



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

MARCOS DE PINHO COTRIM

INFLUÊNCIA OCULTA ENTRE REDES SOCIAIS: UM ESTUDO DO EFEITO DE
GRUPOS DO WHATSAPP NO IMPULSIONAMENTO DE CONTEÚDO DO YOUTUBE

RIO DE JANEIRO
2021



MARCOS DE PINHO COTRIM

INFLUÊNCIA OCULTA ENTRE REDES SOCIAIS: UM ESTUDO DO EFEITO DE
GRUPOS DO WHATSAPP NO IMPULSIONAMENTO DE CONTEÚDO DO YOUTUBE

Dissertação apresentada como requisito parcial
para obtenção do título de mestre pelo
Programa de Pós-Graduação em Informática da
Universidade Federal do Estado do Rio de
Janeiro (UNIRIO).

Orientador: Carlos Eduardo Ribeiro de Mello

Rio de Janeiro
2021

Catálogo informatizada pelo(a) autor(a)

- | | |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| C843 | <p>Cotrim, Marcos de Pinho</p> <p>Influência oculta entre redes sociais: um estudo do efeito de grupos do WhatsApp no impulsionamento de conteúdo do YouTube. / Marcos de Pinho Cotrim. -- Rio de Janeiro, 2021.</p> <p>78 p.</p> <p>Orientador: Carlos Eduardo Ribeiro de Mello.</p> <p>Dissertação (Mestrado) - Universidade Federal do Estado do Rio de Janeiro, Programa de Pós-Graduação em Informática, 2021.</p> <p>1. Inferência Causal. 2. Redes Sociais Online. 3. Matching. 4. WhatsApp. 5. Youtube. I. Mello, Carlos Eduardo Ribeiro de, orient. II. Título.</p> |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

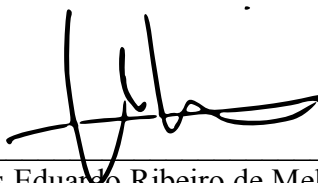
MARCOS DE PINHO COTRIM

INFLUÊNCIA OCULTA ENTRE REDES SOCIAIS: UM ESTUDO DO EFEITO DE
GRUPOS DO WHATSAPP NO IMPULSIONAMENTO DE CONTEÚDO DO YOUTUBE

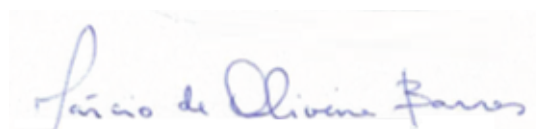
Dissertação apresentada como requisito parcial
para obtenção do título de mestre pelo
Programa de Pós-Graduação em Informática da
Universidade Federal do Estado do Rio de
Janeiro (UNIRIO).

Aprovado em: 19/10/2021.

Banca examinadora:



Prof. Dr. Carlos Eduardo Ribeiro de Mello (Orientador)
Universidade Federal do Estado do Rio de Janeiro – UNIRIO.



Prof. Dr. Márcio de Oliveira Barros
Universidade Federal do Estado do Rio de Janeiro – UNIRIO.



Prof. Dr. Eduardo Fonseca Mendes
Fundação Getúlio Vargas – FGV.

AGRADECIMENTOS

Sou muito grato a todos aqueles que me deram suporte, conselhos e energia durante todo o processo que culminou nesse trabalho.

Aos meus pais, por me transmitirem valores e fundamentos que carrego por toda a vida, sabendo que estarão sempre ao lado para o que for necessário.

À minha esposa e filha, por acreditarem em mim mais do que eu mesmo, por me incentivarem a fazer escolhas e por apoiarem minhas decisões, mesmo quando significava terem que abrir mão de alguma coisa. Passamos juntos por isso. Amo vocês!

Aos meus amigos, pela troca constante de experiências e ideias, mesmo que a ocorrência inesperada de uma pandemia global tenha nos deixado um pouco distantes durante tanto tempo.

À CVM, pela oportunidade de dedicar parte do meu tempo nos últimos anos ao desenvolvimento de conhecimentos que levarei para o resto da vida e que, tenho certeza, serão retribuídos ao longo dos próximos anos.

À UNIRIO e a todos os seus professores e servidores, agradeço pela oportunidade de desenvolver essa pesquisa em uma instituição de luta constante pela manutenção de um ensino público de qualidade.

Ao meu orientador Cadu, também conhecido como Carlos Eduardo Ribeiro de Mello, agradeço pela paciência e por todos os conhecimentos e ensinamentos compartilhados. Se não fossem por suas críticas e palavras de incentivo, esse trabalho não teria chegado até aqui, e sabemos disso! Muito obrigado!

Agradeço também aos membros da banca, professores Márcio de Oliveira Barros e Eduardo Fonseca Mendes, por terem aceitado o convite e por terem se disposto a avaliar meu trabalho. Foi uma honra e uma grande oportunidade de aprendizado.

Por fim, agradeço a todos aqueles que fazem parte da minha vida e que contribuíram direta ou indiretamente para que chegasse até aqui.

Sinceramente, muito obrigado!

COTRIM, Marcos de Pinho. **Influência oculta entre redes sociais: um estudo do efeito de grupos do WhatsApp no impulsionamento de conteúdo do YouTube**. 2021. 73 f. Dissertação (Mestrado em Informática), PPGI, Departamento de Informática Aplicada, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2021.

RESUMO

As redes sociais online (OSN) deixaram de ser um simples espaço de interação entre pessoas e se tornaram um ambiente eficaz para disseminação de informações e mobilização de pessoas. Porém, os atores e as intenções por trás de campanhas de comunicação nem sempre ficam explícitos e muitas vezes os impactos de ações estimuladas em uma rede específica buscam influenciar resultados fora de suas fronteiras. Nesse estudo, métodos clássicos de inferência causal foram aplicados para estimar os efeitos que ações realizadas em uma rede social causam sobre outra. O método incluiu a representação do processo de geração dos dados através de um modelo causal, a identificação e controle de variáveis causadoras de interferências entre as variáveis de tratamento e desfecho, e a estimativa do efeito causal através de *matching* por distância euclidiana e *coarsened exact matching*. Aplicado em um cenário concreto, os resultados sugerem que os vídeos compartilhados em grupos do WhatsApp durante a campanha eleitoral brasileira de 2018 tiveram um aumento aproximado de 30% no número de visualizações e de 20% na quantidade de engajamento que receberam no YouTube. Apresentamos evidências de que o compartilhamento nesses grupos pode ter contribuído para aumentar a remuneração dos canais do YouTube que os publicaram.

Palavras-chave: Inferência Causal. Redes Sociais Online. Matching. WhatsApp. Youtube.

COTRIM, Marcos de Pinho. **Influência oculta entre redes sociais: um estudo do efeito de grupos do WhatsApp no impulsionamento de conteúdo do YouTube.** 2021. 73 f.
Dissertação (Mestrado em Informática), PPGI, Departamento de Informática Aplicada,
Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2021.

ABSTRACT

Online social networks (OSN) are no longer simply a place of human interaction and have become an effective environment for information dissemination and people mobilization. However, the actors and intentions behind communication campaigns are not always explicit, and, frequently, activities in a specific network aim to influence results outside their borders. In this study, classical methods of causal inference were applied to estimate the effects caused by actions performed in one social network over another. The methods involved the representation of the data-generating process through a causal model, the identification and control of confounding factors with influence over treatment and outcome variables, and the estimation of the causal effect through Euclidean Distance Matching (EDM) and Coarsened Exact Matching (CEM). Analysis held over a concrete case suggest that videos shared in WhatsApp groups during the 2018 Brazilian election campaign caused a 30% increase in the number of views and an additional 20% increase in the amount of engagement they received on YouTube. We present evidence that sharing videos in these groups may have contributed to increase the monetization of the YouTube channels that published them.

Keywords: Causal Inference, Online Social Network. Matching. WhatsApp. YouTube.

LISTA DE ILUSTRAÇÕES

Figura 1. Relação implícita entre atores e resultados.	14
Figura 2. Exemplo de efeito causal direto ou através de variável mediadora.	25
Figura 3. Exemplo de <i>back-door</i> entre uma variável tratamento D e o desfecho Y.	26
Figura 4. Efeitos da variável intervenção sobre as variáveis de desfecho.	27
Figura 5. Principais etapas do trabalho.	29
Figura 6. Fluxo simplificados dos conjuntos de dados da origem até a amostra selecionada.	30
Figura 7. Esquema simplificado de seleção da amostra de dados analisada.	32
Figura 8. Média de mensagens e grupos por usuário.	35
Figura 9. Influência do WhatsApp sobre a atenção recebida por conteúdos na Internet.	36
Figura 10. Influência de canais e vídeos sobre o compartilhamento no WhatsApp.	38
Figura 11. Efeitos do número de visualizações e engajamento sobre a monetização.	38
Figura 12. Efeitos da repercussão do vídeo sobre variáveis de desfechos.	39
Figura 13. Relação entre intervenção e desfechos.	40
Figura 14. Diagrama causal da influência do WhatsApp sobre visualizações.	41
Figura 15. Diagrama causal da influência do WhatsApp sobre engajamento.	42
Figura 16. Evolução da situação dos vídeos entre 2018 e 2021.	45
Figura 17. Distribuição do número de visualizações entre grupos de tratamento e controle.	48
Figura 18. Distribuição da quantidade de engajamento entre grupos de tratamento e controle.	48
Figura 19. Correlação entre o número de visualizações e a quantidade de engajamento.	49
Figura 20. Distribuição da quantidade de vídeos por canal entre grupos de tratamento e controle.	50
Figura 21. Distribuição do dia de publicação entre grupos de tratamento e controle.	51
Figura 22. Diagrama causal da influência do WhatsApp sobre visualizações.	54
Figura 23. Distâncias entre todas as observações do grupo de tratamento e controle.	56
Figura 24. Distâncias entre as observações de cada par após o <i>matching</i>	56
Figura 25. Distribuição dos efeitos observados em todos os pares.	57
Figura 26. Diagrama causal da influência do WhatsApp sobre engajamento.	62
Figura 27. Distâncias entre todas as observações do grupo de tratamento e controle.	64
Figura 28. Distâncias entre as observações de cada par após o <i>matching</i>	64
Figura 29. Distribuição dos efeitos observados em todos os pares.	65

LISTA DE TABELAS

Tabela 1. Estatísticas gerais dos vídeos mencionados nos grupos de WhatsApp.	44
Tabela 2. Evolução da situação dos vídeos de 2018 a 2021.....	45
Tabela 3. Variáveis binárias (<i>dummy</i>) representativas dos vídeos.....	50
Tabela 4. Balanceamento entre grupos de tratamento e controle antes de controle de covariantes.	53
Tabela 5. Efeito estimado considerando diferentes valores para corte de observações.....	58
Tabela 6. Balanceamento entre grupos de tratamento e controle após <i>matching</i> ($d < 1$).....	58
Tabela 7. Matriz de correlação entre as variáveis endógenas após <i>matching</i> ($d < 1$).	59
Tabela 8. Quantidade de observações antes e depois do descarte.	60
Tabela 9. Efeito do compartilhamento no WhatsApp sobre número de visualizações estimados por <i>matching</i>	61
Tabela 10. Efeito estimado considerando diferentes valores para corte de observações.....	65
Tabela 11. Balanceamento entre grupos de tratamento e controle após <i>matching</i> ($d < 1$)....	66
Tabela 12. Matriz de correlação entre as variáveis endógenas após <i>matching</i> ($d < 1$).	66
Tabela 13. Quantidade de observações antes e depois do descarte.	67
Tabela 14. Efeito do compartilhamento no WhatsApp sobre quantidade de engajamento estimados por <i>matching</i>	68
Tabela 15. Efeito médio do WhatsApp sobre visualizações e engajamento.....	69

LISTA DE ABREVIATURAS E SIGLAS

ATE – Average Treatment Effect
ATT – Average Treatment Effect on Treated
DAG – Directed Acyclical Graph
EAC – Estudo Aleatório Controlado
OSN – Online Social Network
RCT – Randomized Controlled Trial

SUMÁRIO

1 INTRODUÇÃO	12
1.1 Motivação	13
1.2 Caracterização do Problema	15
1.3 Objetivos	16
1.4 Principais contribuições	16
1.5 Organização do Trabalho	17
2 BIBLIOGRAFIA RELACIONADA	18
3 FUNDAMENTOS METODOLÓGICOS	21
3.1 Inferência causal	21
3.2 Problema fundamental da inferência causal	22
3.3 Experimentos aleatórios controlados (EAC)	22
3.4 Matching	23
3.4.1 <i>Matching</i> por distância euclidiana	24
3.4.2 Coarsened Exact Matching (CEM)	24
3.5 Diagramas causais	24
3.6 Critério de portas dos fundos	25
4 ABORDAGEM DO TRABALHO	27
4.1 Representação do processo de geração dos dados	30
4.2 Canais do YouTube como agrupamentos naturais	30
4.3 Composição dos grupos de tratamento e controle	31
4.4 Estimativa do efeito causal	32
5 PROPOSTA DE MODELAGEM	34
5.1 O processo de geração dos dados	34
5.1.1 Grupos públicos do WhatsApp como rede social	34
5.1.2 Grande volume de URLs para vídeos do YouTube	36
5.1.3 Produção de vídeos para o YouTube	37
5.1.4 Geração de receitas no YouTube	38
5.1.5 Repercussão dos vídeos na Internet	39
5.2 Diagramas causais	39
5.2.1 Número de visualizações (<i>YV</i>)	40
5.2.2 Quantidade de engajamento (<i>YE</i>)	42
6 ESTRUTURAÇÃO DOS DADOS	44
6.1 Construção do conjunto de dados	44
6.1.1 Coleta original de dados do WhatsApp (2018)	44
6.1.2 Recorte temporal	44
6.1.3 Seleção dos grupos de tratamento e controle	45
6.1.4 Descarte de canais com poucas observações	46
6.1.5 Coleta de metadados do YouTube	46
6.1.6 Descarte de observações sem valores	46
6.1.7 Considerações importantes	46
6.2 Descrição do conjunto de dados	47
6.2.1 <i>W</i> – WhatsApp (variável de intervenção)	47
6.2.2 <i>YV</i> – Log(Visualizações) (variável de desfecho)	47
6.2.3 <i>YE</i> – Log(Engajamento) (variável de desfecho)	48
6.2.4 Repercussão	49
6.2.5 Canal	50
6.2.6 Vídeo	50
6.2.7 Dia de Publicação	51

7 ANÁLISE DO EFEITO CAUSAL	52
7.1 Balanceamento da amostra antes de controle do viés de seleção	52
7.2 Efeito do WhatsApp sobre o número de visualizações	54
7.2.1 Matching por distância euclidiana	55
7.2.2 Coarsened Exact Matching (CEM)	59
7.2.3 Resultados	61
7.3 Efeito do WhatsApp sobre a quantidade de engajamento.....	62
7.3.1 <i>Matching</i> por distância euclidiana	63
7.3.1.1 <i>Coarsened Exact Matching (CEM)</i>	67
7.3.2 Resultados	67
8 CONCLUSÕES	69
8.1 Contribuições.....	71
8.2 Limitações.....	71
8.3 Trabalhos futuros	72
REFERÊNCIAS	74

1 INTRODUÇÃO

Nas últimas décadas assistimos as redes sociais online (OSN) se tornarem onipresentes na vida da sociedade. Aos poucos, as experiências individuais passaram a fazer sentido apenas após serem curtidas e compartilhadas. A popularidade de cada pessoa passou a ser medida com base no número de seguidores. A distância entre os habitantes do planeta foi praticamente eliminada, sendo possível acompanhar a vida de desconhecidos em diferentes partes do mundo em um grande *reality show* global (CHRISTAKIS e FOWLER, 2009).

Porém, para além das relações sociais, o espaço online tornou-se um grande mercado em que produtos, serviços e, principalmente, a própria atenção das pessoas são negociados continuamente, na maioria das vezes de forma pouco clara (DAVENPORT e BECK, 2002). Como as ações nas redes não estão sujeitas às mesmas barreiras e costumes que se aplicam ao mundo fora delas, o comportamento *online* às vezes cruza a fronteira do que seria esperado nas relações *offline* (YEE et. al., 2009).

O modelo atual de funcionamento das redes sociais se baseia em contas e perfis, e não em indivíduos. Mesmas pessoas podem navegar por diferentes espaços, sob diferentes identidades e personalidades. Da mesma forma, conteúdos patrocinados são misturados a conteúdos criados espontaneamente (*stealth marketing*). É difícil afirmar se determinada informação é efetivamente a opinião de uma pessoa ou apenas mais uma propaganda. Em outras palavras, não temos garantia nem do mensageiro nem da mensagem (KAIKATI e KAIKATI, 2004; MARTIN e SMITH, 2008).

Por outro lado, temos cada vez mais a consciência de que as ações nas redes sociais causam impactos no mundo concreto. De campanhas globais de desinformação à manipulação de mercados, pequenos grupos de atores organizados conseguem gerar grandes prejuízos, muitas vezes de forma velada, sem que suas ações e intenções sejam percebidas.

Assim se pergunta: será possível estimar o impacto causado por diferentes redes sociais? Mais especificamente, será possível estimar o efeito que ações realizadas em uma rede social específica causam fora dela? Essa resposta interessaria não só às plataformas e a seus clientes do ponto de vista de negócio, mas também aos responsáveis por políticas públicas. Será que as responsabilidades de diferentes redes sociais não deveriam ser proporcionais aos riscos a que elas nos expõem?

1.1 Motivação

As razões pelas quais as pessoas usam as redes sociais têm mudado com a evolução das plataformas e com o surgimento das novas gerações. Se por um lado as redes sociais cresceram nos últimos anos como mera ocupação do tempo livre e como fonte de entretenimento e de notícias, por outro lado continuam sendo usadas para que as pessoas mantenham contato com amigos e compartilhem detalhes das próprias vidas com outros (GWI, 2021).

Da mesma forma, empresas e profissionais estão presentes nas redes sociais visando algum interesse associados a seus negócios como obter novos clientes, comunicar sua marca, desenvolver novos produtos, vender, e conseguir se comunicar diretamente com seus clientes (CULNAN et. al., 2010).

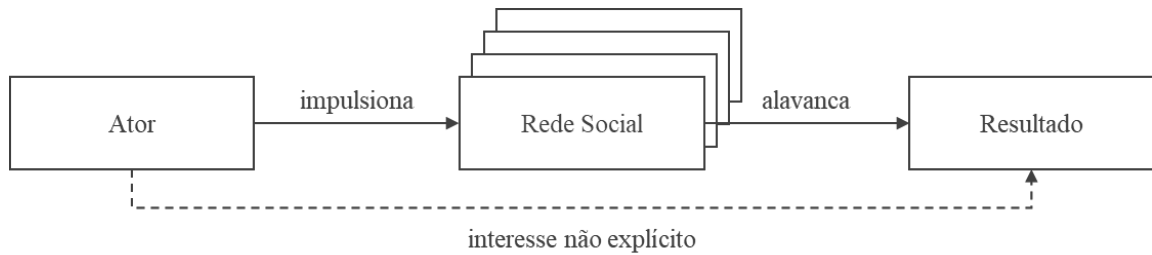
As plataformas de redes sociais (Facebook, Instagram, YouTube, LinkedIn etc.) sustentam essas relações, às vezes atuando de forma complementar, às vezes competindo entre si pelos mesmos usuários, mas sempre com o mesmo desafio: gerar receitas a partir de serviços oferecidos, na maior parte das vezes, gratuitamente.

Assim, é possível supor que os interesses de cada ator variem em função das partes com que se relacionam:

- Pessoas esperam que as plataformas ofereçam funcionalidades e recursos que facilitem suas interações, e que as empresas e profissionais disponibilizem seus produtos e serviços de forma fácil e rápida através das próprias redes.
- Empresas e profissionais esperam que as pessoas comprem mais produtos e serviços, e que as plataformas as ajudem a atingir um número maior e melhor de clientes, se possível com informações que possibilitem identificar novas oportunidades e mercados.
- Plataformas esperam que as pessoas lhes forneçam informações e tempo de atenção, e que as empresas e profissionais paguem pela exposição de suas marcas e serviços.

Dentro desse conjunto de relações possíveis é importante observar que, com exceção das pessoas com objetivos estritamente sociais, em todos os outros relacionamentos há sempre um interesse de negócio latente. Nesse contexto, podemos generalizar uma dinâmica padrão em que *um ator estimula as redes sociais para obter um determinado resultado*. Porém, atores, interesses e resultados nem sempre estão explícitos ou declarados. Essa relação é representada na **Figura 1**.

Figura 1. Relação implícita entre atores e resultados.



Como essas manifestações acontecem em diferentes plataformas, as ações estimuladas em uma rede social específica podem ter efeitos fora de suas fronteiras. Por exemplo, um influenciador digital hipotético (ator explícito) pode usar um vídeo no YouTube para estimular seus seguidores a compartilharem o perfil de uma empresa que o patrocina no Instagram. Neste exemplo, o influenciador digital (ator explícito), motivado por um interesse não declarado (*merchandising*), estimulou sua rede pessoal no YouTube para aumentar o número de seguidores que seu parceiro comercial possui no Instagram (resultado não explícito e externo à rede). Essa interrelação é comum e pode ser observada em diferentes casos recentes.

