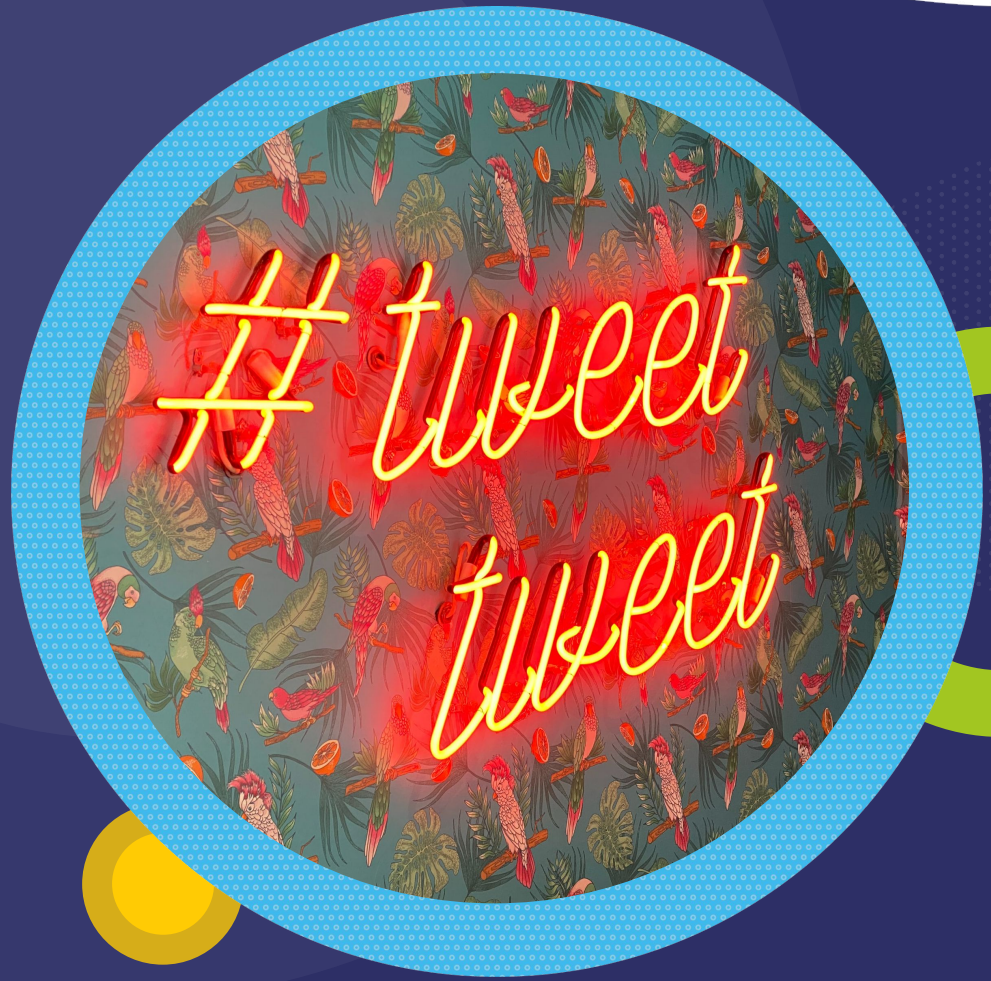
Altmetric

Using sentiment analysis to navigate the research-linked Twittersphere

Who is talking about your research and what are they saying?



Unsplash by Chris J. Davis

Mike Taylor

Data Insights & Customer Analytics, Digital Science

Carlos Areia

Data Scientist, Digital Science

What do we do?

Altmetric tracks online discussions about research and clinical trials.

Altmetric Data Insights harnesses data science to create new information.

