Oct 4, 2019

Disinformation: Detect to Disrupt

Something is wrong on the internet.

Disinformation has undergone a **technological upgrade** with the rise of social media.

Modern Disinformation Tactics

Upgraded tactics

Exploit the ease of creating **false personas**

Upgraded tactics

Exploit the ease of creating **false personas**

Test **messaging** on fringe platforms before moving to mainstream

Upgraded tactics

Exploit the ease of creating **false personas**

Test **messaging** on fringe platforms before moving to mainstream

Manufacturer consensus as a proxy for the truth by **coordinating** many fake accounts

Upgraded tactics

Exploit the ease of creating **false personas**

Test **messaging** on fringe platforms before moving to mainstream

Manufacturer consensus as a proxy for the truth by **coordinating** many fake accounts

Exploit **targeting algorithms** to direct content to those most likely to propagate it

Prior Work

Prior Approaches

Focus on fact checking or bot detection	Requires manually curated data (labelled examples or knowledge base)
Focus on content or accounts in isolation	Blind to larger patterns in content and the behavior of groups of accounts
Focus on small-scale prediction	Little consideration given to the explainability or scalability

Thesis

The **tactics** that make disinformation campaigns effective leave **statistical trails** that make them **possible to detect** without evaluating content-specific properties.

Our Approach

Use **content-agnostic** representations to observe **patterns of media dissemination** across **networks of accounts**.

The Goal

To **disrupt** disinformation campaigns, we must detect them them **before** they achieve **mass reach**.

Technical Challenges

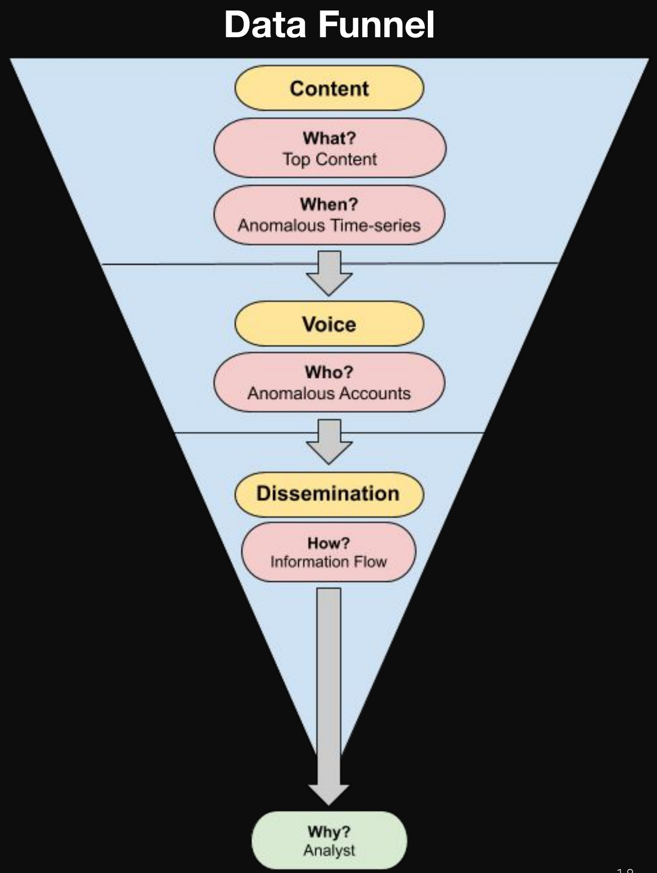
Early detection requires adaptive, **real-time analysis.**

Sufficient coverage requires processing a **high volume** of multi-platform, multimedia data.

Solutions must be **content and platform-agnostic** to address global campaigns and novel domains.