Redes sociais tornaram-se peças centrais em estratégias de campanhas políticas da última década. Equipes de campanha e apoiadores dos candidatos (atores não explícitos) trabalham suas redes sociais, de forma declarada ou não, para estimular o voto no dia da eleição (resultado explícito externo à rede) (SANTOS, 2019).

Em janeiro de 2021, usuários de fórum do Reddit¹ se organizaram “espontaneamente” para manipular o preço de ações da companhia GameStop com o argumento de que estariam agindo contra grandes fundos de investimento (CARLSON, 2021). Aqui, alguns membros do fórum (atores não explícitos) estimularam a rede social para aumentar o preço de cotação das ações da companhia (interesse declarado) e, com isso, causar prejuízo a fundos de investimento (resultado explícito). Ao mesmo tempo, podem ter obtido lucro com a venda da ação a um preço maior do que o de compra (resultado não explícito).

Durante a pandemia da COVID-19, redes sociais foram utilizadas para a disseminação de desinformação sobre tratamentos e vacinas (OMS, 2021). Algumas situações foram associadas posteriormente a interesses não declarados de ganhos financeiros com a comercialização de produtos alternativos (FRENKEL, 2021).

¹ O Reddit (<https://www.reddit.com/>) é uma plataforma formada por comunidades, ou *subreddits*, que podem ser públicas ou restritas, e que permitem aos usuários postar conteúdos, fazer comentários e votar sobre assuntos variados dentro de temáticas específicas.

Esse padrão na dinâmica dos atores e suas redes nos levou a refletir sobre a seguinte questão: será possível medir o impacto causado por uma rede social sobre um resultado externo a ela? Nos exemplos anteriores, qual foi o impacto que o vídeo do nosso influenciador digital hipotético causou sobre o número de seguidores do seu parceiro no Instagram? Quantos votos recebidos por um determinado candidato foram consequência exclusiva das ações realizadas por sua campanha no Facebook? Quanto do aumento no preço das ações da GameStop foi resultado da ação coordenada dos membros do fórum WallStreetBets no Reddit? Ainda, considerando que as campanhas de comunicação e marketing permeiam várias redes sociais de forma simultânea, será possível isolar a influência de cada uma delas sobre os resultados observados?

Essas perguntas pressupõem uma relação de causalidade entre ações em uma rede social e algum resultado externo a ela no sentido de que, se a ação não tivesse acontecido, o resultado observado teria sido outro. Neste sentido, este trabalho tem como objetivo investigar o possível efeito contrafactual associado a tais questões. Essa informação poderia ser usada como medida objetiva do grau de impacto ou responsabilidade de uma rede sobre o efeito de suas atividades.

1.2 Caracterização do Problema

Trazendo as questões apresentadas na seção anterior para um caso concreto, um cenário de estudo foi construído a partir de dados coletados por Bursztyn e Birnbaum (2019) sobre grupos públicos do WhatsApp durante a campanha eleitoral brasileira de 2018. Após a observação de que cerca de 5% das mensagens enviadas nos grupos faziam referência a algum vídeo do YouTube, identificou-se a oportunidade de utilizar esse conjunto de dados como base para um estudo com o objetivo de estimar qual teria sido a popularidade alcançada por esses vídeos caso não tivessem sido compartilhados nos grupos do WhatsApp. Desta forma, buscou-se **investigar a influência que grupos do WhatsApp podem exercer sobre a capacidade de monetização de vídeos no YouTube**.

Dentro do conjunto de dados analisados, membros de grupos públicos do WhatsApp (atores não explícitos) divulgaram e estimularam a divulgação de vídeos do YouTube (intenção declarada) e, com isso, podem ter causado um aumento significativo no número de visualizações e engajamento dos mesmos, o que causaria um aumento na receita gerada pelos canais que os publicaram (resultado não explícito). Até que ponto o compartilhamento de conteúdo em uma rede de alto engajamento como essa teve por interesse contribuir apenas para os objetivos da própria rede? Ou será que se pode considerar a possibilidade de o poder de

engajamento da rede ter sido utilizado também para impulsionar a remuneração de canais particulares?

1.3 Objetivos

Com foco no problema anterior, e considerando o contexto específico de mensagens compartilhadas em grupos públicos do WhatsApp e seus efeitos sobre os vídeos do YouTube divulgados através delas, buscou-se neste trabalho responder à seguinte questão geral:

- O compartilhamento dos vídeos nos grupos de WhatsApp observados influenciou o aumento da monetização desses vídeos no YouTube?

Veremos que, apesar de não ser possível observar a quantidade de monetização dos vídeos do YouTube, é possível aproximá-la com base no número de visualizações e na quantidade de engajamento que cada vídeo recebeu, informações que são públicas. Assim, a questão geral anterior é desmembrada em duas questões mais específicas e acessíveis:

- Qual o efeito causal do compartilhamento nos grupos do WhatsApp observado sobre a popularidade de vídeos do YouTube, medida pelo número de visualizações?
- Qual o efeito causal do compartilhamento nos grupos do WhatsApp observado sobre a reação gerada por vídeos do YouTube, medida pela quantidade de engajamento?

1.4 Principais contribuições

O estudo demonstra que métodos clássicos de inferência causal são suficientes para estimar quantitativamente o efeito que ações em uma determinada rede social geram em resultados fora de suas fronteiras.

Aplicado a um cenário concreto, o trabalho apresenta evidências de que grupos públicos do WhatsApp utilizados durante a campanha presidencial brasileira de 2018 podem ter contribuído para aumentar a receita de canais do YouTube que tiveram vídeos compartilhados nos mesmos.

Para realização da análise, foi estruturado um conjunto de dados integrando informações coletadas a partir de duas redes sociais distintas (WhatsApp e YouTube).

1.5 Organização do Trabalho

Essa dissertação está organizada da seguinte forma:

- O capítulo 2 apresenta uma revisão da bibliografia relacionada com plataformas de redes sociais com destaque para o WhatsApp enquanto plataforma de mobilização e engajamento.
- O capítulo 3 apresenta principais tópicos teóricos relacionados com a abordagem metodológica adotada no trabalho. A abordagem específica do estudo é apresentada no capítulo 4.
- O capítulo 5 contextualiza o fenômeno estudado e representa as premissas do modelo através de diagramas causais.
- No capítulo 6 descrevemos o processo de exploração e estruturação do conjunto de dados utilizado na análise.
- O capítulo 7 descreve a análise dos efeitos causais e outros resultados observados.
- No capítulo 8 são apresentadas as conclusões, limitações e trabalhos futuros potenciais.

2 BIBLIOGRAFIA RELACIONADA

Apesar da grande quantidade de estudos disponíveis sobre o impacto das redes sociais online na sociedade, em áreas como psicologia (STEINFELD, C. et. al., 2008) e comunicação (BROWN et. al., 2007), pesquisas envolvendo redes específicas, como WhatsApp e YouTube, têm ficado restritas ao comportamento de indivíduos dentro das próprias plataformas (ZHOU et. al., 2010; ANGER e KITTL, 2011) e ao uso de plataformas isoladas em contextos específicos como, por exemplo, no apoio à educação (AMRI, 2014; CHTOUKI et. al., 2012).

No caso do WhatsApp há uma complexidade adicional. Considerando se tratar de uma ferramenta de mensagens instantâneas “ponto a ponto”, com foco em preservar a privacidade e sigilo das comunicações dos usuários, sem um “ponto central de observação”, sua própria natureza reduz a quantidade de dados disponíveis e dificulta as possibilidades de estudos e experimentos relacionados.

Não por acaso, a maior parte dos estudos relacionados com o WhatsApp aborda sua aplicação em cenários muito particulares, uma vez que a coleta de dados pessoais depende da autorização explícita dos participantes (LAMBTON-HOWARD, 2019). Esse fato também explica porque a maior parte dos estudos realizados sobre redes sociais gira em torno daquelas que possuem APIs públicas para acesso aos dados, principalmente Twitter² e YouTube³.

Mesmo assim, estudos que cruzam informações de mais de uma rede social são escassos e geralmente contemplam análises descritivas e exploratórias sobre os conjuntos de dados. Thorson et. al. (2012), por exemplo, analisaram a dinâmica entre a publicação de vídeos no YouTube e postagens no Twitter associados ao movimento Occupy, percebendo a necessidade, e dificuldade, de pesquisas envolvendo dados de múltiplas plataformas. Mejova e Srinivasan (2012), por sua vez, analisaram os sentimentos de comentários nas duas plataformas, observando que, enquanto comentários no Twitter são mais influenciados por notícias e novidades originadas em fontes externas, as mensagens no YouTube possuem uma carga de opinião maior, nem sempre alinhada com as opiniões dos próprios vídeos comentados. Abisheva et. al. (2014) conduziram, talvez, o primeiro estudo em “larga escala” (milhões de vídeos e *tweets*), buscando identificar e analisar diferentes perfis de comportamento com base em dinâmicas de publicação de vídeos no YouTube e compartilhamento deles no Twitter.

Outros trabalhos abordaram a análise cruzada entre redes sociais como uma tarefa de predição, buscando prever comportamentos em uma rede a partir de atividades fora de suas

² <https://developer.twitter.com/en/docs/twitter-api>

³ <https://developers.google.com/youtube/v3>

fronteiras. Yu et.al. (2014) propuseram um método para prever o aumento de visualizações em vídeos do YouTube a partir de publicações no Twitter. Os autores não só concluíram ser possível prever o aumento em popularidade de vídeos com base na atividade de apenas uma rede social externa (neste caso o Twitter), mas também observaram que a diversidade de atividades na origem tem mais informação para a previsão de popularidade do que o volume de atividades em si. Zhang e Pennacchiotti (2013) analisaram a correlação entre os interesses manifestados por usuários no Facebook com o comportamento de compras dos mesmos no eBay. O interesse dos autores estava em prever o comportamento de compra a partir de informações de perfil e comportamento de usuários.

Em uma tentativa de romper a barreira de informação do WhatsApp, nos últimos anos alguns trabalhos buscaram desenvolver ferramentas para possibilitar, minimamente, o monitoramento e coleta de dados que circulam em grupos públicos da plataforma (GARIMELLA e TYSON, 2018; RESENDE et. al., 2018). Dados coletados através dessas ferramentas têm viabilizado uma série de novos estudos (como o presente). Porém esses estudos ainda acabam concentrados em países onde o WhatsApp tem maior utilização e com foco em casos em que a plataforma foi associada a ações de mobilização e influência de grandes grupos de pessoas, principalmente em manifestações e processos eleitorais (SANTOS, 2019).

Os dados utilizados nessa dissertação foram coletados originalmente por Bursztyrn e Birnbaum (2019) sobre grupos públicos de WhatsApp ativos durante o período da campanha presidencial brasileira de 2018. Além de descrever o processo de coleta dos dados, os autores observaram uma rede de grupos fortemente conectada, com praticamente todos os grupos em um mesmo componente gigante, e que grupos voltados a campanhas políticas são mais interconectados do que grupos que tratam de política em geral. Também destacaram que uma a cada 20 mensagens compartilhadas fazia referência a algum vídeo do YouTube.

Antes disso, Caetano et. al. (2018) e Resende et. al. (2018) estudaram o comportamento de grupos públicos do WhatsApp com focos distintos. Caetano et. al. (2018) propôs metodologia para caracterização do comportamento de grupos públicos sob perspectivas de mensagens, usuários e grupos. No trabalho observou-se a dinâmica temporal diária das atividades e a aderência das distribuições acumuladas de diferentes estatísticas à Lei de Zipf. Resende et. al. (2018) apresentaram um sistema para a coleta, análise e visualização de informações sobre grupos públicos de WhatsApp⁴ com foco em combate a campanhas de desinformação. Apesar de analisar duas redes de grupos em contextos distintos, os autores

⁴ <http://www.whatsapp-monitor.dcc.ufmg.br>

observaram em comum que, dentre as URLs compartilhadas, o domínio www.youtube.com foi, de longe, o mais referenciado nas mensagens (cerca de 2% do total de mensagens). Ao analisar a dinâmica de propagação de imagens, verificou-se que o reenvio dentro dos grupos acontecia em intervalos menores do que 2 horas em 40% dos casos, e que o tempo de vida do conteúdo nos grupos foi de 24 horas em metade dos casos, chegando a 4 dias em outros 30% de imagens.

Melo et. al. (2019) analisaram a dinâmica de disseminação de des(informação) em redes de grupos públicos do WhatsApp em diferentes países através do rastreamento de imagens. Apesar de 80% das mensagens serem postadas uma única vez, algumas chegaram a ser enviadas centenas de vezes e alcançaram vários grupos. De forma semelhante ao observado por Resende et. al. (2019), a grande maioria dos conteúdos sobreviveu por no máximo 2 dias. Além disso, simulou-se o impacto da restrição do número de destinatários no encaminhamento de mensagens sobre a velocidade e alcance do contágio da rede por um modelo de conteúdo viral. Apesar de pequena redução na velocidade, destacou-se que as ações implementadas pelo WhatsApp para combate à desinformação não são suficientes para impedir a contaminação de toda a rede.

Por fim, considerando o tamanho e capacidade de influência que as redes sociais adquiriram, estudos recentes, com resultados ainda incipientes, têm buscado compreender as estruturas de financiamento e monetização que estão por trás de fenômenos como as celebridades do YouTube (BUDZINSKI e GAENSSLE, 2018; COROMINA et. al., 2020). De certa forma, os resultados da presente dissertação buscam contribuir nesta direção.

3 FUNDAMENTOS METODOLÓGICOS

3.1 Inferência causal

A inferência causal é descrita por Pearl e Marckenzie (2018) como uma “nova ciência”. Não porque a ideia de causalidade seja algo recente, afinal seres humanos inferem naturalmente relações de causa e efeito entre eventos há dezenas milhares de anos. Mas apenas nas últimas décadas surgiram esforços para aproximar o vocabulário utilizado para definição de problemas causais à linguagem utilizada para elaboração de teorias científicas. (PEARL e MACKENZIE, 2018).

Diferentemente de análises estatísticas “puras”, perguntas causais demandam conhecimento sobre o processo por trás da geração dos dados (*data-generating process*). Respostas causais não podem ser computadas simplesmente a partir de dados. Toda afirmação que invoca conceitos causais se apoia em premissas baseadas no conhecimento que se tem sobre determinado fenômeno. São premissas que dependem de conhecimento e não podem ser inferidas ou definidas apenas com base em associações estatísticas. (PEARL, 2009). Para falar de causalidade, precisamos ter um modelo mental do mundo real (PEARL, 2018).

Alguns pesquisadores atribuem o início da inferência causal moderna aos trabalhos de Fischer (1935), Haavelmo (1943) e Rubin (1974). Outros a associam a pioneiros como John Snow (1886). Porém, nas últimas décadas se destaca o trabalho de Pearl (2000, 2009, 2018) que buscou estabelecer uma linguagem estruturada para problemas de causalidade.

Em termos gerais, a inferência causal pode ser compreendida como a aplicação de teorias e conhecimentos sobre um fenômeno para se estimar o impacto que eventos e intervenções selecionadas causam sobre um desfecho de interesse. Os métodos utilizados em estudos que envolvem perguntas causais dependem de conhecimento sobre o problema. (CUNNINGHAM, 2021).

As abordagens da inferência causal interessam a todo tipo de disciplina científica. No entanto, apesar de inovações em métodos e definições nos últimos anos, ainda há grande debate entre cientistas sobre a forma apropriada para se determinar causalidade e de se determinar a validade de resultados.

3.2 Problema fundamental da inferência causal

Considerando que um determinado efeito causal pode variar entre indivíduos, o cenário ideal para se compreender causalidade envolveria conhecer o efeito que uma intervenção específica gera sobre cada unidade de uma população, permitindo conhecer a distribuição completa dos efeitos observados. O problema fundamental da inferência causal reside no fato de que não é possível observar, para uma mesma unidade, dois estados de mundo mutuamente exclusivos: “com a intervenção” e “sem a intervenção”. A existência de um dos estados necessariamente anula a possibilidade de existência do outro (HOLLAND, 1986).

Splawa-Neyman (1923) e Rubin (1974) representaram essa situação através do que chamaram de modelo de resultados potenciais (*potencial outcomes framework*). Segundo esse modelo, efeitos causais podem ser definidos como a comparação dos resultados potenciais observados para mesmas unidades sob diferentes tratamentos. O objeto de análise é a variável de desfecho Y no nível de cada unidade, representada originalmente como $Y_x(u)$ significando “o valor que seria observado para o desfecho Y na unidade u caso o valor do tratamento X tivesse sido x ” (PEARL, 2009).

A ideia de desfechos potenciais mistura-se com a noção de contrafactuais. Em outras palavras, “o que teria acontecido se tivéssemos seguido por outro caminho?”. Assim, Holland (1986) afirma que “é impossível observar o valor de $Y_t(u)$ e $Y_c(u)$ na mesma unidade e, portanto, é impossível observar o efeito de t sobre u ”. Uma alternativa para essa limitação é substituir o efeito de t sobre uma unidade específica, que não pode ser observado, pelo efeito médio de t sobre uma população de unidades, que pode ser estimado (HOLLAND, 1986).

3.3 Experimentos aleatórios controlados (EAC)

Uma vez que, por definição, não é possível medir o efeito causal de uma intervenção sobre cada unidade, experimentos de inferência causal são projetados para que seja possível estimar o efeito médio de um tratamento com base em resultados observados sobre determinada amostra de uma população. Para isso, separa-se a população em dois grupos, um que receberá o tratamento e outro que não receberá o tratamento e funcionará como grupo de controle.

O desafio no desenho de estudos como esse está no processo de seleção de quais observações serão atribuídas ao grupo de tratamento e quais serão atribuídas ao grupo de controle. A menos que a atribuição de indivíduos a um dos grupos seja realizada de forma

aleatória, precisamos considerar que os grupos estarão sujeitos a um viés de seleção (ANGRIST e PISCHKE, 2009).

A melhor forma de se eliminar o viés de seleção é através de experimentos aleatórios controlados (*randomized controlled trials*). Nestes experimentos, a atribuição de uma unidade ao grupo de tratamento e controle é realizada de forma aleatória. Tais estudos experimentais são considerados o ‘padrão-ouro’ em diversas áreas, tais como, Economia, Medicina e Epidemiologia. (ANGRIST e PISCHKE, 2014).

Apesar de experimentos aleatórios controlados serem considerados o padrão ouro da ciência, em muitos casos tais experimentos são impraticáveis e muitas perguntas só podem ser respondidas através de estudos observacionais ou quase-experimentos. Porém, a ausência de aleatoriedade na alocação de indivíduos entre grupos de tratamento ou controle não garante que a distribuição das características desses indivíduos seja comparável, o que pode levar a vieses sistemáticos e a resultados errados. (HAMMER, et. al., 2009).

3.4 Matching

Métodos de *matching* buscam replicar, a partir de dados observacionais (não aleatorizados), distribuições semelhantes a que seriam obtidas em um experimento aleatório controlado para o mesmo conjunto de covariáveis observadas. As técnicas de *matching* visam reduzir o viés de seleção em dados de experimentos não aleatorizados, podendo ser aplicadas tanto no momento de seleção e identificação das observações quanto na estimativa de efeitos causais, após a coleta de dados (STUART, 2010).

São métodos não paramétricos que buscam controlar as influências que variáveis de confusão (*confounding*) causam sobre as variáveis de tratamento em dados observacionais. O principal objetivo dos métodos de *matching* é eliminar observações de forma que os dados restantes sejam mais bem balanceados entre os grupos de tratamento e controle, fazendo com que as distribuições empíricas das covariantes nos grupos sejam mais semelhantes (IACUS, KING e PORRO, 2012).

Em linhas gerais, métodos de *matching* buscam identificar observações nos grupos de tratamento e controle semelhantes entre si. Esses indivíduos semelhantes podem, então, ser comparados como se fossem aproximações dos seus “desfechos potenciais alternativos”. Para isso, a aplicação de *matching* envolve a definição de um critério para seleção e agrupamento de observações dos grupos de tratamento e controle consideradas semelhantes, e de alguma

estimativa do efeito do tratamento a partir das diferenças observadas entre os desfechos das unidades semelhantes selecionadas (STUART, 2010; IACUS, KING e PORRO, 2012).

3.4.1 *Matching* por distância euclidiana

O método de *matching* por distância euclidiana é uma adaptação simples do método *Mahalanobis Distance Matching* (MBM) em que a distância de Mahalanobis é substituída pela distância euclidiana (SPIEL et. al., 2008).

O método se baseia na noção de uma distância entre observações com base nos valores de suas covariantes. As implementações mais comuns do método buscam identificar, para cada unidade do grupo de tratamento, seu vizinho mais próximo no grupo de controle. A seleção de unidades próximas no grupo de controle pode ser realizada com ou sem reposição (KING et. al, 2011).

