do tipo *drop-in* que requer modificações mínimas na infraestrutura existente, o SAF representa a via mais pragmática para a redução das emissões da aviação no curto e médio prazo — embora sua concretização exija abordagens mais sofisticadas de gestão da inovação (Yusaf et al., 2022).

Nossa metodologia oferece as ferramentas analíticas necessárias para preencher lacunas críticas entre a pesquisa técnica em nível micro — como o desenvolvimento de catalisadores —, as capacidades industriais em nível meso — incluindo a integração de biorrefinarias — e os marcos regulatórios em nível macro — como os mandatos de uso de combustíveis renováveis (de Oliveira et al., 2024). Ao iluminar essas conexões e fornecer inteligência acionável para tomadores de decisão, esta pesquisa contribui tanto para o avanço do entendimento acadêmico sobre ecossistemas de inovação quanto para os esforços práticos de aceleração da transição rumo à aviação sustentável.

1.1 CONTEXTO E MOTIVAÇÃO DA PESQUISA

A urgência da descarbonização do setor aéreo tem sido intensificada por descobertas recentes da ciência climática e por avanços nas políticas internacionais. A Conferência das Nações Unidas sobre Mudança do Clima (Conference of the Parties - COP26) estabeleceu metas ambiciosas para a redução das emissões da aviação, com o Combustível Sustentável de Aviação (Sustainable Aviation Fuel - SAF) ocupando papel central na maioria dos roteiros nacionais e setoriais (Stergiou et al., 2023). Paralelamente, a instabilidade geopolítica dos últimos anos ressaltou os riscos associados à dependência de combustíveis fósseis, criando motivações adicionais — tanto econômicas quanto estratégicas — para o desenvolvimento de alternativas sustentáveis (Chou et al., 2023).

O Brasil configura um estudo de caso particularmente relevante devido à sua posição paradoxal no cenário do SAF. Sendo o segundo maior produtor mundial de biocombustíveis e possuindo tecnologias agrícolas e energéticas avançadas, o país reúne todos os componentes necessários para exercer liderança

nesse campo (Ray, 2019). No entanto, apesar de seu potencial, entidades brasileiras têm estado notavelmente ausentes nos depósitos de patentes relacionadas ao SAF, o que sugere a existência de barreiras significativas à transferência de inovação entre o setor nacional de biocombustíveis já consolidado e as tecnologias emergentes voltadas à aviação sustentável (Nicola et al., 2023).

Essa lacuna de pesquisa fundamenta a adoção de uma abordagem analítica dual nesta investigação:

- Primeiramente, empregamos a teoria dos grafos para mapear as redes globais de inovação em SAF a partir da análise de patentes, identificando os principais atores, aglomerados tecnológicos e fluxos de conhecimento na área (Soofi and Abdi, 2020);
- Em segundo lugar, aplicamos a Análise de Redes Sociais (Social Network Analysis - SNA) para examinar os padrões de colaboração entre instituições de pesquisa brasileiras e desenvolvedores internacionais de SAF, buscando elucidar os motivos da posição relativamente atrasada do país, apesar de seus fundamentos favoráveis (Ang et al., 2024).

1.2 OBJETIVOS

Esta dissertação persegue três objetivos interconectados que, em conjunto, ampliam nossa capacidade de compreender e acelerar a inovação sustentável no setor de aviação. Em primeiro lugar, busca-se estabelecer uma metodologia inédita para o monitoramento do desenvolvimento tecnológico que transcenda as análises tradicionais de patentes, integrando a ciência de redes quantitativa com a expertise qualitativa em domínios técnicos. Diferentemente das abordagens convencionais, que examinam patentes de forma isolada, nosso arcabouço analítico captura a interação dinâmica entre inventores, instituições e áreas de pesquisa, revelando não apenas quais tecnologias estão emergindo, mas, sobretudo, como o conhecimento circula dentro dos ecossistemas de inovação. Essa inovação metodológica permite que tomadores de decisão identifiquem tecnologias promissoras em estágios iniciais de desenvolvimento e com maior confiança quanto à sua viabilidade comercial (Soofi and Abdi, 2020; Ang et al., 2024).

Com base nesse fundamento, o segundo objetivo consiste em investigar o paradoxo da posição brasileira no desenvolvimento de Sustainable Aviation Fuel (SAF). Apesar de possuir capacidades de classe mundial na produção de biocombustíveis e abundância de recursos renováveis, o Brasil permanece notavelmente ausente do cenário global de patentes em SAF (Caldeira et al., 2016; Mosquera et al., 2024). Por meio da análise sistemática de redes de colaboração e transbordamentos de conhecimento, buscamos elucidar as barreiras estruturais que têm impedido esse líder em bioenergia de se posicionar estrategicamente na revolução do SAF. Nossa investigação vai além de explicações superficiais, revelando a complexa interação entre fatores tecnológicos, econômicos e institucionais que restringem a participação do Brasil nesse mercado em expansão (Nicola et al., 2023).

O terceiro e mais transformador objetivo visa a ponte entre a pesquisa acadêmica e a tomada de decisão prática, por meio do desenvolvimento de uma ferramenta analítica interativa que democratiza o acesso a dados complexos de inovação. Ao combinar ciência de redes com interfaces baseadas em modelos de linguagem de grande escala (Large Language Models - LLMs), criamos uma plataforma que permite a

formuladores de políticas realizar perguntas em linguagem natural sobre tendências tecnológicas e obter recomendações baseadas em evidências, adaptadas ao seu contexto regional e às suas capacidades industriais (Molina et al., 2020). Trata-se de um avanço significativo em relação a relatórios estáticos e ferramentas convencionais de inteligência de mercado, ao oferecer um sistema dinâmico que aprende com as interações dos usuários e se adapta continuamente aos padrões emergentes de inovação.

Em conjunto, esses objetivos conformam uma agenda de pesquisa coesa que não apenas avança o entendimento acadêmico sobre sistemas de inovação, mas também entrega ferramentas práticas para acelerar a transição rumo a uma aviação sustentável. Ao mapear o panorama global da inovação em SAF ao mesmo tempo em que enfrenta os desafios específicos do Brasil, esta dissertação propõe um modelo adaptável a outras tecnologias sustentáveis e a diferentes contextos regionais, contribuindo, assim, para estratégias mais eficazes de mitigação climática em setores industriais de difícil descarbonização (Stergiou et al., 2023; de Oliveira et al., 2024).

1.3 CONTRIBUIÇÕES

Esta dissertação apresenta contribuições substanciais nos domínios teórico, metodológico e aplicado, promovendo avanços tanto para a literatura acadêmica quanto para a prática em estudos de inovação voltados a tecnologias sustentáveis. No plano teórico, propomos uma reconceituação fundamental de como os ecossistemas de inovação são analisados, ao desenvolver o primeiro arcabouço integrado que combina análise de redes de patentes, classificação de áreas de pesquisa e redes sociais colaborativas. Essa abordagem tripartite supera as limitações de estudos anteriores que trataram essas dimensões de forma isolada, permitindo capturar os complexos ciclos de retroalimentação entre desenvolvimento tecnológico, pesquisa científica e colaboração industrial (Soofi and Abdi, 2020; Nicola et al., 2023). Nossos achados desafiam a visão tradicional de processos lineares de inovação, revelando, em seu lugar, a natureza emergente e em rede do progresso tecnológico em combustíveis sustentáveis para a aviação.

Do ponto de vista metodológico, introduzimos diversas técnicas inovadoras que estabelecem novos padrões para análises de inovação. O desenvolvimento de nosso classificador baseado em aprendizado de máquina, que conecta documentos de patentes com áreas de pesquisa acadêmica, representa um avanço significativo na cartografia do conhecimento, solucionando o desafio persistente de quantificar as conexões entre descobertas científicas e aplicações comerciais (Tsankov, 2021). Igualmente inovador é nosso sistema de interrogação de redes baseado em modelos de linguagem de grande escala (Large Language Models - LLMs), que transforma dados estáticos de patentes em ferramentas dinâmicas de apoio à decisão. Esses avanços metodológicos oferecem aos pesquisadores capacidades inéditas para rastrear fluxos de conhecimento, identificar gargalos de inovação e prever trajetórias tecnológicas emergentes com maior precisão (Ang et al., 2024; Molina et al., 2020).

Para formuladores de políticas públicas e líderes industriais, esta pesquisa oferece ferramentas práticas para navegar no complexo panorama da inovação em SAF. Nossa análise da posição brasileira no desenvolvimento global de SAF fornece recomendações específicas e acionáveis para integrar a expertise nacional em biocombustíveis às tecnologias emergentes de aviação (Caldeira et al., 2016; Mosquera et al.,

2024). A plataforma interativa desenvolvida permite a exploração em tempo real de oportunidades de colaboração, reduzindo significativamente o tempo e a incerteza associados ao planejamento estratégico de P&D. Mais importante ainda, demonstramos como economias emergentes com forte base agrícola podem alavancar capacidades existentes para ingressar em mercados tecnológicos sustentáveis de alto valor — uma constatação com implicações profundas para a justiça climática global e para transições energéticas mais equitativas (Stergiou et al., 2023).

Para além do contexto específico dos combustíveis sustentáveis para a aviação, esta pesquisa estabelece um arcabouço replicável para análise de sistemas de inovação em outros domínios críticos da sustentabilidade. Os princípios que fundamentam nossa metodologia — combinação de análise de redes quantitativa com expertise qualitativa e visualização interativa — podem ser adaptados para o estudo de tecnologias em energias renováveis, soluções de economia circular e outros ecossistemas de inovação enfrentando desafios urgentes de escalabilidade (Wang et al., 2021; de Oliveira et al., 2024). Dessa forma, as contribuições desta dissertação extrapolam os estudos de caso analisados, oferecendo tanto uma nova lente para compreender a mudança tecnológica quanto ferramentas concretas para acelerar a transição rumo a sistemas industriais sustentáveis.

1.4 ESTRUTURA DA DISSERTAÇÃO

Esta dissertação se desenvolve como uma jornada intelectual cuidadosamente estruturada, que progride desde os conceitos fundamentais até inovações metodológicas e, por fim, aplicações práticas. O percurso tem início no Capítulo 2, com uma síntese abrangente de três vertentes críticas da literatura que fundamentam este trabalho: tecnologias de combustíveis sustentáveis para a aviação (SAF), teoria de redes de inovação e o panorama bioenergético brasileiro. Em vez de apenas revisar os estudos existentes, este capítulo constrói ativamente o arcabouço conceitual que orienta nossas análises subsequentes, revelando lacunas importantes na interseção entre esses domínios de pesquisa que nossa metodologia buscará preencher (Akpan and Olanrewaju, 2023; Abrantes et al., 2021).

O Capítulo 3 constitui o núcleo metodológico da pesquisa, onde transformamos conceitos teóricos em ferramentas analíticas operacionais. Aqui apresentamos nossa metodologia integrada de três etapas não como procedimentos isolados, mas como componentes interconectados de um único arcabouço analítico. O capítulo detalha nossa abordagem inovadora para construção de redes de patentes, demonstrando como superamos desafios comuns de limpeza de dados e desambiguação de entidades (Soofi and Abdi, 2020). Em seguida, explica-se a arquitetura do classificador de áreas de pesquisa baseado em aprendizado de máquina, incluindo seu protocolo de treinamento e métricas de validação. Por fim, introduz-se o design de nossa interface baseada em LLMs (Large Language Models), que torna a análise de redes complexas acessível a tomadores de decisão não técnicos (Molina et al., 2020).

O coração analítico da dissertação pulsa no Capítulo 4, onde aplicamos nosso arcabouço para revelar padrões até então ocultos na inovação em SAF. Este capítulo vai além da simples apresentação de resultados, ao contar a história envolvente das oportunidades perdidas e do potencial latente do Brasil em combustíveis sustentáveis para a aviação (Caldeira et al., 2016; Mosquera et al., 2024). Por meio de uma série

de análises interconectadas, mapeamos inicialmente o panorama global de inovação em SAF, para então focar na posição do Brasil dentro dessas redes e, por fim, identificar os entraves institucionais e estruturais específicos que têm limitado o desenvolvimento nacional nessa área. O capítulo apresenta visualizações ricas de redes de colaboração e mapas patente-tecnologia que trazem essas relações complexas à tona.

O Capítulo 5 eleva a discussão dos achados empíricos para implicações mais amplas, articulando como nossos resultados desafiam o senso comum sobre transferência de tecnologia em economias emergentes (Nicola et al., 2023). Este capítulo final não se limita a resumir os resultados, mas constrói uma agenda prospectiva que conecta nossas inovações metodológicas com aplicações práticas para formuladores de políticas públicas, líderes do setor produtivo e pesquisadores. Identificamos caminhos concretos para que o Brasil possa alavancar sua expertise em bioenergia no mercado de SAF, ao mesmo tempo em que delineamos como nosso arcabouço analítico pode ser adaptado a outras tecnologias sustentáveis e a diferentes contextos regionais (Stergiou et al., 2023; de Oliveira et al., 2024). O capítulo conclui discutindo as limitações do estudo não como lacunas, mas como oportunidades produtivas para futuras pesquisas que queiram expandir este trabalho.

1.5 REVISÃO DE LITERATURA

A urgência por combustíveis sustentáveis para aviação (Sustainable Aviation Fuel - SAF) surge dos desafios únicos de descarbonização do setor aéreo, nos quais as alternativas convencionais de biocombustíveis se mostram tecnologicamente inadequadas devido a limitações químicas fundamentais. Etanol e biodiesel contêm moléculas oxigenadas que reduzem a eficiência da combustão enquanto aumentam os riscos de corrosão e instabilidade em baixas temperaturas—limitações críticas para aplicações na aviação (Gray et al., 2021). Esse impasse tecnológico posicionou os SAF como a solução viável no curto prazo, com sete rotas de produção distintas recebendo certificação da ASTM desde 2009 (United State's Department of Energy, 2023). No entanto, a adoção em larga escala continua limitada por barreiras econômicas significativas, com custos de produção atualmente entre 3 a 6 vezes superiores aos do querosene convencional, agravados por processos de certificação rigorosos que exigem de 3 a 5 anos e mais de US\$ 5 milhões em custos de testes (Cabrera and de Sousa, 2022; A.Qasem et al., 2024). Esses desafios são particularmente agudos em economias emergentes como o Brasil, onde, apesar da liderança mundial em etanol de cana-de-açúcar e na produção de biodiesel (Mosquera et al., 2024; Jha and Schmidt, 2021), o desenvolvimento de SAF permanece notavelmente atrasado em relação às tendências globais de inovação.

As rotas tecnológicas para a produção de SAF revelam compensações complexas entre disponibilidade de matéria-prima, custos de produção e benefícios ambientais. A rota de ésteres e ácidos graxos hidroprocessados (Hydroprocessed Esters and Fatty Acids - HEFA) atualmente domina a produção comercial, gerando mais de 1000 litros por tonelada de matéria-prima a aproximadamente US\$ 1,25 por litro como preço mínimo de venda do combustível (MFSP), enquanto alternativas emergentes, como a conversão de açúcares fermentados hidroprocessados em isoparafinas sintéticas (Hydroprocessed Fermented Sugars to Synthetic Iso-Paraffins - HFS-SIP), alcançam preços próximos de US\$ 4,56 por litro (Ng et al., 2021). Resíduos sólidos urbanos e florestais oferecem o maior potencial de redução de gases de efeito estufa (20–95% em comparação ao querosene convencional) por meio da síntese de Fischer-Tropsch, embora limitações na cadeia de suprimentos restrinjam sua escalabilidade (Ângelo et al., 2009; Kandaramath Hari et al., 2015). Esse nexos econômico-tecnológico reforça a importância crítica da análise de redes de inovação para otimizar investimentos em P&D e intervenções de política pública.

A teoria dos grafos evoluiu desde suas origens no século XVIII, com a solução de Euler para o problema das pontes de Königsberg (Henning and van Vuuren, 2022), para se tornar uma ferramenta indispensável na análise de redes complexas de patentes. Aplicações contemporâneas demonstram notável sofisticação, empregando algoritmos de clusterização de grafos para identificar aglomerados de inovação (Maskittou et al., 2021), técnicas de embedding de grafos para análise competitiva (Wang et al., 2020), e padrões frequentes de subgrafos para análise semântica de patentes (Ding et al., 2019). Esses métodos avançados enfrentam desafios persistentes na análise de patentes, incluindo complexidade técnica, crescimento exponencial do volume de dados e limitações inerentes às abordagens convencionais de mineração de texto (Björkqvist and Kallio, 2023). Entre as métricas de centralidade, a centralidade de grau tem se mostrado particularmente valiosa para mapear relações entre códigos da Classificação Internacional de Patentes (International Patent Classification - IPC) em redes de inovação em SAF, revelando trajetórias tecnológicas dominantes (Metcalf and Casey, 2016).

A análise de redes sociais oferece insights complementares sobre as dinâmicas de colaboração que

moldam ecossistemas de inovação. Estudos identificam consistentemente vínculos fracos entre indústria e academia como uma barreira crítica em economias emergentes (Nicola et al., 2023), ao mesmo tempo em que demonstram a maior eficácia de mandatos de mistura em comparação a mecanismos de adoção voluntária (Mäki-Arvela et al., 2022). Trabalhos recentes utilizando modelos estocásticos orientados por atores revelaram padrões de aglomeração em colaborações internacionais de P&D, mostrando como estruturas de rede influenciam a difusão do conhecimento (Tsankov, 2021). Os Estados Unidos exemplificam uma integração bem-sucedida de políticas, onde abordagens híbridas regulatórias-voluntárias fomentaram a inovação mantendo flexibilidade (Korkut and Fowler, 2021), em nítido contraste com o cenário fragmentado de inovação no Brasil, apesar de sua expertise em biocombustíveis.

O aprendizado de máquina revolucionou a classificação de patentes por meio de diversos avanços-chave. A classificação por Naive Bayes multinomial e máquinas de vetor de suporte (Support Vector Machines - SVM) permitem atualmente categorizações precisas de textos complexos de patentes, apesar dos desafios linguísticos e estruturais (Chebil et al., 2023; Guven and Simsir, 2020). Técnicas de Indexação Semântica Latente demonstraram especial potencial para superar lacunas conceituais entre documentos de patentes multilíngues e a pesquisa acadêmica (Yu et al., 2024), enquanto métodos baseados em descida do gradiente estocástica (Stochastic Gradient Descent - SGD) melhoraram significativamente a acurácia de classificação em bases de dados de patentes em larga escala (Nakkiran et al., 2019). Esses avanços computacionais coincidem com o crescente reconhecimento das estruturas bipartidas em sistemas de inovação, nas quais as relações entre tipos distintos de nós—como inventores e instituições—exigem abordagens analíticas especializadas para evitar viés por agregação (Campos et al., 2023).

Tabela 1.1: Dimensões Integradas da Inovação em SAF

Dimensão	Principais Conclusões
Tecnológica	A síntese de Fischer-Tropsch oferece maior flexibilidade de matérias-primas, enquanto a rota HEFA domina a capacidade de produção atual devido à maturidade tecnológica (United State’s Department of Energy, 2023)
Econômica	A intensidade de capital varia significativamente entre as rotas, representando de 50 a 75% do custo total para gasificação-FT em comparação com 20 a 50% para os processos de álcool-para-querosene (Ng et al., 2021)
Política	Marcos regulatórios híbridos que combinam mandatos com incentivos de mercado demonstram desempenho superior em relação a instrumentos puramente regulatórios (Korkut and Fowler, 2021)
Analítica	As medidas de centralidade em redes revelam fluxos de conhecimento assimétricos, com certos códigos IPC influenciando desproporcionalmente as trajetórias de inovação (Metcalfe and Casey, 2016)

A síntese dessas literaturas revela três lacunas críticas de pesquisa que este estudo busca abordar. Primeiro, os trabalhos existentes analisam majoritariamente redes de patentes e tecnologias de SAF de forma isolada, negligenciando suas inter-relações dinâmicas. Segundo, a posição paradoxal do Brasil como líder em biocombustíveis com inovação limitada em SAF permanece pouco explorada, apesar de sua importância estratégica. Terceiro, poucos estudos desenvolveram estruturas integradas que combinem

aprendizado de máquina, teoria de redes e análise de políticas para suporte à decisão. Esta pesquisa contribui significativamente em cada um desses pontos ao desenvolver uma metodologia unificada que mapeia redes de patentes para domínios da pesquisa acadêmica, analisa o ecossistema de inovação do Brasil sob perspectivas tanto tecnológicas quanto colaborativas, e cria ferramentas acionáveis para formuladores de políticas e atores da indústria. O framework resultante não apenas avança a compreensão acadêmica dos sistemas de inovação, como também oferece instrumentos práticos para acelerar a transição para a aviação sustentável.

