CKP8077
Estruturas de Dados

Detecção de Comportamento Coordenado

Material cedido pelo Prof. Marco A. Casanova (PUC-Rio)

Roteiro

- Introdução
- O que é Comportamento Coordenado?
- Conceitos
- Detecção de Comportamento Coordenado

Introdução

- Motivação:
  - Propaganda, desinformação, manipulação e polarização são as doenças modernas de uma sociedade cada vez mais dependente das redes sociais como fonte de notícias.

Conceitos

- Coordenação
  - Coordenação, o processo no qual múltiplos atores conectados estão envolvidos para perseguir objetivos, é um aspecto fundamental na existência de várias formas de vida, incluindo os seres humanos.
  - Com o advento das plataformas de redes sociais, a coordenação também se tornou um componente fundamental das interações online. Os usuários de redes sociais agora dispõem de uma ampla variedade de ferramentas para se coordenar uns com os outros, como hashtags que lhes permitem discutir coletivamente tópicos específicos. As plataformas online se tornaram um ambiente adequado para organizar movimentos sociais e políticos em todo o mundo, dando origem a fenômenos como ativismo online, boicotes e protestos.

Conceitos

- Coordenação
  - Porém, estudiosos encontraram evidências de coordenação online sendo explorada por atores nefastos para todos os tipos de propósitos maliciosos.
  - Por exemplo, campanhas de desinformação frequentemente aproveitam atores que coordenam suas ações para maximizar o alcance de suas narrativas falsas.
  - Da mesma forma, a coordenação é empregada dentro de operações de informação e astroturfing, que envolve criar a falsa aparência de apoio popular para uma causa, produto ou pessoa alvo.
  - Também bots sociais e trolls exploram a coordenação online para amplificar mensagens, manipular tendências ou espalhar desinformação.
  - Finalmente, a coordenação também resulta de, e contribui para, a formação de câmaras de eco e polarização online.

Conceitos

- Coordenação:

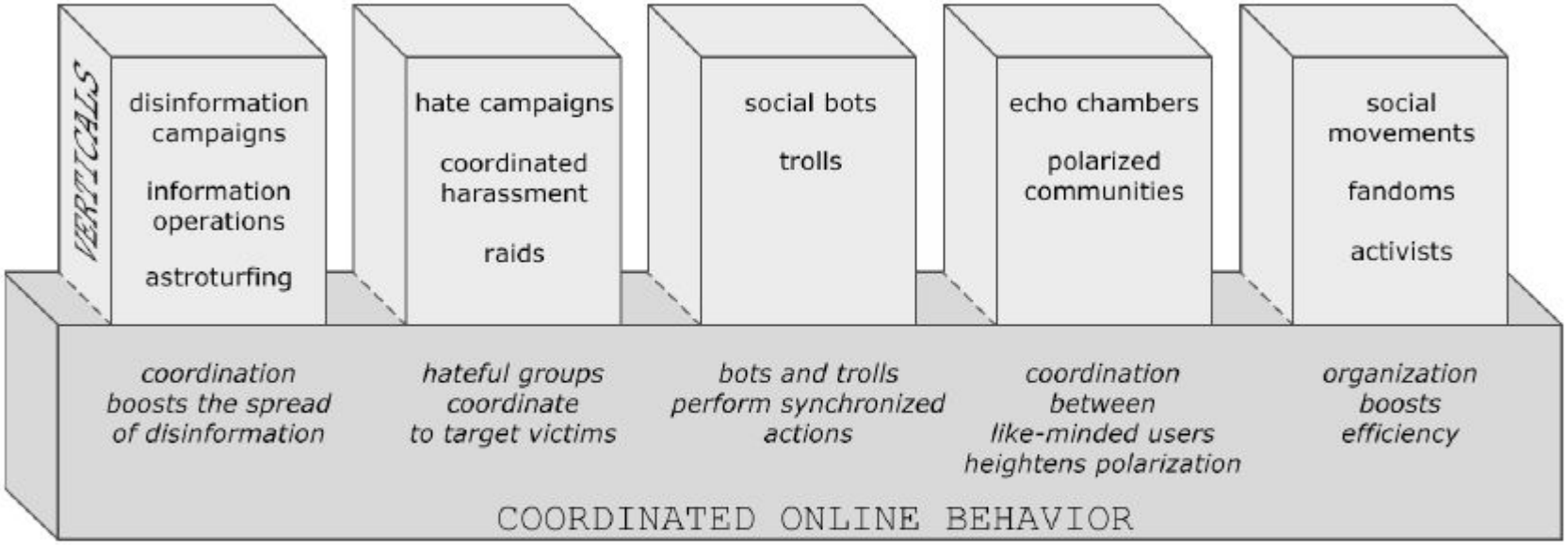


Fig. 1. Coordination is a fundamental aspect of online human interactions and the study of coordinated online behavior can complement the analyses of many other online phenomena.

Conceitos

- Comportamento Inautêntico Coordenado (Coordinated Inauthentic Behavior - CIB):
  - Uma tática de manipulação que utiliza uma rede de contas, páginas ou grupos para coordenar ações, muitas vezes usando contas falsas para amplificar mensagens.
  - Comportamento inautêntico coordenado (CIB) é uma tática de comunicação manipulativa que usa uma mistura de contas de mídia social autênticas, falsas e duplicadas para operar como uma rede adversária (AN) em várias plataformas de mídia social.
  - Atividade online organizada onde uma conta ou grupos de contas, incluindo contas secundárias "falsas" (que existem exclusivamente ou principalmente para se envolver em tais campanhas), agem para enganar as pessoas ou elevar fraudulentamente a popularidade ou visibilidade de conteúdo ou contas, como seguir em massa uma conta para aumentar sua influência.

Conceitos

- Comportamento Inautêntico Coordenado (Coordinated Inauthentic Behavior - CIB):

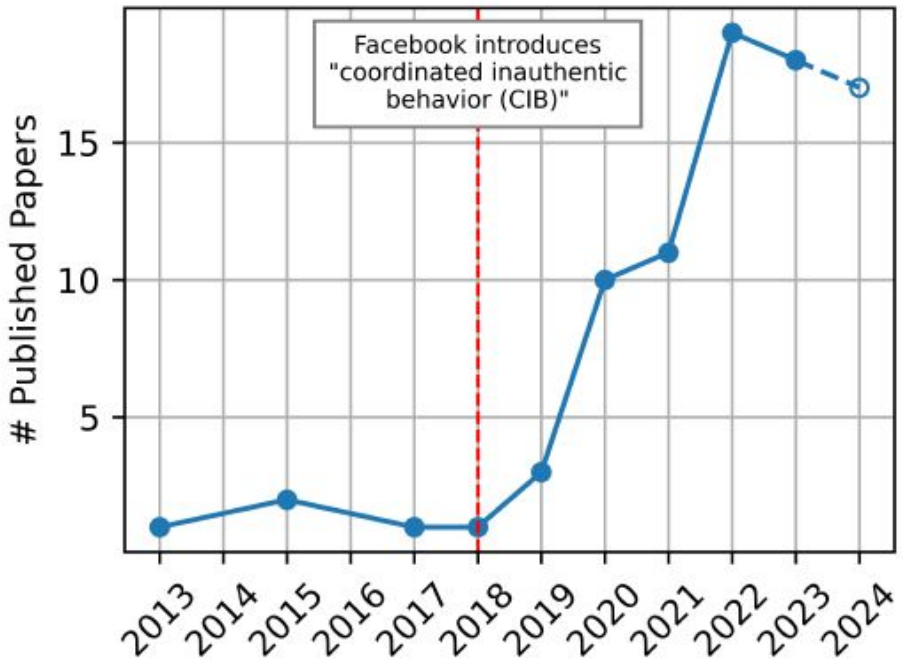


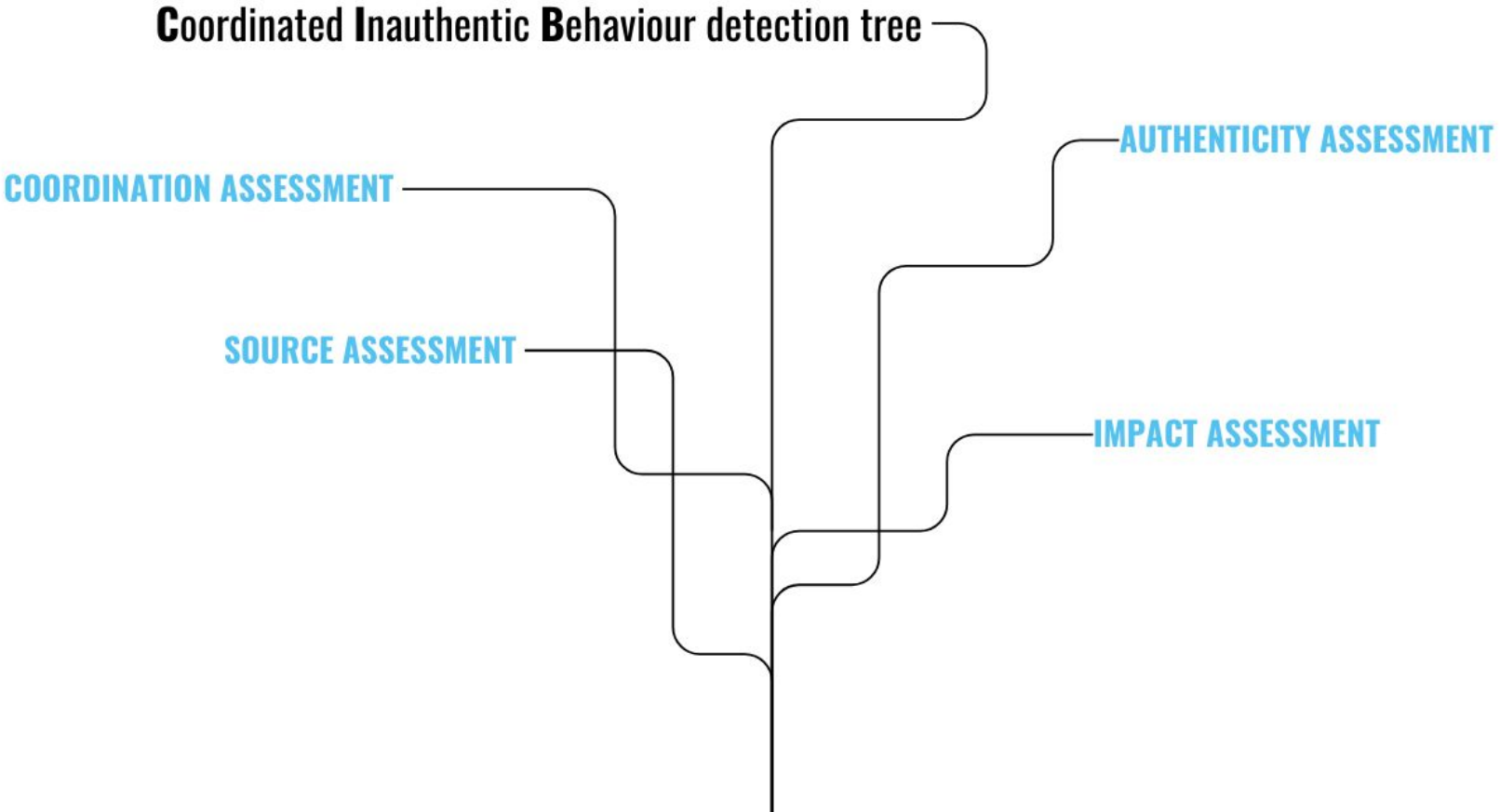
Fig. 2. Number of articles published yearly on coordinated online behavior. A steep rise is observed after Facebook introduced CIB in 2018.

