

Discovering information

Chris Hartgerink ? CC 0

August 8th, 2017

Key points

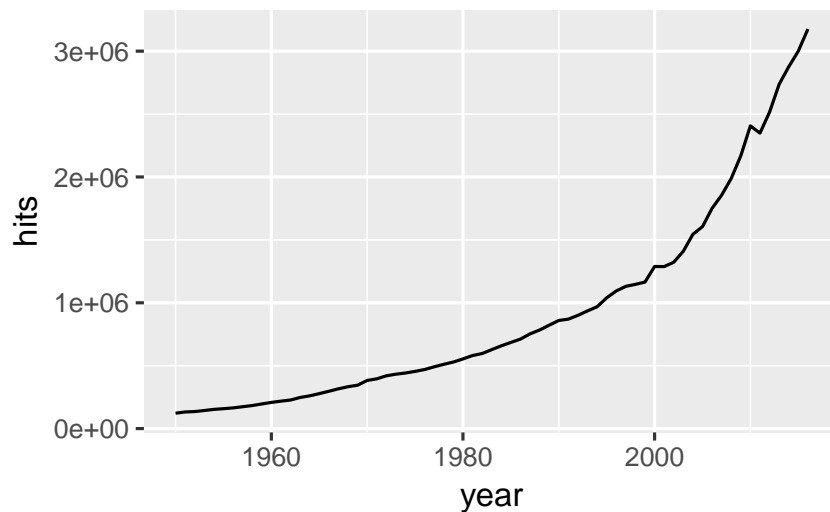
1. Learn about the increased importance of innovation in information discovery
2. Learn about barriers to innovation
3. Learn about an innovative information discovery tool

But much more information available than to cover in 20 minutes!

Importance of discovery

1. Discovery is proportional to production of information
 2. Digital age allows for many ways to discover/consume information
- No longer limited by physical restrictions of paper :-)

Discovery and production



Importance of discovery

1. Production exploded!
 2. Has discovery changed much in last decades?
- digitization of databases
 - speed of discovery has changed
 - Has mode of discovery changed?

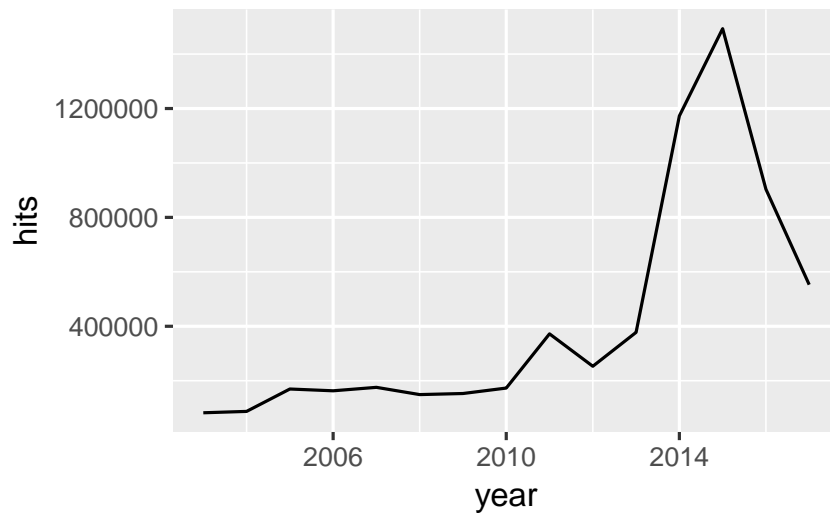
Production changes

1. Previously, discovery primarily related to scholarly reports
2. Now, more and more data discovery!
 - Due to easier options to share
 - Due to increased data production
 - Due to efficiency in research process

DataCite

1. DataCite is the CrossRef of data
2. Aggregates metadata about data sets from ICPSR, DANS, and other trusted repositories
 - See also re3data.org for finding trusted repositories to post data to

