



UNIVERSIDADE FEDERAL FLUMINENSE
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA E DE TELECOMUNICAÇÕES

NICOLLAS RODRIGUES DE OLIVEIRA

Análise e Mitigação de Câmaras de Eco em Plataformas de Redes Sociais: Abordagens Baseadas em Processamento de Linguagem Natural e Meta-heurística de Otimização

NITERÓI

2024

Ficha catalográfica automática - SDC/BEE
Gerada com informações fornecidas pelo autor

D278a De Oliveira, Nicollas Rodrigues
Análise e Mitigação de Câmaras de Eco em Plataformas de
Redes Sociais: Abordagens Baseadas em Processamento de
Linguagem Natural e Meta-heurística de Otimização /
Nicollas Rodrigues De Oliveira. - 2024.
122 f.: il.

Orientador: Diogo Menezes Ferrazani Mattos.
Coorientador: Dianne Scherly Varela De Medeiros.
Tese (doutorado)-Universidade Federal Fluminense, Escola de
Engenharia, Niterói, 2024.

1. Câmaras de Eco. 2. Análise de Redes Sociais. 3. Modelos
de Difusão de Informação. 4. Redes Complexas. 5. Produção
intelectual. I. Mattos, Diogo Menezes Ferrazani, orientador.
II. De Medeiros, Dianne Scherly Varela, coorientadora. III.
Universidade Federal Fluminense. Escola de Engenharia. IV.
Título.

CDD - XXX

NICOLLAS RODRIGUES DE OLIVEIRA

Análise e Mitigação de Câmaras de Eco em Plataformas de Redes Sociais: Abordagens Baseadas em Processamento de Linguagem Natural e Meta-heurística de Otimização

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Telecomunicações da Universidade Federal Fluminense como requisito parcial para a obtenção do título de Doutor em Engenharia Elétrica e de Telecomunicações. Área de concentração: Comunicação de Dados Multimídia.

Aprovada em 3 de Julho de 2024.

BANCA EXAMINADORA

Prof. Diogo M. F. Mattos, D.Sc. – Orientador, UFF

Prof^a. Dianne S. V. de Medeiros, D.Sc. – Coorientadora, UFF

Prof^a. Débora C. M. Saade, D.Sc. – UFF

Prof. Tadeu N. Ferreira, D.Sc. – UFF

Prof. Luís Henrique M. K. Costa, Dr. – UFRJ

Prof. Pedro Henrique C. Caminha, D.Sc. – UFRJ

Niterói

2024

Para minha avó Lenoir (in memoriam)

Agradecimentos

Esta tese é especialmente dedicada à minha avó Lenoir, cujos estímulo e cuidado foram inestimáveis na minha formação e perduraram até sua partida. A convivência com uma ex-professora como ela permitiu-me compreender e, minimamente, absorver o valor da dedicação e do senso de responsabilidade, aspectos marcantes da sua personalidade.

Gostaria de agradecer ao meu companheiro Eric pela cumplicidade e por ser uma base motivacional e tranquilizadora, essenciais para me manter presente e atento ao óbvio, inclusive em momentos de baixa sinapse cognitiva. Agradeço igualmente ao núcleo mais próximo da minha família — irmão, mãe e padrasto — pelo apoio incondicional e acolhedor às minhas decisões pessoais e profissionais.

Agradeço imensamente ao meu orientador e à minha coorientadora pelo suporte teórico e por atuarem complementarmente como uma balança durante toda minha trajetória acadêmica. Em especial, agradeço ao Prof. Diogo M. F. Mattos pela orientação objetiva e perspicaz, sempre sugerindo novas ideias e fomentando a busca contínua por resultados. Em contrapartida, agradeço à Prof^a. Dianne S. V. de Medeiros pela coorientação precisa e atenciosa, sempre destacando pontos negligenciados e estimulando a dupla verificação de códigos e textos.

Minha gratidão estende-se a todos os professores, colaboradores, técnicos e alunos do Laboratório Mídiacom pelo conhecimento compartilhado. Especialmente, agradeço a Ana Carolina Mendes, Guilherme Nasseh, Govinda Bezerra, Hélio Cunha, Lúcio Reis e Yago Rezende pela amizade construída.

Agradeço aos professores e professoras membros da banca examinadora — Débora C. M. Saade, Luís Henrique M. K. Costa, Pedro Henrique C. Caminha e Tadeu N. Ferreira — pelas contribuições relevantes e pela disponibilidade em aceitar o convite. Agradeço também à professora Natalia C. Fernandes pela disponibilidade em participar do exame de qualificação.

Por fim, agradeço a todos os órgãos de fomento — CNPq, CAPES, RNP e FAPERJ — que financiaram direta ou indiretamente a pesquisa desenvolvida nesta tese.

Resumo

O fenômeno social das câmaras de eco (*echo chambers*) relaciona-se à tendência dos usuários de redes sociais de interagirem entre si em grupos homogêneos com ideias e opiniões semelhantes. Os membros de uma câmara de eco tendem a ignorar tentativas de refutar as ideias unilaterais constantemente difundidas em seu interior. Como resultado, as câmaras de eco prejudicam o contraditório e incentivam o viés de confirmação, fomentando ambientes propícios ao discurso de ódio e à propagação de notícias falsas. Nesse sentido, esta tese apresenta três propostas complementares para lidar com câmaras de eco em redes sociais *online*. A primeira proposta consiste em uma metodologia de monitoramento, captura e caracterização de conteúdos compartilhados no *Twitter* (atual *X*), capaz de identificar câmaras de eco em torno de temas controversos. A caracterização foca em *hashtags* e utiliza múltiplas métricas de redes complexas, avaliadas sob diferentes perspectivas estruturais, semânticas e tipográficas. Diferente de trabalhos que usam dados históricos, a proposta é aplicada a um caso de uso real no *Twitter*, analisando métricas como assortatividade, centralidade, diâmetro e densidade. Os resultados mostram tendências de escrita e padrões de relacionamento que sugerem oportunismo midiático na utilização de *hashtags*. A segunda proposta, o Método de Concordância de Rótulos (MCR), quantifica a similaridade entre os resultados de algoritmos que identificam agrupamentos homogêneos, utilizando a decomposição espectral de uma matriz de pertencimento. Ao avaliar a similaridade entre os algoritmos de detecção de comunidades e os algoritmos de aprendizado não supervisionado na identificação de câmaras de eco, o método revelou uma quase total discordância. A terceira proposta consiste em uma Estratégia de Contenção Otimizada da propagação de informação em câmaras de eco baseada em *Simulated Annealing* (ECO-SA). Essa abordagem identifica nós críticos em grafos que, quando bloqueados, reduzem significativamente a difusão de informações. Simulações em 15 exemplares de câmaras de eco, aplicadas a modelos de difusão de informação sob diferentes cenários, demonstram a superioridade da estratégia em comparação com outros métodos. Destaca-se que o bloqueio dos nós estratégicos identificados pela ECO-SA limitou a disseminação a aproximadamente 45% a 60% do total de nós, enquanto o segundo melhor método permitiu a disseminação a cerca de 65% a 75%, respectivamente.

Palavras-chave: Câmaras de Eco, Análise de Redes Sociais, Modelos de Difusão de Informação, Redes Complexas.

Abstract

The social phenomenon of echo chambers is related to the tendency of social media users to interact within homogeneous groups that share similar ideas and opinions. Members of an echo chamber tend to ignore attempts to refute the unilateral ideas constantly disseminated within it. As a result, echo chambers hinder counterarguments and encourage confirmation bias, fostering environments conducive to hate speech and the spread of fake news. This thesis presents three complementary proposals to address echo chambers in online social networks. The first proposal consists of a methodology for monitoring, capturing, and characterizing shared content on Twitter, capable of identifying echo chambers around controversial topics. The characterization focuses on hashtags and uses multiple complex network metrics, evaluated from different structural, semantic, and typographic perspectives. Unlike works that use historical data, the proposal is applied to a real use case on Twitter, analyzing metrics such as assortativity, centrality, diameter, and density. The results show writing trends and relationship patterns that suggest media opportunism in the use of hashtags. The second proposal, the Label Concordance Method (LCM), quantifies the similarity between the results of algorithms that identify homogeneous clusters, using the spectral decomposition of a membership matrix. By evaluating the similarity between community detection algorithms and unsupervised learning algorithms in identifying echo chambers, the method revealed an almost total disagreement. The third proposal consists of an Optimized Containment Strategy for the propagation of information in echo chambers based on Simulated Annealing (ECO-SA). This approach identifies critical nodes in graphs which, when blocked, significantly reduce information diffusion. Simulations on 15 echo chamber samples, applied to information diffusion models under different scenarios, demonstrate the strategy's superiority compared to other methods. It is noteworthy that blocking the strategic nodes identified by ECO-SA limited the dissemination to approximately 45% to 60% of the total nodes, while the second-best method allowed dissemination to about 65% to 75%, respectively.

Keywords: Echo Chambers, Social Network Analysis, Information Diffusion Models, Complex Networks.

Lista de Figuras

2.1	Representação de grafos temporais baseada em séries de instantâneos. . . .	14
2.2	Visão topológica de uma câmara de eco.	15
2.3	Exemplos da estrutura em grafo intitulada k -clique.	17
2.4	Exemplos de grafos e suas representações matriciais de adjacência.	25
2.5	Variação da modularidade de um grafo conexo.	26
2.6	Critérios de ligação e dendrograma do algoritmo hierárquico.	31
2.7	Valores da silhueta para conjunto de dados com cinco comunidades.	35
4.1	Metodologia de monitoramento e caracterização de redes sociais.	44
5.1	Caracterização estrutural e tipográfica do grafo de <i>hashtags</i>	55
5.2	Evolução temporal da distribuição do grau e das <i>hashtags</i> mais centrais. . .	56
5.3	Discretização temporal dos subgrafos mais populosos.	57
5.4	Caracterização semântica do subgrafo mais populoso.	59
5.5	Concordância entre algoritmos no tópico de controle de armas.	61
5.6	Concordância entre algoritmos no tópico de discriminação de minorias. . .	62
5.7	Concordância entre algoritmos no tópico de política.	63
5.8	Modelo <i>Kertesz Threshold</i> aplicado a múltiplos <i>snapshots</i> simultaneamente	64
5.9	Disseminação em câmaras de eco pelo modelo <i>Profile</i>	67
5.10	Variação da estabilidade da propagação em diferentes cenários.	68
5.11	Trivialidade da solução nodal encontrada pelo ECO-SA.	69
5.12	Comparação das estratégias de contenção da propagação.	71
A.1	Visão geral do processo de contratação e gestão de serviços de rede.	89
A.2	Arquitetura Open RAN, conforme especificada pela O-RAN Alliance. . . .	92

A.3	Diagrama de blocos da arquitetura do sistema AGIR.	103
A.4	Variação das centralidades entre as classes gramaticais no grafo de intenções.	107
A.5	Variação da assortatividade entre as classes gramaticais no grafo de intenções.	108
A.6	Desempenho de modelos de redes neurais na detecção de conflitos.	109

Lista de Tabelas

2.1	Definições dos termos e conceitos associados às câmaras de eco.	10
2.2	Trabalhos relacionados ao estudo de câmaras de eco.	11
4.1	Estatísticas da captura de dados relacionados ao episódio final do <i>BBB21</i>	48
4.2	Características, condutância e controvérsia das câmaras de eco analisadas.	53
A.1	Comparação das características dos trabalhos relacionados a IBM.	101

Sumário

1	Introdução	1
1.1	Motivação	3
1.2	Objetivo	5
1.3	Contribuição	5
1.4	Estrutura do Texto	6
2	Fundamentação Teórica	7
2.1	Câmaras de Eco: Definição e Diferenciação	7
2.2	Abordagens de Detecção	10
2.2.1	Estratégia Tradicional de Monitoramento	12
2.3	Análise de Redes Sociais	12
2.3.1	Métricas de Redes Complexas em Comunidades	16
2.3.2	Métricas de Controvérsia em Comunidades	20
2.4	Modelagem da Difusão de Informação	22
2.5	Representação Vetorial de Grafos	24
2.6	Delimitação de Câmaras de Eco em Redes Sociais	24
2.6.1	Algoritmos de Detecção de Comunidade	25
2.6.1.1	<i>Louvain</i>	26
2.6.1.2	<i>Fast Greedy</i>	27
2.6.1.3	<i>WalkTrap</i>	27
2.6.1.4	<i>InfoMap</i>	28
2.6.2	Algoritmos Não Supervisionados	29

2.6.2.1	Algoritmos Baseados no Particionamento	29
2.6.2.2	Algoritmos Hierárquicos	30
2.7	Métricas de Avaliação Extrínsecas e Intrínsecas	31
2.7.1	Métricas de Deslocamento Semântico	34
3	Trabalhos Relacionados	36
3.1	Investigação de Câmaras de Eco	36
3.2	Análise de <i>Hashtags</i>	39
3.3	Contenção da Propagação de Informação	40
4	Propostas de Monitoramento, Análise e Contenção de Câmaras de Eco	43
4.1	Metodologia de Monitoramento de Redes Sociais	43
4.1.1	Etapa de Captura de Dados	44
4.1.2	Etapa de Estruturação	45
4.1.3	Etapa de Atualização	46
4.1.4	Etapa de Extração	46
4.1.5	Etapa de Visualização	47
4.2	Estudo de Caso	47
4.3	Método de Concordância de Rótulos	48
4.4	Estratégia de Bloqueio Otimizada	49
4.5	Base de Dados de Câmaras de Eco	52
5	Resultados e Discussão	54
5.1	Resultados do Monitoramento e Caracterização	54
5.1.1	Análise Temporal e Tipográfica	54
5.1.2	Análise Semântica	58
5.2	Resultados do Método de Concordância de Rótulos	59
5.3	Resultados da Análise da Propagação	64

5.3.1	Avaliação Intra-metodológica	65
5.3.2	Avaliação Inter-metodológica	70
6	Conclusão	73
	Referências	77
	Apêndice A – Sistema de Gerenciamento Baseado em Intenção	87
A.1	Gerenciamento Baseado em Intenção para O-RAN	87
A.1.1	Arquitetura O-RAN	91
A.2	Ciclo de Gerenciamento Baseado em Intenção	94
A.2.1	Ingestão e Expressão de Intenções	95
A.2.2	Tradução de Intenções	96
A.2.3	Resolução de Políticas	97
A.2.4	Orquestração, Implantação e Garantia de Políticas	98
A.3	Trabalhos Relacionados à IBM	99
A.4	Sistema AGIR	102
A.5	Resultados e Discussão	105
A.6	Conclusão	108

Capítulo 1

Introdução

A facilidade de acesso pervasivo e ubíquo para a publicação e consumo de informações torna as plataformas de redes sociais um importante meio de interação pública. Estudos recentes revelam que 71% dos brasileiros recorreram a plataformas como o *Facebook* e o X (ex-*Twitter*) para se informarem¹. Apesar de hospedarem tanto conteúdos úteis para geração de conhecimento, quanto para o entretenimento, é latente a vocação das redes sociais para fomentar e repercutir conteúdos focados no discurso de ódio e em notícias falsas (*fake news*) [1]. Essa tendência de amplificar fenômenos indesejáveis é consequência direta da alteração do modelo de produção e consumo de informação.

Tradicionalmente o processo de seleção de informação era mediado por jornalistas ou editores. Atualmente, com as redes sociais, essa mediação passou a ser exercida por todos e quaisquer usuários, com ou sem formação adequada para divulgação de notícias. Essa supressão – ou ausência da mediação profissional especializada – contribui para degradação dos princípios de imparcialidade e legitimidade da informação, uma vez que usuários não treinados são mais suscetíveis ao fenômeno conhecido como “viés de confirmação”. Tal fenômeno expressa a tendência dos usuários de absorverem e compartilharem informações que se aderem ao seu sistema de crenças, mesmo que essas informações sejam inverídicas. Igualmente recorrente nas redes sociais, a “câmara de eco” é um fenômeno social relacionado à tendência dos usuários de interagirem e ingressarem em grupos homogêneos com ideias semelhantes às suas. Embutido no cerne desse fenômeno está o conceito de *unfriending*, que consiste na exclusão intencional de usuários com opiniões contrárias às adotadas na câmara de eco [2, 3, 4].

Implicitamente o fenômeno da câmara de eco engloba mecanismos que alienam os

¹Disponível em <https://www.kaspersky.com.br/blog/pesquisa-infodemia-impactos-vida-digital/17467/>.

membros da comunidade formada, impedindo o acesso a fontes de informação epistêmicas externas. Assim, as câmaras de eco funcionam sistematicamente para formar padrões ideológicos disfuncionais que impedem os integrantes da câmara de eco de se envolverem em buscas informativas além da sua comunidade intelectual. Além de impedir o acesso à informação, os mecanismos envolvidos na formação das câmaras de eco desacreditam ativamente informações externas à câmara. Como consequência, há um desequilíbrio epistêmico que poderia implicar apenas omissão de pontos de vista contrários, mas que, devido à formação da câmara de eco, implica uma desconfiança preventiva ideológica dando origem ao aprisionamento dos indivíduos em câmaras de eco [5].

Algumas redes sociais, como *Reddit*² e *Gab*³, facilitam o processo de formação de câmaras de eco, uma vez que são naturalmente organizadas em comunidades relacionadas a um determinado tema. Recentemente, o *Whatsapp* divulgou o desenvolvimento da funcionalidade *Whatsapp Comunidades* capaz de agregar vários grupos em um mesmo espaço compartilhado⁴. Embora almejem facilitar a comunicação, tais recursos e plataformas reacendem preocupações sobre os potenciais impactos nocivos da divulgação de conteúdos em ambientes virtuais segregados, ajudando a intensificar o viés de confirmação. Outro facilitador normalmente utilizado pelas plataformas de redes sociais é o algoritmo de recomendação, que fornece aos usuários mais do mesmo conteúdo consumido com base em seus comportamentos passados para moldar a preferência futura. Os usuários tendem a aceitar as recomendações e, adicionalmente, buscam ativamente mais informações sobre o conteúdo de interesse, devido ao viés de confirmação. Os algoritmos de recomendação também podem usar a popularidade de um conteúdo como indicativo de qualidade ou de preferência pessoal por um determinado conteúdo. No entanto, essa métrica de engajamento pode vir às custas da diversidade de opiniões [6]. Assim, o viés de confirmação e os algoritmos de recomendação criam uma espiral de auto-reforço. Como consequência, o ciclo de retroalimentação entre algoritmos de recomendação e aspectos que operam na psicologia humana eventualmente leva a uma câmara de eco que muda a visão de mundo dos usuários [7].

Um exemplo prático e preocupante das consequências da formação de câmaras de eco é o agravamento da polarização política nos últimos anos no mundo. Devido ao mecanismo de auto-reforço, as câmaras de eco ideológicas formadas alimentam ainda mais a polarização ao ampliar as lacunas de conhecimento entre grupos díspares. Recentemente,

²Disponível em <https://www.reddit.com/>.

³Disponível em <https://gab.com/>.

⁴Disponível em <https://blog.whatsapp.com/sharing-our-vision-for-communities-on-whatsapp>.

Fletcher *et al.* [8] avaliaram o número de pessoas em câmaras de eco de notícias politicamente partidárias, abrangendo diferentes países europeus e os Estados Unidos. O estudo focou na contabilização de pessoas que usam apenas fontes de notícias *online* tendenciosas, sejam de esquerda ou de direita. Os resultados obtidos mostraram que, no Reino Unido, cerca de 2% do público pesquisado integram uma câmara de eco inclinada para a esquerda, enquanto cerca de 5% participam de uma câmara de eco inclinada para a direita. Tais proporções são semelhantes aos demais países analisados, com exceção dos EUA, onde estima-se que mais de 10% dos entrevistados confiam apenas em fontes de notícias partidárias.

1.1 Motivação

A interação entre usuários dentro das câmaras de eco é marcada pela disseminação desenfreada de conteúdos enviesados e pela blindagem a qualquer tentativa de refutação por opiniões externas. Além disso, devido à natural baixa capacidade humana de distinguir notícias verdadeiras daquelas de cunho falso [9], as câmaras de eco propiciam também a criação e divulgação de notícias falsas. Todos esses fatores contribuem fortemente para a geração de atritos em diversos campos da sociedade. Ao analisar o âmbito político, percebe-se que a combinação desses fatores promove uma degradação do espírito democrático, estimulando a instabilidade institucional e o radicalismo de opiniões. No âmbito social, o foco dos membros de uma câmara de eco é nomeadamente a desacreditação de instituições de referência e a destruição de reputações. Essa prática é frequentemente direcionada a profissionais como cientistas, jornalistas e médicos com posicionamentos contrários às crenças difundidas dentro da câmara de eco.

A relação de simbiose existente entre redes sociais e eventos ou situações cotidianas permite que qualquer indivíduo externalize suas opiniões, crenças e posicionamentos ideológicos através de postagens em seus perfis. A ausência da mediação de profissionais relacionados à divulgação de notícias aliada à liberdade e facilidade de divulgação de informações nas redes sociais por quaisquer usuários reforça a formação das câmaras de eco [3]. Dessa forma, é fundamental (i) desenvolver soluções capazes de detectar e caracterizar câmaras de eco em torno das repercussões de eventos externos às redes sociais. Arelado ao desafio da caracterização está (ii) a dificuldade de fazê-la dinamicamente, acompanhando a velocidade natural de geração e o volume de dados relacionados a eventos reais que potencializam a polaridade de opiniões. Uma maneira de satisfazer esse imediatismo, sem recorrer a técnicas possivelmente mais demoradas e complexas de pro-

cessamento de linguagem natural e aprendizado de máquina, é através do processamento das palavras-chave indexadoras das mensagens, como as *hashtags* utilizadas nos *tweets*.

A menor complexidade ao processar *hashtags* ao invés de processar todo o texto de uma postagem é acompanhada por três desafios principais. O primeiro é a polissemia, ou seja, uma mesma *hashtag* pode se referir a eventos distintos em diferentes janelas temporais. Como um segundo desafio ao lidar com *hashtags*, a sinonímia está relacionada ao fato de diferentes *hashtags* poderem possuir o mesmo significado. Tal característica está diretamente relacionada à arbitrariedade de sua criação, uma vez que não há um consenso entre os usuários ou orientação das plataformas sobre a padronização da escrita das *hashtags*. O terceiro desafio é a obscuridade que remete à dificuldade de interpretação enfrentada tanto por humanos quanto por algoritmos [10]. Sendo frequentemente compostas por acrônimos, palavras concatenadas, neologismos, abreviações ou combinações das opções anteriores, as *hashtags* podem demorar a serem completamente compreendidas dependendo do conhecimento prévio do leitor sobre o contexto relacionado. Para mitigar esses desafios, a proposta desta tese de doutorado emprega uma estrutura em grafo a fim de prover um entendimento mais baseado no contexto ou tema relacionado às *hashtags*. Uma vantagem adicional da estruturação de *hashtags* em grafos é a associação a ontologias para a representação do conhecimento relacionando termos, palavras, expressões ou axiomas pertencentes a um mesmo domínio de interesse [11]. Dessa forma, garante-se o monitoramento de discussões *online*, sem comprometer a capacidade das *hashtags* de oferecer metadados sobre o texto associado de forma compactada.

Além dos desafios de monitoramento, detecção e caracterização de câmaras de eco, um terceiro desafio emergente envolve a (iii) mitigação de sua expansão e redução de seus efeitos tanto dentro quanto fora da redoma de usuários. Recentemente, o tratamento dessa questão advém de abordagens passivas que visam principalmente o caráter preventivo, buscando a redução da influência de conteúdo tendencioso em usuários por meio da exposição a uma diversidade de perspectivas informacionais. Iniciativas científicas nesse contexto variam desde abordagens de treinamento baseadas na gamificação, como o ChamberBreaker [12], até ferramentas de detecção de enviesamento de perfis, como o Check-my-echo [13], e sistemas de análise e consumo de informações multi-ideológicas, como o StarryThoughts e o HearHere [14, 15]. No entanto, a falta de uma metodologia ativa de atuação para remover usuários membros de câmaras de eco torna essa questão um desafio em aberto tanto no âmbito psicológico quanto computacional. Motivada por

ações judiciais⁵ ou tomada voluntariamente pelos gestores de redes sociais⁶, a solução de combate mais incisiva consiste no bloqueio de contas ou perfis de usuários específicos. Essa prática de regulação visa dismantelar o ciclo de disseminação em estruturas extremamente polarizadas, como as câmaras de eco. Na prática, a seleção de tais contas é normalmente fundamentada na análise do histórico de postagens de perfis em busca dos chamados disseminadores primários, ou seja, aqueles usuários responsáveis pela criação ou compartilhamento inicial em massa de conteúdo falso na rede.

1.2 Objetivo

Este trabalho tem como objetivo geral o desenvolvimento de propostas complementares capazes de monitorar, analisar e mitigar o surgimento de potenciais câmaras de eco em plataformas de redes sociais. A fim de desenvolver tais propostas, os seguintes objetivos específicos são traçados: (i) monitorar e coletar ininterruptamente o fluxo de *hashtags* utilizadas no debate de um tema polarizado no *Twitter*; (ii) estruturar e caracterizar diferentes variações de grafos que representem câmaras de eco ou temas polarizados, empregando métricas de redes complexas e métricas de controvérsia; (iii) identificar padrões recorrentes de compartilhamento e escrita de *hashtags* em redes sociais; (iv) analisar temporalmente o deslocamento semântico entre *hashtags* atreladas ao mesmo tema polarizado; (v) mensurar quantitativamente o grau de concordância entre abordagens algorítmicas capazes de detectar estruturas homogêneas semelhantes a câmaras de eco; (vi) simular a dinâmica de propagação da informação dentro de exemplos de câmaras de eco reais segundo diferentes modelos de difusão de informação; e (vii) identificar e elaborar estratégias de neutralização de pontos críticos na estrutura das câmaras de eco, capazes de promover uma redução no nível de propagação de informações entre seus integrantes.

1.3 Contribuição

Esta tese de doutorado apresenta três contribuições principais relacionadas às câmaras de eco. A primeira contribuição consiste no desenvolvimento de uma metodologia de monitoramento, captura e caracterização de conteúdos compartilhados no *Twitter*, capaz de fornecer indícios sobre o surgimento de câmaras de eco em torno de temas controversos debatidos na plataforma. A segunda contribuição reside em uma abordagem comparativa,

⁵Disponível em <https://www.nbcnews.com/news/world/brazil-supreme-court-justice-orders-elon-musk-investigation-fake-news-rcna146799>.

⁶Disponível em <https://www.reuters.com/article/idUSKBN2BE17T/>.

denominada Método de Concordância de Rótulos, capaz de quantificar o grau de similaridade entre os resultados de algoritmos que identificam estruturas ou agrupamentos homogêneos. A terceira contribuição, intitulada ECO-SA, apresenta-se como uma estratégia otimizada de contenção da propagação de informação em câmaras de eco, baseada na meta-heurística *Simulated Annealing*. Vale ressaltar a originalidade do tópico abordado nesta tese, sendo um dos primeiros trabalhos nacionais dedicados ao fenômeno das câmaras de eco no âmbito virtual através de uma abordagem computacional.

Como resultados desta tese de doutorado, destacam-se duas publicações científicas em congressos nacionais e diretamente relacionadas ao monitoramento de redes sociais e fenômeno das câmaras de eco, sendo elas:

- Artigo intitulado “*Caracterização Sócio-Temporal de Conteúdos em Redes Sociais baseada em Processamento em Fluxo*” e publicado no XXVI Workshop de Gerência e Operação de Redes e Serviços (WGRS 2021);
- Minicurso intitulado “*Identificação de Câmaras de Eco em Redes Sociais Através de Detecção de Comunidade em Redes Complexas: Ferramentas, Tendências e Desafios*” e publicado no XXVIII Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2022).

1.4 Estrutura do Texto

Esta tese de doutorado está organizada da seguinte forma. O Capítulo 2 apresenta uma breve fundamentação teórica, introduzindo o fenômeno das câmaras de eco e revisitando conceitos da teoria de grafos e análise de redes sociais. Este capítulo também aborda algoritmos de detecção de comunidades e suas métricas de avaliação. O Capítulo 3 oferece uma visão abrangente sobre a literatura relacionada às câmaras de eco e termos correlatos, bem como estudos dedicados à análise de *hashtags* e à mitigação da propagação de informação. O Capítulo 4 detalha um conjunto de propostas complementares para a caracterização, análise e mitigação dos efeitos das câmaras de eco em redes sociais. O Capítulo 5 expõe e discute os resultados obtidos em cada proposta. O Capítulo 6 conclui a tese de doutorado e aponta perspectivas para trabalhos futuros. O Apêndice A apresenta um sistema de gerenciamento de rede baseado em intenção, desenvolvido em paralelo com as propostas principais relacionadas às câmaras de eco durante o doutorado.

Capítulo 2

Fundamentação Teórica

Este capítulo apresenta o embasamento teórico que irá fundamentar o conjunto de propostas desta tese de doutorado. Dessa forma, revisitam-se as principais definições bem como os principais procedimentos e técnicas usados no processo de detecção das câmaras de eco em redes sociais. Também são abordados modelos de difusão de informação, algoritmos de detecção de comunidade, algoritmos não supervisionados, métricas de caracterização e, por fim, formas de avaliação da detecção com ou sem dados de referência.

2.1 Câmaras de Eco: Definição e Diferenciação

As câmaras de eco não possuem uma definição única e consensual na literatura. Apesar da multiplicidade de definições [16, 17, 18], em um sentido mais amplo, as **câmaras de eco** são definidas como

“Rede altamente ativa de usuários, na qual ideias semelhantes são amplificadas ou reforçadas por meio de um processo de compartilhamento repetitivo de uma mesma ideia, blindando-a de refutação.”

Diante dessa definição, caracteriza-se uma câmara de eco em relação a três aspectos: (i) a composição estrutural, (ii) o conteúdo divulgado e (iii) o comportamento de seus integrantes perante refutação. Nas redes sociais *online*, as câmaras de eco organizam-se na forma de uma rede de comunicação altamente conectada, composta por diferentes tipos de usuários, porém com posicionamentos ideológicos semelhantes. Dentro das câmaras de eco, há uma prevalência pelo compartilhamento de conteúdos extremamente unilaterais e muito aderentes às posturas e opiniões dos usuários que a integram. Essa ressonância de informação no interior de câmaras de eco é igualmente acompanhada por uma tática ativa

de descredibilização e exclusão de quaisquer fontes ou opiniões externas ou divergentes. Na prática, quando expostos a fatos que desafiam suas crenças, os membros de câmaras de eco rapidamente os rejeitam já que não fortalecem seus posicionamentos originais. Tal atitude reativa a contra-evidências é descrita pelo efeito *backfire* [7], que resulta em membros acreditarem em informação ilegítima, mesmo após a apresentação de evidências de inverdades.

A definição de câmara de eco confunde-se comumente com a de outro fenômeno relacionado, a **bolha de filtro** (*filter bubble*). Embora a bolha de filtro também restrinja o acesso à informação de seus integrantes, o isolamento observado nas bolhas de filtro é derivado do uso de um conjunto extremamente restrito de recomendações. Normalmente, essas sugestões são automaticamente realizadas pelos chamados sistemas de recomendação, empregados em redes sociais *online*. Esses sistemas de recomendação consistem em algoritmos que inferem os interesses ou preferências do próprio usuário por meio da análise de dados subjacentes ao uso da rede, tais como comportamento de cliques, histórico de compras ou registro de pesquisas [19]. Como a superabundância de informações seria esmagadora para qualquer usuário, os sistemas de recomendação têm o papel fundamental de personalizar o fluxo de informações, selecionando e exibindo somente o conteúdo mais relevante para cada usuário. As redes sociais exigem uma quantidade adequada de diversidade para permitir a soberania informacional de seus usuários. No entanto, a imposição de filtros personalizados nas informações pode exacerbar a fragmentação, criando ciclos de realimentação degenerados, nos quais a amplitude da informação é cada vez mais reduzida ao longo do tempo [5].

Comparativamente, a bolha de filtro apresenta uma estrutura relativamente mais frágil que a câmara de eco, uma vez que são permissivas a fontes externas relevantes. Ao contrário da descredibilização adotada nas câmaras de eco, as fontes externas à bolha de filtro são normalmente ignoradas por motivos involuntários do usuário. Diferentemente das câmaras de eco, a exposição dos usuários em bolhas de filtro a informações ou argumentos relevantes, não experienciados antes, permite libertá-los de suas bolhas [20]. Embora menos potencialmente danosas, a superpersonalização intencional de uma bolha de filtro pode eventualmente confinar usuários em câmaras de eco. Outro termo relacionado, o *gatekeeping*, traduz-se como uma prática de filtragem de informação tradicionalmente executada por fontes midiáticas. Quando essa prática é executada por usuários, esses passam a ser intitulados como *gatekeepers*. Os *gatekeepers* consomem conteúdo de múltiplas visões ideológicas, porém reproduzem ou disseminam conteúdos relacionados a uma única visão [21].

A organização dos usuários em grupos homogêneos está intimamente atrelada ao conceito de **homofilia**, ou seja, a tendência dos indivíduos de se associarem a outros semelhantes. Essa semelhança ou similaridade social pode ser classificada em: i) homofilia de *status*, relacionada ao agrupamento de pessoas com base em suas características semelhantes atribuídas (sexo, raça ou etnia) ou adquiridas (educação ou religião); e ii) homofilia de valor, que envolve agrupar pessoas semelhantes com base em seus valores, atitudes. Dependendo da ideologia atrelada, a câmara de eco pode ser formada devido à homofilia de *status*, homofilia de valor ou ambos. As plataformas de redes sociais afrouxaram as fronteiras básicas de homofilia, como a geografia, permitindo que os usuários vinculem relacionamentos homófilos em outras dimensões, como raça, etnia, sexo, gênero e religião [22, 7].

A estrutura da câmara de eco também pode contribuir para a disseminação de **notícias falsas**. Usuários dentro de uma câmara de eco são repetidamente expostos a informações homogêneas, fato que favorece a aceitação acrítica de informações inverídicas que confirmam suas visões pessoais. A seleção de informações consistentes com as crenças e a prevenção de informações contrárias ou contrastantes às crenças reforçam a confiança nas opiniões minoritárias, como notícias fabricadas, mesmo na presença de evidência contrária preponderante. A estrutura da câmara de eco também pode induzir a uma convergência rápida e prematura para soluções sub-ótimas de problemas complexos. Ademais, o limiar para perceber um conteúdo como novo pode ser menor dentro das câmaras de eco em virtude da reduzida diversidade de pontos de vista aos quais as pessoas estão expostas. As tentativas de captura de atenção são desempenhadas pelo uso de conteúdo inflamatório, polêmico, ou emocional [23].

Além das notícias falsas, as câmaras de eco criam uma estrutura que potencializa a crença de que alguma organização secreta, mas influente, se reúne em acordo secreto a fim de alcançar um objetivo malévolo. Na prática, as **teorias conspiratórias** são tentativas de explicar as causas finais de eventos e circunstâncias sociais e políticas significativas. Seus adeptos usam as redes sociais para se encontrarem, disseminarem conteúdos conspiratórios e compartilharem pontos de vista marginais. As teorias da conspiração expressam e amplificam ansiedades e medos de perder o controle da ordem religiosa, política ou social. Na literatura científica, diversos trabalhos debruçam-se sobre o estudo das câmaras de eco considerando a disseminação de teorias conspiratórias em seu interior [24, 25]. Com o objetivo de explicar e evitar o intercambiamento equivocado, a Tabela 2.1 condensa as definições dos principais termos e conceitos associados às câmaras de eco.

Tabela 2.1: Definições dos termos e conceitos associados às câmaras de eco.

Termos	Definição
Câmaras de Eco	Estruturas sociais que excluem sistematicamente as fontes de informação não necessariamente por omissão, mas por ação deliberada.
Bolha de Filtro	Ambiente social superpersonalizado derivado da filtragem excessiva de informação, praticada automaticamente por sistemas de recomendação.
Sistemas de Recomendação	Algoritmos que inferem os interesses ou preferências do próprio usuário por meio da análise de dados subjacentes ao uso da rede, exibindo em um conteúdo personalizado.
<i>Gatekeeping</i>	Prática de filtragem de informação tradicionalmente executada por fontes de mídia, onde certos conteúdos são selecionados, modificados ou omitidos antes de serem transmitidos ao público.
Homofilia	Conceito psicológico relacionado à tendência dos indivíduos de se associarem a outros com características ou valores semelhantes.
Viés de Confirmação	Conceito psicológico relacionado à tendência dos indivíduos procurarem e privilegiarem o recebimento de informações que reforcem seus pontos de vista existentes.
Notícias Falsas	Notícias comprovadamente inverídicas, criadas e disseminadas intencionalmente para confundir ou desacreditar algo ou alguém.
Teoria Conspiratória	Crença na existência de que alguma organização secreta, mas influente, se reúne em acordo secreto a fim de alcançar um objetivo maldoso.

2.2 Abordagens de Detecção

Atualmente, as abordagens de detecção de câmaras de eco distinguem-se em duas grandes classes de abordagens, a baseada em ideologia e a baseada em topologia de rede. A **abordagem baseada na ideologia** infere a existência de um ambiente virtual polarizado ao analisar semanticamente a inclinação do conteúdo compartilhado ou consumido por um usuário [7]. Internamente, essa análise semântica visa medir a similaridade, ou distância, entre representações vetoriais do texto compartilhado, seja o texto derivado de comentários, postagens, notícias ou mensagens. Tais representações vetoriais podem ser obtidas por meio de modelos simples como Frequência do Termo – Inverso da Frequência nos Documentos (*Term Frequency–Inverse Document Frequency*, TF-IDF), ou por meio de algoritmos de incorporação de palavras (*word embeddings*), que empregam redes neurais treinadas com grandes volumes de dados textuais. Algoritmos como *GloVe* e *Word2Vec* são capazes de embutir a semântica das palavras em vetores de baixa dimensão, densos e de tamanho fixo, garantindo que sinônimos ou palavras minimamente relacionadas sejam mapeados em vetores similares [26]. Essa representação vetorial individualizada das palavras é especialmente empregada na detecção de notícias falsas, um conteúdo frequentemente divulgado em câmaras de eco.

A detecção de câmaras de eco segundo a **abordagem baseada em topologia de rede** explora os padrões de propagação de conteúdo entre os usuários. Normalmente, a

Tabela 2.2: Trabalhos relacionados ao estudo de câmaras de eco.

	Abordagem		Tipo de Câmara de Eco		
	Topológica	Ideológica	Microescala	Mesoescala	Macroescala
Cota et al. [2]	✓				✓
Baumann et al. [27]	✓				✓
Bessi et al. [25]		✓			✓
Morini et al. [16]	✓	✓		✓	
Morales et al. [28]	✓	✓			✓
Villa et al. [17]	✓	✓			✓
Williams et al. [29]	✓	✓			✓
Garimella et al. [21]	✓	✓			✓
Barbera et al. [30]	✓	✓			✓
Zollo et al. [4]		✓			✓
Bakshy et al. [31]		✓	✓		
Conover et al. [32]	✓	✓			✓
Gillani et al. [33]	✓				✓
Sasahara et al. [6]	✓				✓
Cinelli et al. [34]	✓	✓			✓
Zannettou et al. [35]		✓			✓

abordagem topológica inicia com a extração de metadados de usuários participantes de um debate controverso nas redes sociais. Em posse dos metadados, é possível inferir os posicionamentos de cada usuário a respeito do tópico selecionado, bem como construir estruturas em grafos que representem as interações digitais entre eles. Dentre as vantagens dessa estruturação em redes de usuários estão: i) a possibilidade de revelar padrões estruturais característicos de câmaras de eco por meio de métricas de redes complexas; e ii) a identificação de grupos de usuários homogêneos a partir da aplicação de algoritmos de detecção de comunidades [7].