Conceitos

- Principais Características do CIB:
  - Coordenação: Pessoas ou grupos agem em conjunto para alcançar um objetivo estratégico compartilhado, frequentemente usando uma combinação de contas autênticas e falsas.
  - Inautenticidade: Contas falsas são centrais para a operação, criadas especificamente para enganar as pessoas sobre quem está por trás da atividade. Em alguns casos, uma única fonte oculta usa muitas contas falsas, enquanto em outras situações, pessoas com contas reais se coordenam de maneira enganosa.
  - Manipulação: O propósito é enganar o público ou influenciar um debate. Isso pode ser alcançado inflando artificialmente a popularidade de uma pessoa ou ideia, criando uma falsa sensação de apoio popular, ou espalhando desinformação.
  - Atividade multiplataforma: CIB frequentemente opera em múltiplas plataformas, como sites de redes sociais, sites de notícias falsas e outros serviços de internet.

Conceitos

- Principais Características do CIB:



Conceitos

- Postagem Coordenada:
  - A postagem coordenada ocorre quando várias contas transmitem a mesma mensagem aparentemente independentes umas das outras de diferentes contas gerenciadas por uma pessoa ou equipe.
  - Um estudo do EU Disinfo Lab mostra que as contas CIB geralmente postam a mesma mensagem com alguns segundos de diferença.
  - A publicação coordenada de muitas imagens iguais ou semelhantes é uma estratégia popular em uma campanha de disseminação de informações de mídia mista que se refere ao uso de vários canais de mídia social e formatos multimídia para disseminar uma narrativa.
  - Agentes de campanha de informação geralmente usam aplicativos como plataformas de agendamento para controlar suas múltiplas contas ou copiar e colar manualmente as mensagens nas contas sob seu controle.

Conceitos

- Repostagem Coordenada:
  - A repostagem coordenada é a maneira mais simples pela qual uma conta falsa ou astroturfing pode amplificar uma mensagem de campanha e requer apenas um clique.
  - A simples repostagem de uma mensagem de campanha é um pouco óbvia e, portanto, não é usada em campanhas do mundo real.
  - Um fenômeno mais comum é quando várias contas amplificam as mensagens em conjunto ao republicar exatamente a mesma mensagem de terceiros.
  - A repostagem coordenada é uma estratégia empregada por muitos agentes e contas de campanha de informação, mesmo os altamente automatizados.

Conceitos

- Hashtagging Coordenado:
  - Hashtagging coordenado é quando várias contas usam a mesma hashtag dentro de um limite de tempo.
  - É uma tática comum em campanhas de desinformação ter postagens com pequenas variações de texto sob a mesma hashtag.
  - As contas tentam ofuscar suas hashtags coordenadas parafraseando textos semelhantes nas mensagens.
  - Outra tática comum em operações de desinformação é usar hashtags genéricas, como #corona e #covid19, provavelmente em uma tentativa de injetar seu conteúdo nas principais conversas online relacionadas à pandemia.

Conceitos

- Menção Coordenada:
  - A menção coordenada ocorre quando várias contas mencionam o mesmo usuário dentro de um limite de tempo, incluindo o nome de tela de outro (ou vários) usuário em uma postagem, permitindo injetar narrativa de desinformação no contexto da discussão.
  - Isso geralmente se aplica a celebridades ou entidades políticas em campanhas de desinformação política.

Conceitos

- Coordenação Genuína:
  - É importante distinguir CIB de coordenação genuína e orgânica ou respostas entusiásticas da comunidade.
  - **Mobilização Comunitária Genuína**: Grupos ou comunidades legítimas podem se coordenar organicamente para compartilhar informações, apoiar uma causa ou responder a um evento. A principal diferença geralmente é a transparência sobre sua afiliação e a autenticidade das contas participantes.
  - **Conteúdo Viral**: Conteúdo que se torna viral naturalmente verá compartilhamento generalizado e rápido de muitas contas independentes. Isso é distinto da amplificação fabricada vista em CIB.
  - **Grupos de Interesse Compartilhado**: Usuários com interesses compartilhados naturalmente discutirão e amplificarão tópicos similares. Isso carece da intenção enganosa ou manipuladora de CIB.

Conceitos

- Coordenação Inautêntica X Genuína:

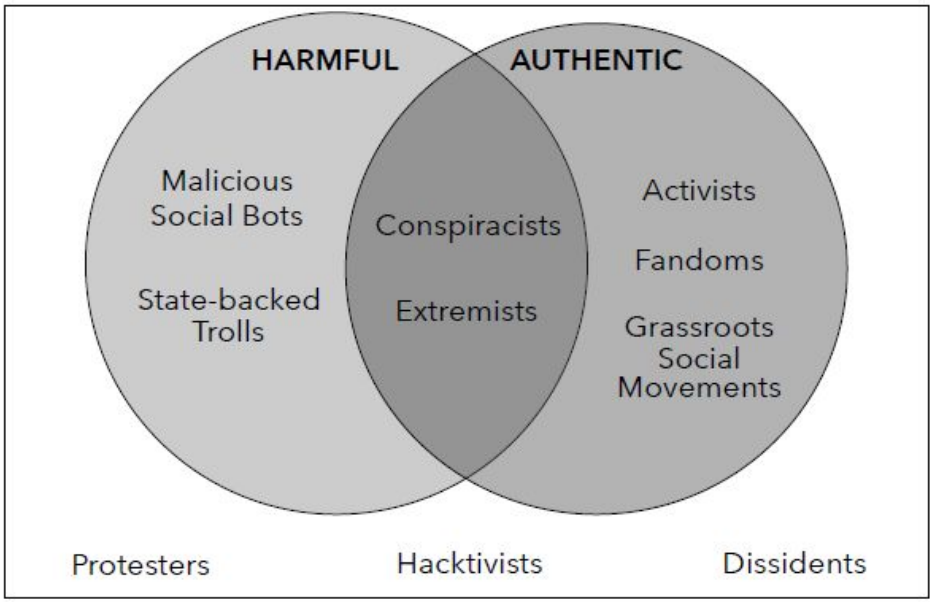


Fig. 3. Taxonomy of coordinated online behavior obtained by considering the dimensions of *harmfulness* and *authenticity* of our conceptual framework. The framework conveniently allows the mapping of disparate instances of online coordination.

Detecção

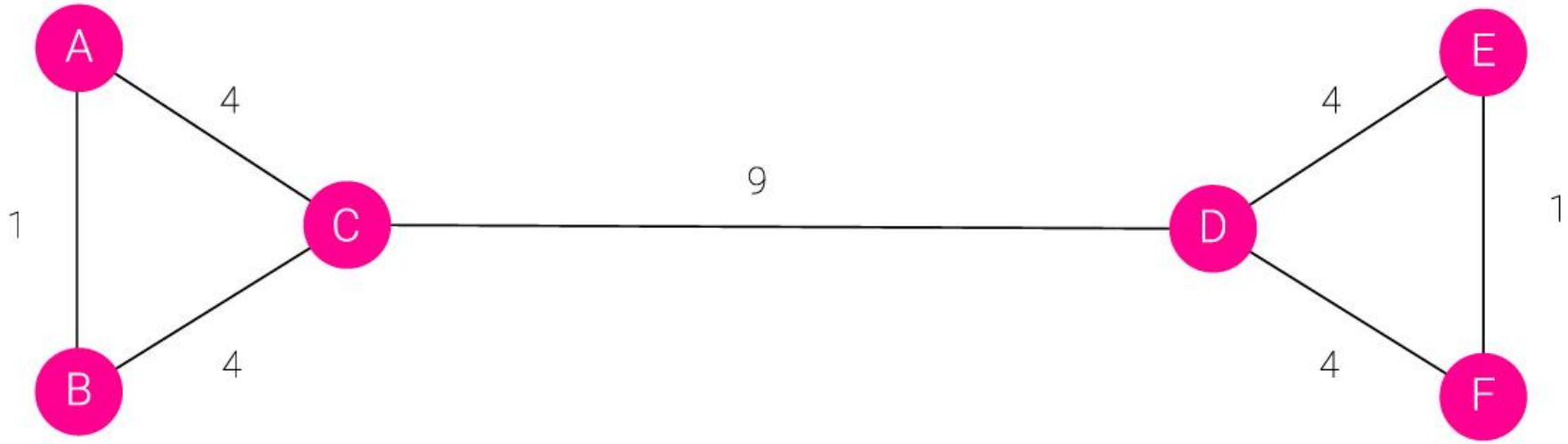
- Como Detectar o Comportamento Inautêntico Coordenado?
  - Modelamos aas relações entre usuários do Telegram na forma de grafos direcionados e valorados, considerando o envio de mensagens em grupos.
  - Nessa modelagem, cada nó representa um usuário e podemos considerar um grafo para cada tipo de mensagem: mensagem em geral, mensagem viral e mensagem com desinformação.

