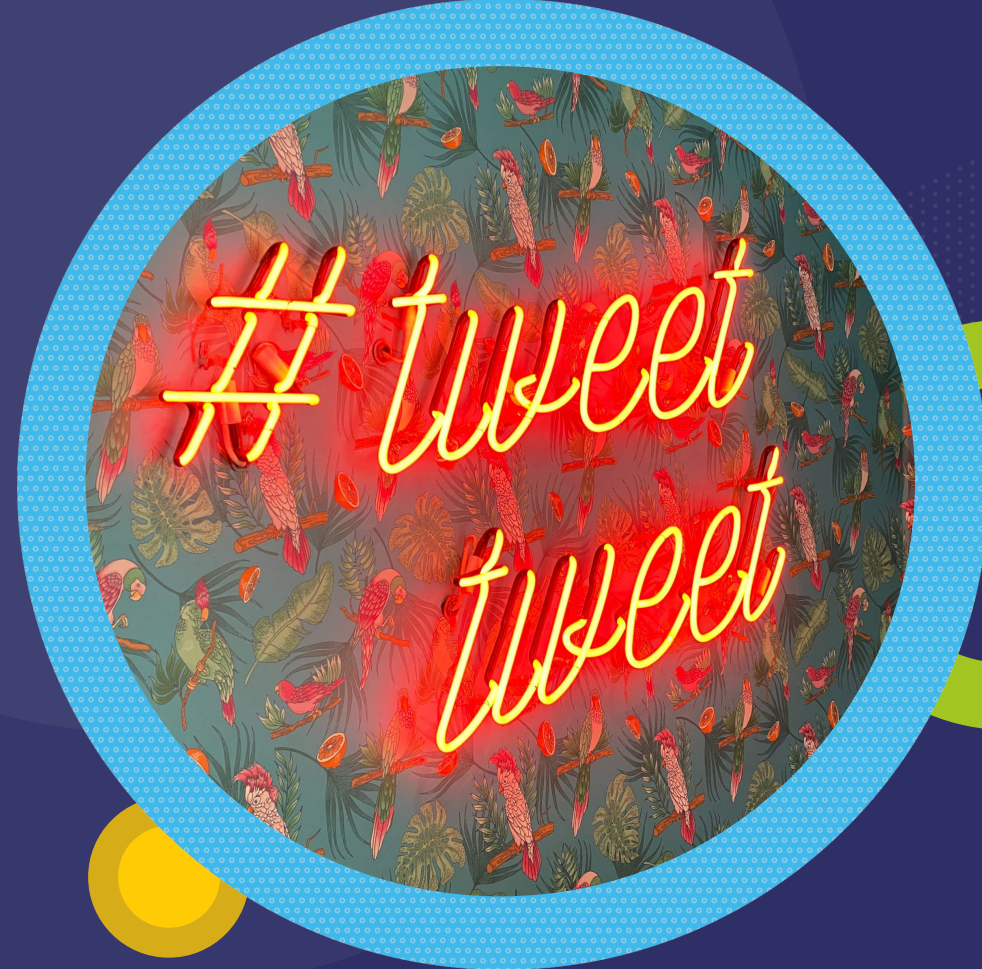


# Using sentiment analysis to navigate the research-linked Twittersphere

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





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.





What are we going to talk about today?





**IAN M. MACKAY, PHD**

@MackayIM

virologist. husband. dad. reader. writer. fixer. bad typist. learner  
referenced. opinions mine alone. not medical advice or endorsement

📍 Brisbane, Australia 🌐 [virologydownunder.com](http://virologydownunder.com) 📅 Joined

4,860 Following 125.3K Followers

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

Twitter users

**Crawford Kilian** @Crof · Jan 1, 2020  
And here's another dull update from Hubei Ribao:

Wuhan Municipal Health Commission announces ...  
Via Hubei Ribao: Wuhan Municipal Health  
Commission Announces Pneumonia Epidemic. ...  
[crofsblogs.typepad.com](http://crofsblogs.typepad.com)

1 2 5

**IAN M. MACKAY, PHD** 🦠 🤔 🧐 🧪 🧬 🧫 🧭 🧭 🧭 @Mack... · Jan 1, 2020  
So a bunch of hospital reported pneumonia cases from the "fish" market ( ) to the Municipal Health Dept. I wonder if they excluded all the usual pathogenic suspects through lab work first, making these particular 27 interesting (only >27 pneumo cases in Wuhan)? Ugh. More info!

