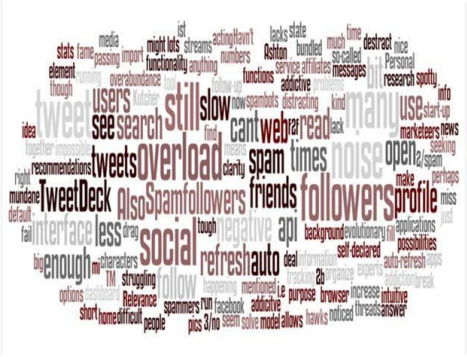
Understanding and Supporting search for scholarly knowledge



Marcos Baez, Daniil Mirylenka and
Cristhian Parra

SCOPUS
Europäisches Patentamt
European Patent Office
Office européen des brevets
Google Scholar BETA
ELSEVIER SCIENCE
Blogger
THOMSON REUTERS
ScienceBlog
science news straight from the source
OpenMED@NIC
Share
arXiv.org
dblp.uni-trier.de
Computer Science Bibliography
Free Medical Journals
Promoting free access to medical journals
PLoS
The Problem
acm
BASE
Bielefeld Academic Search Engine
CiteSeerx beta
Google Docs
crossref
PubMed
citeulike
SpringerLink
COGPRINTS
COGNITIVE SCIENCES EPRINT ARCHIVE
eprints
education research global observatory
scanning the world for open access scholarship in education
RePEc