Morini *et al.* [16] introduzem a visão de classificação das câmaras de eco de acordo com a escala da detecção. Nessa visão, as chamadas câmaras de eco de **microescala** são delimitadas avaliando o comportamento *online* de usuários individualmente. Alternativamente, as câmaras de eco de **macroescala** são delimitadas considerando a rede de interação dos usuários em um nível mais amplo, ou seja, analisando o grafo como um todo. Contudo, este tipo de câmara de eco pode ignorar eventuais diferenças dentro de determinadas áreas da rede. Uma terceira classificação expressa que as câmaras de eco de **mesoescala** podem ser identificadas como apenas um subconjunto de nós na rede geral. Isso implica que, na rede geral de debates, é possível encontrar diversas câmaras de eco com a mesma inclinação ideológica. A Tabela 2.2 apresenta vários trabalhos relacionados ao estudo de câmaras de eco, distinguindo-os segundo a abordagem adotada e o tipo da câmara de eco detectada.

2.2.1 Estratégia Tradicional de Monitoramento

O monitoramento dos conteúdos e entidades digitais potencialmente relacionados às câmaras de eco inicia-se com a seleção de um tópico controverso. Independente do domínio de conhecimento, a discussão sobre questões duais propiciam a polarização de opiniões e podem transformar atitudes divergentes em extremos ideológicos. Questões controversas são igualmente debatidas tanto no âmbito *offline* quanto no *online*. Contudo as plataformas de redes sociais *online* são, provavelmente, o espaço aberto mais utilizado para discussões. Essas plataformas possuem estrutura e funcionalidades que facilitam a identificação de debates *online* sobre uma ampla gama de questões diferentes. Plataformas como o *Twitter*, *Instagram* e *Facebook* permitem que os usuários incluam em suas postagens as *hashtags*, que são palavras, ou frases sem espaçamento, prefixadas com o caractere cardinal “#”. Embora tenham sido originalmente idealizadas para indexar as postagens dos usuários, as *hashtags* atualmente comportam-se como um termômetro de eventos sociopolíticos, comerciais e culturais. Visto que as *hashtags*, assim como qualquer outra manifestação textual, expressam os interesses, opiniões e crenças dos seus autores, elas podem ser usadas como ponto de partida para encontrar grupos de usuários com posicionamentos semelhantes sobre um determinado assunto [3]. De forma diferente, plataformas como *Reddit* e *Gab* são organizadas como fóruns virtuais que restringem o debate por temas de interesse, tirando o foco das *hashtags*. No *Reddit* tais comunidades de interesse são chamadas *subreddits*. Para coletar os dados oriundos dessas redes de forma automática, algumas plataformas disponibilizam APIs (*Application Programming Interface*), como o *Twitter*¹. Além de utilizar a API própria da plataforma, é possível utilizar códigos terceirizados desenvolvidos no formato de APIs, como no caso do *Reddit* e do *Gab*².

2.3 Análise de Redes Sociais

A Análise de Redes Sociais (*Social Network Analysis* - SNA) compreende a investigação de relações ou fenômenos sociais por meio do uso de grafos, que são estruturas matemáticas compostas por vértices e arestas. Um grafo $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ possui $|\mathcal{V}|$ vértices (ou nós) interligados por $|\mathcal{E}|$ arestas. Os grafos são orientados quando a relação entre dois nós é unidirecional e, portanto, para que $v_i \in \mathcal{V}$ esteja conectado a $v_j \in \mathcal{V}$ e v_j esteja conectado a v_i devem existir duas arestas $\{\varepsilon_{i,j}, \varepsilon_{j,i}\} \in \mathcal{E}$. Se o grafo é não orientado, a relação

¹Disponível em <https://developer.twitter.com/en/docs/twitter-api>.

²Disponível em <https://github.com/pushshift/api>.

entre nós é bidirecional, e apenas uma aresta $\varepsilon_{i,j} \in \mathcal{E}$ precisa existir entre v_i e v_j para que estejam conectados entre si. As arestas, sejam bidirecionais ou unidirecionais, podem possuir pesos não unitários $w_{i,j} \in \mathcal{W}$, transformando os grafos em grafos ponderados $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$. Os pesos representam a intensidade da ligação entre os pares de nós adjacentes.

As redes sociais *online* são caracterizadas matematicamente como redes complexas, *i.e.*, grafos detentores de um grande número de nós e interligados por meio de uma topologia complexa, que tende a apresentar uma distribuição de grau de cauda longa, podendo esses grafos serem orientados ou não, dependendo do tipo de relação entre nós. Em uma rede social como *Twitter* ou *Instagram*, os grafos que representam a conexão entre pessoas são orientados. A pessoa v_i pode seguir a pessoa v_j sem que a pessoa v_j siga a pessoa v_i , existindo, assim, apenas a aresta $\varepsilon_{i,j}$. Isso implica que v_i recebe o conteúdo de v_j , mas o inverso não é verdade. Em uma rede social como *Facebook*, para que duas pessoas estejam conectadas, elas precisam ser amigas uma da outra, de forma que há uma relação bidirecional entre todos os nós que fazem parte da rede. Assim, o grafo da rede para essa representação de conexão entre pessoas é não orientado e todas as arestas representam uma relação bidirecional em que v_i recebe conteúdo de v_j e vice-versa.

O problema abstrato de identificar câmaras de eco em redes sociais é tratado na abordagem topológica como um problema de identificar comunidades em um grafo. Na teoria de redes complexas, define-se uma comunidade como um subconjunto de nós densamente conectados entre si e esparsamente conectado aos demais nós [36, 37]. Em uma rede estática, a topologia não se altera ao longo do tempo, facilitando a aplicação direta de algoritmos clássicos de detecção de comunidades. Porém, essa estruturação estática não é capaz de retratar a natureza evolutiva de sistemas reais, como a criação de novos vínculos com outras pessoas e a extinção de antigos vínculos, o que representaria, por exemplo, novas amizades e o término de antigas amizades. Assim, devido à dinamicidade das redes sociais, a topologia da rede varia ao longo do tempo e a rede estática não é a melhor representação. Nesse sentido, é válido utilizar representação de redes temporais para modelar a rede social em observação. As redes temporais têm uma topologia variável no tempo, em que nós e arestas podem estar ativos ou não, dependendo do instante de observação. Existem três modelos principais para representação de redes temporais:

- **Sequência de Contatos**, redes cujas arestas existem por um tempo desprezível e são representadas como um conjunto de sequência de contato (i, j, t) em que i e j são nós, e t é o carimbo de tempo do contato;

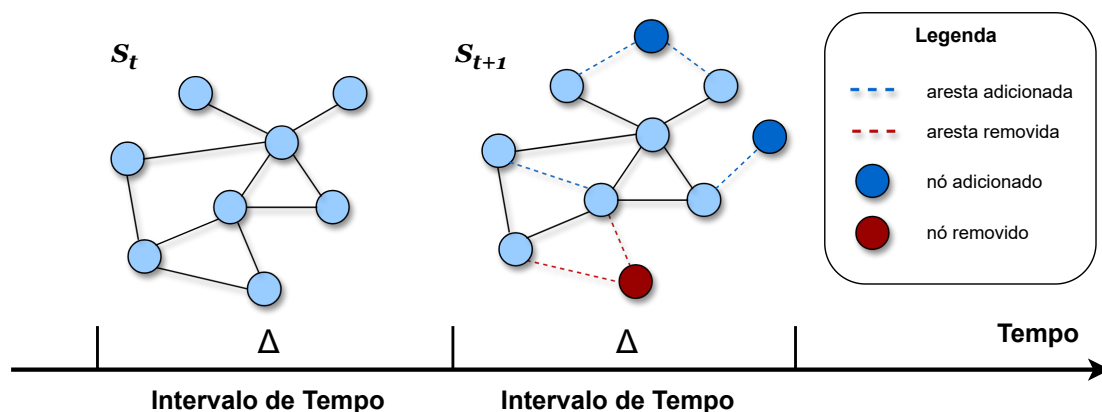


Figura 2.1: Representação de grafos temporais baseada em séries de instantâneos. A evolução de um grafo dinâmico em tempo discreto $\mathcal{G}$ é modelada como uma série de instantâneos discretos em diferentes intervalos de tempo (Δ), i.e., $\mathcal{G} = \{S_1, S_2, \dots, S_T\}$, que contém T instantâneos estáticos. A modelagem de uma câmara de eco segundo esta representação discreta permite que modificações na sua estrutura, como inclusão ou remoção de nós e arestas, sejam analisadas entre *snapshots* consecutivos.

- **Grafo de Intervalo**, redes que registram a duração dos períodos de atividade da uma aresta e através um conjunto de intervalos $T_e = (t_1, t'_1) \dots (t_n, t'_n)$;
- **Séries de Instantâneos**, rede segmentada em um conjunto de janelas de tempo sequenciais (*snapshots*) em que contatos em uma janela de tempo são agregados em arestas. A evolução da rede temporal pode ser estudada através desses *snapshots*, conforme mostrado na Figura 2.1.

Embora as redes temporais de **sequência de contato** e **grafo de intervalo** preservem as informações temporais, a análise desses modelos é mais complexa e depende da aplicação de novas metodologias e algoritmos. A rede temporal representada por uma **série de instantâneos** reduz a complexidade da análise, permitindo o uso de métodos de pesquisa de redes estáticas de forma independente em cada *snapshot*. Contudo, a principal limitação do modelo de séries de instantâneos é a indeterminação quanto ao tamanho ótimo do intervalo de tempo usado para analisar o grafo. Na prática, a escolha do tamanho adequado depende de uma profunda compreensão da rede em questão, sendo vital para evitar a particularização da análise e a obtenção de resultados dissonantes. *Snapshots* muito curtos podem não conter arestas suficientes agregadas em cada instantâneo, resultando em informações incompletas. Por outro lado, o uso de *snapshots* muito longos pode mascarar detalhes da evolução do grafo [38].

Mediante a coleta de dados, muitos autores costumam estruturar as câmaras de eco aproveitando-se dos dados nativos extraídos das plataformas de redes sociais. Com o

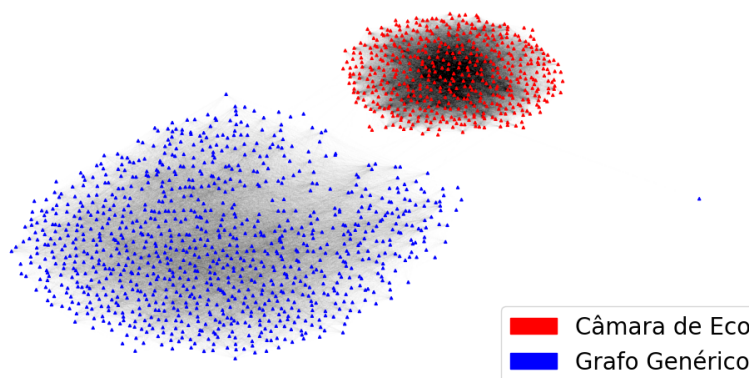


Figura 2.2: Visão topológica de uma câmara de eco. Estruturalmente as câmaras de eco são definidas como um conjunto de nós densamente conectados entre si e fracamente conectados com o resto da rede. A coloração escura no interior do grafo representando uma câmara de eco, ilustra a alta conectividade entre os seus nós integrantes.

Twitter, por exemplo, pode-se empregar metadados descorrelacionados como *retweets*, menções, seguidores ou *hashtags* para construir grafos relacionais. Esses grafos representam redes semânticas em que os nós correspondem entidades específicas e as arestas descrevem a relação entre elas, ou seja, a conexão entre as entidades [39]. Assim, as interações usuário-usuário, como seguir, repostar, bater papo e comentar entre si, podem ser utilizadas para desvendar a difusão de informações e a relação entre os usuários das redes sociais *online*. Uma rede de *retweets* pode ser reproduzida através de um grafo relacional direcionado ponderado em que os nós representam o conjunto de usuários distintos e o peso de uma aresta do nó v_i ao nó v_j representa o número de vezes que o usuário v_i repostou um *tweet* do usuário v_j . Essa mesma lógica pode ser adotada com plataformas que também disponibilizam o número de curtidas (*likes*) nas postagens, como o *Facebook*. Similarmente, uma rede de menção pode ser retratada por um grafo relacional direcionado ponderado na qual os nós representam os usuários e as arestas representam as menções, ou seja, a ação de incluir um nome de usuário em um *tweet*. O peso de uma aresta do nó v_i ao nó v_j representa o número de vezes que o usuário v_i menciona o usuário v_j . Finalmente, a rede de seguidores pode ser construída por meio de um grafo relacional direcional no qual os nós representam usuários e uma aresta do nó v_i ao nó v_j representa o usuário v_i seguindo o usuário v_j . A estruturação de uma câmara de eco como um subgrafo denso e fracamente conectado ao restante da rede é ilustrado na Figura 2.2.

2.3.1 Métricas de Redes Complexas em Comunidades

O mapeamento das redes sociais em grafos permite interpretar cada câmara de eco como uma comunidade, podendo assim revelar características estruturais latentes dessas câmaras a partir de métricas de redes complexas. Para o cálculo da maioria das métricas de redes complexas é necessário conhecer os conceitos de caminho, passeio, passeio aleatório e clique. Um **caminho** é definido como uma sequência ordenada de arestas que unem nós adjacentes, sem repetição dos nós. O tamanho desse caminho é definido pelo número de arestas atravessadas. O **caminho mais curto**, $\delta_{i,j}^*$, entre um par de nós é aquele que atravessa o menor número de arestas possível. De forma semelhante, um **passeio** também é uma sequência ordenada de arestas que unem nós adjacentes, mas sem restrição quanto à repetição de nós e de arestas na sequência ordenada. Já um **passeio aleatório** é definido como um processo aleatório que parte de um nó inicial com destino a um nó final. Em cada nó a partir do nó inicial, há uma determinada probabilidade de seguir para um dos nós adjacentes ao nó atual. No passeio aleatório, a sequência de transições entre nós de um grafo é modelada por uma Cadeia de Markov finita e temporalmente reversível. Ao alcançar o nó final, existe uma sequência ordenada de arestas que atravessa o grafo partindo do nó inicial até o nó final. A essa sequência gerada pelo processo aleatório dá-se o nome de passeio aleatório. Assim como no passeio tradicional, pode haver repetição de vértices e arestas. Por fim, um **clique** é um subconjunto de nós de um grafo não direcionado em que quaisquer dois nós distintos desse subconjunto são sempre adjacentes, isto é, o subconjunto de nós é totalmente conectado, formando uma malha completa. Quando contendo exatamente k nós, o subconjunto é denominado k -clique, conforme, conforme mostra os exemplos na Figura 2.3. É importante destacar e diferenciar os conceitos de clique maximal e clique máximo. No processo de aglutinação de cliques, eventualmente não é mais possível aumentar o tamanho de um determinado clique porque ele não é um subconjunto de um clique maior. Um clique máximo é o maior clique da rede, isto é, aquele que possui a maior quantidade de nós.

Métricas comumente utilizadas na análise de comunidades são a densidade, o diâmetro, a assortatividade, e as diversas variações de centralidades [3]. A **densidade** de um grafo, $D(\mathcal{G})$, indica quão denso um grafo é em termos de conectividade de arestas e é medida em função do número total de arestas $|\mathcal{E}|$ e nós $|\mathcal{V}|$ pertencentes ao grafo, conforme Equação 2.2. O **diâmetro** é expresso pela máxima excentricidade do grafo, ou seja, o maior dentre todos os caminhos mais curtos existentes entre todos os pares de nós v_i e v_j , cujo comprimento é dado por $\delta^*(v_i, v_j)$, em número de saltos. Matematicamente, a

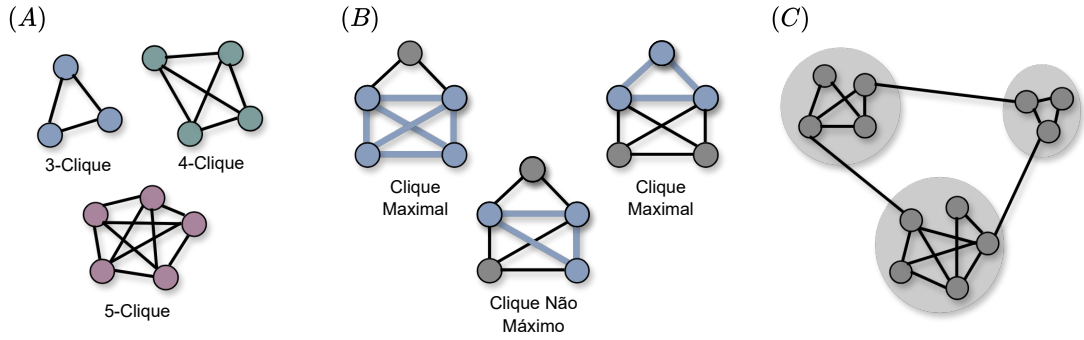


Figura 2.3: Em (A) ilustram-se vários tipos de k -clique. Em (B) mostram-se as diferenças entre clique máximo e maximal em um mesmo grafo. Na prática, o clique máximo inclui o maior número possível de nós do grafo. Já um clique maximal é clique que não pode ser aumentado pela inclusão de mais um nó, tendo alcançado o ápice do seu tamanho. Assim, todo clique máximo é maximal, mas o contrário não. Em (C) exemplifica-se o resultado do agrupamento de triângulos de nós (3-clique) na mesma comunidade quando estes têm $k - 1$ nós em comum.

excentricidade de um vértice v_i , $e(v_i)$, é dada pela Equação 2.3. Assim, o diâmetro, $d(\mathcal{G})$, de um grafo, dado pela Equação 2.2, pode assumir valores inteiros no intervalo de $[1, \infty[$.

$$D(\mathcal{G}) = \frac{2|\mathcal{E}|}{|\mathcal{V}| \cdot (|\mathcal{V}| - 1)} \quad (2.1) \quad d(\mathcal{G}) = \max_{v_i \in \mathcal{V}} e(v_i) \quad (2.2)$$

$$e(v_i) = \max_{v_j \in \mathcal{V}} \delta^*(v_i, v_j) \quad (2.3)$$

Quando transportado para o âmbito de redes complexas, o conceito de homofilia é representado pela métrica de **assortatividade**. Essa métrica, definida entre $[-1, 1]$, expressa a tendência de nós se conectarem a outros nós com valores semelhantes de uma determinada característica. Ao considerar o grau como característica a ser avaliada, por exemplo, valores positivos de assortatividade indicam uma correlação entre nós de graus semelhantes, enquanto valores negativos indicam relações entre nós de graus diferentes. Valores nulos traduzem a completa conexão entre todos os nós em um grafo. Casos extremos, positivos ou negativos, mostram que o grafo exibe padrões de mistura entre ordenamentos perfeitos ou padrões não ordenados, respectivamente [3, 40]. Adicionalmente, as diversas variações de centralidade visam quantificar a importância de cada nó em um grafo. Essa importância assume diferentes sentidos dependendo do tipo de centralidade. Alguns dos principais tipos de métricas clássicas para calcular a centralidade são a proximidade, o grau e a intermediação.

- **Centralidade de Grau** (*Degree*) calcula a importância de um nó a partir do número de arestas conectadas a ele. Assim, a centralidade de grau de um nó é igual à quantidade de arestas que ele possui. Em grafos orientados, o grau do nó é medido em termos de grau de entrada, que considera apenas o número de arestas incidentes no nó, ou grau de saída, que considera apenas o número de arestas partindo do nó. O grau relaciona-se com a popularidade do nó, ou quão bem conectado é o nó. Representa-se a centralidade de grau de um nó v_i por

$$C_{deg}(v_i) = deg(v_i). \quad (2.4)$$

- **Centralidade de Proximidade** (*Closeness*) relaciona-se à rapidez com que um nó alcança todos os outros nós da rede. O cálculo da centralidade de proximidade ($C_C(v_i)$) de cada nó v_i leva em consideração os caminhos mais curtos entre o v_i e todos os outros nós da rede. Matematicamente, a proximidade é dada por

$$C_C(v_i) = \frac{|\mathcal{V}| - 1}{\sum_{j \neq i} \delta^*(v_i, v_j)}, \quad (2.5)$$

em que $|\mathcal{V}|$ é o número total de nós e $\delta^*(v_i, v_j)$ é a distância mais curta, em número de saltos, entre o par de nós v_i e v_j . Naturalmente, a inexistência de uma conexão entre v_i e v_j implica uma distância $\delta^*(v_i, v_j) = \infty$. A utilização de caminhos mais curtos permite que a complexidade de tempo para computar a proximidade seja $O(|\mathcal{V}||\mathcal{E}|)$. Na prática, quando usada em um grafo de usuários, a centralidade de proximidade mede efetivamente o quão próximo, em média, cada usuário está de todos os outros usuários em uma comunidade, ou seja, dentro de um grupo de usuários semelhantes que estão densamente conectados.

- **Centralidade de Intermediação** (*Betweenness*) reflete a fração total de caminhos mais curtos que passam por um nó, usando-o como ponte. Dado um grafo $\mathcal{G}$, conexo ou não, a centralidade de intermediação (C_B) do nó v_i é definida como

$$C_B(v_i) = \sum_{v_s \neq v_t \neq v_i \in \mathcal{V}} \frac{\sigma_{st}(v_i)}{\sigma_{st}}, \quad (2.6)$$

em que $\sigma_{st}(v_i)$ representa o número de caminhos mais curtos do nó v_s para o nó v_t que passam pelo nó v_i e σ_{st} é o total de caminhos mais curtos do nó v_s para o nó v_t . Dessa forma, a relação representa a proporção de caminhos mais curtos entre v_s e v_t que passam por v_i . Implicitamente a centralidade de intermediação relaciona-se com o controle de fluxos ou de informação que são propagados pela rede. Quanto mais central é o nó, maior a proporção de caminhos mais curtos que passam por ele

e, conseqüentemente, maior é o controle exercido sobre os fluxos que trafegam na rede pelos caminhos mais curtos. Realizando buscas em largura, algoritmos podem computar a centralidade de intermediação com uma complexidade de tempo igual a $O(|\mathcal{V}| + |\mathcal{E}|)$.

- **Centralidade de Informação** (*Information*) é uma variante da centralidade de proximidade. A centralidade de informação também é chamada de centralidade de proximidade do fluxo de corrente. O valor da centralidade de informação de um nó é determinado com base no fluxo de informação presente em todos os possíveis caminhos entre pares de nós. Essa métrica considera todos os caminhos existentes entre os pares de nós no grafo. Computacionalmente, o cálculo da centralidade de informação é custoso para grandes redes, visto que sua complexidade de tempo possui uma relação cúbica com o número de nós $O(|\mathcal{V}|^3)$. Matematicamente é definida por

$$C_I(v_i) = \frac{|\mathcal{V}|}{\sum_{v_j \in \mathcal{V}} \frac{1}{I(v_i, v_j)}}, \quad (2.7)$$

em que $|\mathcal{V}|$ é o número de nós no grafo, $\delta^*(v_i, v_j)$ é o custo do caminho mais curto entre v_i e v_j , e $I(v_i, v_j) = \sum_n I_n(v_i, v_j)$ é o somatório da informação medida em todos os caminhos entre v_i e v_j . A informação em um único caminho n é dada por $I_n(v_i, v_j) = 1/\delta_n(v_i, v_j)$, em que $\delta_n(v_i, v_j)$ é o tamanho do caminho n .

- **Centralidade Harmônica** (*Harmonic*) também é uma variante da centralidade de proximidade que privilegia nós que estão mais próximos do nó analisado, atenuando o impacto de nós mais distantes. Para tanto, a centralidade harmônica $C_H(v_i)$ de um nó v_i é computada pela soma dos inversos das distâncias mínimas entre v_i e todos os outros nós v_j da rede, como mostra a Equação 2.8 abaixo. Tal abordagem garante que distâncias menores detenham um peso maior na soma total, enquanto distâncias maiores têm um peso reduzido.

$$C_H(v_i) = \frac{1}{\sum_{j \neq i} \delta^*(v_i, v_j)}, \quad (2.8)$$

- **PageRank** foi originalmente criada para ranquear a importância de uma página *web*, contabilizando a quantidade e a qualidade de *links* que referiam a página [41]. A quantidade é medida em função do grau de entrada, enquanto a qualidade é mensurada em relação à importância das páginas que contêm esses *links*. Por esta razão, essa métrica é frequentemente incorporada na análise de grafos, uma vez que pode ser utilizada para medir a popularidade de um nó na rede. De maneira

simplicada, pode-se calcular o valor de *PageRank* de um nó v_i como

$$C_{PageRank}(v_i) = \alpha \sum_{v_j \in \mathcal{V}} \frac{a_{j,i}}{C_{deg}(v_j)} C_{PageRank}(v_j) + \beta, \quad (2.9)$$

em que $a_{j,i}$ é um elemento da matriz de adjacências $\mathbf{A}$, α é um fator de amortecimento constante, β é um fator de personalização constante e $C_{deg}(v_j)$ é o grau de saída do nó v_j se esse grau for positivo, ou $C_{deg}(v) = 1$ se o grau de saída for nulo. A Equação 2.9 é definida de forma recursiva, visto que a análise do *PageRank* de determinado nó depende do *PageRank* de todos os nós vizinhos. Destaca-se que a vizinhança entre dois nós exerce maior relevância no cálculo do *PageRank* se o nó de origem possuir maior *PageRank*. Contudo, o *PageRank* de um nó é impactado negativamente proporcionalmente ao número de arestas partindo de si. O *PageRank* mais alto pode ser interpretado como proporcional à probabilidade de o nó espalhar uma informação em sua comunidade.

Visando capturar a qualidade de uma comunidade localmente, a métrica de **condutância** remete à porcentagem de arestas que atravessam os limites de uma comunidade. Comunidades bem definidas tendem a ter uma pequena condutância, significando a densa presença de arestas internas à comunidade, enquanto que arestas externas são esparsas. Assim, dado um grafo $\mathcal{G}$ não direcionado e sua matriz de adjacências $\mathbf{A}$, a condutância ϕ de uma comunidade $\mathcal{C} \subset \mathcal{V}$ pode ser calculada pela razão expressa por

$$\phi(\mathcal{C}) = \frac{\sum_{v_i \in \mathcal{C}, v_j \notin \mathcal{C}} a_{i,j}}{\min \{a(\mathcal{C}), a(\bar{\mathcal{C}})\}}, \quad (2.10)$$

em que o numerador representa o número de arestas intercomunitárias e o denominador representa o número total de arestas da comunidade. Os valores $a(\mathcal{C})$ são dados por $\sum_{v_i \in \mathcal{C}} \sum_{v_j \in \mathcal{V}} a_{i,j}$ e $a(\bar{\mathcal{C}})$ por $\sum_{v_i \notin \mathcal{C}} \sum_{v_j \in \mathcal{V}} a_{i,j}$. A condutância do grafo é dada pela condutância mínima de todas as possíveis comunidades $\mathcal{C}$, dada por

$$\phi(\mathcal{G}) = \min_{\mathcal{C} \subseteq \mathcal{V}} \phi(\mathcal{C}). \quad (2.11)$$

2.3.2 Métricas de Controvérsia em Comunidades

A interpretação de comunidades como câmaras de eco tem como vantagem permitir a aplicação das métricas usadas para análise e detecção de comunidades já existentes. Além disso, essa interpretação possibilita obter indícios da presença das câmaras de eco por meio de duas propriedades: a controvérsia e a homogeneidade. Ambas as propriedades

quantificam, de forma complementar, a polarização de um tópico debatido entre usuários de uma rede social. Na prática, a **propriedade de controvérsia** pode ser computada utilizando diversas métricas, dentre elas a *Random Walk Controversy* (RWC) e a *Authoritative Random Walk Controversy* (ARWC).

A métrica *Random Walk Controversy* (RWC), originalmente proposta por Garimella *et al.* [42], considera duas comunidades $\mathcal{X}$ e $\mathcal{Y}$ do grafo $\mathcal{G}$, tal que $\mathcal{X} \cup \mathcal{Y} = \mathcal{V}$, e $\mathcal{X} \cap \mathcal{Y} = \emptyset$. Além disso, são considerados dois passeios aleatórios, um terminando na comunidade $\mathcal{X}$ e o outro terminando na comunidade $\mathcal{Y}$. Definida no intervalo $[0, 1]$, a RWC expressa a diferença das probabilidades de dois eventos, um considerando que ambos os passeios aleatórios começam e terminam na mesma comunidade, e outro assumindo que ambos os passeios aleatórios começam em uma comunidade e terminam na outra. Matematicamente, essa diferença é computada por

$$RWC = P_{\mathcal{X}\mathcal{X}}P_{\mathcal{Y}\mathcal{Y}} - P_{\mathcal{X}\mathcal{Y}}P_{\mathcal{Y}\mathcal{X}}, \quad (2.12)$$

em que $P_{\mathcal{AB}}$ representa a probabilidade condicional de um passeio aleatório começar na comunidade $\mathcal{A}$ e terminar na $\mathcal{B}$, tal que $\mathcal{A}, \mathcal{B} \in \{\mathcal{X}, \mathcal{Y}\}$. Diante disso, espera-se que quanto mais próximo do valor mínimo, maior a probabilidade de migração para a outra comunidade e menor a controvérsia do grafo. Por outro lado, quanto mais próximo do valor máximo, maior a probabilidade de permanência na comunidade original e consequentemente há indícios da presença de controvérsia.

A métrica ARWC, proposta por Villa *et al.* [17], é derivada da RWC, mas distingue-se dela pelo fato de não adotar uma seleção completamente aleatória dos nós iniciais utilizados no passeio aleatório. A ARWC parte apenas dos intitulados nós autoritativos. Tal classificação é atribuída a um conjunto de nós específicos do grafo, com base na centralidade de grau desses nós. Essa alteração do ponto de partida permite capturar a probabilidade de um usuário casual, pertencente a uma determinada comunidade, ser exposto ao conteúdo disseminado por um nó autoritativo da comunidade oposta. No cálculo da ARWC, o passeio aleatório termina quando um nó pertencente ao conjunto de nós autoritativos de uma ou outra partição é alcançado.

Nesse contexto, quanto maiores os níveis de controvérsia entre nós de diferentes comunidades e de homogeneidade entre nós de uma mesma comunidade, maior a probabilidade da existência de câmaras de eco no grafo analisado [17].

2.4 Modelagem da Difusão de Informação

Devido à dificuldade de obtenção de bases de dados reais sobre câmaras de eco, alguns trabalhos empregam modelos estatísticos para simular o processo de disseminação de informações e determinar a influência de um conjunto inicial de nós propagadores. Embora detenham critérios de propagação particulares, os modelos podem ser agrupados em três categorias [43]: (i) Modelos Epidêmicos, aplicáveis em contextos de “contágio social”, onde as pessoas estão tomando decisões para adotar uma nova ideia ou inovação; (ii) Modelos de Dinâmicas de Opiniões, capazes de lidar com as oscilações dos valores de parâmetros, simulando a mutabilidade das opiniões de cada usuário; e (iii) Modelos de Redes Dinâmicas, adaptados para lidar com a evolução temporal de grafos baseados em instantâneos (*snapshots*) discretos.

Apesar da pluralidade de modelos de difusão existentes na literatura, apenas um subconjunto deles é efetivamente aplicável para investigar as câmaras de eco e a difusão de informações e opiniões dentro delas. Essa limitação decorre tanto do modelo possuir uma lógica de funcionamento incoerente com o fenômeno, quanto da obrigatoriedade do uso de parâmetros nativos incompatíveis. Dentre os modelos alinhados ao cenário de difusão observado em câmaras de eco, destacam-se os detalhados abaixo. Em cada modelo, os nós da rede podem assumir três estados possíveis, que representam diferentes níveis de adoção da informação: ativos, inativos (suscetíveis) ou bloqueados.

Kertesz Threshold (KT). Introduzido por Ruan *et al.* [44], este modelo representa um exemplo de modelo de redes dinâmicas capaz de simular a propagação considerando a sensibilidade à pressão dos pares. Essa pressão descreve a resposta dos usuários da rede à influência externa exercida pelos vizinhos, levando-os a adotar uma determinada informação. Neste modelo, a definição do conjunto de nós inicialmente ativos e inicialmente bloqueados pode ser configurada de forma declarativa, especificando o conjunto, ou de forma aleatória, escolhendo-se uma porcentagem. As transições entre esses estados são determinadas por parâmetros específicos do modelo, que incluem:

- **Limiar de Ativação (*Ths*):** parâmetro atribuído a cada nó que reflete a proporção mínima de seus vizinhos ativos necessária para que o nó seja ativado. Definido entre $[0, 1]$, onde 1 impõe a necessidade de adesão unânime dos vizinhos.
- **Fração de Adoção Espontânea (p_s):** parâmetro opcional que simula a probabilidade da ativação exógena, ou seja, uma ativação espontânea que independe da influência de seus vizinhos. Definido entre $[0, 1]$, onde o limite inferior indica uma

probabilidade nula da existência de adotantes espontâneos e o limite superior indica que todos os nós são adotantes espontâneos.

A dinâmica de funcionamento do modelo KT opera segundo cinco etapas: (1) Escolha do conjunto de nós ativos, ou seja, as fontes de propagação, e o conjunto de nós bloqueados; (2) Busca pelo subconjunto dos nós inativos da rodada, que são vizinhos a nós ativos. (3) Alteração do estado de cada nó desse subconjunto, caso a fração de seus vizinhos ativos exceda seu limiar de ativação; (4) Caso $p_s > 0$, adotantes espontâneos podem surgir a qualquer momento da propagação, adicionando novas ativações na rede de forma independente da estrutura local; e (5) Repetição das três etapas anteriores até alcançar o número máximo de iterações estabelecido.

Profile (PFL). Proposto por Milli *et al.* [45], este modelo é um exemplo de modelo epidêmico que considera a premissa de que cada usuário na rede, representado como um nó, detém uma resistência individual à adoção da informação. Essa resistência é específica para cada usuário e o único fator que influencia sua mudança de estado durante a propagação. Assim como no modelo KT, a definição dos nós ativos e bloqueados pode ser feita de maneira declarativa ou aleatória. Em termos de parâmetros, a única diferença em relação ao modelo anterior está na substituição do parâmetro *Ths*, presente no modelo *Kertesz Threshold*, pelo parâmetro descrito abaixo:

- **Perfil do Usuário (*Pfl*)**: parâmetro atribuído a cada nó que reflete sua resistência interna à adesão de informações ou conteúdos. Definido entre $[0, 1]$, onde 1 implica alta resistência interna e 0 indica baixa resistência interna à adesão de informações propagadas por terceiros.

O processo de difusão do modelo *Profile* pode ser sumarizado em cinco etapas: (1) Escolha do conjunto de nós ativos, ou seja, as fontes de propagação, e o conjunto de nós bloqueados; (2) Varredura e identificação do subconjunto de nós inativos contendo vizinhos ativos; (3) Ativação de cada nó do subconjunto caso seu perfil (*Pfl*) seja menor que uma probabilidade gerada aleatoriamente. Tentativas falhas de ativação resultam na alteração do estado inativo para bloqueado; (4) Caso $p_s > 0$, adotantes espontâneos podem surgir a qualquer momento da propagação, independente do perfil; e (5) Repetição das três etapas anteriores até alcançar o número máximo de iterações estabelecido.

2.5 Representação Vetorial de Grafos

A adoção da abordagem topológica para a investigação de câmaras de eco implica a necessidade de modelar a rede como um grafo. A transformação de um grafo em uma representação vetorial auxilia na avaliação quantitativa da qualidade da estrutura da comunidade gerada. Na representação vetorial de grafos, os nós do grafo são mapeados em um espaço vetorial derivado.

Nesse sentido, a matriz de adjacências apresenta-se como uma das estratégias mais simples e tradicionais de vetorização de grafos. A matriz de adjacência permite representar um grafo $\mathcal{G}$ composto por $|\mathcal{V}|$ vértices por meio de uma matriz $\mathbf{A}(\mathcal{G}) = [a_{ij}]$ de dimensão $n \times n$, em que os elementos a_{ij} da matriz dependem das propriedades intrínsecas ao grafo. Supondo um grafo não ponderado, seja direcionado ou não direcionado, cada elemento a_{ij} da matriz $\mathbf{A}$ contém 1, caso v_i e v_j sejam adjacentes e 0 caso contrário. Em grafos ponderados, os elementos a_{ij} retratam o peso da aresta entre v_i e v_j . Essa lógica de preenchimento garante que grafos não direcionados sejam mapeados em matrizes de adjacência simétricas ao longo da diagonal principal. Todavia, tal simetria não é garantida em grafos direcionados. Embora o vetor de adjacências de um nó codifique a estrutura de vizinhança de primeira ordem de um nó, o vetor resultante é esparsa, discreto e de alta dimensionalidade devido à natureza esparsa de redes de grande escala [46, 47].

2.6 Delimitação de Câmaras de Eco em Redes Sociais

As câmaras de eco são estruturas que têm o potencial de ampliar as ideias disseminadas dentro da própria estrutura ao mesmo tempo em que isola essas ideias de refutação. Essas estruturas são formadas por usuários altamente conectados. A comunidade em uma rede social também é uma estrutura altamente conectada, formada por um subconjunto de nós da rede que possuem maior densidade de conexões entre eles do que com o restante dos nós da rede. Assim, estruturalmente, as câmaras de eco podem ser interpretadas como comunidades em um grafo, de forma que algoritmos de detecção de comunidade amplamente difundidos na literatura podem ser empregados para identificar e delimitar as câmaras de eco. A qualidade das comunidades encontradas é avaliada por meio de métricas de avaliação extrínsecas e intrínsecas. Esta subseção apresenta diversos algoritmos utilizados para detecção de comunidade e métricas de avaliação que podem ser aplicados para delimitação e análise de câmaras de eco em redes sociais.

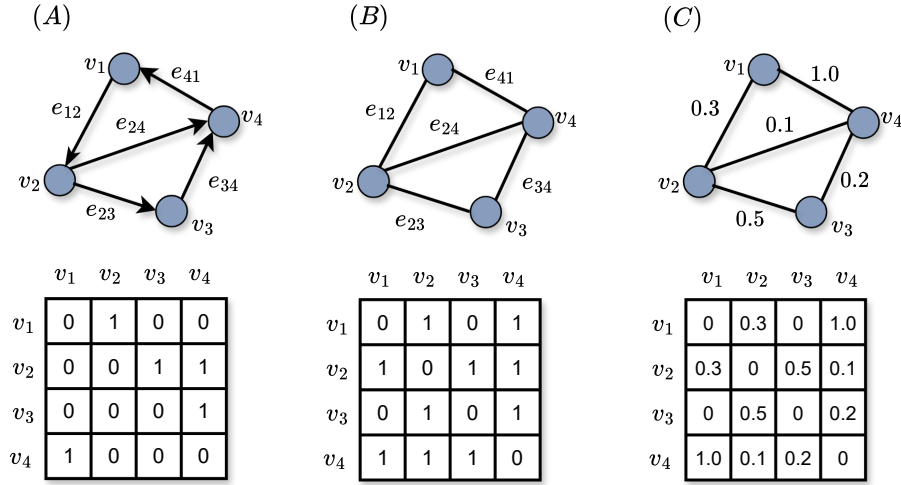


Figura 2.4: Diferentes tipos de grafos e suas respectivas representações matriciais de adjacência. A primeira linha de (A) a (C) são, respectivamente, exemplos de grafos direcionados, não direcionados e ponderados. A principal diferença entre (A) e (B) é que as arestas são direcionadas em (A), mas não direcionadas em (B). (C) mostra um grafo ponderado onde cada aresta é ponderada com um valor específico. As matrizes de adjacência 4×4 correspondentes a cada grafo encontram-se imediatamente abaixo dos mesmos. Convém ressaltar que como nenhum grafo possui auto-laços (*self-loops*), *i.e.*, arestas que se originam e terminam no mesmo nó, as matrizes de adjacência de cada grafo têm todos os elementos da diagonal principal nulos.