2 RASTREANDO A INOVAÇÃO EM BIOCOMBUSTÍVEIS: UMA ANÁLISE BASEADA EM GRAFOS DE PATENTES DE COMBUSTÍVEL SUSTENTÁVEL PARA AVIAÇÃO

Neste capítulo, detalha-se a metodologia empregada na pesquisa para assegurar a robustez e a reprodutibilidade da análise. Esta abordagem integra métodos qualitativos e quantitativos, com o objetivo de oferecer uma compreensão holística do tema. As seções seguintes exploram os processos envolvidos nos critérios de seleção de materiais, na configuração experimental, nas metodologias de coleta de dados e nas análises estatísticas utilizadas para interpretar os resultados em cada etapa.

2.1 COLETA DE DADOS

O primeiro passo realizado foi a execução de uma consulta específica na base de dados Derwent Innovations Index (DII), utilizando as palavras-chave “*Sustainable Aviation Fuel*”. Essa busca foi realizada em 14 de maio de 2024, e resultou em um total de 73 famílias de patentes relacionadas a essa tecnologia emergente. Para garantir a relevância e a aplicabilidade dos dados, foram excluídas do conjunto as patentes iniciadas por “WO” (Written Opinion). Essas entradas “WO” representam pedidos feitos pelo Tratado de Cooperação em Matéria de Patentes (Patent Cooperation Treaty - PCT), que são depósitos iniciais sob o sistema internacional de patentes e ainda não garantiram proteção em nenhum país (World Intellectual Property Organization, 2022). Essa decisão foi baseada no entendimento de que tais patentes podem não avançar para a fase nacional, etapa em que são avaliadas e potencialmente recebem proteção em países individuais.

Após esse refinamento inicial dos dados, foi realizada uma análise subsequente para identificar e compreender as tendências futuras de patentes no campo do combustível sustentável para aviação. A exclusão das patentes “WO” foi crucial, visto que essas aplicações devem avançar para a fase nacional dentro de 30 meses após a data do depósito, para evitar expiração e garantir proteção legal nos países designados (World Intellectual Property Organization, 2022). Esse período permite que os inventores decidam em quais países desejam obter proteção e se adequem às legislações nacionais de patentes. Ao focar em patentes que já ingressaram na fase nacional, a análise baseia-se em patentes mais maduras e com maior potencial de aplicabilidade jurídica, oferecendo um retrato mais claro dos avanços tecnológicos e dos compromissos de mercado no setor de combustíveis sustentáveis para aviação.

Essa abordagem metodológica permitiu refinar o conjunto de dados para incluir apenas aquelas patentes com maior probabilidade de impactar o mercado e os marcos regulatórios. Conseqüentemente, a análise oferece uma previsão mais precisa dos desenvolvimentos futuros e das tendências de investimento em tecnologias de combustível sustentável para aviação.

2.2 EXTRAÇÃO E PRÉ-PROCESSAMENTO DOS DADOS

Após a obtenção dos documentos de patentes relevantes, foi iniciado o pipeline de extração e pré-processamento de dados com base em Python. As etapas de pré-processamento incluíram:

1. **Análise das Informações de Patentes:** Os documentos de patente foram processados e os dados críticos foram extraídos, como os nomes dos depositantes, os códigos de Classificação Internacional de Patentes (International Patent Classification - IPC), as datas de depósito e os IDs das famílias de patentes. A Tabela 2.1 também apresenta a tabela de renomeação.
2. **Filtragem de Dados:** Um procedimento de segregação foi implementado para separar as famílias de patentes propriamente ditas das aplicações exclusivamente WO (entradas que não possuem nenhum ID além daquele com o prefixo WO).
3. **Manipulação de Dados:** A empresa líder depositante foi extraída do campo para agregar companhias pertencentes à mesma organização; o tamanho da família foi determinado pela contagem dos números de patentes individuais, excluindo as aplicações WO; o ano da patente foi extraído do campo “Application Details and Date” e uma coluna “Anos até a expiração” foi criada para as patentes (20 anos a partir da data de publicação subtraindo-se a data de 14 de maio de 2024) e para as aplicações WO (30 meses a partir da mesma data).

2.3 GERAÇÃO DE GRAFOS

2.3.1 Rede IPC-Assignee

1

Com as informações pré-processadas carregadas, foi desenvolvido um grafo de rede. Os nós centrais vermelhos representam os códigos de Classificação Internacional de Patentes (IPC), e os nós azuis adjacentes representam os depositantes das patentes: empresas e universidades. Estas últimas foram configuradas na cor verde para destacar suas características específicas. Os pesos das arestas foram definidos pelo número de patentes que cada depositante registrou nas respectivas classificações IPC.

2.3.2 Rede IPC-Patente-Assignee

2

Esta rede também foi criada com o objetivo de visualizar as patentes como nós e identificar as mais relevantes em determinado campo. Duas métricas principais foram consideradas para atribuir a relevância da patente: o tamanho da família e os anos restantes até a expiração. O tamanho da família determina o

¹https://github.com/Matheusmno/TGP1/blob/main/IPC_Assignee_weighted_DW.html

²https://github.com/Matheusmno/TGP1/blob/main/IPC_Assignee_PN_DW.html

Tabela 2.1: Descrição dos códigos Derwent

Código Derwent	Descrição
PN	Patent Number
TI	Title
AU	Authors or Inventors
AE	Patent Assignee
GA	IDS Number
AB	Abstract / BHTD Critical Abstract
TF	Technology Focus Abstract
EA	Early access date; Equivalent Abstract
DC	Derwent Class Code(s)
MC	Major Concepts or Derwent Manual Code(s)
IP	International Patent Classification
PD	Publication Date; Patent Details
AD	Application Details and Date
FD	Further Application Details
PI	Publisher City; Patent Priority Information
DS	Designated States
FS	Field of Search
CP	Cited Patent(s)
CR	Cited References
DN	DCR Number
MN	Markush Number
RI	ResearcherIDs; Ring Index Number
CI	Derwent Compound Number
RG	Derwent Registry Number

tamanho do nó em forma de triângulo, e sua cor é definida pelo número de anos restantes até a expiração, considerando a regra dos 20 anos. Foram definidos quatro intervalos de cinco anos para determinar a cor do nó triangular: de 0 a 5 anos, verde; de 5 a 10 anos, verde-claro; de 10 a 15 anos, coral-claro; e de 15 a 20 anos, vermelho-claro. Dessa forma, é possível visualizar tanto as maiores famílias de patentes quanto aquelas prestes a expirar, o que pode representar uma oportunidade de exploração pela indústria.

2.3.3 Rede IPC-Assignee-Written Opinion

3

Essa rede foi criada exclusivamente para visualizar as aplicações do tipo WO e obter uma ideia do que está por vir. O princípio de visualização é essencialmente o mesmo utilizado na rede IPC-Patente-Depositante, porém, a regra para a cor dos nós triangulares foi ajustada: cinza para aplicações prestes a expirar (menos de seis meses até a data de expiração) e roxa para aquelas que ainda têm mais de dois anos para se tornarem uma patente.

2.4 BIBLIOTECAS E FERRAMENTAS PYTHON PARA ANÁLISE DE GRAFOS

- **Pandas:** Utilizada para gerenciamento e processamento de dados.
- **NetworkX:** Empregada para criar estruturas de grafos e realizar análises gráficas.
- **PyVis:** Responsável por gerar os grafos do NetworkX como arquivos HTML interativos.

Combinando tanto a utilização da base de dados Derwent Innovations Index quanto algoritmos de análise de grafos escritos em Python (implementados com a biblioteca NetworkX), esta metodologia investigou minuciosamente o universo das patentes de SAF. O arcabouço metodológico ilustrado na Figura 2.1 permitiu visualizar a cooperação entre clusters, tendências tecnológicas e a dinâmica de inovação dentro da indústria de combustíveis sustentáveis para aviação.

2.5 ANÁLISE E RESULTADOS

Esta seção apresenta os resultados da análise dos dados de patentes de SAF e da análise de grafos. Ela também revela um recente aumento na atividade de patentes de SAF, assim como as principais áreas tecnológicas de foco.

2.5.1 Contextualização dos Dados de Patentes de SAF

Antes de utilizar a teoria dos grafos para exibir as correlações entre diferentes patentes, foi realizada uma análise preliminar abrangente para mapear a evolução das solicitações de patentes e identificar suas

³https://github.com/Matheusmno/TGP1/blob/main/IPC_Assignee_PN_WO.html

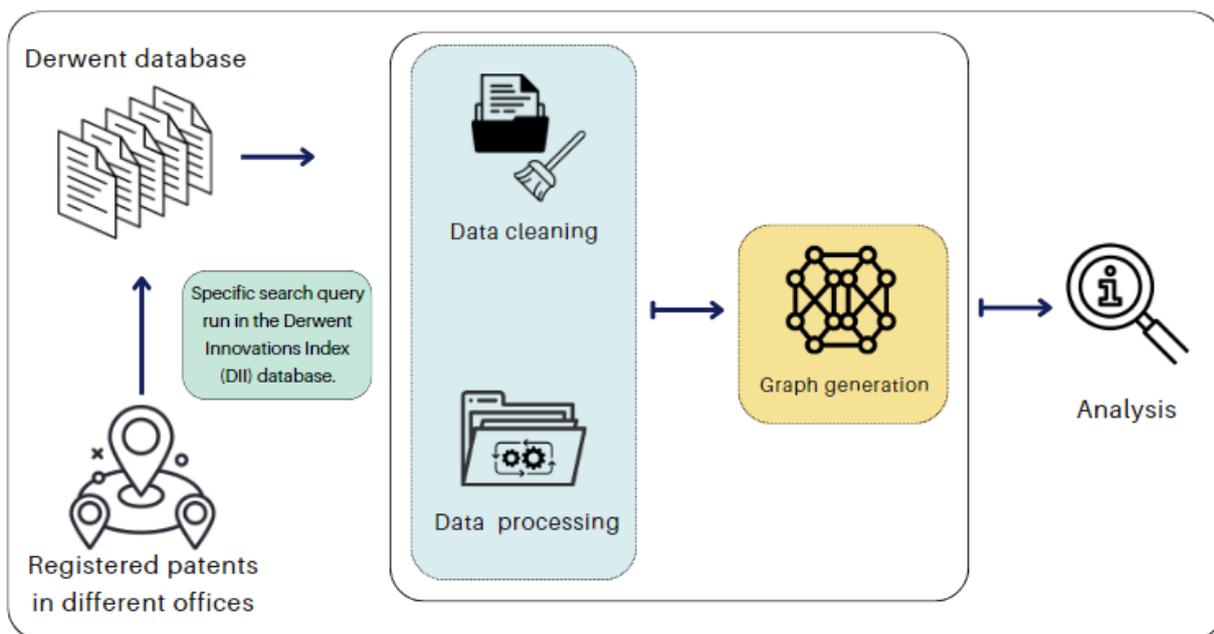


Figura 2.1: Framework Metodológico

principais áreas de pesquisa. A base de dados é composta por um total de 136 patentes, excluindo as aplicações do Tratado de Cooperação em Matéria de Patentes (PCT), com apenas seis patentes tendo data de solicitação anterior a 2019. Esse conjunto de dados oferece uma indicação clara do crescente interesse e dos rápidos avanços no campo dos Combustíveis Sustentáveis de Aviação (Sustainable Aviation Fuel - SAF) nos últimos anos.

Como ilustrado na Figura 2.2, houve um notável aumento nas solicitações de patentes, com um pico em 2022, quando 56,6% de todas as patentes relacionadas a SAF foram depositadas. Esse pico evidencia um crescimento significativo nas atividades de inovação e pesquisa durante esse ano.

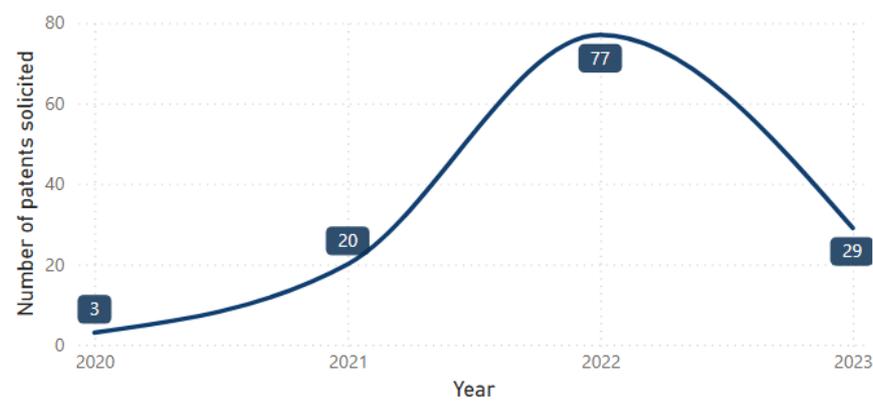


Figura 2.2: Evolução das Patentes Solicitadas

Na Figura 2.3, é possível observar a participação dos países com patentes registradas de SAF. Os Estados Unidos lideram os depósitos de patentes, com 22,4% do total, enquanto China e Reino Unido aparecem logo em seguida, com 13,7% cada. Essa distribuição destaca as contribuições cruciais desses

países para o avanço e a comercialização das tecnologias de SAF, mas também demonstra que não se trata de um mercado fechado, uma vez que é possível notar a presença de outros concorrentes relevantes, como Índia e Coreia do Sul.

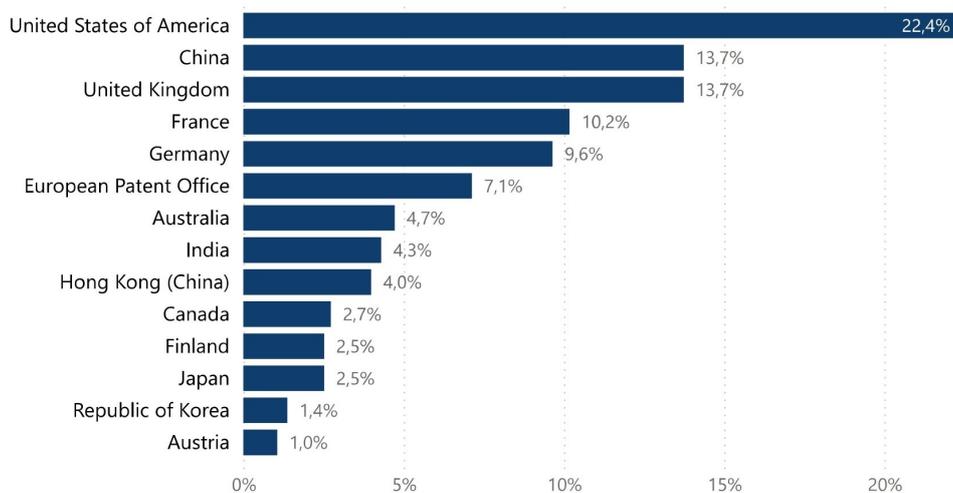


Figura 2.3: Patentes por País

Na Figura 2.4, o debate sobre os perfis dos países nas patentes de SAF é continuado. O gráfico ilustra a distribuição de patentes em várias áreas tecnológicas para diferentes países nos últimos cinco anos (2019–2023), excluindo o ano atual. Os Estados Unidos possuem uma ampla gama de patentes, com participações significativas em Física (39%), Química e Metalurgia (27%) e Engenharia Mecânica (27%).

A China concentra-se fortemente em Química e Metalurgia (42%) e Eletricidade (31%), com participações notáveis em Física (22%) e Engenharia Mecânica (21%). O Reino Unido possui uma distribuição equilibrada, com destaque para Necessidades Humanas (26%) e Construções Fixas (26%). A França enfatiza Química e Metalurgia (41%)

O Escritório Europeu de Patentes (European Patent Office - EPO) tem um foco maior em Física (54%). Na Austrália, Física (61%) e Química e Metalurgia (37%) dominam. A Índia enfatiza Física (54%) e Química e Metalurgia (27%). Hong Kong (China) também se concentra em Física (61%)

Além disso, a Figura 2.5 apresenta o número de patentes solicitadas por cada país por ano de 2020 a 2023, revelando diversas tendências importantes. Em 2020, a China foi o único país com atividade signi-

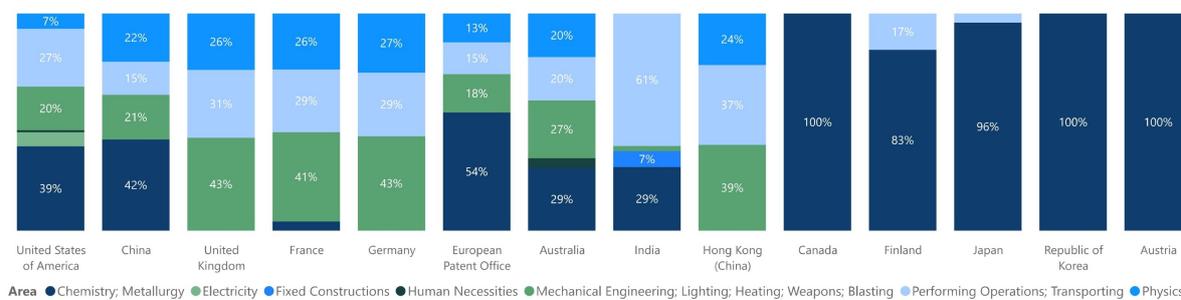


Figura 2.4: Perfil tecnológico por país

ficativa de patentes, mas não manteve essa tendência nos anos seguintes. Os Estados Unidos se destacam como o principal ator nas solicitações de patentes, especialmente em 2022, quando atingiram a maior contagem em todos os anos apresentados. Esse domínio continua em 2023, mantendo sua posição de liderança no cenário de patentes.

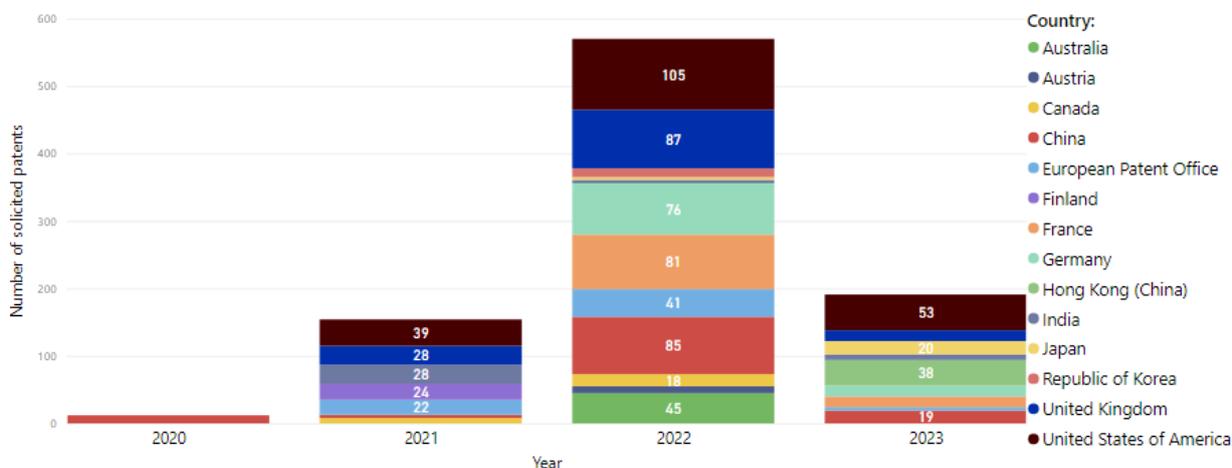


Figura 2.5: Patentes solicitadas por país por ano

Houve um aumento notável nas solicitações de patentes em 2022 em comparação com os outros anos, com um crescimento significativo na atividade de múltiplos países. Esse ano contou com contribuições de uma gama diversificada de países, incluindo Reino Unido, República da Coreia, Alemanha, França, Escritório Europeu de Patentes, China, Canadá e Austrália, indicando um pico na atividade global de patentes. Em contraste, 2021 apresentou uma atividade de patentes moderada, com Estados Unidos, Reino Unido, Índia, Finlândia e o Escritório Europeu de Patentes contribuindo para uma distribuição equilibrada, porém em menor escala em comparação a 2022.