The Goal

To understand and improve the way find Scientific Knowledge

Understanding

Supporting

Understanding

Qualitative

30 Semi-Structured Interviews

Analyzing references on authored papers

Quantitative

Online Survey based on qualitative results

Qualitative Study

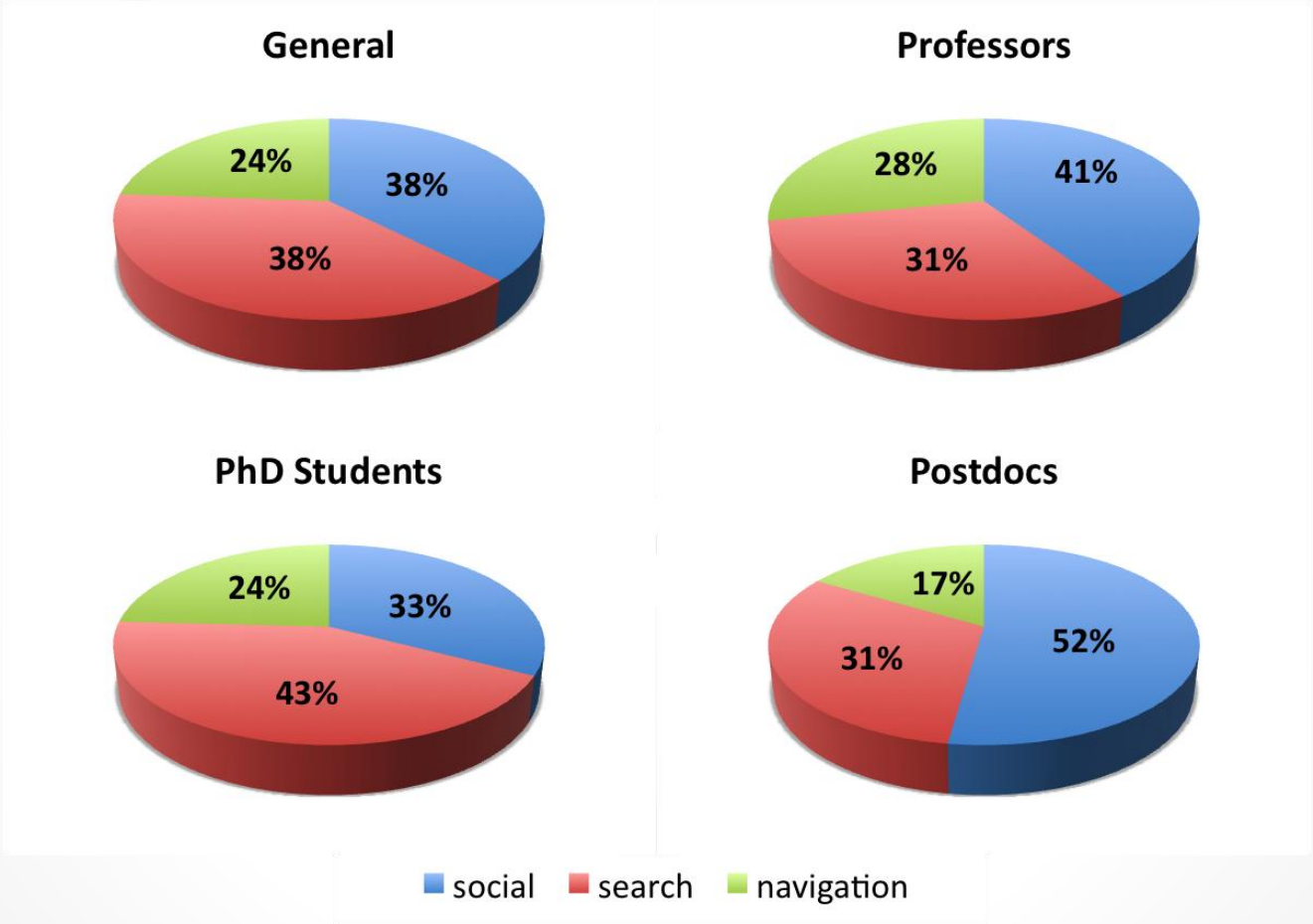
30 Semi-Structured Interviews

For each citation of one authored paper:

How did you get to this paper?

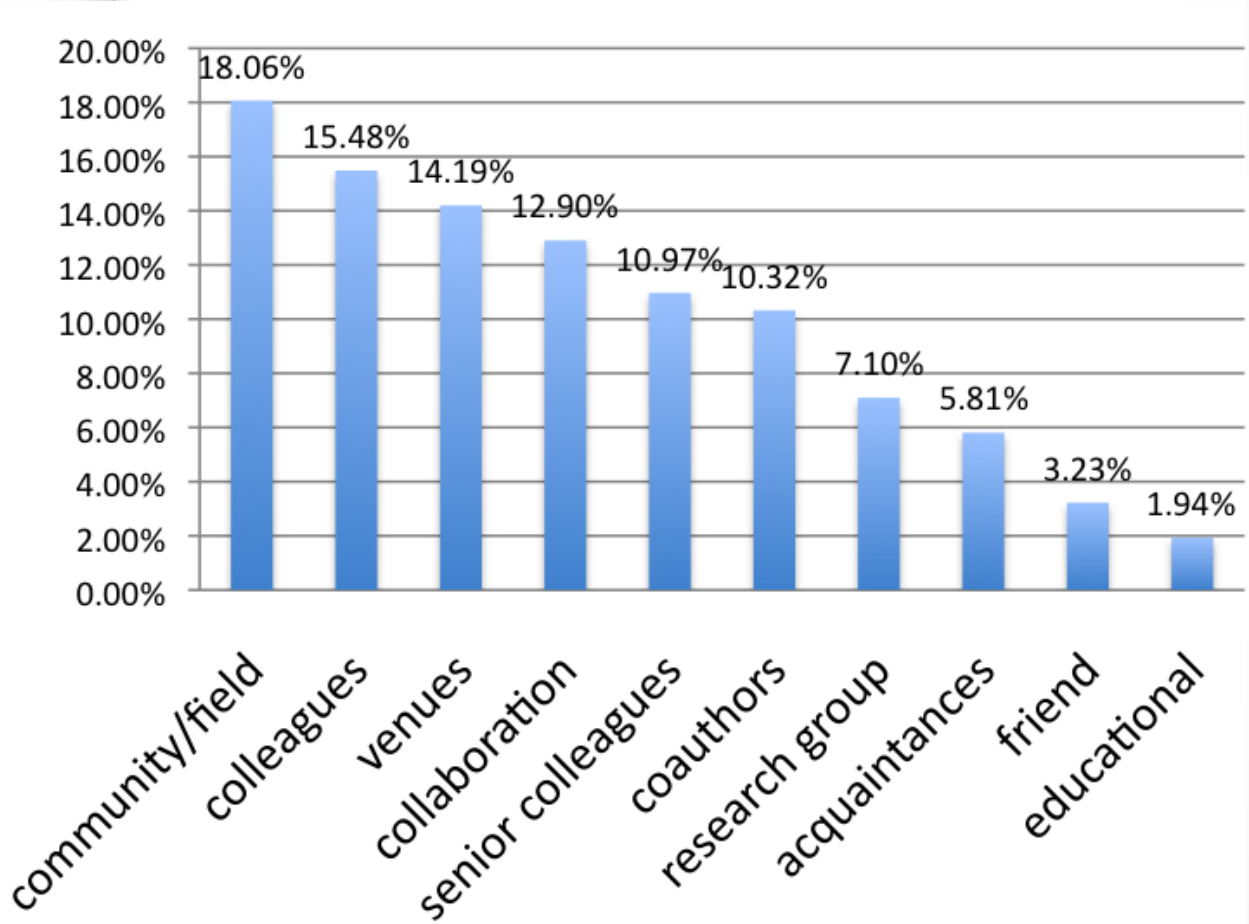
"My advisor suggested it"
"I search for X on Google Scholar"
"I was following some citations in a paper"