Detecção

- Como Detectar o Comportamento Inautêntico Coordenado?
  - Considerando o **grafo de mensagens gerais**, onde cada nó representa um usuário, existe uma aresta direcionada entre o usuário i e o usuário j se o usuário i enviou uma mesma mensagem que o usuário j. O peso dessa aresta é a quantidade de mensagens iguais (ou semelhantes) enviadas tanto pelo usuário i quanto pelo usuário j.
  - Um raciocínio análogo foi aplicado para criar um **grafo** apenas de **mensagens virais**: existe uma aresta direcionada entre o usuário i e o usuário j se tanto o usuário i quanto o usuário j postaram uma mesma mensagem viral e o peso dessa aresta é quantidade de mensagens virais enviadas tanto pelo usuário i quanto pelo usuário j.
  - O mesmo procedimento foi utilizado para criar o **grafo de desinformação**: existe uma aresta direcionada entre o usuário i e o usuário j se tanto o usuário i quanto o usuário j postaram uma mesma mensagem contendo desinformação e o peso dessa aresta é quantidade de mensagens com desinformação enviadas tanto pelo usuário i quanto pelo usuário j.

Detecção

- Como Detectar o Comportamento Inautêntico Coordenado?

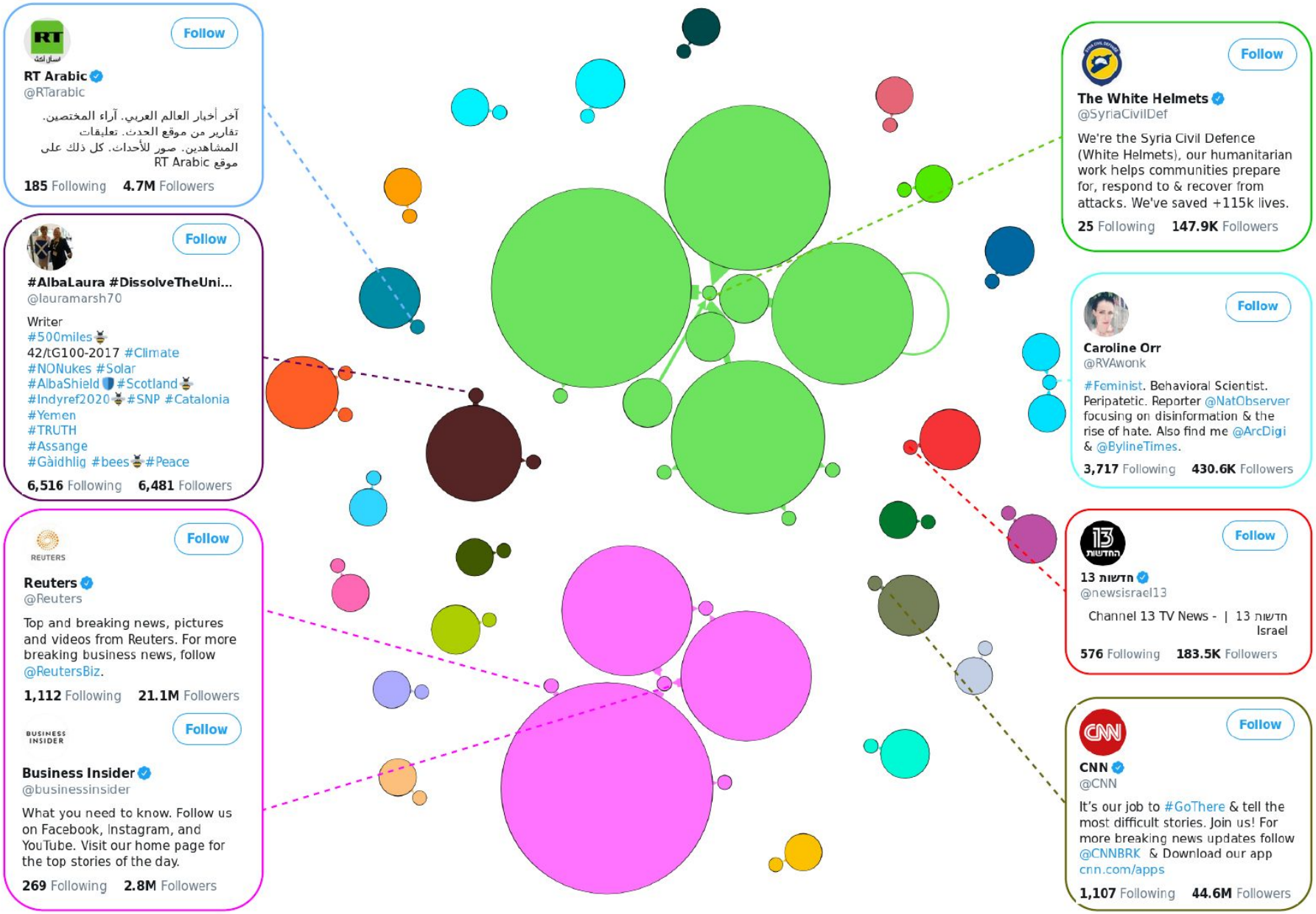


Detecção

- Rapid Retweet Network:
  - Uma primeira abordagem para detectar coordenação é identificar grupos de contas que consistentemente retweetam a mesma fonte.
  - Cria-se uma rede direta desenhando um link ponderado do retweetador (promotor) para a conta que produz a postagem original (promovida).
  - Como retweetar é algo comum, e para evitar rotular relações casuais como suspeitas, mantemos apenas as arestas se o retweet acontecer dentro de dez segundos após a postagem original, e se o promotor retweetar o promovido pelo menos duas vezes.
  - Assim, os pesos das arestas correspondem ao número de retweets rápidos.
  - Finalmente, extraímos os componentes conectados da rede para identificar grupos coordenados de contas promotoras e promovidas.

Detecção

- Rapid Retweet Network:
  - A rede resultante é mostrada na Fig. 1. A maioria dos componentes conectados são díades e tríades. O tamanho de um nó é proporcional à sua força de saída, ou seja, à intensidade de sua atividade promocional. As contas promovidas (nós menores) não estão se engajando em atividade de retweeting e geralmente são contas com muitos seguidores.


Follow
RT Arabic
@RTarabic
آخر أخبار العالم العربي. آراء المختصين. تقارير من موقع الحدث. تعليقات المشاهدين. صور للأحداث. كل ذلك على موقع RT Arabic
185 Following 4.7M Followers
Follow
#AlbaLaura #DissolveTheUni...
@lauramarsh70
Writer
#500miles
42/tG100-2017 #Climate
#NONukes #Solar
#AlbaShield #Scotland
#Indyref2020 #SNP #Catalonia
#Yemen
#TRUTH
#Assange
#Gàidhlig #bees #Peace
6,516 Following 6,481 Followers
Follow
Reuters
@Reuters
Top and breaking news, pictures and videos from Reuters. For more breaking business news, follow @ReutersBiz.
1,112 Following 21.1M Followers
Follow
Business Insider
@businessinsider
What you need to know. Follow us on Facebook, Instagram, and YouTube. Visit our home page for the top stories of the day.
269 Following 2.8M Followers
Follow
The White Helmets
@SyriaCivilDef
We're the Syria Civil Defence (White Helmets), our humanitarian work helps communities prepare for, respond to & recover from attacks. We've saved +115k lives.
25 Following 147.9K Followers
Follow
Caroline Orr
@RVAwonk
#Feminist. Behavioral Scientist. Peripatetic. Reporter @NatObserver focusing on disinformation & the rise of hate. Also find me @ArcDigi & @BylineTimes.
3,717 Following 430.6K Followers
Follow
13 חדשות
@newsisrael13
Channel 13 TV News - Israel | 13 חדשות
576 Following 183.5K Followers
Follow
CNN
@CNN
It's our job to #GoThere & tell the most difficult stories. Join us! For more breaking news updates follow @CNNBRK & Download our app cnn.com/apps
1,107 Following 44.6M Followers

Detecção

- Similar Tweet Network:
  - Ao buscar comportamento suspeito, considerar a similaridade de texto entre tweets originais pode ser mais revelador do que padrões simples de retweeting. A replicação pode ter diferentes motivações, como plágio e intermediação. Ao contrário dos retweets, tweets com conteúdo similar são considerados originais pela plataforma, e os links entre original e cópia são difíceis de detectar.
  - Adotamos uma estratégia baseada em similaridade de texto. Visando identificar grupos de contas que deliberadamente postam conteúdo similar, analisamos o texto de tweets originais, respostas e citações, mas ignoramos retweets. Medimos a similaridade de texto usando o algoritmo de Reconhecimento de Padrões Ratcliff/Obershelp. Consideramos todos os pares de tweets com uma similaridade acima de um limite fixo. Finalmente, construímos uma rede de coordenação de tweets similares de contas onde uma aresta indica que as duas contas conectadas foram responsáveis por pelo menos um par de tweets similares produzidos dentro de um curto intervalo de tempo.