The Data Funnel

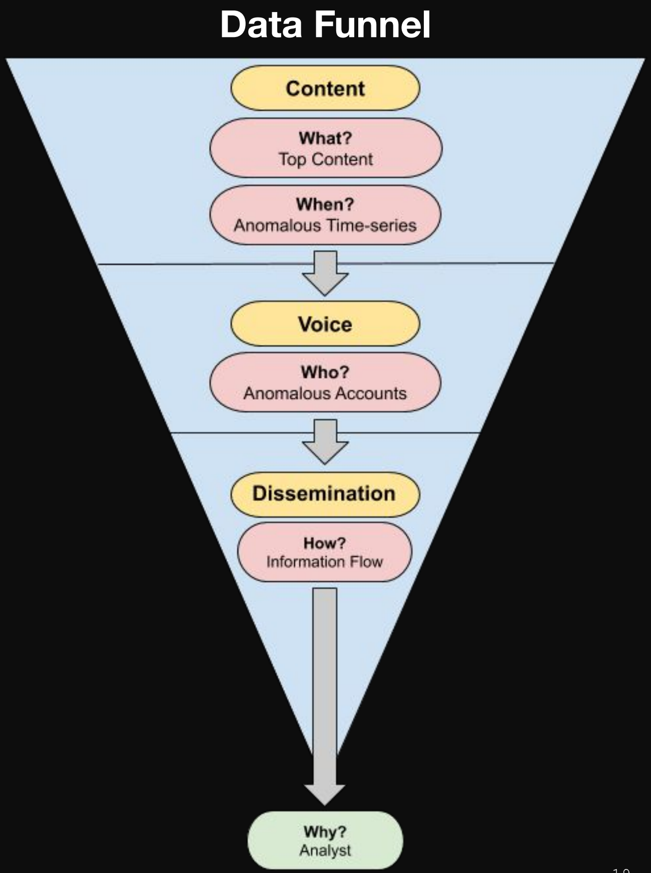
1. Use light-weight, scalable algorithms to identify **potentially anomalous content.**



