

# PAPER HIVE

Making high-quality literature **stand out**

A grayscale photograph of a person sitting at a desk, leaning forward with their head buried in their hands. The person's face is obscured by their hands, and their body language suggests a state of stress, frustration, or exhaustion. The background is a simple office setting with a desk and some papers.

What problems are we solving?

**Researchers read between 12-25 hours a week but a large portion of this time is lost trying to understand complex concepts and repeating others' mistakes.**



What problems are we solving?

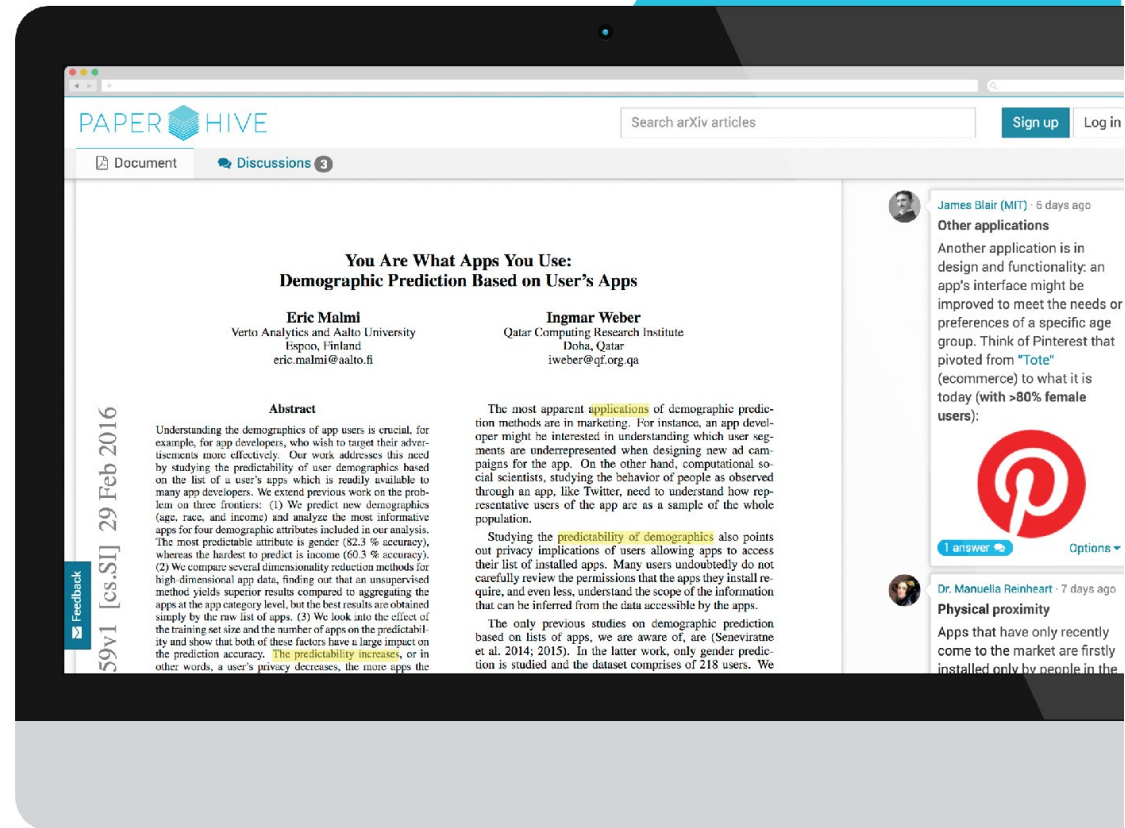
**The massive quantity of new literature makes it hard for individual content to stand out.**

How can PaperHive help?

# PaperHive is an online platform and a cross-publisher layer of interaction on top of research documents

Public and private **contextual discussions** on published content

- collaborative reading
- interactive seminars and lectures
- post-publication peer-review
- personal literature management





# By helping each other with explanations, code, data, and forward references, readers develop a web of interconnected knowledge

Every single **rich-media** annotation increases the value of content for all readers.