Detecção

- Similar Tweet Network:
  - A rede tem dois parâmetros: o limite de similaridade de texto e o limite da janela de tempo. A seguir, determinamos ambos empiricamente.
  - Para reduzir a complexidade computacional implícita em medir a similaridade entre todos os tweets em nosso banco de dados, primeiro consideramos o fluxo longitudinal de tweets e medimos similaridades de texto apenas entre tweets separados por no máximo nove outros tweets ao longo do fluxo. Ordenamos todos os tweets cronologicamente e usamos o Python SequenceMatcher para comparar o texto de tweets consecutivos à distância +1 (vizinhos imediatos), +2 (um tweet no meio), e assim por diante, até +10 (o décimo tweet seguinte).
  - A Fig. 2 mostra que uma quantidade considerável de conteúdo similar é criada nos primeiros dez segundos após a postagem do tweet original. Portanto, usamos esse limite de janela de tempo para limitar as comparações pareadas entre tweets em nossa coleção.

Detecção

- Similar Tweet Network:
  - A Fig. 3 mostra as distribuições da similaridade de texto pareada para todos os pares de tweets criados com menos de dez segundos de diferença. Como esperado, a maioria do conteúdo é dissimilar e a probabilidade de encontrar tweets com conteúdo cada vez mais similar diminui monotonicamente.
  - Uma exceção surpreendente é que a probabilidade aumenta para similaridade maior que 70%. Portanto, definimos o limite de similaridade em 0,7. De agora em diante, por questão de legibilidade, quando nos referimos a tweets similares, queremos dizer pares de tweets que têm tanto conteúdo acima do limite de similaridade quanto horários de criação dentro do limite da janela de tempo.

Detecção

- Similar Tweet Network:

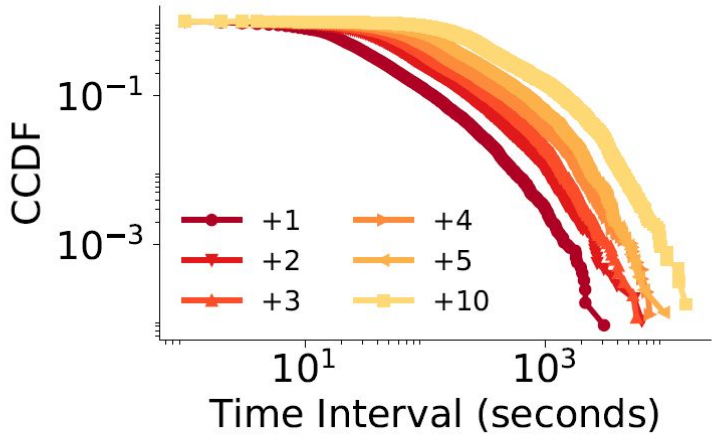


Figure 2: Distributions of time intervals between pairs of tweets with text similarity greater than 0.7. Most tweets replicating content are created a few seconds after the original. The distributions show similar patterns regardless of the sequential distance used to select tweets for comparison.

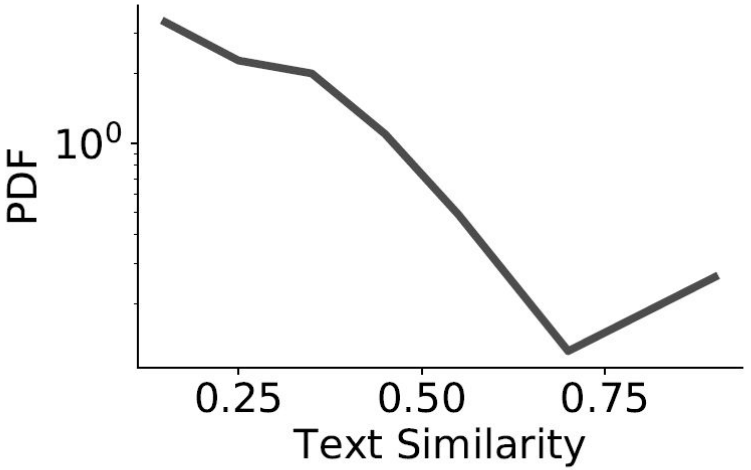
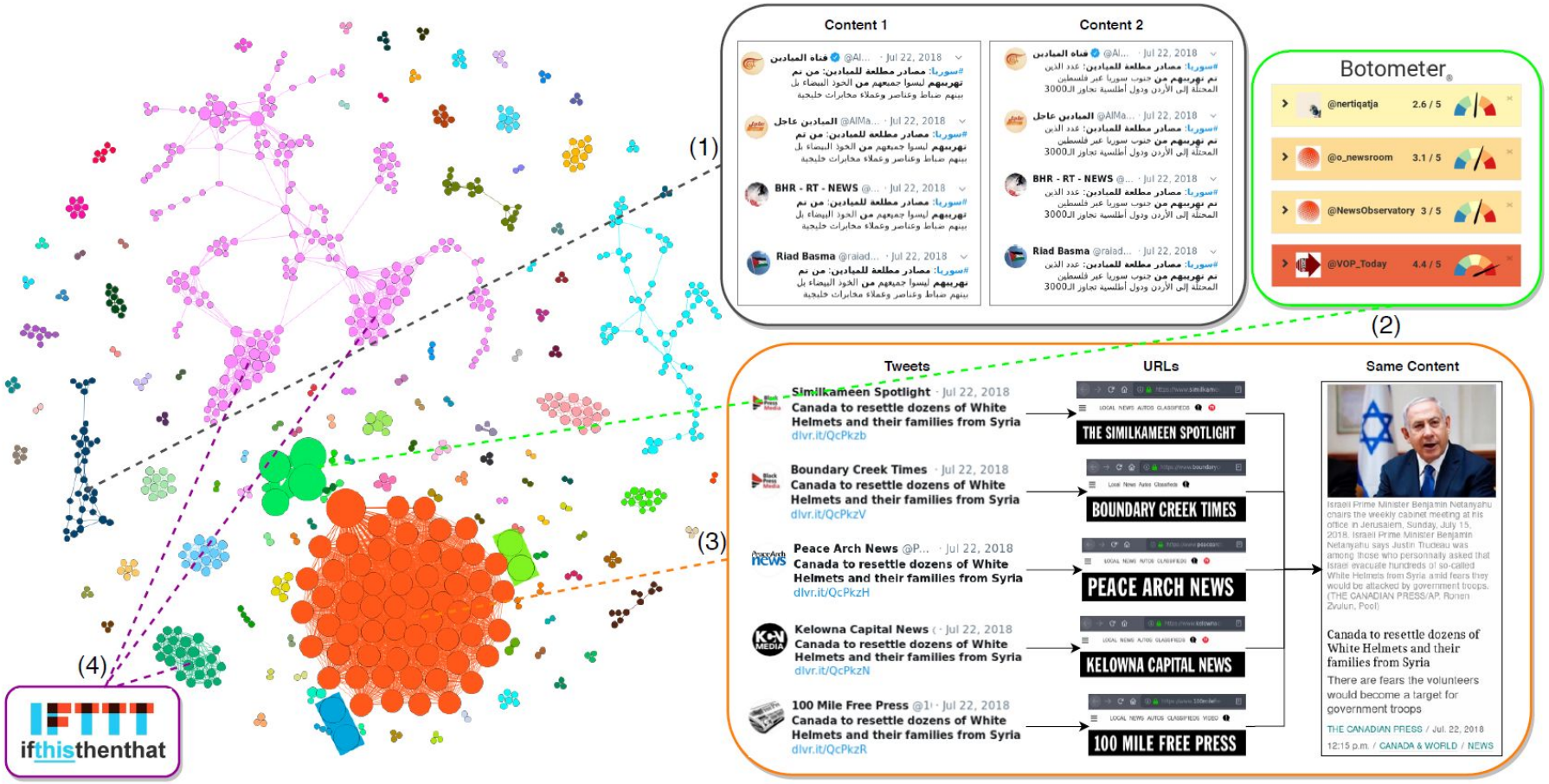


Figure 3: Distribution of text similarity. Pairs of tweets are compared if they were posted less than ten seconds apart.

Detecção

- Similar Tweet Network:



Detecção

- Similar Tweet Network:

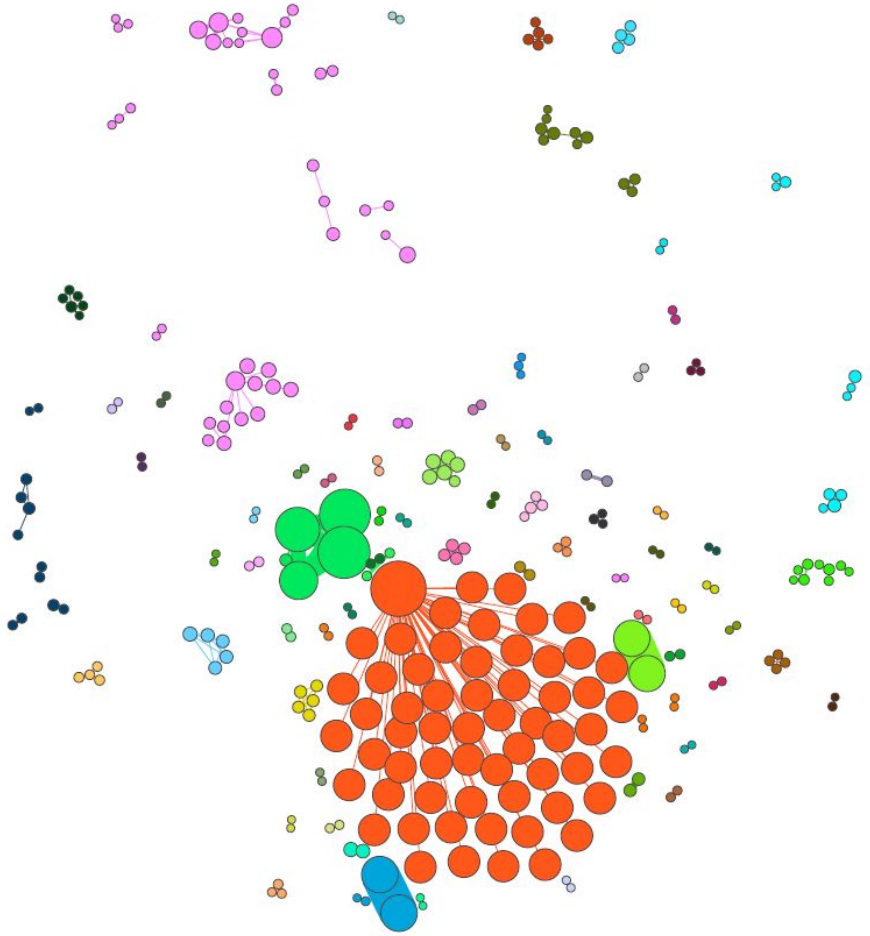


Figure 5: Coordination network as described in Fig. 4, but with a more restrictive minimum edge weight of 2.