Posteriormente, algum critério pode ser adotado para se descartar os pares em que as unidades de tratamento estejam em distância “fora do comum” das unidades de controle selecionadas. O procedimento de *calipers*, por exemplo, propõe o estabelecimento de valores de corte para as maiores distâncias permitidas entre os pares (STUART e RUBIN, 2008).

3.4.2 Coarsened Exact Matching (CEM)

No método de *coarsened exact matching* (CEM), cada covariante é transformada em uma variável categórica de forma que valores próximos possam ser agrupados dentro de mesmas categorias. Em seguida, as observações que possuem mesmos valores para todas as variáveis categóricas são agrupadas em estratos. Por fim, as unidades de controle têm seus valores ponderados em cada estrato para que possam ser consideradas em igual número às observações de tratamento. Estratos que não possuem pelo menos uma observação de controle e uma observação de tratamento são descartados (IACUS, KING e PORROS, 2012).

3.5 Diagramas causais

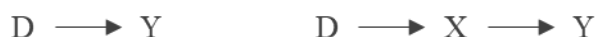
Diagramas causais são uma linguagem gráfica para representação do conhecimento sobre um fenômeno através de relações de influência entre variáveis. As setas (ou caminhos) que ligam as variáveis representam premissas conhecidas dos efeitos causados por uma sobre a outras. É importante observar que, em um diagrama causal, as premissas de causa-efeito estão

representadas não só nas ligações incluídas, mas, principalmente, nas ligações não incluídas. A ausência de uma ligação representa um argumento do pesquisador de que não existe influência entre as variáveis (PEARL, 2009).

A “análise de caminhos” foi utilizada pela primeira vez por Wright (1921) como representação gráfica da relação de influência entre diferentes variáveis e como recurso para responder a perguntas causais a partir de dados (*path diagrams*). Apesar desse primeiro registro há 1 século, o uso de diagramas causais como ferramenta formal de inferência causal é consequência de adaptação realizada por Pearl (2009).

Diagramas causais são representados através de grafos acíclicos dirigidos ou DAG (*directed acyclic graph*). Em um DAG a causalidade entre duas variáveis flui em uma única direção, sempre em direção ao futuro e sem a possibilidade de surgimento de ciclos. Efeitos causais podem acontecer de forma direta como em $D \rightarrow Y$ ou através de uma variável mediadora como em $D \rightarrow X \rightarrow Y$, como exemplificado na **Figura 2** (CUNNINGHAM, 2021).

Figura 2. Exemplo de efeito causal direto ou através de variável mediadora.



Um diagrama causal deve ser uma representação fiel do conhecimento que se tem sobre o fenômeno sob estudo. Ele representa todas as premissas consideradas como verdadeiras na análise de um determinado problema.

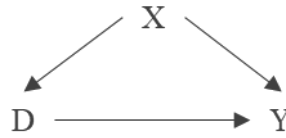
3.6 Critério de portas dos fundos

Para se estimar o efeito que uma variável causa sobre outra, é fundamental compreender a influência que as variáveis endógenas exercem umas sobre as outras, incluindo as variáveis de tratamento e de desfecho. Fatores que afetam tanto o desfecho quanto o tratamento são também chamados de fatores de confusão (*confounding*) e precisam ser identificados e tratados adequadamente. O método *back-door criterion*, ou “critério de portas dos fundos” (tradução nossa), proposto por Pearl (1993) possibilita a identificação, a partir de um diagrama causal, das variáveis endógenas que precisam ser controladas para remoção do efeito de confusão que estas causam sobre a variável de tratamento e a variável de desfecho (PEARL, 2009).

Como exemplo, na **Figura 3** a variável D exerce um efeito causal sobre Y . Da mesma forma, uma terceira variável X exerce influência simultaneamente sobre D e Y . Neste cenário, além da relação direta $D \rightarrow Y$, existe uma relação indireta entre D e Y que flui através do

caminho $D \leftarrow X \rightarrow Y$. A variável X gera confusão na relação observada entre D e Y e, enquanto não for controlada, diz-se que há um caminho aberto através de “portas dos fundos”.

Figura 3. Exemplo de *back-door* entre uma variável tratamento D e o desfecho Y .

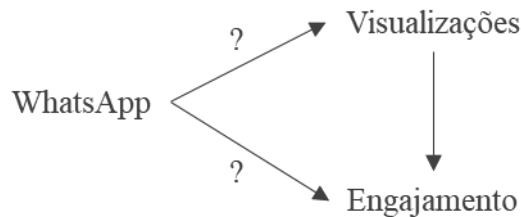


Ao se condicionar (controlar ou estratificar) a relação entre o tratamento D e o desfecho Y através de um conjunto de variáveis endógenas com base no critério de portas dos fundos, o efeito de associações espúrias é bloqueado permitindo que a associação restante medida entre D e Y seja interpretada como uma relação causal (PEARL, 2009).

4 ABORDAGEM DO TRABALHO

Neste estudo, deseja-se estimar o efeito causal que o compartilhamento de vídeos do YouTube em grupos do WhatsApp teve sobre o número de visualizações e a quantidade de engajamento que eles receberam, como representado na **Figura 4**.

Figura 4. Efeitos da variável intervenção sobre as variáveis de desfecho.



Considera-se como engajamento o total de “interações mensuráveis” que os usuários tiveram com os vídeos, complementares à ação de visualização. Especificamente, a quantidade de engajamento foi determinada como a soma da quantidade de ações dos tipos: curtir (*like*), não curtir (*dislike*) e comentar, características de redes sociais.

Define-se o “compartilhamento no WhatsApp” como variável de intervenção ou tratamento (W), binária, que assume valor 1 se o vídeo tiver sido compartilhado em um dos grupos, e 0 caso contrário. Assim, busca-se investigar seu efeito sobre dois possíveis desfechos (Y) diferentes: o número de visualizações (Y_V) e a quantidade de engajamento (Y_E).

Em um cenário hipotético ideal, para medir o efeito que a intervenção W teve sobre cada observação, seria preciso comparar, para cada vídeo, os desfechos que eles tiveram em duas condições antagônicas: com intervenção e sem intervenção. Assim, o efeito do compartilhamento sobre o número de visualizações de um determinado vídeo, por exemplo, seria a diferença entre o número de visualizações observado naqueles compartilhados no WhatsApp e nos que não foram compartilhados, como representado na equação (4.1).

$$Efeito(W) = Y(W = 1) - Y(W = 0) \quad (4.1)$$

Contudo, isto recai no problema fundamental da inferência causal, *i.e.* não é possível conhecer esses dois estados – “compartilhado” e “não compartilhado” – para uma mesma observação. Necessariamente, a ocorrência da intervenção elimina a existência da observação sob o “estado alternativo à não intervenção”. Logo, o efeito da intervenção não pode ser medido individualmente, mas apenas estimado. Todavia, estimar o efeito causal de uma intervenção

sobre o desfecho traz um conjunto de desafios. É necessário garantir que se está comparando coisas semelhantes. Uma vez que coisas semelhantes não são mesmas coisas, por definição, o resultado será sempre uma aproximação.

Essa aproximação realiza-se com a separação das observações em grupos distintos. Define-se como grupo de tratamento aquele formado pelas observações que receberam a intervenção ($W = 1$) e como grupo de controle aquele formado pelas observações que não foram submetidas à intervenção ($W = 0$). Dessa forma, seria possível, em tese, estimar o efeito médio causado pela intervenção comparando-se a diferença entre os desfechos estimados separadamente nos grupos de tratamento e de controle como representado na equação (4.2):

$$\text{Efeito estimado}(W) = E[Y(W = 1)] - E[Y(W = 0)] \quad (4.2)$$

Entretanto, sem mais critérios e cuidados na seleção dos vídeos, é perfeitamente possível que uma diferença entre desfechos observados nos conjuntos de tratamento e controle não seja resultado da intervenção em si, mas apenas consequência das diferenças intrínsecas aos elementos das amostras. Se o que se deseja é comparar o número de visualizações obtidas por vídeos que foram compartilhados nos grupos do WhatsApp, com a quantidade obtida por vídeos que não foram compartilhados, *é preciso garantir que se esteja comparando vídeos de características semelhantes*. Em outras palavras, é necessário controlar outras características dos vídeos, evitando-se assim um efeito endógeno das covariantes. Por exemplo, ao se comparar vídeos de um influenciador digital popular, com vídeos de um canal de baixa popularidade, é bastante provável que seja observada uma grande diferença no número de visualizações causada não pelo compartilhamento no WhatsApp, mas apenas pela diferença no número de inscritos de cada canal.

O padrão ouro para a construção de grupos de tratamento e controle sem viés de seleção são os experimentos aleatórios controlados (EAC). Em um EAC, a definição de quais observações serão submetidas à intervenção é aleatória. Com amostras de tamanho suficiente, a Lei dos Grandes Números (LGN) garante que, na média, as características dos elementos de cada amostra (tratamento e controle) sejam semelhantes (ANGRIST e PISCHKE, 2014).

Entretanto, a realização de um EAC nem sempre é possível ou viável. No caso deste estudo, foram utilizados dados de eventos passados. Mesmo se houvesse a intenção em se coletar novos dados para um experimento com desenho semelhante, não seria possível definir de forma aleatória quais vídeos seriam compartilhados no WhatsApp e quais não, sem que, com isso, outros vieses fossem criados no processo de geração dos dados. Por isso, conduzimos um

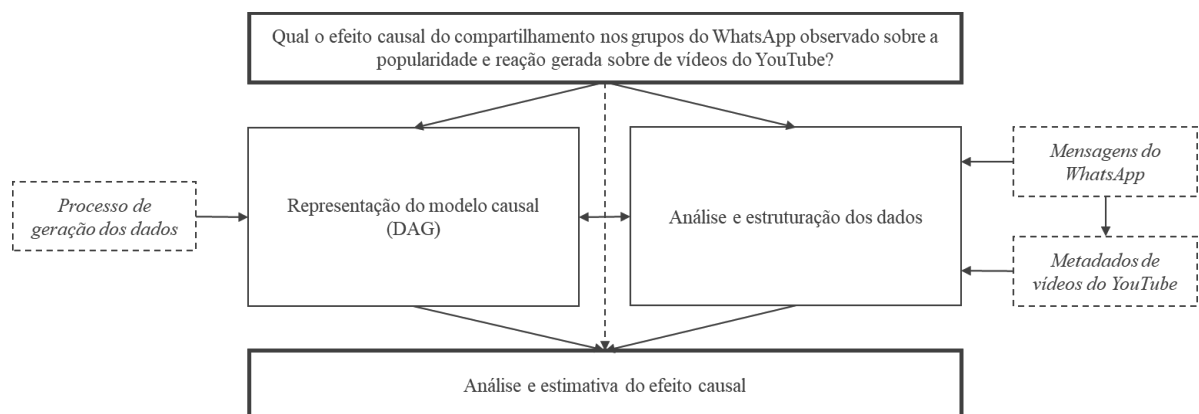
quase-experimento, com cuidado especial na definição e composição de grupos de tratamento e controle de forma a fazê-los o mais semelhantes o possível. Não é possível conhecer simultaneamente os desfechos potenciais de cada observação, nem distribuir os elementos da amostra entre tratamento e controle de forma aleatória. Mas pode-se seleccionar elementos de forma justificada e planejada para que os resultados estimados sejam coerentes, robustos e confiáveis.

O método de trabalho adotado nesse estudo considerou as seguintes atividades:

1. Representação do processo de geração dos dados através de diagramas causais, com identificação de variáveis relevantes e, principalmente, daquelas a serem ajustadas/controladas para isolamento do efeito da intervenção sobre os desfechos;
2. Análise e estruturação dos dados em grupos de tratamento e controle a partir de mensagens postadas nos grupos do WhatsApp e de metadados de vídeos coletados através da YouTube Data API;
3. Controle de covariantes com base no critério de portas dos fundos e estimativa do efeito causal da variável de tratamento sobre as variáveis de desfecho através de técnicas de *matching*.

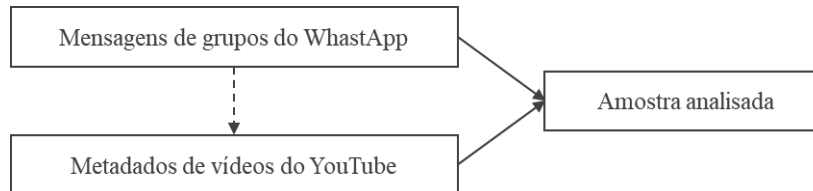
A interação entre as etapas está representada na **Figura 5**.

Figura 5. Principais etapas do trabalho.



A amostra de dados analisada nesse estudo foi gerada a partir de dois conjuntos distintos: o primeiro formado por mensagens postadas em grupos públicos do WhatsApp e o segundo formado por metadados dos vídeos do YouTube mencionados nessas mensagens.

Figura 6. Fluxo simplificados dos conjuntos de dados da origem até a amostra selecionada.



4.1 Representação do processo de geração dos dados

Todas as relações causais relevantes para explicação do efeito de W sobre Y_V e Y_E foram representadas através de grafos acíclicos dirigidos (DAG). O DAG permite representar de forma clara e objetiva o conhecimento teórico que se tem sobre o fenômeno sob estudo.

Aplicando o critério de porta dos fundos (*backdoor criterion*) sobre os caminhos do DAG, foram identificadas as variáveis do modelo causal consideradas relevantes para que, uma vez controladas, permitem isolar os efeitos $W \rightarrow Y_V$ e $W \rightarrow Y_E$ e, dessa forma, estimar os efeitos do compartilhamento de vídeos do WhatsApp sobre número de visualizações e quantidade de engajamento. Da mesma forma, as variáveis descritas e explicadas pelo DAG foram consideradas na construção do conjunto de dados representativos dos grupos de tratamento e controle. O modelo causal é apresentado em detalhes no Capítulo 5.

4.2 Canais do YouTube como agrupamentos naturais

Vídeos publicados em um mesmo canal do YouTube tendem a possuir atributos em comum e a serem mais semelhantes entre si do que quando comparados com vídeos publicados em canais diferentes. Por exemplo, os vídeos de um canal costumam abordar assuntos parecidos, são divulgados aos mesmos usuários (assinantes) quando publicados, são planejados e produzidos pela mesma equipe de criadores de conteúdo, utilizam-se de linguagens semelhantes, sendo avaliados pelo mecanismo de recomendação do YouTube com base em metadados comuns.

Por essa razão, *o canal de publicação pode ser considerado um mecanismo natural de controle dos vídeos* e foi considerado como principal fator para composição dos grupos de tratamento e controle a partir de mensagens compartilhadas nos grupos do WhatsApp.

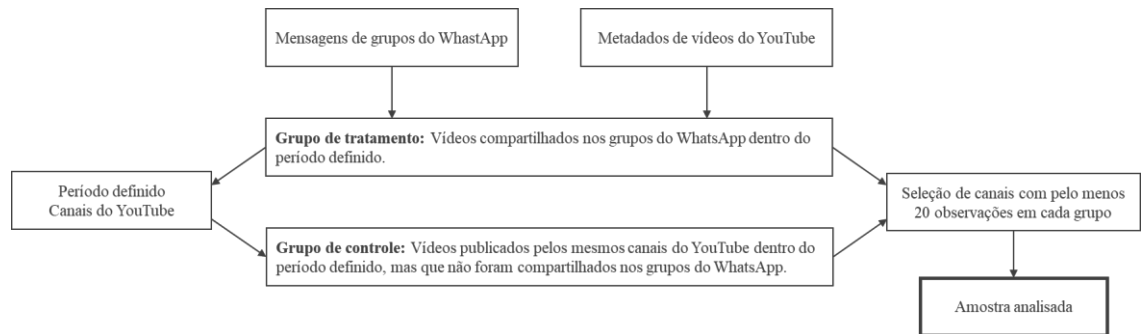
4.3 Composição dos grupos de tratamento e controle

A composição do grupo de tratamento foi realizada a partir dos vídeos mencionados nas mensagens de grupos do WhatsApp coletadas. Todos os vídeos observados no conjunto de mensagens podem ser considerados como tendo recebido a intervenção “compartilhados no WhatsApp” ($W = 1$). Partindo da mesma lógica, o grupo de controle foi composto pelos vídeos publicados pelos mesmos canais que publicaram os vídeos presentes no grupo de tratamento, mas que não receberam a intervenção, ou seja, que não foram compartilhados nos grupos do WhatsApp ($W = 0$). Cabe ressaltar que os vídeos publicados nos grupos de WhatsApp serviram de “proxy” para a seleção dos canais do YouTube onde o efeito causal do tratamento é investigado.

Como medida adicional, apenas vídeos publicados e compartilhados dentro de um mesmo período, comum aos grupos de tratamento e controle, foram considerados. A restrição a vídeos publicados na mesma janela de tempo permite supor que todos estiveram sujeitos a influências de mesmos eventos exógenos. A seleção dos mesmos canais possibilita comparar vídeos produzidos dentro de mesmas condições, em termos de linguagem, orientação, tópicos e recursos de produção, *i.e.* controlando variáveis conhecidas e desconhecidas associadas ao perfil do conteúdo produzido pela canal e pelo seu grupo de seguidores.

Assim, os grupos de tratamento e controle compostos artificialmente podem ser definidos da seguinte forma:

- Grupo de tratamento – vídeos publicados em conjunto definido de canais no YouTube e compartilhados nos grupos de WhatsApp dentro de um período definido;
- Grupo de controle – vídeos publicados no YouTube dentro do mesmo período definido e pelos mesmos canais dos vídeos do grupo de tratamento, mas que não foram compartilhados no WhatsApp.

Figura 7. Esquema simplificado de seleção da amostra de dados analisada.

A seleção foi restrita a canais que possuíam ao menos 20 vídeos em cada grupo (tratamento e controle) em maio/2021, o que resultou em um total de 6.102 vídeos associados a 42 canais. O processo de construção da amostra é apresentado na **Figura 7** e está descrito de forma detalhada no capítulo 6.

4.4 Estimativa do efeito causal

Com base no modelo causal definido para o fenômeno em estudo, aplicou-se o critério de portas dos fundos para identificação e seleção das variáveis a serem controladas para eliminação de eventual viés de seleção nas amostras, possibilitando que o efeito causal intervenção-desfecho pudesse ser isolado.

Para aproximar os dados observacionais utilizados no estudo de dados que seriam obtidos através de experimentos aleatorizados, foram aplicados dois métodos de *matching* distintos: *matching* por distância euclidiana e *coarsened exact matching* (CEM). (Optou-se por manter o nome do método em inglês ao longo do texto uma vez que não foi identificada tradução comumente adotada na língua vernácula.) Ambos os métodos buscam, por vias diferentes, permitir a identificação de um quase-experimento a partir de dados observacionais.

O método de *matching* por distância euclidiana possibilita a construção de grupos de tratamento e controle balanceados para as variáveis de controle. Para isso, cada observação do conjunto de dados é representada como um vetor formado pelos respectivos valores para cada variável de controle. Em seguida, para cada unidade do grupo de tratamento (submetida à intervenção), busca-se seu vizinho mais próximo dentre as unidades do grupo de controle (não submetidas à intervenção). A distância considerada para fins de proximidade, neste caso, foi a distância euclidiana. Ao final do procedimento, obteve-se pares de unidades tratamento-controle semelhantes em termos de covariantes, e apenas distintas em relação ao recebimento

ou não da intervenção. Dessa forma, o vizinho mais próximo no grupo de controle se aproxima do “desfecho potencial alternativo não observável” do seu par do grupo de tratamento. O efeito causal foi estimado como a média do efeito observado nos pares individuais.

No *coarsened exact matching*, as variáveis de controle foram transformadas em categorias e as observações foram distribuídas em grupos de forma que, em cada um deles, todas as observações possuissem exatamente o mesmo valor para todas as variáveis de controle após “categorizadas”. Em seguida, estratos que não contivessem ao menos uma unidade de tratamento e uma unidade de controle foram considerados incompletos e foram descartados. O efeito causal foi estimado como a diferença das médias entre as observações de tratamento e de controle. Para que as médias fossem comparáveis, os valores associados às variáveis de controle foram ponderados com base na diferença das quantidades de unidades de tratamento e controle dentro de cada estrato e em todos os estratos.

Uma análise do balanceamento das variáveis de controle após o *matching* demonstrou que o conjunto de dados se encontrava em condições de comparação ou *ceteris paribus*.

Dessa forma, esses métodos permitiram estimar os efeitos isolados da intervenção W sobre os desfechos Y_V e Y_E e, assim, chegar a conclusões sobre quanto o compartilhamento dos vídeos nesses grupos do WhatsApp contribuiu para o aumento da sua capacidade de geração de receitas dos respectivos canais do YouTube.

5 PROPOSTA DE MODELAGEM

Neste capítulo, o processo teórico que gerou os dados analisados é contextualizado e são apresentados os modelos causais propostos, bem como as premissas que orientaram os seus desenvolvimentos. Começa-se descrevendo o “fenômeno de interesse observado”, destacando relações relevantes entre diferentes variáveis envolvidas. No final, são descritas as relações causais representadas nos grafos acíclicos direcionados (DAG), que, posteriormente, orientarão o restante da análise a fim de identificar relações relevantes para isolamento do efeito do compartilhamento de vídeos no WhatsApp (W) sobre o número de visualizações (Y_V) e quantidade de engajamento (Y_E).