Unsplash by christina-wocintechchat

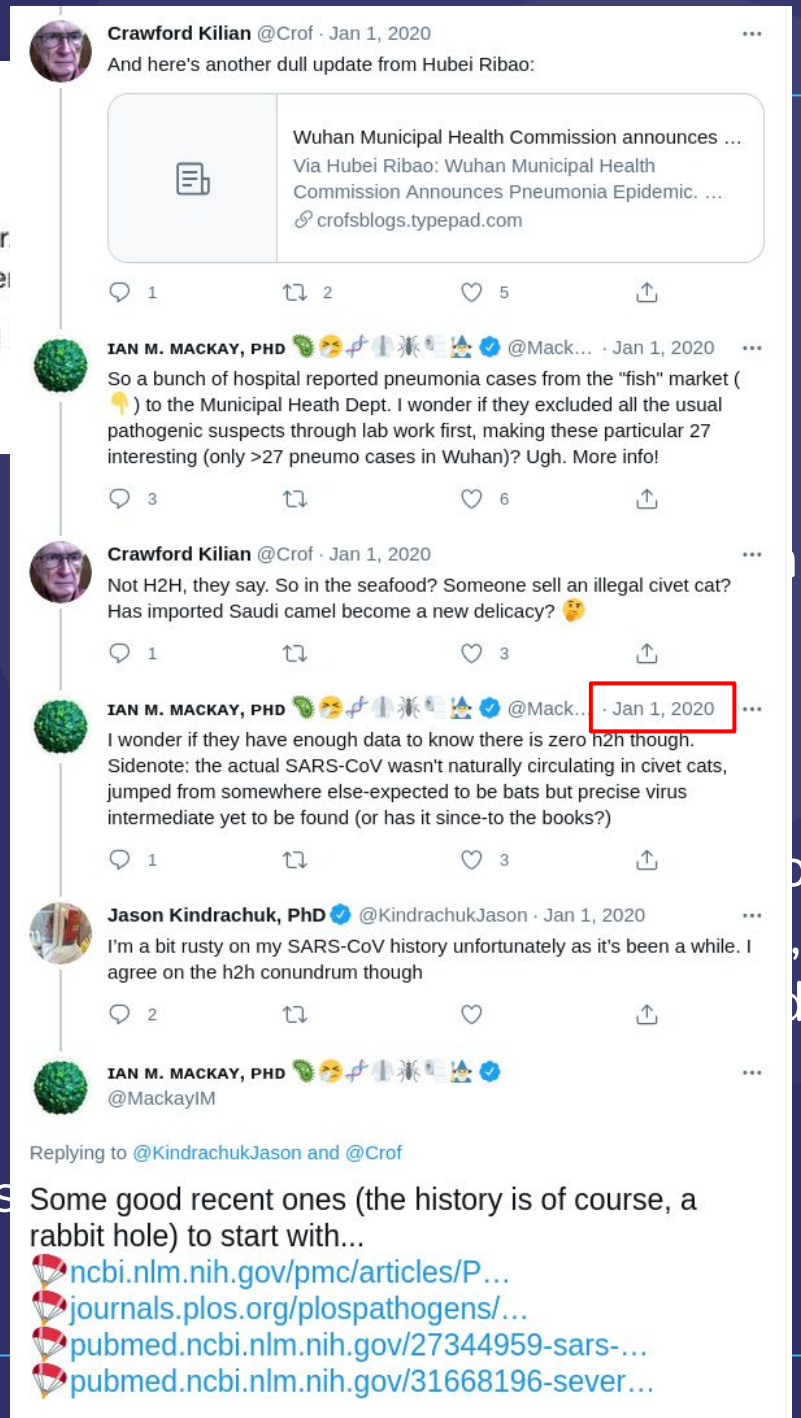
What are we going to talk about today?

How do we go beyond 'the numbers' and generate insights?