Detecção

- Redes Bipartidas:
  - Etapa 1 - Extração de rastros comportamentais:
  - O ponto de partida da detecção de coordenação deve ser uma conjectura sobre comportamento suspeito. Assumindo que usuários autênticos são de certa forma independentes uns dos outros, consideramos uma surpreendente falta de independência como evidência de coordenação.
  - A implementação da abordagem é guiada por uma escolha de rastros que capturam tal comportamento suspeito.
  - Por exemplo, se conjecturamos que contas são controladas por uma entidade com o objetivo de amplificar a exposição de uma fonte de desinformação, poderíamos extrair URLs compartilhadas como rastros. Cenários de coordenação podem estar associados a algumas categorias amplas de rastros suspeitos: conteúdo, atividade, identidade ou uma combinação desses aspectos.

Detecção

- Redes Bipartidas:
  - Etapa 2 - Construção de rede bipartida:
  - O próximo passo é construir uma rede bipartida conectando contas e características extraídas de seus perfis e mensagens.
  - Nesta fase, podemos usar os rastros comportamentais como características, ou criar novas características derivadas dos rastros.
  - Por exemplo, análise de conteúdo pode produzir características baseadas em sentimento, posicionamento e enquadramentos narrativos.

Detecção

- Redes Bipartidas:
  - Etapa 2 - Construção de rede bipartida:
  - Características temporais como hora do dia e dia da semana poderiam ser extrapoladas a partir de metadados de timestamp.
  - Características poderiam ser criadas agregando rastros, por exemplo, agrupando localizações em países ou imagens em perfis de cores.
  - Características mais complexas poderiam ser criadas considerando conjuntos ou sequências de rastros.
  - A rede bipartida pode ser ponderada com base na força da associação entre uma conta e uma característica — compartilhar a mesma imagem muitas vezes é um sinal mais forte do que compartilhá-la apenas uma vez.
  - Os pesos podem incorporar normalização como IDF para considerar características populares; não é suspeito se muitas contas mencionam a mesma celebridade.

Detecção

- Redes Bipartidas:
  - Etapa 3 - Projeção na rede de contas:
  - A rede bipartida é projetada em uma rede onde os nós de contas são preservados, e arestas são adicionadas entre nós com base em alguma medida de similaridade sobre as características.
  - O peso de uma aresta na rede de coordenação não direcionada resultante pode ser calculado via co-ocorrência simples, coeficiente de Jaccard, similaridade de cosseno, ou métricas estatísticas mais sofisticadas como informação mútua ou χ^2.
  - Em alguns casos, cada aresta na rede de coordenação é suspeita por construção.
  - Em outros casos, as arestas podem fornecer sinais ruidosos sobre coordenação entre contas, levando a falsos positivos.

Detecção

- Redes Bipartidas:
  - Etapa 3 - Projeção na rede de contas:
  - Por exemplo, contas compartilhando vários dos mesmos memes não são necessariamente suspeitas se esses memes são muito populares.
  - Nesses casos, curadoria manual pode ser necessária para filtrar arestas de baixo peso na rede de coordenação para focar nas interações mais suspeitas. Uma maneira de fazer isso é preservar arestas com um percentil superior de pesos.

Detecção

- Redes Bipartidas:
  - Etapa 4 - Análise de clusters:
  - O passo final é encontrar grupos de contas cujas ações são provavelmente coordenadas na rede de contas.
  - Algoritmos de detecção de comunidades em redes que podem ser usados para este propósito incluem componentes conectados, k-core, k-cliques, maximização de modularidade e propagação de rótulos, entre outros.

Detecção

- Redes Bipartidas:

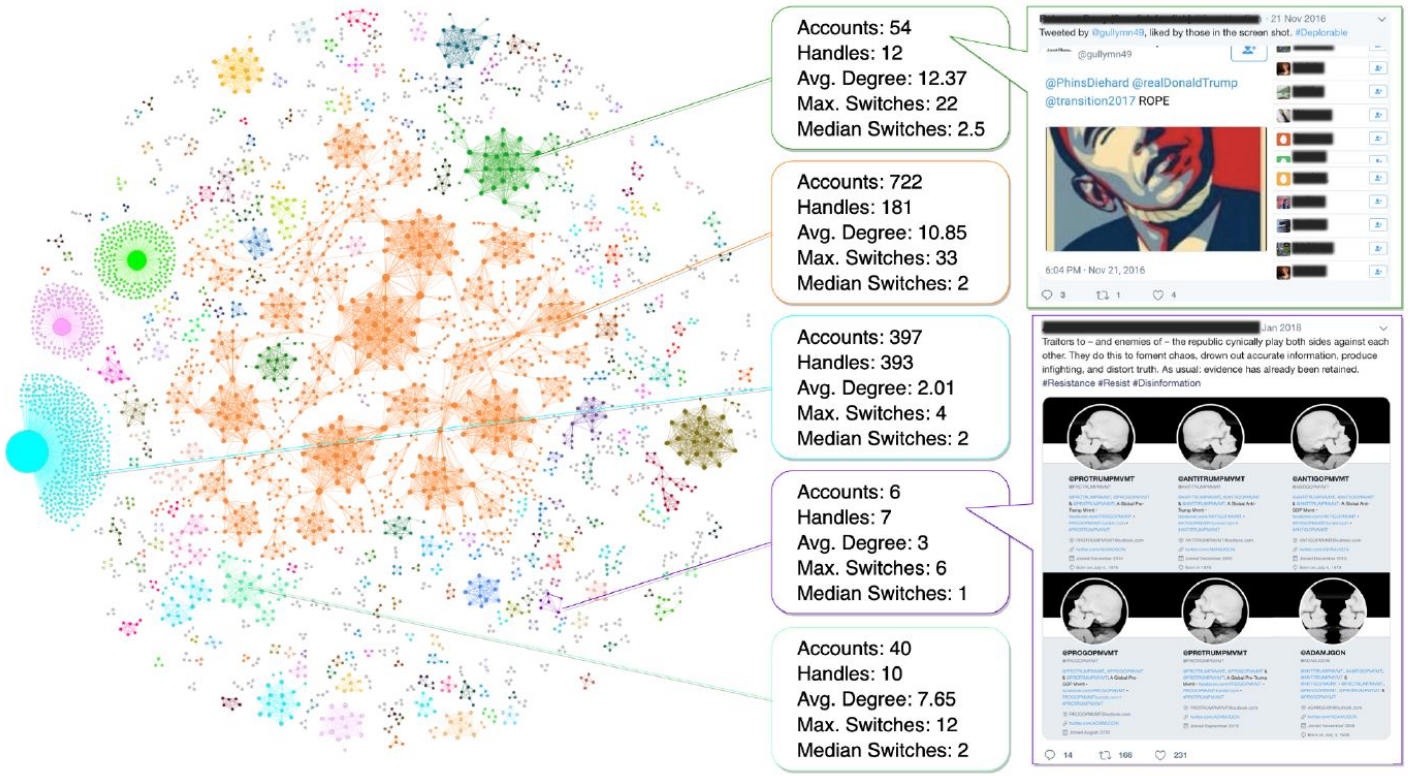
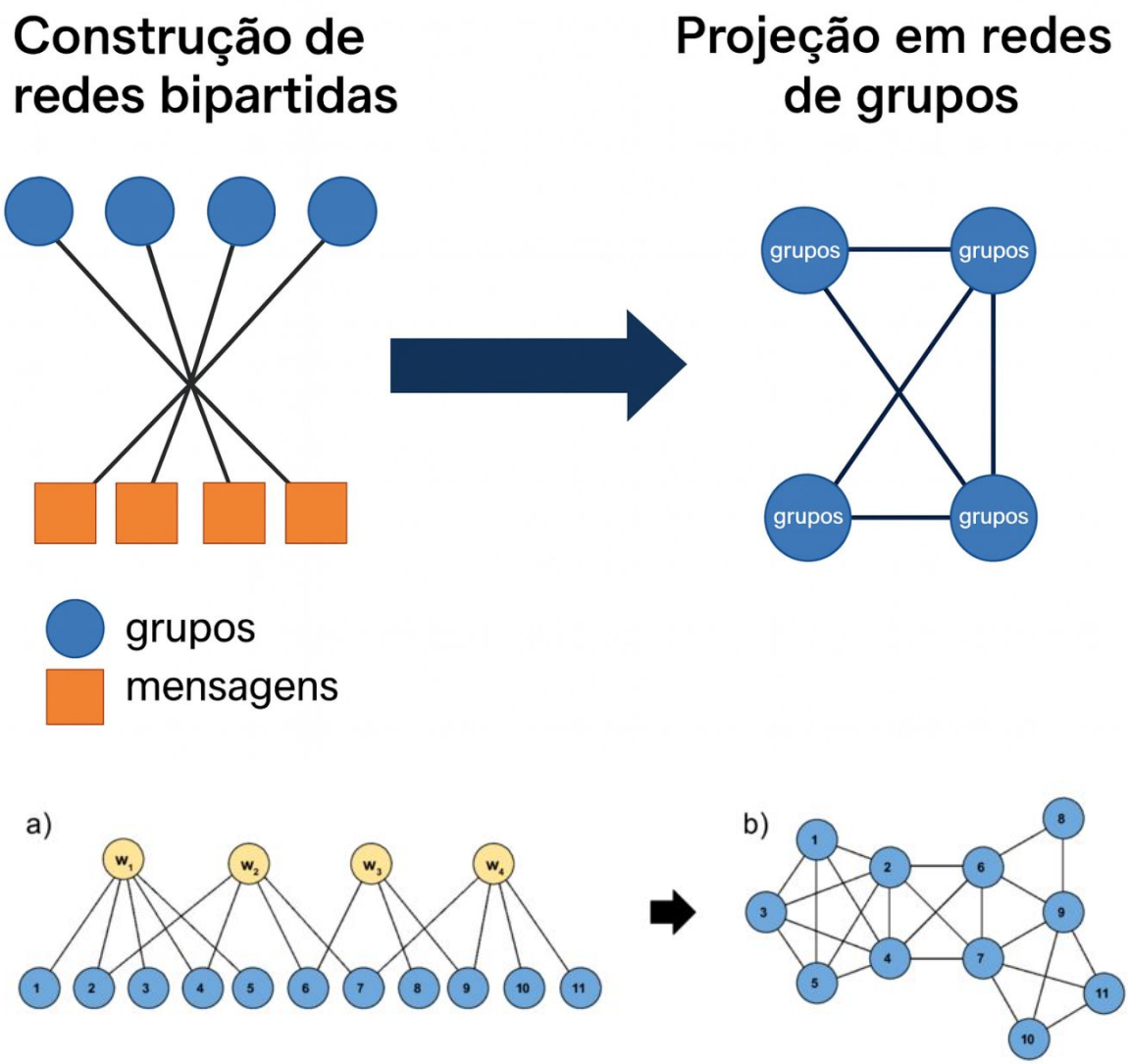