Qualitative Results



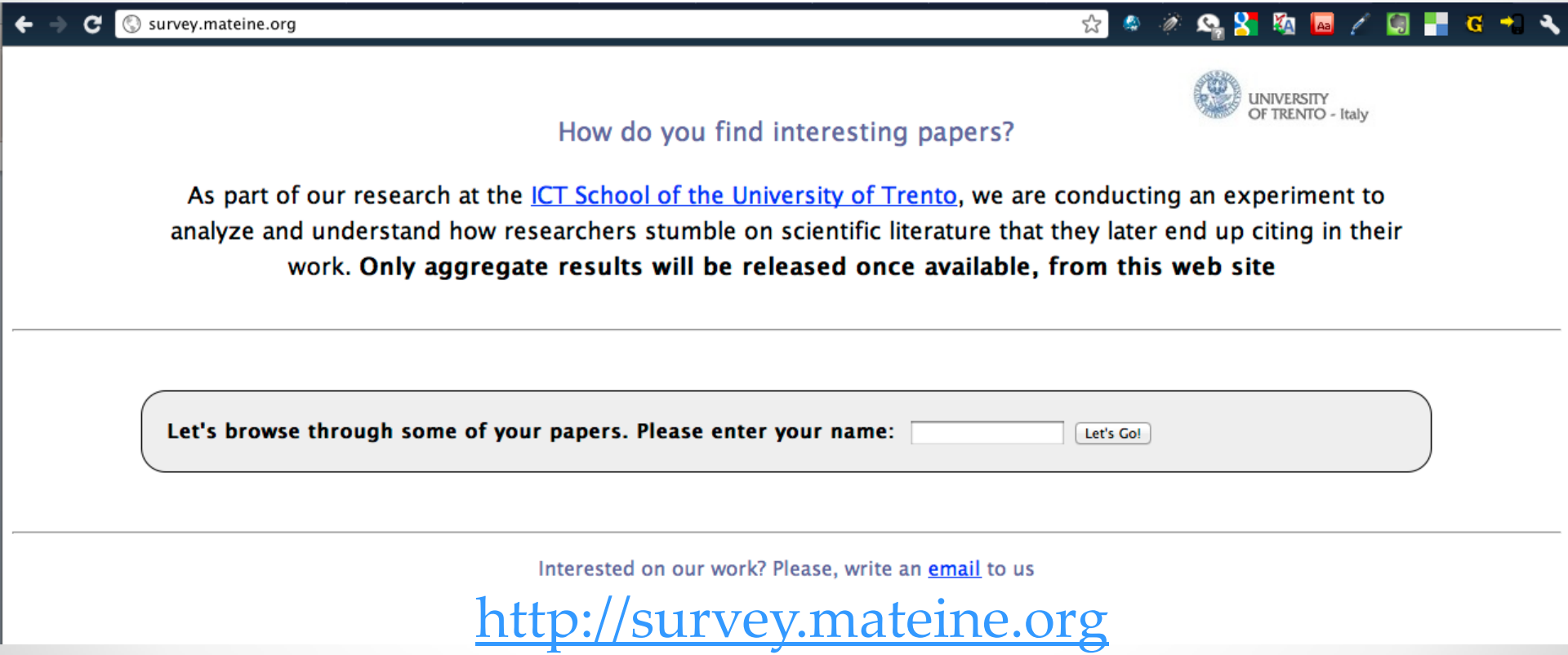
~ 700 interview notes. ~ 250 references. ~16% of Self citations.

