

Title: TBA

Date: Mar 28, 2018 09:45 AM

URL: <http://pirsa.org/18030109>

Abstract:

# Do-It-Yourself Measures for Academic Success

Tobias Mistele, Tom Price, Sabine Hossenfelder

Frankfurt Institute for Advanced Studies (FIAS)

Open Research: Rethinking Scientific Collaboration  
March 28, 2018



**FIAS** Frankfurt Institute  
for Advanced Studies



**FQXi**

FOUNDATIONAL QUESTIONS INSTITUTE

## Measures - what they should be (but usually are not)

→ Measures should work *for* scientists

- ▶ Heterogeneous
- ▶ Transparent, customizable
- ▶ Bottom-up
- ▶ Revised over time

## Measures - get rid of them?

- ▶ Carefully read up on hundreds of applicants?
- ▶ Objectively?

## Measures - have to improve them!

- ▶ Make them useful to the community
- ▶ How?

## Work so far - Keyword analysis and clouds

<http://www.scimeter.org/>

## Work so far - Keyword analysis and clouds

[Keywords](#) [Clouds](#) [Similar authors](#) [About](#)

### Create your own keyword cloud

Create your own keyword cloud:

X

Add author

Go

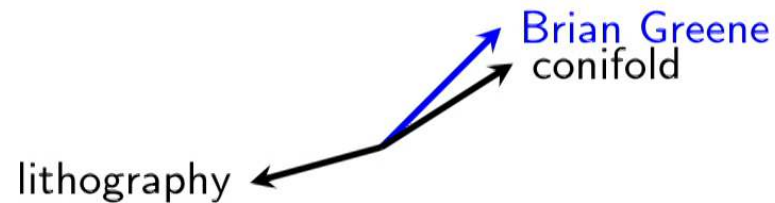


## Work so far - Keyword analysis and clouds

- ▶ Data: from arXiv
- ▶ Identify keywords in titles and abstracts
- ▶ Throw away out “generic” words  
Keep: “Calabi-Yau manifolds”, “conifold”, ...  
Don't keep: “This is”, “Therefore”, ...
- ▶ Generic = little mutual information with arXiv category

## Work so far - Keyword analysis and clouds

- ▶ Vectorization of keywords as well as authors (in the same 200-dim. vector space)
- ▶ Made-up example:



- ▶ Can identify keywords that describe an author well but were never “literally” used by that author

# Work so far - Keyword analysis and clouds

[Keywords](#) [Clouds](#) [Similar authors](#) [About](#)

1. [Vishnu Jejjala](#)
2. [John Stout](#)
3. [Antonio Segui](#)
4. [Wei Gu](#)
5. [Pontus Ahlqvist](#)
6. [Lars Nilse](#)
7. [Frank Saueressig](#)
8. [Nana Cabo Bizet](#)
9. [William Cottrell](#)
10. [Anindya Dey](#)
11. [Chien-Hsun Wang](#)
12. [Aradhya Shukla](#)
13. [Stephen Shenker](#)
14. [Albion Lawrence](#)
15. [Pedro Resco](#)
16. [Morten Ernebjerg](#)
17. [Miguel Montero](#)
18. [David Vegh](#)
19. [Cody Long](#)
20. [Tibra Ali](#)
21. [David Kagan](#)
22. [Ayush Saurabh](#)
23. [Prabwal Phukon](#)
24. [Michal Fabinger](#)
25. [Armen Yeranyan](#)
26. [Sunggeun Lee](#)

## Measures - give them to the community!

### Do-It-Yourself Measures:

- ▶ Users can assemble their own measures weighting different factors (#papers, #citations, #single-authored papers, *h*-index, pagerank, JIF, #coauthors, ...)
- ▶ Apply to list of authors
- ▶ Measures can be shared, combined, blended

## Measures - give them to the community!

- ▶ Addresses traditional issues of measures
  - ▶ Heterogeneous
  - ▶ Transparent, Customizable
  - ▶ Bottom-up
  - ▶ Revised over time

## Measures - give them to the community!

- ▶ Aggregation of which measures other users perceive as useful
- ▶ Include new factors
- ▶ E.g.: How well does a researcher's keywords match those of your institute?

## Measures - Disclaimer

This is *not* about

- ▶ finding the one best measure
- ▶ and then tell everybody to use that

But

- ▶ Creating a community-driven platform
- ▶ to create a *variety* of useful measures
- ▶ and changing the way measures are used

## Work in progress - Neural net predictor as a measure?

- ▶ Another possible new factor to be used in DIY measures
- ▶ Neural net trained to predict e.g. future  $h$ -index or #citations
- ▶ Experiment with input from keyword/author vectorization

## Work in progress - Neural net predictor as a measure?

- ▶ Another possible new factor to be used in DIY measures
- ▶ Neural net trained to predict e.g. future  $h$ -index or #citations
- ▶ Experiment with input from keyword/author vectorization

## Summary

- ▶ Measures have issues, but are necessary
  - ▶ DIY measures may help the community
  - ▶ Currently available: Tools for creating keyword clouds of research interests + finding similar authors
- Measures should work *for* scientists

## Measures - give them to the community!

- ▶ Addresses traditional issues of measures
  - ▶ Heterogeneous
  - ▶ Transparent, Customizable
  - ▶ Bottom-up
  - ▶ Revised over time

## Work in progress - Neural net predictor as a measure?

- ▶ Another possible new factor to be used in DIY measures
- ▶ Neural net trained to predict e.g. future  $h$ -index or #citations
- ▶ Experiment with input from keyword/author vectorization

## Summary

- ▶ Measures have issues, but are necessary
  - ▶ DIY measures may help the community
  - ▶ Currently available: Tools for creating keyword clouds of research interests + finding similar authors
- Measures should work *for* scientists