gesis

Leibniz Institute
for the Social Sciences

Query for Architecture, Click through Military: Comparing the Roles of Search and Navigation on Wikipedia

Dimitar Dimitrov, Florian Lemmerich, Fabian Flöck and Markus Strohmaier

Wikimedia Research Showcase - August 2020

Dimitar Dimitrov, Florian Lemmerich, Fabian Flöck, and Markus Strohmaier, 2018. Query for Architecture, Click through Military: Comparing the Roles of Search and Navigation on Wikipedia. In *Proceedings of the 10th ACM Conference on Web Science*.

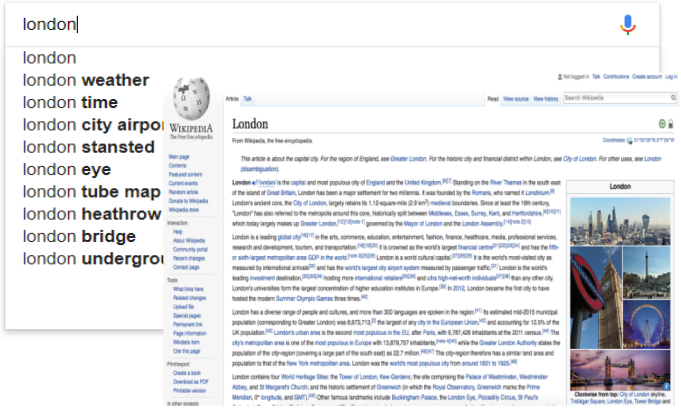
Dimitar Dimitrov, Florian Lemmerich, Fabian Flöck, and Markus Strohmaier, 2019. Different topic, different traffic: How search and navigation interplay on Wikipedia. *The Journal of Web Science*, *6*.

Member of the Leibniz Association

Information Access on Wikipedia

- Search

Google

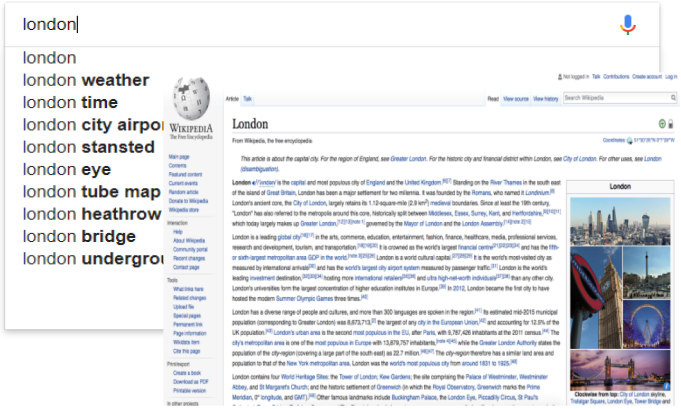


gesis Leibniz Institute for the Social Sciences

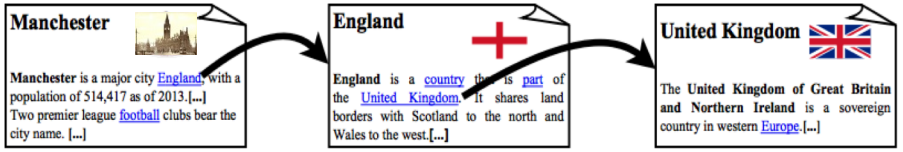
Information Access on Wikipedia

- Search

Google



- Navigation



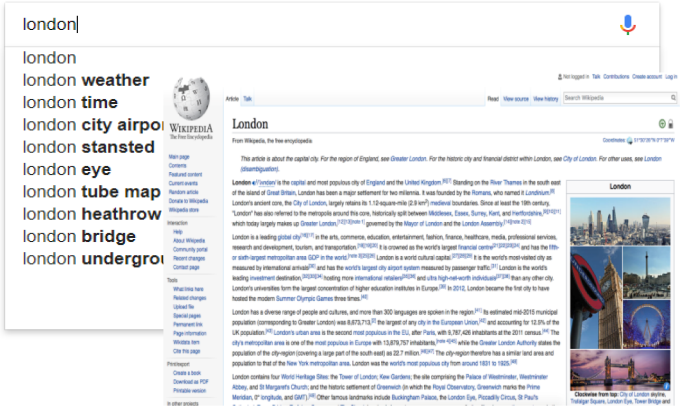
Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

Information Access on Wikipedia

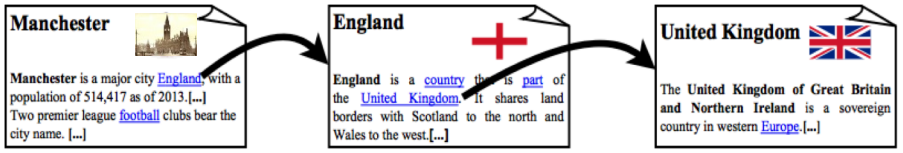
- Search

Google



~70% of all visits

- Navigation



~30% of all visits

Ellery Wulczyn and Dario Taraborelli. Wikipedia clickstream. figshare. doi:10.6084/m9.figshare.1305770. (accessed March. 29, 2018).

Member of the Leibniz Association

Not logged in Talk Contributions Create account Log in

Amsterdam

From Wikipedia, the free encyclopedia

Coordinates: 52°22′N 4°54′E

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes

Amsterdam (/ˈæmstərdæm/;[9][10][11] Dutch: [ɑmstərˈdɑm] (listen)) is the capital and most populous municipality of the Netherlands. Its status as the capital is mandated by the Constitution of the Netherlands,[12] although it is not the seat of the government, which is The Hague.[13] Amsterdam has a population of 851,373 within the city proper, 1,351,587 in the urban area,[14] and 2,410,960 in the Amsterdam metropolitan area.[8] The city is located in the province of North Holland in the west of the country but is not its capital, which is Haarlem. The metropolitan area comprises much of the northern part of the Randstad, one of the larger conurbations in Europe, with a population of approximately 7 million.[15]

Amsterdam's name derives from *Amstelredamme*,[16] indicative of the city's origin around a dam in the river Amstel. Originating as a small fishing village in the late 12th century, Amsterdam became one of the most important ports in the world during the Dutch Golden Age (17th century), a result of its innovative developments in trade. During that time, the city was the leading centre for finance and diamonds.[17] In the 19th and 20th centuries the city expanded, and many new neighbourhoods and suburbs were planned and built. The 17th-century canals of Amsterdam and the 19–20th century Defence Line of Amsterdam are on the UNESCO World Heritage List. Since the annexation of the municipality of Sloten in 1921 by the municipality of Amsterdam, the oldest historic part of the city lies in Sloten (9th century).