The Data Funnel

1. Use light-weight, scalable algorithms to identify **potentially anomalous content.**

2. Look for **patterns** of in the behavior of **groups of accounts** that suggest coordination.

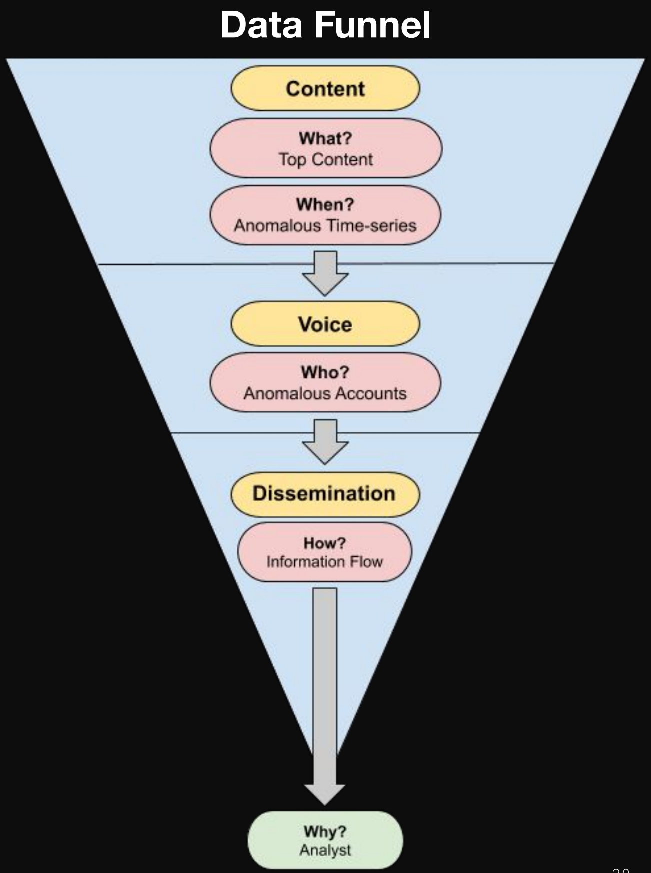


The Data Funnel

1. Use light-weight, scalable algorithms to identify **potentially anomalous content.**

2. Look for **patterns** of in the behavior of **groups of accounts** that suggest coordination.

3. Track how content is **disseminated** for patterns characteristic of a disinformation campaign.



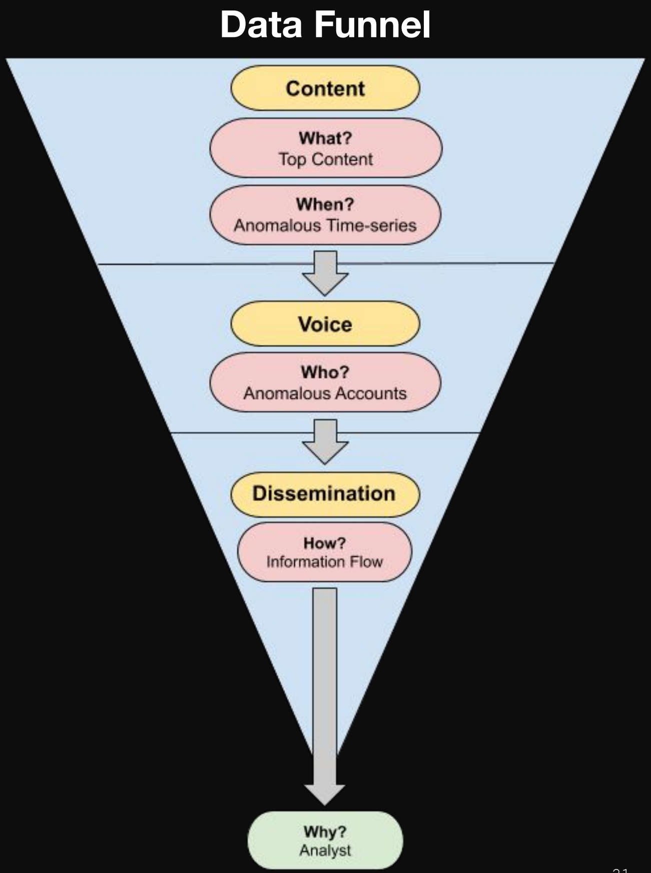
The Data Funnel

1. Use light-weight, scalable algorithms to identify **potentially anomalous content.**

2. Look for **patterns** of in the behavior of **groups of accounts** that suggest coordination.

3. Track how content is **disseminated** for patterns characteristic of a disinformation campaign.

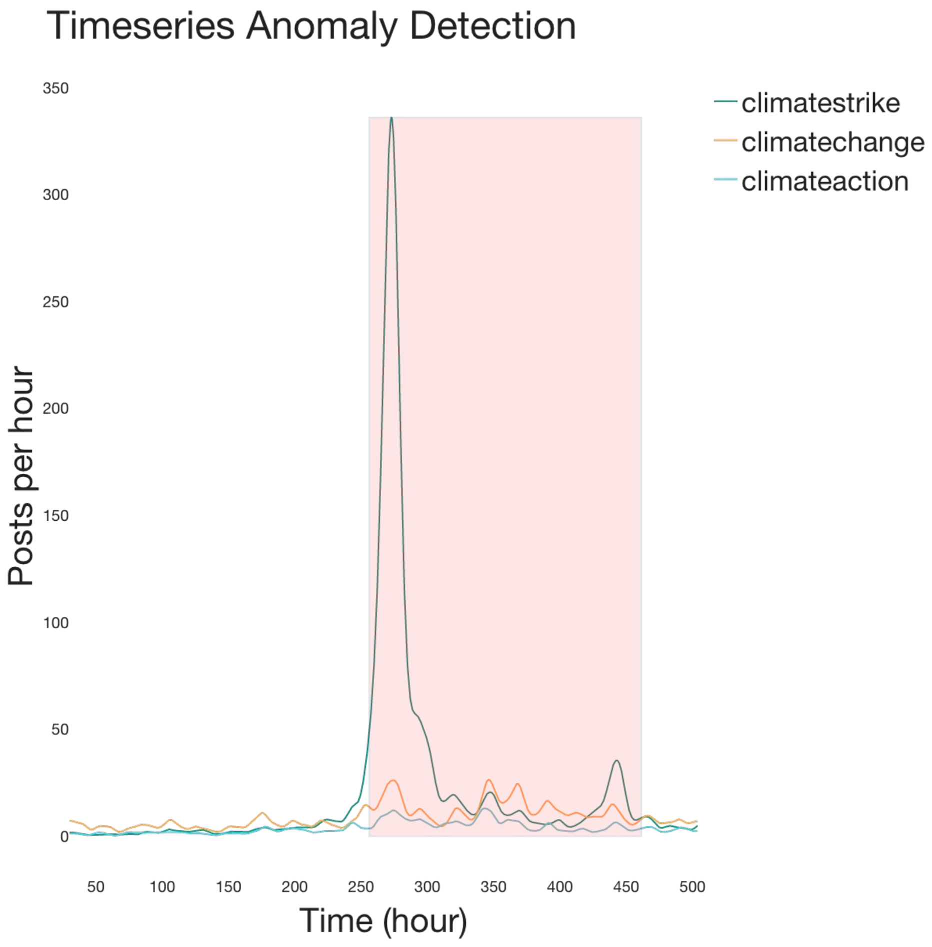
4. Present the collection of analyses to a **human analyst**.