Twitter users

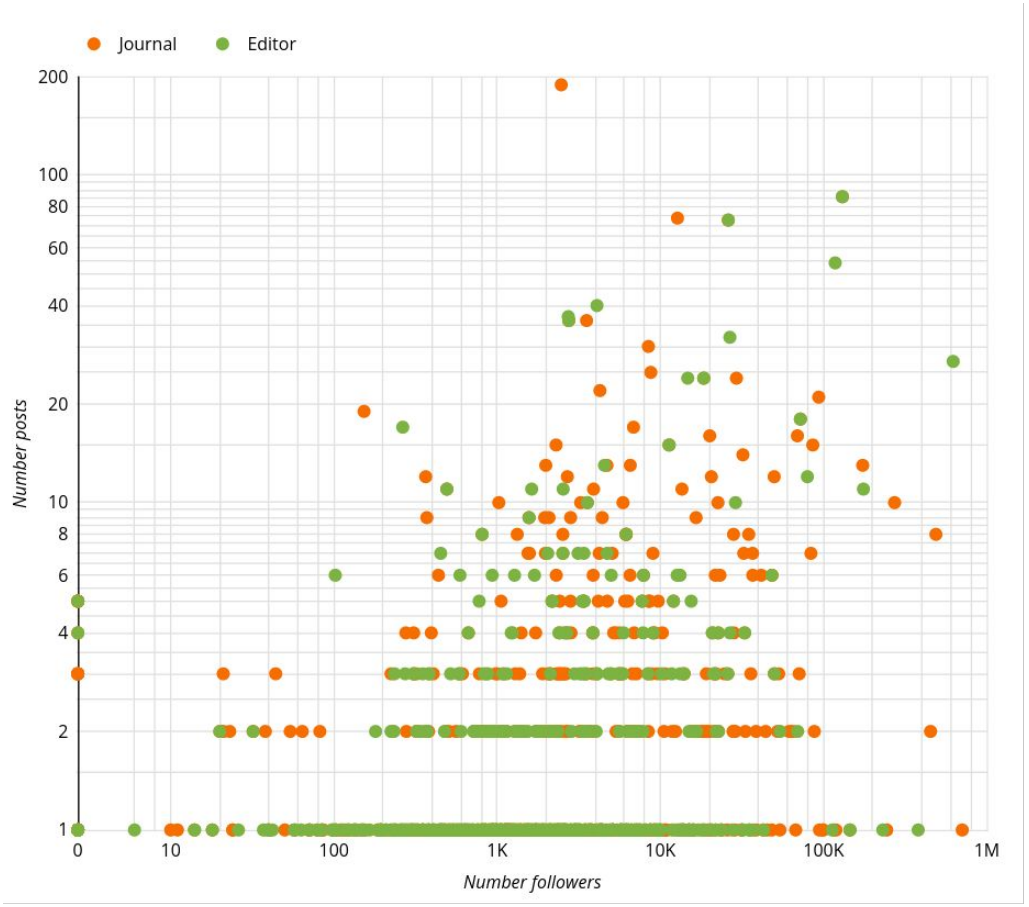
istory

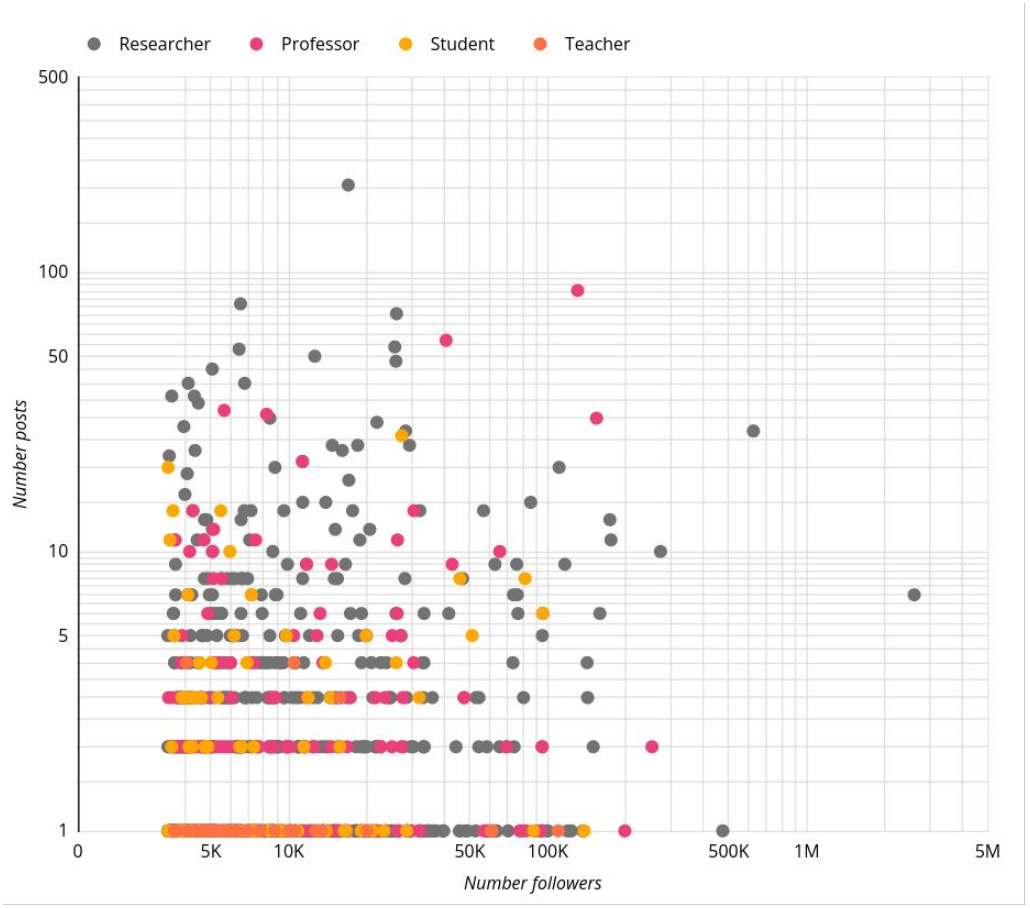
ork effect –
, opinions
der groups



Who is tweeting?

Machine learning approaches allow us to investigate which stakeholder groups are involved in communicating research