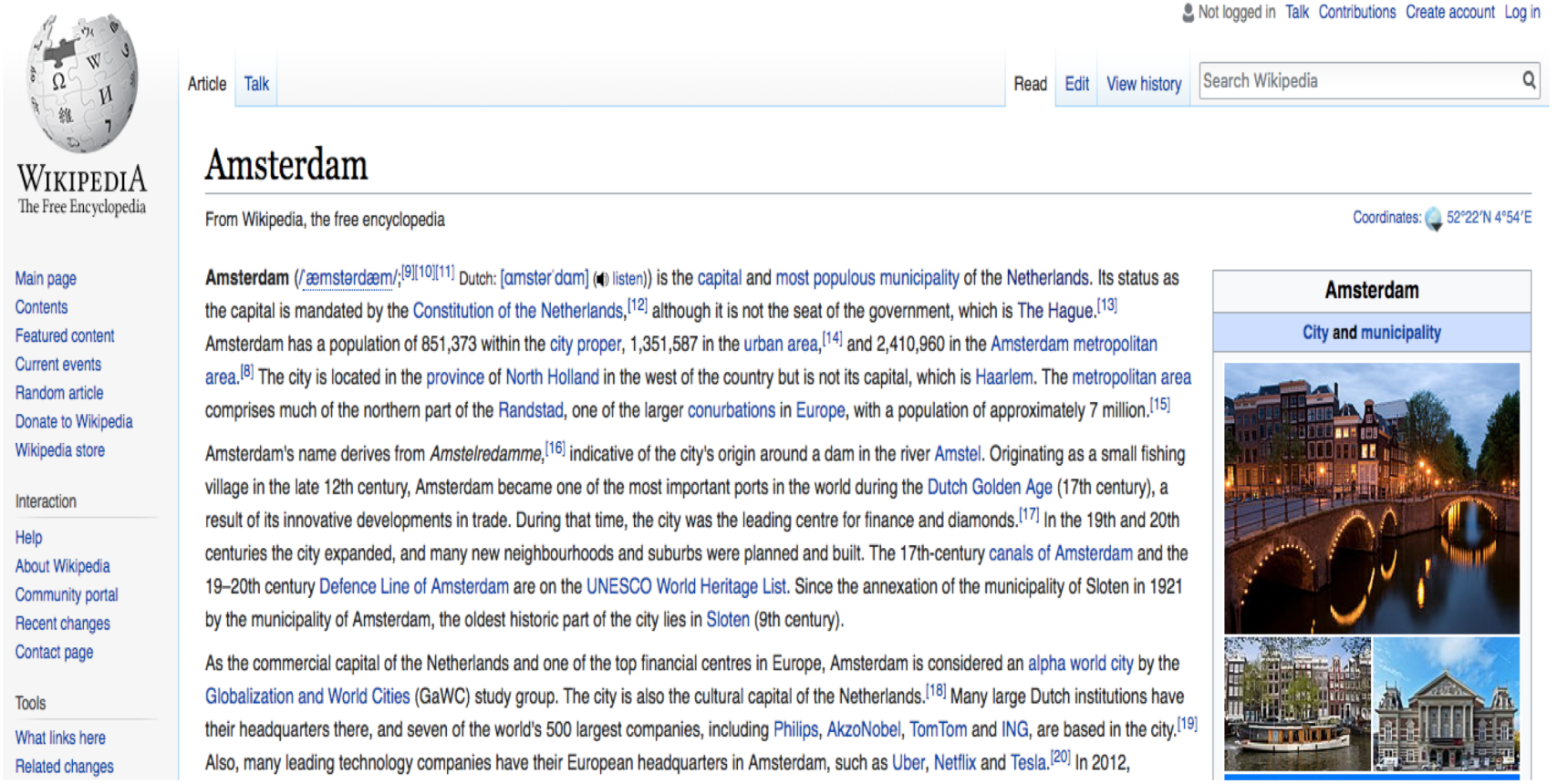
As the commercial capital of the Netherlands and one of the top financial centres in Europe, Amsterdam is considered an alpha world city by the Globalization and World Cities (GaWC) study group. The city is also the cultural capital of the Netherlands.[18] Many large Dutch institutions have their headquarters there, and seven of the world's 500 largest companies, including Philips, AkzoNobel, TomTom and ING, are based in the city.[19] Also, many leading technology companies have their European headquarters in Amsterdam, such as Uber, Netflix and Tesla.[20] In 2012,

Amsterdam

City and municipality

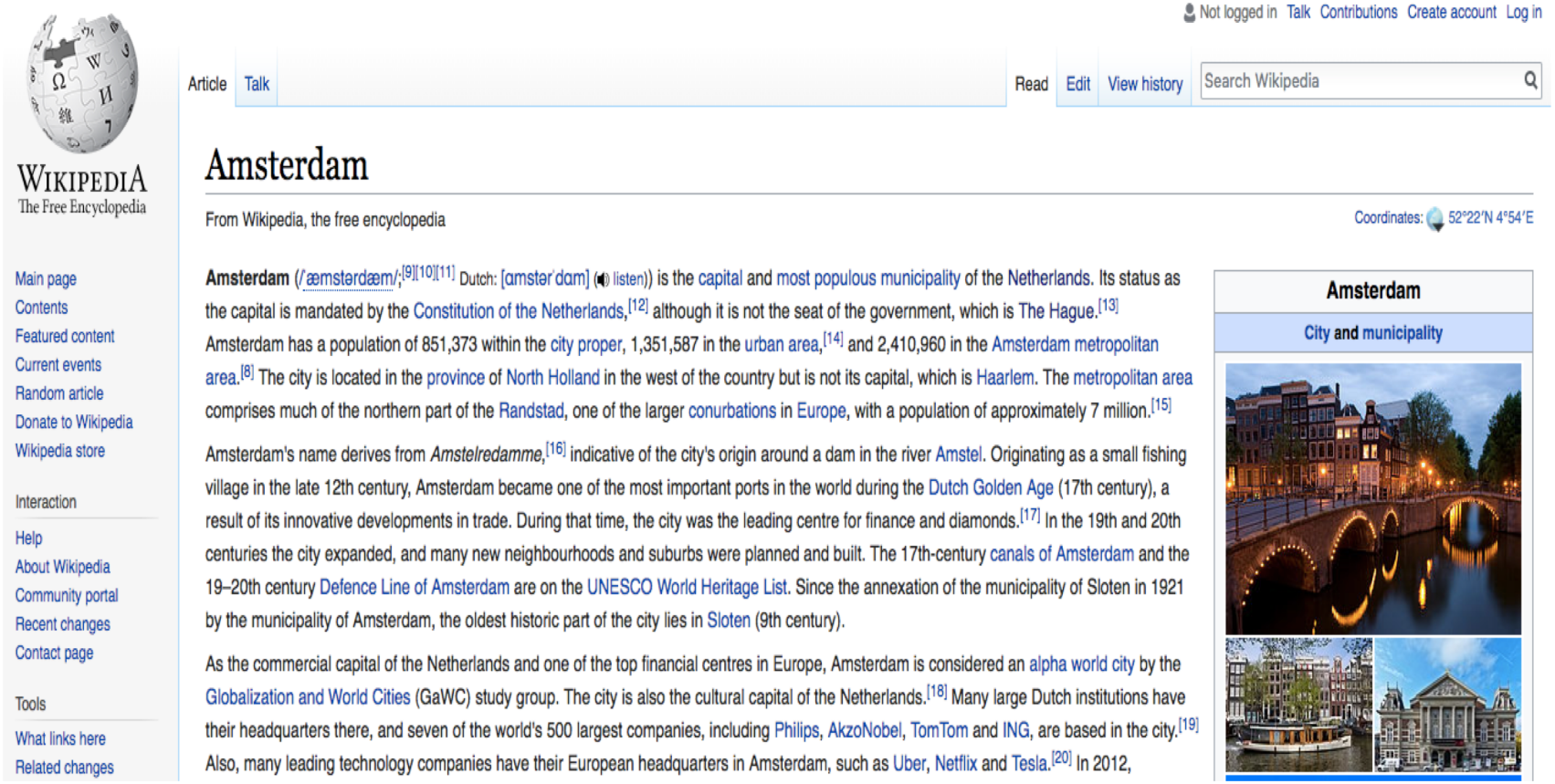
gesis Leibniz Institute for the Social Sciences

How do search and navigation interplay to shape the article traffic on Wikipedia?



How do search and navigation interplay to shape the article traffic on Wikipedia?

Which article properties are indicative of specific information access behavior?



Dataset

- Wikipedia clickstream
- Counts of referrer-resource pairs
- English Wikipedia (August, 2016)
- ~ 3,1 million articles
- ~ 4 billion views

Referrer	Resource	Access form	Counts
Google	Amsterdam	Search	1340
London	Amsterdam	Navigation	230
Paris	Amsterdam	Navigation	360
Yahoo	Amsterdam	Search	270

Ellery Wulczyn and Dario Taraborelli. Wikipedia clickstream. figshare. doi:10.6084/m9.figshare.1305770. (accessed March. 29, 2018).

Methodology

Step 1
- Traffic features and article roles

Step 2
- Topic-specific heatmaps

Step 3
- Modeling access behavior

Which article properties are indicative of specific information access behavior?