- citable comments
- licensed under CC-BY
- support for MathJax, LaTeX, and Markdown
- following W3C Web Annotation standards

**Demographic Prediction Based on User's Apps**

**Eric Malmi**  
Verto Analytics and Aalto University  
Espoo, Finland  
eric.malmi@aalto.fi

**Ingmar Weber**  
Qatar Computing Research Institute  
Doha, Qatar  
iweber@qi.org.qa

**Abstract**

Understanding the demographics of app users is crucial, for example, for app developers, who wish to target their advertisements more effectively. Our work addresses this need by studying the predictability of user demographics based on the list of a user's apps which is readily available to many app developers. We extend previous work on the problem on three frontiers: (1) We predict new demographics (age, race, and income) and analyze the most informative apps for four demographic attributes included in our analysis. The most predictable attribute is gender (82.3 % accuracy), whereas the hardest to predict is income (60.3 % accuracy). (2) We compare several dimensionality reduction methods for high-dimensional app data, finding out that an unsupervised method yields superior results compared to aggregating the apps at the app category level, but the best results are obtained simply by the raw list of apps. (3) We look into the effect of the training set size and the number of apps on the predictability and show that both of these factors have a large impact on the prediction accuracy. **The predictability increases**, or in other words, a user's privacy decreases, the more apps the user has used, but somewhat surprisingly, after 100 apps, the prediction accuracy starts to decrease.

**Introduction**

In 2014, 60 % of internet traffic was estimated to come from

The most apparent **applications** of demographic prediction methods are in marketing. For instance, an app developer might be interested in understanding which user segments are underrepresented when designing new ad campaigns for the app. On the other hand, computational social scientists, studying the behavior of people as observed through an app, like Twitter, need to understand how representative users of the app are as a sample of the whole population.

Studying the **predictability of demographics** also points out privacy implications of users allowing apps to access their list of installed apps. Many users undoubtedly do not carefully review the permissions that the apps they install require, and even less, understand the scope of the information that can be inferred from the data accessible by the apps.

The only previous studies on demographic prediction based on lists of apps, we are aware of, are (Seneviratne et al. 2014; 2015). In the latter work, only gender prediction is studied and the dataset comprises of 218 users. We have obtained a dataset of 3 760 users, which allows us to perform more fine-grained analyses, e.g., looking into the effect of app count on the predictability, and to obtain statistically more reliable results.

James Ily (MIT) 5 minutes ago  
**Other applications**  
Another application is in design and functionality: an app's interface might be improved to meet the needs or preferences of a specific age group. Think of Pinterest that pivoted from "Tote" (e-commerce) to what it is today (with +93% female users).

Histogram of ads

Dr. Manuella Reinherm 8 hours ago  
**Physical proximity**  
Apps that have only recently come to the market are firstly installed only by people in the relative physical proximity of the app creators. A model would go like that:

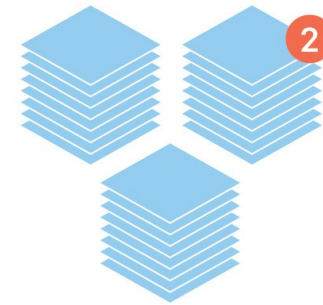
$$\int x^2 dx = \frac{x^3}{3} + C$$

Alexander Neumann 4 months ago  
**Clustering of users and classification**  
Wouldn't the predictability increase even more if different classification models are used for different user types, sup. users with a few apps vs. users with many apps?

# PaperHive embeds communication in the researcher's workflow



Smart **deep links** to specific passages of academic articles and books increase sharing of content.



Researchers and students collect the articles relevant for them in their own **Hives** and receive **personalized notifications** for new updates and comments.

# Benefit 1: Increasing the audience of content among students and interdisciplinary researchers

Editors, authors or publishers can add background information to **increase the accessibility** of content.

## ative geometry and Painlevé equations

Okounkov and Eric Rains

### Abstract

Elliptic Painlevé equation and its higher dimension of line bundles on 1-dimensional sheaves surfaces.

ations are very special 2-dimensional dynamical generalizations (including discretizations) appear their theory is very well developed, in fact, from for example [19] for an introduction. Many of geometric and some can be interpreted in terms of. A full discussion of the relation between the

AMS · 8 minutes ago

### Overview and Background

Paul Painlevé, mathematician and an important political figure of France, discovered the celebrated nonlinear differential equations on the turn of the century. **Typical examples of the Painlevé's equations** are Gauss' hypergeometric functions, Kummer's confluent hypergeometric functions, and various special functions with the name of Airy, Bessel, Hermite, etc.