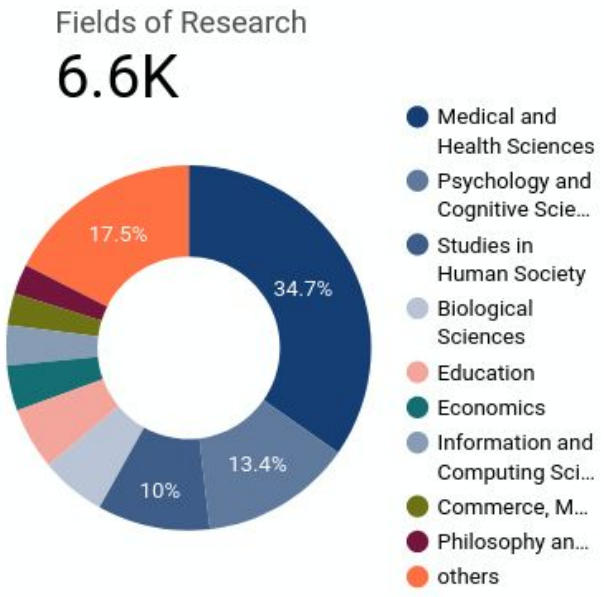


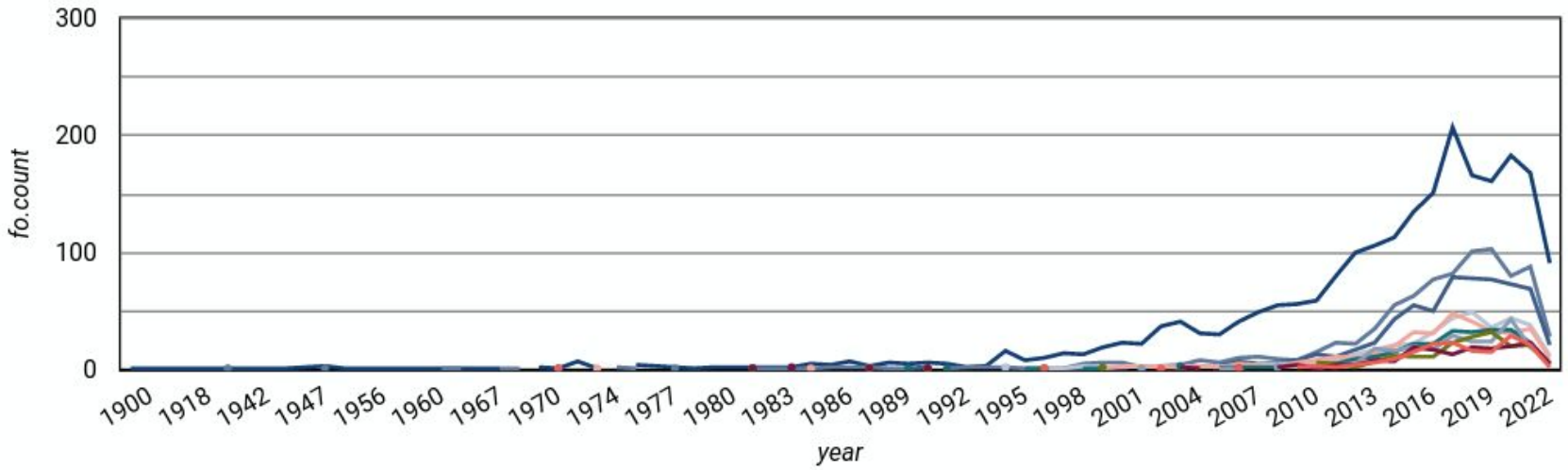
Understanding the stakeholder groups means that we can identify those people sharing your book content: researchers vs clinicians vs patients vs bloggers (etc, etc)

What are they tweeting about?