2.6.1 Algoritmos de Detecção de Comunidade

Os algoritmos de detecção de comunidade são capazes de desvendar a existência de uma organização de rede interna não trivial de modo geral. Isso permite inferir relações especiais entre os nós que podem não ser facilmente acessíveis a partir de testes empíricos diretos e ajuda a entender melhor as propriedades dos processos dinâmicos que ocorrem em uma rede [36]. Embora cada algoritmo detenha uma lógica própria, todos convergem para um objetivo comum de encontrar subgrafos com características estruturais homogêneas. Sendo a métrica base de diversos algoritmos, a modularidade reflete a tendência de formar agrupamentos que podem ser interpretados como comunidades, chamados de módulos da rede. Pode-se definir a modularidade (Q) de um grafo conexo por

$$Q = \frac{1}{2|\mathcal{E}|} \sum_{i,j} \left[w_{i,j} - \frac{C_{deg}(v_i)C_{deg}(v_j)}{2|\mathcal{E}|} \right] \gamma(c_i, c_j) \quad (2.13)$$

em que $|\mathcal{E}|$ é o número total de arestas do grafo, $C_{deg}(v_x)$ é o grau do nó v_x , que em um grafo ponderado é o peso total das arestas conectadas ao nó v_x , $w_{i,j}$ é o peso da aresta entre v_i e v_j . O nó v_x faz parte da comunidade c_x e há uma função indicadora $\gamma(c_i, c_j)$ que assume valor unitário, se $c_i = c_j$, e é nula, se $c_i \neq c_j$, ou seja, a função indicadora é

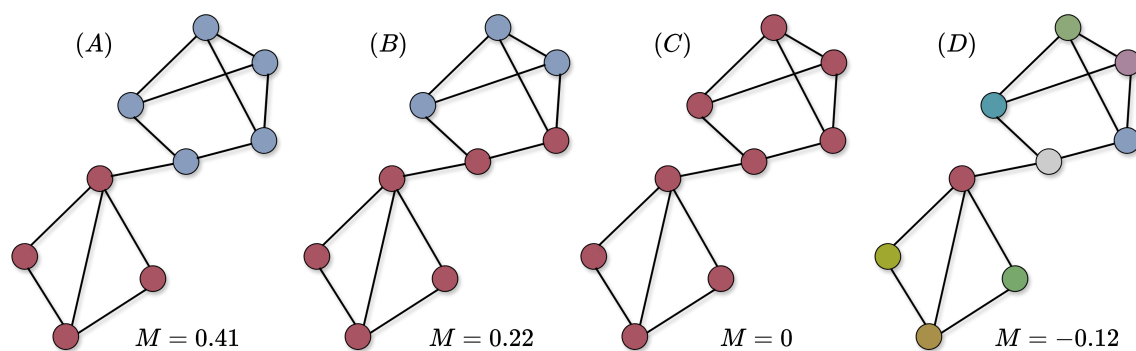


Figura 2.5: Variação da modularidade de um grafo conexo, considerando diferentes cenários de segregação em comunidades. As cores dos nós representam a comunidade à qual estão associados. A modularidade diminui do cenário (A) em direção ao cenário (D). Quanto maior a modularidade da rede, maior a tendência de formar comunidades com forte conectividade entre os membros.

igual a 1 se os nós v_i e v_j estão na mesma comunidade. A Figura 2.5 mostra a variação da modularidade de um grafo conexo, considerando diferentes cenários de segregação em comunidades. Em (A), o surgimento de duas comunidades bem definidas e fracamente conectadas entre si, resulta em um valor ótimo de modularidade. Em compensação, em (B) observa-se um cenário subótimo de divisão em comunidades. Em (C), constata-se a existência de uma única comunidade contendo todos os nós do grafo, implicando uma modularidade nula. Por fim, a modularidade negativa vista em (D) indica que cada nó do grafo é uma comunidade em si. Observa-se que quanto maior a modularidade da rede, maior a tendência de formar comunidades com forte conectividade entre os membros. A modularidade é capaz de medir o quão forte é a conexão de um nó adicionado a uma comunidade em contraste à sua adição a uma comunidade aleatória. Outros algoritmos empregam o passeio aleatório (*random walk*).

2.6.1.1 *Louvain*

O algoritmo *Louvain* é uma abordagem heurística não supervisionada que visa maximizar a modularidade a partir de sucessivas redistribuições de nós entre as várias comunidades de um grafo. Assim, o algoritmo *Louvain* divide-se em 2 fases, otimização da modularidade e agregação de comunidade. Após a otimização da modularidade executa-se a agregação de modularidade. Na fase de otimização da modularidade, o algoritmo ordena os nós aleatoriamente e verifica a modularidade ao remover cada nó de uma comunidade e adicioná-lo em outra comunidade, até que não haja mais um aumento significativo na modularidade. Na fase da agregação de comunidade, todos os nós que pertencem a uma mesma comu-

nidade são fundidos em um único nó representativo dessa comunidade, o nó gigante. Os enlaces que conectam nós gigantes são o conjunto dos enlaces que conectavam os nós que fazem parte do nó gigante. Isso pode gerar auto-laços (*self-loops*) conectando o nó gigante a ele mesmo. Essas duas fases são repetidas até alcançar a convergência, quando nenhum outro remanejamento de nós proporciona um aumento na modularidade total do grafo. A popularidade do algoritmo é reflexo da sua eficiência, sendo capaz de garantir a rápida construção de comunidades mesmo em redes extremamente populosas [7].

2.6.1.2 *Fast Greedy*

Implementando uma lógica baseada no agrupamento hierárquico aglomerativo, o *Fast Greedy* [48] busca segregar o grafo original em comunidades a partir da otimização da métrica de modularidade. Ao optar pelo método aglomerativo, *i.e.* *bottom-up*, o algoritmo inicia atribuindo uma comunidade única a cada nó do grafo. Guiando-se pelas potenciais alterações na modularidade do grafo, o *Fast Greedy* elege as duas comunidades a serem mescladas de cada iteração. Na prática, o par que fornece o máximo de melhoria de modularidade é selecionado para compor uma nova comunidade. Tal procedimento é repetido até que nenhuma fusão de pares de comunidade culmine em um aumento na modularidade. A eficiência do algoritmo reflete em um tempo de execução linear mesmo em redes extremamente grandes. Caso a rede analisada apresente uma estrutura esparsa e hierárquica, o algoritmo pode atingir uma complexidade de tempo de $O(|\mathcal{V}| \log^2 |\mathcal{V}|)$, em que $\mathcal{V}$ é o número de nós.

2.6.1.3 *WalkTrap*

O algoritmo *WalkTrap* [49] fundamenta-se na premissa de que passeios aleatórios de curtas distâncias tendem a permanecer na mesma comunidade. Como o *Fast Greedy*, o algoritmo *Walktrap* aplica uma lógica baseada no agrupamento hierárquico aglomerativo, em que nós, ou agrupamentos de nós, são recursivamente mesclados segundo um critério de união. Para lidar com a alta complexidade computacional de encontrar as comunidades ótimas, é possível empregar uma abordagem de Monte Carlo na estimação das probabilidades para os passeios aleatórios. Destaca-se que o algoritmo *Walktrap* atribui apenas uma comunidade a cada nó e, dessa forma, não há sobreposição entre as comunidades.

O algoritmo *WalkTrap* computa uma matriz de probabilidades de transição a partir de uma matriz de adjacências. Cada elemento da matriz de transição representa a probabilidade do passeio continuar para o nó adjacente com base na intensidade da relação

entre os nós. Utiliza-se um processo de passeio aleatório com uma quantidade pequena de passos para definir a probabilidade de transição entre um nó v_i e um nó v_j . Essa probabilidade é influenciada pelo grau do nó v_j de forma que existe uma maior probabilidade de transicionar para um nó com maior grau. A probabilidade de transição também é maior quando os nós v_i e v_j estão na mesma comunidade. Os passeios aleatórios definem uma distância entre os nós que é, então, generalizada para distância entre comunidades. Considera-se um passeio aleatório que se inicia em uma comunidade a partir de um nó inicial escolhido aleatoriamente e uniformemente dentre os nós da comunidade. Define-se a probabilidade de transição de partir de uma comunidade c para um nó v_j em t passos, a fim de encontrar a distância entre comunidades. Inicialmente, existem $|\mathcal{V}|$ comunidades, uma para cada nó. Computa-se todas as distâncias entre os nós adjacentes. Dados k passos, duas comunidades são escolhidas de acordo com um critério, essas duas comunidades são fundidas e as distâncias entre as comunidades atualmente existentes são computadas. Após $|\mathcal{V}| - 1$ passos, o algoritmo termina. O objetivo é minimizar a média das distâncias quadradas entre cada nó e a comunidade à qual pertence. A modularidade é normalmente usada para determinar a divisão ótima entre as comunidades.

2.6.1.4 *InfoMap*

O algoritmo *InfoMap* foca na otimização da equação de mapa (*map equation*) [50], uma equação que busca minimizar o comprimento da sequência usada para representar um passeio no grafo. Essa minimização é alcançada empregando a codificação de Huffman, um tipo de codificação sem perdas que garante que os nós mais visitados sejam representados por um número menor de bits. Para minimizar o comprimento da caminhada, o grafo pode ser dividido em diferentes módulos, onde cada módulo possui seu próprio livro de códigos (*codebook*). Há também um livro de códigos que representa o movimento entre os módulos, o livro de códigos de índice. O comprimento da descrição de um módulo pode ser representado pela equação do mapa:

$$L(M) = q_{\curvearrowright} H(Q) + \sum_{I=1}^M p^I H(P^I), \quad (2.14)$$

em que a primeira parte é a entropia do movimento entre as comunidades e a segunda parte é a entropia dos movimentos dentro das comunidades. As comunidades detectadas pela equação de mapa podem eventualmente divergir daquelas identificadas por algoritmos baseados na maximização da modularidade. Isso ocorre porque o cerne da equação de mapa está em otimizar o fluxo de informações, enquanto a modularidade baseia-se na

conexão entre nós.

2.6.2 Algoritmos Não Supervisionados

Algoritmos de agrupamento são a forma mais comum de aprendizado não-supervisionado. Apesar de possuírem lógica operacional, caso de uso, escalabilidade e desempenhos distintos, o propósito genérico de usar esses algoritmos é a segregação de termos em grupos (*clusters*) de acordo com suas características semânticas. Esse procedimento de separação em grupos é conhecido como agrupamento.

2.6.2.1 Algoritmos Baseados no Particionamento

Essa classificação é dada àqueles algoritmos que são semelhantes no sentido de que cumprem simultaneamente dois critérios no processo de agrupamento de dados. O primeiro critério expressa a obrigação de ter pelo menos uma amostra em cada agrupamento criado. O segundo refere-se a uma exclusividade de pertencimento, em que cada amostra deve pertencer a somente um agrupamento [51, 52].

Um exemplar clássico desse tipo de algoritmo é o *k-means*, uma heurística capaz de particionar dados em k agrupamentos pela minimização da soma dos quadrados das distâncias em cada agrupamento. Sua lógica de execução parte da escolha aleatória dos centroides de cada agrupamento seguida do cálculo de distância entre cada amostra e os centroides, segundo uma métrica de dissimilaridade ou similaridade. Posteriormente cada amostra é alocada no agrupamento cujo centroide está mais próximo. A cada nova amostra alocada a um agrupamento, o centroide é recalculado podendo ocorrer eventuais redistribuições de amostras para outros grupos. O algoritmo finaliza quando cessam essas alterações na alocação das amostras aos agrupamentos.

Outro exemplo é o algoritmo *k-medoids*, indicado para pequenos conjuntos de dados, e que também particiona os dados em k grupos adotando o critério de minimizar a soma dos quadrados das distâncias em cada grupo. Embora lembre o *k-means*, sua diferença está no fato de escolher efetivamente uma das amostras de entrada como centro dos agrupamentos, não pontos médios como o *k-means*. Essa característica de tomada de decisão se traduz em maior robustez a dados ruidosos e *outliers*, além de uma capacidade de lidar com alta dimensionalidade, útil em representações vetoriais de dados textuais [51, 52]. Outra vantagem do *k-medoids* em relação ao *k-means* está no fato de as saídas do *k-medoids* serem mais facilmente interpretadas, dado que os centros dos agrupamentos são amostras reais, ao contrário do *k-means* que fornece um ponto que pode representar

uma amostra de dados inviável. Contudo ambos os algoritmos, assim como outros, estão sujeitos a uma desvantagem singular: a indeterminação quanto ao número adequado de grupos k .

2.6.2.2 Algoritmos Hierárquicos

Os algoritmos hierárquicos não apenas criam agrupamentos, mas consideram uma lógica multinível e calculam uma representação hierárquica dos dados de entrada. Esta representação é um tipo particular de árvore, em que os nós-folhas expressam dados individuais, e pode ser construída seguindo um método aglomerativo ou divisivo. O método aglomerativo, conhecido também como abordagem *bottom-up*, começa considerando cada amostra como um agrupamento unitário e mescla recursivamente duas ou mais em um novo agrupamento seguindo uma função de ligação escolhida. Tais funções, quando associadas a métricas de distância ou de similaridade, definem critérios únicos que elegem os agrupamentos mesclados de cada iteração. A função de ligação única (*single-linkage*), por exemplo, estabelece a união considerando a distância entre as amostras mais próximos de cada agrupamento. De forma oposta, a função de ligação completa (*complete-linkage*) considera a distância das amostras mais distantes entre cada agrupamento. Paralelamente, a função de ligação média (*average linkage*) calcula a média das distâncias de todas as amostras de um agrupamento em relação a todas as amostras de outro agrupamento. Em especial, o critério de Ward emprega a distância Euclidiana na descoberta do par de agrupamentos que minimizam o aumento na variância total interna após a união.

Por outro lado, o método divisivo, i.e. abordagem *top-down*, inicia com uma estrutura plana em que todas as amostras pertencem ao mesmo agrupamento, ou seja, nível hierárquico. Portanto, a cada iteração, o algoritmo divide um ramo-pai em dois subconjuntos menores, os ramos-filhos. O processo termina quando um critério de parada é atingido, frequentemente, o número k de agrupamentos. No final do algoritmo, é criado um dendrograma de agrupamentos, uma hierarquia de árvore binária [53]. Um possível agrupamento hierárquico considerando a disposição espacial entre as amostras 1-6 da Figura 2.6(a) é ilustrada na Figura 2.6(b). Traçando as retas pontilhadas A-D perpendiculares aos ramos verticais do dendrograma é possível identificar diferentes momentos do processo de agrupamento. Em *A*, nota-se a existência de 6 agrupamentos unitários, ou seja, cada um contendo as amostras. Em *B*, constata-se a existência de 3 agrupamentos: o agrupamento unitário da amostra 1, o agrupamento das amostras 2 e 3, além do agrupamento formado pelas amostras 4, 5 e 6. Em *C*, já é possível identificar o mesmo par de

agrupamentos retratados na Figura 2.6(a). Por fim, em D verificamos a presença de um único agrupamento superpopuloso, contendo todas as amostras iniciais.

2.7 Métricas de Avaliação Extrínsecas e Intrínsecas

O processo de avaliação de heurísticas ou algoritmos de detecção de comunidade pode ser conduzido por meio de duas classes de métricas, as intrínsecas e as extrínsecas. Embora sejam igualmente úteis, a diferença entre ambas está na exigência, ou não, de uma verdade fundamental (*ground truth*), isto é, um rótulo de verdade fundamental atribuído a cada amostra. Tal rótulo funciona como base de comparação entre os resultados esperados e os obtidos. Em um cenário de detecção de câmaras de eco, esses rótulos podem expressar em qual comunidade, *i.e.* câmara de eco, um usuário participa. As métricas extrínsecas requerem obrigatoriamente a presença de amostras rotuladas, sendo portanto capazes de comparar o desempenho entre métodos. Na ausência de *ground truth*, muitos autores inferem o pertencimento, ou não, de um usuário a uma câmara de eco por meio da análise das *hashtags* compartilhadas. Apesar de subjetiva, essa inferência evita que a descoberta do rótulo de cada usuário seja dependente da análise textual com técnicas possivelmente mais demoradas e complexas de processamento de linguagem natural e aprendizado de máquina. Logo, pode-se inferir a qual comunidade um determinado usuário pertence,

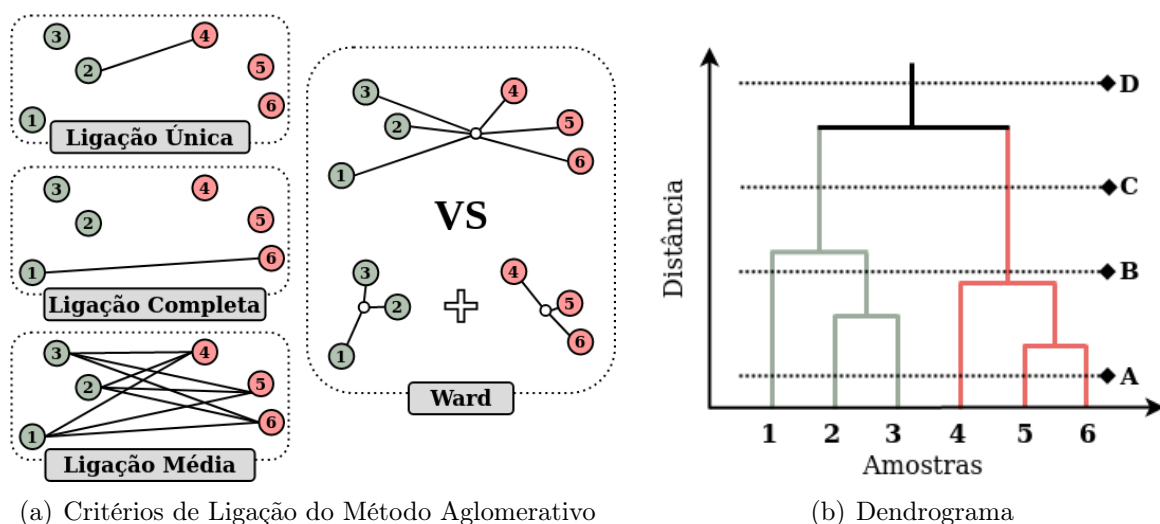


Figura 2.6: Critérios de ligação e dendrograma do algoritmo hierárquico. (a) Representação bidimensional de diferentes critérios de ligação considerando 6 amostras alocadas em dois agrupamentos. (b) Dendrograma resultante da aplicação do algoritmo hierárquico de agrupamento nas amostras 1-6, utilizando o método aglomerativo com critério de ligação única.

avaliando a ocorrência de *hashtags* em suas postagens. Além das medidas de recuperação da informação, *e.g.* acurácia, precisão, revocação, entre outras, as métricas extrínsecas também incluem:

- **Rand Index** (RI) que é um índice que expressa a similaridade entre os resultados previstos e os reais a partir da contabilização dos pares de amostras atribuídos na mesma ou em comunidades diferentes. Assumindo que $\mathcal{X} = \{x_1, \dots, x_n\}$ é o conjunto dos rótulos das n amostras atribuídas às comunidades retornadas pelo algoritmo de detecção de comunidade e $\mathcal{Y} = \{y_1, \dots, y_n\}$ é o conjunto dos rótulos das n amostras atribuídas às comunidades de *ground truth*, o índice é dado por

$$RI(\mathcal{X}, \mathcal{Y}) = \frac{2(a + b)}{n(n - 1)}, \quad (2.15)$$

em que n é o número de amostras, a expressa o número de pares de amostras que mantiveram-se na mesma comunidade em $\mathcal{X}$ e $\mathcal{Y}$, e b é o número de pares de amostras que foram alocadas em comunidades diferentes em $\mathcal{X}$ e $\mathcal{Y}$. A métrica é definida entre $[0, 1]$, em que 0 indica que os dois resultados previstos e reais não concordam em nenhum par de amostras e 1 reflete a completa concordância entre os dois resultados. O Adjusted Rand Index, uma versão corrigida e simétrica da métrica original, que introduz uma normalização estatisticamente induzida, produzindo valores próximos de zero para partições aleatórias.

- **Informação Mútua Normalizada** (NMI) que é a métrica derivada da informação mútua que é calculada entre os rótulos de *ground truth* e os rótulos previstos. Para tal, considera-se que $\mathcal{X} = \{X_1, \dots, X_n\}$ é o conjunto de rótulos das comunidades originalmente atribuídas às n amostras, enquanto $\mathcal{Y} = \{Y_1, \dots, Y_n\}$ é o conjunto de rótulos das comunidades atribuídas às n amostras após a aplicação de um algoritmo. Ao tratar $\mathcal{X}$ e $\mathcal{Y}$ como variáveis aleatórias discretas, a informação mútua normalizada entre ambas é expressa por

$$NMI(\mathcal{X}, \mathcal{Y}) = \frac{2I(\mathcal{X}; \mathcal{Y})}{H(\mathcal{X}) + H(\mathcal{Y})}, \quad (2.16)$$

em que $I(\mathcal{X} : \mathcal{Y})$ é a informação mútua entre $\mathcal{X}$ e $\mathcal{Y}$ e as entropias de $\mathcal{X}$ e $\mathcal{Y}$ são denotadas por $H(\mathcal{X})$ e $H(\mathcal{Y})$, respectivamente.

A escassez de conjuntos de dados previamente rotulados e a dificuldade de rotulagem frequentemente inviabilizam a utilização de métricas de avaliação extrínsecas. Dessa forma, cabe uma avaliação baseada em medidas intrínsecas, também chamadas de índices de

validação de agrupamento, para quantificar a coesão e separação das comunidades geradas [54]. Dentre essas métricas estão:

- **Calinski-Harabasz Index (CH)**: uma razão entre a dispersão dentro da comunidade e a dispersão entre as comunidades. Na prática, tais dispersões são obtidas através da soma de quadrados entre comunidades (SSB_c) e da soma de quadrados dentro de comunidades (SSW_c) expressos respectivamente pelas equações:

$$SSB_c = \sum_{j=1}^{|\mathcal{N}|} \|x_j - b_{c_j}\|^2 \quad (2.17) \quad SSW_c = \sum_{i=1}^{|\mathcal{C}|} \|b_i - \bar{X}\|^2, \quad (2.18)$$

em que $|\mathcal{N}|$ é o tamanho do conjunto das amostras x_i , com $i = 1, \dots, |\mathcal{N}|$ e $|\mathcal{C}|$ denota o número de comunidades. Os centroides das comunidades são denotados como b_i , em que $i = 1, \dots, |\mathcal{C}|$. Adicionalmente, cada amostra x_j pertence a uma comunidade c_j cujo centroide é denotado por b_{c_j} . Diante disso, a métrica pode ser computada através da equação

$$CH = \frac{SSB_c}{SSW_c} \cdot \frac{|\mathcal{N}| - |\mathcal{C}|}{|\mathcal{C}| - 1}. \quad (2.19)$$

Ao contrário do índice anterior, quanto maior o valor obtido, melhor a qualidade das comunidades geradas.

- **Davies-Bouldin Index (DB)**: que é um índice que fornece uma estimativa do grau de sobreposição do agrupamento. É definido como a medida de similaridade média de cada comunidade com seu par mais semelhante, em que a similaridade é a razão entre as distâncias dentro da comunidade e as distâncias entre as comunidades. Considerando $|\mathcal{C}|$ o número de comunidades, o índice é dado por

$$DB = \frac{1}{|\mathcal{C}|} \sum_{i=1}^{|\mathcal{C}|} \max_{i:j \neq i} \frac{S_i + S_j}{d_{i,j}}, \quad (2.20)$$

em que $S_i = \frac{1}{|c_i|} \sum_{x_j \in c_i} \|x_j - v_i\|$ é a medida de espalhamento dentro da comunidade c_i , x_j é um vetor de dimensão n atribuído à comunidade c_i , e $d_{i,j} = \|v_i - v_j\|$ é a distância euclidiana entre os centroides das comunidades c_i e c_j . Ao considerar o pior cenário de similaridade para cada comunidade, espera-se que quanto mais próximo de zero, o valor mínimo, melhor será o índice e, conseqüentemente, melhores serão os resultados do processo de agrupamento;

- **Coeficiente da Silhueta (SC)**: um coeficiente que quantifica a qualidade do agrupamento de dados com base na proximidade e na separação entre as comunidades

geradas. Conforme evidenciado na Equação 2.21, o coeficiente da silhueta é medido em função da média dos valores da silhueta de cada amostra do conjunto de dados. Definida entre $[-1, +1]$, a silhueta $s(i)$ para cada amostra i pode ser computada através da Equação 2.22.

$$SC = \frac{1}{n} \sum_{i=1}^n s(i) \quad (2.21)$$

$$s(i) = \frac{\bar{b}(i) - \bar{a}(i)}{\max(\bar{a}(i), \bar{b}(i))} \quad (2.22)$$

Na Equação 2.22, $\bar{a}(i)$ representa a distância média entre a amostra i em relação a todas as amostras pertencentes à mesma comunidade e $\bar{b}(i)$ reflete a distância média entre a amostra i em relação a todas as amostras. Em uma perspectiva local, $s(i)$ avalia a adequação de cada amostra individualmente. Nesse contexto, valores próximos a 1 expressam que a amostra observada encontra-se distante das comunidades vizinhas, indicando que essa amostra foi adequadamente alocada na comunidade à qual pertence. Valores nulos indicam que a amostra localiza-se no limite ou muito próximo do limite de decisão entre duas comunidades adjacentes. Em compensação, valores negativos indicam que a amostra foi possivelmente atribuída a uma comunidade errada. Em uma perspectiva global, o SC traduz a qualidade do agrupamento em comunidades considerando todo o conjunto de amostras. Assim, quanto mais próximo a 1, melhor a adequação de todas as amostras às comunidades a que pertencem. De maneira oposta, valores negativos informam que o processo de agrupamento não foi adequado. A Figura 2.7 mostra graficamente os reflexos no valor da silhueta quando adota-se uma divisão em comunidades não ótima.

2.7.1 Métricas de Deslocamento Semântico

As métricas de deslocamento semântico visam quantificar distância ou similaridade entre significados de estruturas textuais. Tradicionalmente, métricas como a Distância Euclidiana, Similaridade de Cosseno e a Distância de Jaccard realizam essa quantificação mediante versões vetorizadas de palavras, frases ou documentos. Em contraste a estas métricas, a **Distância Normalizada do Google** (*Normalized Google Distance – NGD*) [55] independe da necessidade de vetorizações prévias. Originalmente, a *NGD* é uma métrica assimétrica de distância semântica entre dois termos não vetorizados (x e y). Essa métrica está definida entre $[0, \infty[$ e pode ser expressa pela Equação 2.23 dada por

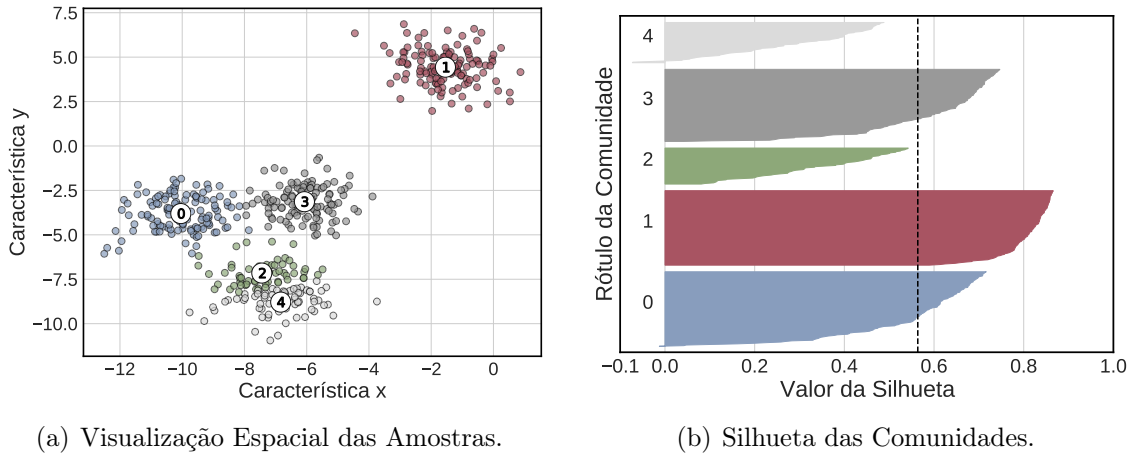


Figura 2.7: Valores da métrica silhueta para um conjunto de dados dividido em cinco comunidades. (a) As comunidades 2 e 4 estão muito próximas entre si, fato que contribui para o baixo valor da métrica nessas comunidades. (b) Cada silhueta, *i.e.*, manchas coloridas, tem comprimento horizontal e espessura diretamente proporcionais ao valor da silhueta das amostras associadas e à quantidade de amostras na comunidade, respectivamente. Em um cenário ideal de agrupamento em comunidades, o valor da silhueta correspondente a cada comunidade deve ser maior que o valor médio dos coeficientes de silhueta (linha pontilhada preta). Além disso, espera-se que nesse cenário as espessuras de cada silhueta sejam semelhantes entre si.

$$NGD(x, y) = \frac{\max\{\log f(x), \log f(y)\} - \log f(x, y)}{\log \mathcal{N} - \min\{\log f(x), \log f(y)\}}, \quad (2.23)$$

em que f_x e f_y são o número de resultados retornados pelo *Google* para os termos pesquisados x e y , respectivamente, e $f_{x,y}$ é o número de páginas da *web* em que x e y estão presentes. Embora a escolha do parâmetro $\mathcal{N}$ não seja pré-estabelecido, utiliza-se um valor grande³. Usualmente dois termos podem ser considerados semanticamente desconcorrelacionados quando a $NGD > 1$.

³Assim como outros trabalhos, esta qualificação adota $\mathcal{N} = 2,527 \times 10^{10}$, o número de resultados encontrados ao pesquisar o artigo “a” no *Google*.

Capítulo 3

Trabalhos Relacionados

Neste capítulo, conduz-se uma revisão abrangente da literatura científica dedicada ao estudo das câmaras de eco, bem como a investigação do papel das *hashtags* nas dinâmicas de interação entre usuários. Além disso, são explorados os principais mecanismos baseados na modelagem da propagação capazes de conter a disseminação descontrolada de informações em plataformas de redes sociais.

3.1 Investigação de Câmaras de Eco

Recentemente, o fenômeno das câmaras de eco tem sido frequentemente associado a múltiplas frentes de investigação, abrangendo temas como a modelagem de opinião, análise de sistemas de recomendação, disseminação de conteúdo falso, polarização ideológica e incidência de homofilia e controvérsia em grupos de usuários.

Em relação à **caracterização em Redes Sociais e Plataformas Web**, vários trabalhos abordam os efeitos e características da formação de câmaras de eco em torno de temas controversos nas redes sociais e em plataformas *web*. Dentre esses trabalhos, Cossard *et al.* analisam no *Twitter* o debate sobre vacinação tendo como premissa a existência de câmaras de eco [56]. Ao caracterizar e distinguir usuários em grupos de acordo com seus posicionamentos defensivo ou cético sobre vacinas, as características relacionais e textuais são aplicadas a algoritmos de classificação. Resultados qualitativos apontam que os céticos e os defensores da vacinação residem em suas próprias câmaras de eco e têm preferências por fontes de informação excludentes entre si. Ge *et al.* investigam a formação e os efeitos de câmaras de eco em sistemas de recomendação de plataformas de *e-commerce* [19]. Empregando uma base de dados real formada pelo histórico de acesso *web* dos usuários, os autores distinguem esses usuários em grupos com base na frequência

em que aceitam, ou ignoram, produtos recomendados. A análise temporal dos resultados sugere que as câmaras de eco apresentam um impacto maior sobre o comportamento de clique do usuário, enquanto têm sua influência mitigada sobre comportamentos de compra. Cinus *et al.* visam avaliar a contribuição dos sistemas de recomendação de pessoas no aumento ou diminuição das câmaras de eco [57]. Ao combinar um modelo de dinâmica de opinião com um algoritmo de recomendação de pessoas, a proposta é capaz de simular o comportamento de indivíduos mudando de opinião como consequência de suas interações com sua vizinhança. Resultados mostram que, caso haja uma homofilia inicial considerável na rede, as recomendações de pessoas podem efetivamente contribuir para a inclusão de novos usuários em câmaras de eco.

Outros trabalhos analisam o **comportamento dos usuários na disseminação de notícias falsas** e traçam relações de causa e efeito entre câmaras de eco e a propagação de desinformação. Törnberg modela a disseminação de notícias falsas nas mídias sociais atribuindo a cada usuário um limite que descreve o quão difícil é convencê-lo sobre uma determinada narrativa [23]. Caso uma fração suficientemente grande de seus vizinhos espalhe um mesmo conteúdo, o usuário é convencido e replicará o conteúdo. O modelo revela que as notícias originadas no interior de câmaras de eco têm uma maior disseminação do que em uma rede sem grupos definidos. Além disso, mostra-se que a simples reunião de usuários com visões homogêneas pode ser suficiente para aumentar a prevalência de desinformação, visto que a viralidade é diretamente proporcional à homofilia da rede. De forma similar, Bessi *et al.* conduzem uma análise quantitativa para investigar o consumo e a propagação de diferentes assuntos de teorias da conspiração dentro de câmaras de eco no *Facebook* [58]. A coleta e análise das curtidas, compartilhamentos e comentários dos usuários revelou que os diferentes assuntos conspiratórios são consumidos de maneira semelhante pelos usuários de câmaras de eco, com divergências apenas no tempo de atividade. Em especial, usuários polarizados em assuntos relacionados à geopolítica são mais ativos nos comentários.

Algoritmos, métodos e métricas de detecção também são propostos na investigação das câmaras de eco. Visando promover a redução da controvérsia em redes sociais, Garimella *et al.* investigam técnicas algorítmicas para interligar usuários de câmaras de eco isoladas [59]. Ao empregar uma métrica de quantificação de controvérsias, o modelo proposto busca pelas chamadas pontes, conexões entre usuários que minimizem essa métrica. Partindo da hipótese de que usuários localizados na borda de uma câmara de eco estão menos fechados a diferentes opiniões, espera-se que conteúdos enviados de fora da câmara de eco possam ser recebidos e eventualmente endossados pelo usuário que os re-

cebeu. Introduzida por Guerra *et al.*, a métrica *Boundary Connectivity* (BC) baseia-se nos conceitos de nós internos e nós limítrofes para mensurar a controvérsia entre duas comunidades [60]. Um nó é considerado limítrofe em uma comunidade caso tenha pelo menos uma conexão com um nó da comunidade oposta e uma conexão com um nó da sua própria comunidade que não esteja ligado à comunidade oposta. Os nós que não atendem a esses critérios são chamados de nós internos. Em cenários de câmaras de eco, os nós limítrofes de cada comunidade são mais fortemente conectados aos nós internos da mesma comunidade do que aos nós limítrofes da comunidade oposta. Similarmente, a métrica *Displacement Random Walk Controversy* (DRWC) é proposta por Villa *et al.* e avalia a controvérsia entre comunidades considerando a relação entre o comprimento total do passeio e o número de passos que resultam em mudança de comunidade [17]. A análise nodal sugere que a permanência de um nó em sua própria comunidade durante todo o passeio aleatório indica uma forte conexão com essa comunidade, o que implica controvérsia entre as comunidades. Por outro lado, mudanças frequentes de comunidade indicam um baixo grau de controvérsia.

A fim de quantificar a presença de câmaras de eco e polarização em redes sociais, Alatawi *et al.* propõem a *Echo Chamber Score* (ECS). A métrica avalia a dinâmica das interações entre usuários, medindo a coesão dentro das comunidades e a separação entre elas no espaço de incorporação [61]. Segundo o ECS, indícios estruturais de que uma comunidade pode ser uma câmara de eco são observados em comunidades de grafos que apresentam alta coesão interna e alta separação em relação às demais. Bara *et al.* investigam a capacidade de previsão métrica de lacuna de influência (*influence gap*) na antecipação de resultados de eleições multipartidárias em redes com estrutura de comunidade [62]. Essa métrica, originalmente proposta por Stewart *et al.*, mede a diferença de influência entre um partido e sua oposição em uma rede social [63]. Para tanto, os autores desenvolvem um modelo baseado no conceito topológico dos k -cliques e no comportamento homofílico dos usuários em cenários partidários. Tais características permitem ao modelo gerar grafos sintéticos com comunidades que podem exibir câmaras de eco e, assim, avaliar a dinâmica de opiniões através da métrica.

As métricas mencionadas, assim como as tradicionais métricas de recuperação da informação, intrínsecas e extrínsecas, enfrentam dificuldades ao tentar mensurar quantitativamente o grau de distinção entre as comunidades geradas por algoritmos categoricamente distintos. Essa dificuldade deriva da ingerência na escolha do número de comunidades geradas por alguns algoritmos. Como consequência, gera-se um conjunto de saídas, ou seja, rótulos atribuídos, que não estão necessariamente padronizados em número ou utilizando

a mesma notação de atribuição. Superando esta lacuna, o Método de Concordância de Rótulos (MCR) proposto baseia-se em uma decomposição espectral de uma matriz de pertencimento gerada pela análise dos rótulos. Tal manipulação projeta os resultados de diferentes algoritmos para uma dimensão comum, garantindo a possibilidade de inferir o nível de concordância entre seus resultados, mesmo que contenham rótulos diferentes ou descompassados.

3.2 Análise de *Hashtags*

As *hashtags* são utilizadas em pesquisas para diversas finalidades, incluindo a detecção [64] e monitoramento de eventos [65], predição de popularidade [66], desenvolvimento de sistemas de recomendação [67, 68], preservação de semântica [69], modelagem e extração de tópicos [70, 71].

Zhang propõe uma análise empírica tridimensional sobre as *hashtags* compartilhadas no *Instagram* [72]. Na dimensão espaço-temporal, os autores empregam um algoritmo não supervisionado para agrupar *hashtags* de acordo com a trajetória temporal. Explorando a dimensão semântica, os autores incorporam o significado de cada *hashtag* ao longo dos anos em vetores distintos e os comparam aplicando a similaridade do cosseno. Na dimensão social, investiga-se a utilização das *hashtags* como características válidas em uma abordagem não supervisionada para inferir relações sociais. Resultados mostram não somente uma maior tendência dos usuários em compartilhar *hashtags* em locais específicos, como também que aproximadamente 10% das *hashtags* sofrem deslocamento semântico. Cui *et al.* propõem uma abordagem para investigar as propriedades das *hashtags* para detecção de eventos repentinos, englobando tanto a caracterização quanto a aplicação de um algoritmo não supervisionado [73]. Para associar cada *hashtag* à ocorrência de um evento específico, os autores estabelecem níveis de intensidade em uma caracterização tridimensional contendo atributos como instabilidade temporal, possibilidade de ser conteúdo humorístico e entropia de autoria. Kowald *et al.* exploram a caracterização temporal para compreender o processo de reutilização de *hashtags* no *Twitter* [74]. Modelando a contribuição temporal de duas perspectivas, uma considerando *hashtags* de usuários antigos e a outra contendo *hashtags* postadas por seguidores do usuário, os autores desenvolvem um algoritmo preditivo para recomendação de *hashtags* inspirado na cognição humana. Os resultados revelam uma queda temporal na probabilidade de reutilização de *hashtags* com picos sazonais diários. Tal padrão expressa uma tendência dos usuários de utilizarem o mesmo conjunto de *hashtags* nesse período.