O número de patentes solicitadas em 2023 foi visivelmente menor que em 2022. Embora os Estados Unidos continuem liderando, o número total de patentes solicitadas por outros países caiu significativamente. Notavelmente, países como Japão e Hong Kong (China) só aparecem nos anos mais recentes, especificamente em 2023, indicando um engajamento mais recente nas atividades de solicitação de patentes.

De forma geral, o gráfico indica que, enquanto a China teve uma forte presença em 2020, os Estados Unidos lideraram consistentemente nos anos subsequentes, com destaque em 2022. O ano de 2022 marcou um ponto alto na atividade global de patentes, com ampla participação internacional, enquanto 2023 apresenta uma redução nas solicitações, sugerindo uma possível contração ou estabilização após o pico do ano anterior.

A análise se estendeu à distribuição de patentes nas áreas da Classificação Internacional de Patentes (International Patent Classification - IPC), com o objetivo de revelar tendências emergentes e focos tecnológicos dentro do domínio do SAF. A Figura 2.6 mostra as diversas categorizações para o mesmo número de patentes. Entre as categorias IPC mais prevalentes estão Química e Metalurgia (área C), Engenharia Mecânica, Iluminação, Aquecimento, Armas, Explosivos (área F) e Operações e Transporte (área B). Essas categorias representam o amplo espectro de inovações tecnológicas que contribuem para o avanço

do SAF e devem ser priorizadas no investimento em infraestrutura no país.

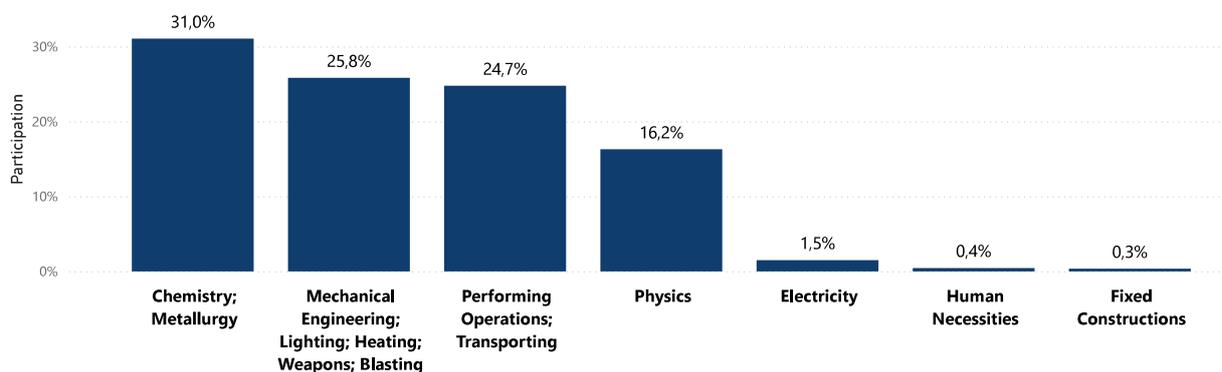


Figura 2.6: Porcentagem das áreas tecnológicas segundo a IPC

Uma análise adicional, representada na Figura 2.7, ilustra a evolução anual das IPCs (International Patent Classification) nas patentes solicitadas, aprofundando a análise anterior. Notavelmente, a área de Química e Metalurgia manteve uma taxa de aplicação consistente ao longo do tempo, indicando interesse contínuo de pesquisa e inovação constante nessas áreas fundamentais. Por outro lado, a categoria que abrange Engenharia Mecânica, Iluminação, Aquecimento, Armas e Detonação apresentou um aumento significativo em 2022, correspondendo a 30,8% de todas as IPCs em patentes de SAF naquele ano. Esse crescimento acentuado indica um foco crescente em soluções de engenharia e mecânicas dentro do campo de SAF.

Em resumo, o exame da base de dados forneceu insights valiosos sobre a distribuição e evolução das patentes de SAF nos últimos anos. O conjunto de dados, apesar de contar com 136 patentes, revela que apenas uma pequena fração é anterior a 2019, indicando um aumento recente na atividade de patentes, reflexo

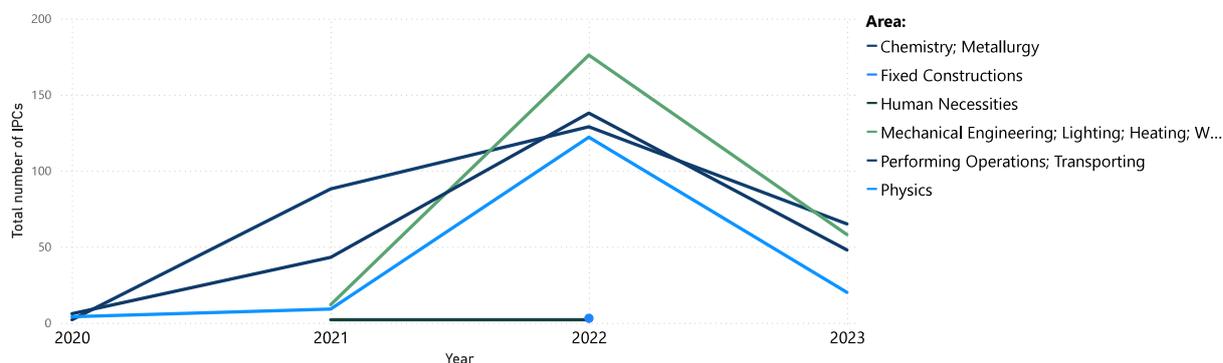


Figura 2.7: Evolução das áreas tecnológicas da IPC nas patentes solicitadas

da aceleração da inovação. Estados Unidos, China e Reino Unido destacaram-se como os principais atores no cenário global de registros de patentes de SAF. A análise das áreas da IPC revela padrões distintos, com Química e Metalurgia, Engenharia Mecânica, Iluminação, Aquecimento, Armas e Detonação, e Realização de Operações e Transporte emergindo como categorias proeminentes. Conforme ilustrado pela evolução anual das IPCs, a natureza dinâmica da distribuição de patentes enriquece ainda mais a compreensão das tendências emergentes e das áreas de inovação dentro do campo.

2.5.2 Análise de Redes Complexas de SAF

O grafo da Figura 2.8 conecta empresas e universidades (cessionários de patentes) aos códigos da Classificação Internacional de Patentes (IPCs) com os quais a patente foi registrada. Os nós verdes representam as universidades, e a largura da aresta representa o número de patentes que cada cessionário registrou na respectiva área.

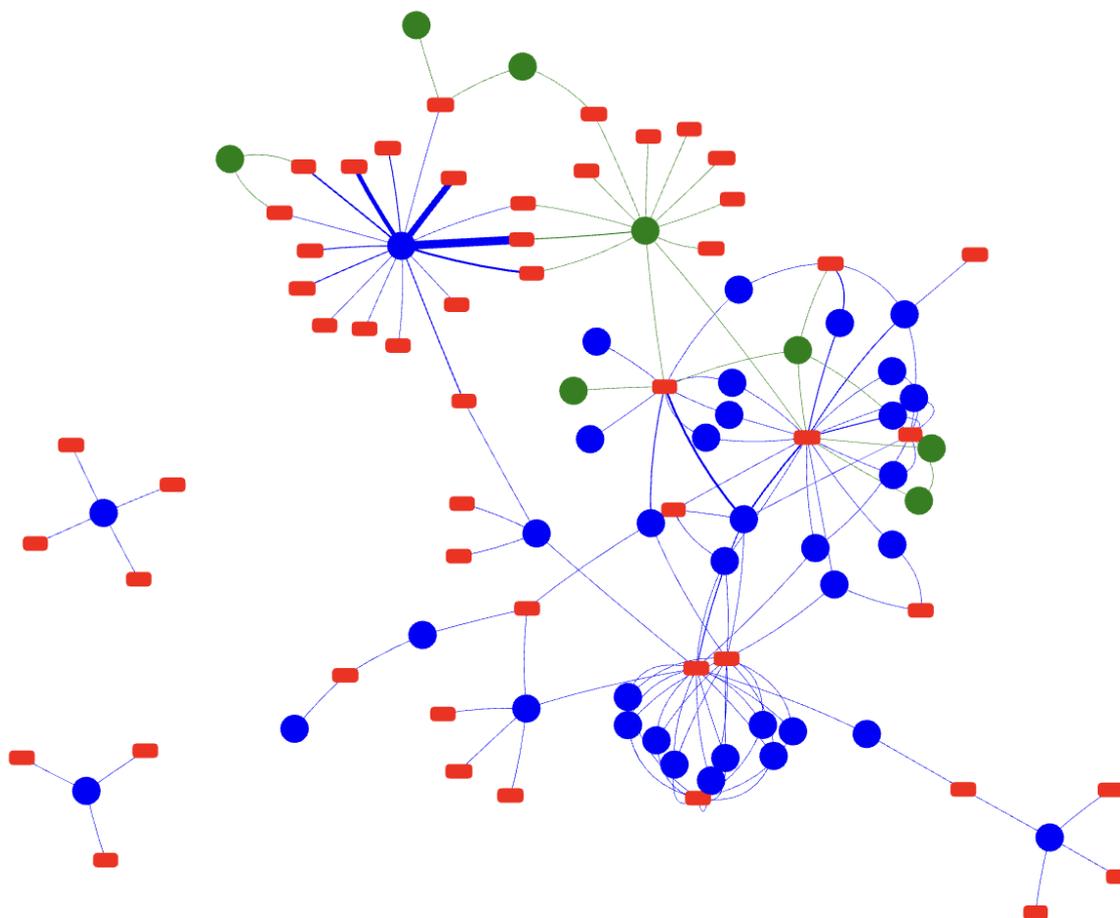


Figura 2.8: Visão geral do grafo de rede social IPC/Cessionário

A espessura de cada aresta que conecta um nó a uma IPC (International Patent Classification) indica o volume de patentes registradas por cada cessionário dentro daquele campo tecnológico específico. Arestas mais espessas representam uma quantidade maior de patentes, revelando, assim, o envolvimento ativo e o

conhecimento especializado dos cessionários na respectiva categoria da IPC.

Essa representação visual serve como uma ferramenta valiosa para obter insights sobre o padrão de distribuição e o nível de concentração das atividades relacionadas a patentes entre as organizações e suas respectivas áreas de foco em pesquisa. Por meio da análise dessas interconexões, torna-se possível identificar os principais contribuintes em setores tecnológicos distintos e apontar oportunidades potenciais de colaboração, bem como áreas que podem gerar competição.

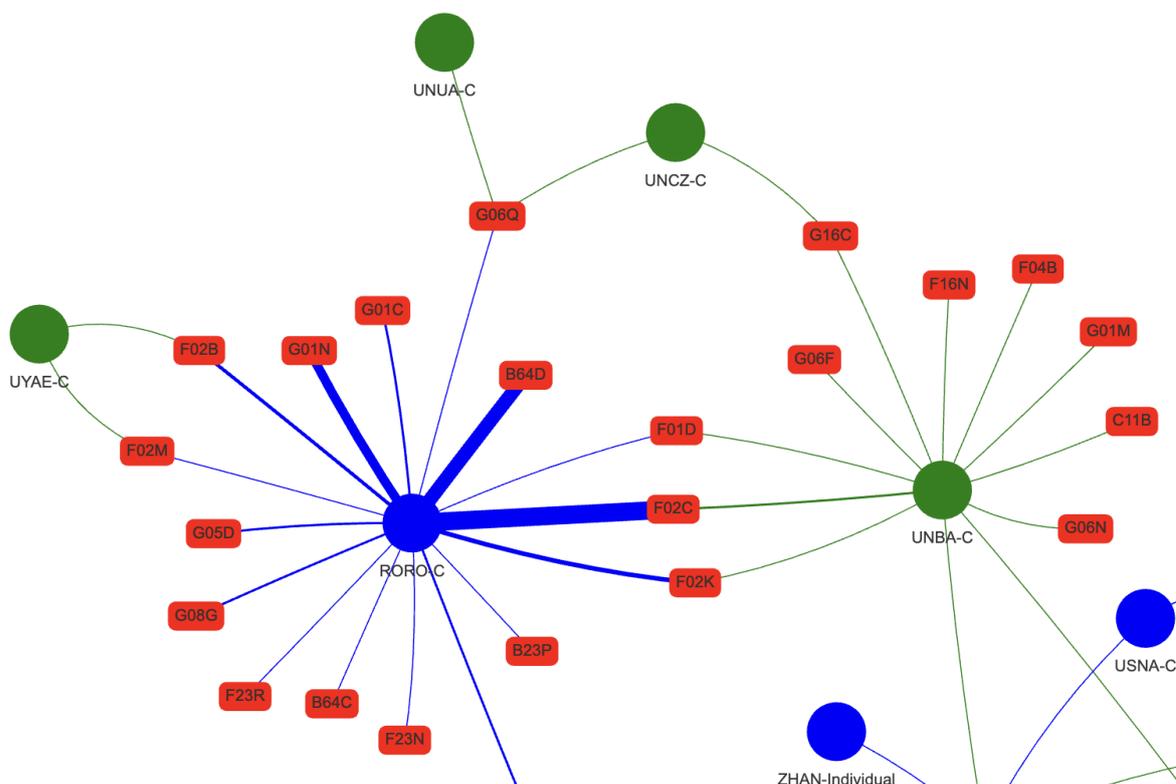


Figura 2.9: Domínio da Rolls-Royce e nós folha de universidades (recorte/ampliação)

Na seção do grafo representada pela Figura 2.9, é possível observar que a empresa Rolls-Royce (RORO-C) possui muitas patentes em múltiplas áreas nas quais não há outros concorrentes, e que essas IPCs (Internacional Patent Classification) estão conectadas a outras IPCs relacionadas que estão sendo exploradas por universidades, sugerindo uma possível cooperação entre essas entidades.

A distribuição demonstra o estabelecimento de uma presença dominante da Rolls-Royce em campos tecnológicos específicos, indicando uma posição sólida conquistada pela empresa. Simultaneamente, observa-se o envolvimento ativo das universidades em áreas de pesquisa intimamente relacionadas às da empresa. A natureza interconectada dessas Classes de Propriedade Intelectual (IPC) evidencia o potencial de esforços colaborativos em pesquisa e desenvolvimento.

Se concretizada, essa colaboração pode possibilitar a transferência de conhecimento entre a indústria e a academia, levando, em última instância, à ampliação dos processos de inovação e ao avanço das tecnologias nas áreas mencionadas. Essa relação simbiótica tem o potencial de gerar o desenvolvimento de tecnologias inéditas e fortalecer os portfólios de propriedade intelectual tanto da Rolls-Royce quanto das instituições acadêmicas envolvidas.

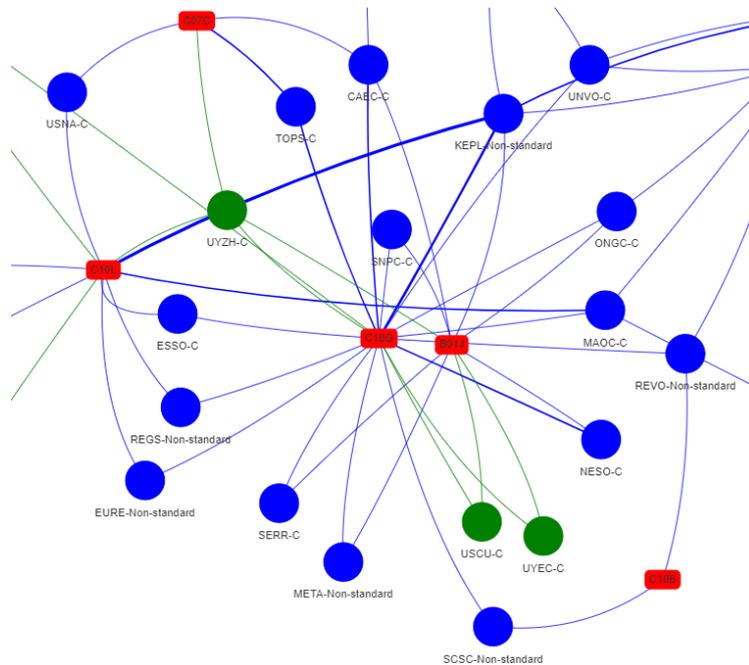


Figura 2.10: IPC mais visada em SAF (recortada/ampliada)

Foi realizada uma análise de centralidade de grau para identificar os principais nós no grafo, com foco específico nos nós de IPC (International Patent Classification). Essa análise teve como objetivo identificar os nós com mais conexões diretas com os demais, destacando assim sua importância e influência dentro da rede.

A análise forneceu dados que oferecem insights sobre os papéis centrais desempenhados por nós específicos de IPC. Os resultados dessa análise de centralidade de grau estão consolidados na tabela a seguir:

Tabela 2.2: Top 5 medidas de centralidade de grau para nós de IPC

Nó IPC	Centralidade de Grau
C10G	0.2021
B01D	0.1596
C01B	0.1383
C10L	0.1170
B01J	0.1064

A Tabela 2.2 ilustra as cinco maiores medidas de centralidade de grau para nós de IPC. Conforme já observado, o nó C10G apresenta o maior grau de centralidade nesta análise. Isso implica que o nó C10G estabelece o maior número de associações imediatas com outros nós dentro da estrutura, fortalecendo assim sua posição central.

A predominância do C10G em relação à centralidade de grau acentua sua relevância e impacto dentro do sistema, indicando que ele provavelmente funciona como um ponto focal fundamental de conexões. Esse status destaca o papel do C10G na viabilização da comunicação e conexões entre os nós IPC.

Adicionalmente, a Tabela 2.3 apresenta as cinco maiores métricas de centralidade de grau para os nós de depositários. A partir dos dados, observa-se que a RORO-C (Rolls-Royce) possui a maior centralidade de grau, demonstrando sua posição proeminente dentro da rede. A centralidade da RORO-C implica uma forte conexão com outros nós e ressalta seu papel na estruturação da rede. De forma semelhante, a instituição UNBA-C apresenta um grau notável de centralidade, reforçando sua relevância e impacto dentro da rede. Esses resultados destacam os papéis críticos desempenhados tanto pela RORO-C quanto pela UNBA-C na rede.

Tabela 2.3: Top 5 medidas de centralidade de grau para nós de depositários

Nó de Depositário	Centralidade de Grau
RORO-C	0.1702
UNBA-C	0.1277
KEPL-Non-standard	0.0638
PARA-Non-standard	0.0532
CAEC-C	0.0426

Conforme a Figura 2.10, o IPC (International Patent Classification) mais popular (centralidade de grau = 0,2021) dentro do conjunto de dados é o C10G, como mostrado na Tabela 2.2 (Craqueamento de óleos hidrocarbonetos; produção de misturas de hidrocarbonetos líquidos, por exemplo, por hidrogenação destrutiva, oligomerização, polimerização (craqueamento para hidrogênio ou gás de síntese C01B; craqueamento ou pirólise de gases hidrocarbonetos para hidrocarbonetos individuais ou misturas de constituição definida ou especificada C07C; craqueamento para coque C10B); recuperação de óleos hidrocarbonetos de xisto betuminoso, areia betuminosa ou gases; refino de misturas compostas principalmente por hidrocarbonetos; reformação de nafta; ceras minerais), possuindo 22 arestas conectando depositários a ele, sugerindo possível exaustão.