It’s possible to analyse individual Twitter accounts to see what research they’ve tweeted about in the past.

From sample (2.5k) of Twitter accounts of people identifying as patients, family and patient advocates in the area of mental health, it’s possible to identify which areas are discussed, and how that varies over time.



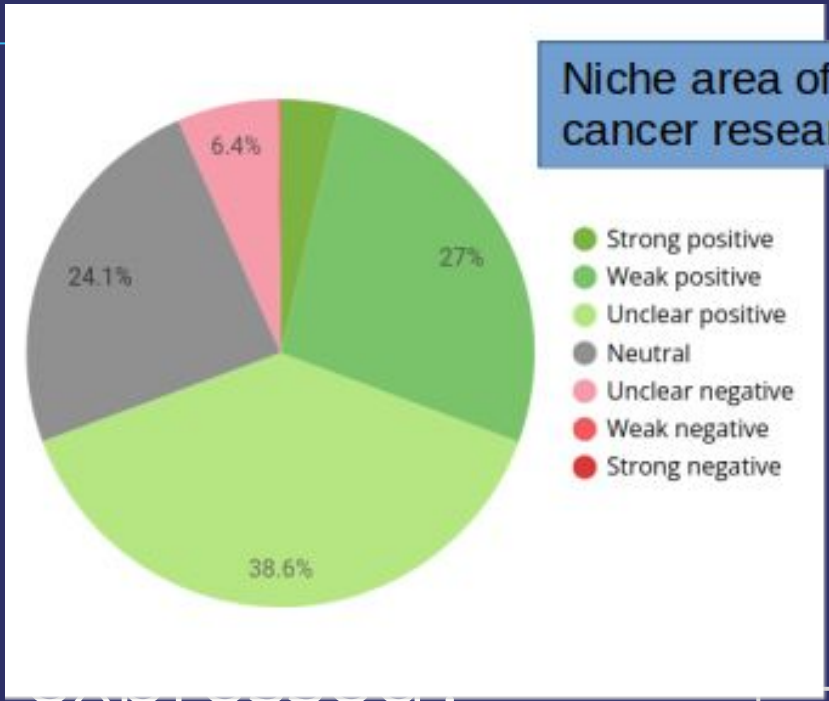


Who are these Twitter accounts reaching?

Different groups cause different forms of behaviour, for example, oncologists tweeting about oncology are over twice as likely to be retweeted as non-oncologist clinicians.

So … we know how they describe themselves, and the research they're sharing - we also know who they're reaching - their followers

And we can even pick up the retweets, and follow them down the line!



To understand the sentiment being expressed on Twitter, we developed a machine learning application trained only on Tweets linking to research and trials.

Previous research indicated that it was largely neutral – this has changed.

A brief exploration of the kinds of dashboard and insights we produce

Scores range from “Strong negative” (-3), to “Weak negative”, “Unclear negative” (typically satire, double negatives, “Neutral” (0) – normally just link/title, “Unclear Positive”, “Weak Positive”, “Strong Positive” (+3) – which contains a clear ‘go and read this’ signal.

The average scores tell us broad view, the distribution tells us about polarity and alerts us to potentially negative reception.

Publications and clinical trials mentions sentiment analysis

Methodology - Data journey

Dimensions

Publications of interest

Altmetric

Mentions information (who and what).

Google Big Query

Process profiles and tweets.

jupyter

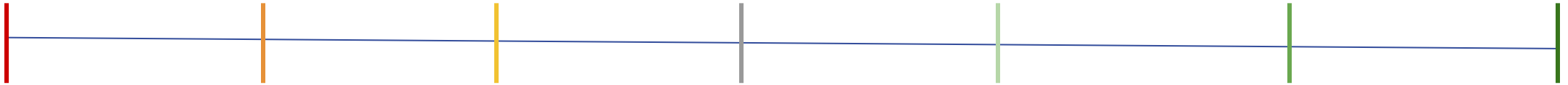
Identify stakeholders.