Ao contrário dos trabalhos anteriores de caracterização do uso de *hashtags*, que empregam dados represados, o sistema proposto prioriza a captura contínua de *hashtags* relacionadas a eventos reais. Essa condição introduz dinamismo à abordagem, possibilitando monitorar tendências tipográficas, padrões de conexão, além de detectar qualquer disjunção de domínios do conhecimento.

3.3 Contenção da Propagação de Informação

Embora permeie diversas finalidades, como controle epidemiológico, gestão de tráfego, gestão de abastecimento hídrico ou marketing viral, a incorporação de estratégias de mitigação da propagação de fluxos em grafos mostra-se eficaz na redução da disseminação de desinformação [43]. A prevalência do compartilhamento de informações inverídicas entre integrantes de câmaras de eco torna essas estratégias uma oportunidade para atenuar esse fenômeno. Ao focar trabalhos relacionados à mitigação da propagação de informação, pode-se distingui-los em dois grupos, dependendo do parâmetro modificado na estrutura do grafo analisado.

As **estratégias baseadas no bloqueio de nós** são também conhecidas como métodos de imunização de nós. Essa classe de estratégias de bloqueio baseia-se na identificação e remoção de um conjunto de nós críticos de um grafo antes do início do processo de propagação. Tais estratégias são consideradas mais disruptivas dado que a remoção de um nó implica a remoção de todas as suas arestas convergentes ou divergentes. Dentro deste escopo, Holme *et al.* propõem a identificação de nós-chave cuja remoção aumenta o comprimento geodésico médio inverso, ou seja, a distância média entre pares de nós no grafo [75]. O impacto da remoção de nós com alto grau e alta intermediação é avaliado considerando modelos de rede com propriedades estruturais e comportamentais distintas. Embora diminua a propagação de desinformação de maneira diversa dependendo da rede, a alta intermediação mostrou-se a solução testada mais eficiente para escolha dos nós bloqueados. Diferentemente, Dey e Roy [76] mostram que as maiores diminuições na propagação de informações na rede são alcançadas mediante a adoção de nós com alta centralidade de proximidade como nós bloqueados. Essa conclusão é fundamentada pela modelagem da propagação segundo o algoritmo de passeio aleatório, ao bloquear os nós mais centrais em termos de diversas centralidades. Paralelamente, Pham *et al.* desenvolvem um método para encontrar um conjunto de nós a serem removidos, considerando o custo associado a cada nó bloqueado [77]. O intuito é garantir que o custo total não exceda um limite e que o número máximo de saltos na propagação seja limitado. A partir da

construção de uma subárvore a partir do nó inicial, calcula-se a influência de cada nó nessa subárvore usando busca em profundidade. Ao aplicar programação dinâmica, garante-se uma solução quase ótima composta pelos nós com maior impacto na disseminação da desinformação.

As **estratégias baseadas no bloqueio de arestas** identificam e removem um conjunto específico de arestas de um grafo antes do início do processo de propagação. Apesar de serem menos disruptivas e permitirem uma canalização mais refinada do fluxo de informação, métodos dessa classe podem ser mais computacionalmente custosos devido à tradicional desproporção entre o número de nós e arestas. Dentre essas estratégias, Kimura *et al.* introduzem um método capaz de estimar e minimizar a contaminação em um grafo segundo três principais etapas [78]. Na primeira etapa, é executada a percolação de ligação, um processo repetido várias vezes no grafo original, resultando em uma série de M variações do grafo original com arestas aleatoriamente removidas. Na segunda etapa, calcula-se a influência de cada nó com base nas conexões presentes nos grafos resultantes. Essa influência é uma medida da extensão em que um nó afeta a propagação da contaminação no grafo. Na terceira etapa, um algoritmo guloso é aplicado para identificar as k arestas que minimizam a contaminação média ou máxima no grafo, selecionando-as de acordo com o objetivo definido. Khalil *et al.* têm como objetivo bloquear um conjunto de k arestas para reduzir a suscetibilidade à propagação na rede, definida como a soma das influências de todos os nós [79]. Sob o modelo de difusão *Linear Threshold*, demonstra-se que o problema formulado é submodular e monotônico. Para resolvê-lo, propõe-se um algoritmo ganancioso que utiliza o método de arestas ativas, garantindo uma solução próxima da ótima. Em outra perspectiva, Tong *et al.* visam identificar um conjunto de arestas cuja remoção minimize os autovalores da matriz de adjacência da rede, utilizados como medida de suscetibilidade à propagação [80]. Com base nos autovetores associados a esses autovalores, calcula-se uma pontuação para cada aresta, e as k arestas com as maiores pontuações são selecionadas como solução.

Perante os trabalhos publicados na literatura que objetivam reduzir a propagação de informação em redes sociais, a proposta desta tese diferencia-se em três aspectos. (i) O primeiro aspecto reside na aplicabilidade específica da proposta. Dentro do escopo bibliográfico pesquisado, não foram encontrados trabalhos prévios que se dediquem especificamente à minimização da propagação de informação dentro de câmaras de eco, tampouco utilizando uma estratégia otimizada baseada no bloqueio de nós críticos. (ii) O segundo diferencial diz respeito à eficácia das soluções encontradas pela estratégia proposta, uma vez que é baseada no *Simulated Annealing*. (iii) O terceiro aspecto relaciona-se

à abrangência das avaliações, uma vez que as modelagens são conduzidas considerando 15 exemplares de câmaras de eco explorados em cenários diversos. Além dessa avaliação intrametodológica, a proposta é comparada com outras estratégias de contenção de propagação, sob as mesmas condições iniciais. Independentemente dos cenários modelados e dos trabalhos comparados, a solução encontrada pela estratégia proposta resulta em um maior impacto na redução da propagação de informação na rede.

Capítulo 4

Propostas de Monitoramento, Análise e Contenção de Câmaras de Eco

Este capítulo detalha três propostas complementares para o enfrentamento dos efeitos disseminatórios em câmaras de eco: uma metodologia de monitoramento de redes sociais, um método para avaliação da concordância entre algoritmos de detecção de estruturas homogêneas e uma estratégia otimizada de seleção de nós impactantes na propagação de informação em redes.

4.1 Metodologia de Monitoramento de Redes Sociais

A primeira proposta compreende uma metodologia de monitoramento de redes sociais, capaz de revelar padrões latentes no comportamento dos usuários ao compartilharem postagens sobre temas controversos. Utilizando um evento preeminente e realista como estudo de caso, a metodologia permite o monitoramento contínuo do fluxo de *hashtags* relacionadas ao evento, seguido de sua caracterização segundo perspectivas temporais, semânticas e tipográficas.

A metodologia apresenta um processo de aplicação composto por cinco etapas desenvolvidas em linguagem Python conforme mostra a Figura 4.1. Para capturar o fluxo (*streaming*) de dados provenientes da rede social monitorada, o *Twitter* (atual *X*), a **Etapas de Captura** emprega a API nativa da plataforma. Ao ser inicializada com uma palavra-chave de entrada, esta etapa visa a captura ininterrupta do fluxo de *tweets* relacionado à entrada. Durante a captura, cada *tweet* é sequencialmente submetido à **Etapas de Estruturação**, cuja função é selecionar as *hashtags* utilizadas e correlacioná-las em um grafo relacional segundo uma lógica incremental. Para introduzir dinamismo e realismo

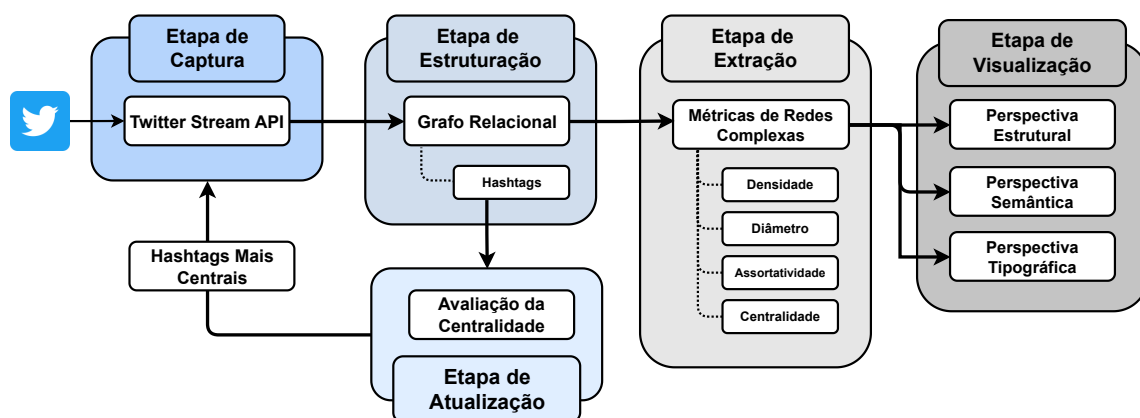


Figura 4.1: O fluxograma da metodologia proposta segue uma abordagem incremental para a caracterização temporal do conteúdo extraído do *Twitter*. A partir de uma palavra-chave de entrada inicial, os *tweets* relacionados são capturados e injetados num fluxo de processamento. Após a extração, as *hashtags* são incorporadas a um grafo relacional, que é atualizado periodicamente usando as *hashtags* mais centrais desse grafo como novas palavras-chave de entrada. Finalmente, o grafo de *hashtags* é analisado temporalmente.

à manutenção incremental do grafo, a **Etapa de Atualização** avalia periodicamente a centralidade dos nós do grafo estruturado e atualiza o foco do monitoramento com base nas *hashtags* mais centrais. No final de cada ciclo de monitoramento, a **Etapa de Extração** extrai diferentes métricas de redes complexas do grafo relacional formado. Por fim, a **Etapa de Visualização** exibe graficamente as métricas calculadas segundo três perspectivas: estrutural, tipográfica e semântica. As Subseções 4.1.1 a 4.1.5 abordam com mais detalhes cada uma das etapas supracitadas.

4.1.1 Etapa de Captura de Dados

A Etapa de Captura de Dados é responsável pelo *web scraping* de dados potencialmente relacionados a câmaras de eco em plataformas de redes sociais. Embora seja adaptável a múltiplas fontes de dados, essa etapa atualmente dedica-se exclusivamente ao monitoramento do *Twitter*. Como descrito na Subseção 2.2.1, a estratégia de monitoramento de câmaras de eco mais difundida na literatura é através da captura de postagens sobre tópicos controversos. Além de favorecer a polarização de opiniões, um tópico controverso costuma ser intensamente debatido nas redes sociais, gerando um alto engajamento e conseqüentemente dados para coleta. Perante essas observações, essa etapa emprega uma palavra-chave para nortear o processo de aquisição de dados. Para automatizar o processo de obtenção dos *tweets* relacionados à palavra-chave selecionada, adota-se um

*script*¹ desenvolvido em *Python* que acessa a API do *Twitter* usando credenciais de desenvolvedor. Em posse dessas credenciais, a biblioteca **Tweepy**² permite a extração de conteúdo textual dos *tweets* de qualquer perfil aberto na rede social segundo duas abordagens. A primeira abordagem destina-se à captura de metadados históricos, postados até o momento de execução do *script*. Apesar da simplicidade e do imediatismo da coleta, a abordagem enfrenta duas limitações impostas pelo *Twitter*, uma temporal e outra quantitativa. A limitação temporal remete a um número máximo de requisições passíveis de serem direcionadas à plataforma a cada janela de tempo de 15 minutos. Paralelamente, há uma limitação na quantidade de *tweets* históricos passíveis de serem coletados. Dependendo da permissão vinculada às credenciais, a obtenção de *tweets* é restringida a um período de até, no máximo, alguns meses passados a contar pela data de execução do *script*. Ciente dessas limitações e visando a dinamicidade e contemporaneidade da captura, a etapa é desenvolvida com base na segunda abordagem de coleta. Diferentemente da abordagem histórica, a abordagem em fluxo prevê a captura ininterrupta do fluxo de *tweets* relacionados à palavra-chave, publicados a partir do momento de execução do *script*. Logo, quaisquer *tweets* futuros publicados contendo a palavra-chave adotada serão coletados.

4.1.2 Etapa de Estruturação

Nutrido pela etapa anterior, a Etapa de Estruturação foca na criação e posterior manutenção incremental das estruturas em grafo relacionadas a cada tipo de metadado capturado. Em *Python*, a criação de estruturas em grafo é facilmente implementável por funções das bibliotecas **igraph**³ ou **NetworkX**⁴. Por ser implementado na linguagem C, o *igraph* detém um desempenho consideravelmente superior ao de bibliotecas puramente desenvolvidas em *Python*, como é o caso da *NetworkX*. Além disso, a *igraph* dispõe de uma gama maior de algoritmos de detecção de comunidades nativamente implementados. Em compensação, a *NetworkX* possui uma documentação mais detalhada e uma comunidade *online* mais ativa. Devido a esses fatores, emprega-se inicialmente a biblioteca *NetworkX* às tarefas de estruturação e extração de características.

Internamente, a Etapa de Estruturação trata cada *tweet* individualmente, respeitando a ordenação no fluxo de captura. Assim, cada *tweet* tem suas *hashtags* extraídas para

¹Disponível em <https://github.com/nicollasro/Echochamber>.

²Disponível em <https://www.tweepy.org/>.

³Disponível em <https://igraph.org/python/>.

⁴Disponível em <http://networkx.lanl.gov/>.

compor incrementalmente um grafo não direcionado de *hashtags*. Nesse grafo relacional, cada nó representa uma *hashtag* distinta e a aresta entre um par de nós indica que as *hashtags* aparecem simultaneamente no mesmo *tweet*. Nós isolados representam quaisquer ocorrências únicas de *hashtags* em um *tweet*. Devido à abordagem incremental, na qual novas capturas são utilizadas para atualizar o grafo relacional, a estrutura cresce continuamente.

4.1.3 Etapa de Atualização

A manutenção do dinamismo e a adaptabilidade a mudanças no rumo da captura de dados são objetivos da Etapa de Atualização. Essa etapa adota uma lógica de atualização do foco de monitoramento. Para tanto, as palavras-chave de entrada são substituídas a cada intervalo de monitoramento de uma hora pelas 20 *hashtags* mais centrais do grafo relacional atual. Essa seleção é determinada pelo cálculo da centralidade de proximidade (c_c) de cada nó (v_i), cujos detalhes estão descritos na Subseção 2.3.1. Quando usada em um grafo de *hashtags*, a centralidade de proximidade mede efetivamente o quão próxima, em média, cada *hashtag* está de todas as outras *hashtags* em uma comunidade, ou seja, dentro de um grupo de *hashtags* semelhantes que estão densamente conectadas. Logo, a *hashtag* mais central é a mais representativa da comunidade.

4.1.4 Etapa de Extração

Pautada pelos ciclos temporais de monitoramento, a Etapa de Extração calcula periodicamente diversas características próprias de cada versão do grafo de *hashtags*. Dentre as características clássicas inerentes a qualquer grafo, destaca-se a contabilização da quantidade de nós e arestas presentes em cada grafo e nas suas componentes mais populosas. Outras características são obtidas utilizando algumas das métricas de redes complexas descritas na Subseção 2.3.1, tais como densidade, diâmetro, assortatividade e centralidade. Com base na centralidade de proximidade, esta etapa visa contabilizar também as modificações no grupo de nós, i.e. *hashtags*, mais centrais do subgrafo principal. Além das características derivadas da análise do grafo em si, extrai-se características semânticas e tipográficas através da análise individualizada do texto que compõe cada *hashtag*. Dessa análise, obtém-se o deslocamento semântico entre *hashtags* segundo a métrica NGD (*Normalized Google Distance*) baseada no resultado de buscas no *Google*, bem como a quantidade de caracteres frequentemente adotada na escrita das *hashtags* capturadas.

4.1.5 Etapa de Visualização

Em posse das diversas características extraídas, a Etapa de Visualização é responsável por exibi-las segundo três perspectivas complementares: (i) estrutural, acompanhando o crescimento incremental do grafo de *hashtags*; (ii) tipográfica, revelando padrões de escrita adotados nas *hashtags* compartilhadas; (iii) semântico, medindo o deslocamento de significados entre as *hashtags* capturadas no monitoramento. A visualização dos resultados é realizada por meio de funções da biblioteca **Matplotlib**⁵ capaz de exibir gráficos em múltiplos formatos.

4.2 Estudo de Caso

Embora as *hashtags* estejam incorporadas em várias plataformas, como *Instagram*, *Facebook* e *LinkedIn*, seu uso como um indicador de um posicionamento ideológico ou preferencial é mais frequente nas postagens do *Twitter*. Visando monitorar um evento real e com opiniões controversas, a metodologia proposta na Seção 4.1 é aplicada em um estudo de caso sobre o episódio final do programa de televisão *Big Brother Brasil 2021*, a versão brasileira do *reality show* Big Brother, exibido em 4 de Maio de 2021 pela Rede Globo de Televisão. O foco nesse estudo de caso justifica-se pela conjunção de diferentes aspectos vantajosos para a análise de potenciais câmaras de eco, tais como:

- **Ampla Participação Pública.** O expressivo interesse por *reality shows* é quantificado pela alta audiência relatada pelo IBOPE (Instituto Brasileiro de Opinião Pública e Estatística)⁶, e pelo alto engajamento alavancado nas redes sociais, alcançando recordes de mensagens no *Twitter*⁷. Tais características contribuem para a geração de um grande montante de postagens analisáveis.
- **Formação de Comunidade Online.** Durante a exibição de *reality shows*, é natural que os usuários-espectadores associem-se aos perfis e postagens mais condizentes às suas preferências dentro do programa. Essa associação provê um cenário interessante para análise comportamental dos usuários-espectadores quando organizados em comunidades on-line intensamente ativas.

⁵Disponível em <https://matplotlib.org/>.

⁶Disponível em <https://www.kantaribopemedia.com/dados-de-audiencia-nas-15-pracas-regulares-com-base-no-ranking-consolidado-26-04-a-02-05-2021/>.

⁷Disponível em <https://www1.folha.uol.com.br/ilustrada/2021/02/bbb-21-e-marco-de-era-em-que-a-internet-domina-e-pauta-o-que-passa-na-tv.shtml>.

- **Uso de *Hashtags* Identificáveis.** Ao expressar seus posicionamentos sobre determinados participantes ou situações ocorridas dentro do programa, os usuários-espectadores costumam adotar *hashtags* específicas em suas postagens. Essa prática de auto-rotulação auxilia na categorização de quais grupos estes usuários pertencem.

Nesse contexto, para coletar *hashtags* relacionadas ao evento específico, emprega-se a API *Twitter Streaming*, que permite a captura todos os *tweets* recebidos relacionados à palavra-chave de entrada arbitrariamente escolhida: *BBB21*. Assim, após 37 horas de captura contínua, é consolidada uma base de dados contendo mais de 1,5 milhão de *tweets*, cujas estatísticas estão expressas na Tabela 4.1.

Tabela 4.1: Estatísticas da captura de dados relacionados ao episódio final do *BBB21*.

Intervalo Total de Captura	18:00 04/05/2021 - 07:00 06/05/2021
Quantidade Total de <i>Tweets</i>	1.737.753
Quantidade de <i>Tweets</i> Contendo <i>Hashtags</i>	443.547
Número de <i>Hashtags</i> Distintas	7175
Vazão Média	782 <i>tweets</i> /min
Vazão Máxima	2998 <i>tweets</i> /min

4.3 Método de Concordância de Rótulos

A segunda proposta consiste em um método de avaliação da concordância entre diferentes algoritmos na tarefa de detectar estruturas semelhantes a câmaras de eco em grafos. O método proposto visa comparar duas abordagens de particionamento de grafos em grupos semelhantes: uma baseada em algoritmos de detecção de comunidades e outra baseada em algoritmos de aprendizado de máquina não supervisionado. Independentemente da abordagem utilizada, esses algoritmos geram rótulos para cada nó com base em sua lógica particular. O conjunto de rótulos resultante da execução de cada algoritmo é o objeto de análise do método e podem ser representados como $\mathcal{R} = \{r_{v_1}, r_{v_2}, \dots, r_{v_n}\}$, em que cada r_{v_i} indica a comunidade ou agrupamento ao qual o nó v_i pertence. O método é composto por três etapas principais sequenciais: (i) mapeamento intra-algorítmico, (ii) decomposição em autovetores e (iii) quantificação da similaridade ou distância.

A primeira etapa do método de concordância proposto consiste em um **Mapeamento Intra-Algorítmico**, onde os rótulos atribuídos a cada par de nós são comparados iterativamente para determinar se pertencem à mesma comunidade. Este mapeamento deriva uma matriz de pertencimento M_p de dimensão $n \times n$, onde n é o número de nós do grafo.

Os elementos m_{ij} da matriz de pertencimento são determinados segundo a formalização matemática dada pela Equação 4.1.

$$\forall v_i, v_j \in \mathcal{V}, \quad m_{ij} = \begin{cases} 1 & \text{se } r_{v_i} = r_{v_j} \\ 0 & \text{caso contrário} \end{cases} \quad (4.1)$$

A segunda etapa compreende a **Decomposição Espectral**, processo pelo qual a matriz de pertencimento M_p é expressa em termos de seus autovalores e autovetores. Sendo M_p uma matriz quadrada de ordem n , pode-se garantir que existem valores escalares λ_i e vetores $\mathbf{v}_i$ tais que

$$M_p \mathbf{v}_i = \lambda_i \mathbf{v}_i. \quad (4.2)$$

onde λ_i são os autovalores e $\mathbf{v}_i$ são os autovetores correspondentes. Como expresso na Equação 4.3, a matriz M_p pode ser diagonalizada se houver um conjunto completo de n autovetores linearmente independentes, formando uma matriz $\mathbf{V}$, e uma matriz diagonal $\mathbf{\Lambda}$ contendo os autovalores.

$$M_p = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^{-1}. \quad (4.3)$$

A terceira etapa, **Quantificação da Similaridade**, concentra-se na aplicação de métricas de distância ou similaridade entre pares de matrizes de autovetores normalizados. A interpretação dos valores obtidos permite inferir o nível de concordância entre os resultados dos algoritmos quando aplicados à mesma câmara de eco, mesmo que os rótulos dos agrupamentos gerados sejam diferentes ou descompassados.

4.4 Estratégia de Bloqueio Otimizada

A terceira proposta visa reduzir a disseminação excessiva de conteúdos em redes sociais, especialmente em câmaras de eco. Para isso, propõe-se a **ECO-SA**, uma **Estratégia de Contenção Otimizada** da propagação em câmaras de eco baseada no *Simulated Annealing*. Essa estratégia é uma variação da meta-heurística *Simulated Annealing* (SA), adaptada para otimizar a busca por um subconjunto estratégico de nós em um grafo, que quando bloqueados minimizam a propagação de conteúdos dentro da rede. Assim, o ECO-SA combina a exploração aleatória de soluções com um mecanismo de aceitação probabilística para encontrar uma solução ótima.

A meta-heurística SA foi originalmente desenvolvida com o propósito de encontrar o mínimo global de uma função, baseando-se nos princípios da mecânica estatística. Esta

busca parte de uma solução inicial gerada aleatoriamente e possui uma convergência comprovadamente assegurada para uma solução global ótima ξ [81]. Em cada iteração, a meta-heurística produz uma solução aleatória ξ' na vizinhança da melhor solução atualmente conhecida. O desempenho da solução candidata ξ é avaliado segundo uma função de perda $f(\xi)$, que no contexto da mitigação da propagação da informação é a função de perda global. Soluções que minimizam a função de perda são sempre aceitas, enquanto que as demais podem ser aceitas probabilisticamente dependendo do parâmetro de temperatura simulada T atual e da degradação ΔE da função de perda. Expresso na Equação 4.4, o ΔE quantifica a diferença entre o valor da função de perda para a solução atual $f(\xi)$ e para a solução candidata $f(\xi')$.

$$\Delta E = f(\xi') - f(\xi) \quad (4.4) \quad P(\Delta E, T) = e^{-\Delta E/T} \quad (4.5)$$

A função de probabilidade dada pela Equação 4.5 mede a probabilidade de transição da solução atual ξ para uma solução candidata pior ξ' , seguindo a distribuição de Boltzmann. Uma temperatura mais baixa reduz a probabilidade de aceitar uma solução pior, favorecendo a melhor solução. O valor negativo da razão ΔE e T na função exponencial reforça essa diminuição de probabilidade. O parâmetro α , definido no intervalo $]0, 1[$, representa a redução gradual da temperatura simulada [82].

Computacionalmente, a estratégia ECO-SA pode ser codificada segundo o Algoritmo 1 e suas funções auxiliares descritas nos Algoritmos 2 e 3. Dentre os parâmetros de entrada admitidos pelo algoritmo estão: o grafo analisado ($\mathcal{G}$), o conjunto de nós propagadores iniciais (V_0), a temperatura inicial (T_0), a temperatura mínima (T_{min}), a taxa de resfriamento (α) e uma solução inicial ($S_{inicial}$). Esta solução é modelada como um conjunto N composto por nós bloqueados escolhidos aleatoriamente. Sendo uma estratégia insensível à fonte, a ECO-SA sempre exclui os nós iniciais desse conjunto de bloqueio.

A inicialização do algoritmo inclui a definição da temperatura atual (T) como sendo igual a T_0 e a determinação do conjunto de nós disponíveis para bloqueio (V_D), dado pela diferença entre o conjunto total de nós do grafo e o conjunto de nós iniciais (linhas 1 e 2). O impacto mitigatório da solução inicial no grafo analisado é modelado segundo a função objetivo **PropagationImpact** (linha 4). Internamente, a função implementa um modelo de difusão de informação iniciado nos nós iniciais e bloqueando os nós definidos pela solução inicial. Assim, o custo relacionado à solução inicial é retornado na forma de nível de propagação nominal, ou seja, na quantidade de nós ativados ao término de um número fixo de iterações do modelo. Essa solução inicial é considerada a melhor solução encontrada até o momento, e seu custo é registrado como o melhor custo (linhas 5 e

Algoritmo 1: Estratégia de Bloqueio de Nós Baseada no *Simulated Annealing*

Entrada: $G = (V, E)$, V_0 , T_0 , T_{min} , α , $S_{inicial}$
Saída : S_{melhor}

```

1  $T \leftarrow T_0$ 
2  $V_D \leftarrow V - V_0$ 
3  $S_{atual} \leftarrow S_{inicial}$ 
4  $C_{atual} \leftarrow \text{PropagationImpact}(G, V_0, S_{atual})$ 
5  $S_{melhor} \leftarrow S_{atual}$ 
6  $C_{melhor} \leftarrow C_{atual}$ 
7 while  $T \geq T_{min}$  do
8    $S_{candidata} \leftarrow \text{UpdateSolution}(V_D, S_{atual})$ 
9    $C_{novo} \leftarrow \text{PropagationImpact}(G, V_0, S_{candidata})$ 
10   $\Delta E \leftarrow C_{novo} - C_{atual}$ 
11   $P = e^{-\Delta E/T}$ 
12  if  $\Delta E < 0$  or  $\text{norm.cdf}(\text{np.random.normal}(0, 1)) < P$  then
13     $S_{atual} \leftarrow S_{candidata}$ 
14     $C_{atual} \leftarrow C_{novo}$ 
15    if  $C_{atual} < C_{melhor}$  then
16       $S_{melhor} \leftarrow S_{atual}$ 
17       $C_{melhor} \leftarrow C_{atual}$ 
18    end
19  end
20   $T \leftarrow \alpha \cdot T$ 
21 end
22 return  $S_{melhor}$ 

```

6). Em seguida, o algoritmo mantém-se em um *loop* principal, enquanto a temperatura atual for maior ou igual à temperatura mínima T_{min} . Dentro do *loop*, diferentes soluções candidatas são exploradas com o objetivo de encontrar aquela que minimize a propagação. A cada iteração, uma solução candidata ($S_{candidata}$) é obtida modificando a solução atual usando a função `UpdateSolution` (linha 8). Essa função é responsável pela substituição aleatória de um nó presente na solução atual por outro pertencente ao conjunto de nós disponíveis para bloqueio. A solução candidata encontrada tem seu custo calculado usando a função objetivo e comparado com o custo atual (linhas 9 e 10). A partir da variação de energia entre a solução atual e a solução candidata e da temperatura atual, calcula-se a probabilidade P (linha 11). Quaisquer soluções que derivem em variações negativas são imediatamente aceitas. Contudo, a solução piorante também pode ser aceita, desde que a probabilidade calculada seja maior que a função de distribuição acumulada de uma variável aleatória normal com média zero e desvio padrão um (linha 12). A solução candidata aceita e seus respectivos custos (linhas 13 e 14) atualizam a solução atual e seu custo, respectivamente. Em seguida, verifica-se se o custo atual C_{atual} é menor do

Algoritmo 2: Função de Atualização do Conjunto de Nós Bloqueados

```

1 Function UpdateSolution( $V_D, S_{candidata}$ ):
2    $v_{removido} \leftarrow \text{RandomChoice}(1, S_{candidata})$ 
3    $S_{candidata} \leftarrow S_{candidata} - \{v_{removido}\}$ 
4    $v_{adicionado} \leftarrow \text{RandomChoice}(1, V_D \setminus S_{candidata})$ 
5    $S_{candidata} \leftarrow S_{candidata} + \{v_{adicionado}\}$ 
6   return  $S_{candidata}$ 
7 End Function

```

Algoritmo 3: Função Objetivo

```

1 Function PropagationImpact( $G, V_0, S$ ):
2    $N_A \leftarrow \text{RunModel}(G, V_0, S)$ 
3   return  $N_A$ 
4 End Function

```

que o custo da melhor solução encontrada até o momento C_{melhor} . Se isso for verdade, a melhor solução S_{melhor} e seu custo C_{melhor} são atualizados com os valores da solução atual e do novo custo, respectivamente. Durante o decorrer do *loop*, a temperatura atual é gradualmente reduzida multiplicando-a pelo fator de resfriamento (α), o que controla a probabilidade de aceitar soluções piores (linha 20). Tal procedimento é realizado para prevenir o aprisionamento em ótimos locais subótimos e promover a exploração abrangente do espaço de soluções. Uma vez atingida a condição de parada ($T < T_{min}$), o algoritmo retorna a melhor solução encontrada S_{melhor} , que representa o conjunto ótimo de nós a serem bloqueados.

4.5 Base de Dados de Câmaras de Eco

As propostas de comparação da concordância e mitigação da propagação descritas nas Seções 4.3 e 4.4 são avaliadas utilizando uma base de dados originalmente apresentada por Morini *et al.* [16]. A base de dados é composta por exemplares de câmaras de eco construídas a partir do monitoramento do uso da rede social *Reddit* por usuários americanos. Embora pouco explorado na literatura, o *Reddit* facilita o processo de construção de câmaras de eco, uma vez que a própria plataforma é organizada em comunidades. As 15 amostras de câmaras de eco permeiam o debate sobre três questões sociopolíticas: controle de armas, discriminação de minorias e política americana, durante cinco semestres consecutivos. Na prática, cada câmara de eco é representada por um grafo não direcionado e ponderado $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$, onde cada nó v_i representa um usuário distinto, cada aresta $\varepsilon_{i,j}$ representa uma interação entre usuários e o peso $w_{i,j}$ representa a intensidade

dessa interação entre eles. Mediante uma análise manual da inclinação ideológica de cada postagem, os autores classificaram cada nó em um dos três possíveis rótulos: Pro-Trump, Anti-Trump ou Neutro. Detalhes estruturais e as métricas de controvérsia RWC e ARWC, e de condutância $\phi(\mathcal{G})$ sobre cada amostra são expressos na Tabela 4.2.

Tabela 4.2: Características, condutância e controvérsia das câmaras de eco analisadas.

Tópico Debatido	Snapshot	Período	# V	# E	Densidade	Diâmetro	$\phi(\mathcal{G})$	RWC	ARWC
Controle de Armas	S_1	01/17 - 07/17	833	4044	0.006	11	0.444	0.172	46.4
	S_2	07/17 - 01/18	847	3925	0.007	11	0.498	0.173	35.8
	S_3	01/18 - 07/18	1054	3942	0.007	10	0.494	0.161	39.3
	S_4	07/18 - 01/19	985	3478	0.006	10	0.551	0.182	39.8
	S_5	01/19 - 07/19	1046	3601	0.006	10	0.459	0.133	36.2
Discriminação de Minorias	S_1	01/17 - 07/17	1040	3765	0.007	11	0.421	0.288	45.6
	S_2	07/17 - 01/18	1004	3465	0.007	10	0.417	0.266	42.4
	S_3	01/18 - 07/18	1170	3832	0.006	10	0.417	0.257	54.8
	S_4	07/18 - 01/19	1113	3594	0.006	10	0.409	0.269	47.2
	S_5	01/19 - 07/19	1126	3405	0.007	11	0.404	0.273	51.5
Esfera Política	S_1	01/17 - 07/17	917	2525	0.007	10	0.639	0.051	31.7
	S_2	07/17 - 01/18	746	1816	0.006	12	0.561	0.061	31.3
	S_3	01/18 - 07/18	825	2179	0.006	13	0.516	0.059	37.3
	S_4	07/18 - 01/19	775	1787	0.005	11	0.543	0.039	36.0
	S_5	01/19 - 07/19	686	1411	0.012	8	0.558	0.047	44.8

Capítulo 5

Resultados e Discussão

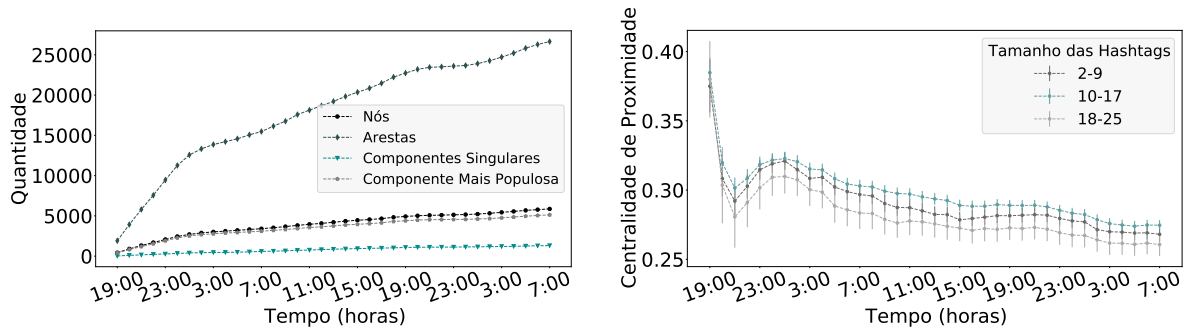
Este capítulo apresenta e discute os resultados obtidos, dividindo-os em três grupos. O primeiro grupo apresenta as análises do monitoramento e caracterização do fluxo de postagens relacionadas a um evento controverso no *Twitter*. O segundo grupo trata da aplicação do método de concordância de rótulos em diferentes algoritmos para a tarefa de detectar câmaras de eco. Por fim, o terceiro grupo revela os resultados da modelagem da propagação de informação dentro de câmaras de eco em diferentes cenários, bem como compara o impacto da proposta com outras estratégias na contenção da disseminação.

5.1 Resultados do Monitoramento e Caracterização

Na caracterização temporal conduzida neste trabalho, são considerados aspectos como (i) o crescimento estrutural do grafo relacional de *hashtags*; (ii) os padrões de escrita das *hashtags*; (iii) métricas de redes complexas como centralidade, densidade, diâmetro e assortatividade; e (iv) a distância entre os significados de *hashtags* pertencentes à componente principal do grafo gerado. Vale ressaltar que os resultados apresentados são médias com intervalo de confiança de 95%.

5.1.1 Análise Temporal e Tipográfica

Em uma primeira análise, a Figura 5.1(a) mostra o progresso temporal do grafo relacional em diferentes perspectivas, uma avaliando o número de arestas totais e outra considerando o número de nós no grafo e seus componentes, i.e. subgrafos. Quando comparadas, essas perspectivas divergem no padrão de crescimento, evidenciando que todas as curvas relacionadas a nós apresentam pouca variação. Além disso, constata-se que a componente



(a) Crescimento temporal do grafo relacional de *hashtags* quando se considera o número de nós e arestas.

(b) Variação temporal da centralidade de proximidade considerando *hashtags* com diferentes quantidades de caracteres.

Figura 5.1: Caracterização temporal da estrutura do grafo gerado e do padrão de escrita das *hashtags* que o integram.

principal concentra a maioria dos nós e, consequentemente, *hashtags* de todo o grafo relacional. Em contraste, a curva de arestas possui um comportamento ascendente ao longo da captura, demonstrando que usuários estão mais propensos a utilizar novas associações de *hashtags* previamente usadas do que propriamente criar novas. A existência de vários componentes não conexos com nó único no grafo gerado, marcada pela presença de nós isolados do componente mais populoso do grafo, pode revelar *hashtags* oportunistas. Tal fenômeno possivelmente ocorre quando o usuário, intencionalmente, ou não, posta *tweets* relacionados ao evento monitorado, mas criam ou usam *hashtags* específicas que não possuem adesão generalizada. Embora componentes desconectados sejam inerentes às primeiras horas de captura, sua permanência até o fim do período de captura reforça a suposição de que essas *hashtags* não pertencem ao domínio de conhecimento do evento, mas representam um comportamento de usuários oportunistas.

Em uma caracterização tipográfica é possível avaliar o padrão de escrita das *hashtags*. A Figura 5.1(b) representa a média da centralidade de proximidade para diferentes intervalos de tamanho de *hashtag*. O tamanho é medido em número de caracteres de cada *hashtag*. Conforme esperado, para as primeiras horas de captura, percebe-se que todas as curvas possuem alta variabilidade dos valores de centralidade, marcada pelos grandes intervalos de confiança. Essa variabilidade reflete diretamente o uso de uma única *hashtag* como entrada, o que inicialmente favorece as *hashtags* presentes nos *tweets* relacionados a essa entrada enviesada. Assim, a entrada inicial é a mais representativa da comunidade formada pelo primeiro conjunto de *hashtags* coletadas. Outra conclusão está relacionada à maior centralidade das *hashtags* que não são muito curtas nem muito longas. A adesão dos usuários ao uso de *hashtags* de tamanho médio pode estar relacionada à praticidade de

sua escrita, por não serem muito longas, e ao fato de poderem transmitir mais informações do que *hashtags* muito curtas.