Qualitative Results



Social Networks

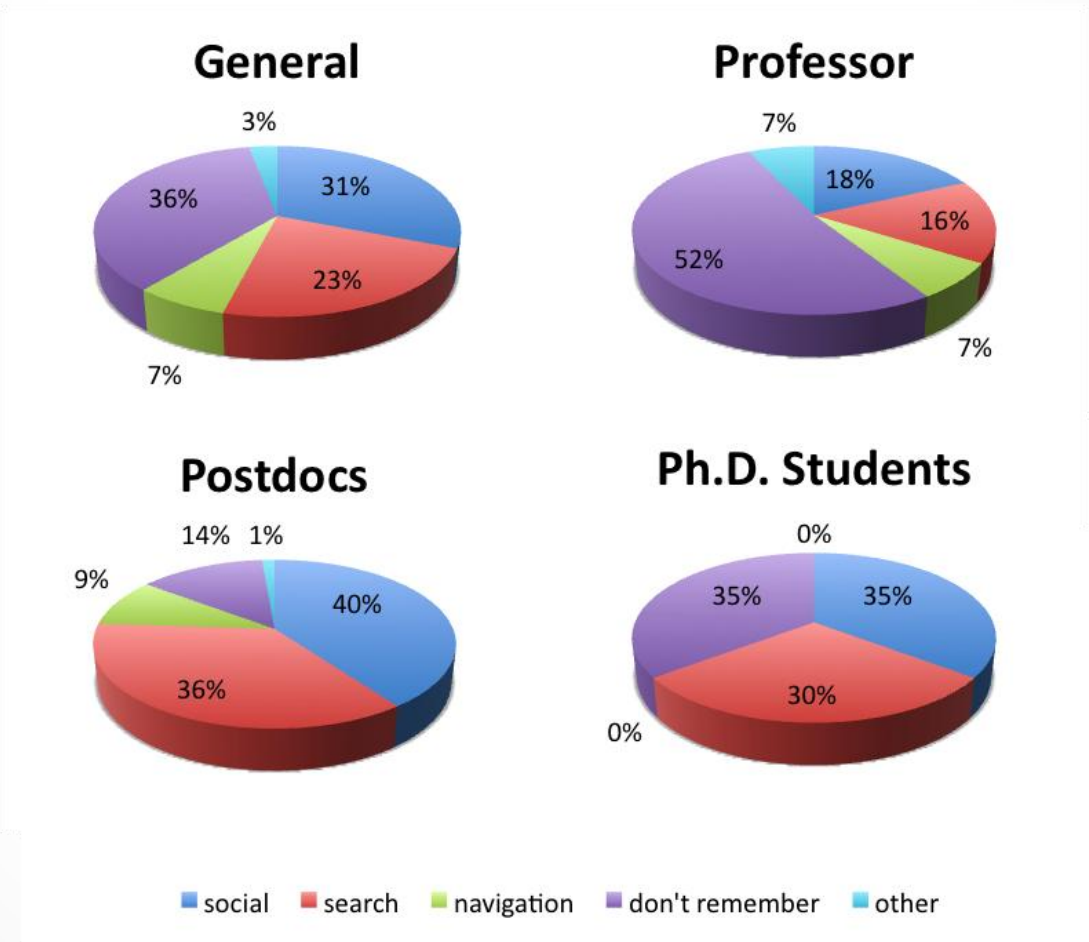
Quantitative Study



http://survey.mateine.org

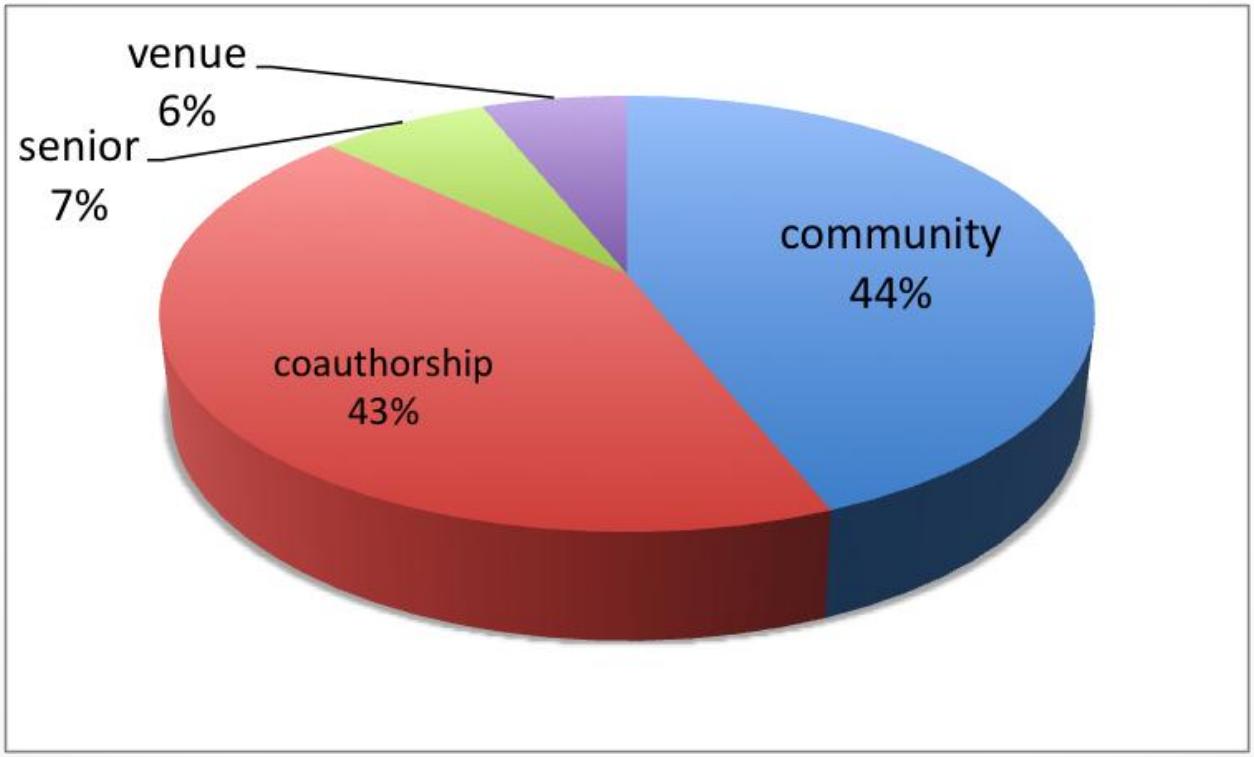
~ 5 x 10^6 publications with references from Microsoft Academic Search