3 6

**Crawford Kilian** @Crof · Jan 1, 2020  
Not H2H, they say. So in the seafood? Someone sell an illegal civet cat? Has imported Saudi camel become a new delicacy? 🤔

1 3

**IAN M. MACKAY, PHD** 🦠 🤔 🧐 🧪 🧬 🧫 🧭 🧭 🧭 @Mack... · Jan 1, 2020  
I wonder if they have enough data to know there is zero h2h though. Sidenote: the actual SARS-CoV wasn't naturally circulating in civet cats, jumped from somewhere else-expected to be bats but precise virus intermediate yet to be found (or has it since-to the books?)

1 3

**Jason Kindrachuk, PhD** @KindrachukJason · Jan 1, 2020  
I'm a bit rusty on my SARS-CoV history unfortunately as it's been a while. I agree on the h2h conundrum though

2

**IAN M. MACKAY, PHD** 🦠 🤔 🧐 🧪 🧬 🧫 🧭 🧭 🧭 @MackayIM  
Replying to @KindrachukJason and @Crof

Some good recent ones (the history is of course, a rabbit hole) to start with...

- [ncbi.nlm.nih.gov/pmc/articles/P...](http://ncbi.nlm.nih.gov/pmc/articles/P...)
- [journals.plos.org/plospathogens/...](http://journals.plos.org/plospathogens/...)
- [pubmed.ncbi.nlm.nih.gov/27344959-sars-...](http://pubmed.ncbi.nlm.nih.gov/27344959-sars-...)
- [pubmed.ncbi.nlm.nih.gov/31668196-sever...](http://pubmed.ncbi.nlm.nih.gov/31668196-sever...)