Apply sentiment analysis to content.

tableau

Build dashboard with data of interest (SoV)

Methodology - Sentiment Analysis labelling



-3 - Strongly negative. “This paper is completely biased”

-2 - Weak negative. “This is preprint so buyer beware but hopefully it holds up.”

-1 - Unclear negative. “Oh boy”

0 - Neutral. “https://t.co/u8hSn3x5Lu”

1 -Unclear positive. “COVID-19 diagnosis and management: a comprehensive review. https://t.co/n3WGYvwwHA”

2 -Weak positive. “New study from Brazil finds "regular use of ivermectin as a prophylactic agent was associated with significantly reduced COVID-19 infection, hospitalization, and mortality rates." https://t.co/vRjVHAb09s”

3 -Strong positive.
“Amazing paper”

Potential applications

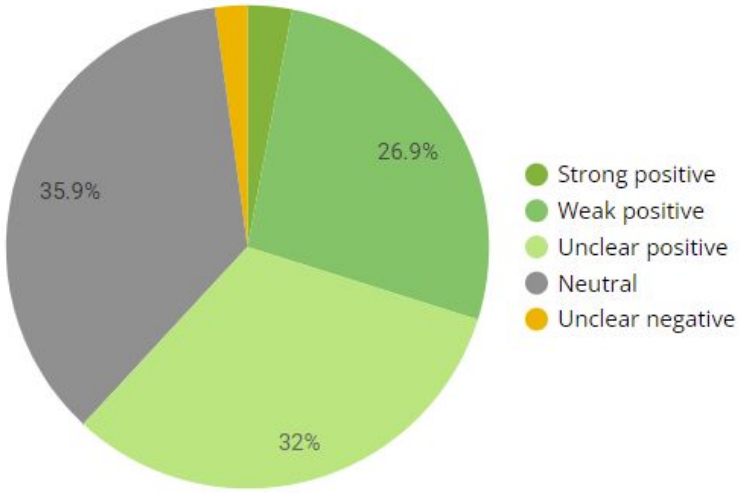
What

Who

When

Examples

1- Overall sentiment of all mentions of a particular drug publications for breast cancer.



Potential applications

What

Who

When

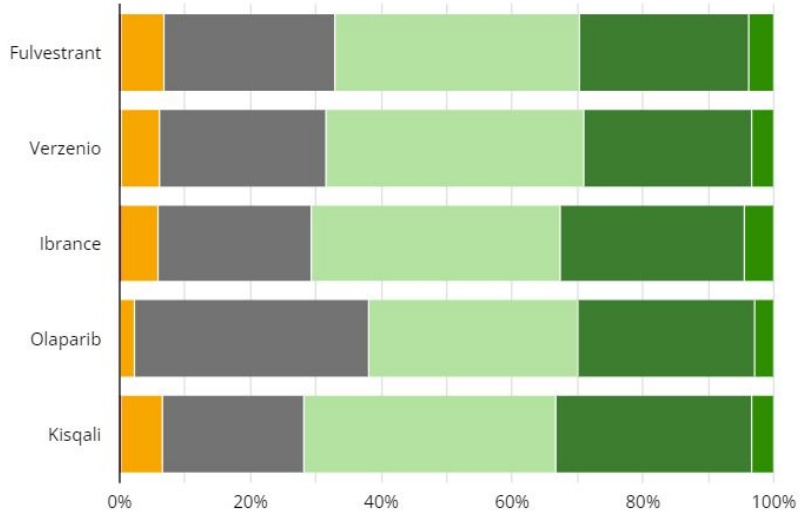
Examples

2-Average sentiment comparison between several breast cancer parp inhibitors.