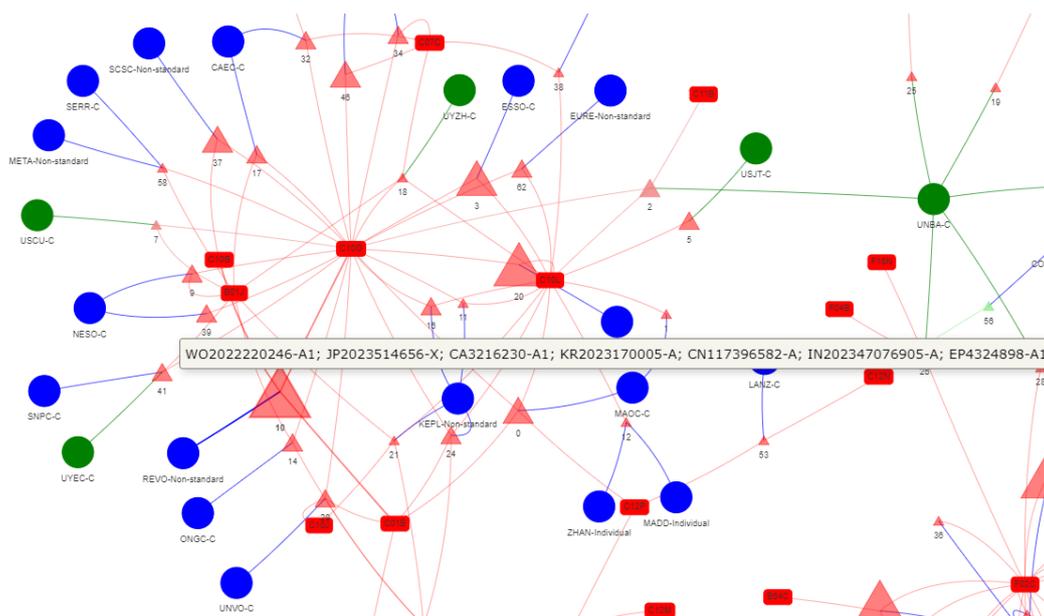


Figura 2.11: Família de patentes mais relevante do C10G destacada

Ao analisar a Figura 2.11, torna-se evidente que a família de patentes destacada 10 — “Produção de combustível líquido, por exemplo, nafta, envolve o fornecimento de matéria-prima orgânica contendo resíduos poliméricos, introdução de um gás inerte, decomposição da matéria-prima e descarte com um gás de arraste, resfriamento dos componentes craqueados não sólidos, reformação e destilação do óleo reformado”(WO2022220246-A1; JP2023514656-X; CA3216230-A1; KR2023170005-A; CN117396582-A; IN202347076905-A; EP4324898-A1), representada pelo maior triângulo que está conectado ao IPC (International Patent Classification) mais visado (C10G), pode ser uma oportunidade para os países interessados na área começarem, já que é um indicativo de força e sucesso da família de patentes.

Tabela 2.4: 5 Maiores Valores de Centralidade de Grau para Nós de Patente

ID do Nó de Patente	Centralidade de Grau
68	0.0779
50	0.0584
48	0.0455
67	0.0390
0	0.0325

Para países que possuem interesse neste campo específico, concentrar-se nesta família de patentes pode representar um ponto de partida estratégico. A presença evidente desta família de patentes implica um grau de criatividade e potencial para crescimento comercial, tornando-se, portanto, um ponto de referência para indivíduos iniciantes na área. Ao utilizar os insights derivados dessa família de patentes dominante, as indústrias têm a oportunidade de alinhar seus esforços de pesquisa e desenvolvimento para capitalizar sobre as forças já existentes e, potencialmente, adquirir uma vantagem competitiva dentro do domínio.

A Figura 2.12 mostra que as maiores famílias de patentes estão associadas à empresa Rolls-Royce, que domina exclusivamente a IPC B64D, conectada à maioria de suas patentes. Apenas duas universidades também possuem patentes que conectam a rede a outras comunidades, indicando, assim, um nível restrito de concorrência dentro deste setor.

Ademais, a presença de duas instituições acadêmicas que possuem patentes conectando a rede a diversas outras comunidades implica um certo nível de engajamento acadêmico; contudo, é evidente que o domínio de mercado repousa predominantemente com a Rolls-Royce e as duas empresas mencionadas. O domínio da Rolls-Royce, juntamente com a participação limitada de outras organizações, ressalta o firme controle da corporação e seu potencial poder de influência sobre as tendências de inovação e desenvolvimento dentro do setor da IPC B64D. Esse argumento sugere fortemente que tal controle de mercado acentua a necessidade de monitorar a dinâmica competitiva para que haja uma participação mais ampla e se cultive um cenário competitivo favorável à inovação.

Relacionado ao subconjunto de pedidos WO (Written Opinion), a Figura 2.13 segrega as aplicações WO para uma visualização mais clara do que está no radar para os próximos anos. Das 73 entradas extraídas do banco de dados, 12,32% (9) são estritamente aplicações WO (Written Opinion) e ainda não foram patenteadas em nenhum lugar do mundo. É importante ressaltar que essas aplicações específicas ainda não foram concedidas em nenhuma jurisdição global. O ponto principal é que essa segregação

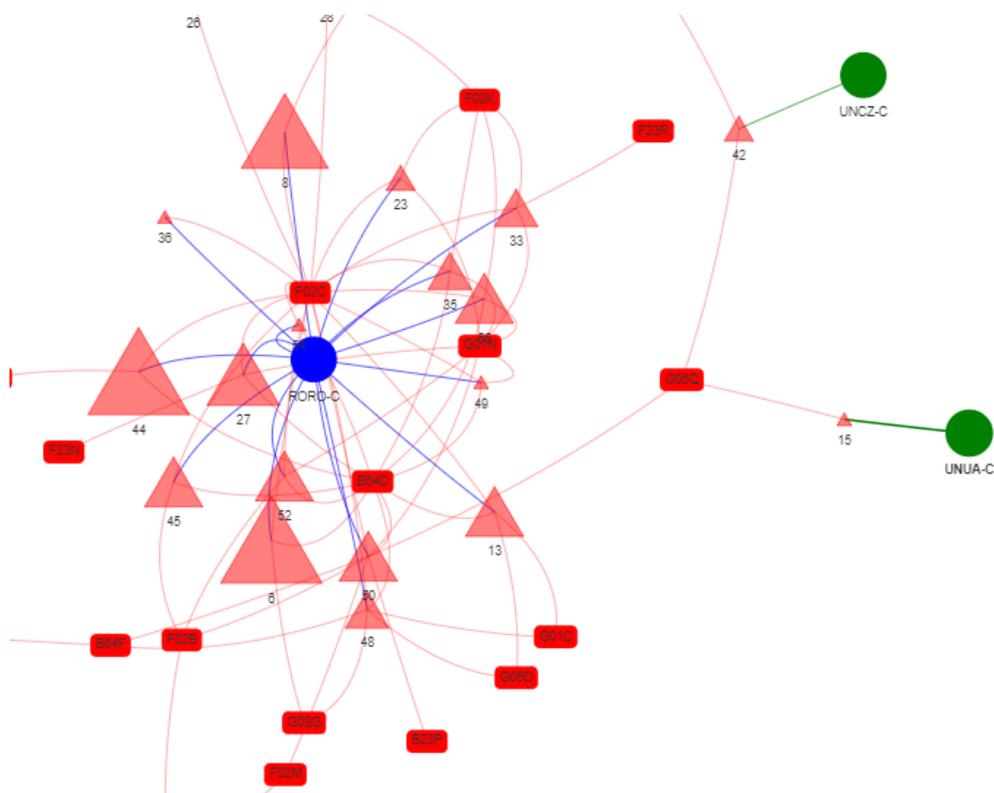


Figura 2.12: Cluster da Rolls-Royce

destaca o potencial de invenções novas e únicas que estão atualmente sob o processo de patenteamento internacional, aguardando exame e aprovação.

Subsequentemente, a Figura 2.14 foca em dois pedidos que estão prestes a expirar e pertencem à mesma empresa sob as mesmas IPCs. O fato de esses pedidos estarem expirando simultaneamente e pertencerem ao mesmo domínio tecnológico pode indicar diversas questões subjacentes. Isso pode sugerir uma possível estagnação nas capacidades inovadoras da empresa, onde não foram desenvolvidos avanços ou melhorias significativas que justificassem novas proteções de patente ou extensões. Alternativamente, pode refletir uma falta de interesse do mercado ou de viabilidade comercial para as tecnologias cobertas por essas patentes, levando a empresa a deixá-las expirar em vez de investir em sua renovação ou em mais inovações.

Com relação aos pedidos WO2024030611A2 e WO2024030611A3, estes são referenciados como os reagentes compostos por lignina e hidrogênio. Embora, nos últimos anos, avanços significativos tenham sido alcançados no desenvolvimento de eletrocatalisadores ativos e na engenharia da configuração das células, sua cinética permanece inferior à envolvida na eletrólise da água tradicional por várias ordens de magnitude González-Cobos et al. (2023). Portanto, esse argumento sugere que esse tipo de tecnologia pode não ser considerado inovador o suficiente para publicação.

Da mesma forma, o pedido WO (Written Opinion) intitulado WO2023222798A1 demonstra o mesmo dilema. Embora esforços tenham sido feitos, Zhu et al. (2020) mostram que o uso de um catalisador de reação reversa gás-água provoca uma reação colateral de metanação em diversos catalisadores. Como resultado, forma-se uma mistura de produtos CO e CH_4 , sobrecarregando, assim, os processos de separação

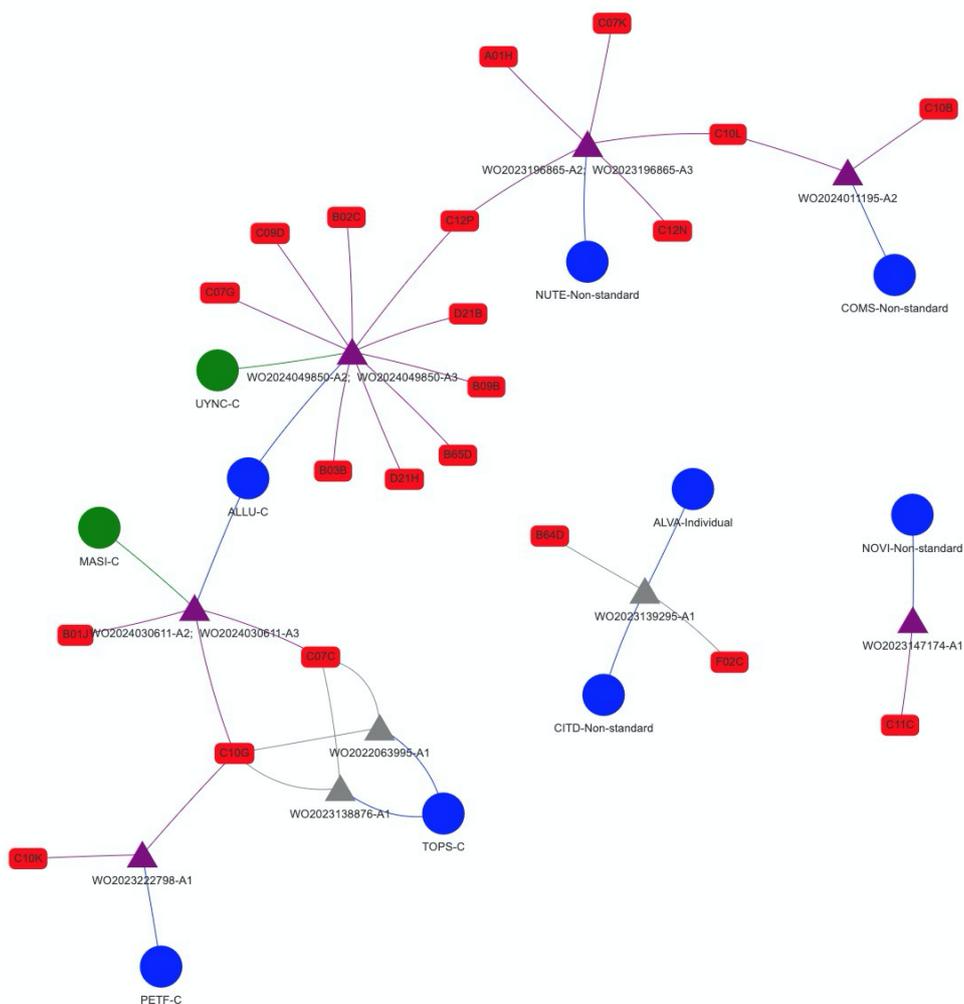


Figura 2.13: Rede IPC-Assignee-Opinião Escrita

subsequentes. Além disso, o autor alerta para que os estudos se concentrem no mecanismo de reação dos catalisadores nesses processos, como os utilizados neste pedido. Esses cenários destacam os desafios potenciais enfrentados pelas empresas para manter sua vantagem competitiva e enfatizam a importância da pesquisa e desenvolvimento contínuos para sustentar a liderança tecnológica e de mercado.

Finalmente, na Figura 2.15, chama-se atenção para a colaboração entre ALLU-C e UYNC-C, que possuem expertise em nove áreas. Uma possível nova parceria poderia ocorrer entre ALLU-C e NUTE-Non-Standard, uma vez que compartilham interesse comum na IPC C12P. Isso pode indicar que, caso sejam concorrentes e não tenham interesse em colaborar entre si, a parceria sugerida seria entre UYNC-C e NUTE-Non-Standard.

Por fim, este subconjunto de pedidos WO (Written Opinion) representa uma parcela significativa das entradas totais e reforça o potencial para inovação futura e depósitos de patentes nessas áreas. Ao isolar esses pedidos WO, pesquisadores e partes interessadas podem concentrar sua atenção em tecnologias e tendências emergentes que ainda não foram plenamente resolvidas ou protegidas. Essa abordagem direcionada possibilita o monitoramento proativo e o planejamento estratégico para capitalizar oportunidades emergentes no domínio das patentes.

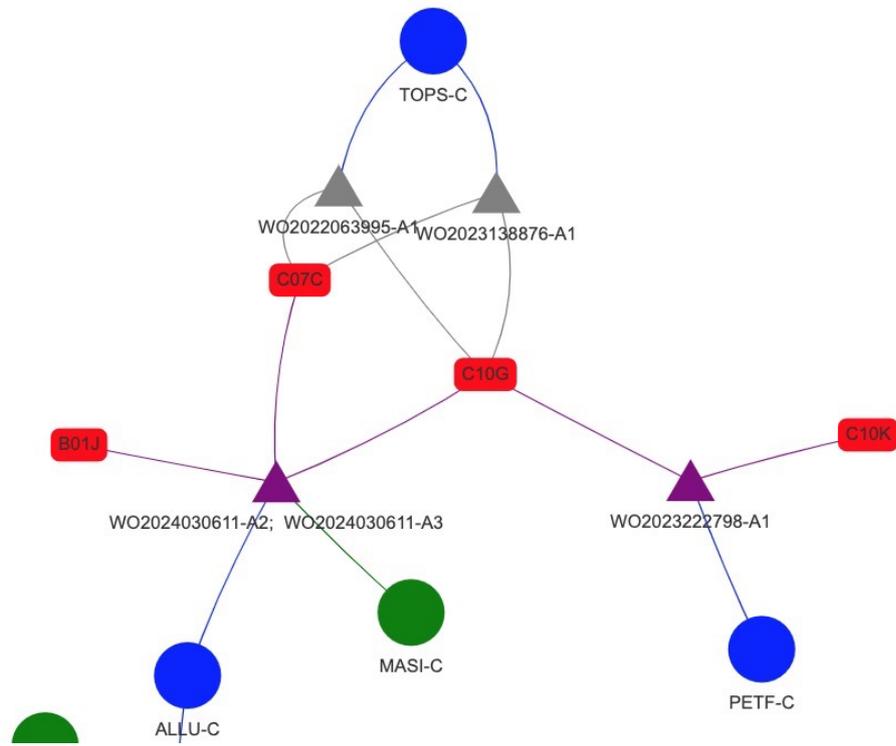


Figura 2.14: Pedidos próximos da expiração

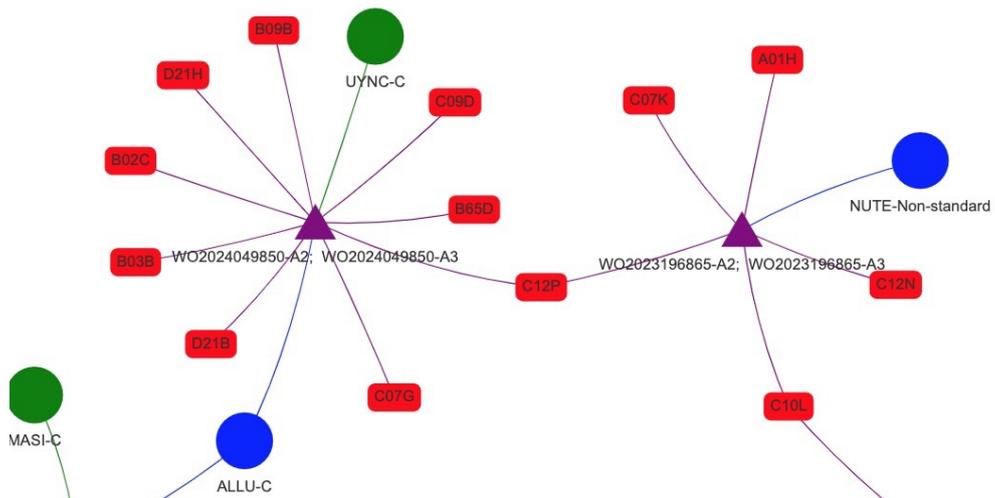


Figura 2.15: Colaboração entre universidade e empresa em múltiplas IPCs

3 O CAMINHO DO BRASIL PARA CÉUS VERDES: ANÁLISE DE COMBUSTÍVEIS SUSTENTÁVEIS PARA AVIAÇÃO POR MEIO DE PATENTES E ARTIGOS UTILIZANDO TEORIA DE REDES

Neste capítulo, a metodologia empregada na pesquisa é detalhada para garantir a robustez e a reprodutibilidade da análise. Essa abordagem integra métodos qualitativos e quantitativos. As subseções a seguir exploram os processos envolvidos nos critérios de seleção do material, na configuração experimental, nas metodologias de coleta de dados e nas análises estatísticas utilizadas para interpretar os resultados em cada etapa. A Figura 3.1 representa o framework metodológico.