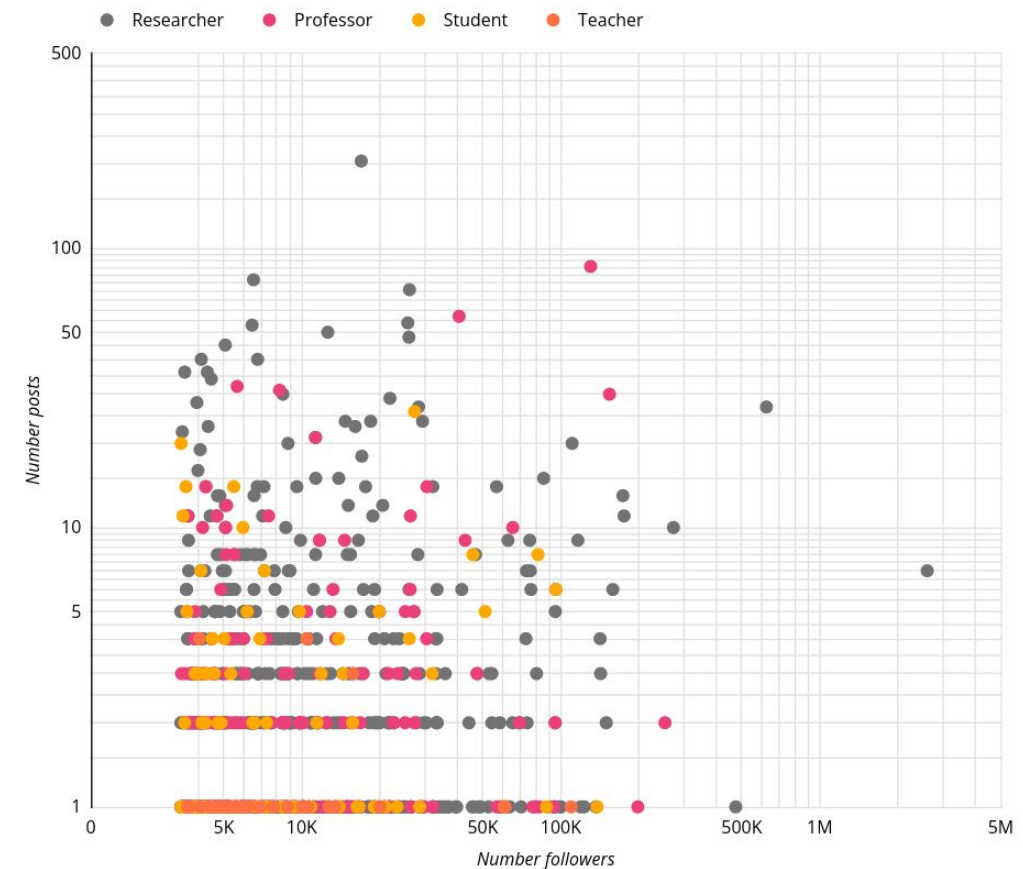
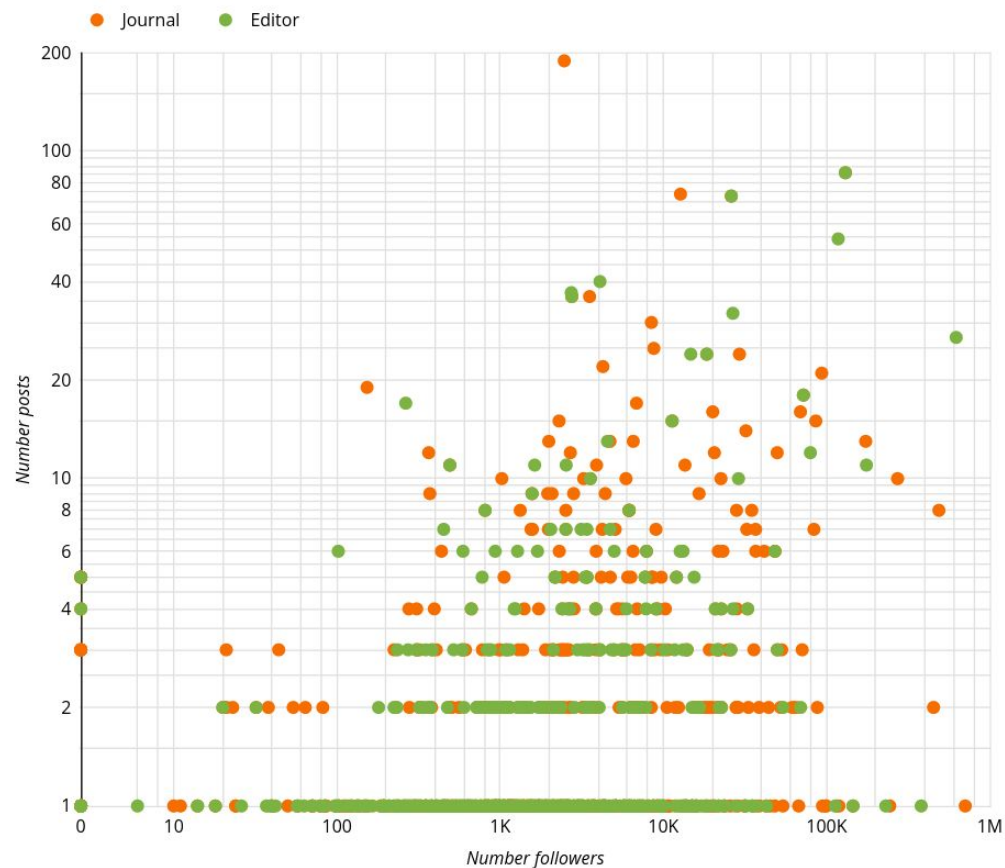
history

ork effect –  
, opinions  
der groups



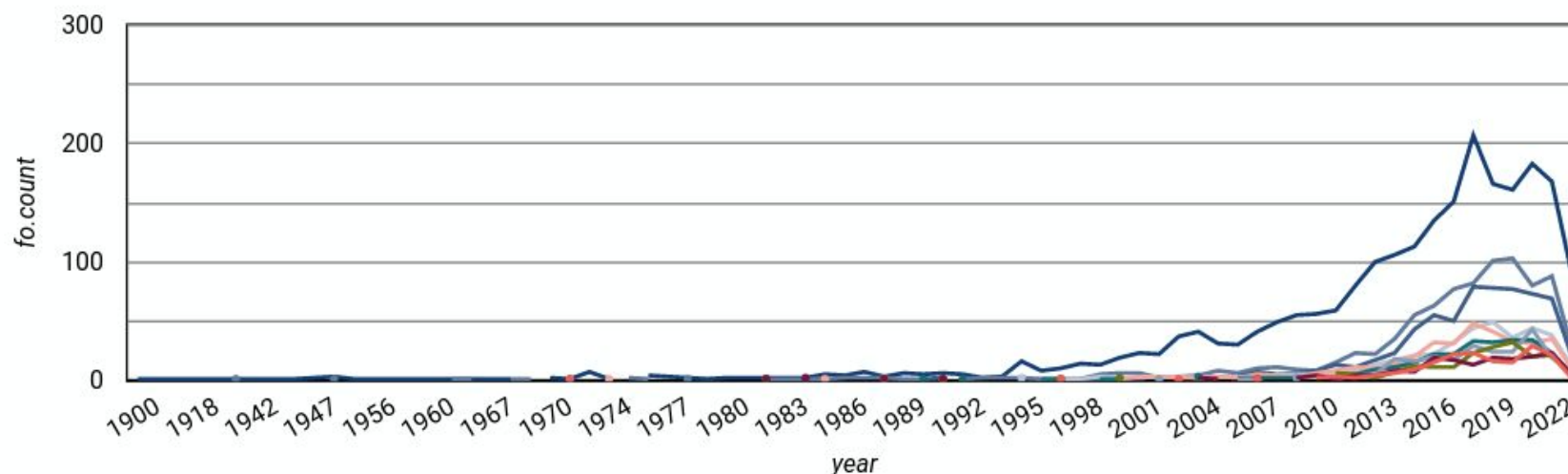
## Who is tweeting?