What are the functional roles articles can assume for the traffic flow?

Step 1

Traffic features

$$searchshare(a) = \frac{in_{se}(a)}{in_{se}(a) + in_{nav}(a)}$$

Incoming search traffic

Incoming navigation traffic

Total pageviews

Searchshare – the amount of views an article received by search

Step 1

Traffic features

$$searchshare(a) = \frac{in_{se}(a)}{in_{se}(a) + in_{nav}(a)}$$

Incoming search traffic

Incoming navigation traffic

Total pageviews

Searchshare – the amount of views an article received by search

$$resistance(a) = 1 - \frac{out_{nav}(a)}{in_{se}(a) + in_{nav}(a)}$$

Outgoing navigation traffic

Total pageviews

Resistance – the ability of an article to relay traffic to other Wikipedia articles

Step 1

Traffic features

$$searchshare(a) = \frac{in_{se}(a)}{in_{se}(a) + in_{nav}(a)}$$

Incoming search traffic

Incoming navigation traffic

Total pageviews

Searchshare – the amount of views an article received by search

$$resistance(a) = 1 - \frac{out_{nav}(a)}{in_{se}(a) + in_{nav}(a)}$$

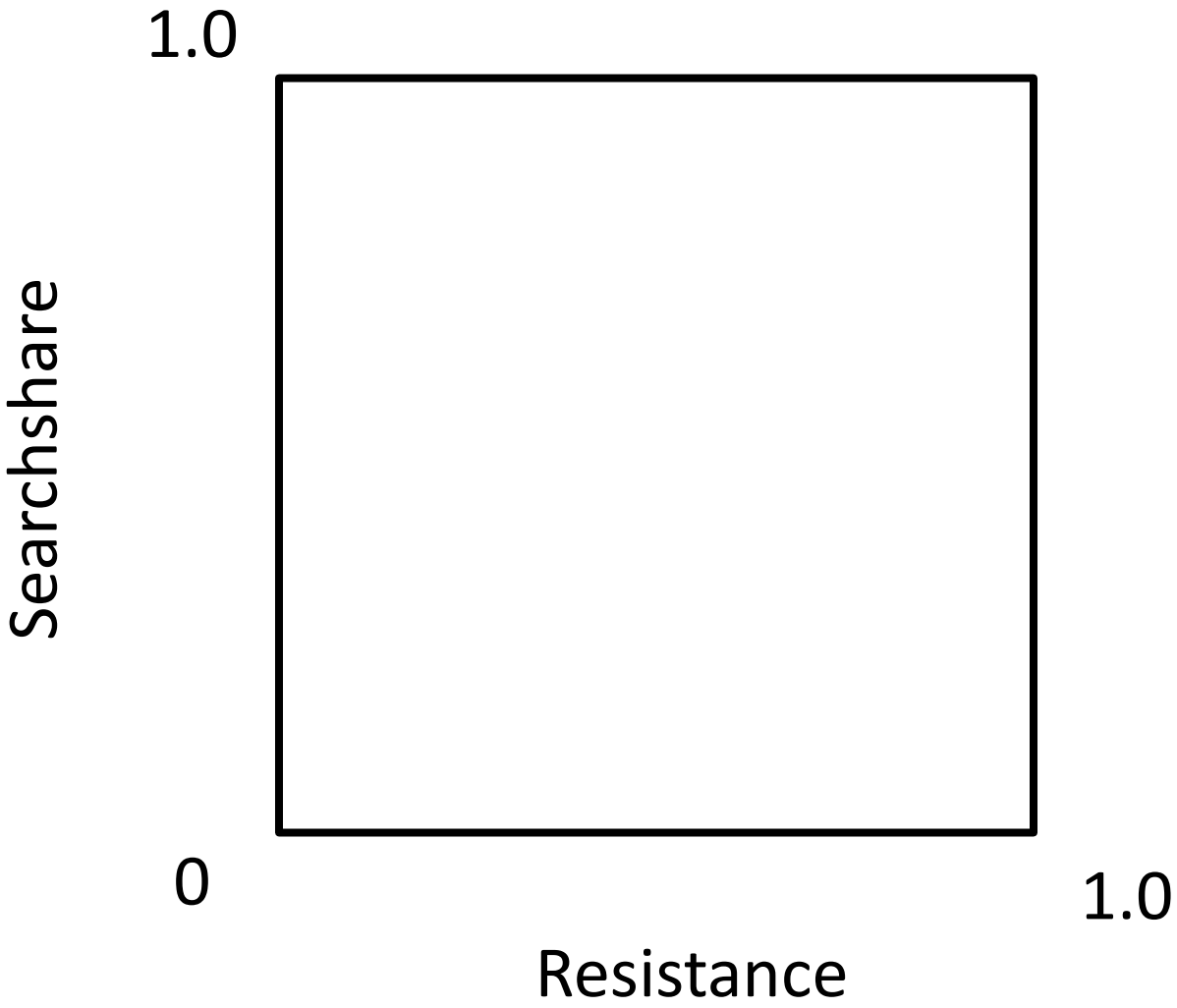
Outgoing navigation traffic

Total pageviews

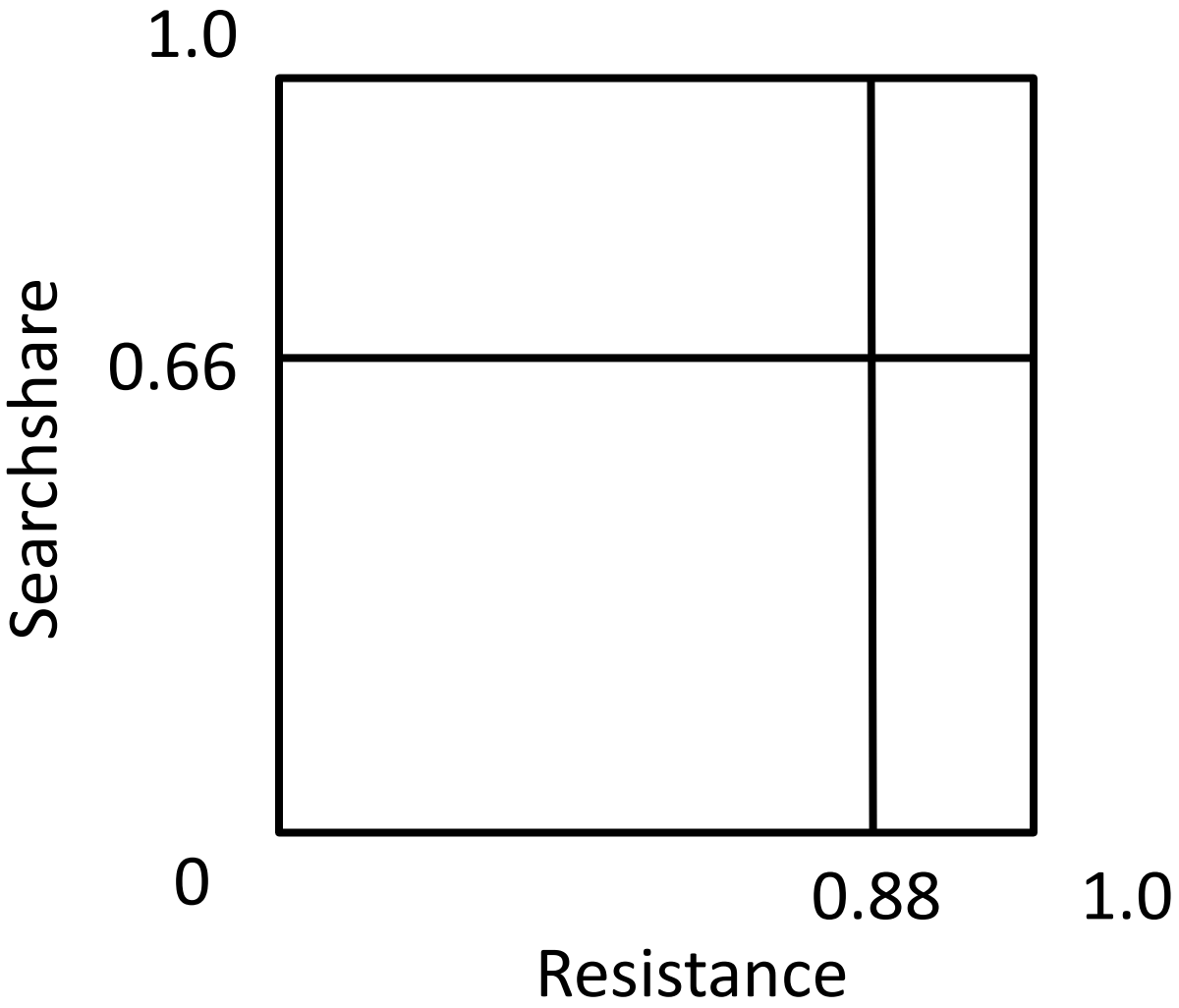
Resistance – the ability of an article to relay traffic to other Wikipedia articles