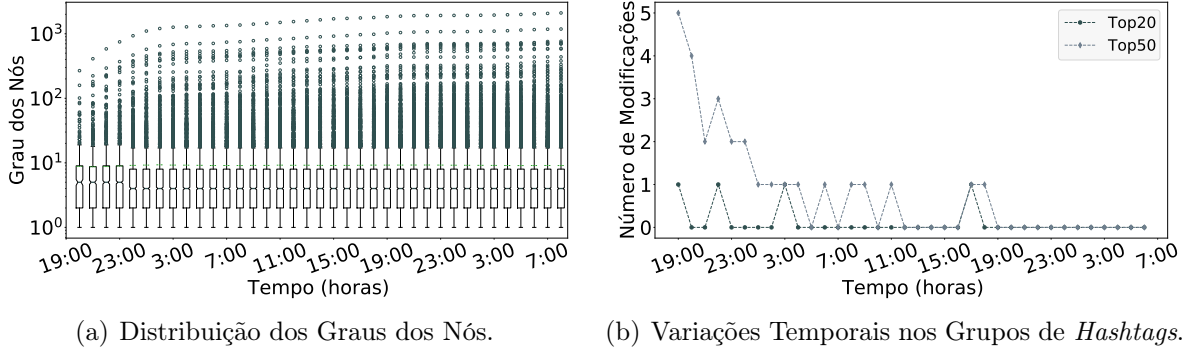


Figura 5.2: Evolução da distribuição de grau e das alterações nos grupos de *hashtags* mais centrais ao longo do tempo. Na Figura 5.2(b) é notória a presença de um grupo mais coeso de *hashtags* mais centrais (Top20), uma vez que este apresenta apenas três modificações durante toda a captura do fluxo de dados.

Considerando todo o grafo relacional, analisa-se também a distribuição do grau dos nós, aspecto fundamental para compreender a dinâmica das associações entre *hashtags*. O diagrama de caixa na Figura 5.2(a) torna evidente a disparidade entre os graus dos nós do grafo. Uma vez que o grau de um nó em um grafo não orientado refere-se ao número de arestas que se conectam a ele, é possível afirmar que o valor médio do grau ao longo do tempo é constante e próximo a 10, como mostra a marcação em verde. No entanto, alguns nós estão muito distantes desse padrão de conectividade, variando entre 11 e mais de 1.000 conexões. Em particular, ao observar o nó com o grau mais discrepante em relação à média (*#BBB21*), corrobora-se a hipótese de viés no monitoramento.

A Figura 5.2(b) fornece uma visão discretizada do número de mudanças no grupo mais central de *hashtags* entre cada intervalo de captura de 1 hora. Analisando o grupo das 50 *hashtags* mais centrais (Top50), nota-se um declínio perceptível no número de alterações ao longo do tempo, enfatizado pelo período a partir da 24^a hora de captura, na qual não ocorrem mais alterações. Em contraste, a curva das 20 *hashtags* mais centrais (Top20) expressa um comportamento mais estável, com apenas quatro mudanças em todo o tempo de captura. Comparando as duas curvas, é possível inferir que (i) como o maior número de mudanças (87%) ocorre fora do Top20, há um núcleo quase imutável de *hashtags* mais centrais. Esse fenômeno fortalece a hipótese de que para entender um evento no *Twitter* de forma plena, basta monitorar um grupo restrito de *hashtags* relacionadas àquele evento, evitando desperdício de tempo e recursos; (ii) a estabilização do número de mudanças observadas nas duas curvas pode ser interpretada como uma saturação do

tema monitorado.

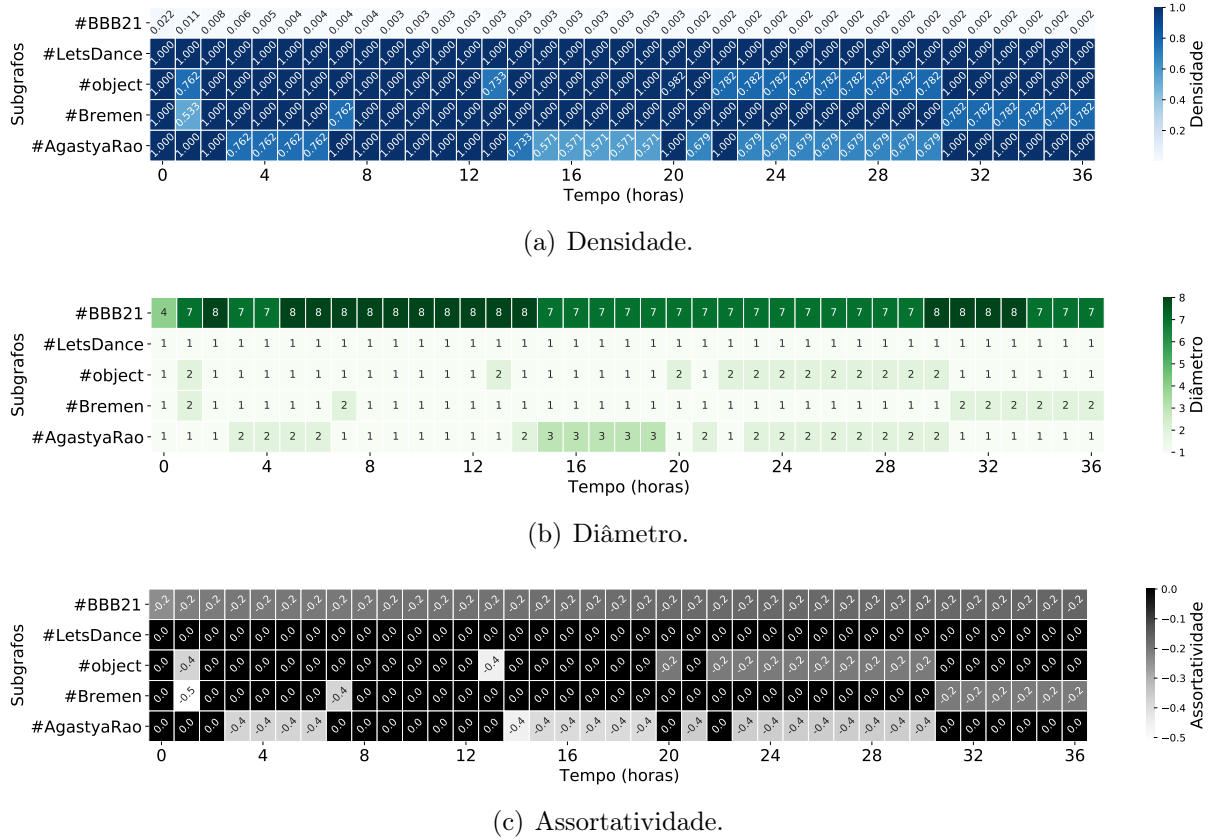


Figura 5.3: Discretização temporal das métricas de redes complexas referentes aos cinco subgrafos mais populosos.

A existência de componentes desconexas no grafo de *hashtags*, como discutida na análise da Figura 5.1(a), motiva a realização de uma análise temporal individualizada para cada subgrafo. Ao considerar a análise de métricas de redes complexas, como densidade, diâmetro e assortatividade, torna-se possível quantificar diversas características latentes que integram um grafo. Diante de eventuais uniões de subgrafos ao longo da captura, essa análise foca nos cinco subgrafos mais populosos que se mantêm disjuntos durante todo o processo. Em uma primeira perspectiva, a Figura 5.3(a) apresenta os resultados da análise para a métrica de densidade, previamente explicada na Seção 2.3.1. O indicativo de baixa densidade, dado pela tonalidade clara observada no subgrafo associado à *hashtag* #BBB21, evidencia uma constante prevalência da quantidade de nós em detrimento da quantidade de arestas. Isso reitera a suposição feita na análise da Figura 5.1(a), de que usuários ao debaterem um determinado tema preferem publicar *tweets* com novas associações de *hashtags* previamente usadas a inventar novas *hashtags*. Ressalta-se ainda que a baixa densidade do subgrafo principal está relacionada ao uso de combinações prevalentes de *hashtags*.

Em uma segunda perspectiva comparativa, são dispostos temporalmente os diâmetros (d_{ia}) de cada subgrafo, calculando-os com base na Equação 2.2 expressa na Seção 2.3.1. A Figura 5.3(b) mostra que há uma diferença significativa entre os diâmetros de cada subgrafo avaliado ao longo do tempo. Enquanto que os subgrafos *#LetsDance*, *#object*, *#Bremen* e *#AgastyaRao* apresentam diâmetros que não ultrapassam três saltos ($d_{ia} < 3$), o subgrafo principal (*#BBB21*), relacionado ao evento monitorado, apresenta diâmetros variando entre 7 e 8. Mesmo sendo medida em saltos, essa distância entre nós pertencentes ao mesmo subgrafo revela *hashtags* relacionadas a assuntos ou nichos semânticos completamente díspares entre si.

Em uma terceira avaliação, observa-se o comportamento temporal dos subgrafos sob a perspectiva da assortatividade de grau. Dessa maneira, pode-se constatar o recorrente padrão disassortativo do subgrafo (*#BBB21*), evidenciando que a maioria das conexões das *hashtags* mais centrais provêm de *hashtags* com grau baixo. Em contraponto, subgrafos menos populosos exibem predominantemente valores nulos de assortatividade, uma consequência direta do tipo de estrutura formada por poucos nós completamente conectados.

5.1.2 Análise Semântica

Diante da possível existência de domínios de conhecimento distintos dentro do mesmo subgrafo principal, é importante realizar uma caracterização semântica para medir a distância semântica entre a *hashtag* mais central (*#BBB21*) e as demais *hashtags* que compõem o subgrafo principal. Contudo, a ausência do caractere espaço, informando o início e o término de uma palavra, dificulta a compreensão do significado de algumas *hashtags*. Uma vez que métodos computacionais tradicionalmente usados na medição da distância semântica dependem da aplicação de técnicas de tokenização¹ ou incorporação de palavras, opta-se pelo cálculo da Distância Normalizada do Google (*NGD*). Sendo uma métrica assimétrica, o resultado de similaridade é consideravelmente impactado pela ordem de pesquisa dos termos. Por essa razão, o conjunto de valores utilizados nas Figuras 5.4(a) e 5.4(b) advém da média de ambas as configurações.

A Figura 5.4(a) retrata a dispersão dos valores da distância semântica entre a *#BBB21* e as demais *hashtags* do mesmo subgrafo em três momentos distintos: após a 1^a, 18^a e 37^a hora de monitoramento. Embora a média e a mediana dos valores da *NGD*, nos três

¹Procedimento de discretização de textos ou frases contíguas (*string*), segmentando-a em *tokens* segundo um caractere delimitador arbitrário.

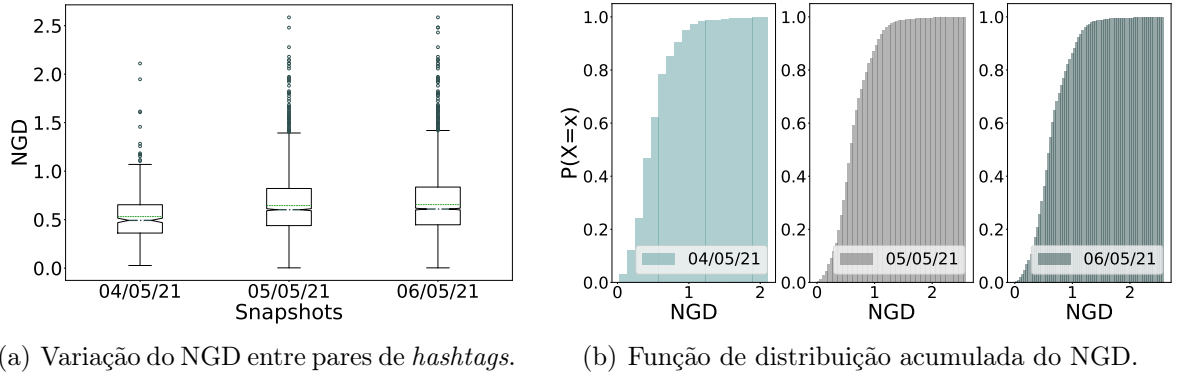


Figura 5.4: Caracterização semântica do subgrafo mais populoso em três momentos da captura.

momentos, circundem próximas a 0.55, percebe-se que nos dois últimos dias de monitoramento há um aumento da ocorrência de *hashtags* com valores de NGD superiores aos demais. Nesse sentido, cogita-se a hipótese de que *hashtags* cujas distâncias semânticas até a *hashtag* central são maiores do que 0.6, possuem uma centralidade menor. Para comprovar a hipótese e consequentemente refutar a hipótese nula, aplica-se o *t*-test de Welch, uma adaptação do *t*-test de Student indicado para conjuntos de amostra com diferentes variâncias ou tamanhos. Com uma significância estatística superior a 95% (valor $p \ll 0.05$), o resultado do teste de hipótese mostra que *hashtags* mais distantes semanticamente da *hashtag* central ($NGD > 0.6$), de fato possuem uma centralidade menor comparado àquelas relacionadas ao tema monitorado. Suportado por essa verificação estatística e correlacionando-a com a Figura 5.4(b), pode-se inferir a percentagem total de *hashtags* com significados suficientemente distantes do tema monitorado. Dependendo do momento observado, constata-se que o grafo relacional de *hashtags* obtém até 78% de coesão semântica.

5.2 Resultados do Método de Concordância de Rótulos

O método de avaliação da concordância proposto foi aplicado aos diferentes exemplares de câmaras de eco descritos na Tabela 4.2, gerando resultados segmentados por tópico de interesse: controle de armas, discriminação de minorias e política. As Figuras 5.5, 5.6 e 5.7 representam, respectivamente, cada um desses tópicos e contém cinco mapas de calor referentes a cada *snapshot* analisado. Entre os algoritmos comparados, estão quatro algoritmos de detecção de comunidade, i.e. Louvain (Lou), InfoMap (Inf), Walktrap (Wkt) e três algoritmos de agrupamento, i.e., K-means (KM), K-medoids (KMed) e Hierárquico

(Hier).

A geração da matéria-prima de análise do método, ou seja, os rótulos aos quais cada nó pertence, segue fluxos de pré-processamento distintos, dependendo da abordagem do algoritmo. Algoritmos de detecção de comunidade são naturalmente aplicáveis para encontrar estruturas homogêneas em grafos, não requerendo manipulação prévia do grafo de entrada. Em contraste, a abordagem baseada em algoritmos de clusterização depende primeiro da extração da representação vetorial do grafo inicial. Ao optar pela matriz de adjacência, é provável que a matriz obtida seja esparsa e de alta dimensionalidade, o que motiva uma redução dimensional. Para isso, utilizamos a técnica de Análise de Componentes Principais (PCA), visando reduzir o custo computacional das etapas do método proposto. Essa redução é orientada pela necessidade de preservar pelo menos 90% da variância das características originais da matriz de adjacência. Assim, a aplicação da PCA sobre cada matriz garante uma diminuição de 10% no número de características.

Os mapas de calor abaixo demonstram o quanto os resultados dos pares de algoritmos diferem ao serem avaliados na mesma câmara de eco. Cada célula do mapa representa a concordância entre um par específico de algoritmos, quantificada pela distância Euclidiana entre as matrizes de autovetores extraídas da matriz de pertencimento de cada algoritmo.

Os mapas de calor do primeiro tópico, i.e., controle de armas, apresentam um comportamento particular dependendo do *snapshot* observado. As Figuras 5.5(a), 5.5(b), 5.5(d), 5.5(e) denotam uma divergência alta entre as interpretações de cada algoritmo na classificação dos nós pertencentes a uma câmara de eco em cada *snapshot*. Contudo, esse comportamento difere do apresentado na Figura 5.5(c). Em particular, o terceiro *snapshot* mostra que os algoritmos Hierárquico e K-means apresentam uma distância Euclidiana nula ($d_{eu} = 0$), o que indica que ambos os algoritmos convergiram totalmente sobre o pertencimento de cada nó em comunidades. Em contrapartida, na Figura 5.6, todos os mapas de calor retratam um mesmo padrão discordante entre todos os algoritmos aplicados aos grafos relacionados ao segundo tópico, e.g., discriminação de minorias. Por fim, ao analisar o terceiro tópico, e.g., política, encontra-se um cenário semelhante ao do primeiro tópico. Enquanto as Figuras 5.7(b), 5.7(d) e 5.7(e) revelam uma discordância total entre os algoritmos, seja de agrupamento ou detecção de comunidade, as Figuras 5.7(a) e 5.7(c) mostram que os algoritmos de agrupamento tendem a apresentar resultados idênticos.

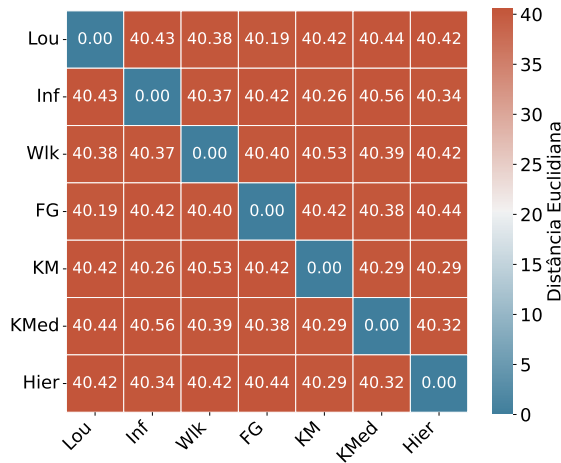
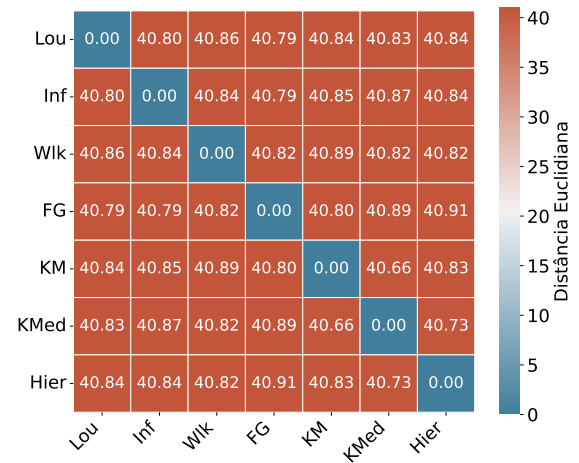
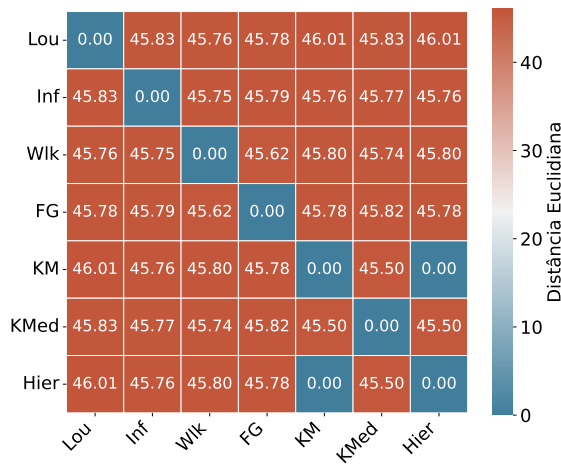
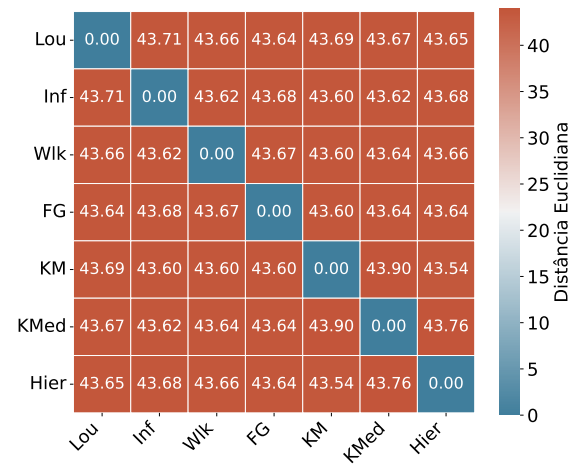
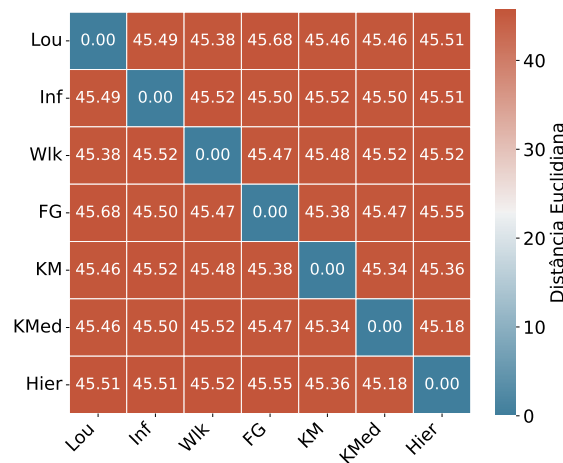
(a) Primeiro *Snapshot*.(b) Segundo *Snapshot*.(c) Terceiro *Snapshot*.(d) Quarto *Snapshot*.(e) Quinto *Snapshot*.

Figura 5.5: Concordância entre diferentes algoritmos na identificação de câmaras de eco relacionadas ao tópico de controle de armas.

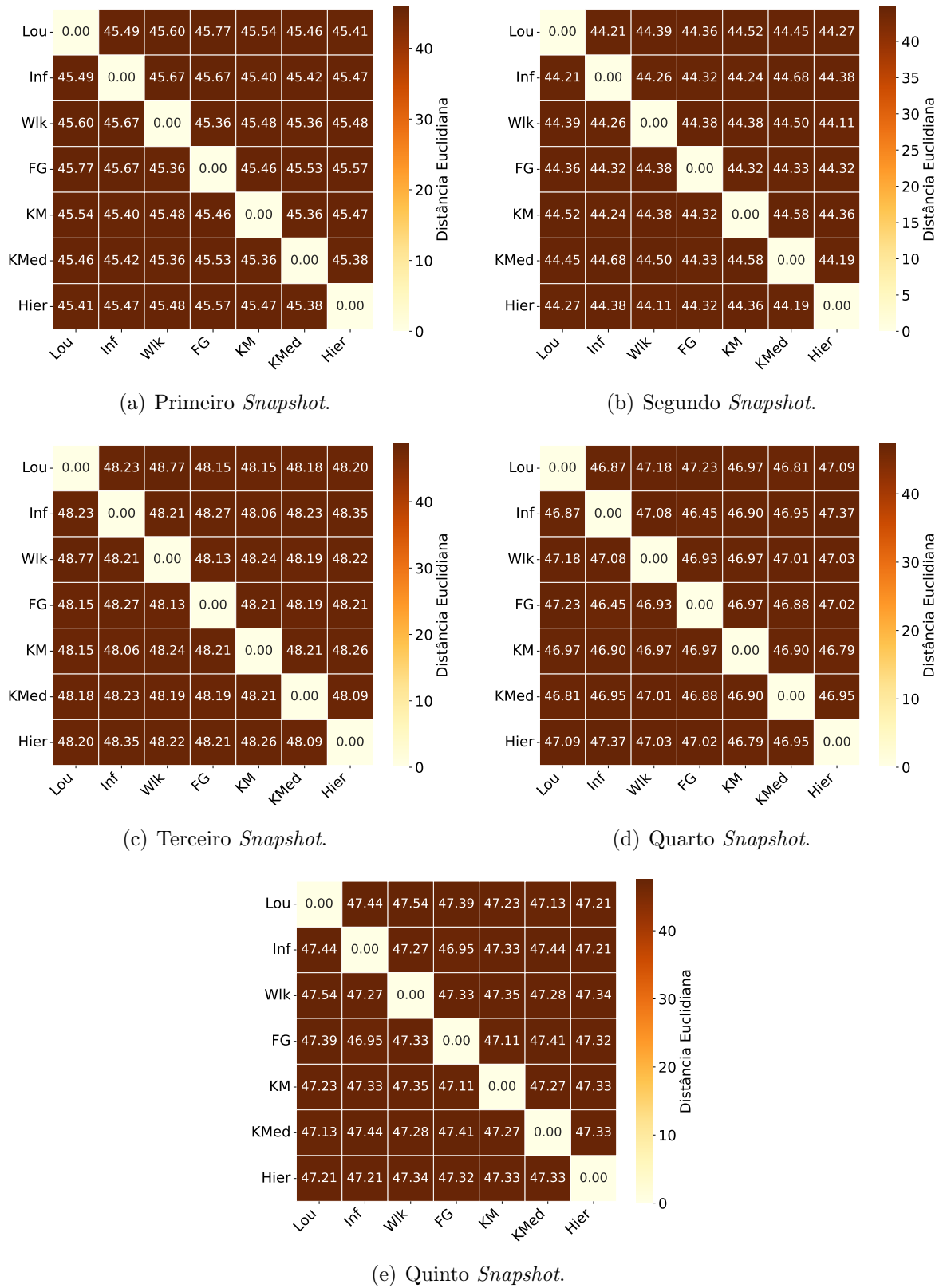


Figura 5.6: Concordância entre diferentes algoritmos na identificação de câmaras de eco relacionadas ao tópico de discriminação de minorias.

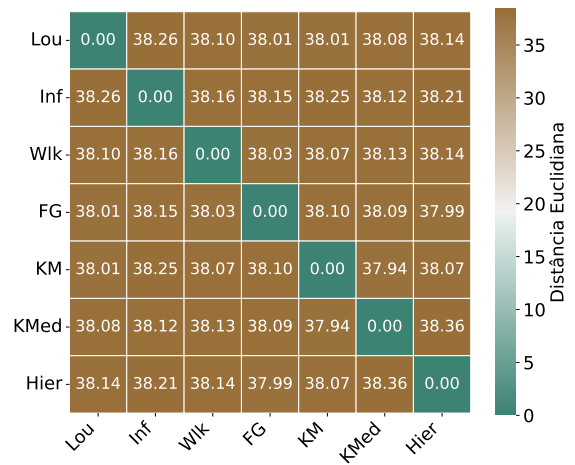
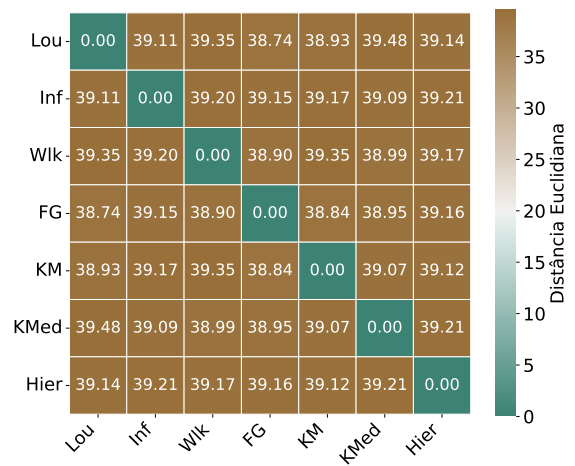
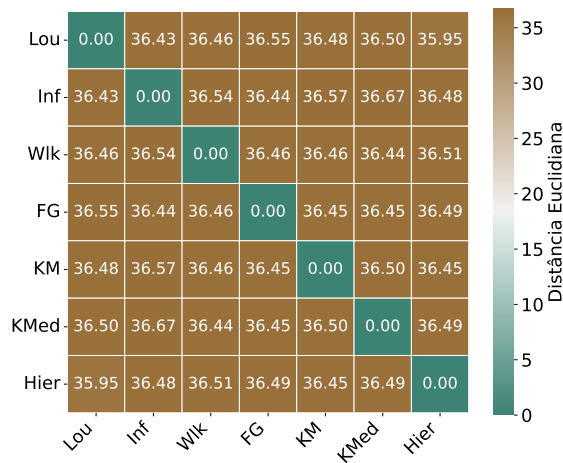
(a) Primeiro *Snapshot*.(b) Segundo *Snapshot*.(c) Terceiro *Snapshot*.(d) Quarto *Snapshot*.(e) Quinto *Snapshot*.

Figura 5.7: Concordância entre diferentes algoritmos na identificação de câmaras de eco relacionadas ao tópico de política.

5.3 Resultados da Análise da Propagação

O primeiro conjunto de simulações visa observar a disseminação de informação sem mecanismos de contenção e sob uma perspectiva dinâmica, ou seja, ao longo da evolução das câmaras de eco apresentadas na Seção 4.5. Para tanto, emprega-se o modelo *Kertesz Threshold*, que possibilita que a modelagem da propagação de informação considere os múltiplos *snapshots* relacionados a um tema específico. Dentre os parâmetros variados na simulação estão o critério de escolha dos nós iniciais (V_0) e o parâmetro *Threshold* (Ths), que define o limiar de adesão de cada nó. Esse limiar estabelece a porcentagem mínima de vizinhos que já aderiram, necessária para que um nó também adira à informação propagada. A quantidade de nós iniciais foi mantida em 5% do total de nós da rede para aproximar a modelagem a um cenário real recorrente em redes sociais, onde poucos usuários são responsáveis pela geração dos fluxos iniciais.

A Figura 5.8(a) mostra o comportamento ascendente da propagação iniciada pelos nós aleatoriamente selecionados (V_0^A) ao longo da evolução das diferentes câmaras de eco. Esse comportamento decorre do baixo limiar configurado ($Ths = 0.05$) por todos os nós do grafo, onde apenas 5% de seus vizinhos precisam ter aderido previamente. Conforme visto nas Figuras 5.8(b) e 5.8(c), o aumento desse limiar impacta diretamente a propaga-

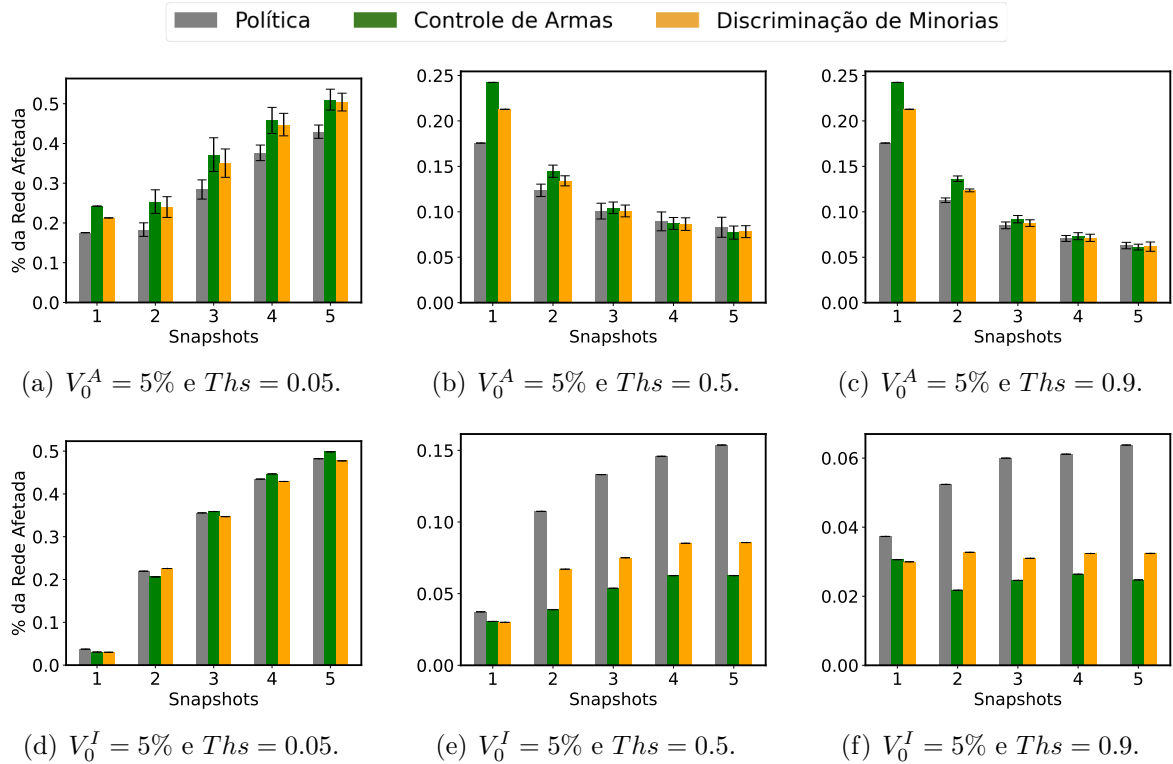


Figura 5.8: Modelo *Kertesz Threshold* aplicado a múltiplos *snapshots* simultaneamente.

ção, refletindo tanto em um comportamento descendente quanto em uma diminuição pela metade no percentual da rede afetada ao longo dos *snapshots*. A Figura 5.8(d) apresenta o mesmo padrão ascendente observado em 5.8(a). Contudo, a seleção de nós iniciais com base na centralidade de informação (V_0^I) tornou a propagação mais assertiva, reduzindo as variações na porcentagem de rede afetada após múltiplas execuções do modelo. Nas Figuras 5.8(e) e 5.8(f), o potencial disseminatório da escolha de nós iniciais com base na métrica de centralidade é contrabalançado pelos limiares relativamente altos necessários para a adesão entre vizinhos ($Ths = 0.5$ e $Ths = 0.9$). Dessa forma, o alcance da propagação é consideravelmente reduzido, atingindo no máximo 6 e 15% da rede, dependendo do limiar.

5.3.1 Avaliação Intra-metodológica

O segundo conjunto de simulações considera cenários contendo estratégias de mitigação da disseminação de informação em grafos, adotando o modelo *Profile*. Diferentemente do modelo anterior, o modelo *Profile* admite apenas um *snapshot* por execução, permitindo analisar a propagação de forma individualizada e independente dos demais *snapshots*. Ao variar o critério de escolha dos nós iniciais (V_0) e a presença e o critério de escolha dos nós bloqueados (V_B), foram originados quatro cenários distintos:

- **Cenário 1.** Propagação iniciada por 2,5% dos nós da rede, escolhidos aleatoriamente ($V_0^A = 2,5\%$) e sem bloqueio de nós.
- **Cenário 2.** Propagação iniciada por 2,5% dos nós mais centrais da rede, escolhidos com base na centralidade de informação ($V_0^I = 2,5\%$) e sem bloqueio de nós.
- **Cenário 3.** Propagação iniciada por 2,5% dos nós mais centrais da rede, escolhidos com base na centralidade de informação ($V_0^I = 2,5\%$), e com bloqueio de 5% dos nós da rede, selecionados aleatoriamente ($V_B^A = 5\%$).
- **Cenário 4.** Propagação iniciada por 2,5% dos nós mais centrais da rede, escolhidos com base na centralidade de informação ($V_0^I = 2,5\%$), e com bloqueio de 5% dos nós da rede, selecionados estrategicamente pelo ECO-SA ($V_B^S = 5\%$).

Cada cenário foi testado em todos os *snapshots* (S_1 a S_5) de cada tópico relacionado às câmaras de eco. Além disso, todos os cenários foram avaliados considerando dois perfis de usuários: aqueles com alta resistência à adesão de conteúdos compartilhados por vizinhos e aqueles com média resistência à adesão. Esses perfis foram inseridos no modelo por meio

do parâmetro inicial *Profile* (Pfl), com valores de 0.9 e 0.5, respectivamente. Os efeitos dos cenários na modelagem computacional da propagação de informações dentro das câmaras de eco são observados de forma complementar nas Figuras 5.9 e 5.10. As Figuras 5.9(a)–(l) exibem a porcentagem da rede afetada ao longo das 100 iterações do modelo, de modo que as curvas referentes a cada *snapshot* representam as médias extraídas após 30 execuções. Paralelamente, as Figuras 5.10(a)–(l) complementam a análise do processo disseminatório, detalhando a variação do ponto de estabilidade de cada curva após as múltiplas execuções.

Analisando as Figuras 5.9(a)–(c), percebe-se que no Cenário 1 há uma sobreposição de todas as curvas de propagação relativas ao mesmo perfil de usuário, independentemente do tópico analisado. Contudo, é notório e esperado que as curvas relativas ao perfil com média resistência ($Pfl = 0.5$) ascendam mais rapidamente que as do perfil com alta resistência ($Pfl = 0.9$). Tal comportamento é corroborado pelas Figuras 5.10(a)–(c), que mostram que as curvas do perfil com média resistência estabilizam entre 10 e 20 iterações aproximadamente, enquanto que as curvas relativas ao perfil com alta resistência demoram no mínimo 60 iterações até a estabilização. Vale destacar que as curvas, especialmente as verdes, exibem um leve padrão sigmoidal, demonstrando que a aleatoriedade na escolha dos nós iniciais atrasou a propagação na câmara de eco. No Cenário 2, observa-se um comportamento muito semelhante ao cenário anterior. No entanto, a seleção dos nós iniciais com base na centralidade de informação resultou na perda da característica sigmoidal das curvas. No Cenário 3, percebe-se que a escolha aleatória dos nós bloqueados limita a propagação a, no máximo, 85% do total de nós da rede. Considerando que o percentual máximo de adesão num cenário sem nós bloqueados é de 95%, o impacto desse bloqueio aleatório foi de aproximadamente 10%.

Por fim, as Figuras 5.9(j)–(l) exibem os efeitos na propagação ao adotar o Cenário 4, onde a seleção do grupo de nós bloqueados é orientada pela estratégia de mitigação proposta. Mesmo assim, percebe-se que todas as curvas, independentemente do perfil, foram reduzidas a no máximo 65% da rede. Isto corresponde a uma queda de aproximadamente 30% na porcentagem da rede afetada em comparação com os cenários anteriores. Ademais, é perceptível que a inclinação das curvas, sobretudo as verdes, foi suavizada, corroborando a percepção de redução do efeito em cascata na propagação entre os nós. A Figura 5.9(l) expressa um resultado similar, contudo a redução na propagação foi mais variável, alternando entre 15% a 35% de redução na disseminação dentro de câmaras de eco.

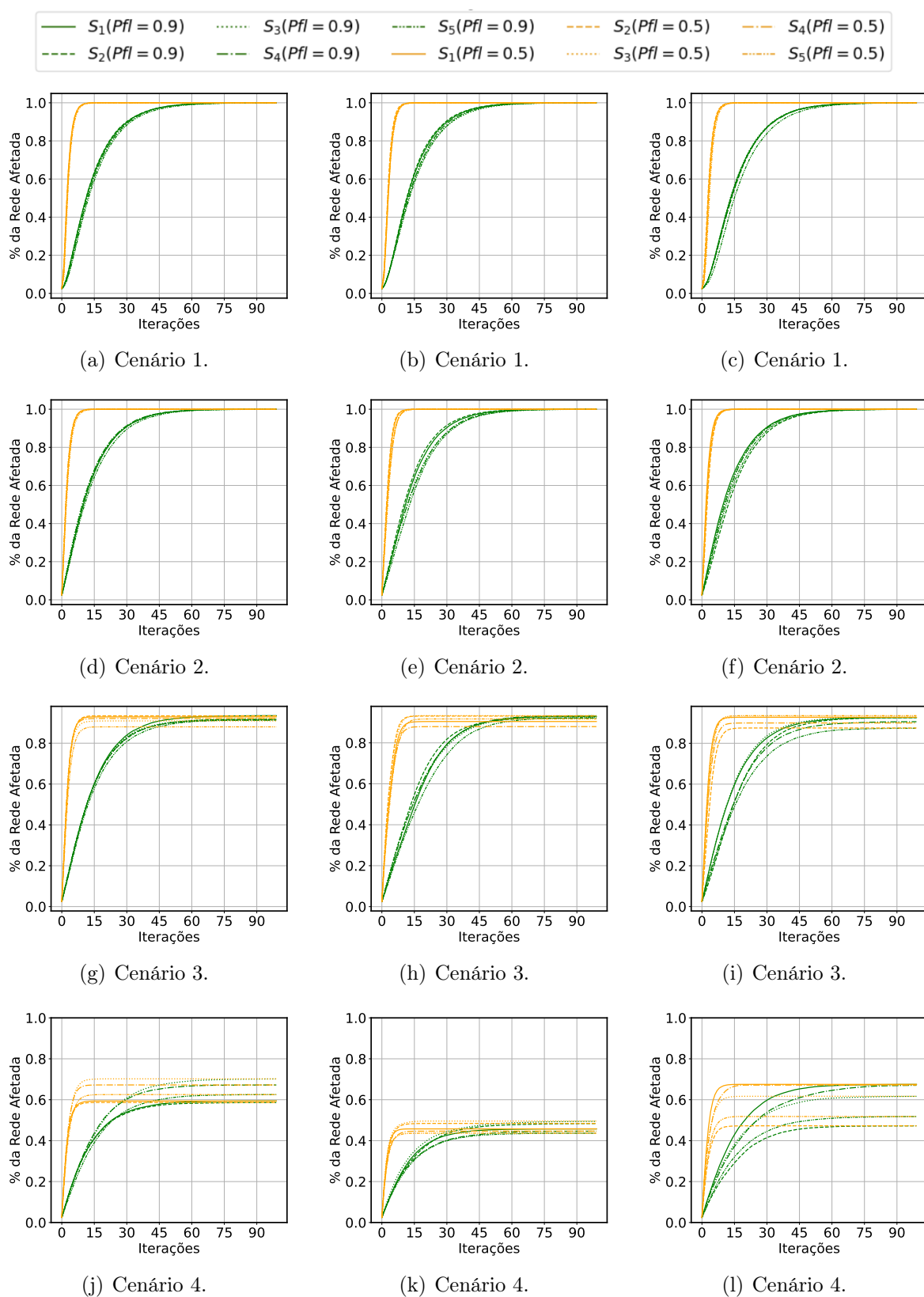


Figura 5.9: Disseminação em câmaras de eco pelo modelo *Profile* com e sem estratégias de contenção. As colunas de gráficos, da esquerda para a direita, representam os temas: controle de armas, discriminação de minorias e esfera política, respectivamente.