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)









## Who are these Twitter accounts reaching?

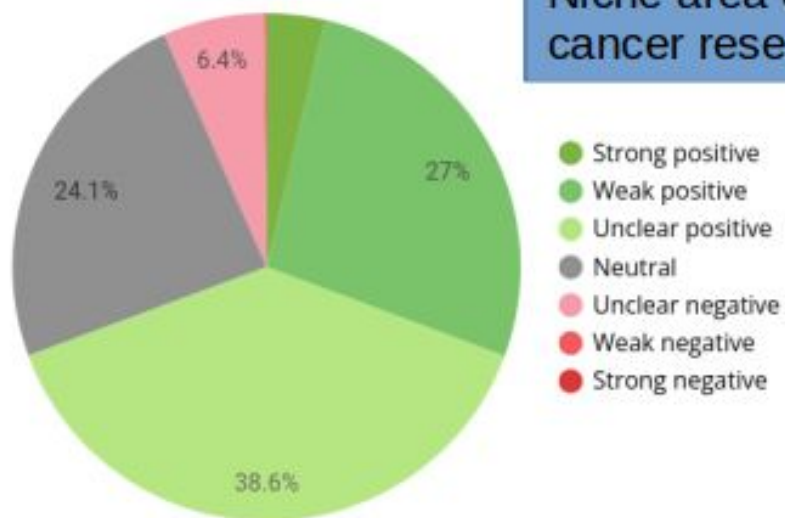
Different groups cause different forms of behaviour, for example, oncologists tweeting about oncology are over twice as likely to be retweeted as non-oncologist clinicians.

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!



### Niche area of breast cancer research



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



Publications of  
interest

Mentions information  
(who and what).

Process profiles and  
tweets.

Identify stakeholders.

Build dashboard with  
data of interest (SoV)

Apply sentiment  
analysis to content.