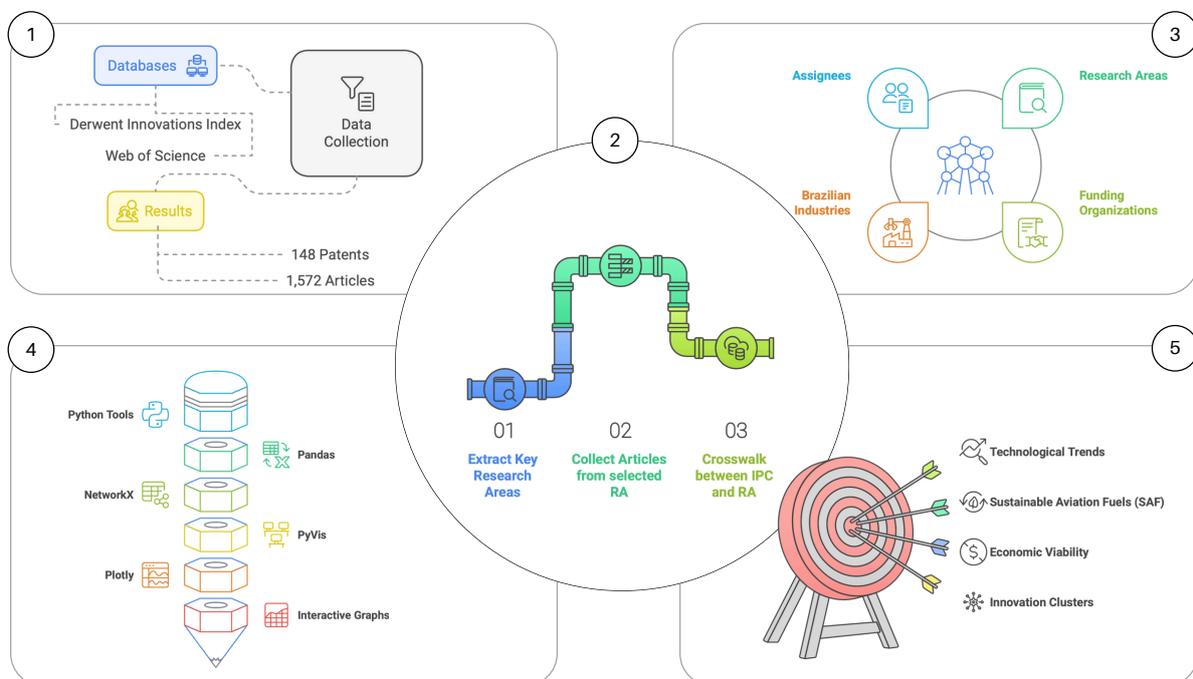


Figura 3.1: Framework Metodológico

A Figura 3.1 apresenta o framework metodológico abrangente desenvolvido para conectar sistematicamente a análise de patentes com a pesquisa acadêmica, possibilitando uma compreensão holística dos ecossistemas de inovação. Essa abordagem inovadora aborda a lacuna crítica de tradução do conhecimento entre a P&D industrial e a descoberta científica, ao estabelecer vínculos entre os desenvolvimentos tecnológicos (patentes) e os avanços científicos (artigos de pesquisa). O framework não apenas ilumina toda a cadeia de inovação — da pesquisa fundamental à aplicação comercial — como também fornece insights acionáveis para fomentar a colaboração estratégica entre indústria e academia, um fator crucial para o progresso tecnológico em combustíveis sustentáveis para aviação e áreas correlatas.

A metodologia descrita envolve um processo de múltiplas etapas que utiliza diversas fontes de dados

e ferramentas analíticas para gerar uma tabela de ligação entre a Classificação Internacional de Patentes (International Patent Classification - IPC) e as Áreas de Pesquisa (Research Areas - RAs). Inicialmente, consultas específicas são realizadas nas bases de dados Derwent Innovations Index (DII) e Web of Science (WoS) para coletar os dados. Esses dados passam por processos de limpeza e tratamento para garantir precisão e consistência. Os dados limpos são então utilizados para gerar grafos, que são subsequentemente analisados para extrair insights e relações.

A metodologia também integra Máquinas de Vetores de Suporte (Support Vector Machines - SVM) para classificar patentes nas principais áreas de pesquisa identificadas nos resultados da Web of Science para a consulta realizada. Subsequentemente, grafos bipartidos são construídos para associar artigos e patentes às suas respectivas áreas de pesquisa, possibilitando a identificação de oportunidades potenciais de colaboração entre a indústria e a academia. O resultado final desse processo é uma tabela de ligação abrangente que mapeia os códigos IPC para as áreas de pesquisa correspondentes, fornecendo uma estrutura sistemática para analisar as conexões entre patentes e áreas acadêmicas, além de apoiar futuras investigações acadêmicas e o financiamento por parte de organizações interessadas.

3.1 COLETA DE DADOS

A primeira etapa da pesquisa consistiu na realização de uma busca direcionada nas bases de dados Derwent Innovations Index (DII) e Web of Science (WoS), utilizando as palavras-chave “*Sustainable Aviation Fuel*” e “*Alternative Aviation Fuel*”. Essa busca teve como objetivo identificar patentes e artigos relevantes que contribuíssem para o desenvolvimento e a inovação em tecnologias de SAF. As consultas foram executadas em 17 de janeiro de 2025 para os artigos e em 22 de janeiro de 2025 para as patentes. Os resultados totalizaram 148 famílias de patentes extraídas da DII e 1.572 artigos da WoS, abrangendo o período de 2020 a 2025.

Um dos principais desafios dessa fase foi estabelecer a ligação entre as áreas de pesquisa (RAs) associadas aos artigos e os códigos da Classificação Internacional de Patentes (International Patent Classification - IPC) atribuídos às patentes. O sistema IPC categoriza patentes em oito classes principais (A–H), enquanto os artigos são classificados em 252 áreas de pesquisa distintas.

Para enfrentar esse desafio, modelos de aprendizado de máquina (Machine Learning - ML) foram empregados para criar uma conexão sistemática entre os códigos IPC e as áreas de pesquisa. A abordagem envolveu a extração das principais áreas de pesquisa identificadas na busca da WoS, a recuperação de 1.000 artigos recentes (dos últimos cinco anos) para cada classe IPC, e o treinamento de múltiplos modelos de aprendizado de máquina para avaliar seu desempenho nesse contexto. O conjunto de dados de treinamento foi composto pelos títulos, resumos e palavras-chave dos artigos, concatenados. A tabela de ligação resultante, que funciona como uma ponte entre as classificações IPC e as áreas de pesquisa, assim como todos os modelos treinados, estão disponíveis em de Oliveira et al. (2025).

Correspondentemente, as Classificações Internacionais de Patentes (International Patent Classification - IPCs) podem ser sistematicamente vinculadas às Classificações Internacionais Padrão de Todas as Atividades Econômicas (International Standard Industrial Classification - ISIC). Essa vinculação permite a

identificação e análise dos setores industriais dentro do Brasil, destacando, assim, áreas-chave de atividade industrial relacionadas à tecnologia SAF. Ao estabelecer essa conexão, os pesquisadores podem identificar oportunidades promissoras de crescimento e desenvolvimento em economias emergentes.

O framework ISIC, mantido e disponibilizado pela Divisão de Estatísticas das Nações Unidas (United Nations Statistics Division - UNSD), serve como um padrão reconhecido globalmente para a categorização de atividades econômicas. Ele facilita a análise de dados industriais e econômicos entre diferentes países. Integrar IPCs com ISICs aprimora a compreensão dos panoramas industriais e apoia o planejamento estratégico e a formulação de políticas em mercados emergentes.

Ademais, vários critérios foram utilizados para garantir a precisão e a utilidade do conjunto de dados. Patentes com o prefixo “WO” foram excluídas do conjunto. O termo “patentes WO” refere-se a pedidos publicados pela Organização Mundial da Propriedade Intelectual (World Intellectual Property Organization - OMPI) no âmbito do Tratado de Cooperação em Matéria de Patentes (Patent Cooperation Treaty - PCT), frequentemente depositados internacionalmente. A exclusão das patentes “WO” foi necessária para restringir o estudo a patentes específicas de jurisdições nacionais, em vez daquelas ainda em processo de consideração para proteção internacional.

3.1.1 Recuperação e Pré-processamento dos Dados

Após a conclusão da recuperação e harmonização dos documentos de patentes e artigos, a próxima fase envolveu a implementação de um pipeline de recuperação e pré-processamento de dados baseado em Python. Esse pipeline foi utilizado para preparar os dados para as análises subsequentes, e as etapas de pré-processamento foram projetadas para garantir a precisão e a relevância dos dados processados. As etapas de pré-processamento abrangeram:

1. **Extração de Informações de Patentes:** Os documentos de patente foram processados, com elementos críticos de dados extraídos, como o titular da patente, códigos da Classificação Internacional de Patentes (IPC), datas de publicação e citações de patentes.
2. **Filtragem de Dados:** Foi aplicada uma abordagem de segregação para isolar informações sobre titulares de patentes atuando com combustíveis sustentáveis para aviação e seus respectivos códigos IPC.
3. **Normalização:** Os dados extraídos foram padronizados, promovendo uniformidade e consistência em todo o conjunto de dados. Essa etapa é crucial para garantir comparabilidade e coerência nas análises subsequentes.
4. **Controle de Qualidade e Validação:** Para verificar a precisão e a confiabilidade dos dados processados. Isso envolveu a verificação cruzada das entradas, a correção de inconsistências e a validação da integridade do conjunto de dados para manter altos padrões de qualidade.

É imperativo destacar que esses processos são parte integrante da metodologia geral de pesquisa, pois asseguram que o conjunto de dados seja preciso, relevante e esteja pronto para análises detalhadas. A

abordagem sistemática de recuperação e pré-processamento de dados estabelece as bases para as fases subsequentes da pesquisa.

3.2 ANÁLISE DE GRAFOS

Esta seção examina duas estruturas de rede utilizadas na pesquisa, com foco na construção e análise de redes bipartidas: redes de afiliação entre depositantes e áreas de pesquisa (RA), redes entre organizações financiadoras, depositantes e RAs, redes entre ISICs de organizações financiadoras brasileiras, depositantes e RAs, e redes multilayer 3D. Essas análises baseadas em grafos são essenciais para compreender as nuances das relações e interações complexas dentro do conjunto de dados, proporcionando insights sobre a natureza interconectada entre patentes, depositantes e entidades relacionadas.

3.2.1 Redes Bipartidas

Segundo Giamphy et al. (2023), um grafo bipartido ($G = (U, V, E)$) é composto por dois conjuntos de nós de domínios (ou tipos) diferentes (U e V), e um conjunto de arestas ($E \subset U \times V$). Grafos bipartidos consideram apenas arestas entre domínios: dentro de um domínio (U ou V), dois nós não podem estar conectados. Seja ($n = |U|$) e ($m = |V|$). Seja (u_i) o (i)-ésimo nó de (U), ($i = 1, 2, \dots, n$) e (v_j) o (j)-ésimo nó de (V), ($j = 1, 2, \dots, m$). Cada aresta ((u_i, v_j)) pode conter um peso (w_{ij}) que descreve a força da relação entre os nós (u_i) e (v_j).

Para Gabow (2016) Gabow (2016), se um grafo não possui pesos, considera-se que todos os pesos são iguais a 1. Da mesma forma, se dois nós (i) e (j) não estiverem conectados, define-se ($w_{ij} = 0$). Portanto, utiliza-se uma matriz ($n \times m$) ($W = (w_{ij})$) para representar todos os pesos na rede bipartida. Precisamente, (W) é denominada matriz de adjacência do domínio (U) e a matriz de adjacência do domínio (V) é a transposta de (W): ($W^T \in \mathbb{R}^{m \times n}$).

3.2.2 Rede de Afiliações–Depositantes–Áreas de Pesquisa

1

Um grafo de rede foi desenvolvido com base nas informações pré-processadas para o projeto. Os nós centrais em vermelho representam as Áreas de Pesquisa, enquanto os nós adjacentes em azul representam os depositantes das patentes: empresas e universidades. Os nós verdes foram utilizados para representar as afiliações dos artigos. Os pesos de cada aresta representam o número de documentos que cada entidade possui no conjunto de IPCs da rede bipartida.

¹https://github.com/Matheusmno/TGP2/blob/main/Affiliation_Assignee_RA.html

3.2.3 Rede de Organizações Financiadoras–Depositantes–Áreas de Pesquisa

2

Este grafo ilustra as conexões entre as Áreas de Pesquisa (Research Areas - RAs), as agências de fomento localizadas no Brasil que financiaram os artigos, e os depositantes de patentes derivados do conjunto de dados. Ao mapear essas relações, o grafo revela quais agências de fomento estão mais ativas no apoio a áreas específicas de estudo, fornecendo assim insights sobre tendências de financiamento e prioridades no panorama da pesquisa brasileira.

3.2.4 Rede de ISICs de Organizações Financiadoras Brasileiras–Depositantes–Áreas de Pesquisa

3

Este grafo ilustra as conexões entre as RAs, as agências de fomento localizadas no Brasil que financiaram os artigos, e os depositantes de patentes derivados do conjunto de dados. A única diferença entre este grafo e o anterior é que apenas as RAs relacionadas às atividades das empresas são exibidas, a fim de fornecer uma visualização mais precisa do cenário das empresas brasileiras.

3.2.5 Redes Multicamadas

Para redes multicamadas (ou multiplex), Maulana and Atzmueller (2021) define um conjunto de camadas – modelando conjuntos de arestas correspondentes a relações, denotadas por $E_l \subseteq E$, $l \in \{1, \dots, m\}$, onde m indica o número de camadas. Uma rede multiplex G_M pode então ser representada formalmente como: $[G_M = (G_1, G_2, \dots, G_l, \dots, G_m)]$, onde $G_i = (V_i, E_i)$, $V_i \subseteq V$.

Cada rede G_l é representada pela matriz de adjacência A_l , com os elementos $a_{l_{ij}}$, sendo $a_{l_{ij}} > 0$ se houver um peso positivo no vínculo entre o par de nós v_{il} e v_{jl} , $v_{il}, v_{jl} \in E_l$ na camada l , e $a_{l_{ij}} = 0$ caso contrário (Maulana and Atzmueller, 2021). Para simplificar a formalização de redes multiplex ponderadas, pode-se considerar apenas valores inteiros positivos ou zero com relação ao vínculo entre quaisquer pares desses nós v_{il} e v_{jl} na camada l .

3.2.6 Bibliotecas Python e Ferramentas de Visualização

A pesquisa emprega diversas bibliotecas e ferramentas Python para facilitar a gestão de dados, geração de grafos e análise interativa. Entre as principais bibliotecas utilizadas está o Pandas, que atua como uma ferramenta poderosa para o gerenciamento e processamento de dados. O Pandas oferece estruturas de dados versáteis e funções para manipulação de tabelas numéricas e séries temporais, permitindo o tratamento eficiente e a análise de grandes conjuntos de dados, essenciais para o processo de pesquisa. Além disso, o NetworkX desempenha um papel central na criação e análise de estruturas de grafos complexas. Essa biblioteca facilita a construção, manipulação e exploração da estrutura, dinâmica e funções de redes in-

²https://github.com/Matheusmno/TGP2/blob/main/FundingOrgs_Assignee_RA.html

³https://github.com/Matheusmno/TGP2/blob/main/BR_ISICS_Affiliation_Assignee_RA.html

tricadas. Seu framework robusto para aplicações em ciência de redes a torna indispensável para a análise gráfica realizada neste estudo. Para complementar essas funcionalidades, PyVis e Plotly são utilizadas para renderizar os grafos gerados pelo NetworkX em arquivos HTML interativos. Essas ferramentas permitem a criação de visualizações dinâmicas e interativas, possibilitando que os pesquisadores explorem e analisem as estruturas de grafos de forma intuitiva. A natureza interativa dessas visualizações aprimora a compreensão dos dados subjacentes e dos relacionamentos capturados no grafo.

Ao integrar essas bibliotecas e ferramentas, o estudo assegura um tratamento de dados eficiente, geração robusta de grafos e visualizações envolventes, que são vitais para atingir os objetivos da pesquisa, como ilustrado pela Figura 3.1.

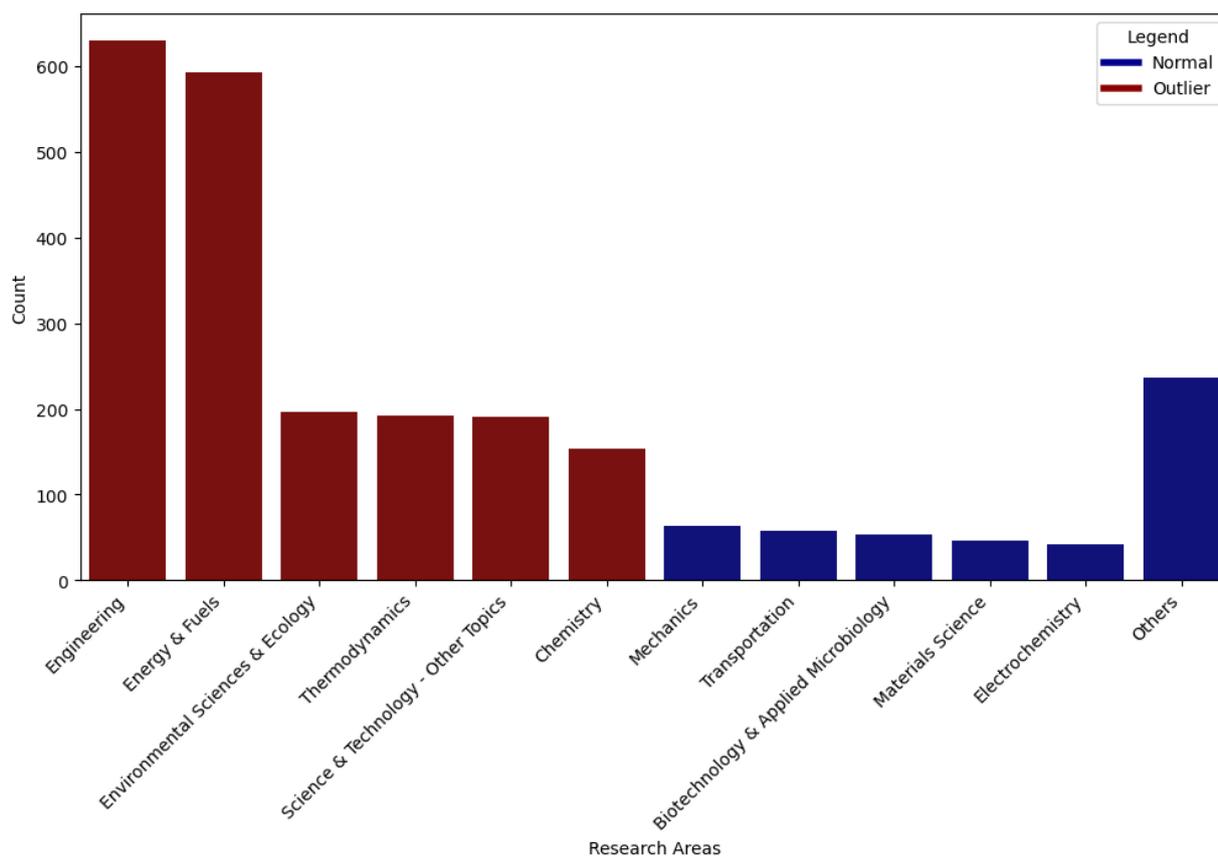


Figura 3.2: Histograma das áreas de pesquisa relacionadas aos artigos consultados

Essa metodologia, que utiliza as plataformas Derwent Innovation e Web of Science, aliada a técnicas de ciência de dados, tornou possível uma investigação aprofundada do universo de patentes e artigos sobre SAF. O arcabouço metodológico permitiu identificar colaborações entre clusters, tendências tecnológicas e aglomerados inovadores dentro da indústria de aviação sustentável.

3.3 CORRESPONDÊNCIA PATENTE-ARTIGO

Neste estudo, exploramos cinco modelos distintos de classificação de texto — Naive Bayes, Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), Árvore de Decisão e Floresta Aleatória (Random Forest). Esses modelos foram treinados utilizando aproximadamente 1.000 artigos de cada uma das classes mais frequentes (extraídas da Figura 3.2) dentro do conjunto de dados de artigos, conforme ilustrado pela Figura 3.3 (os artigos consultados que não possuíam título, resumo ou palavras-chave dos autores foram descartados), e avaliados para identificar padrões nos dados. Entre eles, a Support Vector Machine (Support Vector Machine - SVM) demonstrou a maior acurácia (76,8%), superando os demais classificadores. Essa análise destaca a eficácia da SVM em lidar com a complexidade do conjunto de dados e fornece uma avaliação comparativa do desempenho de diversas abordagens de aprendizado de máquina no contexto da pesquisa sobre SAF. Os resultados, apresentados por meio de matrizes de confusão detalhadas na Figura 3.4, oferecem insights valiosos sobre o desempenho de classificação de cada modelo.