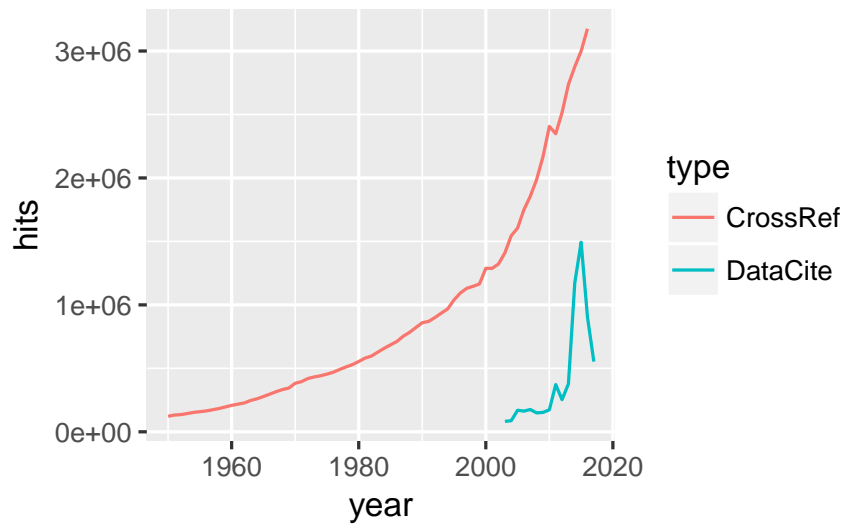
Data production



Data discovery

1. Data can be discovered reactively when reading a paper
2. Or by central searches for relevant data
3. What is more common?
4. How much data that is in papers can be found with central searches?

Combining data and reports




Barriers to improved discovery

1. Innovation requires diversity
 - Diversity of innovators
 - Diversity of content
 - Diversity of business models
 - Diversity of ...?
2. As such, for maximum innovation in information discovery, content needs to be available
3. For metadata, already possible with CrossRef, DataCite.
4. For more in-depth discovery (beyond abstract), copyrighted material limits innovation


Innovations in information discovery

1. Open Access content gives us a glimpse into the potential for innovation
2. Because content is still limited, innovation is only in early stages
 - Predict that we'll see more innovation as more content becomes reusable also for commercial purposes
3. Some examples already present

Example: OpenKnowledge Maps



SearchAboutTeamNewsGet in touchNewsletter



BETA

VISUALIZE A RESEARCH TOPIC

Choose a library:


☐ PubMed (biomedicine) ⓘ

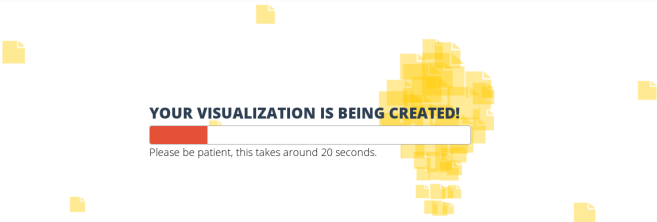
☒ BASE (all disciplines) ⓘ

GO

Options

Example: OpenKnowledge Maps






YOUR VISUALIZATION IS BEING CREATED!


Please be patient, this takes around 20 seconds.

Example: OpenKnowledge Maps


HOW IT WORKS

1

An Open Knowledge Maps visualization presents you with a topical overview for your search term. It is based on the 100 most relevant documents for your search term.

2

We use text similarity to create the knowledge maps. The algorithm groups those papers together that have many words in common.


3

The visualization is intended to give you a head start on your literature search. You can also use Open Knowledge Maps to stay up-to-date - just limit your search to the most recent papers in the options.

Example: OpenKnowledge Maps



Example: OpenKnowledge Maps



Scholar

About 6,340,000 results (0.06 sec)

Articles

Case law

My library

Any time

Since 2017

Since 2016

Since 2013

Custom range...

Sort by relevance

Sort by date

☒ Include patents

☒ Include citations

☒ Make alert

An open, large-scale, collaborative effort to estimate the reproducibility of psychological science

Open Science Collaboration - ... on **Psychological Science**, 2012 - journals.sagepub.com

Reproducibility is a defining feature of **science**. However, because of strong incentives for innovation and weak incentives for confirmation, direct replication is rarely practiced or published. The Reproducibility Project is an **open**, large-scale, collaborative effort to

Cited by 255 Related articles All 16 versions Cite Save

[\[HTML\] sagepub.com](#)

Estimating the reproducibility of psychological science

Open Science Collaboration - **Science**, 2015 - science.sciencemag.org

RESULTS We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and optimal materials when available. There is no single standard for evaluating replication success. Here, we

Cited by 1154 Related articles All 48 versions Cite Save

[\[PDF\] osf.io](#)

The open science grid

R Porces, D Petrucci, B Kramer, D Olson. - Journal of Physics: ... 2007 - iopscience.iop.org