Quantitative Results



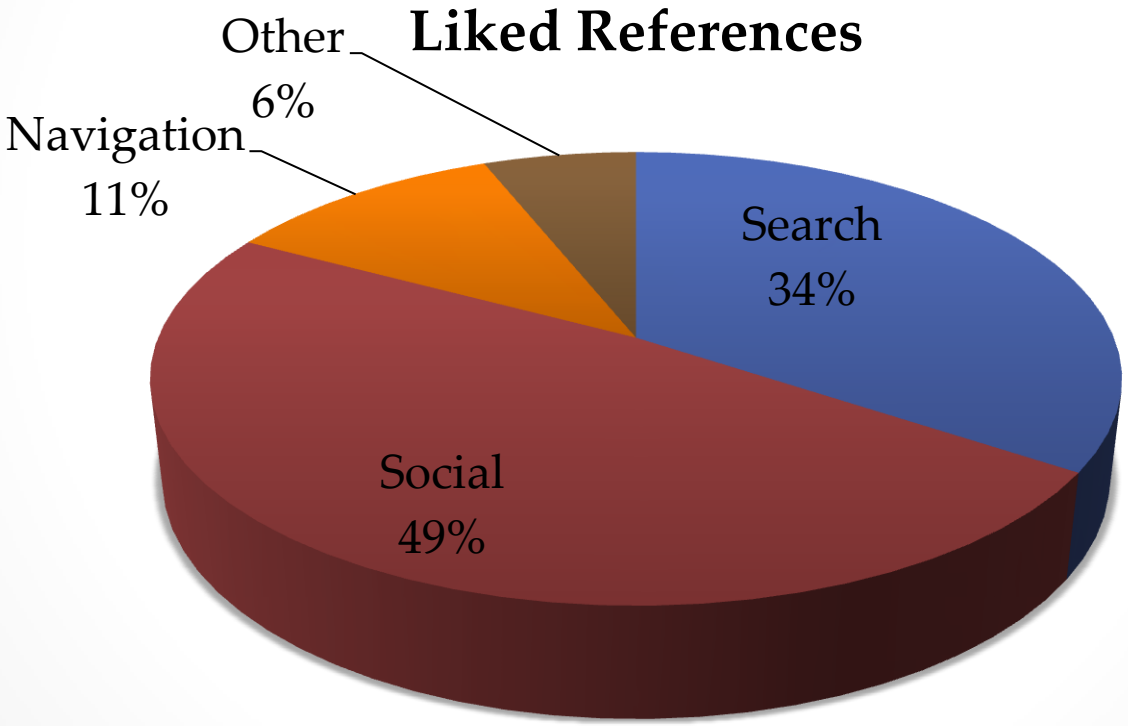
~ 226 References. ~ 23 different publications. ~ 25 different authors

Quantitative Results



Social Networks

Quantitative Results



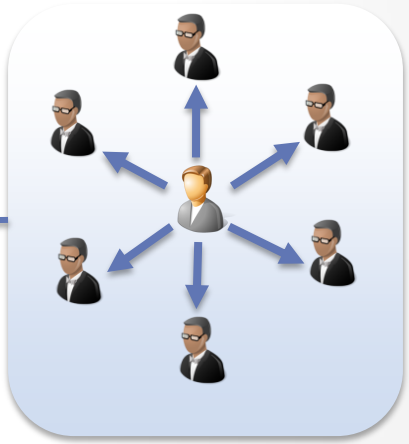
Our Understanding

Finding scientific knowledge has a strong social component to be leveraged

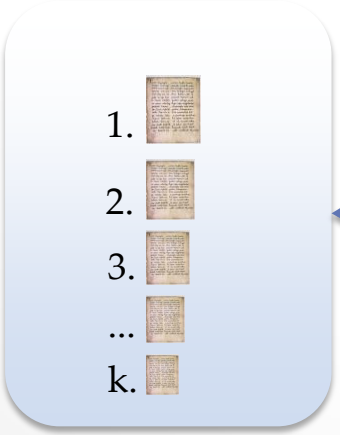
Supporting: Our Goal

Incorporating the social component in scholarly search by using the knowledge of researcher's social networks