## 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/n3WGYvwvHA>”

**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

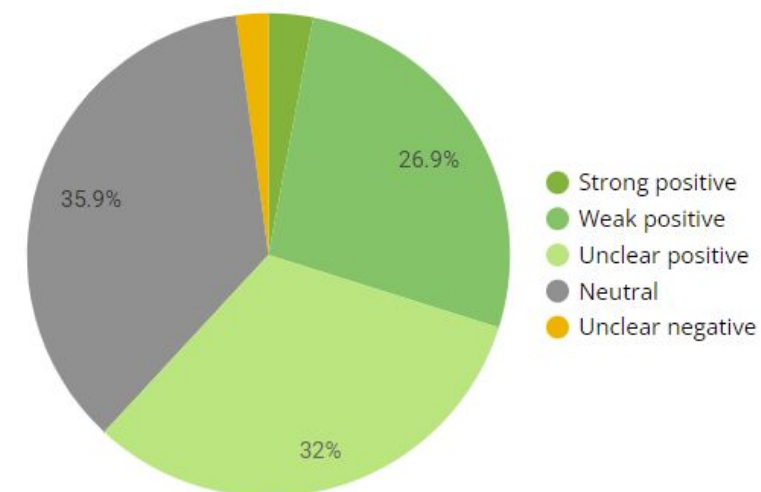
What

Who

When

### Examples

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





# Potential applications

What

Who

When

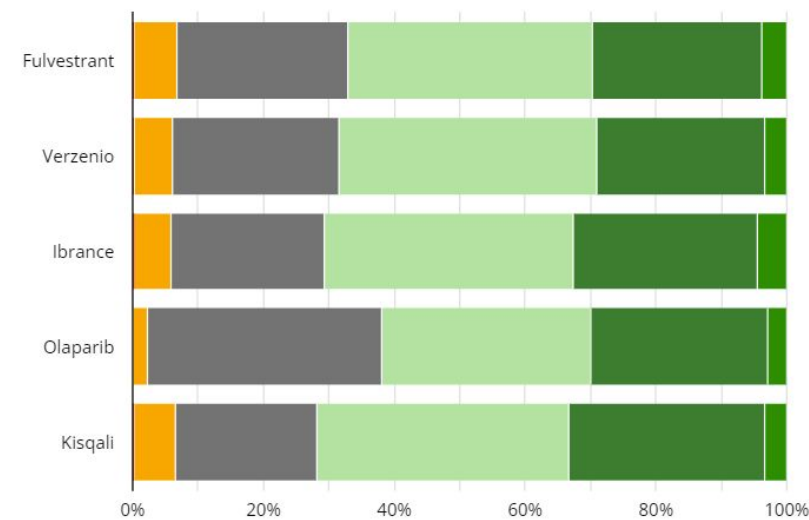
## Examples

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

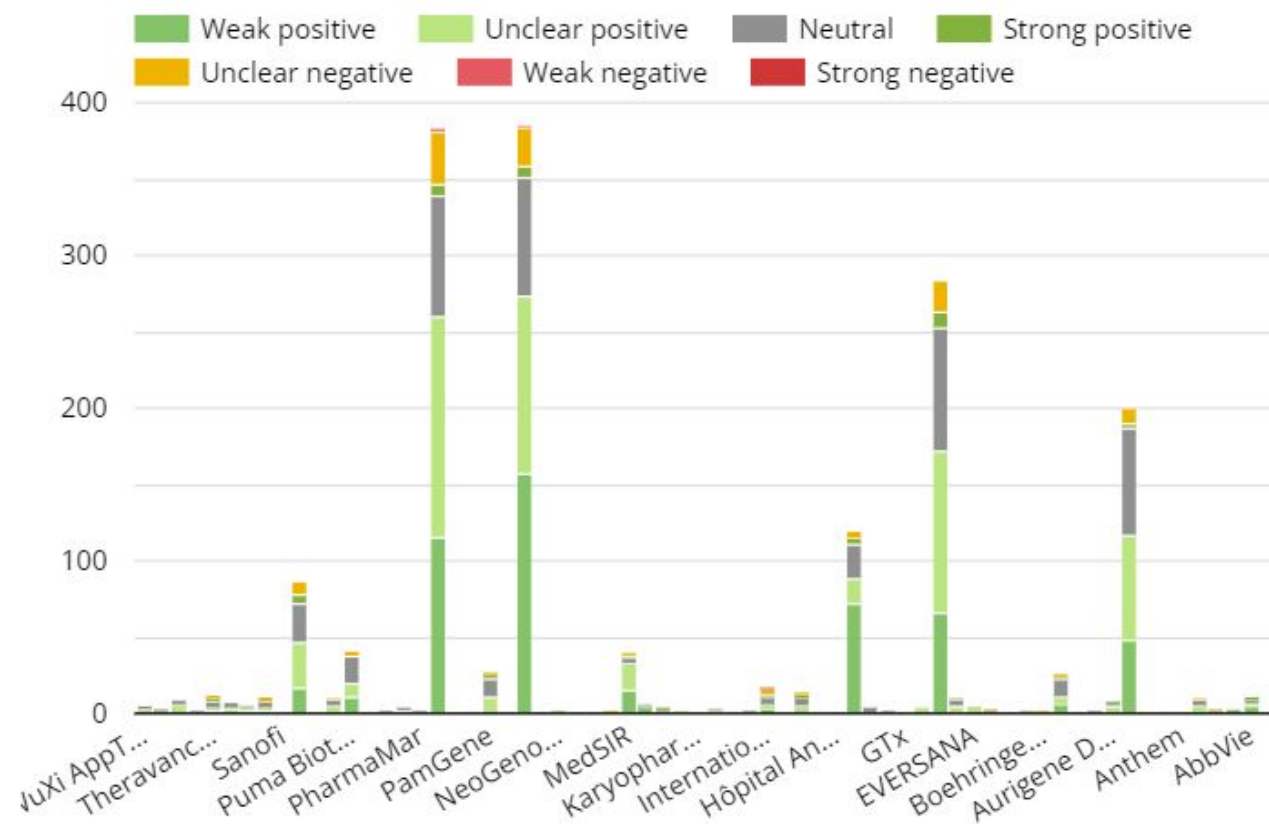
Who

When

Exam

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

Distribution of Sentiment by Company







And much more

As broad or as specific as needed.