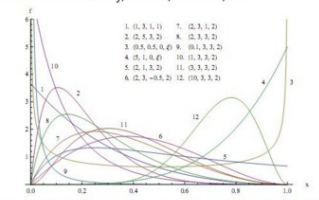



Figure 1: Graph of the Gauss hypergeometric density  $f_{21}(z; \alpha, \beta, \gamma, \delta)$  for different values of  $(\alpha, \beta, \gamma, \delta)$ .

PaperHive · a month ago

### Interview with Prof. Okounkov



Muñoz & Persson: How did you get interested in mathematics?

Okounkov: The most important part of becoming a mathematician is learning from one's teachers. Here I was very fortunate. Growing up in Kirillov's seminar, I had in its participants, especially in Grisha Olshanski, wonderful



## Benefit 2: Extending the life time of articles

Replies and new discussions create **incentives for readers to return** to documents they previously worked with.



Alexander Naydenov · 2 hours ago

### **Clustering of users and classification**

Wouldn't the predictability increase even more if different classification models are used?

1 answer 

Options 



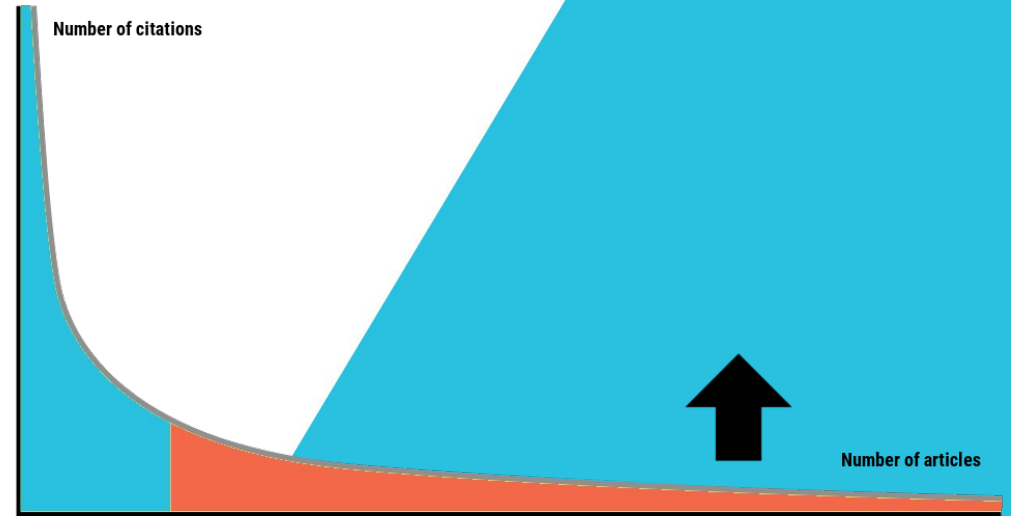
James Blair · 1 hour ago

An increase of accuracy could also be achieved by controlling for the device type.

## Benefit 3: Boosting the usage of articles in the long-tail

Discussions on less frequently cited and read articles increase their value **over-proportionally** relative to popular content.

- niche experts are brought closer together
- benefit from each others' knowledge and results



# Case Study 1: Engagement with great books

## Knowledge Unlatched

- organization helping publishers and libraries worldwide to work together for a sustainable open future for scholarly books
- starts using PaperHive for enabling **communication on selected book titles** in Humanities and Social Sciences (May 1, 2016)



# Case Study 2: Proofreading & peer review

## LangSci Press

- an open access publisher at Freie Universität Berlin
- starts using PaperHive for their **community proofreading** and open peer review projects on articles and books (May 31, 2016)



# Integration and partners

- Easy and fast integration: PaperHive only requires **basic article meta data** from publisher
- All article traffic still goes to publisher
- Publishers and repositories we are already working with include:



- Discussions are safely archived with trusted preservation service providers
- PaperHive is financially supported by:





Put great literature  
in the limelight at  
PaperHive.org!

Alexander Naydenov | Dr. André Gaul  
0049176 9783 5783 | [info@paperhive.org](mailto:info@paperhive.org)  
[@paperhive](#) | Ackerstr. 76, Berlin