Researcher's network



Coauthorship, Venue, Topic



Social-Aware Search Results

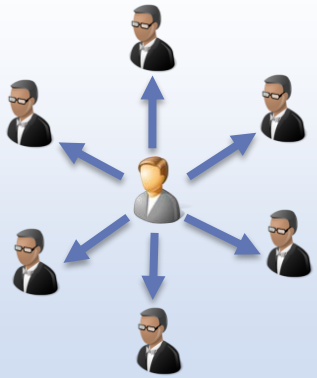
Experiment setup

T = year 2005

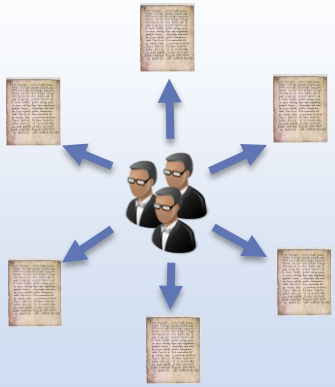
"past" "future"

Sample of 1000 researchers

Researcher's network at time T



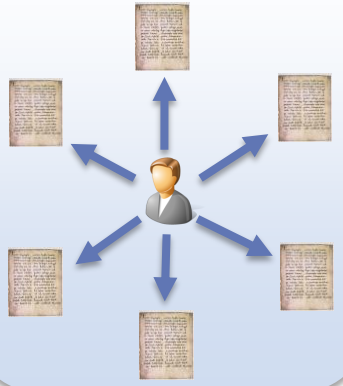
"Past" references of the network



Top k recommendations

1.
2.
3.
...
k.

Researcher's "future" references



Calculate:

precision

recall

depending on

the **number of recommendations**

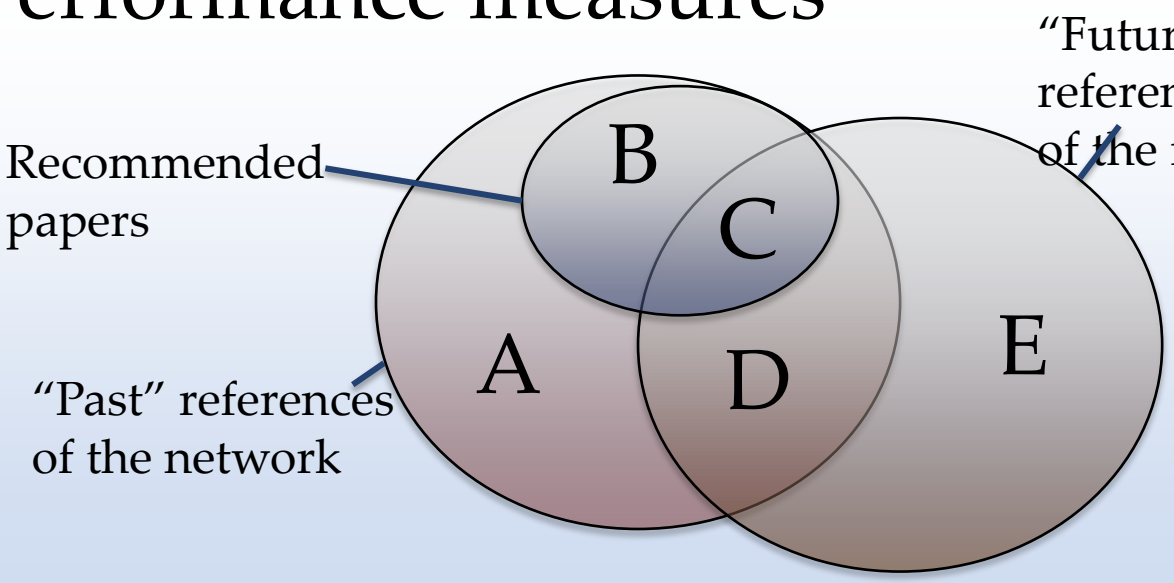
popularity measure

Experiment setup

Popularity metrics:

network – papers most cited in the network, **overall** – papers most cited overall, and **random** – random set of papers