What

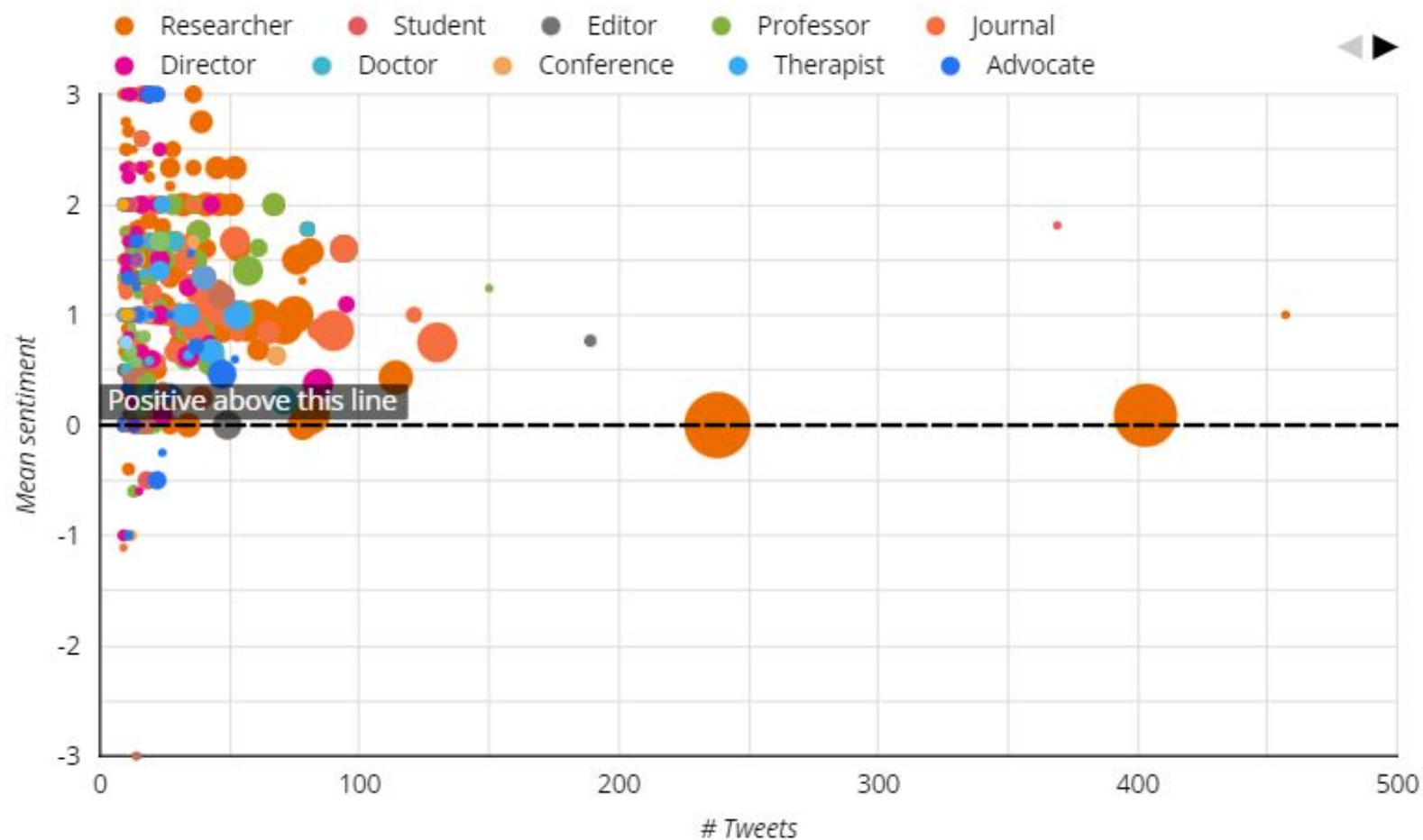
Who

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