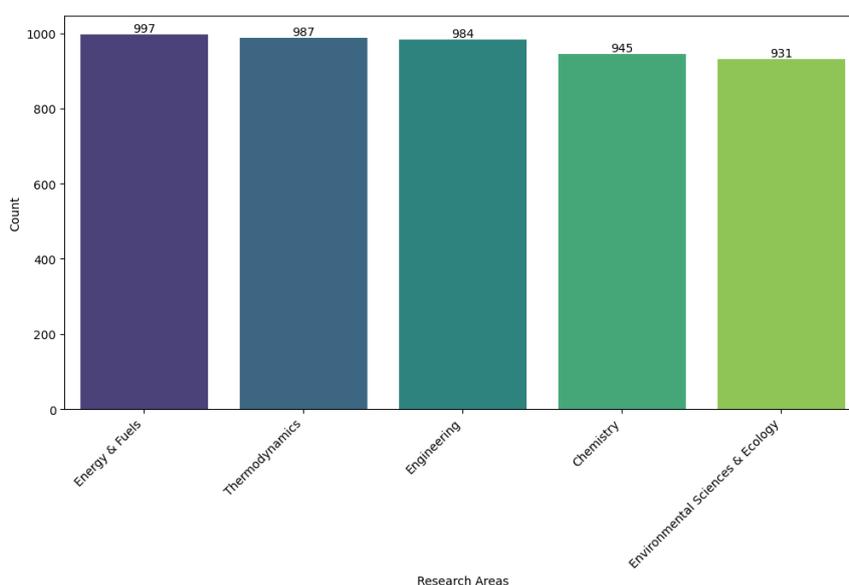
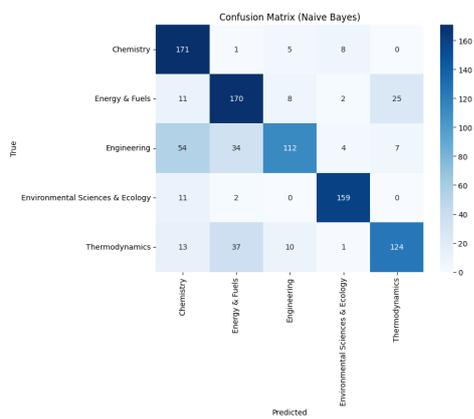


Figura 3.3: Áreas de pesquisa com maior ocorrência utilizadas para o treinamento dos modelos

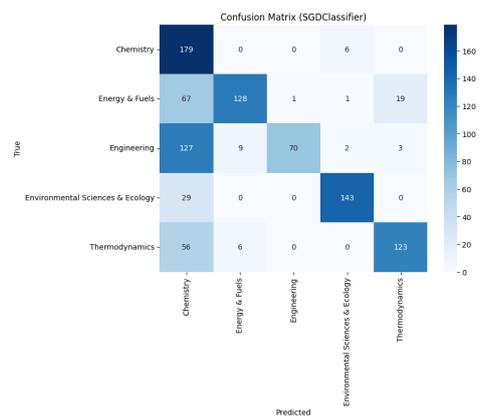
3.4 ESTRUTURA DA REDE SOCIAL

A rede apresentada na Figura 3.5 delinea as afiliações de artigos acadêmicos, os responsáveis por patentes e as respectivas Áreas de Pesquisa (Research Areas - RA). O grafo gerado demonstra que as afiliações dos artigos acadêmicos estão representadas em verde, os responsáveis pelas patentes são mostrados em azul, e as RAs são indicadas por caixas vermelhas (abrir os arquivos HTML para melhor visualização).

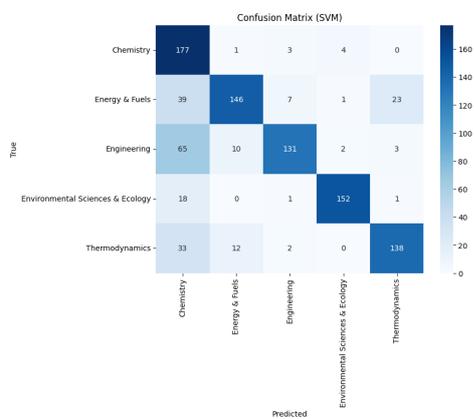
A representação em cores escolhida permite compreender os relacionamentos e interações entre essas entidades. Dessa forma, o mapeamento desses elementos em uma rede torna possível analisar a interconectividade e os padrões de colaboração no contexto de artigos científicos e patentes. Além disso, esse tipo de



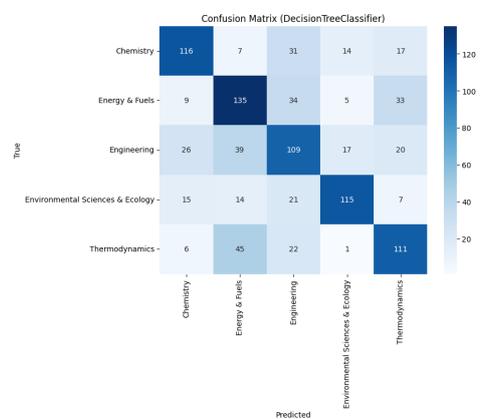
(a) Matriz de Confusão para Naive Bayes



(b) Matriz de Confusão para SGD



(c) Matriz de Confusão para SVM



(d) Matriz de Confusão para Árvore de Decisão



(e) Matriz de Confusão para Floresta Aleatória

Figura 3.4: Matrizes de Confusão para todos os modelos treinados

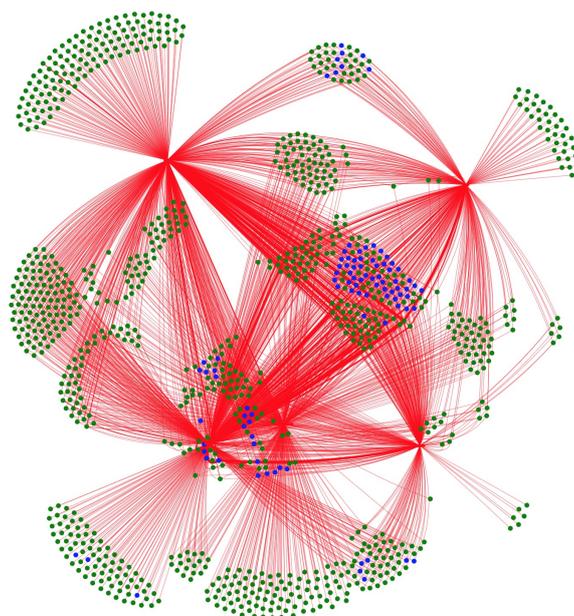


Figura 3.5: Visão geral da rede de responsáveis por patentes e afiliações de artigos

Área de Pesquisa	Centralidade de Grau
Engenharia	0.6669
Energia & Combustíveis	0.6324
Ciências Ambientais & Ecologia	0.3470

Tabela 3.1: Três maiores medidas de centralidade de grau para nós de Área de Pesquisa na rede de responsáveis por patentes e afiliações de artigos

visualização facilita a identificação de atores-chave e o fluxo de conhecimento entre diferentes domínios, fornecendo informações sobre a estrutura e a dinâmica dos ecossistemas de inovação.

A Tabela 3.1 apresenta os três principais nós de RA classificados de acordo com suas medidas de centralidade de grau dentro da rede. A centralidade de grau, que representa o número de conexões diretas que um nó possui, é um indicador crucial da influência de um nó na rede.

O nó de Área de Pesquisa (RA) “Engenharia” apresenta a maior centralidade de grau, com 0,6669, destacando seu papel fundamental e ampla conectividade dentro da rede. Em seguida, “Energia & Combustíveis” possui uma centralidade de grau de 0,6324, refletindo sua influência significativa. Em contraste, o nó “Ciências Ambientais & Ecologia” apresenta uma centralidade bem menor, de 0,3470, indicando menor centralidade e influência em comparação com “Engenharia” e “Energia & Combustíveis”. Esses achados ressaltam os diferentes níveis de conectividade e importância entre os nós de RA, com “Engenharia” e “Energia & Combustíveis” se destacando como centros críticos da rede.

3.5 OPORTUNIDADES POTENCIAIS NO MERCADO BRASILEIRO

O número de artigos publicados pode indicar o grau de interesse e atividade em determinada tecnologia. Essa correlação decorre do fato de que publicações acadêmicas e industriais frequentemente refletem tendências emergentes e inovações. Quando uma nova tecnologia é introduzida, pesquisadores e profissionais exploram seu potencial, levando a um aumento inicial de artigos que antecede sua aplicação comercial em larga escala e adoção generalizada. Assim, um aumento no número de publicações indica crescente curiosidade, financiamento e investimento intelectual. As publicações durante essa fase frequentemente focam em superar desafios técnicos, melhorar a eficiência e encontrar aplicações práticas.

A Figura 3.6 ilustra as conexões entre organizações financiadoras de pesquisas brasileiras (representadas em verde) e responsáveis por patentes (em azul), ligadas por meio de suas respectivas Áreas de Pesquisa (Research Areas - RAs), mostradas em vermelho. Essa rede em camadas foi desenvolvida para fornecer uma visualização tridimensional das relações facilitadas pelas RAs, oferecendo uma visão abrangente da interação entre financiamento, pesquisa e propriedade intelectual.

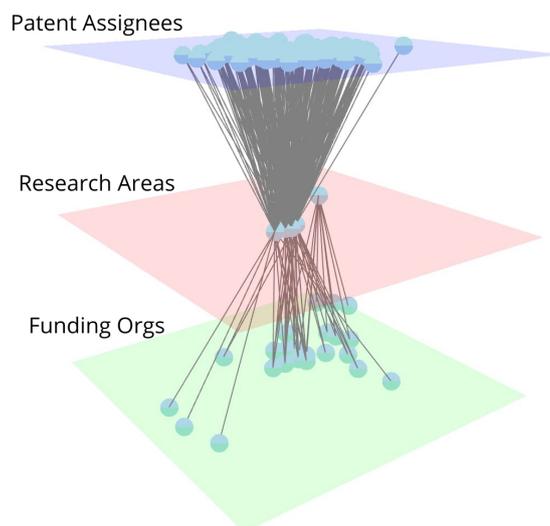


Figura 3.6: Rede em camadas de organizações financiadoras brasileiras e responsáveis por patentes

Uma vez que a tecnologia amadurece e demonstra viabilidade comercial, o foco se desloca da publicação de artigos para o depósito de patentes. As patentes são essenciais para proteger a propriedade intelectual e garantir vantagens competitivas no mercado. A transição de um alto volume de artigos para um aumento no número de patentes marca o avanço da tecnologia da pesquisa teórica para a implementação prática. No caso dos SAFs no Brasil, nenhuma patente está registrada no banco de dados Derwent Innovation. Ainda assim, é possível observar um aumento no número de artigos discutindo as perspectivas de implementação dessa tecnologia no cenário brasileiro. Isso indica um interesse crescente pelo tema que pode levar ao desenvolvimento comercial de biocombustíveis para aviação no Brasil, especialmente se houver cooperação com empresas internacionais do setor.

A partir da análise das entidades financiadoras nos novos artigos brasileiros, é possível perceber quais

áreas estão mais próximas do interesse comercial. Isso pode ser explorado em futuras colaborações ou investimentos. Para melhor visualização, a Figura 3.7 mostra a mesma rede em camadas apresentada na Figura 3.6, mas em formato 2D, o que permite uma identificação mais fácil de clusters e conexões.

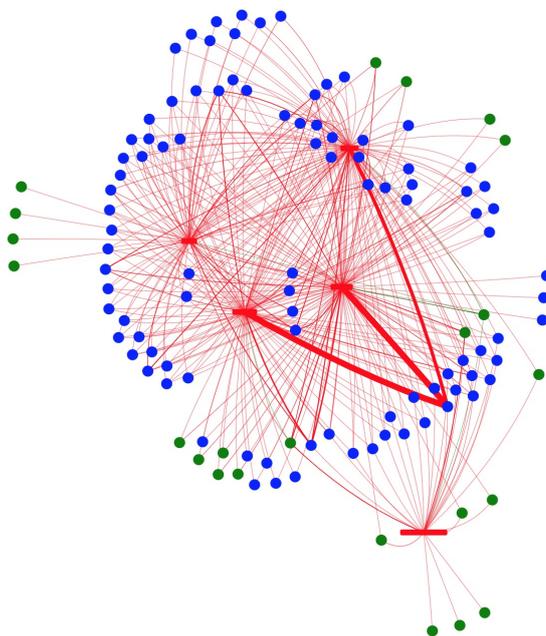


Figura 3.7: Rede de organizações financiadoras brasileiras e responsáveis por patentes

A Figura 3.8 ilustra o caso singular da empresa Rolls Royce (RORO-C), que está conectada às principais Áreas de Pesquisa da Tabela 3.1 e apresenta arestas espessas, indicando sua predominância na área. Diferentemente dos demais responsáveis por patentes, que geralmente possuem no máximo quatro patentes registradas, a Rolls Royce conta com 30 patentes registradas relacionadas a SAF.

A rede revela que “Ciências Ambientais & Ecologia” continuam recebendo contribuições mais significativas da academia em comparação com a indústria. Isso sugere que a indústria de SAFs deve prestar atenção aos temas abordados em artigos recentes publicados nessa área de pesquisa. Curiosamente, a Rolls Royce não possui patentes registradas relacionadas a esse campo, e a rede representada na Figura 3.8 pode fornecer orientações valiosas sobre possíveis direções para seus futuros esforços de pesquisa e desenvolvimento.

Os nós azuis representam empresas individuais, instituições de pesquisa e entidades não padronizadas, enquanto os nós verdes denotam organizações e órgãos financiadores. As arestas vermelhas indicam conexões fortes entre os nós de Área de Pesquisa (Research Area - RA) e outras entidades, destacando a posição central de “Engenharia” e “Energia & Combustíveis” na facilitação de interações na rede. Essa centralidade ressalta sua importância na estrutura e dinâmica do cenário brasileiro de financiamento à pesquisa, evidenciando sua influência na promoção de colaborações e alocação de recursos.

Adicionalmente, a Figura 3.9 apresenta uma rede de áreas de pesquisa conectadas aos ISICs das organizações financiadoras brasileiras, com foco específico nos ISICs registrados 191, 192 e 201. Essa visualização fornece uma visão geral do panorama atual dentro desses setores, que pode ser de interesse

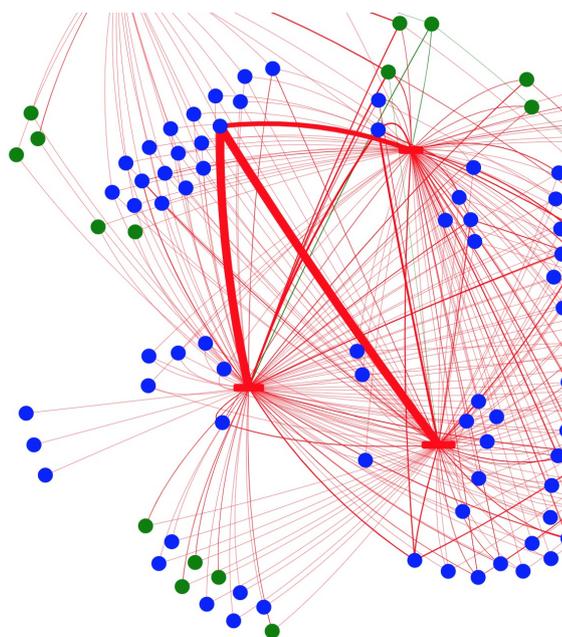


Figura 3.8: Rede de organizações financiadoras brasileiras e responsáveis por patentes com foco em RORO-C

particular para essas organizações. O nó vermelho central representa a principal área de pesquisa (RA), com arestas vermelhas se estendendo para diversos nós verdes, indicando forte atividade de pesquisa global nesses campos. No entanto, o número relativamente baixo de patentes registradas nessas áreas evidencia uma possível lacuna entre a produção científica e a aplicação comercial, sugerindo oportunidades para inovação e desenvolvimento de propriedade intelectual.

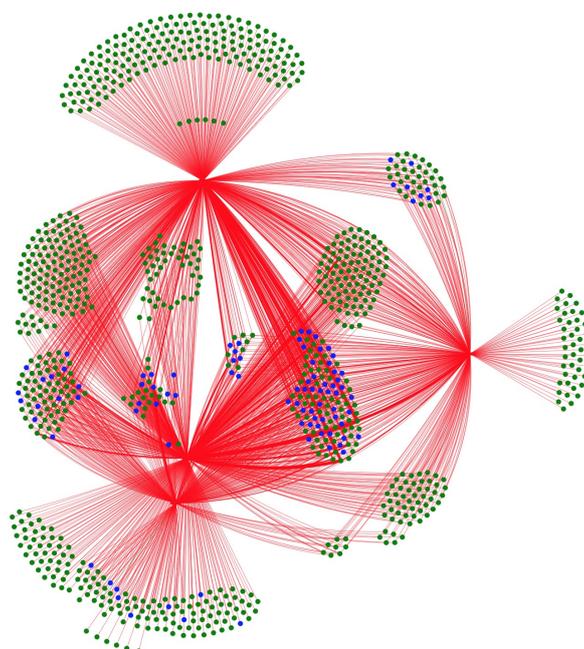


Figura 3.9: Rede de responsáveis por patentes e afiliações de artigos filtrada pelos ISICs correspondentes das organizações financiadoras brasileiras

Área de Pesquisa	Centralidade de Grau
Energia & Combustíveis	0,8644
Engenharia	0,7966
Termodinâmica	0,7627

Tabela 3.2: Top 3 medidas de centralidade de grau para nós de Área de Pesquisa da rede de organizações financiadoras brasileiras e responsáveis por patentes

O denso agrupamento central destaca o papel fundamental da Área de Pesquisa (Research Area - RA) na coordenação e ligação de diversas entidades dentro dos ISICs especificados, ressaltando sua importância na promoção da colaboração e na distribuição de recursos nesses setores-alvo. Essa representação detalhada auxilia na compreensão das complexas relações e interações entre os principais atores nessas áreas especializadas.

Para identificar essas oportunidades com mais precisão, foi medida a centralidade de grau das áreas de pesquisa (RAs) nos artigos brasileiros, a fim de determinar as mais proeminentes. As três principais RAs estão detalhadas na Tabela 3.2, que inclui suas descrições e medidas de centralidade de grau. Entre elas, a área “Energia & Combustíveis” apresenta a maior centralidade, com 0,8644, indicando que é a mais interconectada e potencialmente a mais influente da rede. Adicionalmente, os nós verdes na Figura 3.9 representam as organizações financiadoras brasileiras que apoiaram pesquisas durante o período analisado. Isso comprova que o Brasil possui uma atuação mais expressiva na área de pesquisa “Energia & Combustíveis” em comparação com as demais.

Outro ponto interessante revelado pela rede da Figura 3.9 é seu potencial para servir como um mapa para que as organizações financiadoras acompanhem para onde seus investimentos estão sendo direcionados. Além disso, atua como uma ferramenta valiosa para orientar empresas nacionais em direção a regiões onde o conhecimento está concentrado em áreas de pesquisa específicas de interesse. Essa rede também pode ajudar a identificar patentes que ainda não foram protegidas dentro de sua jurisdição, iluminando, assim, novas oportunidades para inovação e desenvolvimento.

4 DO DADO À DECISÃO: ANÁLISE DE GRAFOS COM MODELOS DE LINGUAGEM

O uso de modelos de linguagem de grande porte (Large Language Models - LLMs) tem se tornado cada vez mais presente no cotidiano da sociedade, tanto na resolução de tarefas complexas quanto em consultas simples e acessíveis por meio de mecanismos de busca como o Google Search. Diante desse avanço, propôs-se a introdução dos LLMs na análise de grafos como estratégia para facilitar a interpretação de dados complexos por parte de usuários não especializados.

Com base nos resultados obtidos nos capítulos 2 e 3, os grafos gerados foram fornecidos a um LLM, permitindo que tomadores de decisão realizassem perguntas diretamente sobre a rede e extraíssem insights das conexões, sem a necessidade de conhecimento técnico prévio. Isso é possível porque os modelos atuais são capazes de interpretar dados estruturados, o que viabiliza a análise automatizada de visualizações interativas e a geração de pareceres técnicos em linguagem acessível.