3-Average sentiment comparison between breast cancer parp inhibitors of last month posts

4-Average sentiment comparison between breast cancer parp inhibitors of last month posts from oncologists

5-Average sentiment comparison between breast cancer parp inhibitors of **X** company of studies published in the past 2 years



Potential applications

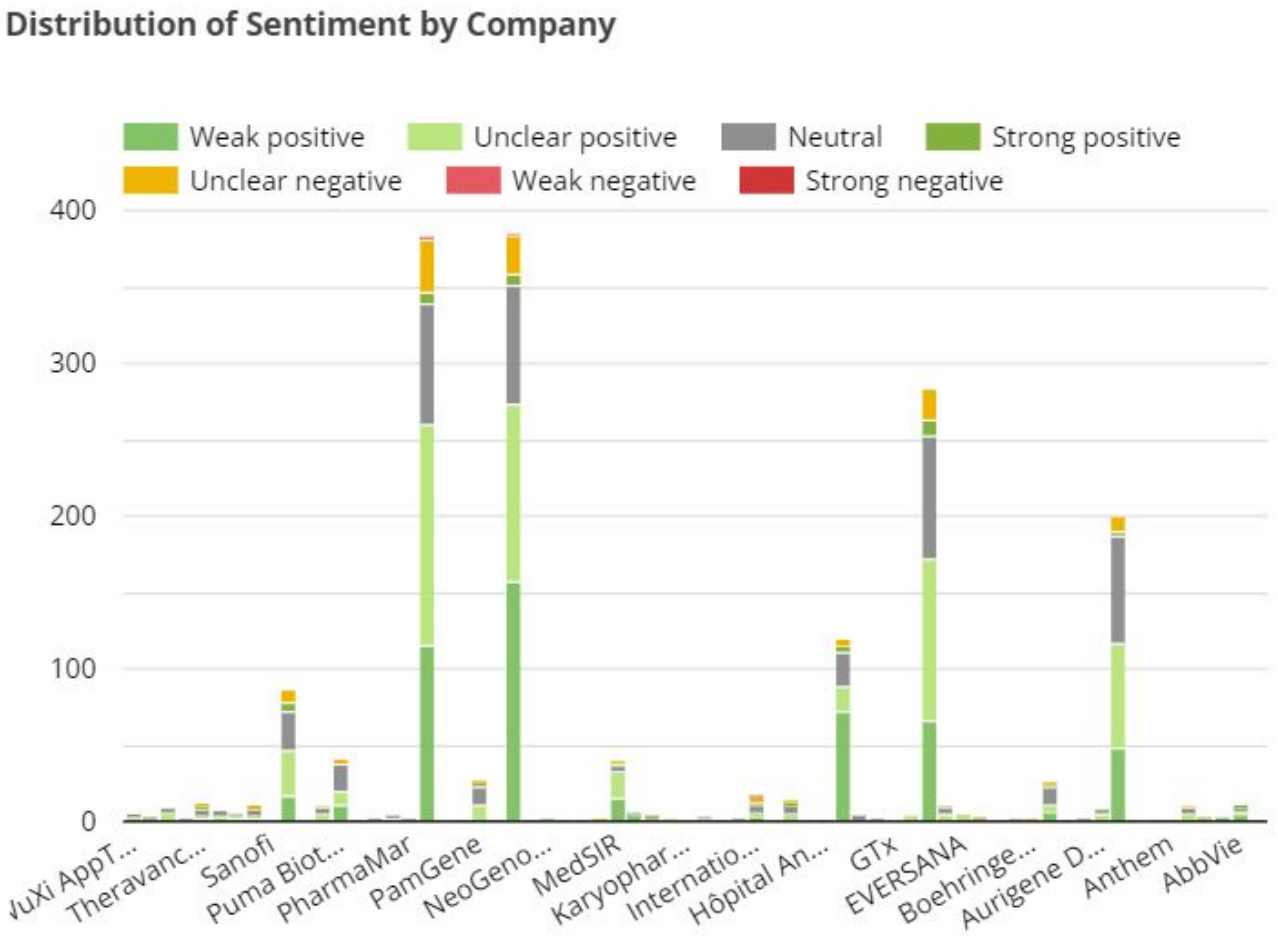
What

Who

When

Exam

6-Average sentiment comparison of breast cancer parp inhibitors by company of studies published in the past 10 years