	Searchshare	***Resistance***
Mean	0.66	0.88

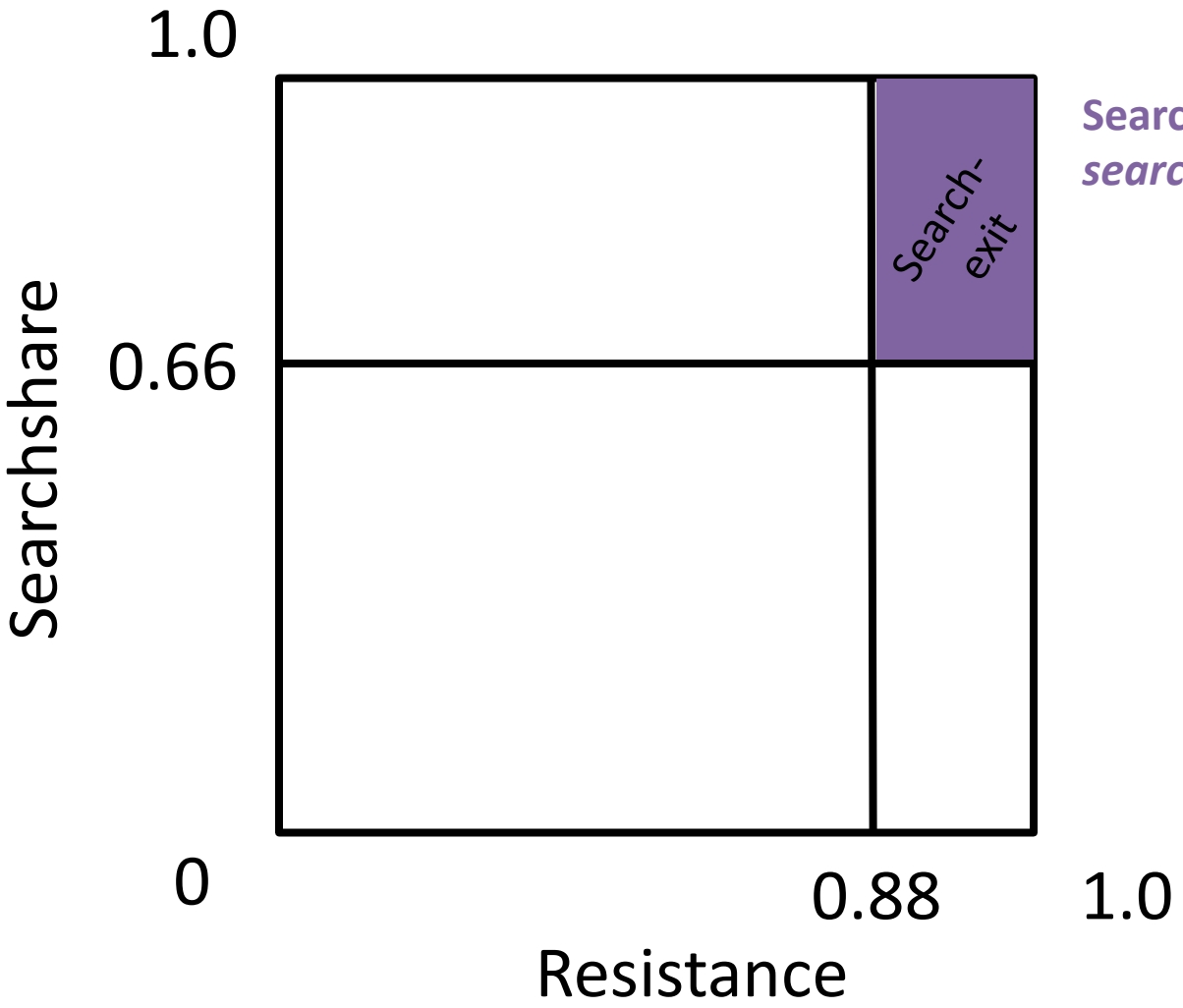
Article roles



Article roles

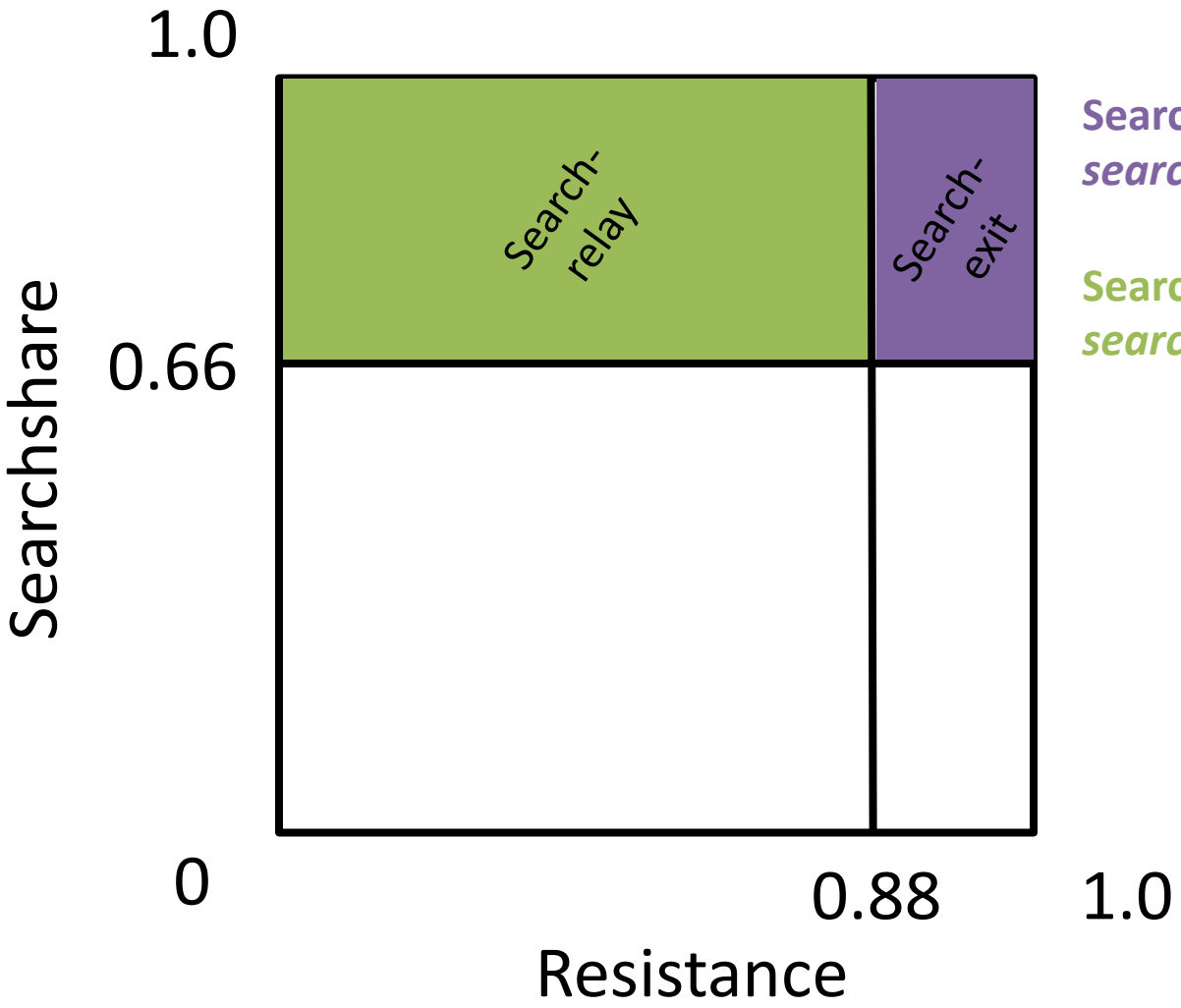


Article roles



Search-exit – articles accessed by *search* that are *unable* to relay traffic

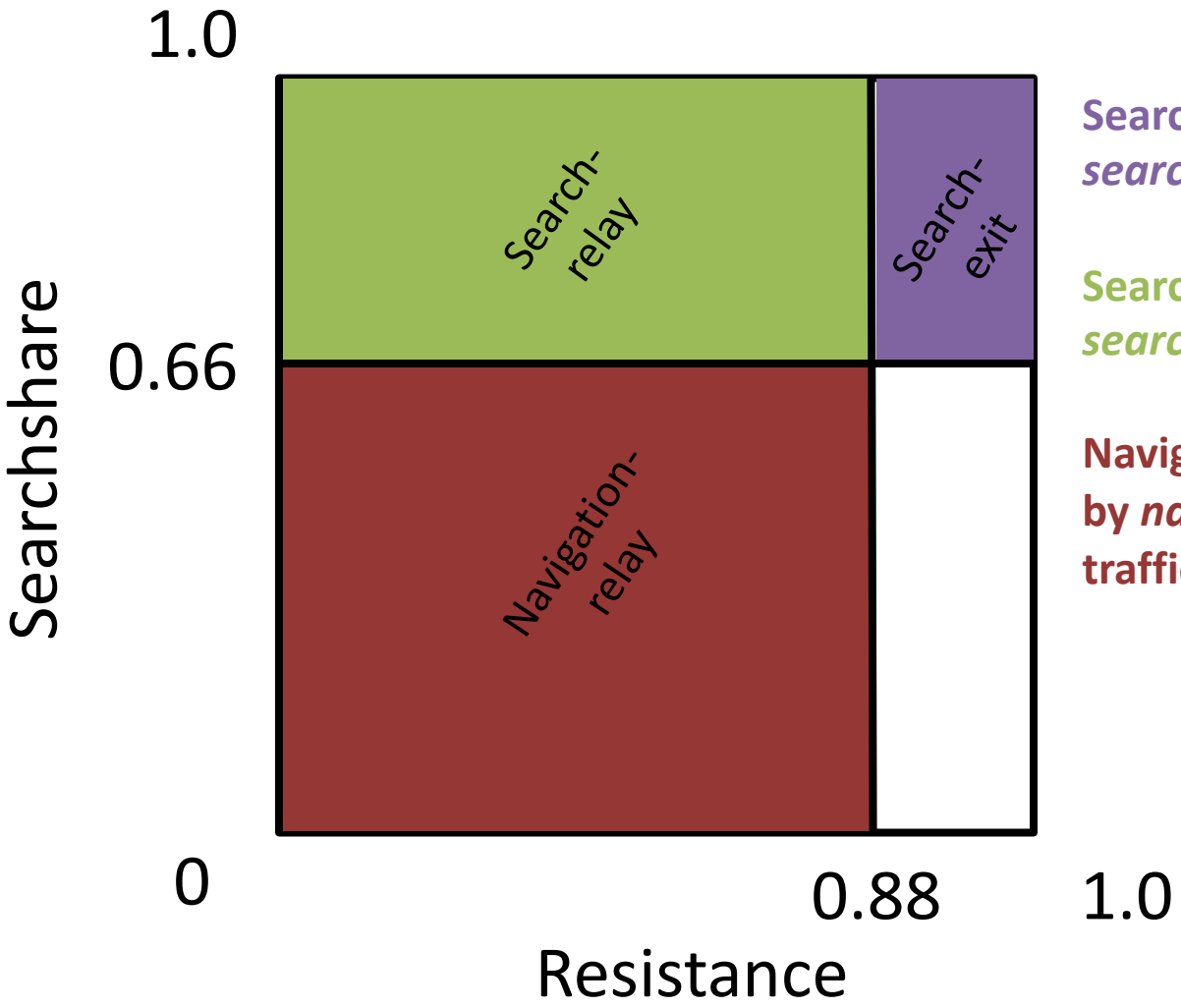
Article roles



Search-exit – articles accessed by *search* that are *unable* to relay traffic