Essa abordagem representa um avanço significativo na forma como se acessam e interpretam estruturas complexas de dados. Ao permitir que grafos sejam explorados por meio de linguagem natural, os LLMs reduzem barreiras de entrada e democratizam o acesso a análises sofisticadas. Como resultado, gestores públicos, formuladores de políticas e demais profissionais podem identificar padrões, gargalos e oportunidades diretamente das redes, sem depender exclusivamente de especialistas em ciência de dados ou análise de redes.

Adicionalmente, a utilização de LLMs como interfaces interativas para consulta e exploração de dados em rede inaugura uma nova forma de interação entre pesquisadores e resultados analíticos, aproximando ciência e prática. A combinação entre visualizações avançadas e respostas contextuais automatizadas abre caminho para sistemas de apoio à decisão mais dinâmicos e responsivos. Essa integração tem o potencial de transformar a maneira como instituições operam com dados estratégicos, abrangendo desde diagnósticos de inovação até o acompanhamento de políticas públicas, e consolidando a inteligência artificial como uma aliada no planejamento de ações concretas e embasadas em dados.

4.1 LEIAUTE DA SOLUÇÃO

A solução proposta foi desenvolvida utilizando a biblioteca Streamlit, que permite a criação rápida de interfaces web interativas em Python. A aplicação foi estruturada com o objetivo de possibilitar que usuários façam o upload de arquivos contendo informações dos grafos gerados por bibliotecas como Pyvis e NetworkX. Esses grafos são renderizados diretamente na interface, ao mesmo tempo em que seus dados estruturais — como nós e arestas — são extraídos, permitindo que tais informações sirvam como contexto para interações via linguagem natural com o modelo de linguagem da *OpenAI*.

Uma vez carregado o grafo, o usuário pode enviar perguntas diretamente relacionadas à rede apre-

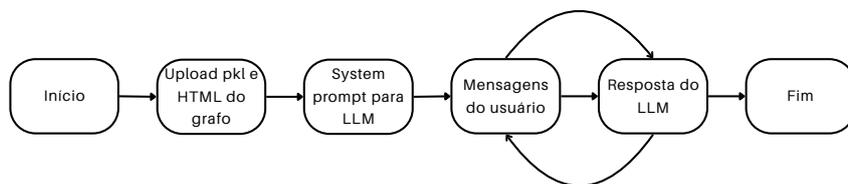


Figura 4.1: Framework do aplicativo web

sentada, utilizando um campo de texto na interface. Internamente, o modelo é instruído com um prompt de sistema que fixa o conteúdo do grafo como base para todas as respostas subsequentes. O histórico de mensagens é mantido em sessão para garantir fluidez na conversa, e o modelo retorna interpretações que facilitam o entendimento dos dados apresentados, mesmo por profissionais sem conhecimento técnico prévio em análise de redes. Dessa forma, a ferramenta atua como uma ponte entre representações complexas e a tomada de decisão baseada em dados. A Figura 4.1 ilustra o fluxograma da aplicação web, que faz interações com o usuário indefinidamente.

4.2 INTEGRAÇÃO ENTRE VISUALIZAÇÃO DE REDES E MODELOS DE LINGUAGEM

Para apoiar a análise interativa de redes complexas formadas por áreas de pesquisa e instituições financiadoras, foi desenvolvido um aplicativo web utilizando a biblioteca *Streamlit* em *Python*. O objetivo central desse aplicativo é proporcionar ao usuário um ambiente no qual seja possível visualizar a estrutura de um grafo bipartido, bem como interagir com um modelo de linguagem natural (LLM) para interpretar essa rede de forma assistida.

O sistema permite que o usuário carregue dois arquivos: um grafo previamente salvo em formato `.pkl` (como objeto `networkx.Graph`) e sua respectiva visualização em HTML (gerada com a biblioteca *PyVis*). O grafo carregado deve possuir a propriedade `bipartite` associada a cada nó, em que o valor 1 indica uma área de pesquisa (*Research Area*) e o valor 0 representa uma organização financiadora (*Funding Organization*).

Uma vez carregado o grafo, o sistema extrai automaticamente métricas de centralidade para os nós — incluindo grau, *betweenness*, *closeness* e centralidade de autovetor (*eigenvector centrality*) — e identifica

os nós mais centrais da rede. Além disso, uma análise adicional mapeia as conexões diretas entre áreas de pesquisa e suas respectivas instituições financiadoras. Para cada área de pesquisa presente no grafo, é gerada uma lista das organizações que estão diretamente conectadas a ela via arestas.

Essas informações são então estruturadas em formato JSON e enviadas ao modelo de linguagem como parte do *system prompt*. O modelo é instruído a tratar os nós do tipo `Research Area` como tópicos científicos e os do tipo `Funding Organization` como atores financiadores. Com base nessa estrutura, o LLM é capaz de responder a perguntas complexas sobre a rede, tais como identificar os tópicos mais centrais, as instituições com maior capacidade de fomento, ou sugerir oportunidades estratégicas de cooperação entre atores e áreas específicas.

Para otimizar o uso de tokens e respeitar os limites da API do modelo de linguagem, apenas os 15 nós mais centrais da rede (segundo a soma normalizada de todas as métricas de centralidade - `score_composto`) são incluídos no contexto enviado ao LLM. Essa estratégia garante um equilíbrio entre riqueza informacional e desempenho computacional.

O *system prompt* utilizado para configurar o modelo de linguagem na solução proposta foi cuidadosamente elaborado para orientar a interpretação contextual da rede. O conteúdo da instrução inicial foi o seguinte:

```
Você é um especialista em análise de redes científicas.
```

```
O grafo analisado é bipartido, com dois tipos de nós:
```

- `Research Areas` (`bipartite = 1`)
- `Funding Organizations` (`bipartite = 0`)

```
As conexões representam relações de financiamento ou colaboração.
```

```
Resumo do grafo:
```

- Total de nós: `{num_nos}`
- Total de arestas: `{num_arestas}`

```
Top {top_k} nós com maior score composto:
```

```
{top_nos_txt}
```

```
Nó mais conectado: {top_node}
```

- Vizinhos principais: `{vizinhos_txt}`

```
Conexões relevantes entre áreas e organizações:
```

```
{conexoes_txt}
```

```
Com base nas informações acima, forneça:
```

- Os hubs mais importantes da rede e o motivo
- Quais áreas estão fortemente ou fracamente conectadas

- Oportunidades de cooperação que se destacam
- Qualquer padrão ou concentração de funding que possa ser explorado

Baseie suas respostas exclusivamente no contexto fornecido.

Essa formulação tem como objetivo maximizar a coerência e a profundidade das respostas geradas pelo modelo, ao mesmo tempo em que garante que a análise seja restrita aos dados estruturais do grafo, sendo uma abordagem que permite ao usuário explorar a estrutura da rede de forma interativa e assistida, combinando visualização gráfica e interpretação contextual com apoio de inteligência artificial.

Tabela 4.1: Descrição das variáveis utilizadas no prompt de análise de redes científicas

Variável	Descrição
num_nos	Número total de nós no grafo bipartido, incluindo tanto áreas de pesquisa quanto organizações de fomento.
num_arestas	Número total de arestas no grafo, representando conexões de financiamento ou colaboração entre áreas e organizações.
top_k	Quantidade de nós com maior score composto a serem destacados no resumo da análise.
top_nos_txt	Texto com os top_k nós de maior score composto, incluindo seus nomes, tipos e valores de score.
top_node	Nome do nó com o maior score composto, considerado o mais relevante da rede.
vizinhos_txt	Lista textual dos principais vizinhos conectados ao top_node, separados por vírgulas.
conexoes_txt	Texto com pares do tipo <i>Área de Pesquisa: Organização</i> , representando conexões relevantes entre nós dos dois tipos.

A Tabela 4.1 fornece a descrição de cada uma das variáveis calculadas para a criação do *system prompt* acima referido.

A Figura 4.2 mostra o leiaute da interface de usuário da solução proposta, com uma caixa para apresentar o grafo e uma caixa de texto para passar as perguntas do usuário para a LLM.

ANÁLISE DOS RESULTADOS DA INTERAÇÃO COM A LLM

A seguir, são apresentados trechos de interações realizadas com o modelo de linguagem, com base no grafo carregado (FundingOrgs_Assignee_Research Areas). As perguntas foram elaboradas por um tomador de decisão, e as respostas geradas automaticamente pelo LLM da *OpenAI* com base no contexto da rede. Abaixo estão listadas algumas das perguntas efetuadas para o modelo gpt-3.5-turbo sobre a referida rede:



Figura 4.2: Dashboard de análise de redes assistida por LLM

Pergunta: “Quais são os principais atores apoiando a pesquisa em Thermodynamics?”

Resposta da LLM:

Os principais atores apoiando a pesquisa em Thermodynamics são:

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Embraer, Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), Klabin, Petrobras, Suzano, ACHA-Individual, ALLU-C, BAGW-Individual e BATT-C.

Essas organizações estão fortemente conectadas à área de pesquisa em Termodinâmica, indicando um alto nível de investimento e interesse nessa área.

Avaliação da resposta da LLM:

Todos os atores citados na resposta aparecem no grafo como nós proeminentes no setor de Termodinâmica, resposta coerente.

Pergunta: “Quais áreas têm maior potencial de colaboração com instituições brasileiras?”

Resposta da LLM:

As áreas de pesquisa com maior potencial de colaboração com instituições brasileiras são “Energy & Fuels”, “Engineering”, “Thermodynamics” e “Chemistry”. Essas áreas possuem um alto grau de conexões no grafo, indicando que estão fortemente ligadas a várias organizações de financiamento.

A área de “Energy & Fuels” tem o maior grau de conexões (102), seguida por “Engineering”(94),

”Thermodynamics”(90) e ”Chemistry”(71). Isso sugere que essas áreas são de grande interesse para as organizações de financiamento e, portanto, têm um alto potencial para colaborações.

Além disso, a área de ”Energy & Fuels”tem a maior centralidade de autovetor (0.3857), o que indica que está conectada a nós muito bem conectados no grafo. Isso sugere que essa área pode ser um ponto de acesso importante para outras áreas de pesquisa e organizações de financiamento.

As organizações de financiamento brasileiras mais conectadas a essas áreas de pesquisa incluem a ”Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)”, ”Conselho Nacional de Pesquisa e Desenvolvimento Científico (CNPq)”, ”Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP)”e ”Financiadora de Estudos e Projetos (FINEP)”. Essas organizações têm um alto potencial para colaborar com pesquisadores nessas áreas.

Avaliação da resposta da LLM:

Todas as instituições citadas de fato são grandes potenciais colaboradores da indústria, visto que fomentam a pesquisa aplicada. As áreas citadas foram corretamente identificadas e os números de arestas batem com os metadados da rede em questão.

Pergunta: “Quem são os hubs de financiamento que atuam em diversas frentes?”

Resposta da LLM:

Os hubs de financiamento que atuam em diversas frentes são as organizações que possuem conexões com várias áreas de pesquisa. Com base nos dados, as seguintes organizações se destacam:

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES): Esta organização tem conexões com as áreas de ”Energy & Fuels”, ”Engineering”, ”Chemistry”, ”Environmental Sciences & Ecology”e ”Thermodynamics”.

Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP): Esta organização tem conexões com as áreas de ”Energy & Fuels”, ”Engineering”, ”Chemistry”e ”Environmental Sciences & Ecology”.

Conselho Nacional de Pesquisa e Desenvolvimento Científico (CNPq): Esta organização tem conexões com as áreas de ”Energy & Fuels”, ”Engineering”, ”Chemistry”e ”Environmental Sciences & Ecology”.

Financiadora de Estudos e Projetos (FINEP): Esta organização tem conexões com as áreas de ”Energy & Fuels”, ”Engineering”e ”Chemistry”.

Essas organizações são hubs de financiamento importantes porque estão conectadas a várias áreas de pesquisa, o que indica que elas financiam uma ampla gama de estudos e projetos.

Avaliação da resposta da LLM:

Todas as conexões indicadas na resposta existem e estão acuradas.

Pergunta: ”Quais as principais conexões do grafo e quais os principais *players* inovadores que poderiam cooperar com a indústria brasileira?”

Resposta da LLM:

As principais conexões do grafo estão centradas em torno de cinco áreas de pesquisa: Energy & Fuels, Engineering, Thermodynamics, Chemistry e Environmental Sciences & Ecology. Essas áreas têm o maior número de conexões com organizações de financiamento, indicando que são áreas de pesquisa ativas e potencialmente inovadoras.

Os principais players inovadores no grafo são as organizações de financiamento que têm o maior número de conexões com essas áreas de pesquisa. Entre eles estão a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), MOHA-Individual, UNBA-C, Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) e o Conselho Nacional de Pesquisa e Desenvolvimento Científico (CNPq).

Essas organizações têm um alto grau de centralidade no grafo, o que indica que elas têm um papel central na rede de pesquisa e inovação. Elas estão bem posicionadas para cooperar com a indústria brasileira, pois têm acesso a uma ampla gama de áreas de pesquisa e podem facilitar a transferência de conhecimento e tecnologia.

Além disso, algumas organizações internacionais, como a Boeing e a German Agency for International Cooperation (Deutsche Gesellschaft für Internationale Zusammenarbeit - GiZ), também aparecem como conexões importantes para algumas áreas de pesquisa. Essas organizações podem representar oportunidades de cooperação internacional para a indústria brasileira.

Em particular, a área de Energy & Fuels tem conexões com uma ampla gama de organizações de financiamento, incluindo a ANP, a Boeing, a Embraer e a FINEP. Isso sugere que essa área de pesquisa é de particular interesse para a indústria brasileira e pode oferecer muitas oportunidades para inovação e cooperação.

Avaliação da resposta da LLM:

As informações apresentadas pela LLM estão condizentes com o grafo passado para análise e há sugestões razoáveis e condizentes com a realidade do mercado. Foi apontada uma possibilidade de cooperação internacional em projetos de infraestrutura e desenvolvimento de tecnologia entre a Boeing e a empresa GiZ, que é uma empresa pública alemã que trabalha em cooperação internacional com países em desenvolvimento. A GiZ oferece financiamento e apoio técnico em diversas áreas, como agricultura, saúde, educação e proteção do meio ambiente. Portanto, demonstra-se coerente a sugestão da LLM em firmar uma parceria entre as duas empresas.

Ao analisarmos os resultados obtidos acima, observa-se que a LLM foi capaz de interpretar corretamente os dados da rede bipartida carregada e de gerar respostas coerentes e fundamentadas. As respostas apresentaram não apenas os principais nós (organizações de fomento ou áreas de pesquisa) com base em métricas de centralidade e grau de conexão, mas também realizaram inferências plausíveis quanto ao potencial de colaboração, relevância temática e oportunidades de inovação.

A capacidade do modelo de linguagem em identificar corretamente os principais atores e áreas mais conectadas do grafo, bem como sugerir colaborações estratégicas com base nessas conexões, demonstra o potencial de sua aplicação como ferramenta auxiliar na análise de redes complexas voltadas ao planejamento de políticas públicas de ciência, tecnologia e inovação.

Além disso, vale destacar que a LLM não apenas recupera informações explícitas do grafo, mas também apresenta sínteses e contextualizações que podem contribuir com a tomada de decisão, como visto na identificação de possíveis parcerias internacionais e no destaque a áreas com alto potencial inovador.

Portanto, conclui-se que a interação entre um tomador de decisão e uma LLM, quando orientada por dados estruturados e bem modelados, como um grafo bipartido de organizações e áreas de pesquisa, pode representar uma importante ferramenta de apoio à inteligência estratégica em ciência e tecnologia.

5 CONCLUSÃO

Este estudo avançou a compreensão sobre a inovação em Combustíveis Sustentáveis de Aviação (Sustainable Aviation Fuel - SAF) por meio de uma integração inédita entre análise de redes de patentes e redes sociais de colaboração. Ao unir metodologias desenvolvidas em dois estudos fundamentais, foi criado um arcabouço unificado capaz de revelar tanto as trajetórias tecnológicas quanto as dinâmicas institucionais que moldam o desenvolvimento do SAF globalmente, com especial destaque para o potencial ainda inexplorado do Brasil. Os achados demonstram que o aparente paradoxo da posição secundária do Brasil na inovação em SAF, apesar de sua liderança em biocombustíveis, não decorre de limitações tecnológicas, mas sim de lacunas estruturais na tradução do conhecimento entre academia e indústria—uma constatação só possível graças à abordagem analítica dual adotada.

A análise de redes de patentes, fundamentada em teoria dos grafos desde os trabalhos pioneiros de Euler até aplicações contemporâneas com aprendizado de máquina (Henning and van Vuuren, 2022; Maskittou et al., 2021), identificou três principais clusters de inovação no desenvolvimento global de SAF: (1) tecnologias de hidroprocessamento para matérias-primas lipídicas, (2) rotas via gaseificação e síntese Fischer-Tropsch para materiais lignocelulósicos e (3) métodos emergentes de conversão eletroquímica. De forma crítica, as medidas de centralidade aprimoradas revelaram que menos de 12% dos principais códigos da Classificação Internacional de Patentes (International Patent Classification - IPC) presentes nesses clusters apresentam conexões significativas com instituições de pesquisa brasileiras, o que explica a ausência do país nos depósitos de patentes de SAF, apesar de sua reconhecida expertise em biocombustíveis (Mosquera et al., 2024; Metcalf and Casey, 2016).

Complementarmente, a análise de redes sociais das colaborações expôs desconexões críticas no ecossistema de inovação brasileiro. Embora o país mantenha redes acadêmicas robustas em pesquisa de bioenergia (Jha and Schmidt, 2021), estas operam de forma isolada da P&D industrial—em contraste com as parcerias universidade-indústria integradas que impulsionam a inovação em SAF nos Estados Unidos e na Europa (Korkut and Fowler, 2021). Os modelos bipartidos de rede mostraram ainda que empresas brasileiras priorizam melhorias incrementais no biodiesel em detrimento de tecnologias disruptivas como o SAF, em função de estruturas de incentivos desalinhadas e lacunas nos marcos regulatórios (Nicola et al., 2023).

A sinergia metodológica entre essas abordagens gerou ferramentas práticas para superar esses entraves. O classificador por aprendizado de máquina, treinado com dados de patentes e artigos acadêmicos, atingiu 76,8% de acurácia na correspondência entre domínios de pesquisa e aplicações industriais—um avanço significativo em relação a métodos anteriores de mineração textual (Ding et al., 2019). Integrado ao sistema de apoio à decisão baseado em LLM (Large Language Model), esse modelo permite a identificação de oportunidades de colaboração de alto potencial, como o redirecionamento da expertise brasileira em etanol de cana para rotas do tipo alcohol-to-jet (ATJ), onde o país poderia conquistar vantagem competitiva (Ng et al., 2021).

Três principais implicações de política pública emergem dessa análise integrada. Primeiramente, o

Brasil precisa de intervenções direcionadas para superar seu “vale da morte” da inovação em SAF, como editais que incentivem parcerias academia-indústria em rotas específicas de produção, como HEFA (Hydroprocessed Esters and Fatty Acids) ou ATJ (Alcohol-to-Jet). Em segundo lugar, redes de colaboração internacional devem ser estrategicamente cultivadas, especialmente com parceiros europeus que enfrentam desafios semelhantes quanto à disponibilidade de biomassa. Por fim, os resultados apontam para a necessidade de marcos regulatórios híbridos que combinem mandatos de uso com créditos fiscais à produção, espelhando o modelo norte-americano bem-sucedido (Korkut and Fowler, 2021), mas adaptados ao contexto institucional brasileiro.

Essa pesquisa abre novas frentes para o estudo de ecossistemas de inovação além do SAF. A metodologia desenvolvida mostra como a combinação entre análise de patentes e teoria de redes sociais pode revelar barreiras ocultas à adoção tecnológica em outras transições sustentáveis, como hidrogênio verde e captura de carbono. A escalabilidade do arcabouço foi validada por meio de aplicação preliminar a redes de inovação em diesel renovável, indicando seu potencial uso em setores de difícil descarbonização. Estudos futuros devem explorar a modelagem dinâmica de redes para captar a evolução temporal dos clusters de inovação, especialmente à medida que rotas como power-to-liquid se aproximam da viabilidade comercial.