Case Study: Climate Change

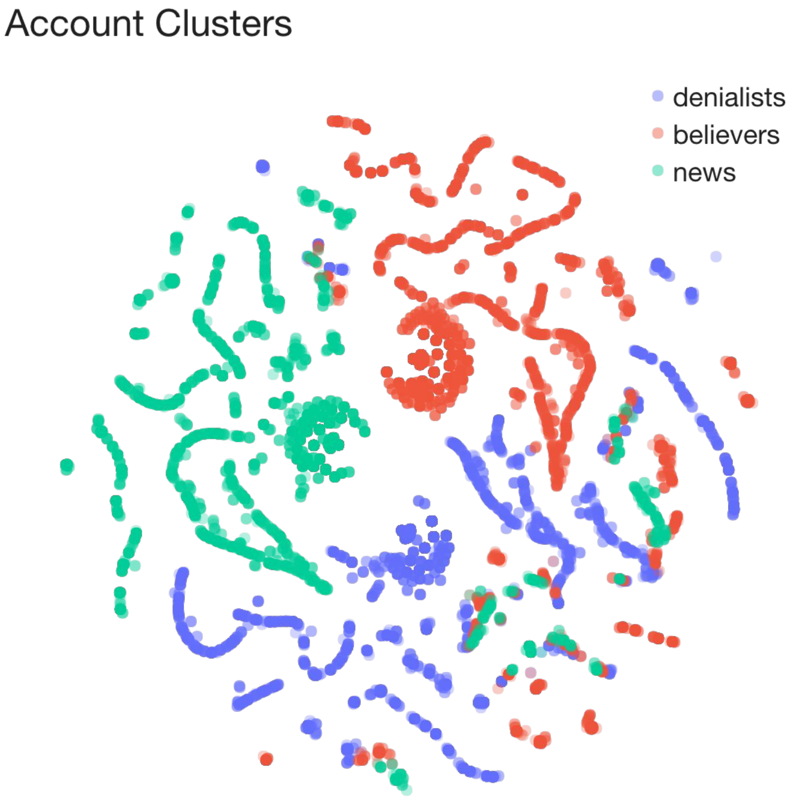
Detect anomalous content

Timeseries anomaly detection provides insight into **when** an influence campaign is most likely to be involved.



Identify groups of accounts

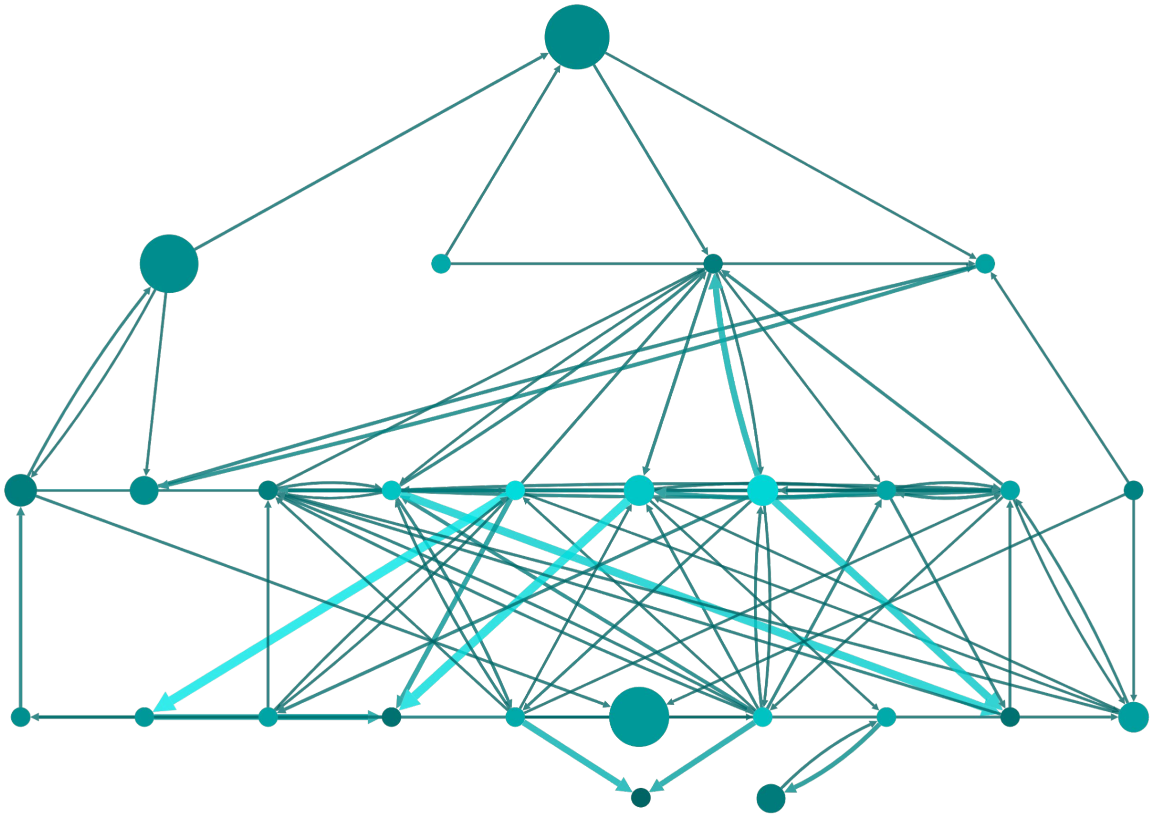
Clustering accounts according to their posting behavior provides indicators of **who** may be part of an organized, intentional faction.