Search-relay – articles accessed by *search* that are *able* to relay traffic

Article roles

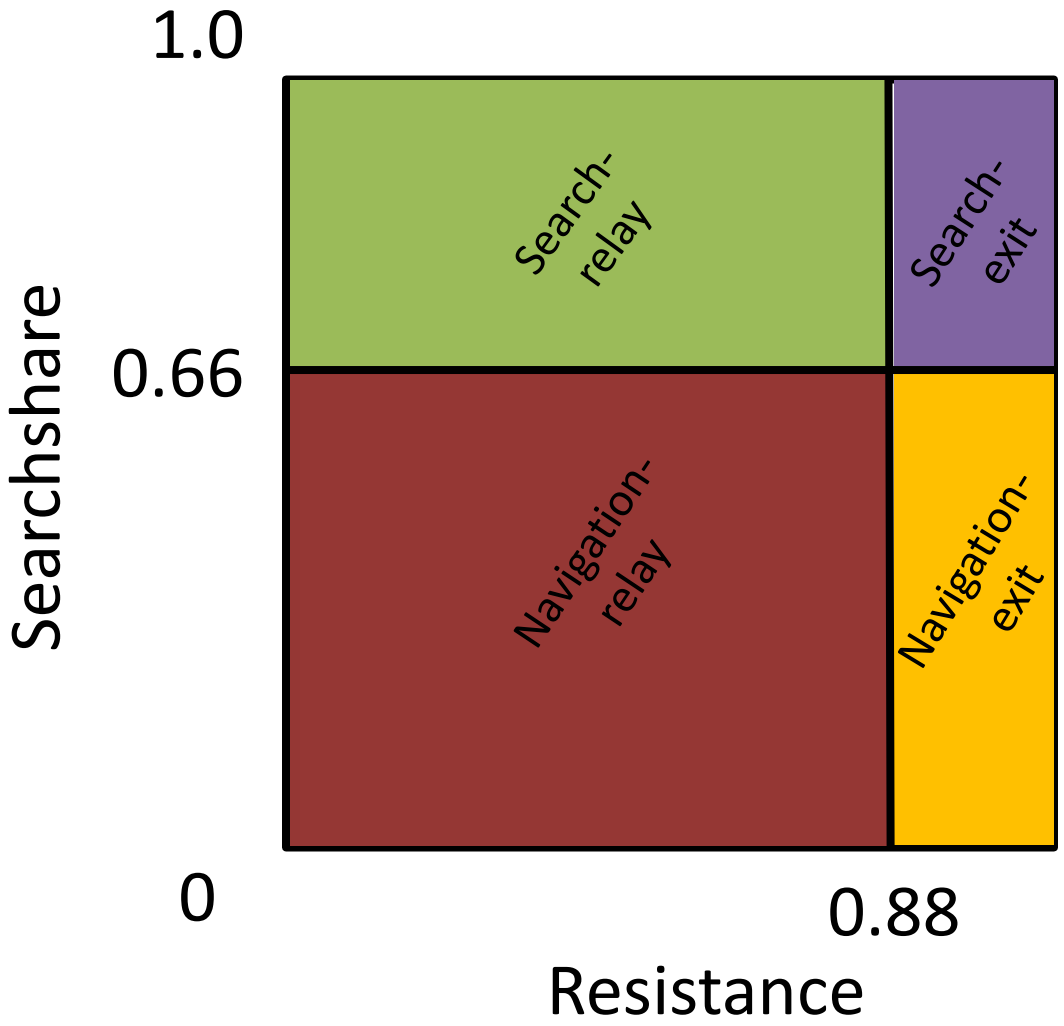


Search-exit – articles accessed by *search* that are *unable* to relay traffic

Search-relay – articles accessed by *search* that are *able* to relay traffic

Navigation-relay – articles accessed by *navigation* that are *able* to relay traffic