Performance measures

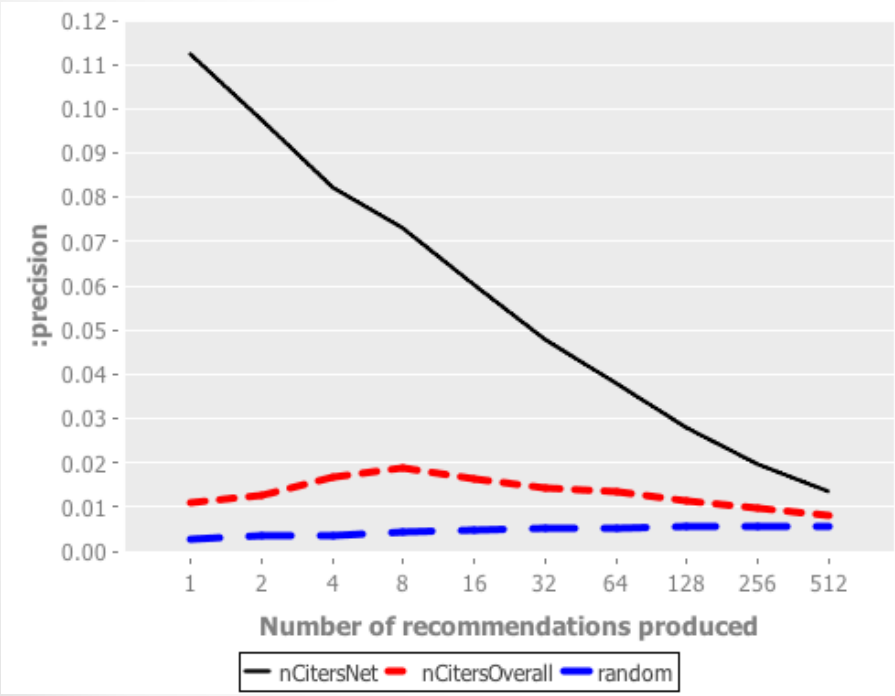


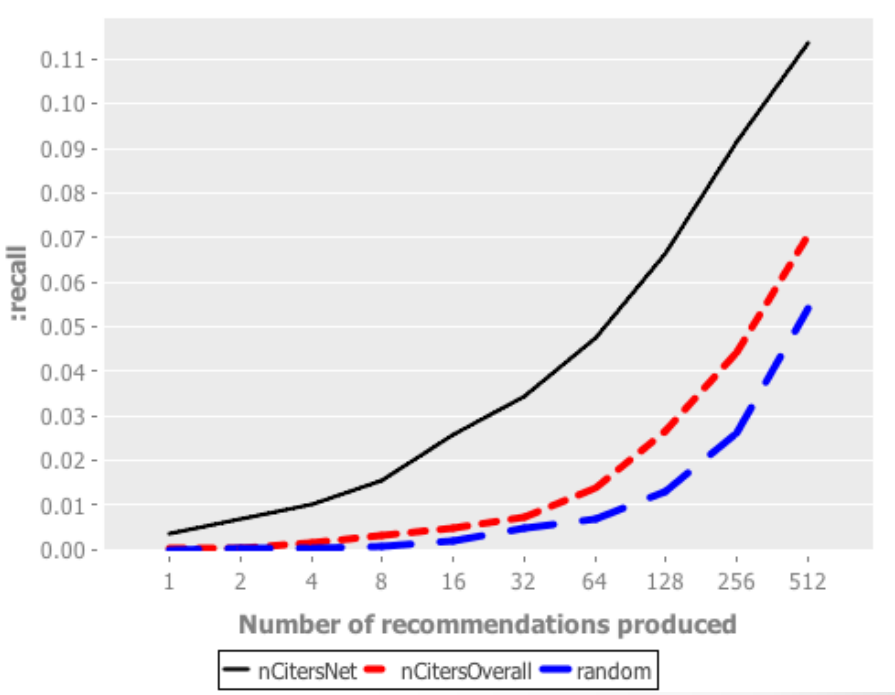
precision: $\frac{|C|}{|C+B|}$

recall: $\frac{|C|}{|C+D+E|}$

Results

11% = P(you will cite the paper most cited between your coauthors)
20% of the papers you are going to cite have been cited by your coauthors





Ideas: **Limiting** the number of coauthors
Weighting the recommendation with the number of coauthored papers
DID NOT improve the precision

Thanks for your Attention