5.1 O processo de geração dos dados

5.1.1 Grupos públicos do WhatsApp como rede social

Em 2018, as plataformas de redes sociais foram amplamente utilizadas durante as eleições brasileiras. Dentre elas, o WhatsApp e o YouTube se destacaram por razões distintas. Enquanto o WhatsApp foi adotado como principal rede de comunicação das equipes de campanha, o YouTube se estabeleceu como principal plataforma de publicação de vídeos (MONT’ALVERNE e MITOZO, 2019).

Apesar do WhatsApp não ser uma plataforma de redes sociais em termos estritos, no Brasil, a proliferação e uso generalizado de grupos temáticos abertos ao público fez com que a ferramenta de mensagens instantâneas adquirisse características de fóruns públicos de conversa.

Dessa forma, o WhatsApp foi amplamente utilizado como ferramenta de comunicação e marketing durante a campanha presidencial de 2018. Parte dessas comunicações se deu em grupos públicos, divulgados abertamente em catálogos especializados ou através das próprias redes sociais de partidos e candidatos (RESENDE et. al., 2018; CAETANO et. al., 2018; BURSZTYN e BIRNBAUM, 2019).

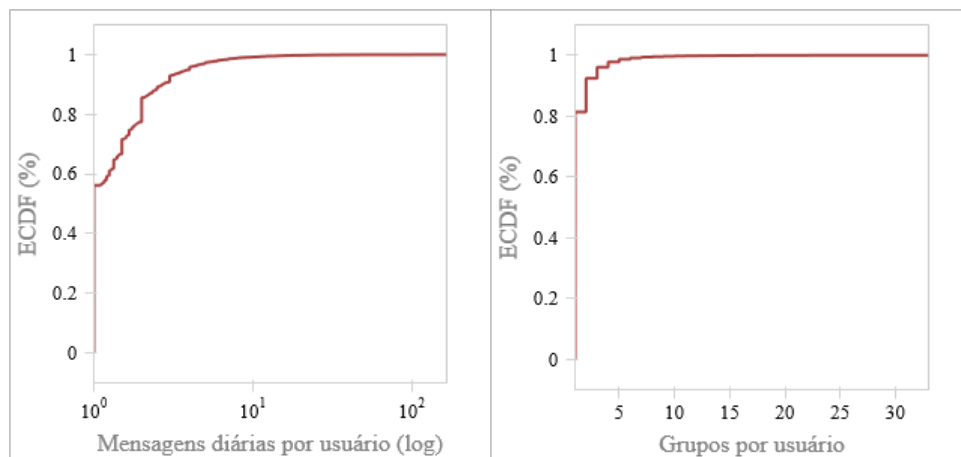
Qualquer interessado podia solicitar o ingresso em um desses grupos. Uma vez em um grupo, a pessoa passava a receber todas as mensagens postadas e tinha a possibilidade de postar mensagens a outros integrantes. Convites para novos grupos também eram divulgados nos grupos existentes e uma mesma pessoa podia participar de vários grupos ao mesmo tempo. Se por um lado, passaria a receber mensagens repetidas, por outro, poderia compartilhar conteúdos com mais pessoas de uma só vez.

Apesar de cada grupo ter “existência própria”, é coerente considerá-los interconectados, seja através de usuários ou de conteúdos comuns, por exemplo. Por essa razão, o conjunto desses grupos é tratado como uma rede, mesmo que as ligações entre os grupos não existam de forma explícita na plataforma. Essas ligações são consideradas “fracas” uma vez que nem todos os participantes de um grupo conhecem os demais grupos “vizinhos”.

Os dados considerados no estudo incluem postagens realizadas em 230 grupos públicos do WhatsApp durante um período de 6 semanas, de 10/09/2018 a 21/10/2018. Nesse período, cada grupo recebeu 243 mensagens em média por dia.

A distribuição da atividade dos usuários apresenta uma cauda longa com assimetria bastante acentuada à direita como se vê na **Figura 8**. Enquanto mais da metade dos usuários enviou uma média de até uma mensagem por dia, um pequeno grupo de 22 usuários chegou a enviar mais de 100 mensagens diárias em média. Da mesma forma, enquanto cerca de 80% dos usuários participaram de apenas 1 grupo, um grupo de apenas 64 usuários postou mensagens em mais de 10 grupos diferentes.

Figura 8. Média de mensagens e grupos por usuário.



Esse fenômeno de cauda longa em distribuições de atividades em redes sociais é comum e foi observado anteriormente em diferentes contextos (CLAUSET et. al, 2009). Considerando que os grupos observados fizeram parte de estruturas de campanha e militância política, é natural imaginar que determinados usuários atuaram ativamente na distribuição de informações e comunicações alinhadas aos objetivos dos grupos.

À época da coleta, de acordo com as regras do WhatsApp, cada grupo podia comportar no máximo 256 membros. Como os dados disponíveis para o estudo incluem apenas as mensagens postadas e seus respectivos remetentes, não foi possível aferir quantas pessoas

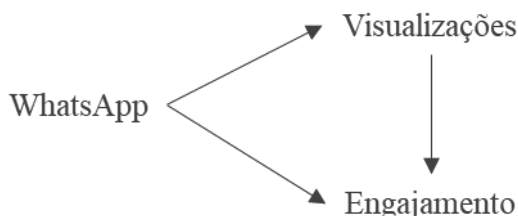
efetivamente receberam cada conteúdo. Sabe-se apenas que o alcance de uma informação em uma determinada conversa estava limitado ao número máximo de 256 pessoas.

Apesar disso, é preciso considerar que os participantes desses grupos participam de outros grupos em suas esferas privadas de relacionamento: família, trabalho, escola, igreja, academia etc. Assim, a rede observada é presumida aqui como apenas a “ponta de um iceberg”, no sentido que, por maior que possa parecer, seu verdadeiro alcance ocorreu fora da nossa capacidade de observação, à medida em que os membros poderiam encaminhar mensagens a seus grupos e contatos privados.

Portanto, pode-se assumir que essa “rede de grupos” em específico teve uma grande capacidade de engajamento e que conteúdos que foram compartilhados nesses grupos podem ter recebido mais atenção do que outros conteúdos semelhantes que não tiveram a mesma “sorte”. Independentemente das intenções e objetivos que motivaram os compartilhamentos, deseja-se entender até que ponto o simples fato de terem sido compartilhados causou um maior número de acessos.

Essas relações estão representadas na **Figura 9**.

Figura 9. Influência do WhatsApp sobre a atenção recebida por conteúdos na Internet.



5.1.2 Grande volume de URLs para vídeos do YouTube

Com relação aos tipos de conteúdo que circularam nos grupos do WhatsApp, de acordo com Bursztyn e Birnbaum (2019), 44% das mensagens postadas nos grupos (N= 1.243.149) continham conteúdo multimídia (imagem, vídeo ou áudio), enquanto 12% (N=329.804) apresentavam alguma URL externa, sendo que 55% dessas apontavam para vídeos publicados no YouTube (N=179.586). Cerca de 1 em cada 20 mensagens faziam menção a um vídeo publicado no YouTube.

Foi exatamente essa observação que estimulou a questão “qual teria sido o efeito desses compartilhamentos sobre a quantidade de atenção recebida pelos vídeos?”. Principalmente, porque, considerando que o YouTube é uma plataforma que remunera os produtores de

conteúdo, o uso de uma rede de grupos engajada como essa poderia ser uma forma de impulsionar a monetização de determinados canais.

Importante destacar que não se considerou que os usuários que compartilharam mensagens nos grupos do WhatsApp tenham sido, necessariamente, os respectivos autores. Principalmente no caso de conteúdo multimídia e URL, assume-se que a grande maioria dos conteúdos foi produzida por terceiros e compartilhada na rede por algum dos membros.

Mesmo no caso de usuários com alto volume de postagens, entende-se que não estavam, necessariamente, envolvidos com a produção dos conteúdos, apesar do padrão da atividade sugerir um interesse maior dessas pessoas na difusão das informações. É perfeitamente plausível que esse grupo de usuários tenha atuado como “curadores” de conteúdos produzidos de maneira independente como sugerido por Mello (2020).

5.1.3 Produção de vídeos para o YouTube

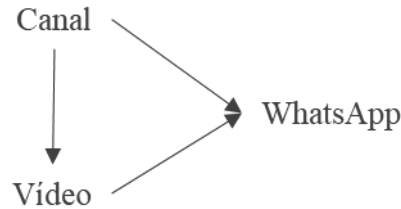
Em relação ao processo de produção dos vídeos, não se imagina que tenha sido impactado em função das atividades nos grupos do WhatsApp. Em outras palavras, apesar de se considerar que um vídeo tenha recebido mais atenção por ter sido compartilhado na rede de grupos, não se assume que um vídeo tenha sido produzido por causa dos grupos do WhatsApp.

No YouTube, criadores de conteúdo publicam vídeos em canais, que são espaços particulares voltados à divulgação de vídeos e geralmente associados a uma temática específica. Qualquer pessoa, natural ou jurídica, pode criar canais e publicar vídeos sem custo. Grupos de mídia costumam utilizar canais para divulgação de conteúdos relacionados com a sua programação. Empresas e organizações usam seus canais para divulgação de vídeos institucionais ou com informações relacionadas a seus produtos e projetos. Artistas e influenciadores digitais (*youtubers*) utilizam a plataforma para divulgação de seus trabalhos e para falar diretamente com seu público-alvo. Políticos adotam a plataforma para manter uma comunicação direta com seus eleitores e como canal para prestação de contas de seus mandatos.

Os vídeos compartilhados nos grupos do WhatsApp tiveram origem em canais diferentes, alguns especificamente ligados a candidatos e a veículos de comunicação políticos. Contudo, muitos canais que tiveram vídeos compartilhados não tinham relação direta com o “evento eleições” e pertenciam a artistas, influenciadores ou mesmo a particulares. Mesmo assim, entende-se que tanto o conteúdo dos vídeos quanto o tema do canal foram determinantes para que fossem compartilhados.

Essas relações estão representadas na **Figura 10**.

Figura 10. Influência de canais e vídeos sobre o compartilhamento no WhatsApp

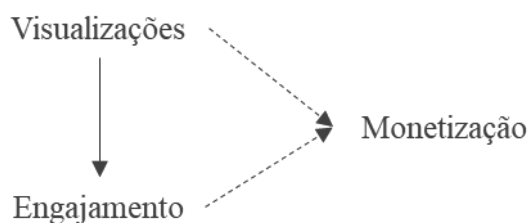


5.1.4 Geração de receitas no YouTube

Apesar do YouTube oferecer outros meios de geração de receita, a fonte mais simples e utilizada de monetização é a habilitação de inserção de anúncios durante a exibição de vídeos. Um canal com mais de 1.000 inscritos e mais de 4.000 horas de exibição nos 12 meses anteriores está apto a monetizar seus conteúdos através da veiculação de anúncios do Google AdSense (desde que não produza conteúdos em desacordo com as políticas do YouTube). O AdSense é o serviço de publicidade do Google. Por razões compreensíveis, a remuneração dá-se com base nos anúncios efetivamente visualizados. Como a plataforma permite que usuários “pulem” um anúncio, a simples inserção de uma propaganda não garante que ela será visualizada.

Mesmo sem acesso à renda efetivamente gerada por um canal, pode-se considerar que um vídeo colabora para a receita na proporção do número de visualizações e da quantidade de engajamento que gera. Quanto mais visualizações, mais anúncios serão exibidos. Quanto mais engajamento gerado, maiores as chances do anúncio ser efetivamente visto. Como não é possível observar a monetização, as influências sobre ela são representadas com setas tracejadas na **Figura 11**.

Figura 11. Efeitos do número de visualizações e engajamento sobre a monetização.

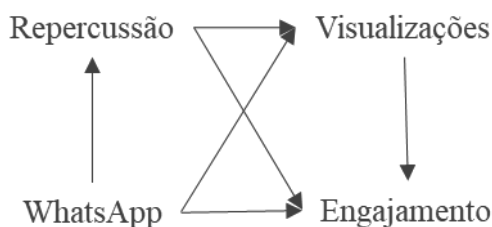


5.1.5 Repercussão dos vídeos na Internet

Para além do que é possível observar, entende-se que todo vídeo publicado no YouTube gera alguma repercussão. Define-se como repercussão o impacto que o vídeo causa sobre as pessoas em sentido geral, e que se reflete tanto em visualizações quanto em engajamento. Seja ela fruto do conteúdo do vídeo, ou consequência natural do canal que o publica, é razoável considerar que a repercussão também influencie decisões de compartilhar (ou não) um determinado vídeo nos grupos do WhatsApp.

No entanto, no contexto específico dos dados analisados, é preciso considerar que, enquanto a decisão de compartilhar ou não um vídeo no WhatsApp ocorreu em um momento passado (2018), a repercussão observada foi estimada em um momento futuro (2021). Dessa forma, o processo de geração dos dados permite considerarmos apenas a influência causada pelo compartilhamento no WhatsApp sobre a Repercussão, e podemos desconsiderar possíveis efeitos de causalidade reversa da Repercussão sobre a decisão de compartilhamento dos vídeos. Essa relação é apresentada na **Figura 12**.

Figura 12. Efeitos da repercussão do vídeo sobre variáveis de desfechos.



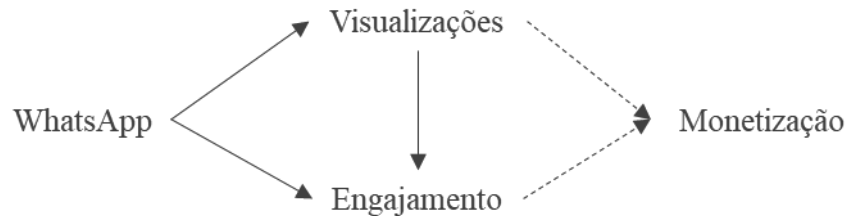
5.2 Diagramas causais

O estudo busca estimar o efeito que o compartilhamento dos vídeos nos grupos do WhatsApp teve sobre a quantidade de visualizações e engajamentos que receberam e, conseqüentemente, sobre o aumento da capacidade de monetização dos canais que os publicaram. Portanto, considerando que essa relação de impulsionamento guarda um efeito causal, acredita-se que, se não fosse pelo compartilhamento, a monetização desses vídeos teria sido menor. Em outras palavras, partiu-se da premissa de que o compartilhamento de um vídeo no WhatsApp (W) tem um efeito positivo sobre a sua monetização (M).

A monetização, apesar de ser uma quantidade que não pode ser observada, é influenciada pelo número de visualizações do vídeo (Y_V) e por sua quantidade de engajamento

(Y_E). Como o engajamento também sofre influência do número de visualizações, existe uma relação entre essas três variáveis como apresentado na **Figura 13**.

Figura 13. Relação entre intervenção e desfechos.



Dado que não é possível observar a monetização, a questão geral do estudo foi dividida em duas perguntas mais específicas:

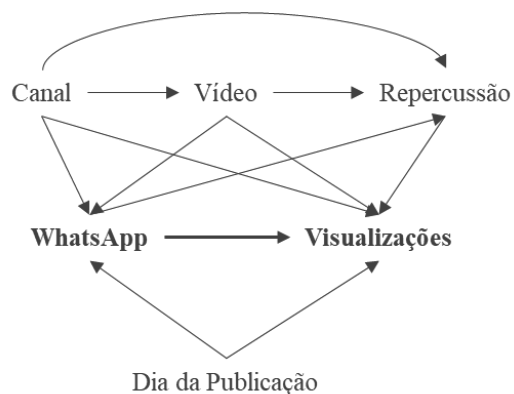
- Qual o efeito causal do compartilhamento nos grupos do WhatsApp observado sobre a popularidade de vídeos do YouTube, medida pelo número de visualizações?
- Qual o efeito causal do compartilhamento nos grupos do WhatsApp observado sobre a reação gerada por vídeos do YouTube, medida pela quantidade de engajamento?

Um ponto importante: entende-se que não existe uma influência do compartilhamento no WhatsApp sobre a monetização além daquela que acontece através das visualizações e do engajamento. Ainda, considerou-se que, no contexto do estudo, nenhuma outra variável gera influência sobre a monetização que não seja através da visualização (Y_V) e engajamento (Y_E). Com base nessas premissas, é possível remover a monetização do diagrama sem impactar a relação entre as demais variáveis. Com a remoção da monetização do modelo, torna-se possível trabalhar apenas com a estimativa do efeito do WhatsApp sobre as duas variáveis de desfecho, o que será abordado de forma separada.

5.2.1 Número de visualizações (Y_V)

A **Figura 14** apresenta o diagrama causal completo, representando a influência do compartilhamento nos grupos do WhatsApp sobre o número de visualizações dos vídeos no YouTube.

Figura 14. Diagrama causal da influência do WhatsApp sobre visualizações.



A variável Canal (C) representa todos os fatores inerentes ao canal de publicação do vídeo que afetaram tanto a chance de compartilhamento nos grupos do WhatsApp quanto o número de visualizações recebidas no YouTube, como a popularidade do canal, seu tempo de existência, categoria, volume de produção, linha editorial, linguagem etc. De forma semelhante, a variável Vídeo (V) representa características específicas do vídeo com impacto sobre o compartilhamento (W) e o número de visualizações (Y_V), como, por exemplo, o assunto abordado, se foi uma transmissão ao vivo (*live*) etc. O caminho $C \rightarrow V$ representa a influência que um canal exerce sobre as características do vídeo que publica.

A variável repercussão (R) do vídeo representa o impacto causado pelo vídeo nas redes sociais em geral. Entende-se que a repercussão é consequência não só das características inerentes ao vídeo, sendo também influenciada pelo canal de publicação. Conteúdos de maior repercussão foram naturalmente mais visualizados. Da mesma forma, a repercussão também foi impactada pelo compartilhamento dos vídeos no WhatsApp. Na prática, a variável Repercussão busca reunir todos os fatores não observados inerentes ao “comportamento das redes sociais”, assumindo uma função importante no modelo. É através dela que se faz possível separar o efeito direto causado pelo WhatsApp do efeito indireto causado por outras redes sociais.

Por fim, como se observou um fator temporal no compartilhamento de vídeos em determinados canais, a variável Dia de Publicação (D) foi considerada para indicação do momento em que a publicação do vídeo foi realizada. Como o fenômeno se relaciona com eventos bem determinados no tempo, por exemplo, a data de votação ou o dia em que um debate foi realizado, é natural imaginar que, dependendo da data, um vídeo teve mais chances de ser compartilhado ou visualizado.

As variáveis Canal, Vídeo e Dia da Publicação abrem portas dos fundos (*backdoors*) e precisam ser controladas para que se possa isolar o efeito direto $W \rightarrow Y_V$:

- $W \leftarrow C \rightarrow Y_V$
- $W \leftarrow V \rightarrow Y_V$
- $W \leftarrow D \rightarrow Y_V$

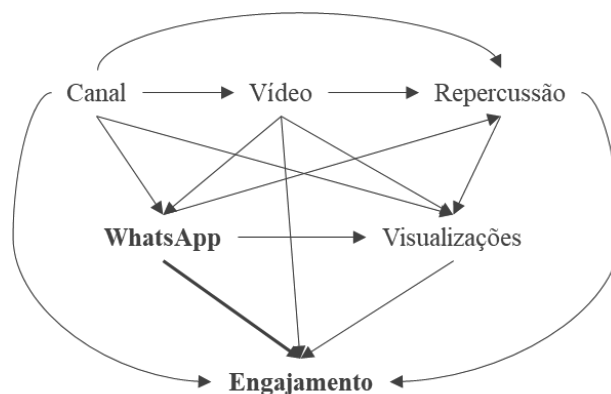
A variável Repercussão exerce uma função de mediação $W \rightarrow R \rightarrow Y_V$ entre a variável de intervenção e de desfecho e foi incluída no conjunto de variáveis de controle uma vez que estamos interessados apenas nos efeitos diretos $W \rightarrow Y_V$

Uma última consideração, ao observar a **Figura 13**, percebe-se a existência do caminho $W \rightarrow Y_E \leftarrow Y_V$ que foi omitido na **Figura 14**. O engajamento (Y_E) é uma variável de colisão (*collider*) entre W e Y_V e como tal não deve ser controlada. Se o engajamento fosse incluído na nossa estratégia de controle para o número de visualizações, um caminho de dependência entre W e Y_V que não existe no modelo causal seria criado.

5.2.2 Quantidade de engajamento (Y_E)

A **Figura 15** apresenta o diagrama causal completo para a influência do compartilhamento nos grupos do WhatsApp sobre a quantidade de engajamento recebida pelos vídeos.

Figura 15. Diagrama causal da influência do WhatsApp sobre engajamento.