Article roles



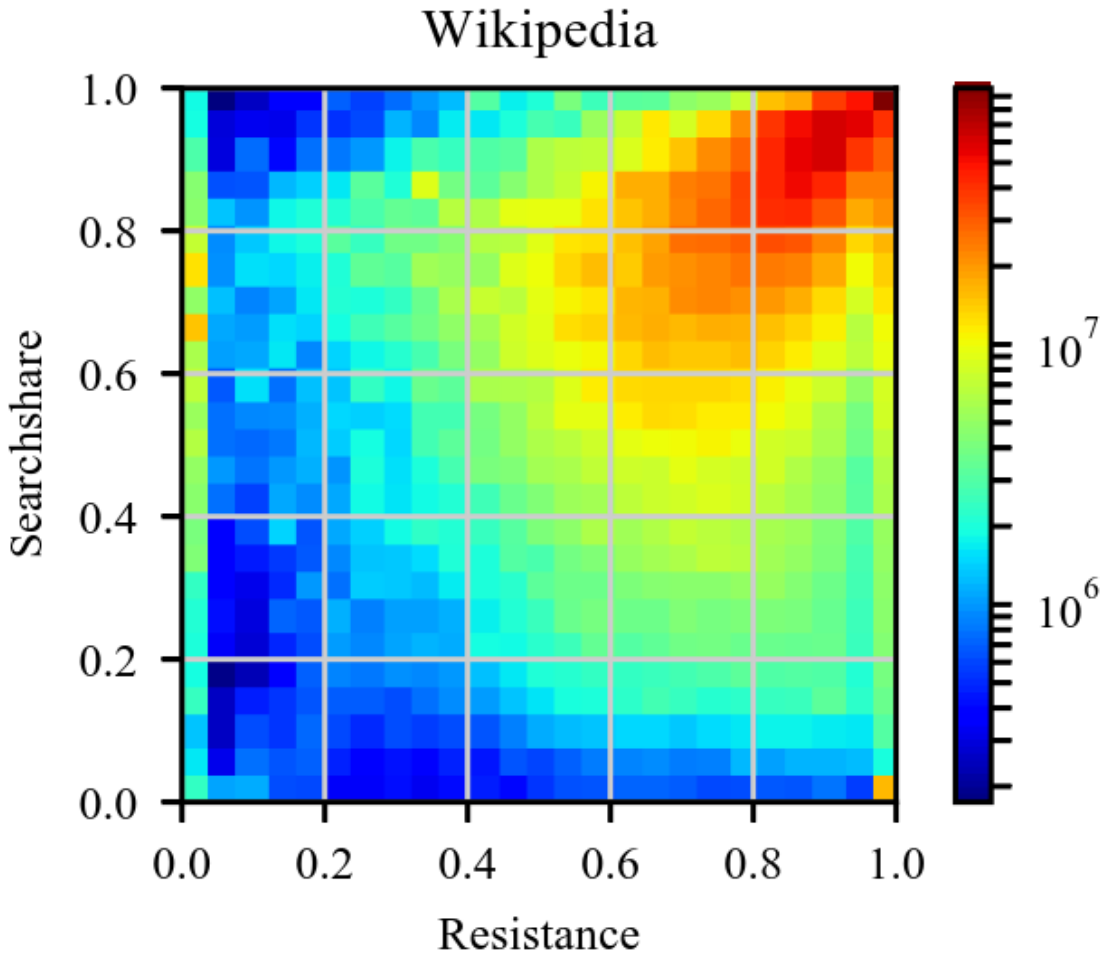
Search-exit – articles accessed by *search* that are *unable* to relay traffic

Search-relay – articles accessed by *search* that are *able* to relay traffic

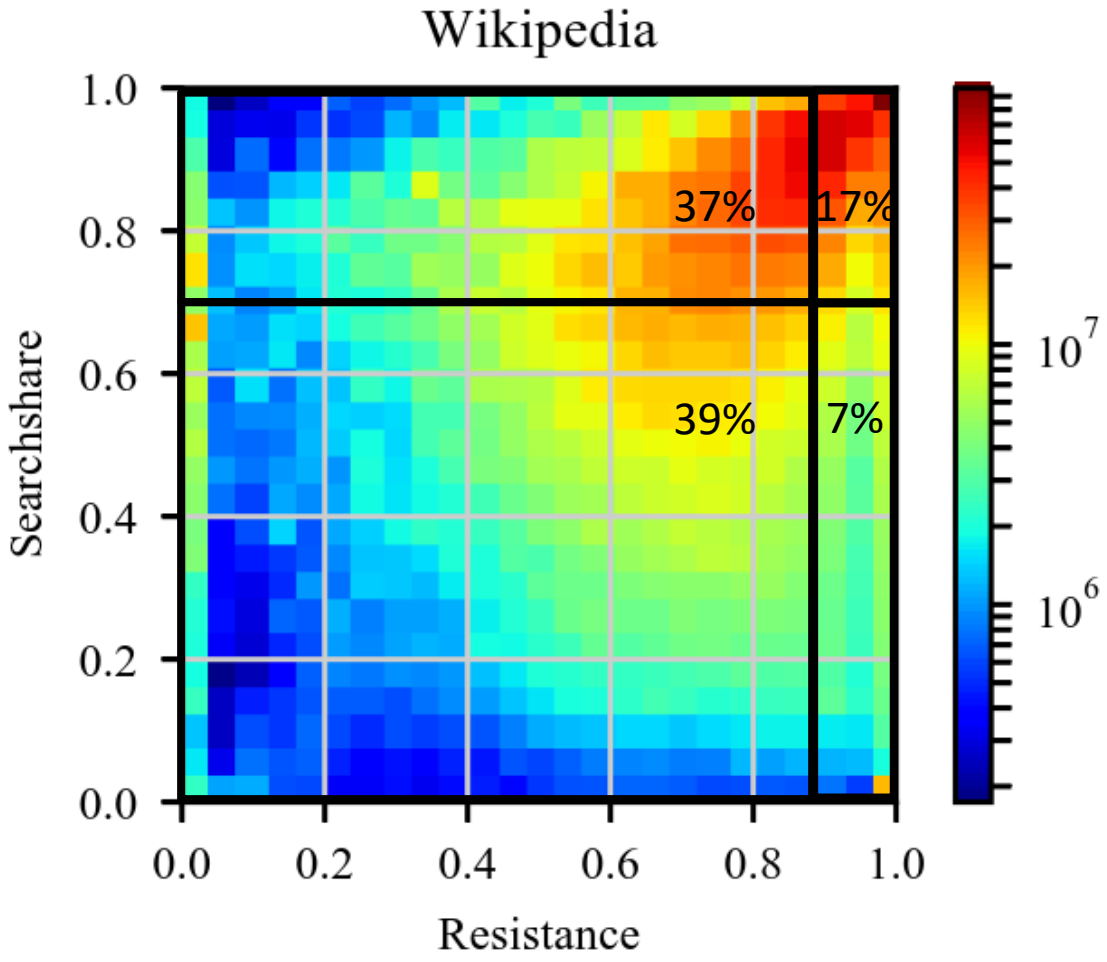
Navigation-relay – articles accessed by *navigation* that are *able* to relay traffic

Navigation-exit – articles accessed by *navigation* that are *unable* to relay traffic

Number of views



Number of views



Which article properties are characteristic for specific article roles?