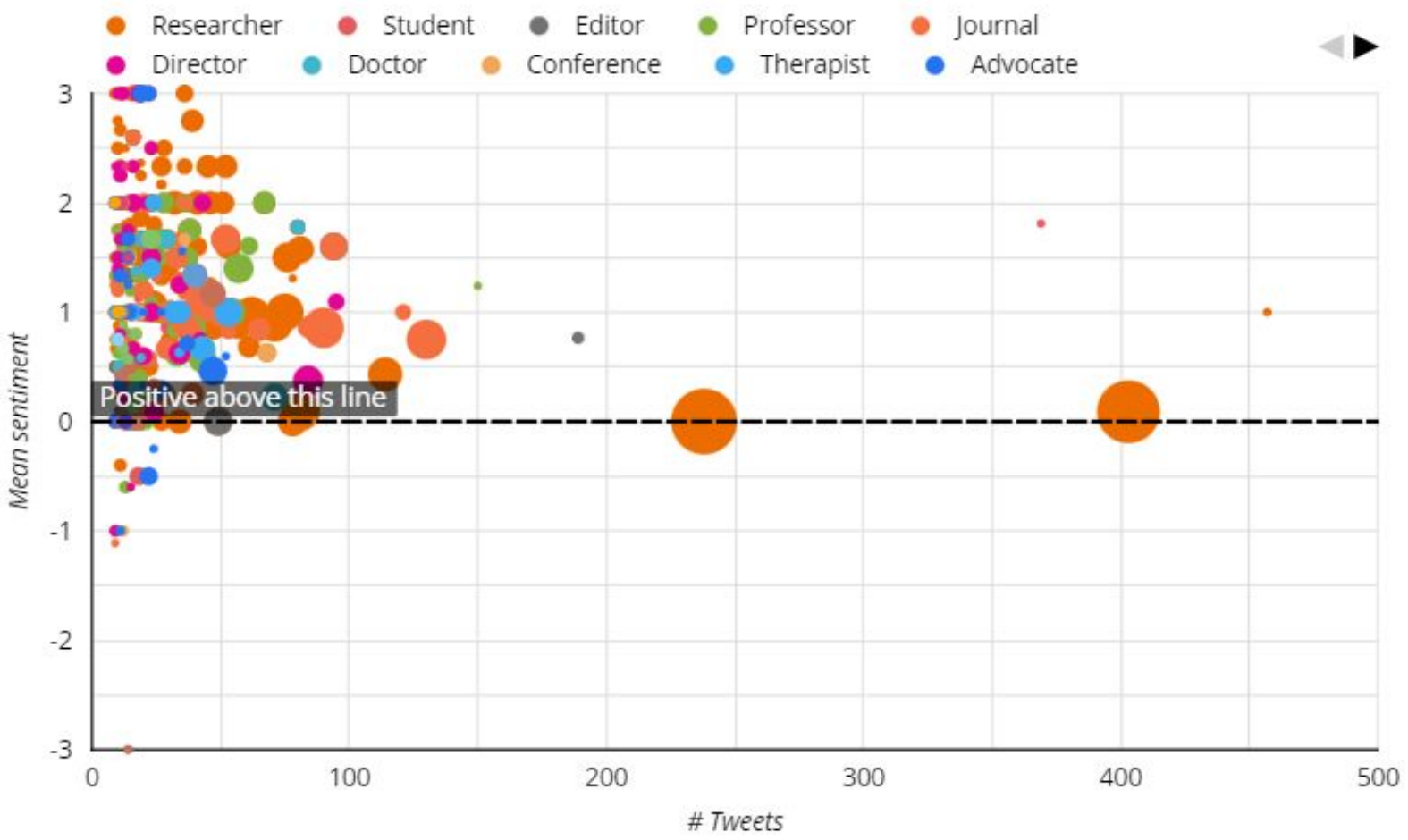
And much more
As broad or as specific as needed.

What | Who | When

7-Sentiment of each twitter profile sharing **X**

8-Sentiment of **X organisation publications**

9-Sentiment of each twitter profile sharing **X**



In summary...

Tracking the diffusion of research and trials across the internet is challenging - and requires much more advanced approaches than social media monitoring.

It is possible to utilize new technologies (cloud computing, pattern matching, phrase extraction, natural language processing, sentiment analysis) to go beyond the simple numbers and to produce real, actionable insights.

Digital Science combines these insights with share-of-voice analyses; publication strategy dashboards and KOL / DOL insights - automating reports that were previously time-consuming and expensive.

Thank you!

@altmetric
@herrison
altmetric.com