Pode-se perceber que as relações na **Figura 15** são muito semelhantes às apresentadas na **Figura 14** com duas diferenças principais. Ao se analisar a influência do compartilhamento no WhatsApp (W) sobre o engajamento (Y_E), o número de visualizações (Y_V) torna-se uma nova variável de confusão (*confounder*) que precisa ser controlada para isolamento do efeito direto

de W sobre Y_E . Por outro lado, a variável Dia de Publicação (D) não exerce influência sobre Y_E . Não faz sentido imaginar que um usuário deixaria de comentar ou curtir um vídeo pelo simples fato de ser quinta ou sexta-feira, por exemplo. Por essa razão, a variável T foi desconsiderada no modelo uma vez que o caminho secundário (*backdoor*) $W \leftarrow T \rightarrow Y_V \rightarrow Y_E$ será fechado quando o modelo for controlado pela variável Y_V .

É possível identificar os seguintes caminhos de portas dos fundos (*backdoors*) que precisam ser controladas para isolamento o efeito direto $W \rightarrow Y_E$:

- $W \leftarrow C \rightarrow Y_E$
- $W \leftarrow V \rightarrow Y_E$

Além destes, consideramos a Repercussão e o número de visualizações no conjunto de variáveis de controle para eliminar os efeitos de medição causados por ambas as variáveis entre a variável de intervenção e de desfecho $W \rightarrow Y_E$.

6 ESTRUTURAÇÃO DOS DADOS

Neste capítulo descreve-se o processo de coleta, organização e estruturação do conjunto de dados utilizado para a análise da influência do compartilhamento de vídeos do WhatsApp sobre o número de visualizações e taxa de engajamento atingidos no YouTube. Algumas análises preliminares acerca das características das variáveis consideradas são apresentadas.

6.1 Construção do conjunto de dados

6.1.1 Coleta original de dados do WhatsApp (2018)

Os dados de mensagens do WhatsApp utilizados neste trabalho derivam de conjunto coletado originalmente por Bursztyn e Birnbaum (2019). Entre agosto e outubro de 2018, os autores coletaram um total de 2.822.686 mensagens a partir de 230 grupos públicos de WhatsApp relacionados com a campanha presidencial brasileira.

A **Tabela 1** sumariza estatísticas gerais da base de dados. De um total de 2.822.686 mensagens coletadas, 136.620 (4,8%) continham uma ou mais URLs para vídeos do YouTube, resultando em 172.144 menções únicas. No total, 26.968 vídeos diferentes foram mencionados por 12.422 usuários em 230 grupos do WhatsApp.

Tabela 1. Estatísticas gerais dos vídeos mencionados nos grupos de WhatsApp.

# de mensagens	2.822.686
# de mensagens com link para vídeo do YouTube	136.620
# de menções individuais a vídeos do YouTube	172.144
# de vídeos	26.968
# de usuários	12.422
# de grupos	230

Para este estudo, considerou-se apenas as mensagens que continham links do YouTube e extraiu-se uma relação de todos os vídeos mencionados ao menos uma vez no conjunto original e a informação de quando foram mencionados.

6.1.2 Recorte temporal

Selecionou-se apenas mensagens postadas durante as 6 semanas centrais do período coberto pela coleta original (10/09/2018 a 21/10/2018). Dessa forma tentou-se garantir que todos os grupos de WhatsApp observados estivessem representados ao longo de todo o período.

Antes de 10/09/2018, parte dos grupos ainda não estava presente por conta do processo de inserção incremental adotado na coleta (Bursztyrn e Birnbaum, 2019). A partir de 21/10/2018 percebe-se a ausência de atividades em alguns grupos, possivelmente por conta do encerramento das campanhas de candidatos derrotados no 1º turno.

6.1.3 Seleção dos grupos de tratamento e controle

Como mencionado no capítulo 3, na impossibilidade de realização de um experimento aleatório controlado, é preciso buscar um conjunto de observações que se aproxime ao máximo dos contrafactuais que não podem ser observados.

Ao conferir o status dos vídeos mencionados nas mensagens, percebeu-se que, após quase 3 anos da coleta original, apenas metade dos vídeos ainda seguiam disponíveis no YouTube. Cerca de 45% dos vídeos haviam sido removidos, a maior parte sem indicação da razão associada (a mensagem do YouTube indica apenas “Vídeo Indisponível”).

Figura 16. Evolução da situação dos vídeos entre 2018 e 2021.

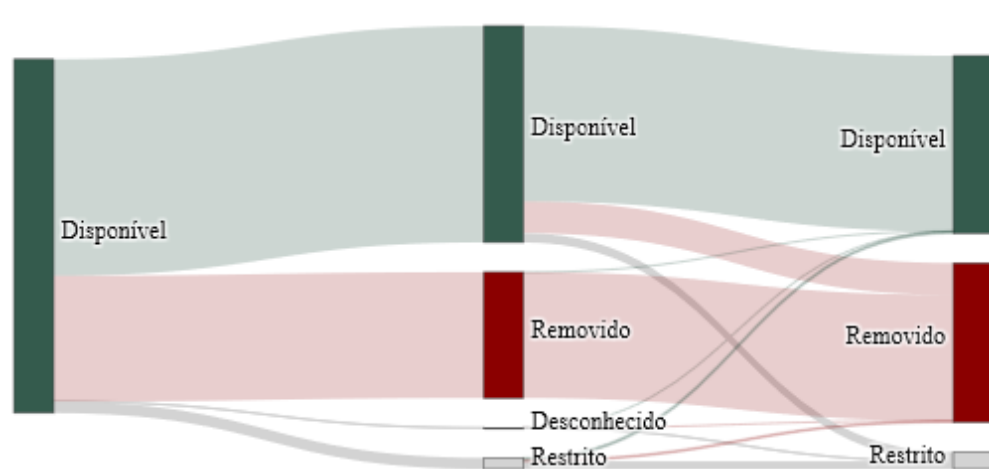


Tabela 2. Evolução da situação dos vídeos de 2018 a 2021

	Em DEZ/2019	%	Em MAI/2021	%
# de vídeos Disponíveis	16.501	61%	13.532	50%
# de vídeos Removidos	9.599	36%	12.142	45%
# de vídeos Restritos	855	3%	1.294	5%
# de vídeos c/ situação Desconhecida	13	0%	-	-
# total de vídeos	26.968	100%	26.968	100%

Para construção do grupo de tratamento, foram identificados todos os vídeos compartilhados no período selecionado que ainda se encontravam disponíveis no YouTube, junto a informações dos respectivos canais. Em seguida, para formação do grupo de controle,

foram selecionados todos os vídeos publicados pelos mesmos canais dos vídeos do grupo de tratamento, mas que não foram mencionados em mensagens do WhatsApp. Os vídeos deste último grupo (controle) foram considerados como contrafactuais potenciais daqueles que foram compartilhados no WhatsApp, selecionados no primeiro grupo (tratamento).

6.1.4 Descarte de canais com poucas observações

Para minimizar distorções por questões de tamanho de amostras, foram descartados os canais que possuíam menos de 20 vídeos publicados e compartilhados nos grupos do WhatsApp no período selecionado de 10/09/2018 a 21/10/2018, ou menos de 20 vídeos publicados no mesmo período, mas que não foram compartilhados nos grupos do WhatsApp. Vídeos de canais que não atingiram essas quantidades foram desconsiderados.

6.1.5 Coleta de metadados do YouTube

Os metadados dos vídeos remanescentes no conjunto de dados foram coletados através de consultas à YouTube Data API para obtenção do título do vídeo e quantidades de visualizações, *likes*, *dislikes* e comentários. Essas três quantidades foram utilizadas para cálculo da quantidade de engajamento obtida por cada vídeo.

6.1.6 Descarte de observações sem valores

Ao final, um pequeno conjunto de vídeos (2%) que não possuíam visualizações ou engajamento foi descartado. Com isso foi possível garantir uma amostra em que todas as observações possuem algum valor para as variáveis de desfecho.

6.1.7 Considerações importantes

A amostra considera mensagens postadas durante um período específico. Os grupos onde foram postadas já existiam antes do início do período e continuaram existindo depois dele. Esse conjunto de grupos também não representa todos os grupos que cobriram o fenômeno “eleições brasileiras de 2018”.

Por isso, vídeos publicados dentro do período observado podem ter sido compartilhados nos grupos após 21/10/2018. Apesar de terem sido compartilhados no WhatsApp, eles aparecem como “não compartilhados” na amostra.

Os vídeos da amostra também podem ter sido publicados em outros grupos públicos do WhatsApp fora da rede de grupos observada. As análises e conclusões obtidas a partir deste conjunto de dados são limitadas ao efeito específico deste conjunto de grupos do WhatsApp, e não devem ser associados à plataforma WhatsApp como um todo.

Por fim, os dados também estão limitados aos vídeos “sobreviventes” no YouTube após quase 3 anos da sua publicação. Quase metade dos vídeos publicados em 2018 não estavam mais disponíveis em meados de 2021. Considerou-se que o efeito do compartilhamento no WhatsApp sobre os vídeos indisponíveis foi semelhante ao que se observa sobre os vídeos sobreviventes.

6.2 Descrição do conjunto de dados

O conjunto de dados final contém 6.024 observações e um conjunto de covariáveis representativo das variáveis relevantes consideradas nos diagramas causais da seção 5.2. Cada observação no conjunto de dados representa um vídeo publicado no YouTube que pode ter sido compartilhado nos grupos do WhatsApp ou não.

6.2.1 W – WhatsApp (variável de intervenção)

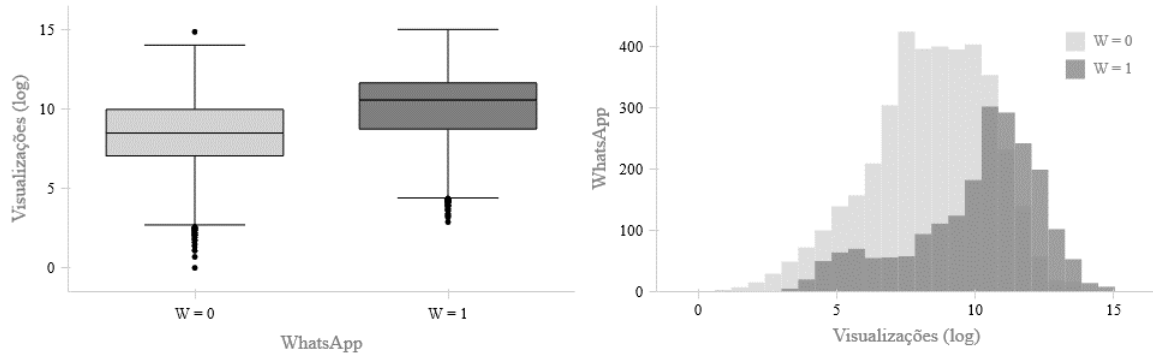
A variável WhatsApp (W) indica se o vídeo foi compartilhado nos grupos do WhatsApp ($W = 1$) ou não ($W=0$). O valor 1 foi atribuído aos vídeos observados nas mensagens compartilhadas nos grupos do WhatsApp. O valor 0 foi atribuído aos vídeos publicados pelos mesmos canais durante o mesmo período, mas que não foram observados nos mesmos grupos.

6.2.2 Y_V – Log(Visualizações) (variável de desfecho)

Essa é a primeira variável de desfecho do estudo e representa o número de visualizações que cada vídeo recebeu no YouTube, em escala logarítmica. A transformação em log atende a dois propósitos: aproxima os valores da distribuição em função da cauda longa característica e possibilita a interpretação da diferença entre os desfechos de forma relativa (ANGRIST e PISCHKE, 2014).

A **Figura 17** apresenta a distribuição observada para a variável entre os grupos de tratamento ($W = 1$; $\mu = 9,99$; $\sigma = 2,36$) e controle ($W = 0$; $\mu = 8,36$; $\sigma = 2,19$).

Figura 17. Distribuição do número de visualizações entre grupos de tratamento e controle.



6.2.3 Y_E – Log(Engajamento) (variável de desfecho)

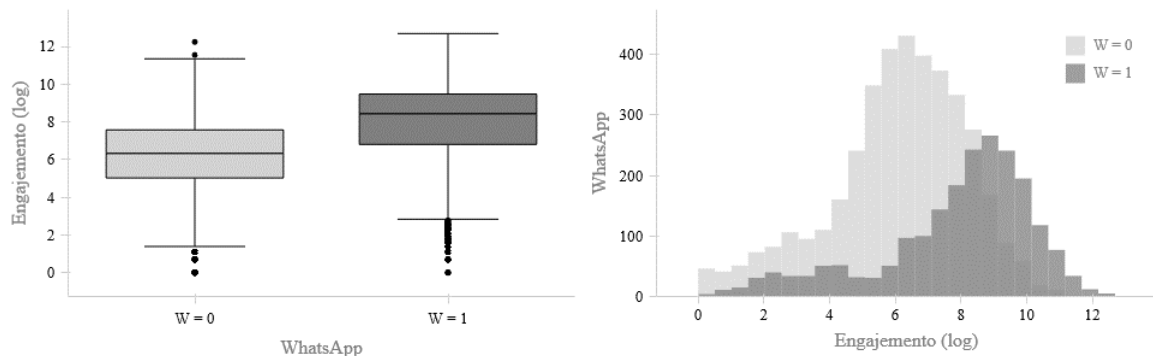
A quantidade de engajamento é a segunda variável de desfecho do estudo e representa a quantidade de interações recebida por cada vídeo, calculada da seguinte forma:

$$\text{Log(Engajamento)} = \text{Log}(\#likes + \#dislikes + \#comments)$$

As quantidades de *likes*, *dislikes* e comentários de cada vídeo foram obtidas a partir do YouTube. Assim como no caso do número de visualizações, a transformação em log aproxima os valores da distribuição e permite interpretar a diferença entre os desfechos de forma relativa (ANGRIST e PISCHKE, 2014).

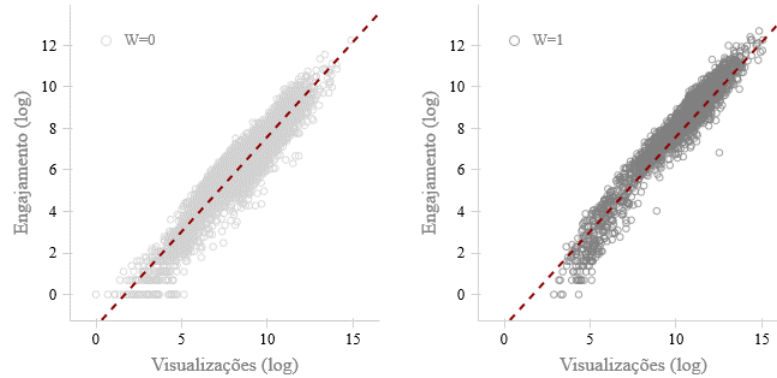
A **Figura 18** apresenta a distribuição observada para a variável entre os grupos de tratamento ($W = 1$; $\mu = 7,84$; $\sigma = 2,38$) e controle ($W = 0$; $\mu = 6,10$; $\sigma = 2,10$).

Figura 18. Distribuição da quantidade de engajamento entre grupos de tratamento e controle.



As duas variáveis de desfecho também são altamente correlacionadas entre si, como se vê na **Figura 19** a seguir. Dentro dos grupos de tratamento e controle a correlação é, respectivamente, de $r = 0,97$ e $r = 0,95$.

Figura 19. Correlação entre o número de visualizações e a quantidade de engajamento.



6.2.4 Repercussão

A variável categórica Repercussão busca representar, de forma aproximada, todo o impacto que um determinado vídeo causou nas redes sociais em geral. Uma vez que não é possível medir a repercussão em termos objetivos, a popularidade do vídeo, medida pelo número de visualizações recebidas, foi utilizada como *proxy*. Para evitar uma correlação perfeita com nossa variável de desfecho, a seguinte transformação foi aplicada:

$$\text{Repercussão} = \text{Inteiro}(\text{Log}_{10}(\text{Visualizações}))$$

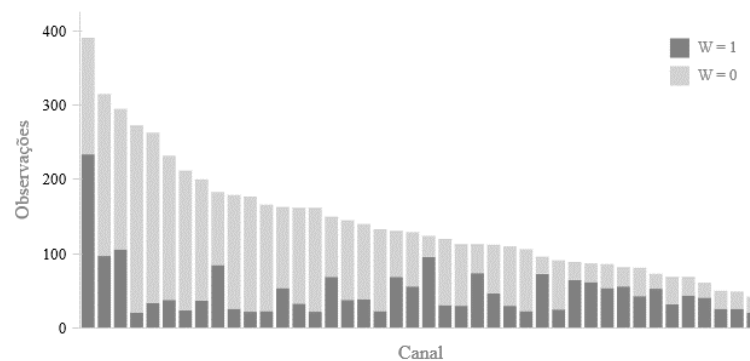
Dessa forma, a repercussão passa a ser tratada como uma “ordem de grandeza” **categórica** da atenção recebida por cada vídeo. Por exemplo, vídeos que receberam centenas de visualizações terão Repercussão = 2, e vídeos que receberam milhões de visualizações terão Repercussão = 5. Assim, torna-se possível comparar entre si apenas vídeos que geraram impactos de mesma proporção.

6.2.5 Canal

A identificação do canal é uma variável categórica com valor único para cada canal em que o vídeo foi publicado. O conjunto contém vídeos de 42 canais do YouTube diferentes.

A distribuição das observações entre os canais não é uniforme, como se vê na **Figura 20**. Alguns canais tiveram uma quantidade bastante superior de vídeos compartilhados do que outros. Também se observa que o conjunto de dados possui quase duas vezes mais vídeos no grupo de controle ($N = 3.922$; $\mu = 93,38$; $\sigma = 66,34$; μ e σ por canal) do que no grupo de tratamento ($N = 2.102$; $\mu = 50,05$; $\sigma = 36,99$; μ e σ por canal).

Figura 20. Distribuição da quantidade de vídeos por canal entre grupos de tratamento e controle.



6.2.6 Vídeo

Para caracterização dos vídeos, um conjunto de variáveis binárias (*dummy*) foi construído com o objetivo de representar diferentes fatores relevantes dentro do contexto das relações causais representadas nos diagramas da seção 5.2. Cada uma das variáveis recebeu valor 1 ou 0 indicando se a característica é observada ou não em cada vídeo. As seguintes variáveis foram incluídas no conjunto de dados representando características particulares:

Tabela 3. Variáveis binárias (*dummy*) representativas dos vídeos.

Alarmante	Indica se o título do vídeo busca causar alarme, com todas as letras em maiúscula ou incluindo pelo menos um dos seguintes termos: flagrante, grave, gravíssimo, gravíssima, bomba, explosivo, explosiva, explosivos, explosivas, importante, urgente, alerta, denuncia, atencao, polemica.
-----------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

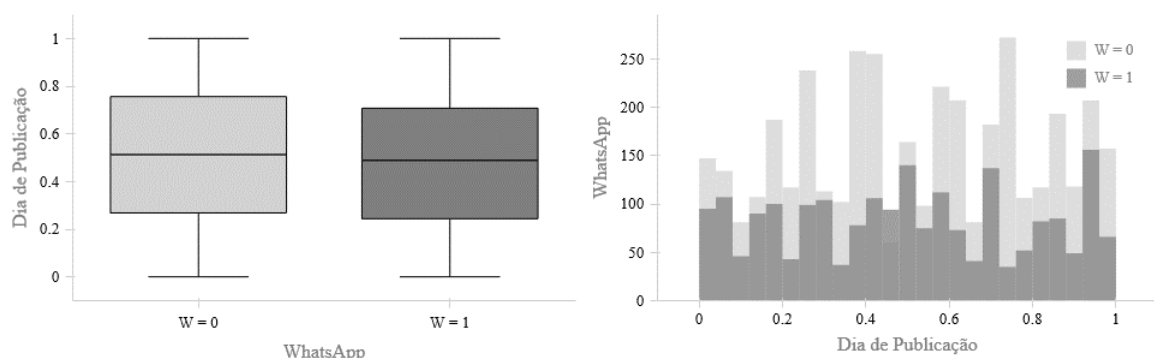
Eleições	Indica se o vídeo está associado com temas ligados ao processo eleitoral. O valor 1 foi atribuído quando o título continha pelo menos um dos seguintes termos: eleicoes, eleicao, eleitoral, turno, campanha, campanhas, carreata, carreatas, presidente, presidencial, presidenciais, debate, debates, pesquisa, pesquisas, urna, urnas, tse, voto, votos, ibope, datafolha, sensus.
Live	Indica se o vídeo foi uma transmissão ao vivo (<i>live</i>).
Grupo 1	Indica se o vídeo está associado a partidos e candidatos de um dos grupos políticos envolvidos na disputa eleitoral de 2018, definido como “grupo 1”.
Grupo 2	Indica se o vídeo está associado a partidos e candidatos de um dos grupos políticos envolvidos na disputa eleitoral de 2018, definido como “grupo 2”.

6.2.7 Dia de Publicação

A variável Dia de Publicação foi construída como um “ranking” da data de publicação de cada vídeo dentro do período observado, em uma escala de 0 a 1. Vídeos publicados no primeiro dia apresentam valor 0. Vídeos publicados no último dia apresentam valor 1. Vídeos publicados em outros dias aparecem distribuídos em posições intermediárias da escala.