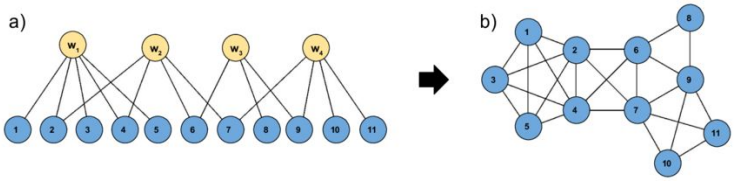
Figure 2: A handle sharing network. A node represents a Twitter account and its size is proportional to the number of accounts with which it shares handles. The weight of an edge is the number of unique handles shared by two accounts. Suspicious coordinated groups are identified by different colors. We illustrate the characteristics of a few coordinated groups, namely the number of accounts, number of shared handles, average number of accounts with which handles are shared, and the maximum and median number of times that a handle is switched among accounts. The number of switches is a lower-bound estimate on the basis of our data sample. We also show tweets by independent parties who uncovered the malicious activity of a couple of the coordinated groups, discussed in the main text.

Detecção

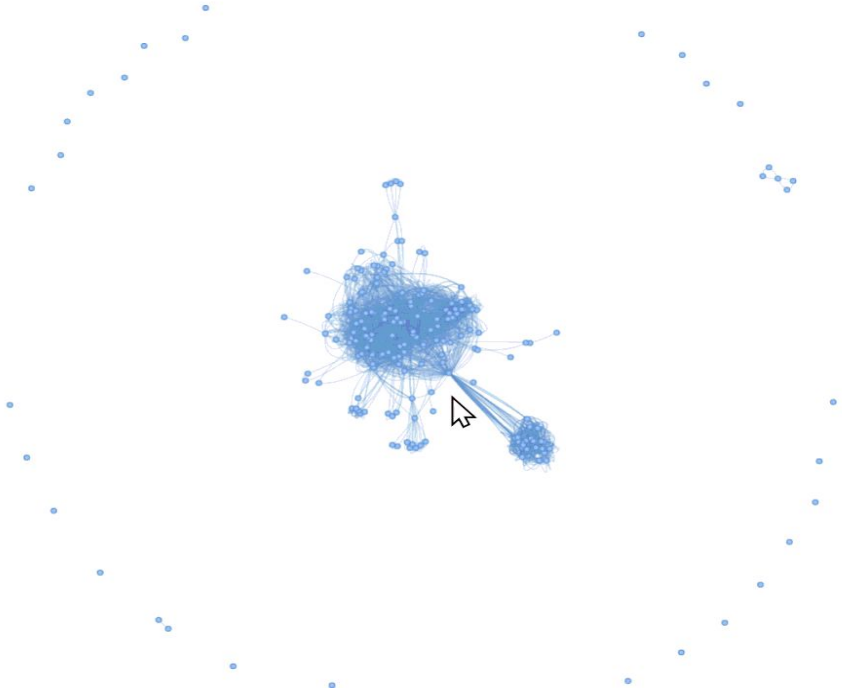
- Redes Bipartidas:



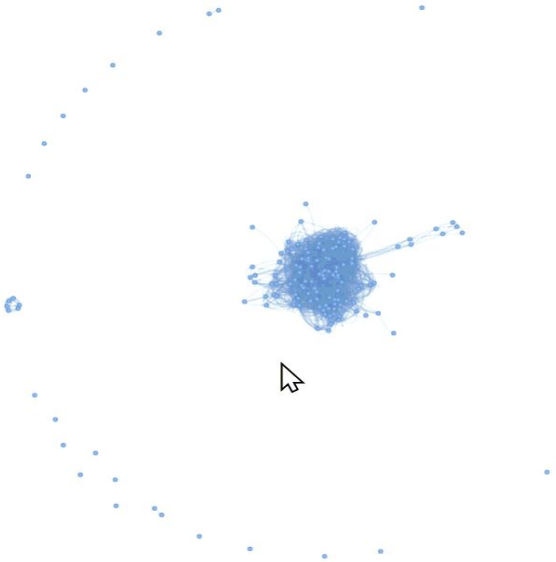
Redes Projetadas



Período Pré-eleitoral.



Período Pós-eleitoral.



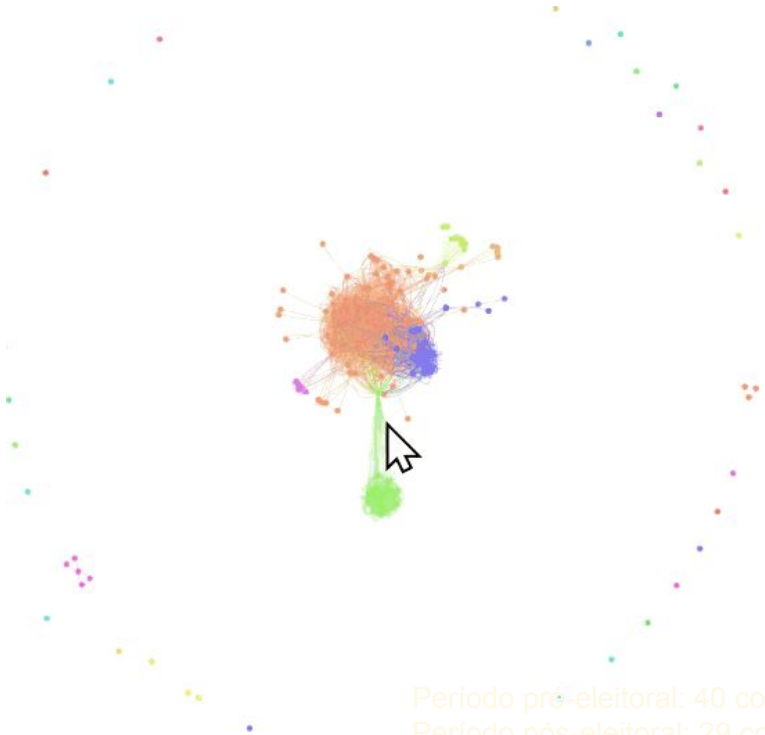
Decomposição Núcleo-Periferia

Estatística	Pré-eleitoral	Pós-eleitoral
Total de Grupos Conectados	154	**192**
Grupos no Núcleo	23	**39**
Grupos na Periferia	131	**153**

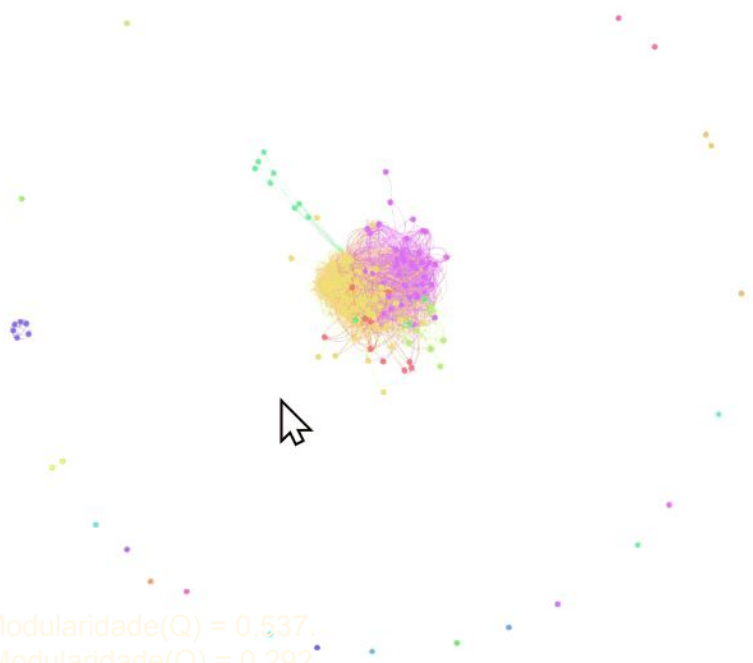
- Dos grupos conectados nas redes projetadas, 116 estavam presentes em ambos os períodos, dos quais 9 migraram da periferia para o núcleo, e 5 fizeram o caminho inverso. Além disso, 38 grupos deixaram a rede projetada, enquanto outros 76 entraram na rede projetada.
- Vale lembrar que a quantidade de grupos observados diminuiu de 327 para 241, indicando que alguns grupos foram removidos logo após a divulgação dos resultados das eleições de 2022.

Detecção de Comunidades

Período Pré-eleitoral.