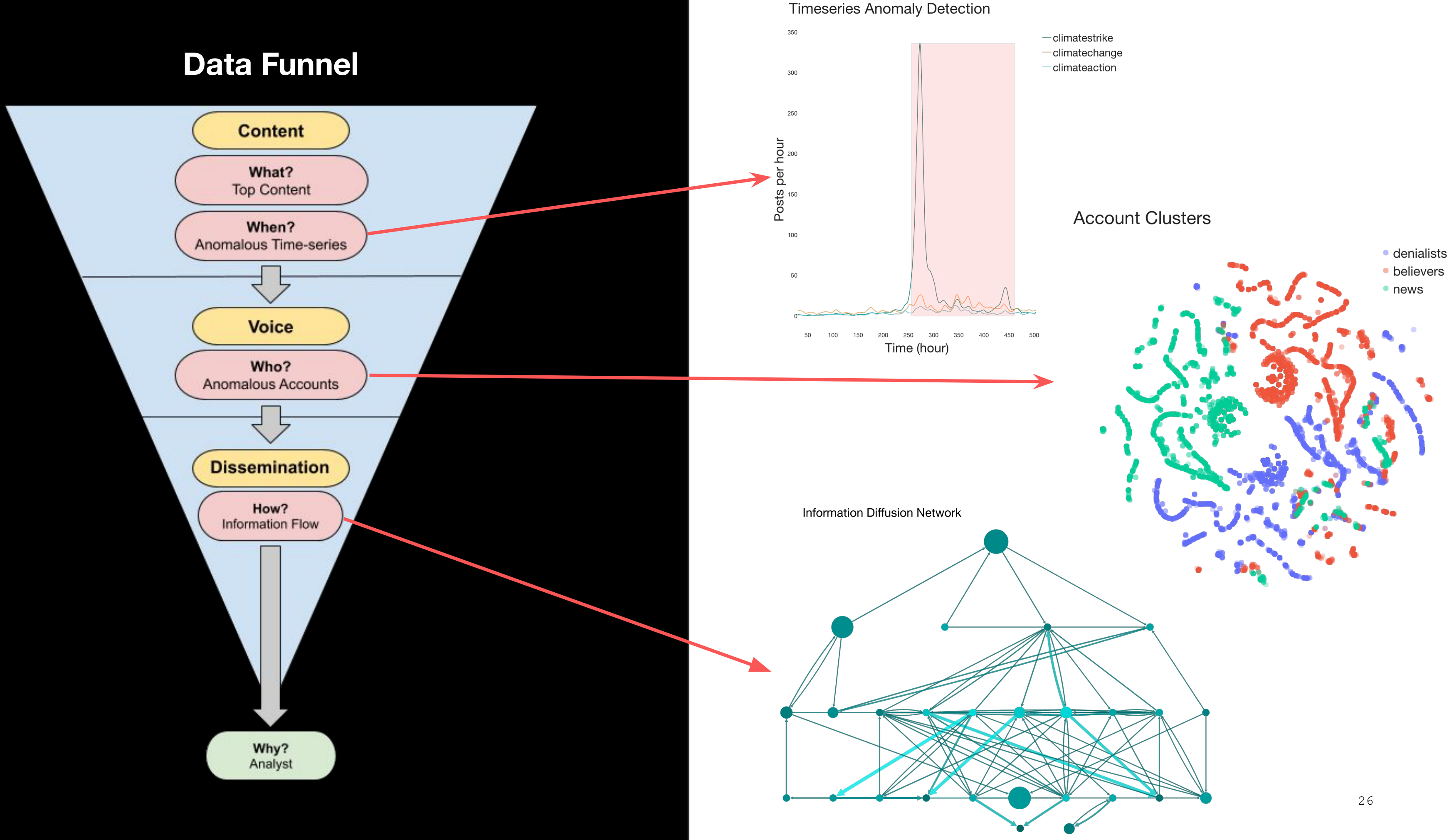
Examine dissemination

Inspecting the diffusion of content across a network of accounts provides insight into **how** a campaign is operating, its tactics, and the roles of the accounts involved.

Information Diffusion Network



Data Funnel



Timeseries Anomaly Detection

Account Clusters

Information Diffusion Network

New Knowledge

careers@newknowledge.com