A **Figura 21** apresenta a distribuição observada para a variável entre os grupos de tratamento ($W = 1$; $\mu = 3,83$; $\sigma = 1,05$) e controle ($W = 0$; $\mu = 3,14$; $\sigma = 0,99$). É possível observar que as distribuições são bastante homogêneas para os dois grupos (tratamento e controle).

Figura 21. Distribuição do dia de publicação entre grupos de tratamento e controle.



7 ANÁLISE DO EFEITO CAUSAL

Nesta seção são apresentados o desenvolvimento e resultados da análise do efeito causal do compartilhamento de vídeos em grupos públicos de WhatsApp sobre o número de visualizações e engajamento no YouTube. Cabe ressaltar que todas as análises e conclusões limitaram-se ao conjunto de dados descritos e estruturados no Capítulo 6.

Como as variáveis de desfecho do estudo estão representadas em escala logarítmica, os efeitos estimados podem ser interpretados como uma mudança percentual através da equação (7.1), em que δ representa o valor do efeito estimado como diferença entre logs e $\Delta\%$ indica a mudança percentual observada na variável de desfecho correspondente à mesma diferença.

$$\Delta\% = \exp(\delta) - 1 \quad (7.1)$$

Quando δ é menor que cerca de 0,2, os valores de δ e $\Delta\%$ são próximos o suficiente para que a própria diferença em log possa ser interpretada como uma mudança percentual aproximada (ANGRIST e PISCHKE, 2014).

7.1 Balanceamento da amostra antes de controle do viés de seleção

Como primeiro passo da análise, as distribuições das variáveis nos grupos de tratamento e controle foram comparadas para avaliar em que grau os vídeos distribuídos entre os grupos eram comparáveis entre si. A verificação do balanceamento é uma etapa importante sempre que se deseja estimar efeitos causais (ANGRIST e PISCHKE, 2014).

A **Tabela 4** compara valores médios de cada variável do modelo causal para o conjunto de vídeos pertencente aos grupos de tratamento e controle, separadamente. A coluna (1) apresenta valores médios para as características observadas no grupo de controle e a coluna (2) apresenta os valores correspondentes ao grupo de tratamento. A coluna (3) apresenta a diferença entre as colunas (2) e (1). Valores entre parênteses indicam o erro padrão da amostra em relação à população. Valores entre colchetes representam o desvio padrão observado na amostra. Diferenças destacadas em negrito na coluna (3) são significativas e, portanto, dificilmente seriam observadas em uma amostra como essa por acaso. Considerou-se como significativas, de maneira aproximada, diferenças superiores a duas vezes o desvio padrão.

Tabela 4. Balanceamento entre grupos de tratamento e controle antes de controle de covariantes.

	Controle (1)	Tratamento (2)	Tratamento – Controle (3)
A. Desfechos			
Visualizações (log)	8,36 [2,19]	9,99 [2,36]	1,63 (0,06)
Engajamento (log)	6,10 [2,10]	7,84 [2,38]	1,74 (0,06)
B. Covariantes			
Repercussão	3,14 [0,99]	3,83 [1,05]	0,69 (0,03)
Dia de Publicação	0,52 [0,28]	0,50 [0,29]	-0,02 (0,01)
Conteúdo alarmante	0,31	0,34	0,03 (0,01)
Conteúdo eleitoral	0,31	0,30	-0,01 (0,01)
Transmissão ao vivo	0,19	0,20	0,01 (0,01)
Menciona grupo 1	0,24	0,27	0,03 (0,01)
Menciona grupo 2	0,39	0,51	0,12 (0,01)
Número de observações (N)	3.922	2.102	6.024

No caso das variáveis *dummy*, os valores nas colunas (1) e (2) devem ser interpretados apenas como a proporção de unidades em que a variável possui valor = 1.

É possível observar no painel B da tabela anterior a existência de diferenças significativas (superior a 2 desvios padrão) entre os grupos de tratamento e controle para algumas das covariantes. Essas diferenças eram esperadas uma vez que os dados não provêm de um experimento aleatório controlado. Dessa forma, não é possível descartar a existência de viés de seleção entre as observações e os grupos não são comparáveis da forma como estão.

Para que seja possível estimar efeitos causais a partir de dados que não provêm de um experimento controlado, precisaremos garantir o controle prévio das covariantes, de forma a isolar os efeitos da variável de intervenção sobre as variáveis de desfecho.

7.2 Efeito do WhatsApp sobre o número de visualizações

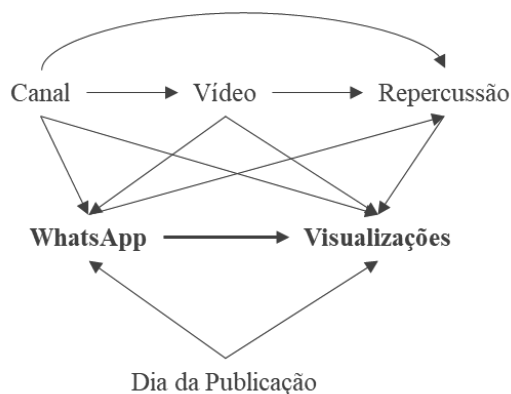
Nessa seção investigamos a primeira pergunta do estudo:

- Qual o efeito causal do compartilhamento nos grupos do WhatsApp observado sobre a popularidade de vídeos do YouTube, medida pelo número de visualizações?

Como forma de contornar o fato de os dados não serem provenientes de um experimento aleatório controlado, foram aplicados métodos de *matching* para a construção de grupos de tratamento e controle balanceados e comparáveis. Os pares resultantes dos métodos de *matching* se aproximam de pares que seriam observados em um experimento aleatório com bloqueio, i.e., em que as unidades do experimento são separadas em grupos (blocos) com características semelhantes. Assim, busca-se comparar a variável de desfecho apenas entre unidades de tratamento e controle semelhantes, considerando-se como semelhantes aquelas que possuem valores similares para as variáveis de controle.

As variáveis de controle foram selecionadas a partir do DAG representado na **Figura 22**. Esta figura foi apresentada detalhadamente na seção 5.2.1 e representa as premissas assumidas para o fenômeno sob estudo. Assim, para eliminar eventuais efeitos de confusão e isolar o efeito causal da variável intervenção WhatsApp sobre o número de visualizações (desfecho) é preciso controlar as variáveis Canal (C), Vídeo (V), Repercussão (R) e Dia de Publicação (D) para fechar os caminhos por portas dos fundos representados no modelo.

Figura 22. Diagrama causal da influência do WhatsApp sobre visualizações.



O efeito causal foi estimado através de dois métodos distintos, *matching* por distância euclidiana e *coarsened exact matching*.

7.2.1 Matching por distância euclidiana

Considerando-se a disposição de todas as observações (tratamento e controle) em um espaço euclidiano em que cada dimensão representa uma das variáveis de controle, o método de *matching* por distância euclidiana consiste em identificar, para cada unidade do grupo de tratamento, a unidade do grupo de controle mais próxima a ela, i.e. com menor distância euclidiana entre as duas.

As variáveis Canal (C) e Repercussão (R) foram utilizadas para fins de estratificação das observações. Dessa forma, o método de *matching* foi aplicado separadamente sobre cada estrato e, portanto, só foram apuradas distâncias entre vídeos publicados em mesmos canais e com mesmo grau de repercussão. Com isso, a distância entre duas observações i e j de um canal pode ser dada pela equação (7.2) a seguir:

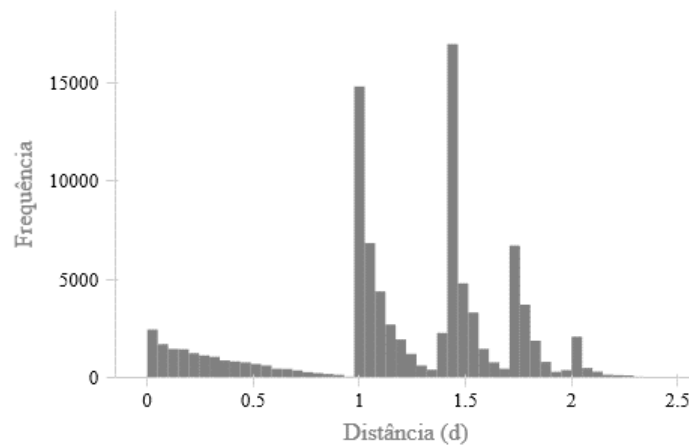
$$d_{i,j} = \sqrt{\sum_{k=1}^5 (V_{ki} - V_{kj})^2 + (D_i - D_j)^2} \quad (7.2)$$

A variável Vídeo (V) está sendo considerada 5 vezes porque, por construção, foi representada através de 5 indicadores independentes (conteúdo alarmante, conteúdo eleitoral, transmissão ao vivo, menciona grupo 1 e menciona grupo 2).

Por construção, todas as variáveis de controle possuem valores no intervalo $[0,1]$. Assim, a menor distância $d_{i,j}$ possível entre dois vídeos i e j é $d_{i,i} = 0$, quando suas covariantes forem idênticas, e igual a $d_{i,j} = \sqrt{1^2 * 5 + 1^2} = \sqrt{6} \cong 2,5$ quando todas as variáveis de controle estiverem em extremidades opostas nos respectivos intervalos de valores.

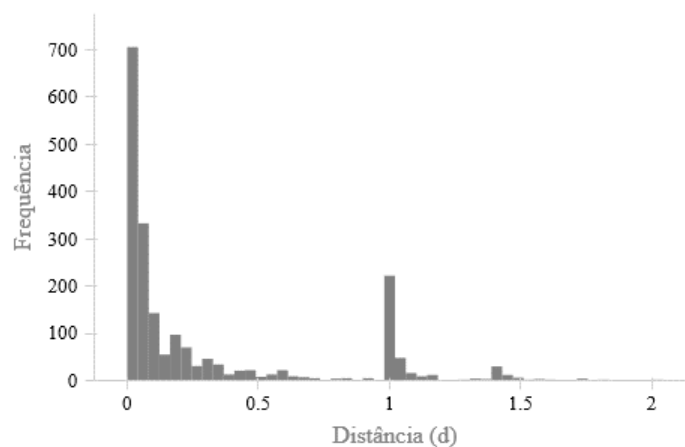
A **Figura 23** apresenta todas as distâncias observadas entre unidades do grupo de tratamento e do grupo de controle, antes do *matching*. Observa-se que, apesar da escala cobrir o intervalo de $[0,2.5]$, a maioria das distâncias entre unidades está concentrada no intervalo $[1,2]$.

Figura 23. Distâncias entre todas as observações do grupo de tratamento e controle.



A **Figura 24** apresenta a distância entre as observações de cada par após o *matching* das unidades do grupo de tratamento com o vizinho mais próximo de cada uma no grupo de controle. Cabe destacar que foi aplicado processo de seleção de vizinhos com reposição e, portanto, cada observação do grupo de controle pode ser selecionada como *match* de mais de uma unidade do grupo de tratamento.

Figura 24. Distâncias entre as observações de cada par após o *matching*.



Ao se comparar as duas figuras anteriores, observa-se que o método teve sucesso em identificar unidades próximas para praticamente todas as observações do grupo de tratamento. Apesar de mais de 80% (82,1%) das distâncias serem superiores a um ($d > 1$) na **Figura 23**, mais de 80% (81,2,4%) dos pares resultantes do método de *matching* foram formados por observações que estão a uma distância inferior a um ($d < 1$), como se vê na **Figura 24**.

Para cálculo do efeito estimado, considerou-se a média da diferença entre os desfechos observados em cada par (*match*). Uma vez que os pares foram construídos a partir de observações do grupo de tratamento, é preciso considerar que o efeito causal foi estimado com base nas unidades tratadas. Dessa forma, ao invés do efeito “geral” do tratamento, estima-se o

efeito do tratamento sobre os tratados ou ATT (*average treatment effect on treated*) com base nas seguintes equações:

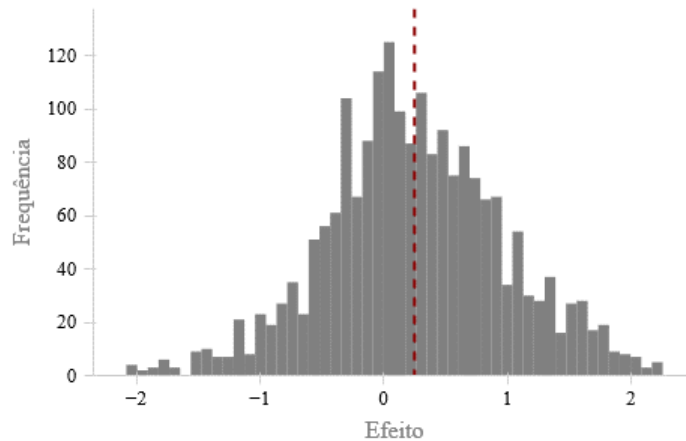
$$TE_m = (Y_m^1 - Y_m^0) \quad (7.3)$$

$$ATT = \frac{1}{N} \sum_{m=1}^N TE_m \quad (7.4)$$

A diferença $TE_m = (Y_m^1 - Y_m^0)$ em (7.3) representa o efeito observado entre a unidade de tratamento Y_m^1 e a unidade de controle Y_m^0 em cada par m . O ATT estimado em (7.4) é a média simples dos efeitos TE observados para todos os pares m .

A **Figura 25** apresenta a distribuição dos efeitos observados em todos os pares. A linha tracejada em destaque marca o ATT estimado de 0,25 ($N = 2.030$; $\sigma = 0,73$; $SE = 0,02$). Cabe destacar que o efeito foi positivo em 63% dos casos.

Figura 25. Distribuição dos efeitos observados em todos os pares.



Apesar da proximidade geral dos pares formados pelo *matching* por distância euclidiana, percebe-se que há pares foram formados por observações com distância entre si superiores às dos demais. Pela simples inspeção visual da **Figura 24** é possível identificar facilmente um grupo de pares com distância superior a um ($d > 1$).

É possível reduzir o impacto que esses pares menos semelhantes causam sobre o efeito estimado com o descarte dos mesmos (STUART e RUBIN, 2008). A **Tabela 5** apresenta o impacto que o descarte de pares exerce sobre o ATT estimado. Em cada linha da tabela, um valor de corte t indica o valor considerado para descarte de pares, i.e., foram mantidos apenas os pares com distância inferior ao valor de corte ($d < t$).

Tabela 5. Efeito estimado considerando diferentes valores para corte de observações.

	N	ATT	Desvio padrão	Erro padrão	Unidades de tratamento descartadas	Unidades de controle descartadas
Todas as observações	2.030	0,25	0,73	0,02	72	2.980
$d < 1$	1.649	0,26	0,73	0,02	453	3.062
$d < 0,75$	1.635	0,26	0,73	0,02	467	3.067
$d < 0,5$	1.571	0,26	0,73	0,02	531	3.081
$d < 0,25$	1.404	0,26	0,71	0,02	698	3.114
$d = 0$	342	0,30	0,72	0,04	1.760	3.639

É interessante observar na segunda linha da tabela anterior que, com o descarte de cerca de 20% (N=453) dos pares menos similares, o ATT estimado praticamente não se altera, passando de 0,25 para 0,26. O valor do efeito permanece estável conforme mais pares são descartados. Quando considerados apenas 342 pares idênticos ($d = 0$), o ATT tem um pequeno aumento para 0,30.

Como verificação adicional, a **Tabela 6** apresenta o balanceamento entre os grupos de tratamento e controle selecionados através do *matching*, mantidos apenas os pares com $d < 1$.

Tabela 6. Balanceamento entre grupos de tratamento e controle após *matching* ($d < 1$).

	Controle (1)	Tratamento (2)	Tratamento – Controle (3)
Dia de Publicação	0,53 [0,27]	0,49 [0,28]	-0,04 (0,01)
Conteúdo alarmante	0,31	0,31	0,00
Conteúdo eleitoral	0,27	0,27	0,00
Transmissão ao vivo	0,18	0,18	0,00
Menciona grupo 1	0,24	0,24	0,00
Menciona grupo 2	0,50	0,50	0,00
Número de observações (N)	1.649	1.649	3.298

Com exceção da variável Dia de Publicação, para todas as demais não há diferença entre as médias das distribuições. Mesmo no caso do Dia de Publicação, a diferença é pequena e pode ser ignorada.

A **Tabela 7** apresenta as correlações entre as covariantes após o procedimento de *matching*. É possível observar na primeira linha (destacada em negrito) que não há correlação entre a variável de intervenção e demais covariantes.

Tabela 7. Matriz de correlação entre as variáveis endógenas após *matching* ($d < 1$).

	W	D	V_1	V_2	V_3	V_4	V_5	Y_V
WhatsApp (W)	1,00	-0,07	0,00	0,00	0,00	0,00	0,00	0,06
Dia de Publicação (D)		1,00	-0,05	0,05	0,04	0,04	0,01	0,05
Conteúdo alarmante (V_1)			1,00	-0,05	-0,12	0,01	0,02	-0,34
Conteúdo eleitoral (V_2)				1,00	0,08	-0,05	0,03	0,01
Transmissão ao vivo (V_3)					1,00	0,03	-0,10	0,08
Menciona grupo 1 (V_4)						1,00	-0,32	-0,01
Menciona grupo 2 (V_5)							1,00	0,08
Número de visualizações (Y_V)								1,00

O procedimento de *matching* por distância euclidiana resultou em uma amostra balanceada e independente da variável de intervenção, resultado semelhante ao que seria obtido através de um experimento aleatório controlado. Dessa forma, é possível considerar que os grupos de tratamento e controle resultantes do *matching* estão em condição de *ceteris paribus*.

7.2.2 Coarsened Exact Matching (CEM)

Como análise adicional do efeito causal, aplicou-se o método *coarsened exact matching* (CEM) para fins de comparação do resultado obtido com o efeito estimado através do método de *matching* por distância euclidiana. O CEM agrupa as observações em estratos de forma que as unidades em cada um deles possuam valores iguais para todas as variáveis de controle. Para que uma estratificação “útil” seja possível, cada variável precisa ser convertida em um número pequeno, porém significativo, de categorias (*bins*), o que torna sua escala de valores menos fina, ou mais “grosseira” (*coarsened*), processo que dá nome ao método.

Com exceção da variável Dia de Publicação, todas as demais são naturalmente categóricas e não demandaram transformação adicional para construção dos estratos. No caso da variável Dia de Publicação, optou-se pela conversão da mesma em 6 categorias com base em intervalos uniformes na escala de $[0,1]$. Dessa forma, cada uma das 6 categorias equivale a uma semana dentro do período contido no conjunto de dados. Dessa forma, a variável passa a ser interpretada com “semana da publicação”.

As 6.024 observações do conjunto de dados foram agrupadas inicialmente em 2.569 estratos diferentes. Em cada estrato, todas as observações possuem exatamente o mesmo valor para todas as variáveis de controle. Em seguida, estratos que não possuíam ao menos uma

unidade de tratamento e uma unidade de controle foram descartados. No fim, foram selecionadas 2.552 observações distribuídas entre 558 estratos. A **Tabela 8** apresenta a quantidade de observações e estratos antes e depois do descarte.

Tabela 8. Quantidade de observações antes e depois do descarte.

	Controle	Tratamento	Total	Estratos	Unidades de tratamento descartadas	Unidades de controle descartadas
N antes do descarte	3.922	2.102	6.024	2.659	0	0
N depois do descarte	1.048	1.504	2.552	558	2.418	1.054

Vale observar que a quantidade de unidades de controle descartadas pelo CEM foi inferior à quantidade descartada pelo método de *matching* por distância euclidiana, mesmo antes do descarte dos pares com distâncias superiores a um valor de corte (**Tabela 5**).

Para possibilitar a estimativa do efeito de tratamento, as observações de controle foram ponderadas para que fossem proporcionalmente equivalentes à quantidade de observações de tratamento em cada estrato (KING, 2011). O peso associado a cada unidade de controle foi estabelecido com base na equação a seguir (IACUS, KING e PORRO, 2012):

$$w_i = \begin{cases} 1, & i \in T^s \\ \frac{m_C}{m_T} \frac{m_T^s}{m_C^s}, & i \in C^s \end{cases} \quad (7.5)$$

Na equação (7.5), w_i denota o peso aplicável a cada unidade i . o Conjunto de todos os estratos é definido por S , e T^s e C^s representam, respectivamente, as unidades de tratamento e de controle em cada estrato $s \in S$. De forma semelhante, m_T e m_C representam a quantidade total de unidades de tratamento e de controle, e m_T^s e m_C^s representam a quantidade de unidades de tratamento e controle dentro de cada estrato $s \in S$. As observações de tratamento recebem peso $w_i = 1$, porém as observações de controle são ajustadas em função do número de observações de tratamento em cada estrato e da relação entre a quantidade de observações de tratamento e controle em todos os estratos.

Após a ponderação da variável de desfecho pelos pesos calculados através da equação (7.5), o efeito do tratamento pode ser estimado como a diferença entre as médias da variável de desfecho entre as observações de tratamento e controle.

O ATT estimado pelo método CEM descrito nesta seção foi de **0,28**, em linha com o efeito estimado através de *matching* por distância euclidiana.