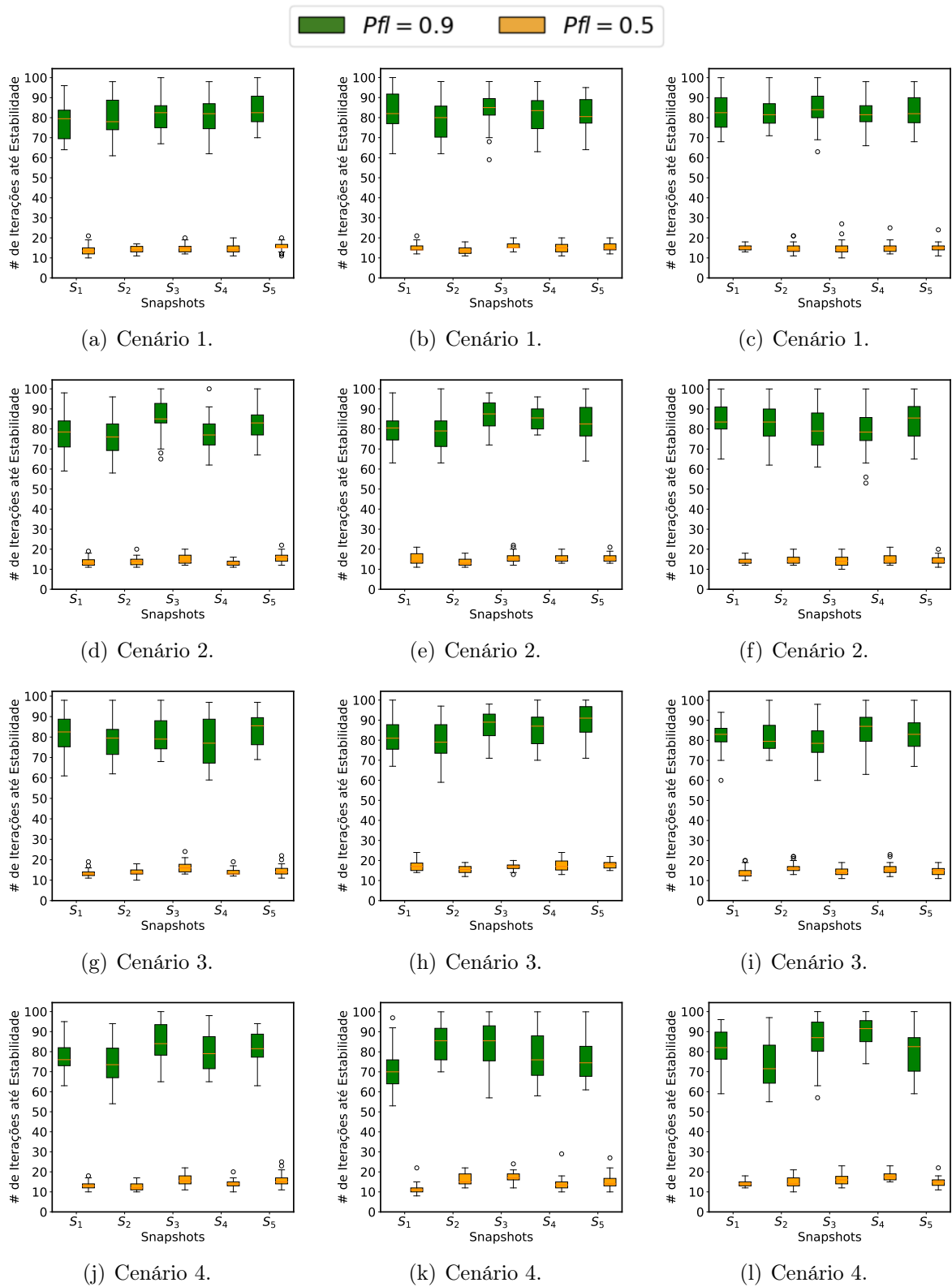


Figura 5.10: Variação da estabilidade da propagação medida em iterações em diferentes cenários. As colunas de gráficos, da esquerda para a direita, representam os temas: controle de armas, discriminação de minorias e esfera política, respectivamente.

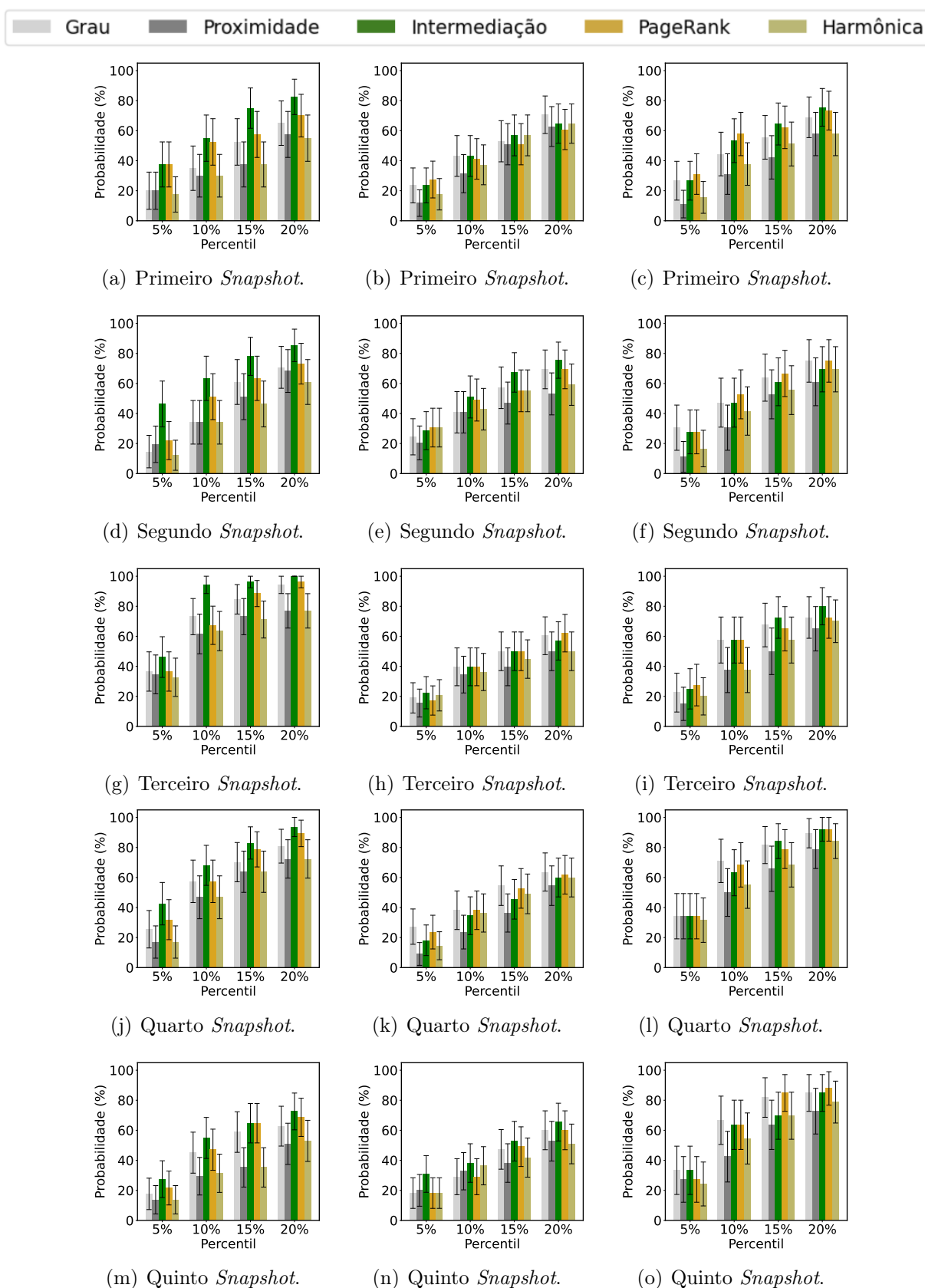


Figura 5.11: Trivialidade do conjunto de nós determinados pelo ECO-SA em vários percentis de centralidade. As colunas de gráficos, da esquerda para a direita, representam os temas: controle de armas, discriminação de minorias e esfera política, respectivamente.

As Figuras 5.11(a)–(o) revelam o nível de trivialidade da solução nodal alcançada pela ECO-SA, com base na probabilidade de encontrar o mesmo conjunto de nós em diferentes percentis de centralidade. Nessa verificação envolvem-se as centralidades de grau, proximidade, intermediação, PageRank e harmônica, observando quatro faixas de percentis, i.e., 5%, 10%, 15% e 20%. Analisando primeiramente a faixa de 5%, mesma porcentagem de nós bloqueados adotada nos Cenários 3 e 4, constata-se uma baixa probabilidade de encontrar o mesmo grupo de nós descoberto pelo ECO-SA. Independentemente do *snapshot* ou tema observados, essa probabilidade não ultrapassa os aproximadamente 47% atingidos pela centralidade de intermediação na Figura 5.11(d). Tal probabilidade é naturalmente maior nas faixas de 10% e 15%, alcançando aproximadamente 90% e 95% dos nós pela centralidade de intermediação na Figura 5.11(g), respectivamente. Embora seja um casamento alto, este somente ocorre no terceiro *snapshot* do tópico de controle armas, enquanto que nos demais *snapshots* o casamento alcança no máximo os 70% e 80% atindigos nas Figuras 5.11(o) e 5.11(l). Vale destacar que, mesmo na faixa de 20%, somente na Figura 5.11(i) identificou-se um tipo de centralidade que atingiu a probabilidade máxima, ou seja, 100%. Nas demais figuras, poucas centralidades obtiveram um casamento maior que 80% com a solução encontrada. Tais resultados ressaltam a especificidade da solução de minimização gerada pela estratégia proposta.

5.3.2 Avaliação Inter-metodológica

O terceiro grupo de simulações atua na avaliação do desempenho do ECO-SA perante trabalhos prévios na tarefa de conter a propagação no interior de câmaras de eco. Diante da ausência de estratégias ou métodos específicos para esse cenário na literatura, optou-se por comparar a proposta com trabalhos dedicados ao controle da difusão de informações em redes sociais. Dentre os trabalhos usados como base comparativa, estão: Holme *et al.* [75], Dey *et al.* [76], Tong *et al.* [80] e Kimura *et al.* [78]. Tais trabalhos foram apresentados na Seção 3.3 e propõem estratégias distintas para minimizar a contaminação, seja pelo bloqueio de nós ou de arestas. Para compará-los sob as mesmas perspectivas, os dois últimos foram minimamente adaptados para o cenário de identificação de nós críticos. Em Kimura *et al.*, essa adaptação foi realizada por meio da percolação de nós em vez de arestas. Assim, ao executar o processo de percolação 500 vezes ($M = 500$) e calcular o grau de contaminação pelo pior caso, identifica-se um grupo de nós estratégicos. Em Tong *et al.*, modificou-se o parâmetro analisado, fazendo com que o novo objetivo seja localizar o conjunto de nós do grafo cuja remoção minimize os autovalores da matriz de adjacência da rede.

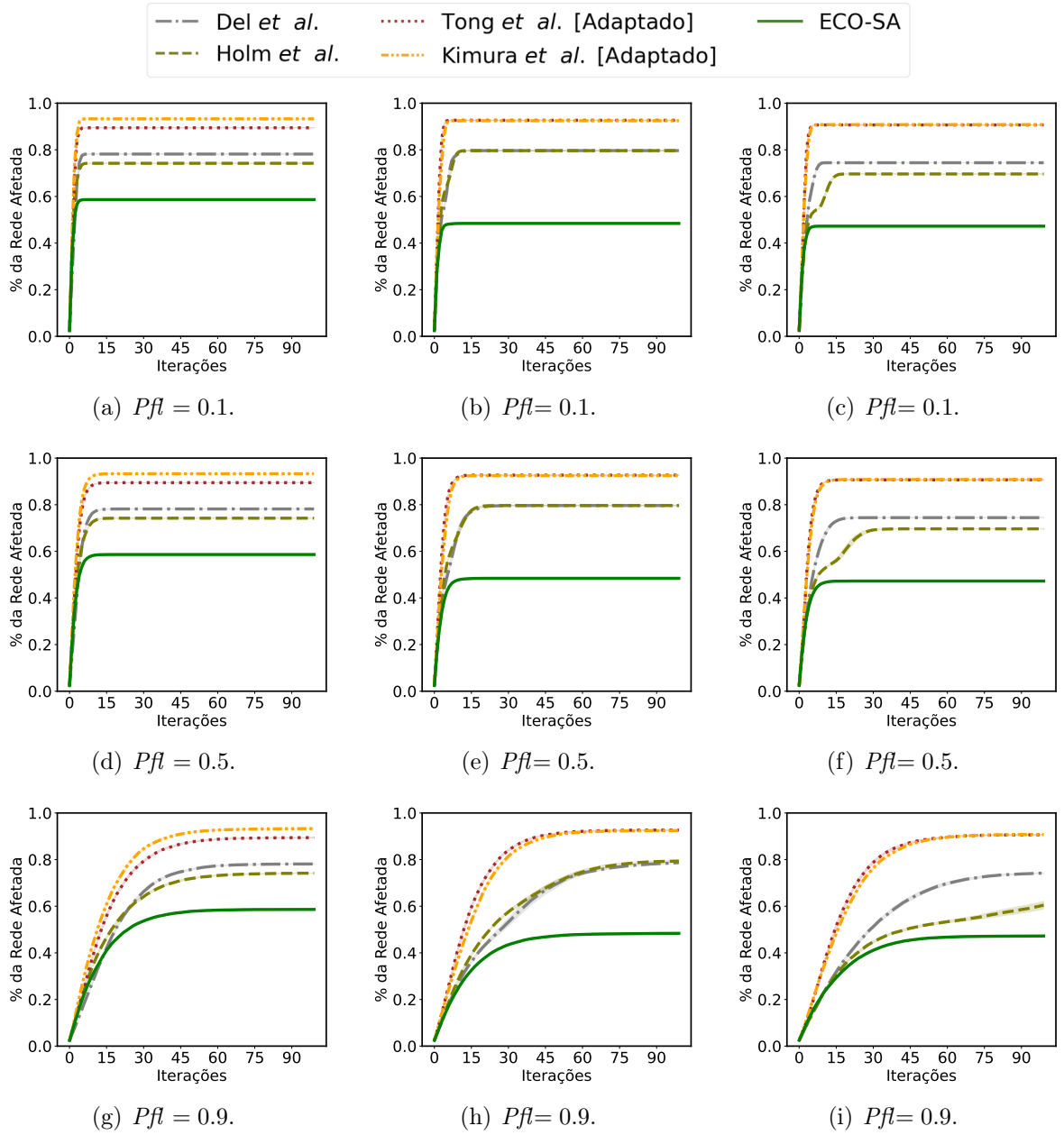


Figura 5.12: Comparação das estratégias de contenção da propagação de informação dentro de câmaras de eco utilizando o modelo *Profile*. As colunas de gráficos, da esquerda para a direita, representam os temas: controle de armas, discriminação de minorias e esfera política, respectivamente. Independentemente do cenário testado, constata-se que a ECO-SA supera as demais propostas, limitando a disseminação entre os nós a aproximadamente 45% a 60% do total.

Adotando o modelo *Profile*, as simulações empregaram o segundo *snapshot* (S_2) de cada tema relacionada a câmaras de eco. Cada simulação foi executada sob as mesmas condições iniciais: (i) propagação iniciada por 2,5% dos nós mais centrais da rede, escolhidos com base na centralidade de informação ($V_0^I = 2,5\%$); (ii) com bloqueio de 5% dos nós da rede, selecionados por cada estratégia comparada. O parâmetro *profile* (Pfl)

foi alternado entre 0.1, 0.5 e 0.9, refletindo perfis de usuários com baixa, média e alta resistência à adesão de conteúdos compartilhados por vizinhos.

A Figura 5.12 ilustra comparativamente o impacto percentual na propagação modelada pelo modelo *Profile*, utilizando os nós bloqueados derivados de cada estratégia. Cada curva representa a porcentagem acumulada da rede afetada ao longo de 100 iterações do modelo, calculada com base na média de 30 execuções. Independente do perfil ou tópico analisado, constata-se uma superioridade da estratégia otimizada proposta em relação às demais. O maior desempenho na contenção é refletido tanto de maneira evolutiva, estabilizando mais rapidamente a propagação entre nós, quanto nominal, reduzindo drasticamente a porcentagem de nós afetados. Conforme observado nas Figuras 5.12(a) e 5.12(f), o bloqueio dos nós estratégicos identificados pela ECO-SA limitou a disseminação a aproximadamente 45% a 60% do total de nós, enquanto que o segundo melhor método restringiu a disseminação a cerca de 65% a 75%, respetivamente.

Capítulo 6

Conclusão

Esta tese de doutorado aprofundou-se na investigação das câmaras de eco (*echo chamber*), um fenômeno relacionado às estruturas sociais homogêneas cujos membros excluem sistematicamente opiniões, crenças e fontes de informação que divirjam daquelas disseminadas entre eles. Para mitigar equívocos semânticos a respeito do tema, busca-se inicialmente descrever as definições e características das câmaras de eco e dos principais termos correlatos.

A partir dessa base conceitual, esta tese propôs três propostas complementares para atuar no monitoramento, análise e contenção de câmaras de eco no cenário virtual. A primeira proposta abrange uma metodologia de monitoramento, captura e caracterização do fluxo de *hashtags* compartilhadas por usuários na plataforma *Twitter*, capaz de fornecer indícios sobre o surgimento de câmaras de eco em torno de temas controversos debatidos na plataforma. Internamente, a metodologia inclui etapas desenvolvidas em *Python* que permitem capturar ininterruptamente quaisquer *hashtags* relacionadas ao tópico monitorado e correlacioná-las por meio de uma estrutura de grafo dinâmica e incrementalmente atualizada. Diferentemente de trabalhos relacionados, que privilegiam a coleta de dados históricos, a proposta foca no monitoramento contínuo e futuro de um evento real e com grande engajamento em redes sociais, durante a exibição do mesmo. Esse dinamismo foi alcançado pelo desenvolvimento de um módulo de captura integrado à API de Streaming do *Twitter*. A caracterização temporal desempenhada emprega múltiplas métricas de redes complexas avaliadas sob diferentes perspectivas, tanto no âmbito estrutural, quanto semântico e tipográfico. Avaliando flutuações em métricas como assortatividade, centralidade, diâmetro e densidade das componentes do grafo de *hashtags*, os resultados revelam que os usuários do *Twitter* tendem a reutilizar combinações de *hashtags* em vez de criar novas. As *hashtags* são preferencialmente publicadas adotando tamanho mediano que

varia entre 10 e 17 caracteres. A presença de componentes desconexas durante a captura sugere a existência de *hashtags* oportunistas em *tweets* relacionados ao tema monitorado. Além disso, percebe-se que mesmo com monitoramento direcionado a um evento específico, o grafo pode incluir *hashtags* atreladas a assuntos ou nichos semânticos completamente díspares entre si. Esse descolamento semântico entre *hashtags* pode representar ideias ou crenças diametralmente opostas, um forte indício de usuários em câmaras de eco. No entanto, essa suposição precisa ser fundamentada na análise do comportamento textual dos usuários quando confrontados com ideias contrárias, um aspecto que não foi abordado nesta tese.

A segunda proposta apresenta uma abordagem comparativa chamada Método de Concordância de Rótulos (MCR), projetada para avaliar o grau de similaridade entre as interpretações de algoritmos que identificam comunidades ou agrupamentos homogêneos a partir de resultados categóricos. Ao considerar as comunidades geradas por cada algoritmo como conjuntos de usuários com posicionamentos ou ideias semelhantes, o método permite comparar a detecção de estruturas análogas a câmaras de eco por pares de algoritmos. Na prática, o método é composto por três etapas sequenciais. A primeira, denominada Mapeamento Intra-Algorítmico, realiza um processo iterativo para avaliar se pares de nós pertencem à mesma comunidade, resultando em uma matriz de pertencimento. Em seguida, na etapa de Decomposição Espectral, a matriz de pertencimento é expressa em termos de seus autovalores e autovetores. Por fim, na etapa de Quantificação da Similaridade, são aplicadas métricas para medir a distância ou similaridade entre pares de matrizes de autovetores normalizados, representando a distância entre interpretações. Essa manipulação permite avaliar o grau de concordância entre os resultados de diferentes algoritmos de detecção de comunidades, mesmo quando os rótulos das comunidades geradas são distintos ou desalinhados. O MCR possibilita que conjuntos de rótulos distintos ou descompassados gerados por diferentes algoritmos sejam projetados para uma dimensão comum, facilitando a comparação. Resultados da análise da concordância entre algoritmos de detecção de comunidades e algoritmos de aprendizado não supervisionado na identificação de câmaras de eco revelam uma quase total discordância entre os algoritmos na classificação das comunidades. A exceção foi a presença de uma concordância completa entre os algoritmos de clusterização k-means, k-medoids e hierárquico em alguns dos conjuntos de dados testados.

A terceira proposta, intitulada ECO-SA, apresenta-se como uma estratégia otimizada para conter a propagação de informação em câmaras de eco, utilizando a meta-heurística *Simulated Annealing*. Tradicionalmente utilizada para resolver problemas complexos, essa

meta-heurística combina a exploração aleatória de soluções com um mecanismo de aceitação probabilística para encontrar uma solução ótima. Adaptando-a ao cenário de contenção da difusão de informação em grafos, a ECO-SA visa encontrar o conjunto ótimo de nós críticos em uma câmara de eco que, quando bloqueados, minimizam a contaminação nominal, ou seja, a quantidade total de nós afetados na propagação da informação. Experimentos foram conduzidos segundo três conjuntos de simulações adotando 15 exemplares de câmaras de eco aplicadas a diferentes modelos de difusão de informação sob diferentes cenários. O primeiro conjunto de simulações visou analisar a disseminação de informação sem mecanismos de contenção, sob uma perspectiva dinâmica, ou seja, ao longo da evolução de uma câmara de eco. Para tanto, emprega-se o modelo *Kertesz Threshold*, que possibilita que a modelagem da difusão de informação considere os múltiplos *snapshots* relacionados a um tema específico. Resultados dessa análise mostraram que a pressão externa exercida pela vizinhança é um fator impactante na propagação de informações dentro de câmaras de eco. No segundo conjunto de simulações, é conduzida uma avaliação intrametodológica que compara tanto cenários de disseminação submetidos à estratégia de mitigação proposta quanto aqueles sem ela, utilizando o modelo *Profile*. As simulações mediram a porcentagem acumulada de nós afetados e o número mínimo de iterações até a estabilidade, considerando perfis de usuários com resistência média e baixa à adesão de conteúdos compartilhados por vizinhos. Ao analisar o nível de trivialidade da solução nodal alcançada pela ECO-SA, observa-se a baixa probabilidade de encontrar o mesmo conjunto de nós em diferentes percentis de centralidade. O terceiro e último conjunto de simulações focou na avaliação intermetodológica, comparando a ECO-SA com outros métodos de mitigação da propagação de informação em redes. Utilizando as mesmas condições iniciais, demonstrou-se a superioridade da ECO-SA em relação aos demais métodos. Destaca-se que o bloqueio dos nós estratégicos identificados pela ECO-SA limitou a disseminação a aproximadamente 45% a 60% do total de nós, enquanto o segundo melhor método restringiu a disseminação a cerca de 65% a 75%.

Como trabalhos futuros, pretende-se estender as análises de propagação de informação em câmaras de eco, incorporando outras bases de dados e incluindo modelos de disseminação simultânea de múltiplas opiniões. Outra vertente possível é o desenvolvimento de um algoritmo de detecção de comunidades dedicado à identificação de câmaras de eco, considerando tanto dados topológicos quanto textuais.

Por fim, esta tese de doutorado fomenta a análise crítica e motiva pesquisadores a desenvolver soluções que auxiliem na identificação e prevenção de câmaras de eco em redes sociais, mitigando assim os seus efeitos danosos. É importante destacar que promover a

interação saudável entre usuários de redes sociais *online* é uma responsabilidade conjunta da comunidade científica, formuladores de políticas digitais, administração e da sociedade em geral.

Referências

- [1] OLIVEIRA, N. R. de; MEDEIROS, D. S.; MATTOS, D. M. A sensitive stylistic approach to identify fake news on social networking. *IEEE Signal Processing Letters*, IEEE, v. 27, p. 1250–1254, 2020.
- [2] COTA, W.; FERREIRA, S. C.; PASTOR-SATORRAS, R.; STARNINI, M. Quantifying echo chamber effects in information spreading over political communication networks. *EPJ Data Science*, SpringerOpen, v. 8, n. 1, p. 1–13, 2019.
- [3] OLIVEIRA, N. R. de; MEDEIROS, D. S.; MATTOS, D. M. Caracterização sócio-temporal de conteúdos em redes sociais baseada em processamento em fluxo. In: SBC. *Anais do XXVI Workshop de Gerência e Operação de Redes e Serviços*. [S.l.], 2021. p. 54–67.
- [4] ZOLLO, F.; BESSI, A.; VICARIO, M. D.; SCALA, A.; CALDARELLI, G.; SHEKHTMAN, L.; HAVLIN, S.; QUATTROCIOCCI, W. Debunking in a world of tribes. *PloS one*, Public Library of Science San Francisco, CA USA, v. 12, n. 7, p. e0181821, 2017.
- [5] DONKERS, T.; ZIEGLER, J. The dual echo chamber: Modeling social media polarization for interventional recommending. In: *Fifteenth ACM Conference on Recommender Systems*. New York, NY, USA: Association for Computing Machinery, 2021. (RecSys '21), p. 12–22.
- [6] SASAHARA, K.; CHEN, W.; PENG, H.; CIAMPAGLIA, G. L.; FLAMMINI, A.; MENCZER, F. Social influence and unfollowing accelerate the emergence of echo chambers. *Journal of Computational Social Science*, Springer, v. 4, n. 1, p. 381–402, 2021.
- [7] ALATAWI, F.; CHENG, L.; TAHIR, A.; KARAMI, M.; JIANG, B.; BLACK, T.; LIU, H. A survey on echo chambers on social media: Description, detection and mitigation. *arXiv preprint arXiv:2112.05084*, 2021.
- [8] FLETCHER, R.; ROBERTSON, C. T.; NIELSEN, R. K. How many people live in politically partisan online news echo chambers in different countries? *Journal of Quantitative Description: Digital Media*, v. 1, 2021.
- [9] SHU, K.; SLIVA, A.; WANG, S.; TANG, J.; LIU, H. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, ACM, v. 19, n. 1, p. 22–36, 2017.
- [10] STILO, G.; VELARDI, P. Hashtag sense clustering based on temporal similarity. *Comput. Linguist.*, MIT Press, Cambridge, MA, USA, v. 43, n. 1, p. 181–200, abr. 2017. ISSN 0891-2017.

- [11] OLIVEIRA, N. R. de; MEDEIROS, D. S. V.; MATTOS, D. M. A syntactic-relationship approach to construct well-informative knowledge graphs representation. In: *2020 4th Conference on Cloud and Internet of Things (CIoT)*. [S.l.: s.n.], 2020. p. 75–82.
- [12] JEON, Y.; KIM, B.; XIONG, A.; LEE, D.; HAN, K. Chamberbreaker: Mitigating the echo chamber effect and supporting information hygiene through a gamified inoculation system. *Proceedings of the ACM on Human-Computer Interaction*, ACM New York, NY, USA, v. 5, n. CSCW2, p. 1–26, 2021.
- [13] BAIL, C. A.; ARGYLE, L. P.; BROWN, T. W.; BUMPUS, J. P.; CHEN, H.; HUNZAKER, M. F.; LEE, J.; MANN, M.; MERHOUT, F.; VOLFOVSKY, A. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 115, n. 37, p. 9216–9221, 2018.
- [14] KIM, H.; KIM, H.; JO, K. J.; KIM, J. Starrythoughts: Facilitating diverse opinion exploration on social issues. *Proceedings of the ACM on Human-Computer Interaction*, ACM New York, NY, USA, v. 5, n. CSCW1, p. 1–29, 2021.
- [15] JEON, Y.; KIM, J.; PARK, S.; KO, Y.; RYU, S.; KIM, S.-W.; HAN, K. Hearhere: Mitigating echo chambers in news consumption through an ai-based web system. *arXiv preprint arXiv:2402.18222*, 2024.
- [16] MORINI, V.; POLLACCI, L.; ROSSETTI, G. Toward a standard approach for echo chamber detection: Reddit case study. *Applied Sciences*, v. 11, n. 12, 2021. ISSN 2076-3417.
- [17] VILLA, G.; PASI, G.; VIVIANI, M. Echo chamber detection and analysis. *Social Network Analysis and Mining*, Springer, v. 11, n. 1, p. 1–17, 2021.
- [18] ARGUEDAS, A. R.; ROBERTSON, C.; FLETCHER, R.; NIELSEN, R. Echo chambers, filter bubbles, and polarisation: A literature review. Reuters Institute for the Study of Journalism, 2022.
- [19] GE, Y.; ZHAO, S.; ZHOU, H.; PEI, C.; SUN, F.; OU, W.; ZHANG, Y. Understanding echo chambers in e-commerce recommender systems. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, New York, NY, USA, p. 2261–2270, 2020.
- [20] NGUYEN, C. T. Echo chambers and epistemic bubbles. *Episteme*, Cambridge University Press, v. 17, n. 2, p. 141–161, 2020.
- [21] GARIMELLA, K.; MORALES, G. D. F.; GIONIS, A.; MATHIOUDAKIS, M. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In: *Proceedings of the 2018 World Wide Web Conference - WWW '18*. [S.l.: s.n.], 2018. p. 913–922.
- [22] COLLEONI, E.; ROZZA, A.; ARVIDSSON, A. Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data. *Journal of communication*, Oxford University Press, v. 64, n. 2, p. 317–332, 2014.

- [23] TÖRNBERG, P. Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS one*, Public Library of Science San Francisco, CA USA, v. 13, n. 9, p. e0203958, 2018.
- [24] VICARIO, M. D.; VIVALDO, G.; BESSI, A.; ZOLLO, F.; SCALA, A.; CALDARELLI, G.; QUATTROCIOCCHI, W. Echo chambers: Emotional contagion and group polarization on facebook. *Scientific reports*, Nature Publishing Group, v. 6, n. 1, p. 1–12, 2016.
- [25] BESSI, A. Personality traits and echo chambers on facebook. *Computers in Human Behavior*, Elsevier, v. 65, p. 319–324, 2016.
- [26] OLIVEIRA, N. R. de; PISA, P. S.; LOPEZ, M. A.; MEDEIROS, D. S. V. de; MATOS, D. M. F. Identifying fake news on social networks based on natural language processing: Trends and challenges. *Information*, v. 12, n. 1, 2021. ISSN 2078-2489.
- [27] BAUMANN, F.; LORENZ-SPREEN, P.; SOKOLOV, I. M.; STARNINI, M. Modeling echo chambers and polarization dynamics in social networks. *Phys. Rev. Lett.*, American Physical Society, v. 124, p. 048301, Jan 2020.
- [28] MORALES, G. D. F.; MONTI, C.; STARNINI, M. No echo in the chambers of political interactions on reddit. *Scientific reports*, Nature Publishing Group, v. 11, n. 1, p. 1–12, 2021.
- [29] WILLIAMS, H. T.; MCMURRAY, J. R.; KURZ, T.; LAMBERT, F. H. Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global environmental change*, Elsevier, v. 32, p. 126–138, 2015.
- [30] BARBERÁ, P.; JOST, J. T.; NAGLER, J.; TUCKER, J. A.; BONNEAU, R. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science*, Sage Publications Sage CA: Los Angeles, CA, v. 26, n. 10, p. 1531–1542, 2015.
- [31] BAKSHY, E.; MESSING, S.; ADAMIC, L. A. Exposure to ideologically diverse news and opinion on facebook. *Science*, American Association for the Advancement of Science, v. 348, n. 6239, p. 1130–1132, 2015.
- [32] CONOVER, M.; RATKIEWICZ, J.; FRANCISCO, M.; GONÇALVES, B.; MENCZER, F.; FLAMMINI, A. Political polarization on twitter. In: *Proceedings of the International AAAI Conference on Web and Social Media*. [S.l.: s.n.], 2011. v. 5, p. 89–96.
- [33] GILLANI, N.; YUAN, A.; SAVESKI, M.; VOSOUGHI, S.; ROY, D. Me, my echo chamber, and i: introspection on social media polarization. In: *Proceedings of the 2018 World Wide Web Conference*. [S.l.: s.n.], 2018. p. 823–831.
- [34] CINELLI, M.; MORALES, G. D. F.; GALEAZZI, A.; QUATTROCIOCCHI, W.; STARNINI, M. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 118, n. 9, p. e2023301118, 2021.
- [35] ZANNETTOU, S.; BRADLYN, B.; CRISTOFARO, E. D.; KWAK, H.; SIRIVIANOS, M.; STRINGINI, G.; BLACKBURN, J. What is gab: A bastion of free speech or an alt-right echo chamber. In: *Companion Proceedings of the The Web Conference 2018*. [S.l.: s.n.], 2018. p. 1007–1014.

- [36] YANG, Z.; ALGESHEIMER, R.; TESSONE, C. J. A comparative analysis of community detection algorithms on artificial networks. *Scientific reports*, Nature Publishing Group, v. 6, n. 1, p. 1–18, 2016.
- [37] TERREN, L.; BORGE-BRAVO, R. Echo chambers on social media: a systematic review of the literature. *Review of Communication Research*, v. 9, p. 99–118, 2021.
- [38] ROSSETTI, G.; CAZABET, R. Community discovery in dynamic networks: A survey. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 51, n. 2, feb 2018.
- [39] MEYER-BAESE, A.; SCHMID, V. Chapter 2 - feature selection and extraction. In: MEYER-BAESE, A.; SCHMID, V. (Ed.). *Pattern Recognition and Signal Analysis in Medical Imaging (Second Edition)*. Second edition. Oxford: Academic Press, 2014. p. 21–69. ISBN 978-0-12-409545-8.
- [40] WANDEL, S.; SHI, X.; SUN, X. Complex network metrics: Can deep learning keep up with tailor-made reference algorithms? *IEEE Access*, IEEE, v. 8, p. 68114–68123, 2020.
- [41] PAGE, L.; BRIN, S.; MOTWANI, R.; WINOGRAD, T. *The PageRank citation ranking: Bringing order to the web*. [S.l.], 1999.
- [42] GARIMELLA, K.; MORALES, G. D. F.; GIONIS, A.; MATHIOUDAKIS, M. Quantifying controversy on social media. *Trans. Soc. Comput.*, Association for Computing Machinery, New York, NY, USA, v. 1, n. 1, jan 2018. ISSN 2469-7818.
- [43] ZAREIE, A.; SAKELLARIOU, R. Minimizing the spread of misinformation in online social networks: A survey. *Journal of Network and Computer Applications*, Elsevier, v. 186, p. 103094, 2021.
- [44] RUAN, Z.; INIGUEZ, G.; KARSAI, M.; KERTÉSZ, J. Kinetics of social contagion. *Physical review letters*, APS, v. 115, n. 21, p. 218702, 2015.
- [45] MILLI, L.; ROSSETTI, G.; PEDRESCHI, D.; GIANNOTTI, F. Information diffusion in complex networks: The active/passive conundrum. In: SPRINGER. *Complex Networks & Their Applications VI: Proceedings of Complex Networks 2017 (The Sixth International Conference on Complex Networks and Their Applications)*. [S.l.], 2018. p. 305–313.
- [46] CUI, P.; WANG, X.; PEI, J.; ZHU, W. A survey on network embedding. *IEEE Transactions on Knowledge and Data Engineering*, v. 31, n. 5, p. 833–852, 2019.
- [47] XU, M. Understanding graph embedding methods and their applications. *SIAM Review*, SIAM, v. 63, n. 4, p. 825–853, 2021.
- [48] CLAUSET, A.; NEWMAN, M. E.; MOORE, C. Finding community structure in very large networks. *Physical review E*, APS, v. 70, n. 6, p. 066111, 2004.
- [49] PONS, P.; LATAPY, M. Computing communities in large networks using random walks. In: SPRINGER. *International symposium on computer and information sciences*. [S.l.], 2005. p. 284–293.

- [50] ROSVALL, M.; AXELSSON, D.; BERGSTROM, C. T. The map equation. *The European Physical Journal Special Topics*, Springer, v. 178, n. 1, p. 13–23, 2009.
- [51] XU, R.; WUNSCH, D. Survey of clustering algorithms. *IEEE Transactions on neural networks*, Ieee, v. 16, n. 3, p. 645–678, 2005.
- [52] FAHAD, A.; ALSHATRI, N.; TARI, Z.; ALAMRI, A.; KHALIL, I.; ZOMAYA, A. Y.; FOUFOU, S.; BOURAS, A. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *IEEE transactions on emerging topics in computing*, IEEE, v. 2, n. 3, p. 267–279, 2014.
- [53] GOVENDER, P.; SIVAKUMAR, V. Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmospheric Pollution Research*, Elsevier, v. 11, n. 1, p. 40–56, 2020.
- [54] CURISKIS, S. A.; DRAKE, B.; OSBORN, T. R.; KENNEDY, P. J. An evaluation of document clustering and topic modelling in two online social networks: Twitter and reddit. *Information Processing & Management*, Elsevier, v. 57, n. 2, p. 102034, 2020.
- [55] CILIBRASI, R. L.; VITANYI, P. M. The google similarity distance. *IEEE Transactions on knowledge and data engineering*, IEEE, v. 19, n. 3, p. 370–383, 2007.
- [56] COSSARD, A.; MORALES, G. D. F.; KALIMERI, K.; MEJOVA, Y.; PAOLOTTI, D.; STARNINI, M. Falling into the echo chamber: The italian vaccination debate on twitter. In: *Proceedings of the International AAAI Conference on Web and Social Media*. [S.l.: s.n.], 2020. v. 14, p. 130–140.
- [57] CINUS, F.; MINICI, M.; MONTI, C.; BONCHI, F. The effect of people recommenders on echo chambers and polarization. In: *Proceedings of the International AAAI Conference on Web and Social Media*. [S.l.: s.n.], 2022. v. 16, p. 90–101.
- [58] BESSI, A.; ZOLLO, F.; VICARIO, M. D.; SCALA, A.; CALDARELLI, G.; QUATTROCIOCHI, W. Trend of narratives in the age of misinformation. *PloS one*, Public Library of Science San Francisco, CA USA, v. 10, n. 8, p. e0134641, 2015.
- [59] GARIMELLA, K.; MORALES, G. D. F.; GIONIS, A.; MATHIOUDAKIS, M. Reducing controversy by connecting opposing views. In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. [S.l.: s.n.], 2017. p. 81–90.
- [60] GUERRA, P.; JR, W. M.; CARDIE, C.; KLEINBERG, R. A measure of polarization on social media networksbased on community boundaries. In: *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013*. [S.l.: s.n.], 2013. p. 215–224.
- [61] ALATAWI, F.; SHETH, P.; LIU, H. Quantifying the echo chamber effect: An embedding distance-based approach. In: *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*. [S.l.: s.n.], 2023. p. 38–45.
- [62] BARA, J.; LEV, O.; TURRINI, P. Predicting voting outcomes in the presence of communities, echo chambers and multiple parties. *Artificial Intelligence*, Elsevier, v. 312, p. 103773, 2022.