Article features

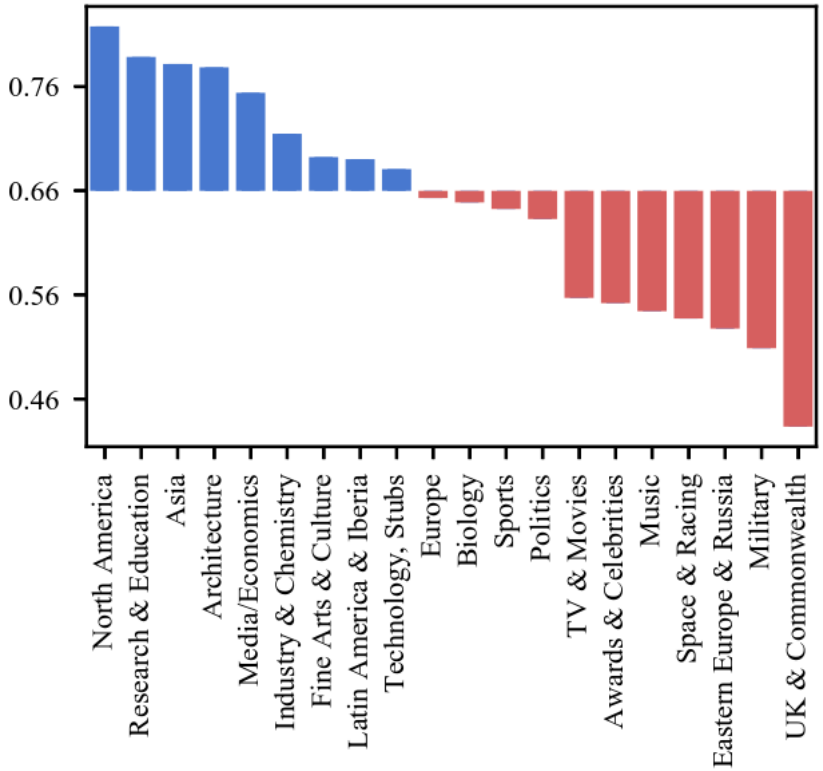
- Network features
  - In, out, and total degree
  - K-Core
- Content and edit features
  - Editors, revisions, age
  - Size(kB), sections, tables
- Topics
  - 20 topics discovered using LDA
  - Topics are labeled by five researchers using the top words and articles for each topic

Article features

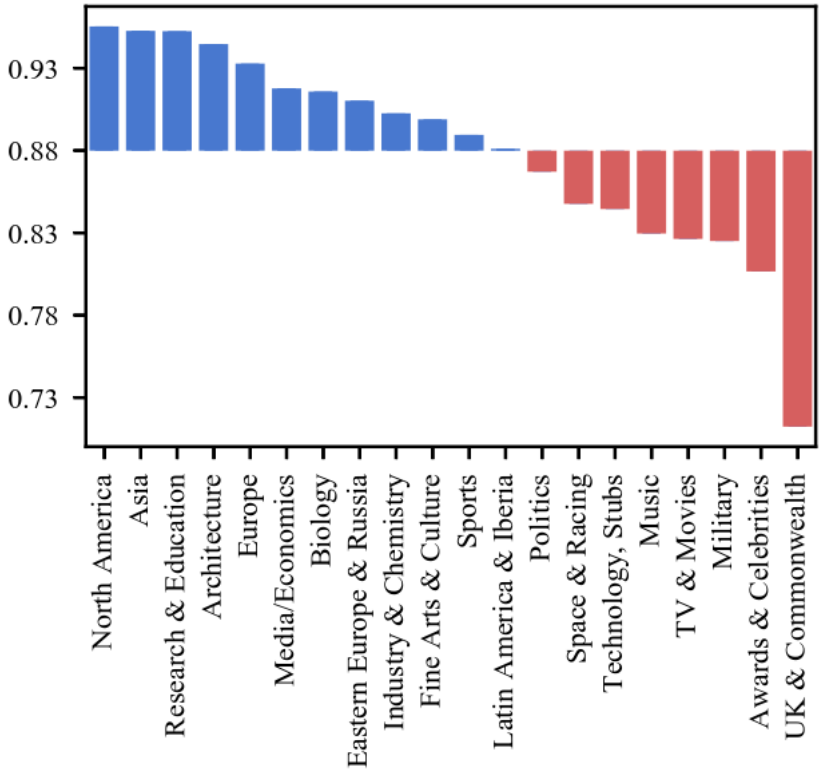
- Network features
  - In, out, and total degree
  - K-Core
- Content and edit features
  - Editors, revisions, age
  - Size(kB), sections, tables
- Topics
  - 20 topics discovered using LDA
  - Topics are labeled by five researchers using the top words and articles for each topic

Are topics accessed differently?

Searchshare



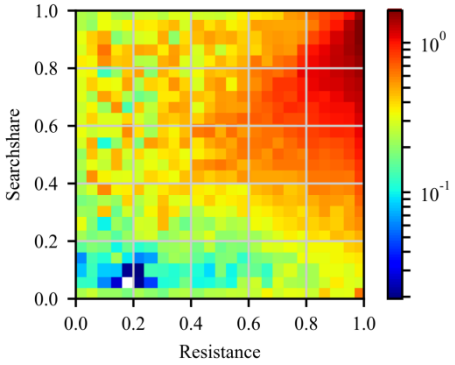
Resistance



Step 2

Topic-specific heatmaps

Topic „X“

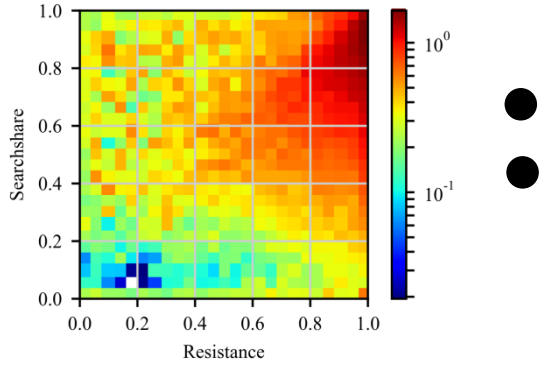


- Create a topic-specific heatmap

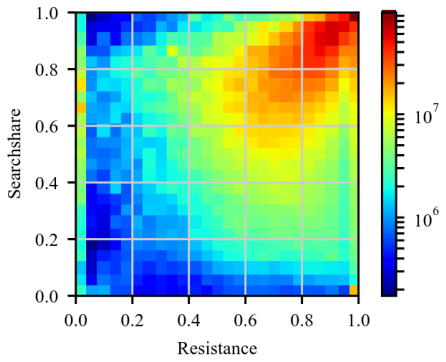
Step 2

Topic-specific heatmaps

Topic „X“



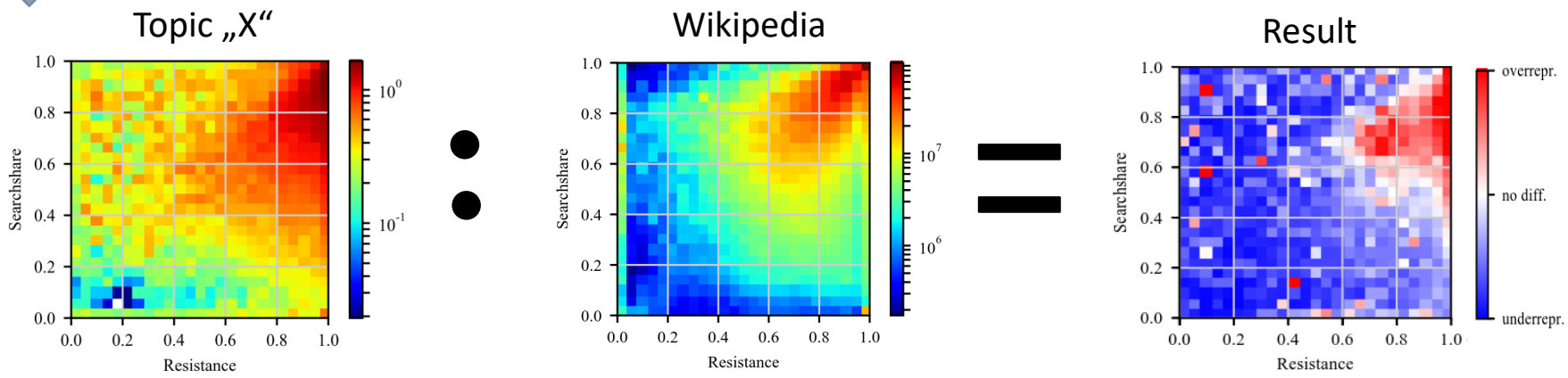
Wikipedia



- Create a topic-specific heatmap
- Bin-wise division of topic-specific and Wikipedia heatmaps