7.2.3 Resultados

A **Tabela 9** consolida o efeito estimado sobre o número de visualizações no YouTube, causado pelo compartilhamento dos vídeos nos grupos do WhatsApp. As estimativas foram obtidas por dois métodos diferentes de *matching* considerando o controle das observações por todas as covariantes identificadas no modelo causal. Os valores de $\Delta_{\%}$ foram calculados de acordo com a equação (7.1).

Tabela 9. Efeito do compartilhamento no WhatsApp sobre número de visualizações estimados por *matching*.

	Efeito estimado	$\Delta_{\%}$
<i>Matching</i> por distância euclidiana ($d < 1$)	0,26	29,7%
<i>Coarsened Exact Matching</i> (CEM)	0,28	32,3%

Nos dois casos, foram comparados apenas vídeos de mesmos canais, com mesmas características, publicados na mesma época e que obtiveram repercussões de mesma ordem de grandeza. Todas as variáveis foram controladas para eliminar eventual viés de seleção e efeitos de interferência das mesmas sobre a variável de intervenção.

Pode-se concluir que, considerando o modelo causal definido neste estudo, o compartilhamento de vídeos nos grupos do WhatsApp causou um aumento no número de visualizações no YouTube de 30% aproximadamente.

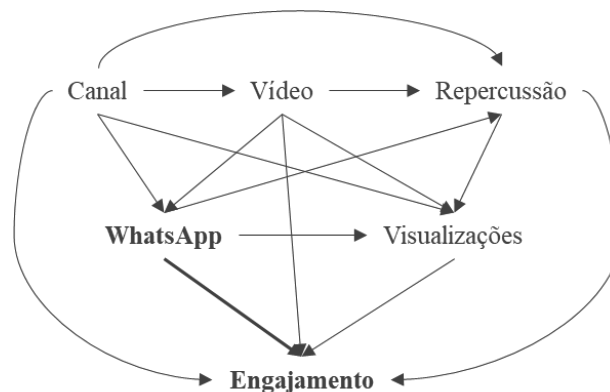
7.3 Efeito do WhatsApp sobre a quantidade de engajamento

Nessa seção investigamos a segunda pergunta do estudo:

- Qual o efeito causal do compartilhamento nos grupos do WhatsApp observado sobre a reação gerada por vídeos do YouTube, medida pela quantidade de engajamento?

As análises apresentadas a seguir são semelhantes em natureza e métodos às que foram apresentadas na seção anterior com uma diferença fundamental: na relação entre a intervenção WhatsApp (W) e o desfecho Engajamento (Y_E), o número de visualizações, que foi a variável de desfecho na análise anterior, passa a exercer um efeito de mediação entre W e Y_E , conforme representado na **Figura 26**. Por essa razão, o número de visualizações passa a fazer parte do conjunto de covariantes a serem controladas.

Figura 26. Diagrama causal da influência do WhatsApp sobre engajamento.



Para fins de simplificação e objetividade do texto, alguns trechos explicativos presentes na análise apresentada na seção 7.2 (efeito do WhatsApp sobre o número de visualizações) foram intencionalmente omitidos nesta seção. Este tópico mantém o foco nos resultados e características particulares associadas ao efeito do WhatsApp sobre o desfecho “quantidade de engajamento”. Cabe apenas reforçar que os métodos utilizados foram os mesmos descritos na seção anterior.

Como antes, o efeito causal foi estimado através de dois métodos distintos, *matching* por distância euclidiana e *coarsened exact matching*.

As variáveis de controle foram selecionadas a partir do DAG representado na **Figura 26**. Esta figura foi apresentada detalhadamente na seção 5.2.2 e representa as premissas assumidas para o fenômeno sob estudo. Assim, para eliminar eventuais efeitos de confusão e isolar o efeito causal da variável intervenção WhatsApp sobre a quantidade de engajamento

(desfecho) é preciso controlar as variáveis Canal (C), Vídeo (V), Repercussão (R) e Visualizações (Y_V) para fechar os caminhos por portas dos fundos representados no modelo.

7.3.1 *Matching* por distância euclideana

Por serem categóricas, as variáveis Canal (C) e Repercussão foram utilizadas para fins de estratificação das observações. Dessa forma, o método de *matching* foi aplicado separadamente sobre cada estrato e só foram apuradas distâncias entre vídeos publicados em mesmos canais e com mesmo grau de repercussão. Com isso, a distância entre duas observações i e j de um canal pode ser dada pela equação (7.6) a seguir:

$$d_{i,j} = \sqrt{\sum_{k=1}^5 (V_{ki} - V_{kj})^2 + (Y_{Vi} - Y_{Vj})^2} \quad (7.6)$$

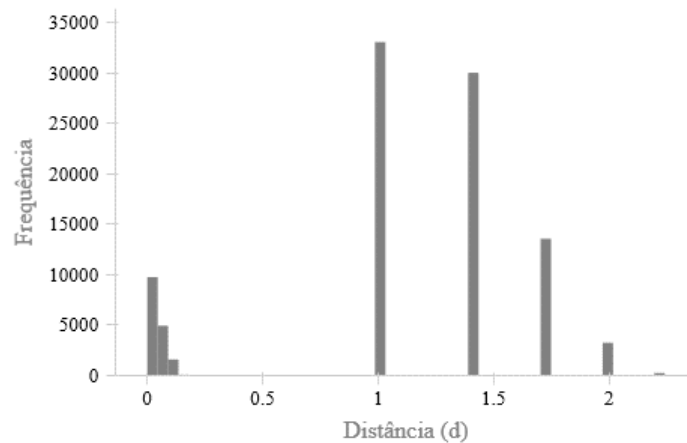
Novamente, a variável Vídeo (V) está sendo considerada 5 vezes porque, por construção, foi representada através de 5 indicadores independentes (conteúdo alarmante, conteúdo eleitoral, transmissão ao vivo, menciona grupo 1 e menciona grupo 2).

Por construção, as variáveis indicativas de características de vídeos (V) possuem valores no intervalo $[0,1]$. Por sua vez, a variável número de visualizações (Y_V) possui valores no intervalo $[0,227]$. Para evitar que essa diferença de escala influencie de forma desproporcional o cálculo da distância euclidiana, optou-se pela transformação linear da variável Y_V em uma escala de $[0,1]$.

Dessa forma, a menor distância $d_{i,j}$ possível entre dois vídeos i e j é $d_{i,i} = 0$, quando suas covariantes forem idênticas, e igual a $d_{i,j} = \sqrt{1^2 * 5 + 1^2} = \sqrt{6} \cong 2,5$ quando todas as variáveis de controle estiverem em extremidades opostas nos respectivos intervalos de valores.

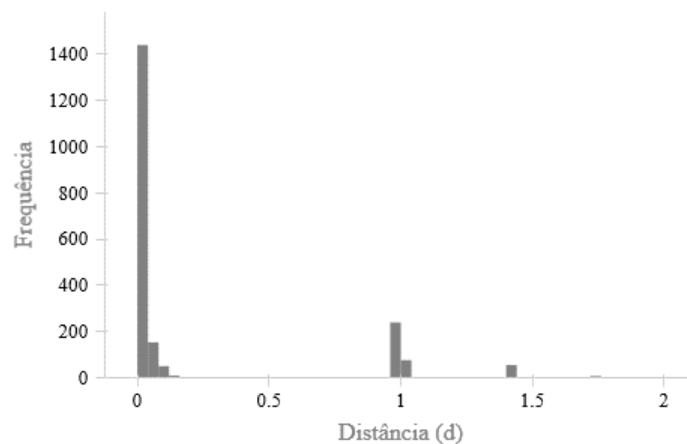
A **Figura 27** apresenta todas as distâncias observadas entre unidades do grupo de tratamento e do grupo de controle, antes do *matching*. Observa-se que, apesar da escala permitir valores no intervalo de $[0,2.5]$, a maioria das distâncias entre unidades está concentrada no intervalo $[1,2]$.

Figura 27. Distâncias entre todas as observações do grupo de tratamento e controle.



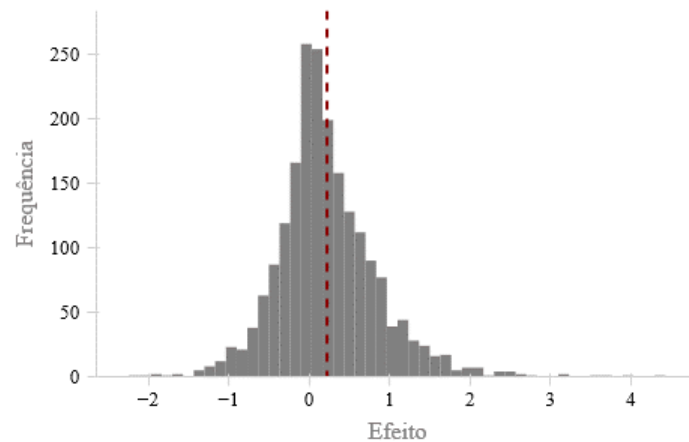
A **Figura 28** apresenta a distância entre as observações de cada par após o *matching* das unidades do grupo de tratamento com o vizinho mais próximo de cada uma no grupo de controle. Cabe destacar que foi aplicado processo de seleção de vizinhos com reposição e, portanto, cada observação do grupo de controle pode ser selecionada como *match* de mais de uma unidade do grupo de tratamento.

Figura 28. Distâncias entre as observações de cada par após o *matching*.



Observa-se que o método teve sucesso em identificar unidades próximas para praticamente todas as observações do grupo de tratamento. Apesar de 83% das distâncias serem superiores a um ($d > 1$) na **Figura 27**, mais de 80% (81,3%) dos pares resultantes do método de *matching* foram formados por observações que estão a uma distância inferior a um ($d < 1$), como se vê na **Figura 28**.

A **Figura 29** apresenta a distribuição dos efeitos observados em todos os pares. A linha tracejada em destaque marca o ATT estimado de 0,22 ($N = 2.030$; $\sigma = 0,63$; $SE = 0,01$). O efeito foi positivo em 63% dos casos. O ATT foi calculado conforme as equações (7.3) e (7.4) apresentadas anteriormente.

Figura 29. Distribuição dos efeitos observados em todos os pares.

Pela inspeção visual da **Figura 28** é possível identificar um pequeno grupo de pares com distância superior a um ($d > 1$). A **Tabela 10** apresenta o impacto que o descarte desses pares exerce sobre o ATT estimado. Em cada linha da tabela, um valor de corte t indica o valor considerado para descarte de pares, i.e., foram mantidos apenas os pares com distância inferior ao valor de corte ($d < t$).

Tabela 10. Efeito estimado considerando diferentes valores para corte de observações.

	N	ATT	Desvio padrão	Erro padrão	Unidades de tratamento descartadas	Unidades de controle descartadas
Todas as observações	2.030	0,22	0,63	0,01	72	2.913
$d < 1$	1.651	0,20	0,59	0,01	451	3.009
$d < 0,75$	1.651	0,20	0,59	0,01	451	3.009
$d < 0,5$	1.651	0,20	0,59	0,01	451	3.009
$d < 0,25$	1.651	0,20	0,59	0,01	451	3.009
$d = 0$	10	0,41	0,56	0,18	2.092	3.913

Após o descarte dos pares com distância superior a um ($d \geq 1$), todos os pares restantes estão em distância inferior a 0,25. O efeito estimado permanece estável em 0,20 para valores de corte entre 0,25 e 1,00. Uma observação importante, há apenas 10 pares com observações idênticas ($d = 0$). Como a variável número de visualizações está em escala contínua, são raras as unidades que possuem exatamente o mesmo valor para essa variável.

A **Tabela 11** apresenta o balanceamento entre os grupos de tratamento e controle selecionados através do *matching*, mantidos apenas os pares com $d < 1$.

Tabela 11. Balanceamento entre grupos de tratamento e controle após matching ($d < 1$).

	Controle (1)	Tratamento (2)	Tratamento – Controle (3)
Número de visualizações	9,64 [2,25]	9,77 [2,30]	0,13 (0,08)
Conteúdo alarmante	0,31	0,31	0,00
Conteúdo eleitoral	0,27	0,27	0,00
Transmissão ao vivo	0,18	0,18	0,00
Menciona grupo 1	0,24	0,24	0,00
Menciona grupo 2	0,50	0,50	0,00
Número de observações (N)	1.651	1.651	3.302

Com exceção do número de visualizações na primeira linha da tabela anterior, em todas as demais não há diferença entre as médias das distribuições. Mesmo no caso dessa variável, considera-se a diferença pequena, podendo ser ignorada.

A **Tabela 12** apresenta as correlações entre as covariantes após o procedimento de *matching*. É possível observar na primeira linha (destacada em negrito) que praticamente toda a correlação que havia entre a variável de intervenção e demais covariantes foi eliminada.

Tabela 12. Matriz de correlação entre as variáveis endógenas após *matching* ($d < 1$).

	W	Y_V	V_1	V_2	V_3	V_4	V_5	Y_E
WhatsApp (W)	1,00	0,03	0,00	0,00	-0,00	-0,00	-0,00	0,04
Número de visualizações (Y_V)		1,00	-0,34	0,00	0,08	-0,02	0,09	0,97
Conteúdo alarmante (V_1)			1,00	-0,05	-0,12	0,01	0,02	-0,31
Conteúdo eleitoral (V_2)				1,00	0,08	-0,05	0,03	0,00
Transmissão ao vivo (V_3)					1,00	0,03	-0,10	0,10
Menciona grupo 1 (V_4)						1,00	-0,32	-0,04
Menciona grupo 2 (V_5)							1,00	0,10
Quantidade de engajamento (Y_E)								1,00

O procedimento de *matching* por distância euclidiana resultou em uma amostra balanceada e independente da variável de intervenção, resultado semelhante ao que seria obtido através de um experimento aleatório controlado. Dessa forma, é possível considerar que os grupos de tratamento e controle resultantes do *matching* estão em condição de *ceteris paribus*.

7.3.1.1 Coarsened Exact Matching (CEM)

Para concluir a análise do efeito causal, aplicou-se o método *coarsened exact matching* (CEM) para fins de comparação do resultado obtido com o efeito estimado através do método de *matching* por distância euclidiana.

Com exceção da variável Número de Visualizações, todas as demais são categóricas e não demandaram transformação adicional para construção dos estratos. A variável foi convertida em 10 categorias com base em intervalos uniformes na sua escala de valores.

As 6.024 observações do conjunto de dados foram agrupadas inicialmente em 1.646 estratos diferentes. Em cada estrato, todas as observações possuem exatamente o mesmo valor para todas as variáveis de controle. Em seguida, estratos que não possuíam ao menos uma unidade de tratamento e uma unidade de controle foram descartados. No fim, foram selecionadas 3.709 observações distribuídas entre 497 estratos. A **Tabela 13** apresenta a quantidade de observações e estratos antes e depois do descarte.

Tabela 13. Quantidade de observações antes e depois do descarte.

	Controle	Tratamento	Total	Estratos	Unidades de tratamento descartadas	Unidades de controle descartadas
N antes do descarte	3.922	2.102	6.024	1.646	0	0
N depois do descarte	1.410	2.299	3.709	497	692	1.623

Vale observar que a quantidade de unidades de controle descartadas pelo CEM foi novamente inferior à quantidade descartada pelo método de *matching* por distância euclidiana, mesmo antes do descarte dos pares com distâncias superiores a um valor de corte (**Tabela 10**).

Após a ponderação da variável de desfecho pelos pesos calculados através da equação (7.5), o efeito do tratamento pode ser estimado como a diferença entre as médias da variável de desfecho entre as observações de tratamento e controle.

O ATT estimado pelo método CEM descrito nesta seção foi de **0,18**, em linha com o efeito estimado através de *matching* por distância euclidiana.

7.3.2 Resultados

A **Tabela 14** consolida o efeito estimado sobre a quantidade de engajamento no YouTube, causado pelo compartilhamento dos vídeos nos grupos do WhatsApp. As estimativas foram obtidas por dois métodos diferentes de *matching* considerando o controle das

observações por todas as covariantes identificadas no modelo causal. Os valores de $\Delta_{\%}$ foram calculados de acordo com a equação (7.1).

Tabela 14. Efeito do compartilhamento no WhatsApp sobre quantidade de engajamento estimados por *matching*.

	Efeito estimado	$\Delta_{\%}$
<i>Matching</i> por distância euclidiana ($d < 1$)	0,20	22,1%
<i>Coarsened Exact Matching</i> (CEM)	0,18	19,7%

Nos dois casos, foram comparados vídeos de mesmos canais, com mesmas características, mesmo volume de visualizações e que obtiveram repercussões de mesma ordem de grandeza. Todas as variáveis foram controladas para eliminar eventual viés de seleção e efeitos de interferência das mesmas sobre a variável de intervenção.

Pode-se concluir que, considerando o modelo causal definido neste estudo, o compartilhamento de vídeos nos grupos do WhatsApp causou um aumento na quantidade de engajamento no YouTube de 20% aproximadamente.

Cabe reforçar que este impacto de 20% foi complementar ao impacto de 30% causado sobre o número de visualizações e desconsidera os efeitos do último.

8 CONCLUSÕES

Neste trabalho, utilizou-se um conjunto de métodos de inferência causal para estimar o impacto que o compartilhamento de vídeos do YouTube teve sobre visualizações e engajamento. Após a representação do processo de geração dos dados em um modelo causal, o critério de portas dos fundos foi utilizado para seleção das variáveis a serem controladas para que fosse possível isolar o efeito da variável de intervenção sobre os diferentes desfechos de interesse (visualizações e engajamento).

Os efeitos foram estimados por dois métodos distintos: *matching* por distância euclideana e *Coarsened Exact Matching* (CEM). Os resultados são apresentados na **Tabela 15**.

Tabela 15. Efeito médio do WhatsApp sobre visualizações e engajamento.

	Visualizações		Engajamento	
	Log	$\Delta\%$	Log	$\Delta\%$
<i>Matching</i> por distância euclideana	0,26	29,7%	0,20	22,1%
<i>Coarsened Exact Matching</i> (CEM)	0,28	32,3%	0,18	19,7%

Com as variáveis controladas com base no critério de portas dos fundos, os resultados estimados através de dois métodos diferentes foram semelhantes tanto para estimativa do efeito sobre o número de visualizações quanto sobre a quantidade de engajamento. Essa coerência reforça nossa segurança sobre os efeitos causais estimados.

Também reforça a percepção de que as variáveis de controle identificadas e representadas no modelo causal representam o fenômeno sob estudo de forma satisfatória, permitindo chegar a conclusões causais mesmo a partir de dados observacionais, não provenientes de um experimento aleatorizado.

Para estimativa do efeito do WhatsApp sobre o número de visualizações, foram comparados vídeos de mesmas características publicados em um mesmo canal, em datas próximas e que causaram repercussões de mesma ordem de grandeza. Para estimativa do efeito do WhatsApp sobre a quantidade de engajamento, foram comparados vídeos de mesmas características publicados em um mesmo canal e que alcançaram números de visualizações e graus de repercussão próximos.

Dessa forma, estima-se que o compartilhamento de vídeos do YouTube em grupos públicos do WhatsApp durante a campanha eleitoral de 2018 tenha gerado um **aumento médio aproximado de 30% no número de visualizações e de 20% na quantidade de engajamento.**

Respondendo à pergunta geral do trabalho, pode-se considerar, com isso, que **os grupos do WhatsApp observados impulsionaram a monetização dos vídeos do YouTube compartilhados.**

Mas em que grau? Infelizmente não é possível medir o efeito efetivo do compartilhamento sobre a monetização gerada por esses vídeos, mas é possível fazer algumas inferências, mesmo que de forma imprecisa.

Considerando um aumento médio nas visualizações de cerca de 30% é coerente imaginar que as visualizações de anúncios tenham crescido na mesma proporção. Logo, apenas o efeito sobre as visualizações pode ter causado um aumento na monetização desses vídeos na mesma ordem de grandeza de 30%. Considerando que mais engajamentos sugerem usuários mais dispostos a interagir com o vídeo, parece razoável pensar que esse aumento no engajamento também possa ter se refletido em uma maior “taxa de visualização” de anúncios.

Não é possível transformar essas inferências em quantidades objetivas, mas pode-se considerar de forma conservadora que os vídeos compartilhados contribuíram pelo menos 30% mais, em média, para a monetização dos seus canais que os demais vídeos. Esse aumento pode ter sido ainda maior se considerarmos que o aumento no engajamento também se refletiu em mais visualizações de anúncios. Considerando que a média de vídeos compartilhados por canal foi de 40%, é possível dizer que o compartilhamento dos vídeos nos grupos fez com que esses canais aumentassem suas receitas em pelo menos 12% na média. Como média, esse número não representa algumas situações extremas. Alguns canais tiveram menos vídeos compartilhados, enquanto outros tiveram mais de 60% dos seus vídeos postados nos grupos. Mas a simples possibilidade de se estimar o impacto causado por uma rede de grupo como essa pode ser suficiente para subsidiar decisões de investimento em ações de marketing direcionada a eles, por exemplo. Mesmo que imprecisos, esses números parecem bastante plausíveis, dados os efeitos sobre visualizações e engajamento estimados separadamente. Também trazem uma percepção de como ações de impulsionamento em redes sociais podem alavancar audiências e interações.