Período Pós-eleitoral.



Período pré-eleitoral: 40 comunidades e Modularidade(Q) = 0,537
Período pós-eleitoral: 29 comunidades e Modularidade(Q) = 0,292

Análise de Centralidade

- Nenum grupo no ranking dos top 10 grupos de maior centralidade (Degree Centrality) no período pré-eleições 2022 se manteve neste ranking no período pós-eleições 2022.
- Esses grupos perderam importância (centralidade).

Detecção

- Medindo a Coordenação:
  - Para a etapa de seleção de usuários, restringimos nossa análise aos superdisseminadores, definidos como o 1% superior dos usuários que compartilharam mais retweets.
  - Medimos a similaridade entre superdisseminadores em termos de co-retweets, a fim de destacar usuários que frequentemente recompartilham as mesmas mensagens.
  - Para cada superdisseminador, calculamos um vetor ponderado por TF.IDF dos IDs de tweets que ele/ela retweetou.
  - Usar ponderação TF.IDF desconta tweets virais de influenciadores e usuários populares, enquanto enfatiza retweets de tweets impopulares.
  - Em seguida, calculamos a similaridade entre todos os pares de superdisseminadores como a similaridade de cosseno entre seus vetores correspondentes, obtendo assim uma rede de similaridade de usuários ponderada e não direcionada.

Detecção

- Medindo a Coordenação:
  - Filtramos a rede calculando sua espinha dorsal multiescala, que permite reter apenas estruturas de rede estatisticamente significativas.
  - Então, aplicamos o conhecido algoritmo de detecção de comunidades Louvain para agrupar usuários em comunidades de rede.
  - Finalmente, aplicamos desmontagem de rede, que atribui uma pontuação de coordenação a cada usuário na rede. Realizamos esta última etapa removendo iterativamente arestas e nós da rede com base em um limiar móvel de peso de aresta. Em cada iteração, removemos todas as arestas cujo peso bruto era inferior a um limiar, e tais que acabaram sendo desconectados do maior componente conectado. O limiar aumentou a cada iteração, até que a rede fosse completamente desmontada, ou seja, nenhum nó conectado permanecesse. Para cada nó, atribuímos uma pontuação de coordenação como o valor do limiar que desconectou aquele nó do resto da rede. Normalizamos a pontuação de coordenação no intervalo [0, 1], com 1 indicando coordenação máxima.

Detecção

- Principais Definições Operacionais:

Table 1. Examples of operational definitions used in recent academic literature. No definition is general enough to comprehensively describe coordinated online behavior. However, each definition grasps one or more relevant properties (highlighted in bold) that we leverage in our framework.

reference	definition
Nizzoli et al. [98]	Unexpected, suspicious, or exceptional **similarity** between a number of users
Cinelli et al. [19]	The number of times two accounts **behaved similarly**, such as when they **repeatedly** retweet the same post
Giglietto et al. [47]	The act of making people and/or things be involved in an **organized cooperation**
Magelinski and Carley [73]	**Many instances** of a tweet-behavior, i.e. tweeted hashtag [...] within a **small predetermined time window**
Weber and Neumann [137]	Anomalous levels of **coincidental behavior**
Magelinski et al. [74]	Users [that] take the **same actions within minutes** of one another
Zhang et al. [142]	Accounts that **co-appear**, or are **synchronized** in time
Pacheco et al. [101]	[Users exhibiting a] surprising **lack of independence**
Hristakieva et al. [55]	Coordination between users implies a **shared intent**
Keller et al. [62]	A group of people who want to **convey specific information** to an audience

Detecção

- Pipeline de Pesquisa em CIB:

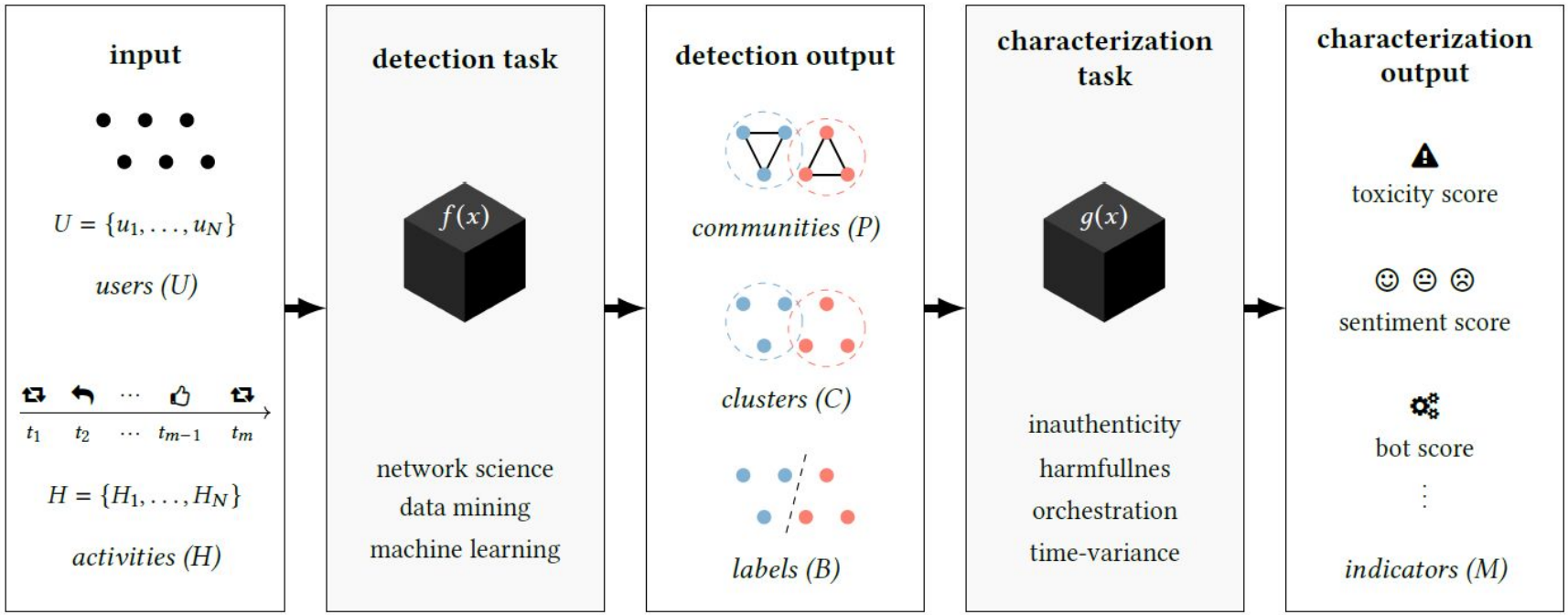


Fig. 4. The analytical process of studying coordinated online behavior, involving the *detection* and *characterization* tasks. The input to the overall process is a set of users U and their activities H on one or more platforms. The output of the detection task is either a set of binary labels B, clusters C, or network communities G that differentiate coordinated and non-coordinated users. The characterization task receives these in input and outputs a set of indicators M.

Detecção

- Pipeline de Pesquisa em CIB:

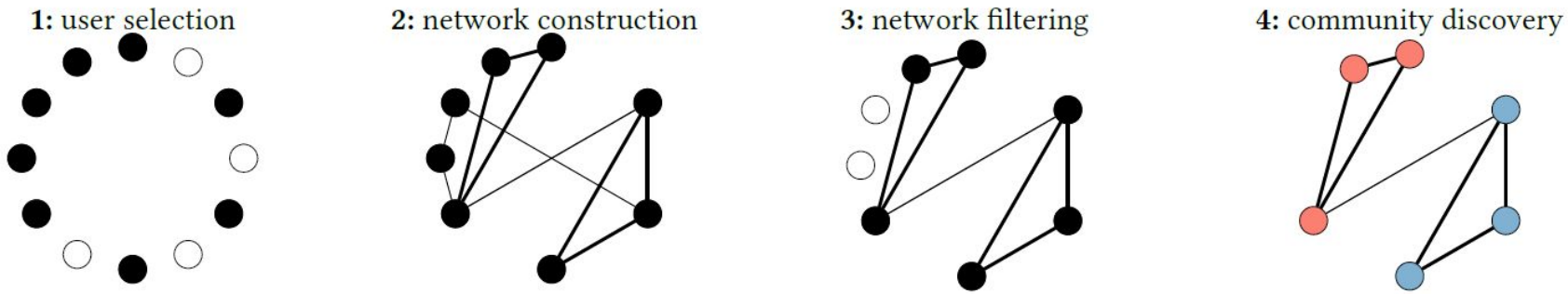


Fig. 5. Main steps of the network science methods for the detection of coordinated online behavior. **1:** The selected users become nodes in a network. **2:** User similarities are computed with a similarity function and assigned to the edge weights of the network. **3:** The network is filtered so as to retain only similarities with given properties. **4:** Community discovery is performed to detect groups of strongly coordinated users.

Detecção:

- Pipeline de Pesquisa em CIB:

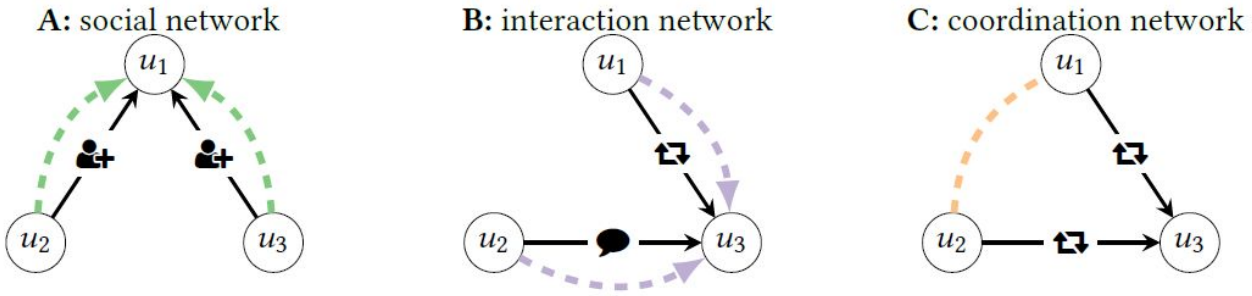


Fig. 6. Differences between social (**A**), interaction (**B**), and coordination (**C**) networks. Solid black edges represent actions on the online platform, while dashed colored edges show how actions are translated into edges in the corresponding type of network. Coordination networks are typically undirected and link users performing similar actions at around the same time. Differently to social and interaction networks, coordination networks allow connecting users even if they never directly interact with one another.