- [63] STEWART, A. J.; MOSLEH, M.; DIAKONOVA, M.; ARECHAR, A. A.; RAND, D. G.; PLOTKIN, J. B. Information gerrymandering and undemocratic decisions. *Nature*, Nature Publishing Group UK London, v. 573, n. 7772, p. 117–121, 2019.
- [64] CHEN, X.; ZHOU, X.; SELLIS, T.; LI, X. Social event detection with retweeting behavior correlation. *Expert Systems with Applications*, v. 114, p. 516–523, 2018. ISSN 0957-4174.
- [65] GIRIDHAR, P.; WANG, S.; ABDELZAHER, T.; AMIN, T. A.; KAPLAN, L. Social fusion: Integrating twitter and instagram for event monitoring. In: *2017 IEEE International Conference on Autonomic Computing (ICAC)*. [S.l.: s.n.], 2017. p. 1–10.
- [66] HUANG, J.; TANG, Y.; HU, Y.; LI, J.; HU, C. Predicting the active period of popularity evolution: A case study on twitter hashtags. *Information Sciences*, v. 512, p. 315–326, 2020. ISSN 0020-0255.
- [67] ALSINI, A.; DATTA, A.; HUYNH, D. Q. On utilizing communities detected from social networks in hashtag recommendation. *IEEE Transactions on Computational Social Systems*, v. 7, n. 4, p. 971–982, 2020.
- [68] YANG, C.; WANG, X.; JIANG, B. Sentiment enhanced multi-modal hashtag recommendation for micro-videos. *IEEE Access*, v. 8, p. 78252–78264, 2020.
- [69] LIU, J.; HE, Z.; HUANG, Y. Hashtag2vec: Learning hashtag representation with relational hierarchical embedding model. In: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. [S.l.]: International Joint Conferences on Artificial Intelligence Organization, 2018. p. 3456–3462.
- [70] WANG, Y.; LIU, J.; HUANG, Y.; FENG, X. Using hashtag graph-based topic model to connect semantically-related words without co-occurrence in microblogs. *IEEE Transactions on Knowledge and Data Engineering*, v. 28, n. 7, p. 1919–1933, 2016.
- [71] MAO, Q.; LI, X.; LIU, B.; GUO, S.; HAO, P.; LI, J.; WANG, L. Attend and select: A segment selective transformer for microblog hashtag generation. *Knowledge-Based Systems*, v. 254, p. 109581, 2022. ISSN 0950-7051.
- [72] ZHANG, Y. Language in our time: An empirical analysis of hashtags. In: *The World Wide Web Conference*. New York, NY, USA: Association for Computing Machinery, 2019. (WWW '19), p. 2378–2389. ISBN 9781450366748.
- [73] CUI, A.; ZHANG, M.; LIU, Y.; MA, S.; ZHANG, K. Discover breaking events with popular hashtags in twitter. In: *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. New York, NY, USA: Association for Computing Machinery, 2012. (CIKM '12), p. 1794–1798.
- [74] KOWALD, D.; PUJARI, S. C.; LEX, E. Temporal effects on hashtag reuse in twitter: A cognitive-inspired hashtag recommendation approach. In: *Proceedings of the 26th International Conference on World Wide Web*. [S.l.]: International World Wide Web Conferences Steering Committee, 2017. (WWW '17), p. 1401–1410. ISBN 9781450349130.
- [75] HOLME, P.; KIM, B. J.; YOON, C. N.; HAN, S. K. Attack vulnerability of complex networks. *Physical review E, APS*, v. 65, n. 5, p. 056109, 2002.

- [76] DEY, P.; ROY, S. Centrality based information blocking and influence minimization in online social network. In: IEEE. *2017 IEEE international conference on advanced networks and telecommunications systems (ANTS)*. [S.l.], 2017. p. 1–6.
- [77] PHAM, C. V.; THAI, M. T.; DUONG, H. V.; BUI, B. Q.; HOANG, H. X. Maximizing misinformation restriction within time and budget constraints. *Journal of Combinatorial Optimization*, Springer, v. 35, n. 4, p. 1202–1240, 2018.
- [78] KIMURA, M.; SAITO, K.; MOTODA, H. Blocking links to minimize contamination spread in a social network. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, ACM New York, NY, USA, v. 3, n. 2, p. 1–23, 2009.
- [79] KHALIL, E.; DILKINA, B.; SONG, L. Cuttingedge: Influence minimization in networks. In: *Proceedings of Workshop on Frontiers of Network Analysis: Methods, Models, and Applications at NIPS*. [S.l.: s.n.], 2013. p. 1–13.
- [80] TONG, H.; PRAKASH, B. A.; ELIASSI-RAD, T.; FALOUTSOS, M.; FALOUTSOS, C. Gelling, and melting, large graphs by edge manipulation. In: *Proceedings of the 21st ACM international conference on Information and knowledge management*. [S.l.: s.n.], 2012. p. 245–254.
- [81] LAARHOVEN, P. J. M. van; AARTS, E. H. L. Simulated annealing. In: _____. *Simulated Annealing: Theory and Applications*. Dordrecht: Springer Netherlands, 1987. p. 7–15.
- [82] NETO, H. N. C.; DUSPARIC, I.; MATTOS, D. M.; FERNANDE, N. C. FedSa: Accelerating intrusion detection in collaborative environments with federated simulated annealing. In: IEEE. *2022 IEEE 8th International Conference on Network Softwarization (NetSoft)*. [S.l.], 2022. p. 420–428.
- [83] ANDREONI, M.; BARBOSA, G. N. N.; MATTOS, D. M. F. New barriers on 6g networking: An exploratory study on the security, privacy and opportunities for aerial networks. In: *2022 1st International Conference on 6G Networking (6GNet)*. [S.l.: s.n.], 2022. p. 1–6.
- [84] COUTO, R. S.; MATTOS, D. M.; MORAES, I. M.; CAMINHA, P. H. C.; MEDEIROS, D. S. V. de; SOUZA, L. A. C. de; CAMPISTA, M. E. M.; COSTA, L. H. M. K. Gerenciamento e orquestração de serviços em o-ran: Inteligência, tendências e desafios. *Minicursos do XLI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, 2023.
- [85] MATTOS, D. M. F.; DUARTE, O. C. M. B. Xenflow: Seamless migration primitive and quality of service for virtual networks. In: *2014 IEEE Global Communications Conference*. [S.l.: s.n.], 2014. p. 2326–2331.
- [86] KUKLIŃSKI, S.; TOMASZEWSKI, L.; KOŁAKOWSKI, R. On o-ran, mec, son and network slicing integration. In: *2020 IEEE Globecom Workshops (GC Wkshps)*. [S.l.: s.n.], 2020. p. 1–6.
- [87] FOSTER, N.; MCKEOWN, N.; REXFORD, J.; PARULKAR, G.; PETERSON, L.; SUNAY, O. *Using deep programmability to put network owners in control*. [S.l.]: ACM New York, NY, USA, 2020. 82–88 p.

- [88] GÓMEZ, S. G.; RUEDA, J. L.; CHIMENO, A. E. Management of the business slas for services econtracting. In: SPRINGER. *Service Level Agreements for Cloud Computing*. [S.l.], 2011. p. 209–224.
- [89] JACOBS, A. S.; PFITSCHER, R. J.; RIBEIRO, R. H.; FERREIRA, R. A.; GRANVILLE, L. Z.; WILLINGER, W.; RAO, S. G. Hey, lumi! using natural language for {intent-based} network management. In: *2021 USENIX Annual Technical Conference (USENIX ATC 21)*. [S.l.: s.n.], 2021. p. 625–639.
- [90] CLEMM, A.; ZHANI, M. F.; BOUTABA, R. Network management 2030: Operations and control of network 2030 services. *Journal of Network and Systems Management*, Springer, v. 28, n. 4, p. 721–750, 2020.
- [91] BANERJEE, A.; MWANJE, S. S.; CARLE, G. Contradiction management in intent-driven cognitive autonomous ran. In: IEEE. *2022 IFIP Networking Conference (IFIP Networking)*. [S.l.], 2022. p. 1–6.
- [92] KIRAN, M.; POUYOUL, E.; MERCIAN, A.; TIERNEY, B.; GUOK, C.; MONGA, I. Enabling intent to configure scientific networks for high performance demands. *Future Generation Computer Systems*, v. 79, p. 205–214, 2018. ISSN 0167-739X.
- [93] RACE, N.; ECKLEY, I.; PARLIKAD, A.; ROTSOS, C.; WANG, N.; PIECHOCKI, R.; STILES, P.; PAREKH, A.; BURBRIDGE, T.; WILLIS, P. Industry-academia research toward future network intelligence: The NG-CDI prosperity partnership. *IEEE Network*, IEEE, v. 36, n. 1, p. 18–24, 2022.
- [94] OLIVEIRA, N. R. de; PISA, P. S.; LOPEZ, M. A.; MEDEIROS, D. S. V. de; MATTOS, D. M. Identifying fake news on social networks based on natural language processing: trends and challenges. *Information*, v. 12, n. 1, p. 38, 2021.
- [95] OLIVEIRA, N. R. de; MEDEIROS, D. S.; MATTOS, D. M. A syntactic-relationship approach to construct well-informative knowledge graphs representation. In: IEEE. *2020 4th Conference on Cloud and Internet of Things (CIoT)*. [S.l.], 2020. p. 75–82.
- [96] ARNAZ, A.; LIPMAN, J.; ABOLHASAN, M.; HILTUNEN, M. Toward Integrating Intelligence and Programmability in Open Radio Access Networks: A Comprehensive Survey. v. 10, p. 67747–67770, 2022.
- [97] POLESE, M.; BONATI, L.; D'ORO, S.; BASAGNI, S.; MELODIA, T. Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges. p. 1–1, 2023.
- [98] JACOBS, A. S.; PFITSCHER, R. J.; RIBEIRO, R. H.; FERREIRA, R. A.; GRANVILLE, L. Z.; WILLINGER, W.; RAO, S. G. Hey, Lumi! Using Natural Language for Intent-Based Network Management. In: . [S.l.: s.n.], 2021. p. 625–639.
- [99] ZHENG, X.; LEIVADEAS, A.; FALKNER, M. Intent based networking management with conflict detection and policy resolution in an enterprise network. *Computer Networks*, v. 219, p. 109457, 2022. ISSN 1389-1286.
- [100] CLEMM, A.; CIAVAGLIA, L.; GRANVILLE, L. Z.; TANTSURA, J. *Intent-Based Networking - Concepts and Definitions*. [S.l.]: IETF, out. 2022. RFC 9315. (Request for Comments, 9315).

- [101] ZEYDAN, E.; TURK, Y. Recent advances in intent-based networking: A survey. In: IEEE. *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*. [S.l.], 2020. p. 1–5.
- [102] WEI, Y.; PENG, M.; LIU, Y. Intent-based networks for 6g: Insights and challenges. *Digital Communications and Networks*, Elsevier, v. 6, n. 3, p. 270–280, 2020.
- [103] BAKHSHI, T.; GHITA, B. Towards dynamic network policy composition and conflict resolution in software defined networking. In: *2017 International Conference on Information and Communication Technologies (ICICT)*. [S.l.: s.n.], 2017. p. 34–39.
- [104] MARSICO, A.; SANTUARI, M.; SAVI, M.; SIRACUSA, D.; GHAFOR, A.; JUNIQUE, S.; SKOLDSTROM, P. An interactive intent-based negotiation scheme for application-centric networks. In: *2017 IEEE Conference on Network Softwarization (NetSoft)*. [S.l.: s.n.], 2017. p. 1–2.
- [105] MAHTOUT, H.; KIRAN, M.; MERCIAN, A.; MOHAMMED, B. Using machine learning for intent-based provisioning in high-speed science networks. In: *Proceedings of the 3rd International Workshop on Systems and Network Telemetry and Analytics*. New York, NY, USA: Association for Computing Machinery, 2020. (SNTA '20), p. 27–30. ISBN 9781450379809.
- [106] FEMMINELLA, M.; PERGOLES, M.; REALI, G. Simplification of the design, deployment, and testing of 5g vertical services. In: *NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium*. [S.l.: s.n.], 2020. p. 1–7.
- [107] RAFIQ, A.; MEHMOOD, A.; KHAN, T. A.; ABBAS, K.; AFAQ, M.; SONG, W.-C. Intent-based end-to-end network service orchestration system for multi-platforms. *Sustainability*, MDPI, v. 12, n. 7, p. 2782, 2020.
- [108] LEIVADEAS, A.; FALKNER, M. Vnf placement problem: A multi-tenant intent-based networking approach. In: *2021 24th Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*. [S.l.: s.n.], 2021. p. 143–150.
- [109] OTTER, D. W.; MEDINA, J. R.; KALITA, J. K. A survey of the usages of deep learning for natural language processing. *IEEE Transactions on Neural Networks and Learning Systems*, v. 32, n. 2, p. 604–624, 2021.
- [110] HUANG, J.; YANG, C.; KOU, S.; SONG, Y. A brief survey and implementation on ai for intent-driven network. In: *2022 27th Asia Pacific Conference on Communications (APCC)*. [S.l.: s.n.], 2022. p. 413–418.
- [111] JACOBS, A. S.; PFITSCHER, R. J.; FERREIRA, R. A.; GRANVILLE, L. Z. Refining network intents for self-driving networks. In: *Proceedings of the Afternoon Workshop on Self-Driving Networks*. New York, NY, USA: Association for Computing Machinery, 2018. (SelfDN 2018), p. 15–21. ISBN 9781450359146.
- [112] RIFTADI, M.; KUIPERS, F. P4i/o: Intent-based networking with p4. In: *2019 IEEE Conference on Network Softwarization (NetSoft)*. [S.l.: s.n.], 2019. p. 438–443.
- [113] COMER, D.; RASTEGATNIA, A. Osdf: An intent-based software defined network programming framework. In: *2018 IEEE 43rd Conference on Local Computer Networks (LCN)*. [S.l.: s.n.], 2018. p. 527–535.

- [114] ZHENG2021, X.; LEIVADEAS, A. Network assurance in intent-based networking data centers with machine learning techniques. In: *2021 17th International Conference on Network and Service Management (CNSM)*. [S.l.: s.n.], 2021. p. 14–20.
- [115] MATTOS, D. M. F.; DUARTE, O. C. M. B.; PUJOLLE, G. Reverse update: A consistent policy update scheme for software-defined networking. *IEEE Communications Letters*, IEEE, v. 20, n. 5, p. 886–889, 2016.
- [116] MATTOS, D. M. F.; DUARTE, O. C. M. B.; PUJOLLE, G. A resilient distributed controller for software defined networking. In: *2016 IEEE International Conference on Communications (ICC)*. [S.l.: s.n.], 2016. p. 1–6.
- [117] FILHO, R. H. S.; FERREIRA, T. N.; MATTOS, D. M.; MEDEIROS, D. S. An efficient and decentralized fuzzy reinforcement learning bandwidth controller for multitenant data centers. *Journal of Network and Systems Management*, Springer, v. 30, n. 4, p. 53, 2022.

APÊNDICE A – Sistema de Gerenciamento Baseado em Intenção

Este apêndice discorre sobre um sistema de gerenciamento de rede baseado em intenção, desenvolvido paralelamente às propostas relacionadas a câmaras de eco durante o período do doutorado. O sistema resulta de uma investigação sobre a aplicação de processamento de linguagem natural no gerenciamento de redes de acesso via rádio abertas. Como objetivos específicos dessa investigação, destacam-se: (i) caracterizar textualmente intenções de rede, empregando técnicas de processamento de linguagem natural; (ii) avaliar a adequação de modelos tradicionais de rede neural na detecção de perfis de conflito entre intenções. Esta investigação culminou em duas publicações científicas em congresso e revista internacionais, sendo elas:

- Artigo intitulado “*Agile Conflict-Solving Framework for Intent-Based Management of Service Level Agreement*” e publicado na 2nd International Conference on 6G Networking (6GNet) em 2023;
- Artigo intitulado “*Towards Intent-Based Management for Open Radio Access Networks: An Agile Framework for Detecting Service Level Agreement Conflicts*” e publicado na revista Annals of Telecommunications em 2024.

A.1 Gerenciamento Baseado em Intenção para O-RAN

As implantações atuais das Redes de Acesso Rádio (*Radio Access Networks* – RANs) dependem de soluções monolíticas e fechadas, tornando os operadores de rede e provedores de serviços dependentes de soluções proprietárias [83]. Iniciativas de redes abertas, como a Rede de Acesso via Rádio Aberta (*Open Radio Access Network* – Open RAN), têm ganhado destaque e possibilitam interoperabilidade e adicionar novas funcionalidades à

RAN [84]. A iniciativa visa modificar a arquitetura da rede para permitir que os operadores implementem soluções de múltiplos fornecedores para RANs usando componentes não proprietários. Isso é possível devido à desagregação de hardware e software, interfaces abertas e interoperáveis, e virtualização [85]. Além de aumentar a autonomia dos operadores para criar soluções que não dependem de um único fornecedor, o Open RAN também se concentra em adicionar inteligência e permitir a programabilidade da RAN. Definir uma nova arquitetura baseada em padrões abertos é fundamental para alcançar os objetivos do Open RAN. Os múltiplos objetivos e casos de uso do Open RAN abrangem várias áreas, como (i) gestão e otimização de tráfego, englobando orientação de tráfego entre camadas e otimização da Qualidade de Experiência (*Quality of Experience* – QoE); (ii) conectividade e suporte a *handover*, incluindo *handover* proativo para comunicação *Vehicle-to-Everything* (V2X) e alocação de recursos dinâmica baseada em rota de voo para veículos aéreos não tripulados; (iii) eficiência energética e operações, envolvendo economias de energia na RAN, operações de IoT com consciência energética, otimização da RAN entre camadas e automação das operações da RAN; e (iv) tecnologias avançadas visando otimizar a conformação de feixe (*beamforming*) em sistemas de Múltiplas Entradas e Múltiplas Saídas (*Multiple-Input and Multiple-Output* – MIMO) [86].

A Aliança O-RAN foi criada em um esforço de padronização, sendo assim um consórcio internacional composto por mais de 30 operadores e mais de 200 fornecedores de telecomunicações. As especificações da Aliança O-RAN fornecem esquemas detalhados para construir soluções RAN que atendam aos requisitos do Open RAN, tornando a orquestração e o gerenciamento da rede mais eficientes. Esse cenário favorece a criação de redes mais confiáveis, seguras e de alto desempenho, capazes de operar de forma autônoma, governadas por laços de controle fechados e verificáveis [87].

A orquestração da RAN depende da implementação de políticas complexas, tipicamente descritas como objetivos de alto nível ou intenções de negócios nas redes de comunicação móvel atuais. Conforme expresso na Figura 4.5, os objetivos de alto nível são acordados em Acordos de Nível de Serviço (*Service Level Agreements* – SLAs), um contrato formal entre o provedor de serviços e o operador que determina a qualidade e a funcionalidade dos serviços acordados. A partir dos SLAs, é possível determinar os Objetivos de Nível de Serviço (*Service Level Objectives* – SLOs) e os Indicadores-Chave de Desempenho (*Key Performance Indicators* – KPIs) para a medição de desempenho. Os SLOs são definidos como metas de desempenho específicas que o provedor de serviços se compromete a atingir, e podem incluir métricas como tempo de resposta, disponibilidade e taxa de erro [88]. Como indicadores quantificáveis que monitoram e avaliam o

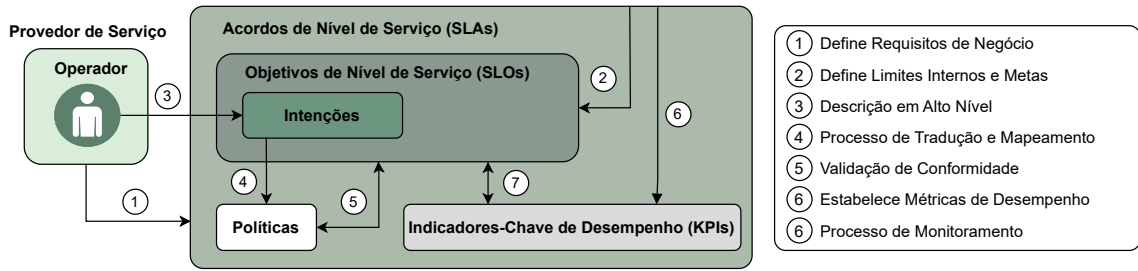


Figura A.1: Visão geral do processo de contratação e gerenciamento de serviços de rede. Os SLAs formalizam objetivos de alto nível acordados entre provedores de serviços e operadores, estabelecendo expectativas claras de qualidade. Os SLOs delineiam limites e objetivos internos específicos, comprometendo o provedor a atender aos padrões. As intenções expressam as expectativas do operador em uma linguagem de alto nível, orientando a derivação de políticas operacionais por meio de um processo de tradução, mapeamento e validação. Os KPIs quantificam o sucesso no cumprimento dos SLOs estabelecidos, permitindo o monitoramento contínuo das operações.

sucesso no cumprimento dos SLOs estabelecidos nos SLAs, os KPIs permitem aos gestores acompanhar o progresso das operações enquanto abstraem os detalhes específicos da gestão e operação da rede. Os operadores realizam a tarefa complexa e propensa a erros de decompor cada objetivo de alto nível em ações de baixo nível a serem implantadas em dispositivos físicos ou virtuais relevantes [89]. As intenções expressam as expectativas do operador em relação à gestão da rede e dos serviços. Uma intenção consiste em um conjunto de SLOs operacionais que a rede deve alcançar e os resultados que a rede deve entregar, definidos de forma declarativa sem especificar como alcançá-los ou implementá-los [90]. Os objetivos e o comportamento esperado da rede podem ser definidos por uma ou várias intenções [91]. Nesse contexto, a tecnologia RAN atual desafia os operadores a orquestrar e traduzir objetivos de negócios de alto nível em políticas e ações de rede de baixo nível.

O Open RAN aproveita o Gerenciamento Baseado em Intenções (*Intent-Based Management* – IBM) para automatizar a configuração e o monitoramento da rede. O IBM envolve a declaração de intenções de alto nível que definem o comportamento da rede de acordo com as especificações dos operadores. Essas especificações são fornecidas através de metas ou KPIs, sem a necessidade de programação explícita da rede para alcançar os SLOs. Quando aplicado à RAN, o IBM atua na simplificação do processo de configuração da rede, desobrigando a necessidade de ajuste fino de parâmetros técnicos e passando a interpretar definições de alto nível. Tal modificação permite que provedores de serviços especifiquem os serviços de conectividade com base em intenções de negócios. A adoção de intenções como instruções de entrada no gerenciamento de redes não envolve direta-

mente a especificação de detalhes como o planejamento de roteamento ou comutação. Em vez disso, a rede baseada em intenções oferece um método simples para articular os requisitos dos usuários. As declarações podem assumir uma forma declarativa ou prescritiva. A forma declarativa é preferida para declarações de intenção, pois oferece uma representação de nível mais alto do problema ou necessidade sem especificar o processo passo a passo. Diferentemente, uma forma prescritiva solicita soluções detalhadas para resolver o problema, especificando os passos exatos, procedimentos ou configurações para atingir um determinado estado ou resultado do sistema. Por exemplo, uma solicitação prescritiva pode incluir instruções como “*Conectar banco de dados ao servidor B com uma rota segura através do switch 28*”. No entanto, solicitações como essa podem culminar em resultados subótimos, com respostas como “*switch 28 indisponível*”. Uma abordagem mais eficaz é a declarativa, permitindo que os usuários definam o contexto do problema, juntamente com quaisquer regras ou restrições, e o resultado desejado. Portanto, a forma declarativa do exemplo acima seria “*Garantir uma conexão segura entre o banco de dados e o servidor B*”. Esses detalhes capacitam as redes a discernir e implementar soluções ótimas para atender às intenções expressas [92].

O uso da linguagem natural para expressar intenções facilita a descrição de metas para indivíduos não técnicos ou aqueles que não estão familiarizados com termos técnicos em um domínio específico. Não obstante, a linguagem natural também é propensa à ambiguidade, tornando desafiador para o sistema capturar a intenção do operador de maneira inequívoca e precisa [93]. Para aproveitar a liberdade trazida pelas descrições em linguagem natural, é necessário extrair fatos e indicadores relevantes das intenções inseridas em sistemas IBM. O Processamento de Linguagem Natural (PLN), também conhecido como linguística computacional, enfrenta esse desafio de forma adequada, extraíndo informações numéricas ou textuais das intenções. Sendo uma subárea da inteligência artificial, o PLN engloba modelos e processos computacionais para resolver problemas práticos relacionados à compreensão e manipulação de linguagens humanas [94]. Expressar intenções diretamente em linguagem natural abstrai as interfaces de gestão através de dispositivos diversos, minimizando o risco de erros humanos durante a tradução manual de políticas em comandos de configuração de equipamentos. Após esse procedimento de processamento textual, as ações necessárias para atingir os objetivos de gerenciamento são inferidas [95]. Todavia, os sistemas IBM não garantem a resiliência da rede, pois podem não abranger todas as possíveis situações que possam surgir.

Diante desse contexto, propõe-se o **AGIR** (***AG**ility in **I**ntent-based management for service level **R**efinement*), um sistema que implementa o gerenciamento de rede baseada

em intenções, capaz de detectar conflitos entre intenções e políticas, mitigá-los e fornecer o nível de serviço acordado para a RAN. O sistema AGIR refina a política de rede de acordo com as intenções extraídas dos SLAs, recebendo intenções e dados de monitoramento da rede e enviando instruções de configuração e otimização para as entidades da rede. O projeto do sistema segue uma arquitetura modular, constituída por quatro módulos para executar o processo: Aplicação Inteligente (iApp), Tradutor, Resolvedor de Conflitos e Agente de Rede. As contribuições deste trabalho são condensadas em quatro aspectos, sendo eles:

- Análise comparativa da literatura científica focada em IBM, destacando as principais técnicas e estratégias adotadas em cada estágio do ciclo IBM;
- Desenvolvimento do sistema AGIR, uma implementação prática de gerenciamento de rede baseado em intenções, projetado para detectar conflitos entre intenções e políticas, mitigá-los e garantir níveis de serviço acordados para a RAN;
- Caracterização abrangente de um conjunto de dados de intenções bem conhecido a partir de uma perspectiva de grafos, empregando várias métricas de redes complexas. Esta análise compara a assortatividade e diferentes variações de centralidade de grupos de rótulos de classe gramatical, extraídos da análise sintática das intenções de rede usando técnicas de PLN;
- Avaliação da aplicabilidade de modelos de redes neurais profundas, *Long Short-Term Memory* (LSTM) e *Gated Recurrent Unit* (GRU), na detecção de intenções conflitantes. O desempenho dos modelos foi comparado segundo métricas de recuperação de informação, i.e., acurácia, precisão, sensibilidade e F1-score, adotando diferentes conjuntos de neurônios;

A.1.1 Arquitetura O-RAN

A arquitetura O-RAN, especificada pela O-RAN Alliance, é mostrada na Figura A.2. A especificação segue a desagregação das três unidades funcionais do gNodeB proposta pela 3GPP, introduzindo assim os seguintes componentes: unidade central O-RAN (O-RAN *Central Unit* – O-CU), unidade distribuída O-RAN (O-RAN *Distributed Unit* – O-DU) e unidade de rádio O-RAN (O-RAN *Radio Unit* – O-RU). A O-RU é um nó lógico que hospeda a camada física inferior (*Low-PHY*) e o processamento de sinal de radiofrequência. A O-DU é um nó lógico que hospeda a camada física superior (*High-PHY*), a subcamada de Controle de Acesso ao Meio (*Medium Access Control* – MAC) e a

subcamada de Controle de Enlace de Rádio (*Radio Link Control* – RLC). A O-CU é um nó lógico responsável pelo controle de mobilidade, compartilhamento da RAN, gerenciamento de sessão e transferência de dados do usuário [96].

A arquitetura O-RAN introduz o conceito de Controladores Inteligentes de RAN (RAN *Intelligent Controllers* – RICs), compreendendo o RIC não tempo-real (*Non-Real-Time* RIC – Non-RT RIC) e o RIC quase tempo-real (*Near-Real-Time* RIC – Near-RT RIC) [97]. Esses controladores oferecem uma visão centralizada da rede, avaliam métricas de desempenho e aplicam algoritmos de aprendizado de máquina para tarefas de otimização automatizadas, como fatiamento de rede, balanceamento de carga e *handovers*. O Near-RT RIC, localizado na borda da rede, opera controles de ciclo fechado com periodicidade entre 10 milissegundos e 1 segundo. O Non-RT RIC complementa o Near-RT RIC para operação inteligente e otimizada da RAN em uma escala de tempo superior a 1 segundo. A arquitetura O-RAN não define o controle em tempo real, ou seja, para uma periodicidade menor que 10 milissegundos. O Near-RT RIC é composto por aplicativos chamados xApps, e os serviços necessários para sua execução. O xApp é um microsserviço que gerencia os recursos de rádio através de interfaces padronizadas e modelos de serviço. O Non-RT RIC suporta aplicativos de terceiros, chamados rApps, que fornecem serviços de valor agregado, facilitando a otimização da RAN e melhorando a qualidade dos serviços fornecidos [87].

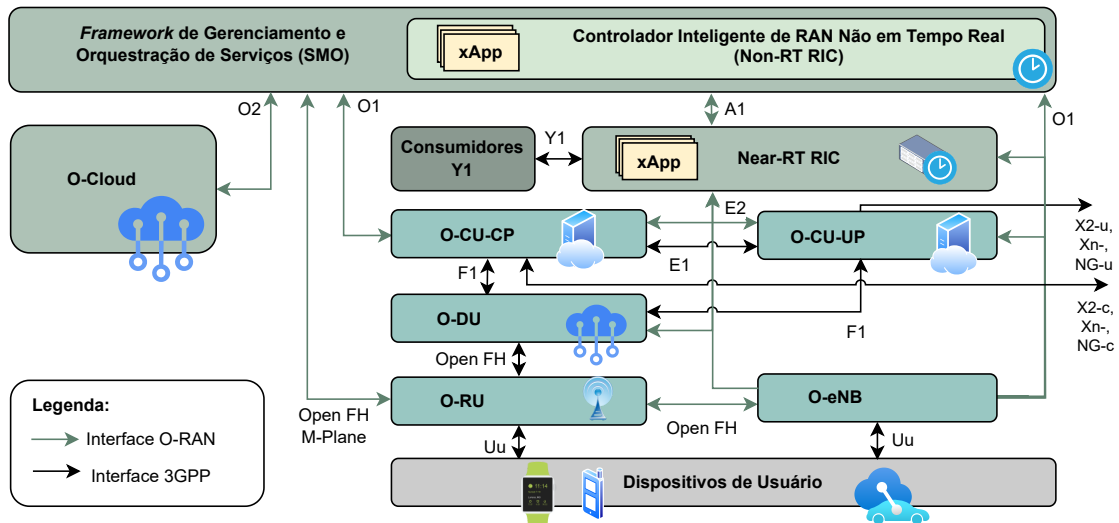


Figura A.2: Arquitetura Open RAN, conforme especificada pela O-RAN Alliance. A O-RAN Alliance padronizou as interfaces abertas para permitir a comunicação entre os componentes Open RAN. As interfaces 3GPP são usadas para comunicação entre componentes herdados de outras gerações de RAN.

O O-Cloud é a infraestrutura de nuvem RAN, composta por componentes de infraestrutura que executam as funcionalidades necessárias, como as Funções de Rede Virtual (*Virtual Network Functions* – VNFs) encontradas no O-CU e os rApps Non-RT RIC [96]. Para adicionar flexibilidade, a arquitetura O-RAN especifica interfaces abertas, como A1, E2, Open Fronthaul (Open FH), Y1, R1, Interface de Notificação O-Cloud, O1 e O2, e também usa interfaces 3GPP, como E1, F1, X2, Xn, NG e Uu. Cada interface interconecta componentes específicos. A interface O1 é usada para comunicação entre o *framework* de Gerenciamento e Orquestração de Serviço (*Service Management and Orchestration* – SMO) e outros componentes da arquitetura O-RAN. Por exemplo, o SMO usa a interface O1 para se comunicar com o Near-RT RIC. A comunicação entre o SMO e o O-Cloud é realizada via interface O2, permitindo suporte para funcionalidades em execução na nuvem. A Interface de Notificação O-Cloud fornece informações críticas da infraestrutura de nuvem para outras funções O-RAN, como o Near-RT RIC, O-CU, O-DU e O-RU. O Non-RT RIC usa a interface A1 para enviar informações para o Near-RT RIC, como dados sobre casos de uso e enriquecimento de informações, e a interface R1 para se comunicar com os xApps. A interface E2 permite a comunicação entre o Near-RT RIC e elementos gerenciados, como O-CU, O-DU e O-eNB. O O-eNB é um eNodeB habilitado para Open-RAN para fornecer serviço LTE. A interface Y1 expõe informações analíticas RAN para os consumidores Y1, que podem ser funções internas ou externas. A interface Open FH permite interação entre O-RU e O-DU. Alguns componentes herdados de gerações anteriores de RAN usam as mesmas interfaces usadas nas arquiteturas dessas gerações. Um exemplo é a interface E1, que conecta os planos de controle e usuário do O-CU. A interface F1 conecta elementos O-CU e O-DU para troca de informações sobre compartilhamento de recursos de rádio e outros status da rede. As interfaces X2 e Xn auxiliam na interoperabilidade entre nós de diferentes gerações, e a interface NG conecta nós 5G à rede principal quando a rede opera em modo autônomo, ou seja, puro 5G [96]. A interface Uu permite a interação entre o Equipamento de Usuário (*User Equipment* – UE) e o O-RU ou O-eNB.

Além da interação facilitada entre os componentes, a orquestração RAN ainda é uma tarefa complexa que depende de políticas derivadas de metas ou intenções de negócios de alto nível. Tais políticas frequentemente descrevem as expectativas do operador em relação à gestão de rede e serviços e os KPIs, permitindo que os gerentes monitorem as operações sem lidar com as complexidades da gestão de rede. No entanto, a orquestração bem-sucedida requer que os operadores decomponham cada política em ações de baixo nível a serem implantadas em dispositivos físicos ou virtuais relevantes [98]. A complexidade e

propensão a erros em redes de comunicação móvel tornam a orquestração RAN desafiadora para os operadores. Com o intuito de simplificar essa orquestração, Open RAN utiliza o gerenciamento baseado em intenção.

A.2 Ciclo de Gerenciamento Baseado em Intenção

O Gerenciamento Baseado em Intenções (IBM) envolve instruções de alto nível que definem o comportamento da rede de acordo com as especificações do operador, expressas por meio de metas comerciais ou KPIs. A característica distintiva do IBM reside em sua capacidade de alcançar SLOs sem programação explícita. Em redes habilitadas para IBM, parâmetros técnicos embutidos nas intenções do usuário são extraídos por meio de um processo de tradução automatizado. Esse processo desmembra as intenções em configurações precisas de nível de dispositivo, garantindo a realização dos recursos de rede desejados pelos clientes. Essa abordagem permite que os operadores evitem configurações propensas a erros, uma ocorrência comum com manipulações tradicionais e demoradas da Interface de Linha de Comando (*Command Line Interface* – CLI) que podem ter repercussões ao longo do tempo e em diferentes espaços de rede [99].

Embora políticas e intenções sejam instrumentos para coordenar o gerenciamento de uma rede baseada em intenções, elas diferem em dois aspectos principais: o nível de granularidade e a ênfase explícita nos resultados [100]. Uma **política** consiste em um conjunto de regras geralmente estruturadas em torno de variações de eventos, condições e ações. Essas regras formam ciclos de controle simples que os dispositivos podem executar sem intervenção de sistemas externos. Consequentemente, as políticas permitem que os usuários estipulem ações sob circunstâncias específicas sem detalhar o resultado desejado. Em contraste, uma **intenção** é uma meta declarativa de negócios de alto nível que opera no nível de rede e serviço, em vez de focar em dispositivos individuais. Ao contrário das políticas, as intenções definem resultados operacionais e objetivos gerais, omitindo os procedimentos ou ações detalhadas necessários para alcançar esses resultados ou satisfazer os objetivos. As intenções permanecem independentes do hardware subjacente, garantindo adaptabilidade entre tecnologias. Além disso, as intenções inerentemente oferecem contexto, tornando-as adequadas para construir implementações de serviços livres de conflitos. Uma característica adicional é que as intenções podem exibir tanto um comportamento persistente, indicando ação contínua ou uma conexão constante, quanto adotar uma natureza transitória, representando ações de curto prazo [101].

Um Sistema Baseado em Intenções (*Intention-Based System* – IBS) geralmente com-

preende cinco etapas funcionais essenciais de processamento, constituindo um sistema de automação de ciclo fechado [100, 99, 102]. A etapa inicial, **Ingestão de Intenções**, permite reconhecer e coletar intenções de rede por meio de interações do usuário. Posteriormente, na etapa de **Tradução de Intenções**, as intenções coletadas são interpretadas e convertidas em um formato compreensível pelos dispositivos e sistemas subjacentes. Essa conversão envolve decompor a intenção original em comandos, parâmetros de configuração ou expressões com uma sintaxe predefinida. Como terceira etapa, a **Resolução de Políticas** valida as intenções em relação às políticas administrativas existentes para detectar e mitigar conflitos, gerando uma nova política operacional para implementação. Esse processo de validação é crucial, pois políticas conflitantes podem desencadear ações destrutivas, resultando em comportamento incorreto da rede. Na etapa de **Orquestração de Políticas**, as políticas são implementadas e transformadas em ações tangíveis na infraestrutura de rede. Essa implementação eficaz, coordenada em toda a rede, garante que a configuração desejada esteja alinhada perfeitamente com as intenções originais dos operadores. Finalmente, a etapa de **Garantia de Políticas** depende dos dados de telemetria de rede para verificar a conformidade contínua de desempenho com os requisitos da intenção ao longo de seu ciclo de vida. A detecção de quaisquer desvios de desempenho promove o refinamento e readaptação da intenção, garantindo a entrega contínua dos níveis de qualidade de serviço desejados. Essa abordagem de ciclo fechado garante a expressão precisa e a tradução da intenção do operador, aliadas ao monitoramento, adaptação e refinamento contínuos para manter o desempenho ideal da rede.