Cabe perguntar até que ponto as pessoas que compartilharam esses vídeos estavam apenas interessadas na disseminação da informação que eles traziam, ou se esses grupos foram também usados para aumentar a remuneração recebida pelos canais. Não é possível responder a essa questão apenas com os dados disponíveis, mas ela sugere um “modelo de negócio” perfeitamente possível: grupos são organizados com o objetivo declarado de discutir um “assunto secundário”, mas na prática existem e são usados para alavancar resultados latentes não percebidos pelos seus membros.

Se não se pode evidenciar as intenções, pelo menos conseguimos estimar o efeito causado pelas ações.

8.1 Contribuições

Esse estudo apresentou evidências de que grupos públicos do WhatsApp utilizados durante a campanha presidencial brasileira de 2018 podem ter contribuído para aumentar as receitas de canais do YouTube que tiveram vídeos compartilhados nos mesmos. Estima-se que o compartilhamento de vídeos nessa rede de grupos aumentou em 30% o número de visualizações e em 20% a quantidade de engajamento, em média. Por consequência, a remuneração obtida por esses canais teria sido menor se não fosse pela alavancagem gerada pelos grupos.

Foi utilizado um método clássico de inferência causal para estimar o efeito direto que determinada intervenção em uma rede de pessoas causa sobre resultados em outra rede social a partir de dados observacionais. O método envolveu a representação do fenômeno em um diagrama causal, reflexão sobre contrafactuais e controle de variáveis endógenas específicas para redução do viés de seleção e de fatores de confusão. Os efeitos causais foram estimados através de métodos de *matching* por distância euclidiana e *Coarsened Exact Matching*.

Um conjunto de dados foi estruturado para o estudo integrando informações coletadas em duas redes sociais distintas (WhatsApp e YouTube) consolidando para cada observação, informações relativas à intervenção, desfechos e variáveis relevantes ao modelo causal.

8.2 Limitações

No desenvolvimento do trabalho foram tomadas decisões e assumidas premissas que restringem suas conclusões a um contexto específico de análise, mas que não impedem sua reprodução em outros cenários. No entanto, algumas limitações aplicáveis a nossas conclusões precisam ser destacadas.

Uma limitação criada deliberadamente foi consequência da forma como a variável repercussão foi construída. Ao representar a repercussão como uma escala discreta a partir do número de visualizações, e ao utilizá-la como variável para fins de controle e pareamento, acabou-se restringindo a estimativa dos efeitos a vídeos que alcançaram visualizações na mesma ordem de grandeza. Dessa forma, o efeito estimado pode ter sido limitado pelo próprio desenho do experimento. Por exemplo, um vídeo que, eventualmente, tenha atingido um grau

de visualização muito superior a outro por causa específica do compartilhamento no WhatsApp acabou comparado a outro vídeo com mesma popularidade. Parte considerável do efeito pode ter sido perdida no processo para algumas observações. Por isso, acredita-se que o efeito efetivo será sempre superior àquele que observamos através dessa variável. A solução para essa limitação passa por identificar uma outra variável para representar a repercussão do vídeo, que não derive diretamente do número de visualizações.

A rede de grupos observada no estudo possui características particulares que podem ser diferentes de outras redes. Por exemplo, é natural pensar que esses grupos tenham favorecido um maior engajamento dos seus usuários uma vez que foram formados com a finalidade explícita de mobilização de ações em uma campanha eleitoral. Outros grupos públicos no WhatsApp podem ter dinâmicas diferentes e em proporções maiores ou menores. Outras plataformas de redes sociais podem ser mais ou menos eficazes no impulsionamento de ações. Uma eventual reprodução da mesma metodologia em outros contextos possivelmente observará efeitos diferentes.

Também cabe destacar que o efeito foi estimado com base em apenas uma parte dos canais e vídeos compartilhados nessa rede de grupos. A seleção dos canais foi feita visando assegurar amostras com um tamanho mínimo para cada canal, mas outros canais mais ou menos favorecidos pelo comportamento dos grupos podem ter sido deixados de fora no processo.

Ainda, o estudo incluiu apenas vídeos sobreviventes no YouTube passados 30 meses de sua publicação, o que representa cerca de metade dos vídeos postados em 2018. A outra metade não estava mais disponível no momento de coleta dos metadados e foi descartada. Considerou-se que a ausência desses vídeos não compromete as estimativas obtidas no trabalho, mas de alguma forma eles contribuíram para a monetização de seus canais enquanto estiveram disponíveis.

Por fim, nem todos os canais do YouTube se habilitam para o programa de parcerias e geram receitas. As estimativas apresentadas não podem ser interpretadas como um aumento real de receitas, mas apenas como referência de ganhos potenciais.

8.3 Trabalhos futuros

Percebe-se um amplo conjunto de ações que seriam possíveis em continuidade a análise apresentada nessa pesquisa. Sobre esse mesmo conjunto de dados, é importante avaliar em que grau a quantidade de compartilhamentos influencia no aumento da popularidade e capacidade de monetização dos vídeos. Também é possível analisar a dinâmica temporal da influência da

rede de grupos. Em outras palavras, o impulsionamento gerado por esses grupos foi constante ao longo do período ou apresentou variações? Também pode-se analisar de forma isolada a contribuição de cada grupo do WhatsApp sobre os efeitos observados nos canais. Que grupos causaram mais visualizações e engajamento? É possível identificar propriedades nesses grupos que ajudem a explicar a diferença observadas entre seus efeitos?

Do ponto de vista metodológico, é importante avaliar os resultados obtidos utilizando-se de outros métodos e ferramentas de inferência causal para confirmar se os resultados observados permanecem estáveis em outras condições, ou se alguma das premissas do modelo causal deve ser revista.

Seria importante realizar novos estudos sobre combinações diferentes de redes sociais e desfechos de interesse. Qual a diferença entre os novos efeitos observados e os estimados nesse estudo? Que redes sociais são mais eficazes para alavancar diferentes tipos de ações? Que ajustes são necessários na abordagem proposta para que possa ser aplicada em outros contextos?

Por fim, considerando apenas desfechos associados a ganhos financeiros, será possível estimar um ponto de equilíbrio com base no potencial de impulsionamento de diferentes redes sociais, a partir do qual o investimento em intervenções em cada rede deixaria de ser vantajoso?

REFERÊNCIAS

- ABADIE, A.; IMBENS, G. W. Bias-Corrected Matching Estimators for Average Treatment Effects. *Journal of Business & Economics Statistics*. vol. 20, p. 1-11, 2011. DOI: <https://doi.org/10.1198/jbes.2009.07333>
- AMRI, A. B. The impact of whatsapp mobile social learning on the achievement and attitudes of female students compared with face to face learning in the classroom. *European Scientific Journal*. Vol.10, No.22. aug. 2014.
- ANGER, I.; KITTL, C. Measuring influence on Twitter. *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies (i-KNOW '11)*. ACM, New York, NY, USA, Article 31, p. 1–4. 2011. DOI:<https://doi.org/10.1145/2024288.2024326>
- ANGRIST, J. D.; KRUEGER, A. B. Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, 15(4), 69–85. 2001. <https://doi.org/10.1257/jep.15.4.69>
- ANGRIST, J.; PISCHKE, J. Mastering 'Metrics: The Path from Cause to Effect. Princeton University Press; 2014.
- ANGRIST, J.; PISCHKE, J. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press; 2009.
- BRADSHAW, S.; HOWARD, P. N. The Global Disinformation Disorder: 2019 Global Inventory of Organised Social Media Manipulation. *Project on Computational Propaganda*, Oxford, UK. 2019. Disponível em: <https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf>.
- BROWN, J.; BRODERICK, A. J.; LEE, N. Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of Interactive Marketing*. Vol. 21(3), p. 2-20. 2007. <https://doi.org/10.1002/dir.20082>.
- BUDZINSKI, O.; GAENSSLE, S., The Economics of Social Media Stars: An Empirical Investigation of Stardom, Popularity, and Success on Youtube. *Ilmenau Economics Discussion Papers*, Vol. 21, No. 112, jan. 2018.
- BURSZTYN, V. S.; BIRNBAUM, L. Thousands of Small, Constant Rallies: A Large-Scale Analysis of Partisan WhatsApp Groups. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2019, p. 484-488, doi: 10.1145/3341161.3342905.
- CAETANO, J. A.; OLIVEIRA, J. F. de; LIMA, H. S., MARQUES-NETO, H. T.; MAGNO, G.; MEIRA, W.; ALMEIDA, V. A. F. Analyzing and characterizing political discussions in WhatsApp public groups. 2018. Disponível em: <http://arxiv.org/abs/1804.00397>
- CARLSON, B. What Investors Should Learn From Reddit, GameStop, Market Manipulation And Other Recent Headlines. *Forbes*. 11 fev. 2021. Disponível em:

<https://www.forbes.com/sites/bobcarlson/2021/02/11/what-investors-should-learn-from-reddit-gamstop-market-manipulation-and-other-recent-headlines>.

CHTOUKI, Y.; HARROUD, H.; KHALIDI, M.; BENNANI, S. The impact of YouTube videos on the student's learning. *International Conference on Information Technology Based Higher Education and Training (ITHET)*. 2012. doi: 10.1109/ITHET.2012.6246045.

CHRISTAKIS, N. A.; FOWLER, J. H. Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives -- How Your Friends' Friends' Friends Affect Everything You Feel, Think, and Do. Little, Brown and Company. 2009.

CLAUSET, A.; SHALIZI, C. R.; NEWMAN, M. E. J. Power-law distributions in empirical data. *SIAM Review*, 51(4), 661–703. 2009. <https://doi.org/10.1137/070710111>

COROMINA, O.; MATAMOROS-FERNÁNDEZ, A.; RIEDER, B. FOLLOW THE MONEY: A LARGE-SCALE INVESTIGATION OF MONETIZATION AND OPTIMIZATION ON YOUTUBE. *AoIR Selected Papers of Internet Research*, 2020. <https://doi.org/10.5210/spir.v2020i0.11194>

CULNAN, M. J.; MCHUGH, P. J.; ZUBILLAGA, J. I. How Large U.S. Companies Can Use Twitter and Other Social Media to Gain Business Value. *MIS Quarterly Executive*, vol. 9, n. 4, p. 243-259, dez. 2010.

CUNNINGHAM, S. Causal Inference: The Mixtape. Yale University Press. 2021.

DAVENPORT, T. H.; BECK, J. C. Attention Economy: Understanding the New Currency of Business. Harvard Business Review Press. 2002.

FISHER, R. A. The design of experiments. 1. ed. Oliver and Boyd. 1935.

FRENKEL, S. The Most Influential Spreader of Coronavirus Misinformation Online. *The New York Times*. 24 jul. 2021. Disponível em: <https://www.nytimes.com/2021/07/24/technology/joseph-mercola-coronavirus-misinformation-online.html>.

GARIMELLA, K.; TYSON, G. WhatsApp Doc? A First Look at WhatsApp Public Group Data. *Twelfth International AAAI Conference on Web and Social Media*. AAAI Publications, 2018.

HAAVELMO, T. The statistical implications of a system of simultaneous equations. *Econometrica*, vol. 11(1), p. 1–12. jan.1943.

HOLLAND, P. W. Statistics and Causal Inference. *Journal of the American Statistical Association*, vol. 81, p. 945-960. 1986.

IACUS, S. M., KING, G., PORRO, G. Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, Vol. 20(1), p. 1–24. 2012. DOI: <https://doi.org/10.1093/pan/mpr013>

KAIKATI, A. M.; KAIKATI, J. G. Stealth Marketing: How to Reach Consumers Surreptitiously. *California Management Review*. vol. 46(4), p. 6-22. 2004, doi:10.2307/41166272

KING, G.; NIELSEN, R.; COBERLEY, C.; POPE, J. E.; WELLS, A. Comparative Effectiveness of Matching Methods for Causal Inference. Harvard. 2011. Disponível em: <https://gking.harvard.edu/files/psparadox.pdf>.

LAMBTON-HOWARD, D.; ANDERSON, R.; MONTAGUE, K.; GARBETT, A.; HAZELDINE, S.; ALVAREZ, C.; SWEENEY, J. A.; OLIVIER, P.; KHARRUFA, A.; NAPPEY, T. WhatFutures: Designing Large-Scale Engagements on WhatsApp. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, Paper 159, p. 1–14. 2019. DOI:<https://doi.org/10.1145/3290605.3300389>

LIKKANEN, L.A. Three Metrics for Measuring User Engagement with Online Media and a YouTube Case Study. p. 2–5. 2014. Disponível em: <https://arxiv.org/abs/1312.5547>

MACHADO, C.; KIRA, B.; NARAYANAN, V.; KOLLANVI, B. HOWARD, P. A Study of Misinformation in WhatsApp groups with a focus on the Brazilian Presidential Elections. *Companion Proceedings of The 2019 World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 1013–1019. mai. 2019. DOI:<https://doi.org/10.1145/3308560.3316738>

MARTIN, K. D.; SMITH, N. C. Commercializing Social Interaction: The Ethics of Stealth Marketing. *Journal of Public Policy & Marketing*. vol. 27(1), p. 45-56. 2008. doi:10.1509/jppm.27.1.45

MEJOVA, Y.; SRINIVASAN, P. Political speech in social media streams: YouTube comments and Twitter posts. In *Proceedings of the 4th Annual ACM Web Science Conference (WebSci '12)*. ACM, New York, NY, USA, p. 205–208. 2012. DOI:<https://doi.org/10.1145/2380718.2380744>

MELLO, P. C. A máquina do ódio. Companhia das Letras. 2020.

MELO, P. de F.; VIEIRA, C. C.; GARIMELLA, K.; MELO, P.O.S.V. de; BENEVENUTO, F. Can WhatsApp Counter Misinformation by Limiting Message Forwarding? CHERIFI, H.; GAITO, S.; MENDES, J.; MORO, E.; ROCHA, L. (eds). *Complex Networks and Their Applications VIII. COMPLEX NETWORKS 2019. Studies in Computational Intelligence*, vol 881. Springer, Cham. https://doi.org/10.1007/978-3-030-36687-2_31

MELO, P.; MESSIAS, J.; RESENDE, G.; GARIMELLA, K.; ALMEIDA, J.; BENEVENUTO, F. WhatsApp monitor: A fact-checking system for WhatsApp. *Proceedings of the 13th International Conference on Web and Social Media, ICWSM 2019*, p. 676–677. 2019.

MONT'ALVERNE, C.; MITOZO, I. MUITO ALÉM DA MAMADEIRA ERÓTICA: As notícias compartilhadas nas redes de apoio a presidenciáveis em grupos de WhatsApp nas eleições. *VIII Congresso Da Associação Brasileira de Pesquisadores Em Comunicação e Política*, 1–25. mai. 2019. Disponível em: http://ctpol.unb.br/compolitica2019/GT4/gt4_Montalverne_Mitozo.pdf

OMS. Fighting misinformation in the time of COVID-19, one click at a time. 27 abr. 2021. Disponível em: <https://www.who.int/news-room/feature-stories/detail/fighting-misinformation-in-the-time-of-covid-19-one-click-at-a-time>

PEARL, J. [Bayesian Analysis in Expert Systems]: Comment: Graphical Models, Causality and Intervention. *Statistical Science*, 8(3), p 266-269. 1993. Disponível em: <http://www.jstor.org/stable/2245965>

PEARL, J. Causal inference in statistics: An overview. *Statistics Surveys*. sep. 2009, 96–146. <https://doi.org/10.1214/09-SS057>

PEARL, J. Causality: models, reasoning and inference. 1. ed. Cambridge University Press; 2000.

PEARL, J. Causality: models, reasoning and inference. 2. ed. Cambridge University Press; 2009.

PEARL, J; MACKENZIE, D. The book of why: the new science of cause and effect. New York, USA: Basic Books; 2018.

RESENDE, G.; MELO, P.; SOUSA, H.; MESSIAS, J.; VASCONCELOS, M.; ALMEIDA, J.; BENEVENUTO, F. (Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures. *The World Wide Web Conference (WWW '19)*. Association for Computing Machinery. New York, NY, USA, 818–828, mai. 2019. DOI: <https://doi.org/10.1145/3308558.3313688>

RESENDE, G.; MESSIAS, J.; SILVA, M.; ALMEIDA, J., VASCONCELOS, M.; BENEVENUTO, F. A System for Monitoring Public Political Groups in WhatsApp. *Anais do Simpósio Brasileiro de Sistemas Multimídia e Web (WEBMEDIA)*, p. 387–390. out. 2018. DOI: <https://doi.org/10.5753/webmedia.2018.4605>

RUBIN, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*. 66(5), p. 688-701. 1974. doi: <https://doi.org/10.1037/h0037350>.

SANTOS, J. G. B. dos; FREITAS, M.; ALDÉ, A.; SANTOS, K.; CUNHA, V. C. C. WhatsApp, política mobile e desinformação: a hidra nas eleições presidenciais de 2018. *Revista Comunicação & Sociedade*, São Bernardo do Campo, v. 41, n. 2, p. 307-334, maio-ago. 2019. Disponível em: <https://www.metodista.br/revistas/revistas-metodista/index.php/CSO/article/view/9410>.

SNOW, J. On the Mode of Communication of Cholera. *Edinburgh Medical Journal*, London, UK, 1 jan. 1856, p 668-670. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5307547/>.

SPIEL, C.; LAPKA, D.; GRANFINGER, P.; ZODLHOFFER, E.; M., REIMANN, R.; SCHOBBER, B.; WAGNER, P.; EYE, A. von. A Euclidean Distance-Based Matching Procedure for Nonrandomized Comparison Studies. *European Psychologist*. Vol. 13:3, p. 180-187, 2008. DOI: <https://doi.org/10.1027/1016-9040.13.3.180>.

- SPIRITES, P. Introduction to Causal Inference. *Journal of Machine Learning Research*, 11, p. 1643-1662, 2010. Disponível em: <https://www.jmlr.org/papers/volume11/spirtes10a/spirtes10a.pdf>.
- SPLAWA-NEYMAN, J. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. *Annals of Agricultural Sciences*. p. 1-51. 1923.
- STEINFELD, C.; ELLISON, N. B.; LAMPE, C. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology*. Vol. 9(8), p. 434-445. nov. 2008. <https://doi.org/10.1016/j.appdev.2008.07.002>.
- STUART, E. A. Matching methods for causal inference: A review and a look forward. *Statistical Science*, Vol. 25(1), p. 1-21. fev. 2010. <https://doi.org/10.1214/09-STS313>.
- STUART, E. A.; RUBIN, D. B. Best practices in quasi-experimental designs: matching methods for causal inference. *Osborne, J. (Ed.), Best practices in quantitative methods*. p. 155-176. SAGE Publications. 2008. DOI: <https://www.doi.org/10.4135/9781412995627>
- SZABO, G.; HUBERMAN, B. A. Predicting the popularity of online content. *Communications of the ACM*, Vol. 53(8), p. 80-88. 2010. <https://doi.org/10.1145/1787234.1787254>.
- THORSON, K.; DRISCOLL, K.; EKDALE, B.; EDGERLY, S.; THOMPSON, L. G.; SCHROCK, A.; SWARTZ, L., VRAGA, E. K.; WELLS, C. YOUTUBE, TWITTER AND THE OCCUPY MOVEMENT: Connecting content and circulation practices, *Information, Communication & Society*, Vol. 16:3, p. 421-451. 2013. DOI: 10.1080/1369118X.2012.756051
- WRIGHT, S. Systems of Mating. I. the Biometric Relations between Parent and Offspring. *Genetics*, 6(2), p. 111-123. mar. 1921.
- WU, S.; RIZOIU, M. A.; XIE, L. Beyond views: Measuring and predicting engagement in online videos. *12th International AAAI Conference on Web and Social Media, ICWSM 2018*, p. 434-443. 2018.
- YEE, N.; BAIENSON, J. N.; DUCHENEAUT, N. The Proteus Effect: Implications of Transformed Digital Self-Representation on Online and Offline Behavior. *Communication Research*. 36(2), p. 285-312. 2009. doi:10.1177/0093650208330254
- YU, H.; XIE, L.; SANNER, S. Twitter-driven YouTube Views: Beyond Individual Influencers. *Proceedings of the 22nd ACM international conference on Multimedia (MM '14)*. ACM, New York, NY, USA, p. 869-872. 2014. DOI:<https://doi.org/10.1145/2647868.2655037>
- ZHANG, Y.; PENNACCHIOTTI, M. Predicting purchase behaviors from social media. *Proceedings of the 22nd international conference on World Wide Web (WWW '13)*. ACM, New York, NY, USA, p. 1521-1532. 2013. DOI:<https://doi.org/10.1145/2488388.2488521>

ZHOU, R.; KHEMMARAT, S.; GAO, L. The impact of YouTube recommendation system on video views. *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement (IMC '10)*. ACM, New York, NY, USA, 404–410. 2010.
DOI:<https://doi.org/10.1145/1879141.1879193>