Abstract: The **Open Science Grid** (OSG) provides a distributed facility where the Consortium members provide guaranteed and opportunistic access to shared computing and storage resources. OSG provides support for and evolution of the infrastructure through activities that

Cited by 270 Related articles All 12 versions Cite Save

[\[PDF\] inspirehep.net](#)

Common agency contracting and the emergence of "open science" institutions

PA David - The American Economic Review, 1998 - JSTOR

The Cold War's ending has brought mounting pressures to recognize national science and technology research systems. Yet, by comparison with what has been learned already concerning institutional arrangements and business strategies affecting corporate R&D

Cited by 233 Related articles All 8 versions Cite Save

Toward a new economics of science

D Partha, PA David - Research policy, 1994 - Elsevier

... and Nelson in examining the implications of the characteristics of information for allocative efficiency in research activities, on the one hand, with the functionalist analysis of institutional structures, reward systems and behavioral norms of **open science** communities-associated ...

Cited by 3043 Related articles All 7 versions Cite Save

Can "open science" be protected from the evolving regime of IPR protections?

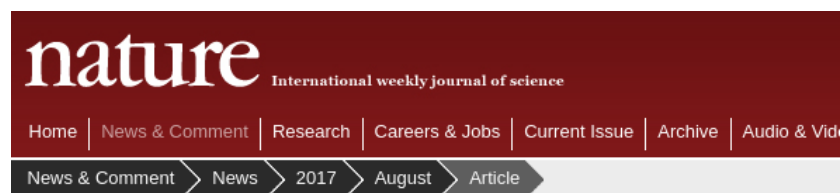
PA David - Journal of Institutional and Theoretical Economics (... 2004 - JSTOR

Increasing access charges and transactions costs arising from monopoly rights in data and information adversely affect the conduct of **science**, especially exploratory research programs. The latter are critical for the sustained growth of knowledge-driven economies,

Cited by 187 Related articles All 23 versions Cite Save

[\[PDF\] philo.at](#)

Example: *papr*



NATURE | NEWS

Swipe right for science: Papr app is 'Tinder for preprints'





App lets researchers rate life-sciences abstracts by swiping across a screen.


Lindsay McKenzie

16 June 2017



Example: *papr*

Swipe abstract to rate the paper

-  Exciting and Probable
-  Exciting and Questionable
-  Boring and Probable
-  Boring and Questionable

Rate papers & level up:
 Undergrad

Download your ratings:

Tell someone about papr:
 

InterPred: A pipeline to identify and model protein-protein interactions

Protein-protein interactions (PPI) are crucial for protein function. There exist many techniques to identify PPIs experimentally, but to determine the interactions in molecular detail is still difficult and very time-consuming. The fact that the number of PPIs is vastly larger than the number of individual proteins makes it practically impossible to characterize all interactions experimentally. Computational approaches that can bridge this gap and predict PPIs and model the interactions in molecular detail are greatly needed. Here we present InterPred, a fully automated pipeline that predicts and model PPIs from sequence using structural modelling combined with massive structural comparisons and molecular docking. A key component of the method is the use of a novel random forest classifier that integrate several structural features to distinguish correct from incorrect protein-protein interaction models. We show that InterPred represents a major improvement in protein-protein interaction detection with a performance comparable or better than experimental high-throughput techniques. We also show that our full-atom protein-protein complex modelling pipeline performs better than state of the art protein docking methods on a standard benchmark set. In addition, InterPred was also one of the top predictors in the latest CAPRI37 experiment. InterPred source code can be downloaded from <http://wallnerlab.org/InterPred>

Example: papr

Most similar Papr users to you.

Anna Gogleva: [Follow @useless_ulysses](#)
 Tanai: [Follow @Tanai_Cardona](#)
 Ryan: [Follow @rynamoshow](#)
 Steve Buyske: [Follow @ste](#)
 Jakub: [Follow @jmazur41](#)

Using your past paper rankings, these are the users that have the most similar taste in preprints to you.

Follow them, tweet at them, collaborate with them, write a new paper for people to rate on Papr!

Exercise: information discovery

1. Determine a topic you want to do some literature search on.
2. Conduct a literature search in a database of your choice (Utrecht University has access to Scopus)
3. Conduct the same literature search in OpenKnowledge Maps
4. What is the added value of each search?
5. Do you think the searches are alternatives or complements in discovery of information?
6. In a utopia, how would you like to be able to discover information?
Get funky with ideas, nothing's too crazy!