Table 2. Network science methods for detecting coordinated behavior based on *single layer* user networks. For each group of works we report the considered co-actions, similarity functions, filtering criteria, and community detection methods.

reference	action	similarity	filters†	community detection
[18]	retweet	cardinality	threshold, ADJ	modularity clustering
[50]	retweet	cardinality	EDO	Louvain
[65]	retweet	cardinality	threshold, ADO	Louvain
[67]	retweet	cardinality	backbone, ADJ	Louvain
[108]	retweet	cardinality	EDO	
[19, 28, 55, 69, 98, 120]	retweet	cosine similarity TF-IDF	backbone	Louvain
[121]	retweet	cosine similarity TF-IDF	backbone, EDO	Leiden
[136]	retweet	cardinality	threshold, ADJ	
[17]	retweet, tweet	cardinality	threshold, EDO	Louvain, connected components
[62]	retweet, tweet	cardinality	threshold, EDO	Louvain
[109]	retweet, tweet	cardinality	threshold, EDO	
[26]	tweet	cardinality	ADO	
[66]	tweet	cardinality	threshold	cohesive campaign
[100]	tweet	text similarity	threshold, ADO	
[92, 94]	parley	cardinality	threshold, kNN graph	Leiden
[132]	text	cardinality	backbone	Louvain
[91]	tweet, URL	cardinality,cosine similarity	threshold, EDO	Louvain
[93]	tweet, parley, URL, username	text similarity, cardinality, cosine similarity	threshold, kNN graph	Louvain
[72]	retweet, tweet, URL, hashtag	cosine similarity TF-IDF, text similarity	threshold, ADO	
[137, 138]	retweet, URL, hashtag, mention, reply	cardinality	ADJ	focal structures
[130]	retweet, tweet, URL, hashtag, mention	cosine similarity TF-IDF	ADJ	Leiden
[101]	retweet, hashtag, image, handle change, synchronization	Jaccard coefficient, cardinality, cosine similarity	threshold, EDO	
[14]	URL	cardinality	kNN graph	Louvain
[11, 44, 46, 47, 52, 105]	URL	cardinality	threshold, ADO	connected components
[1]	URL, hashtag, mention	unweighted	EDO	Leiden
[90]	URL, hashtag, mention	cardinality	threshold, EDO	Louvain
[45, 115]	URL, text-image	cardinality	threshold, ADO	connected components
[95]	image	cardinality	threshold, kNN graph	
[140]	image, video	cardinality	threshold, ADO	connected components
[13]	hashtag	cardinality	threshold	connected components
[88]	hashtag	cardinality	threshold, backbone	Louvain
[134]	hashtag	cardinality		
[57]	like	cardinality	threshold	
[63]	comment	cardinality	threshold	k-means, hierarchical clustering
[87]	mention	cosine similarity TF-IDF	threshold	Louvain

† ADJ: adjacent time window, EDO: evenly distributed overlapping time window, ADO: action-driven overlapping time window

Ferramentas

- Ferramentas para Detecção de CIB:

VI. CIB DETECTION TOOLKIT

The following table gathers the most common tools for identifying CIB and their applicability throughout the various branches.

Tools	Description	Coordination	Authenticity	Source	Impact
Vera AI Alerts[4]	Helps identify the coordination of networks based on account interactions.	x			x
InVID-WeVerify Plugin	Knowledge verification platform. Key features: access to contextual information on Facebook and YouTube videos; reverse image search; fragmentation of videos into keyframes; enhancement of keyframes through a magnifying lens; video and image metadata reading; forensic filters on still images; voice cloning detection; synthesis images; and deep fakes detection. The Xnetwork tool is available to accomplish cross-network queries. The plugin also has two data analysis tools: for X (working with past data) and Facebook/Instagram (relying on CSV files exported from Crowdtangle).	x	x	x	x
Gephi Graph Viz	Network analysis and visualisation tool.	x		x	x
Cytoscape	Network analysis, visualisation and data integration tool.	x			x
NodeXL	Network analysis and visualisation tool (it is an add-in for Microsoft Excel extending the spreadsheet software's capabilities).	x		x	x
CooRnet	Detects coordinated link sharing behaviour (CLSB) and outputs the network of entities that performed such behaviour				
Coordination Network Toolkit	Detects coordination networks in Twitter and other social media data				

Ferramentas

- Ferramentas para Detecção de CIB:

CooRTweet	Coordinated Networks Detection on Social Media				
Maltego	Graphical analysis and data integration tool. Key features: mapping out relationships and networks between various entities.	x		x	
DataMiner	Web scraping tool designed for extracting data from websites and online platforms.	x		x	
Graphika	Tool designed to analyse and visualise large-scale social networks: identifying patterns, clusters, and anomalies within social media data	x		x	
Nisos	Identification and analysis tool of complex adversarial behaviours, including disinformation and coordinated inauthentic activity.	x	x		
Cyber Triage	Digital forensics and incident response tool. Key features: automated data collection, triage analysis, artefact analysis and creating a timeline of events.			x	x
WHOIS	IP address and domain analysis tool. Key features: domain information retrieval, IP address lookup, and registration details.			x	
ICAAN	IP address and domain analysis tool (coordinating and managing the global Domain Name System). The registration data lookup tool includes searching for registration data, domain names and Internet number resources.			x	
CrowdTangle[5]	Social media analysis tool. Key features: track, analyse and report what is happening on Facebook, Instagram, Reddit, and X.	x	x	x	x

Ferramentas

- Ferramentas para Detecção de CIB:

Social Blade	Social media analysis tool. Key features: statistics, analytics and estimated earnings for content creators and influencers using YouTube, Twitch, Instagram and X.		x	x	x
Openmeasures (former SMAT)	Social media analysis tool. Key features: analysis of timeline, activity and link counter. API access to the complete library of sources.		x	x	x
Hootsuite	Social media management platform. It offers some features for research purposes: analytics and reporting tools, news, trends, and identifying other relevant content.		x	x	x
BuzzSumo	Content virality tracking tool. Key features: researching and social media analytics. Identification of trending topics. Content performance analysis.				x
NewsWhip	Content virality tracking tool. Key features: track and analyse news stories and social media content.				x
TextBlob	Provides a simple API for performing various natural language processing tasks, including sentiment analysis. It offers a pre-trained sentiment analysis model to classify text as positive, negative, or neutral.		x		x
Lexalytics	Offers a range of features for sentiment analysis, entity recognition, theme extraction, and more.		x		x
TinEye	Reverse image search engine.		x		
Bing Image	Reverse image search engine.		x		
Yandex Image	Reverse image search engine.		x		
Google Image	Reverse image search engine.		x		

Ferramentas

- Ferramentas para Detecção de CIB:

GoodCreator	Analytics tool to track content performance across platforms.		x		
MISP	Malware information sharing platform. Key feature: collaborative sharing of analysis and correlation (beyond efficient analysis, useful to grasp Tactics, Techniques, and Procedures (TTPs), related campaigns and attribution).			x	x

Ferramentas

- 10 Ferramentas para Monitoramento de Redes Sociais:
  - Hootsuite
  - Google Alerts
  - HubSpot
  - TweetReach
  - BuzzSumo
  - Mention
  - Stilingue
  - Vtracker
  - Buzzmonitor
  - Knewin

Referências

- Coordinated inauthentic behavior and information spreading on Twitter. Matteo Cinelli, Stefano Cresci, Walter Quattrociocchi, Maurizio Tesconi, Paola Zola (2022).
- Unveiling Coordinated Groups Behind White Helmets Disinformation. Diogo Pacheco, Alessandro Flammini, Filippo Menczer (2020).
- Detection and Characterization of Coordinated Online Behavior: A Survey (2024).
- Coordinated Inauthentic Behavior on TikTok: Challenges and Opportunities for Detection in a Video-First Ecosystem. Luca Luceri, Tanishq Vijay Salkar, Ashwin Balasubramanian, Gabriela Pinto, Chenning Sun, Emilio Ferrara (2025).
- Uncovering Coordinated Networks on Social Media: Methods and Case Studies. Diogo Pacheco, Pik-Mai Hui, Christopher Torres-Lugo, Bao Tran Truong, Alessandro Flammini, Filippo Menczer (2021).
- Identifying Coordinated Accounts on Social Media through Hidden Influence and Group Behaviours (2021).

Referências

- A Synchronized Action Framework for Detection of Coordination on Social Media. Thomas Magelinski, Lynnette Hui Xian Ng and Kathleen M. Carley.
- CATCH ME IF YOU CAN: ON THE DETECTION OF COORDINATED INAUTHENTIC BEHAVIOR ON SOCIAL MEDIA AND ITS LIMITS. Christopher Torres Lugo (2023).
- Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign. Franziska B. Keller, David Schoch, Sebastian Stier & JungHwan Yang (2020).
- The Spread of Propaganda by Coordinated Communities on Social Media (2022).
- Leonardo Nizzoli, Serena Tardelli, Marco Avvenuti, Stefano Cresci, and Maurizio Tesconi. 2021. Coordinated behavior on social media in 2019 UK general election. In AAAI ICWSM.
- Temporal Dynamics of Coordinated Online Behavior: Stability, Archetypes, and Influence. Serena Tardelli, Leonardo Nizzoli, Maurizio Tesconi, Mauro Conti, Preslav Nakov, Giovanni Da San Martino, Stefano Cresci.

Referências

- Amplifying influence through coordinated behaviour in social networks. Derek Weber and Frank Neumann (2021).