A.2.1 Ingestão e Expressão de Intenções

A ingestão de intenções em um IBS pode apoiar interações não convencionais entre humanos e máquinas, promovendo diálogos ou estruturas interativas para esclarecimentos, explicações e refinamentos. Tradicionalmente, a ingestão de intenções segue duas abordagens: indireta ou direta. Na abordagem de ingestão indireta, as intenções são construídas automaticamente com base na observação e compreensão da rede e precisam apenas ser validadas pelos operadores [103]. Isso é possível aplicando uma análise algorítmica de padrões, vinculada à identificação de necessidades e à geração automática de intenções, oferecendo automaticamente uma alternativa para operadores que preferem delegar a interpretação das intenções à inteligência do sistema, sem exigir especificação direta. Assim, o IBS pode inferir automaticamente intenções relacionadas à otimização da QoS ou à alocação eficiente de recursos, analisando o tráfego de dados e os padrões de uso. Em contraste, a abordagem de ingestão direta permite que os usuários, incluindo operadores,

gerentes ou máquinas inteligentes, tenham controle explícito sobre as intenções inseridas no sistema. Os usuários podem expressar essas intenções tanto em linguagem natural quanto em requisitos de aplicação de alto nível, especificando *endpoints*, largura de banda, latência e disponibilidade [104]. Um método direto comumente usado envolve o uso de assistentes conversacionais, como *chatbots* baseados em Reconhecimento de Entidades Nomeadas (REN) [89, 105] ou em ontologia [92]. Esses *chatbots* permitem que os usuários interajam de forma natural, comunicando suas intenções de rede por meio de conversas. Embora menos interativas, as Interfaces Gráficas de Usuário (*Graphical User Interfaces* – GUI) também são usadas para selecionar intenções predefinidas. Essas intenções podem alterar serviços de rede, níveis de QoS e configurações de segurança, instanciar máquinas virtuais e instalar pacotes de software [106]. Quando adaptadas para usuários experientes, as GUIs fornecem informações detalhadas, como endereços MAC e IP, números de porta e IDs de VLAN [107]. Outra opção de ingestão de intenções é a adoção de formulários baseados em modelos para especificar intenções explicitamente [108]. Este método exige que os usuários articulem suas intenções preenchendo estruturas predefinidas, facilitando a formulação de políticas, mas potencialmente limitando a interpretação devido à sua natureza rígida. Independentemente do tipo, tais interfaces de entrada de intenções fornecem flexibilidade e atendem às preferências e níveis de conhecimento dos usuários.

A.2.2 Tradução de Intenções

A tradução de intenções concentra-se em converter instruções ou solicitações fornecidas pelo usuário em um formato uniforme para análise por máquina. Dependendo da estrutura original da intenção, esse processo de tradução e análise pode ocorrer de duas formas, baseado em processamento textual ou baseado em regras.

Quando expressas em linguagem natural, as intenções de rede exigem processamento inteligente para extrair fatos e indicadores capazes de deduzir as ações necessárias para alcançar objetivos de gerenciamento. Nesse cenário, o Processamento de Linguagem Natural (PLN) é fundamental, pois envolve modelos computacionais e processos para entender e manipular a linguagem natural [109]. Independentemente de sua manifestação em texto ou fala, a linguagem natural refere-se à comunicação humana diária. Essa definição exclui linguagens de programação e notações matemáticas, consideradas linguagens artificiais. Expressar intenções diretamente em linguagem natural oferece várias vantagens, como garantir interfaces de gerenciamento abstratas em diferentes equipamentos, evitar a necessidade de aprender novas linguagens de programação e reduzir erros humanos na

tradução de políticas. No entanto, a evolução constante das linguagens naturais torna cada vez mais desafiador estabelecer regras explícitas para computadores [94]. Essa flexibilidade pode introduzir ambiguidade, desafiando os sistemas a capturar as intenções dos operadores de forma inequívoca. Para lidar com isso, a combinação de PLN com algoritmos de Aprendizado Profundo (*Deep Learning*) aprimora o processo de tradução de intenções, tornando-o mais robusto e adaptável a diversas estruturas linguísticas, independentemente do nível de abstração. Alternativamente, a abordagem de análise baseada em regras destaca-se por sua estrutura predefinida de regras gramaticais, que orientam o processo de interpretação textual. Esta abordagem estabelece padrões específicos para identificar elementos-chave no texto, fornecendo uma análise mais estruturada e direcionada. Sua principal vantagem está na implementação simples e na previsibilidade do processo, uma vez que as regras fornecem um roteiro claro para a interpretação. No entanto, essa abordagem é limitada a uma única forma de entrada [110].

Independentemente da abordagem de tradução escolhida, as intenções processadas podem ser estruturadas em várias sintaxes. Por exemplo, a sintaxe da Linguagem de Intenção de Rede (*Network Intent Language* – NILE) [111] compreende uma linguagem especializada projetada para articular requisitos de rede com precisão e eficiência. Nesse sentido, o NILE atua como uma camada intermediária entre intenções em linguagem natural e políticas de nível inferior [112]. O formato baseado em Linguagem de Marcação Extensível (*eXtensible Markup Language* – XML) emprega XML para codificar intenções de rede. A natureza estruturada e extensível desse formato facilita uma representação clara dos diversos aspectos da rede. O formato baseado na Linguagem de Ontologia da Web (*Web Ontology Language* – OWL) é indicado quando uma estrutura semântica rica de intenções de rede é desejada [92]. Esse formato permite uma representação diferenciada de relacionamentos e restrições dentro da rede. A Notação de Objetos JavaScript (*JavaScript Object Notation* – JSON) pode ser usada como um formato que combina a simplicidade e flexibilidade do JSON na estruturação de cenários de comunicação em rede. Por outro lado, sintaxes personalizadas oferecem um formato de representação exclusivo para acomodar requisitos ou preferências específicas.

A.2.3 Resolução de Políticas

O processo de resolução de políticas vai além de simplesmente mapear intenções processadas para políticas existentes [108], envolvendo também a criação de novas políticas com base nos parâmetros das intenções. Embora opcional, é recomendável incorporar uma etapa de detecção de conflitos em ambos os cenários. Essa etapa visa identificar potenci-

ais inconsistências entre a entrada proposta e as políticas armazenadas no banco de dados ou aquelas atualmente ativas no sistema. Existem duas estratégias principais para detecção de conflitos. A primeira envolve a intervenção de operadores humanos especializados, que é uma estratégia direta, mas potencialmente demorada e menos escalável. A segunda alternativa é a adoção de rotinas comparativas automatizadas. Essas rotinas podem ser baseadas em vários critérios, como verificação de recursos disponíveis [105], comparações temporais-espaciais [99], comparações baseadas em operações lógicas, ou até avaliação do domínio afetado por uma proposta de intenção.

A existência de conflitos pode acionar um procedimento de resolução de conflitos capaz de propor ações de consenso para minimizar intitulados efeitos de desvio (*drift effects*) no comportamento da rede. Esses efeitos de desvio representam alterações indesejadas no estado ou comportamento do sistema em relação às intenções originalmente definidas. Estratégias resilientes são implantadas para resolver esses conflitos, abrangendo (i) priorização de interesses com base em parâmetros específicos; (ii) negociação entre partes interessadas; (iii) adaptação dinâmica de políticas; ou (iv) exclusão da política conflitante [113]. Casos complexos nos quais a resolução de divergências gera mudanças significativas na rede podem ser precedidos pela emissão automática de relatórios. Tais relatórios são projetados para alertar os operadores sobre inconsistências, fornecendo informações detalhadas sobre as políticas ou intenções envolvidas, uma lista de ações sugeridas e suas consequências.

A.2.4 Orquestração, Implantação e Garantia de Políticas

A orquestração de políticas refere-se à coordenação sinérgica de várias políticas, garantindo seu alinhamento e coordenação contínuos antes da implementação efetiva. Esse processo estabelece um plano estratégico que considera a interdependência das políticas, otimizando o comportamento do sistema. Uma vez orquestradas, a implantação emerge, executando efetivamente as políticas validadas no ambiente operacional. A operação direciona a alocação e o agendamento de recursos na rede, exercendo influência direta no comportamento do sistema em tempo real [110].

O estágio de garantia de políticas envolve uma variedade de práticas de monitoramento padrão para acompanhar vigilante o comportamento da rede. Essas práticas incluem (i) monitorar a rede em busca de eventos e discrepâncias de desempenho; (ii) realizar medições para avaliar os níveis de serviço entregues; e (iii) gerar e coletar dados de telemetria [114]. Observando continuamente o comportamento da rede, as funções de garantia detectam desvios de intenção pós-implantação e fornecem indícios valio-

sos para avaliar a saúde geral da rede. Essa abordagem proativa identifica prontamente possíveis problemas, anomalias ou degradação de desempenho. Os dados de telemetria coletados e as métricas de desempenho se tornam entradas essenciais para fases subsequentes de garantia, facilitando uma compreensão holística de quão bem a rede se alinha com o objetivo original pretendido.

A.3 Trabalhos Relacionados à IBM

Na literatura, a pesquisa relacionada à IBM pode ser amplamente categorizada em duas classes principais: (i) trabalhos dedicados à criação de sistemas abrangentes; e (ii) trabalhos focados no desenvolvimento de módulos ou rotinas lógicas adaptadas para fases específicas do ciclo da IBM. Jacobs *et al.* [89] propõem o LUMI, um modelo de linguagem projetado explicitamente para redes de campus universitários, oferecendo aos operadores a capacidade de comunicar suas intenções para a rede usando linguagem natural. O LUMI implanta um *chatbot* e depende de algoritmos Reconhecimento de Entidades Nomeadas (REN) para extrair e rotular com precisão entidades a partir da entrada do operador em linguagem natural. Para melhorar a compreensão contextual e considerar frases em ambas as direções, esquerda para direita e direita para esquerda, o sistema utiliza o técnica *Word2Vec* para incorporação de palavras e emprega o modelo Bi-LSTM (*Bidirectional Long Short-Term Memory*). Os autores avaliam o desempenho do LUMI avaliando sua precisão na extração de informações e capacidades de aprendizado através de *feedbacks* do operador e medindo os tempos de compilação e implantação em várias topologias de rede de campus.

Mahtout *et al.* [105] introduzem o EVIAN, uma plataforma de renderização de intenções que utiliza processamento de linguagem natural para envolver usuários, extrair requisitos de rede através de conversas em inglês e traduzi-los em chamadas de API de rede. Operando em um modelo cliente-servidor-renderizador, o EVIAN capacita os usuários com um controle aprimorado sobre os recursos da rede enquanto mantém políticas de rede. Na prática, o EVIAN compreende três estágios: a admissão de intenção do cliente através de uma interface de *chatbot* inteligente usando o *framework* RASA, a resolução de conflitos e a verificação de políticas tratadas pelo lado do servidor e o mapeamento de entidades de intenção em ações de dispositivos de rede realizadas pelo renderizador. Essas interações ocorrem principalmente por meio de troca de grafos RDF (*Resource Description Framework*). Antes da implantação da ação, o EVIAN pode se comunicar com o engenheiro de rede para solicitar aprovação para a execução.

Kiran *et al.* [92] propõem o iNDIRA, que visa interpretar demandas de QoS de rede e automatizar a tradução para aplicações científicas. O sistema emprega raciocínio de máquina para extrair e interpretar comandos específicos de rede de consultas em sentenças completas. Internamente, as intenções passam por um tratamento textual aprofundado facilitado, que inclui a remoção de palavras desnecessárias e o reconhecimento e eventual substituição de palavras-chave usando um dicionário sinônimo. Durante o processo de provisionamento, o iNDIRA utiliza a extensibilidade de grafos RDF para criar relacionamentos mais complexos, aprimorando a representação e a recuperação de parâmetros de rede e entidades associadas à intenção. No entanto, os autores não detalharam a lógica de detecção e resolução de conflitos dentro do esquema ontológico.

Mattos, Duarte e Pujolle [115] propõem o mecanismo de Atualização Reversa, que introduz uma abordagem prática para atualizar regras em redes definidas por software. O mecanismo envolve a marcação de versões de políticas na rede e a atualização de regras de encaminhamento em sentido inverso ao longo do fluxo estabelecido durante as atualizações de política em andamento. Os autores enfatizam a importância da identificação de conflitos entre políticas, que podem ser parciais ou totais, com base no domínio de aplicação da política. A identificação precisa de conflitos, especialmente conflitos parciais, é crucial, uma vez que podem resultar em regras de encaminhamento contraditórias na rede, exigindo a rejeição de uma das políticas.

Outros trabalhos são dedicados à detecção de conflitos e à resolução de políticas em IBMs. Em particular, Zheng *et al.* [99] propõem duas abordagens algorítmicas para detectar e resolver conflitos entre intenções de rede. A tarefa de detecção de conflitos é abordada aplicando comparações par-a-par entre intenções de rede e posteriormente filtrando regras. Posteriormente, um algoritmo de resolução de políticas incorpora uma análise dimensional de tempo e espaço com base em *endpoints* e períodos de intenção. A avaliação dos algoritmos revela sua capacidade de modificar intenções conflitantes minimizando mudanças globais, fornecendo resumos detalhados de resolução.

Em contraste, Bakhshi e Ghita [103] propõem um *framework* automatizado para criar e validar políticas com base nos perfis de tráfego dos usuários. O *framework* captura e vetoriza inicialmente as tendências de aplicativos dos usuários e as separa usando um algoritmo de aprendizado de máquina não supervisionado. Os perfis de tráfego resultantes da aglutinação são traduzidos em declarações de interesse que o operador de rede pode personalizar. A resolução de conflitos ou sobreposições entre políticas envolve a comparação de declarações de interesse validadas com interesses pré-existentes usando uma função

de *hash*. A avaliação é realizada em dois ambientes acadêmicos, analisando a latência na geração de ações de declaração de rede.

Comer e Rastegarnia [113] introduzem um *framework* de redes definidas por software para programação de rede que simplifica a configuração e o monitoramento por meio de operações de serviço de alto nível. O *framework* incorpora um módulo de gerenciamento de conflitos de política que examina conjuntos de políticas, detecta conflitos e usa um algoritmo de resolução de conflitos para fornecer sugestões de alto nível ao administrador. O algoritmo pode recomendar a fusão, remoção ou substituição por uma união ou interseção de uma política equivalente com base em sua classe: redundância, sobreposição, generalização, correlação e sobreposição. A Tabela A.1 resume as principais características associadas aos trabalhos anteriores relacionados ao refinamento de intenções de rede.

Tabela A.1: Comparação das características dos trabalhos relacionados a IBM.

		Lumi[89]	EVIAN[105]	iNDIRA[92]	OSDF[113]	Leivadeas <i>et al.</i> [108]	Bakhshi and Ghita[103]	Zheng <i>et al.</i> [99]
Categorias	Abordagens e Estratégias Aplicadas							
Ingestão de Intenção	Direta	✓	✓	✓	✓	✓		✓
	Indireta						✓	
Expressão de Intenção	Requisitos de Aplicação em Alto Nível				✓	✓	✓	✓
	Linguagem Natural	✓	✓	✓				✓
Interface de Ingestão de Intenção	<i>Chatbot</i> baseado em NER	✓	✓					
	<i>Chatbot</i> baseado em Ontologia			✓				
	Aplicação Web (GUI)						✓	✓
	Formulário baseado em Modelos					✓		
	Não Aplicado/Detalhado				✓			
Tradução de Intenção/Política	PLN integrado a técnicas de Aprendizado Profundo	✓	✓					
	Estrutura de Ontologia com Raciocínio de Máquina			✓				
	Não Aplicado/Detalhado				✓	✓		✓
Sintaxe de Intenção/Política	NILE	✓						
	Formato baseado em XML			✓		✓		
	Formato baseado em OWL			✓				
	Formato JSON		✓				✓	
	Sintaxe Personalizada				✓			
	Não Especificado							✓
Detecção de Conflitos	Checação baseada no Operador	✓						
	Checação de Disponibilidade de Recursos		✓					
	Comparação Temporal-Espacial Pareada							✓
	Comparação baseada em Operações Lógicas				✓		✓	
	Não Aplicado/Detalhado			✓		✓		
Resolução de Política	Associação com Cadeia de Serviços de <i>Blueprint</i>					✓		
	Comparação Interativa usando Operações Lógicas							✓
	Não Especificado/Aplicado	✓	✓	✓	✓		✓	
Métrica de Avaliação	Latência			✓		✓	✓	✓
	Taxa de Sucesso na Resolução de Conflitos							✓
	Precisão na Extração de Informações	✓						
	Usabilidade da Interface do Usuário	✓						
	Não Aplicado		✓					

Ao contrário dos trabalhos anteriores, o sistema AGIR proposto utiliza a identificação de conflitos para avaliar se as intenções propostas podem ser aceitas ou devem ser rejeitadas devido a conflitos com políticas já implementadas na rede. Um fator crítico no sistema AGIR é o processo de identificação de conflitos, que deve identificar corretamente o domínio afetado por uma proposta de intenção. Para esse fim, caracterizamos as intenções em linguagem natural e propomos um modelo de aprendizado profundo que identifica com precisão os conflitos entre as propostas de intenção.

A.4 Sistema AGIR

O sistema AGIR implementa o gerenciamento de rede baseado em intenções para Open RAN, considerando a arquitetura especificada pela O-RAN Alliance. O sistema segue uma arquitetura modular, conforme mostrado na Figura A.3, e é baseado em NLP para extrair informações das intenções. O **módulo iApp** prevê uma interface gráfica simples para que os usuários da aplicação possam interagir e inserir suas intenções. Além de ser o ponto de contato dos usuários, o iApp também é responsável por encaminhar as intenções recebidas para o módulo Tradutor, receber dados de monitoramento da rede, validar constantemente se as políticas criadas são adequadas para alcançar os objetivos estabelecidos nas intenções e indicar ao módulo Tradutor a necessidade de adaptação nas políticas mapeadas.

O **módulo Tradutor** transforma a intenção expressa em linguagem natural em declarações de configuração. O módulo implementa algoritmos de NLP para traduzir a intenção, convertendo-a em uma política a ser implementada por meio de um controlador de Redes Definidas por Software (*Software-Defined Networks* – SDN) [116]. A tradução da intenção é baseada na análise lexical e morfológica das entidades textuais. Isso permite identificar palavras-chave que compõem a intenção, que podem então ser usadas para gerar um modelo de regras que pode ser transformado em uma política. As palavras-chave da intenção são comparadas com uma base de conhecimento pré-existente na rede para classificá-las de acordo com os rótulos dos elementos-chave no modelo. O modelo é uma expressão da intenção apresentada em um formato de objeto-ação-resultado. A ideia é formular estratégias, que são as ações para os recursos atuais da rede, que são os objetos do modelo, para alcançar o objetivo da intenção, que é o resultado. Ao converter as palavras-chave em declarações estruturadas, a tradução do controlador SDN em políticas a serem executadas torna-se mais fácil.

Internamente, a tradução de intenções pode seguir uma sequência de técnicas de lim-

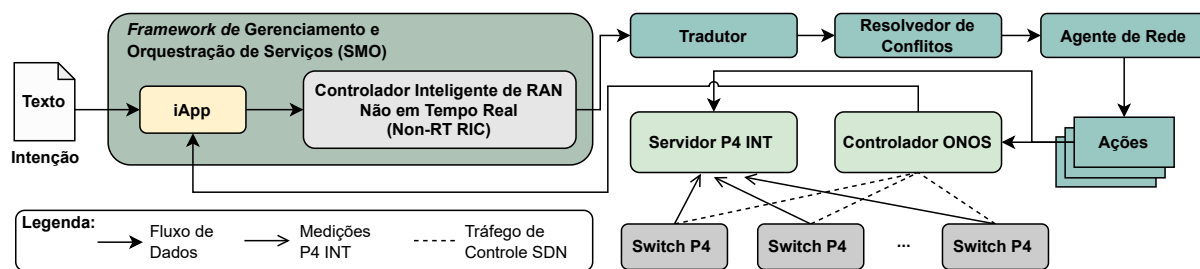


Figura A.3: Diagrama de blocos da arquitetura do sistema AGIR integrado com os elementos da arquitetura especificada pela O-RAN Alliance. O sistema modular captura e processa intenções em linguagem natural através de interações com operadores. Depois de capturadas, as intenções são traduzidas e validadas em relação ao estado atual da rede. Se aprovadas, são convertidas em políticas aplicáveis e executadas nos dispositivos periféricos de telemetria.

peza e formatação de texto. Entre essas técnicas são (i) tokenização, (ii) remoção de pontuação, caracteres especiais e palavras irrelevantes, (iii) correção ortográfica, (iv) reconhecimento de entidades nomeadas específicas e (v) lematização ou radicalização (*stemming*). Seguindo a ordem mencionada, o texto que compõe cada intenção é primeiro submetido à tokenização, um procedimento de discretização. Usando o caractere de espaço como critério delimitador, por exemplo, a tokenização transforma cada sentença contínua de uma intenção em uma lista de *tokens*, permitindo a manipulação individual. Cada *token* é visto como uma sequência de caracteres. Em seguida, características ortográficas como pontuação e caracteres especiais são removidas de cada *token*. Como não contribuem para o entendimento semântico da intenção, palavras irrelevantes, como conjunções, artigos e pronomes, também são removidas do texto. Para lidar com possíveis erros tipográficos, a correção ortográfica é aconselhável, um procedimento realizado comparando cada *token* com seu equivalente mais próximo em um dicionário personalizado. O Reconhecimento de Entidades Nomeadas (REN) visa identificar nomes de software, hardware ou quaisquer nomes próprios relacionados à intenção. Finalmente, a adoção de lematização ou radicalização é comum para reduzir o processamento desnecessário causado por possíveis redundâncias entre palavras, seja por inflexões ou derivações. Na lematização, o objetivo é eliminar possíveis variantes ou plurais da mesma palavra, reduzindo-as aos mesmos lemas, conhecidos como formas de dicionário. Em contraste, essa redução é feita na radicalização transformando cada palavra em sua raiz. No entanto, a lematização ou a radicalização são etapas opcionais no processamento de intenções, pois as declarações de intenções são um domínio fechado de conhecimento e agnóstico às variações de uma raiz de palavra.

Expandindo o processamento textual para outros estágios linguísticos, algumas técnicas de PLN podem realizar análise morfosintática em diferentes níveis de complexi-

dade. Em um nível fundamental, a rotulagem de classe gramatical (*Part-of-Speech tagging*) é uma técnica de análise morfológica que retorna o rótulo gramatical associado, representando a camada mais baixa da análise linguística. Assim, cada palavra em uma sentença recebe metadados, identificando sua classificação gramatical e atributos de conjugação [94]. A adoção das técnicas mencionadas acima permite converter uma intenção descrita em linguagem natural, que pode ser imprecisa e ambígua, em uma ação interpretável pela rede.

O **módulo Resolvedor de Conflitos** é responsável por detectar e mitigar conflitos e é o foco deste trabalho. As intenções do operador podem entrar em conflito, criando contradições quando os objetivos especificados requerem modificações no mesmo parâmetro, mas em quantidades diferentes [91]. A existência desses conflitos torna a gestão baseada em intenções no RAN difícil. A resolução de conflitos no sistema AGIR é baseada na metodologia de atualizações reversas, um esquema de atualização de políticas para SDN que garante a consistência nos compromissos de políticas. A atualização reversa baseia-se na atualização das regras de processamento e encaminhamento de fluxo na direção reversa do caminho de fluxo já instalado, garantindo que um fluxo sempre alcance a configuração de rede mais atualizada [115]. É importante destacar que a etapa antes da resolução de conflitos é a identificação do conflito. O sistema AGIR utiliza a identificação de conflitos para avaliar se as intenções propostas podem ser aceitas ou devem ser rejeitadas devido a conflitos com políticas já implementadas na rede. Um fator crítico no processo de identificação de conflitos do sistema AGIR é a identificação correta do domínio afetado pela proposta de intenção. Em essência, este módulo aborda a tarefa de identificação de conflitos como um problema de classificação baseado em algoritmos de aprendizado profundo.

A detecção de conflitos é fundamentada na modelagem de uma rede neural composta por camadas de Embedding, LSTM e Densa usando a biblioteca Keras¹. A camada de Embedding mapeia sentenças tokenizadas para vetores densos de tamanho fixo usando um modelo *Word2Vec* pré-treinado. Esta técnica de incorporação de palavras (*word embedding*) captura efetivamente o significado semântico e as relações entre as representações significativas das palavras injetadas na próxima camada LSTM. A segunda camada é composta por um modelo LSTM, um algoritmo capaz de aprender dependências de longo alcance e padrões sequenciais nos dados. Esta capacidade de manter e atualizar estados de célula garante que o LSTM retenha informações relevantes em diferentes intervalos, tornando-o especialmente adequado para tarefas relacionadas a texto. Composta por ape-

¹<https://keras.io/>.

nas um neurônio, a camada Densa adota uma função de ativação sigmoideal. A escolha é justificada pela adequação da função sigmoideal a problemas de classificação binária, permitindo o mapeamento da saída em probabilidades no intervalo $[0, 1]$. O modelo também é treinado usando o otimizador Adam sobre a função de perda de entropia cruzada binária. Na prática, o modelo de rede neural visa analisar e classificar a existência de conflitos semânticos entre sentenças, tornando-o uma ferramenta valiosa para automatizar o processo baseado em intenções escritas em uma linguagem de alto nível.

O **módulo Agente de Rede** comunica as políticas obtidas das intenções traduzidas para o controlador SDN. O controlador ONOS é considerado para o controle da rede, comunicando-se com o Agente de Rede através da interface norte. Os *switches* P4 não são componentes explícitos do sistema AGIR, mas são usados para fornecer informações ao sistema proposto. Esses *switches* são responsáveis pela Telemetria de Rede *In-band* (*Network In-band Telemetry* – INT), fornecendo informações de rede ao controlador e a um servidor de telemetria P4 (INT P4 *Server*). Este servidor envia relatórios de medição ao iApp, permitindo o monitoramento contínuo do desempenho da rede. A otimização do desempenho da rede será realizada usando mecanismos de aprendizado de máquina, como Aprendizado por Reforço Profundo (*Deep Reinforcement Learning* – DRL) ou Aprendizado por Reforço Fuzzy (*Fuzzy Reinforcement Learning* – FRL) [117]. Isso permite validar se os objetivos da intenção estão sendo alcançados e adaptar as políticas, se necessário. Vale ressaltar que o sistema proposto está relacionado ao Non-RT RIC, o que significa que o ciclo de controle que governa o processo dura mais de 1 segundo.

A.5 Resultados e Discussão

Os resultados obtidos foram derivados de duas abordagens: uma baseada numa perspectiva em grafos, que se concentra na caracterização de um conjunto de intenções segundo métricas de redes complexas, e outra baseada numa perspectiva algorítmica, que avalia o desempenho de modelos neurais na identificação de conflitos entre intenções. Ambas as perspectivas foram conduzidas adotando um dos conjuntos de dados sinteticamente construídos por Jacobs *et al.* [89]. Esse conjunto de dados compreende 10 mil pares de intenções de rede, devidamente equilibrados e rotulados como conflitantes ou não conflitantes. Cada intenção, escrita em linguagem natural, contém as ações e parâmetros instruídos pelo operador da rede, abrangendo tarefas como configuração de roteadores, monitoramento de tráfego, diagnóstico de falhas e otimização de desempenho. Considerando as intenções do conjunto de dados, o exemplo abaixo ilustra um cenário onde há divergências relacionadas a parâmetros de QoS.

- **Intenção 1:** Para o grupo (*'dormitórios'*), define cota (*'download', '7', 'GB/semana'*)
- **Intenção 2:** Do endpoint (*'19.16.1.1'*) ao endpoint (*'172.16.1.2'*), define cota (*'qualquer', '1', 'GB/semana'*)

Apesar de aparentarem envolver configurações de cotas, as intenções 1 e 2 possuem características conflitantes. Enquanto a primeira intenção se concentra em um grupo específico (*'dormitórios'*) e limita o download a 7 GB/semana, a segunda trata de uma limitação genérica de fluxo de tráfego entre dois pontos de extremidade sem especificar um grupo. Com base nessas discrepâncias identificadas entre atributos, como alvos, tipos de tráfego e restrições de recursos, algoritmos podem inferir possíveis conflitos de configuração entre novas intenções e as previamente implementadas. Os algoritmos e técnicas descritos neste trabalho são desenvolvidos em *Python* usando um computador pessoal equipado com um processador Intel Core i7 4770, com 8 GB de RAM e 1 TB de armazenamento. Nota-se que os resultados apresentados são valores médios com um intervalo de confiança de 95%.

Na prática, ambas as abordagens são precedidas por pipeline abrangente de pré-processamento de texto contendo diferentes funções da biblioteca NLTK. O objetivo é reduzir a pluralidade de palavras, garantindo que apenas as palavras essenciais e mais informativas sejam ingeridas nas etapas subsequentes. Liderado pela tokenização, o pré-processamento começa convertendo as sentenças contínuas em listas de *tokens* e removendo pontuações e caracteres especiais. No entanto, a existência de endereços IP e unidades de medida impõe a necessidade de detectar e preservar essas estruturas específicas, uma vez que são parâmetros comumente passados em intenções. Essa necessidade é atendida integrando funções de expressões regulares. Uma vez identificadas essas estruturas, a rotulagem gramatical e a filtragem de palavras irrelevantes podem ser aplicadas. Esta ordem de implementação busca mitigar possíveis erros de rotulagem que ocorrem ao analisar sentenças extremamente truncadas, ou seja, sem conectivos, artigos ou pronomes. Dentre as classes gramaticais resultantes desse procedimento de limpeza e conformação textual estão: adjetivos (ADJ), substantivos (SUBS), numerais (NUM), verbos (VRB) e advérbios (ADV).

Na perspectiva em grafos, a caracterização realizada iniciou com estruturação de um grafo não direcionado e ponderado $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, baseado em intenções pré-processadas e convertidas em vetores de *tokens* significativos. Neste grafo, cada nó representa um *token* distinto, e a aresta entre um par de nós contém um peso proporcional à frequência

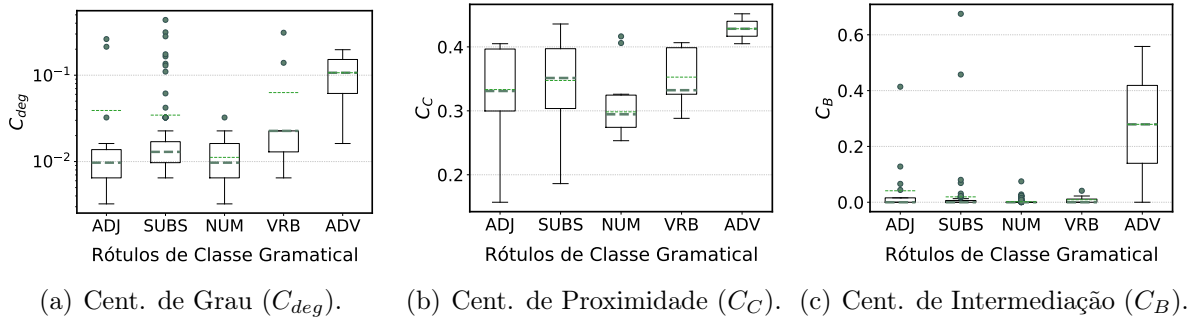


Figura A.4: Resultados da aplicação das métricas de centralidade de grau, proximidade e intermediação sobre no grafo de palavras gerado a partir da rede de intenções pré-processada. Entre os diferentes grupos de classes gramaticais, os advérbios (ADV) demonstram um comportamento mais central independentemente da métrica avaliada.

de coocorrência dos respectivos *tokens* no *corpus* de intenções. Ao analisar a variação de métricas clássicas de centralidade, como grau, proximidade e intermediação, considerando diferentes rótulos de classe gramatical, é possível revelar várias características imperceptíveis no formato textual.

As Figuras A.4(a), A.4(b) e A.4(c) mostram as métricas de centralidade de grau, proximidade e intermediação como *boxplots* para cada grupo de palavras de classe gramatical presentes no grafo de intenções. Independentemente do tipo de centralidade, as palavras ou termos classificados como numerais demonstram consistentemente valores baixos de centralidade. Esse fenômeno ocorre porque os valores numéricos usados nas intenções são específicos e precisos, uma vez que normalmente compreendem parâmetros como endereços IP e valores de largura de banda. Assim, nós de valores numéricos são colocados na periferia do grafo, tornando-os menos centrais. Como visto na Figura A.5, um padrão disassortativo recorrente ($A_c < 0$) é observado em todos os grupos de classes gramaticais, mostrando que a maioria dos nós de palavras centrais vem de palavras com graus mais baixos. Outro achado interessante é que vários nós são classificados como substantivos, demonstrando semelhança com uma malha completa, por exemplo, conectados a todos os nós do grafo. Esse fato é corroborado pela Figura A.4(a), já que os substantivos têm valores mais altos de centralidade de grau ($C_{deg} \approx 1$).

A segunda abordagem explora a perspectiva algorítmica, comparando o desempenho de diferentes modelos de redes neurais e parâmetros. A avaliação considera dois algoritmos de aprendizado profundo, LSTM e GRU. Ambos os algoritmos foram treinados usando 80% do conjunto de dados de intenções e, em seguida, testados com o restante das amostras para detectar conflitos semânticos entre pares de intenções. As Figuras A.6(a), A.6(b) e A.6(c) mostram os valores das métricas de recuperação de informação, ou seja,

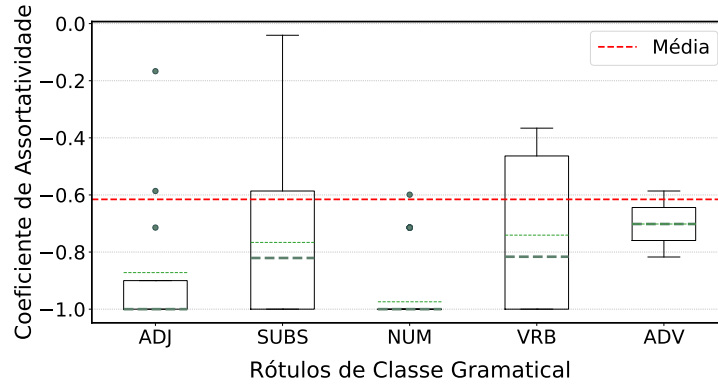


Figura A.5: Variação da métrica de assortatividade entre as classes gramaticais no grafo de intenções. Resultados revelam um padrão disassortativo recorrente ($A_c < 0$) em todas as classes gramaticais, indicando que palavras centrais estão principalmente conectadas a palavras de baixo grau.

acurácia, precisão, sensibilidade e F1-score, ao classificar pares de intenções. As figuras mostram configurações de rede neural com 64, 128 e 256 neurônios na camada oculta. A análise dos resultados indica um consenso de que o modelo LSTM superou o modelo GRU em cada cenário. Especialmente em testes adotando 64 neurônios, observa-se um desempenho superior do modelo LSTM em comparação com o modelo GRU em todas as métricas avaliadas. Embora essa diferença de desempenho diminua em testes adotando 128 neurônios, ela retorna consideravelmente no teste com o número máximo de neurônios, ou seja, 256. No entanto, vale ressaltar que o algoritmo LSTM fornece consistentemente modelos com alta sensibilidade para detectar conflitos entre intenções, com sensibilidade superior a 80%, independentemente da quantidade de neurônios na camada oculta. A sensibilidade é inerentemente uma métrica que quantifica a capacidade de um sistema em identificar todos os casos positivos em relação ao número total de casos positivos existentes, ou seja, a soma de verdadeiros positivos e falsos negativos. Na detecção de conflitos, instâncias positivas surgem quando existem conflitos genuínos entre intenções. Consequentemente, a eficácia dos modelos baseados em LSTM reflete a capacidade do sistema AGIR de identificar no mínimo 80% dos conflitos reais, reduzindo a ocorrência de alertas indevidos ou ações de mitigação desnecessárias.

A.6 Conclusão

Este trabalho propôs o sistema AGIR, um sistema de gerenciamento baseado em intenção para o refinamento do nível de serviço, projetado para lidar com as complexidades das Redes de Acesso por Rádio Abertas (Open RAN). O AGIR apresenta uma abordagem

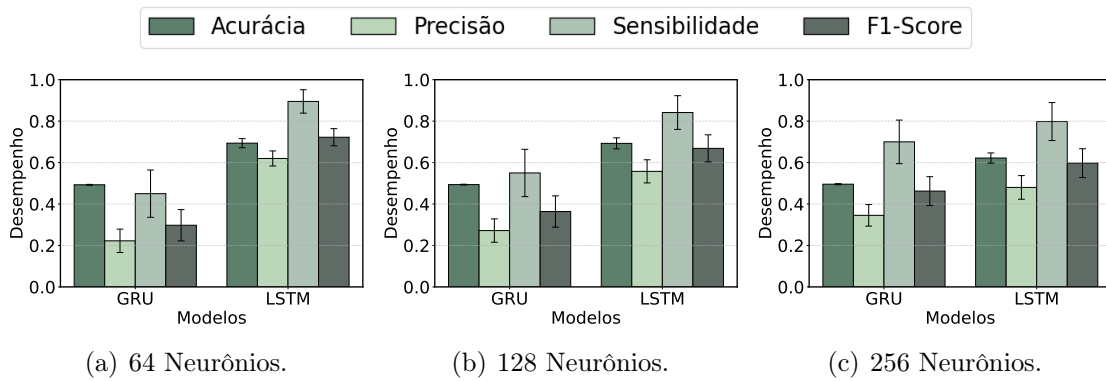


Figura A.6: A segunda abordagem de avaliação envolve a comparação de desempenho de diferentes modelos de redes neurais durante a detecção de conflitos semânticos entre intenções de rede. Os resultados são apresentados por meio de métricas de recuperação de informações em diferentes configurações de neurônios na camada oculta. O LSTM consistentemente resulta em modelos com maior recall.

inovadora para o gerenciamento de rede baseada em intenções, permitindo que os operadores definam e imponham SLOs precisos dentro do ambiente RAN. O sistema atua como uma ponte entre intenções e ações, traduzindo de forma contínua as intenções de alto nível dos operadores dos SLOs em instruções de rede acionáveis. Essa capacidade transformadora aprimora a flexibilidade e escalabilidade da rede e reduz potenciais erros humanos. Composto por quatro módulos integrais, i.e., Aplicação Inteligente (iApp), Tradutor, Resolvedor de Conflitos e Agente de Rede, o sistema AGIR opera de forma coesa para executar seu processo central. O processo de avaliação focou principalmente nos módulos Tradutor e Resolvedor de Conflitos em uma abordagem dupla. A primeira abordagem compara as métricas de centralidade e assortatividade de cada grupo de rótulos de classe gramatical, através da construção de uma estrutura em grafo das intenções analisadas. Os resultados revelam que as declarações de intenções mostram padrões dissassortativos recorrentes e que algumas entidades de rede são mais propensas a ser o sujeito de uma intenção do que outras. A segunda abordagem comparou o desempenho de dois modelos de redes neurais profundas, LSTM e GRU, submetidos à classificação de intenções como conflitantes. A superioridade do modelo LSTM foi verificada em todos os cenários e métricas testadas. Variações do modelo LSTM alcançaram mais de 80% de sensibilidade na classificação de intenções conflitantes. Em trabalhos futuros, pretende-se testar outros modelos de redes neurais e adaptar o reconhecimento de entidades nomeadas relacionadas ao escopo de redes com base no treinamento com *corpora* de intenções mais realistas e mais populosas.