Em última instância, esta dissertação transforma nossa compreensão sobre a inovação sustentável na aviação ao demonstrar que soluções tecnológicas, por si só, não bastam—é igualmente vital construir as redes que conectam criadores de conhecimento a seus implementadores. Para o Brasil e outras nações líderes em biocombustíveis, essa constatação representa tanto um alerta quanto uma oportunidade: as competências existentes podem se converter em protagonismo no SAF, desde que acompanhadas de esforços intencionais para fortalecer os tecidos conectivos dos ecossistemas de inovação. As ferramentas e achados aqui apresentados oferecem um roteiro para essa integração crucial, contribuindo tanto para a literatura acadêmica quanto para o desafio urgente de descarbonizar a aviação global.

5.1 TRABALHOS FUTUROS

Este estudo estabelece as bases para diversos desdobramentos promissores que podem aprofundar e ampliar a metodologia integrada aqui desenvolvida para análise de redes de inovação em Combustíveis Sustentáveis de Aviação (Sustainable Aviation Fuel - SAF). A partir do presente trabalho, futuras pesquisas devem priorizar o desenvolvimento de modelos dinâmicos de redes capazes de capturar a evolução temporal dos clusters de inovação. Técnicas como os Modelos Exponenciais de Grafos Temporais (Temporal Exponential Random Graph Models - TERGMs) permitiriam acompanhar como avanços tecnológicos e mudanças regulatórias afetam a topologia das redes ao longo do tempo, superando a limitação da análise transversal em retratar padrões longitudinais (Interdonato et al., 2020).

Uma extensão crítica envolve a construção de arcabouços de redes multilayer que analisem simultaneamente: (i) vínculos tecnológicos via citações de patentes e co-classificações IPC (International Patent Classification), (ii) laços colaborativos por co-invenção e coautoria, e (iii) difusão de políticas entre jurisdições. Essa abordagem integrada permitiria verificar empiricamente os ciclos de retroalimentação entre política e tecnologia hipotetizados neste trabalho, além de oferecer uma visão mais holística dos ecossistemas de

inovação. As bases teóricas para essa análise já foram estabelecidas na literatura recente de redes multilayer (Liu et al., 2022), mas requerem adaptação ao contexto específico das transições energéticas sustentáveis.

Avanços nas técnicas de classificação também podem aprimorar significativamente a precisão dos sistemas de correspondência entre patentes e artigos. A incorporação de arquiteturas transformer como variantes do BERT adaptadas a domínios técnicos pode melhorar a ponte conceitual entre documentos acadêmicos e industriais, enquanto redes neurais em grafos (Graph Neural Networks - GNNs) tendem a oferecer desempenho superior na construção de *embeddings*, conjuntos de redes de patentes e publicações. Esses aprimoramentos computacionais devem ser acompanhados do desenvolvimento de grafos de conhecimento específicos para matérias-primas e tecnologias de conversão de SAF, a fim de compor arcabouços analíticos semanticamente mais ricos.

A ampliação do escopo geográfico representa outra prioridade. Estudos comparativos com economias produtoras de biocombustíveis no Sudeste Asiático (óleo de palma) ou na África (cana-de-açúcar) podem revelar se as barreiras estruturais observadas no Brasil são fenômenos localizados ou padrões recorrentes em contextos ricos em recursos, porém com restrições à inovação. Tais análises possibilitariam recomendações políticas mais adaptadas a diferentes realidades institucionais.

Outra frente promissora reside na integração entre métricas de redes e parâmetros de análise técnico-econômica, com o objetivo de estabelecer relações preditivas entre centralidade de nós e trajetórias de custo de produção. Essa síntese permitiria simular como intervenções de política pública alterariam posições nas redes e influenciariam taxas de adoção tecnológica, criando uma ferramenta para formulação de políticas baseadas em evidências. Os modelos resultantes poderiam identificar caminhos ótimos de inovação que conciliem viabilidade técnica com posicionamento estratégico na rede.

O sistema de apoio à decisão baseado em LLM desenvolvido neste trabalho também apresenta inúmeras oportunidades de aprimoramento. Versões futuras poderiam incorporar fluxos em tempo real de patentes e publicações para identificação proativa de oportunidades, utilizar IA generativa para simulações de cenários políticos, e oferecer interfaces interativas mais avançadas para mapeamento participativo de redes junto a *stakeholders* industriais. Essas melhorias transformariam a ferramenta de uma plataforma analítica estática para um ecossistema dinâmico de apoio à decisão.

Por fim, essa agenda de pesquisa deve abordar lacunas críticas na interface entre sustentabilidade e economia dentro das redes de inovação. Áreas prioritárias incluem: desenvolvimento de métricas em rede para quantificar efeitos indiretos de mudança de uso da terra, análise da integração da economia circular nas cadeias de valor do SAF e investigação sobre como posições de centralidade afetam desfechos em justiça ambiental. Tais esforços contribuiriam para uma teoria sistêmica mais completa sobre inovação sustentável, que contemple dimensões tecnológicas e sociais de forma integrada.

Coletivamente, essas direções futuras prometem avançar tanto o conhecimento acadêmico quanto a implementação prática de soluções sustentáveis para a aviação. As oportunidades de maior impacto imediato parecem residir na integração de redes *multilayer* e na modelagem técnico-econômica com métricas de rede, pois abordam diretamente limitações identificadas no presente trabalho e oferecem subsídios acionáveis para formuladores de políticas e atores industriais.

Referências

- I. Abrantes, A. F. Ferreira, A. Silva, and M. Costa. Sustainable aviation fuels and imminent technologies - co2 emissions evolution towards 2050. *Journal of Cleaner Production*, 313:127937, 2021. ISSN 0959-6526. doi: <https://doi.org/10.1016/j.jclepro.2021.127937>. URL <https://www.sciencedirect.com/science/article/pii/S0959652621021557>.
- J. Akpan and O. Olanrewaju. Sustainable energy development: History and recent advances. *Energies*, 16(20), 2023. ISSN 1996-1073. doi: 10.3390/en16207049. URL <https://www.mdpi.com/1996-1073/16/20/7049>.
- E. Ang, P. Bhattacharya, and A. Lim. Estimating policy effects in a social network with independent set sampling, 2024.
- N. A. A.Qasem, A. Mourad, A. Abderrahmane, Z. Said, O. Younis, K. Guedri, and L. Kolsi. A recent review of aviation fuels and sustainable aviation fuels. *Journal of Thermal Analysis and Calorimetry*, 2024. doi: <https://doi.org/10.1007/s10973-024-13027-5>.
- R. E. Bailis and J. E. Baka. Greenhouse gas emissions and land use change from *jatropha curcas*-based jet fuel in brazil. *ENVIRONMENTAL SCIENCE & TECHNOLOGY*, 44(22):8684–8691, NOV 15 2010. ISSN 0013-936X. doi: 10.1021/es1019178.
- S. Björkqvist and J. Kallio. Building a graph-based patent search engine. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3300–3304, 2023.
- E. Cabrera and J. M. M. de Sousa. Use of sustainable fuels in aviation—a review. *Energies*, 15(7), 2022. ISSN 1996-1073. doi: 10.3390/en15072440. URL <https://www.mdpi.com/1996-1073/15/7/2440>.
- A. A. Caldeira, M. D. Wilbert, T. B. S. Moreira, and A. L. M. Serrano. Brazilian state debt sustainability: An analysis of net debt and primary balance; [sustainability of brazilian state debt: An analysis of the relationship between net debt and primary result]; [sustainability of the debt of the brazilian states: An analysis of the net debt and the primary balance]. *Public Administration Magazine*, 50(2):285–306, 2016. ISSN 0034-7612. doi: 10.1590/0034-7612151140. URL <https://doi.org/10.1590/0034-7612151140>.
- P. Campos, E. Pinto, and A. Torres. Rating and perceived helpfulness in a bipartite network of online product reviews. *Electronic Commerce Research*, 2023. ISSN 1572-9362. doi: 10.1007/s10660-023-09725-1. URL <https://doi.org/10.1007/s10660-023-09725-1>.
- W. Chebil, M. Wedyan, M. Alazab, R. Alturki, and O. Elshaweesh. Improving semantic information retrieval using multinomial naive bayes classifier and bayesian networks. *INFORMATION*, 14(5), MAY 3 2023. doi: 10.3390/info14050272.

- C.-H. Chou, S. L. Ngo, and P. P. Tran. Renewable energy integration for sustainable economic growth: Insights and challenges via bibliometric analysis. *Sustainability*, 15(20), 2023. ISSN 2071-1050. doi: 10.3390/su152015030. URL <https://www.mdpi.com/2071-1050/15/20/15030>.
- M. N. de Oliveira, L. R. Mosquéra, P. H. d. S. Martins, A. L. M. Serrano, G. D. Bispo, G. F. Vergara, G. M. Saiki, C. Neumann, and V. P. Gonçalves. Tracking biofuel innovation: A graph-based analysis of sustainable aviation fuel patents. *Energies*, 17(15), 2024. ISSN 1996-1073. doi: 10.3390/en17153683. URL <https://www.mdpi.com/1996-1073/17/15/3683>.
- M. N. de Oliveira, L. Mosquéra, and A. L. Marques Serrano. Ipc to research area models and cross-walk table. *Zenodo*, 2025. doi: 10.5281/zenodo.14806070. URL <https://doi.org/10.5281/zenodo.14806070>.
- W. Ding, J. Wang, and H. Zhu. Using graph representations for semantic information extraction from chinese patents. In *Proceedings of the 2019 3rd International Symposium on Computer Science and Intelligent Control*, pages 1–5, 2019.
- H. N. Gabow. Data structures for weighted matching and extensions to b -matching and f -factors, 2016.
- E. Giamphy, J.-L. Guillaume, A. Doucet, and K. Sanchis. A survey on bipartite graphs embedding. *Social Network Analysis and Mining*, 13(1):54, Mar 2023. ISSN 1869-5469. doi: 10.1007/s13278-023-01058-z. URL <https://doi.org/10.1007/s13278-023-01058-z>.
- J. González-Cobos, M. S. Prévot, and P. Vernoux. Electrolysis of lignin for production of chemicals and hydrogen. *Current Opinion in Electrochemistry*, 39:101255, 2023. ISSN 2451-9103. doi: <https://doi.org/10.1016/j.coelec.2023.101255>. URL <https://www.sciencedirect.com/science/article/pii/S2451910323000480>.
- N. Gray, S. McDonagh, R. O’Shea, B. Smyth, and J. D. Murphy. Decarbonising ships, planes and trucks: An analysis of suitable low-carbon fuels for the maritime, aviation and haulage sectors. *Advances in Applied Energy*, 1:100008, 2021. ISSN 2666-7924. doi: <https://doi.org/10.1016/j.adapen.2021.100008>. URL <https://www.sciencedirect.com/science/article/pii/S2666792421000019>.
- I. Guven and F. Simsir. Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ann) and support vector machines (svm) methods. *COMPUTERS & INDUSTRIAL ENGINEERING*, 147, SEP 2020. ISSN 0360-8352. doi: 10.1016/j.cie.2020.106678.
- M. A. Henning and J. H. van Vuuren. Eulerian graphs. In *Graph and Network Theory: An Applied Approach using Mathematica®*, pages 293–322. Springer, 2022.
- R. Interdonato, M. Magnani, D. Perna, A. Tagarelli, and D. Vega. Multilayer network simplification: approaches, models and methods, 2020.
- P. Jha and S. Schmidt. State of biofuel development in sub-saharan africa: How far sustainable? *Renewable and Sustainable Energy Reviews*, 150:111432, 2021. ISSN 1364-0321. doi: <https://doi.org/10.1016/j.rser.2021.111432>. URL <https://www.sciencedirect.com/science/article/pii/S1364032121007152>.

- T. Kandaramath Hari, Z. Yaakob, and N. N. Binitha. Aviation biofuel from renewable resources: Routes, opportunities and challenges. *Renewable and Sustainable Energy Reviews*, 42:1234–1244, 2015. ISSN 1364-0321. doi: <https://doi.org/10.1016/j.rser.2014.10.095>. URL <https://www.sciencedirect.com/science/article/pii/S1364032114009204>.
- E. Korkut and L. B. Fowler. Regulatory and policy analysis of production, development and use of sustainable aviation fuels in the united states. *Frontiers in energy research*, 9:750514, 2021.
- M. Liu, Z. Tu, X. Xu, Z. Wang, and Y. Wang. A data-driven approach for constructing multilayer network-based service ecosystem models. *Softw. Syst. Model.*, 22(3):919–939, 2022. ISSN 1619-1366. doi: [10.1007/s10270-022-01029-6](https://doi.org/10.1007/s10270-022-01029-6). URL <https://doi.org/10.1007/s10270-022-01029-6>.
- P. Mäki-Arvela, A. Aho, I. Simakova, and D. Yu. Murzin. Sustainable aviation fuel from syngas through higher alcohols. *ChemCatChem*, 14(23):e202201005, 2022.
- M. Maskittou, A. El Haddadi, and H. Routaib. Feature learning of patent networks using tensor decomposition. In *Innovations in Smart Cities Applications Volume 4: The Proceedings of the 5th International Conference on Smart City Applications*, pages 1375–1390. Springer, 2021.
- A. Maulana and M. Atzmueller. Many-objective optimization for anomaly detection on multi-layer complex interaction networks. *Applied Sciences*, 11(9), 2021. ISSN 2076-3417. doi: [10.3390/app11094005](https://doi.org/10.3390/app11094005). URL <https://www.mdpi.com/2076-3417/11/9/4005>.
- L. Metcalf and W. Casey. Chapter 5 - graph theory. In L. Metcalf and W. Casey, editors, *Cybersecurity and Applied Mathematics*, pages 67–94. Syngress, Boston, 2016. ISBN 978-0-12-804452-0. doi: <https://doi.org/10.1016/B978-0-12-804452-0.00005-1>. URL <https://www.sciencedirect.com/science/article/pii/B9780128044520000051>.
- J. A. Molina, A. Ferrer, D. Iñiguez, A. Rivero, G. Ruiz, and A. Tarancón. Network analysis to measure academic performance in economics. *Empirical Economics*, 58(3):995–1018, March 2020. doi: [10.1007/s00181-018-1546-0](https://doi.org/10.1007/s00181-018-1546-0).
- L. R. Mosquéra, M. N. de Oliveira, P. H. d. S. Martins, G. D. Bispo, R. V. Borges, A. L. M. Serrano, F. M. Pompermayer, C. Neumann, V. P. Gonçalves, and C. A. S. Bork. Biofuel dynamics in brazil: Ethanol–gasoline price threshold analysis for consumer preference. *Energies*, 17(21), 2024. ISSN 1996-1073. doi: [10.3390/en17215265](https://doi.org/10.3390/en17215265). URL <https://www.mdpi.com/1996-1073/17/21/5265>.
- P. Nakkiran, G. Kaplun, D. Kalimeris, T. Yang, B. L. Edelman, F. Zhang, and B. Barak. Sgd on neural networks learns functions of increasing complexity, 2019. URL <https://arxiv.org/abs/1905.11604>.
- K. S. Ng, D. Farooq, and A. Yang. Global biorenewable development strategies for sustainable aviation fuel production. *Renewable and Sustainable Energy Reviews*, 150:111502, 2021. ISSN 1364-0321. doi: <https://doi.org/10.1016/j.rser.2021.111502>. URL <https://www.sciencedirect.com/science/article/pii/S1364032121007814>.
- G. D. Nicola, C. Fritz, M. Mehrl, and G. Kauermann. Dependence matters: Statistical models to identify the drivers of tie formation in economic networks, 2023.

- P. Ray. Renewable energy and sustainability. *Clean Technologies and Environmental Policy*, 21:1517–1533, 2019.
- A. S. Soofi and M. Abdi. The social network analysis of techno-economic systems: Comparing results based on binary and weighted networks. *Journal of Systems Science and Systems Engineering*, 29: 68–84, 2020.
- E. Stergiou, N. Rigas, E. Zaroutieri, and K. Kounetas. Energy, renewable and technical efficiency convergence: a global evidence. *Economic Change and Restructuring*, 56(3):1601–1628, 2023.
- P. Tsankov. Overview of network-based methods for analyzing financial markets. *Proc. Tech. Univ. Sofia*, 71:1–7, 2021.
- United State’s Department of Energy. *Alternative Fuels Data Center- Sustainable Aviation Fuel*. <https://afdc.energy.gov/fuels/sustainable-aviation-fuel>, 2023. Accessed: 2024-05-16.
- D. R. Vardon, B. J. Sherbacow, K. Guan, J. S. Heyne, and Z. Abdullah. Realizing “net-zero-carbon” sustainable aviation fuel. *Joule*, 6(1):16–21, 2022.
- F. Wang, J. D. Harindintwali, Z. Yuan, M. Wang, F. Wang, S. Li, Z. Yin, L. Huang, Y. Fu, L. Li, S. X. Chang, L. Zhang, J. Rinklebe, Z. Yuan, Q. Zhu, L. Xiang, D. C. Tsang, L. Xu, X. Jiang, J. Liu, N. Wei, M. Kästner, Y. Zou, Y. S. Ok, J. Shen, D. Peng, W. Zhang, D. Barceló, Y. Zhou, Z. Bai, B. Li, B. Zhang, K. Wei, H. Cao, Z. Tan, L. bin Zhao, X. He, J. Zheng, N. Bolan, X. Liu, C. Huang, S. Dietmann, M. Luo, N. Sun, J. Gong, Y. Gong, F. Brahusi, T. Zhang, C. Xiao, X. Li, W. Chen, N. Jiao, J. Lehmann, Y.-G. Zhu, H. Jin, A. Schäffer, J. M. Tiedje, and J. M. Chen. Technologies and perspectives for achieving carbon neutrality. *The Innovation*, 2(4):100180, 2021. ISSN 2666-6758. doi: <https://doi.org/10.1016/j.xinn.2021.100180>. URL <https://www.sciencedirect.com/science/article/pii/S2666675821001053>.
- Y. Wang, R. Richard, and D. McDonald. Competitive analysis with graph embedding on patent networks. In *2020 IEEE 22nd Conference on Business Informatics (CBI)*, volume 1, pages 10–19. IEEE, 2020.
- World Intellectual Property Organization. *Patent cooperation treaty (PCT)*. <https://www.wipo.int/export/sites/www/pct/en/docs/texts/ai.pdf>, 2022. Accessed: 2024-05-12.
- L. Yu, B. Liu, Q. Lin, X. Zhao, and C. Che. Semantic Similarity Matching for Patent Documents Using Ensemble BERT-related Model and Novel Text Processing Method. *arXiv e-prints*, art. arXiv:2401.06782, 2024. doi: 10.48550/arXiv.2401.06782.
- T. Yusaf, L. Fernandes, A. R. Abu Talib, Y. S. M. Altarazi, W. Alrefae, K. Kadrigama, D. Ramasamy, A. Jayasuriya, G. Brown, R. Mamat, H. A. Dhahad, F. Benedict, and M. Laimon. Sustainable aviation—hydrogen is the future. *Sustainability*, 14(1), 2022. ISSN 2071-1050. doi: 10.3390/su14010548. URL <https://www.mdpi.com/2071-1050/14/1/548>.
- M. Zhu, Q. Ge, and X. Zhu. Catalytic reduction of co 2 to co via reverse water gas shift reaction: Recent advances in the design of active and selective supported metal catalysts. *Transactions of Tianjin University*, 26:172–187, 2020.

H. Ângelo, A. N. De Almeida, and A. L. M. Serrano. Determinants of brazil's demand of pulpwood; [determinantes da demanda de madeira em toras para celulose no brasil]. *Scientia Forestalis/Forest Sciences*, 2009. URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-73349114301&partnerID=40&md5=8211462eab43ef60e0ad6f051bdb5640>.