Step 2

Topic-specific heatmaps

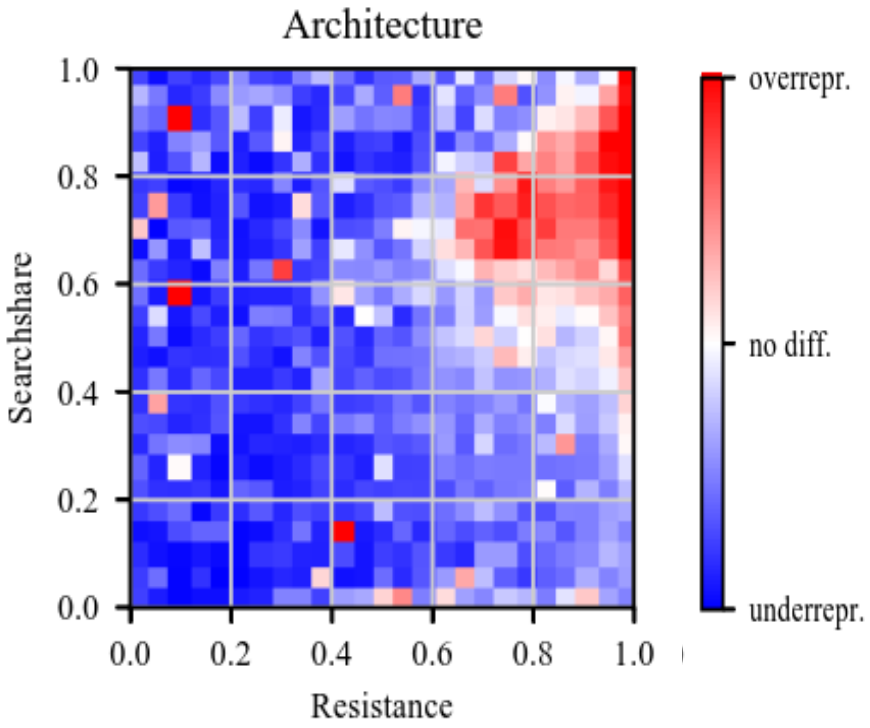


- Create a topic-specific heatmap
- Bin-wise division of topic-specific and Wikipedia heatmaps
- Result: heatmap showing the deviation of the topic-specific from the general access behavior

What is the role of a topic for the traffic?

gesis Leibniz Institute for the Social Sciences

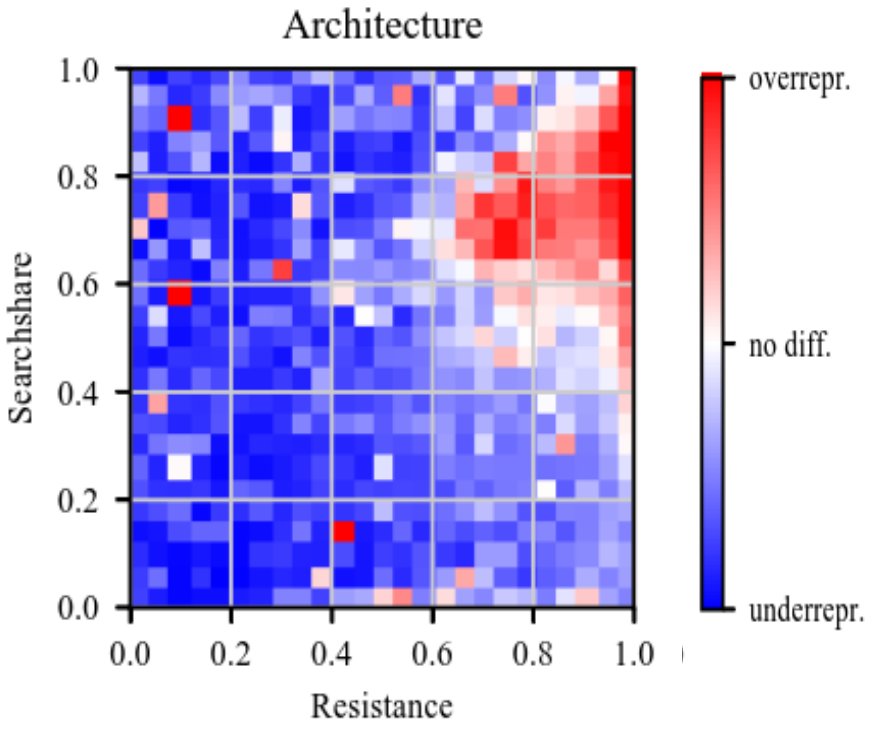
What is the role of a topic for the traffic?

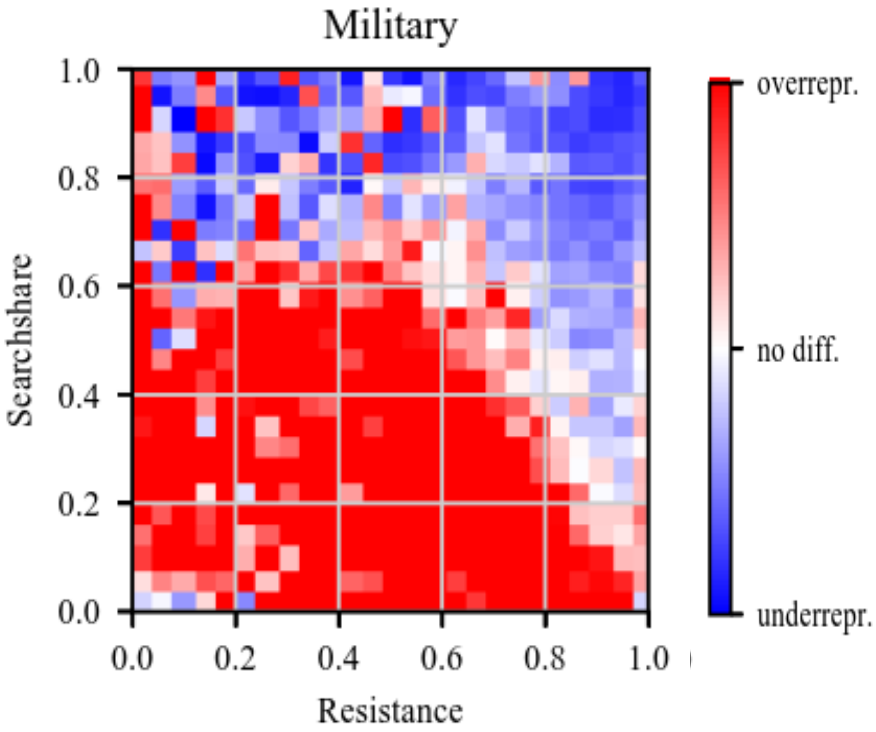


Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

What is the role of a topic for the traffic?

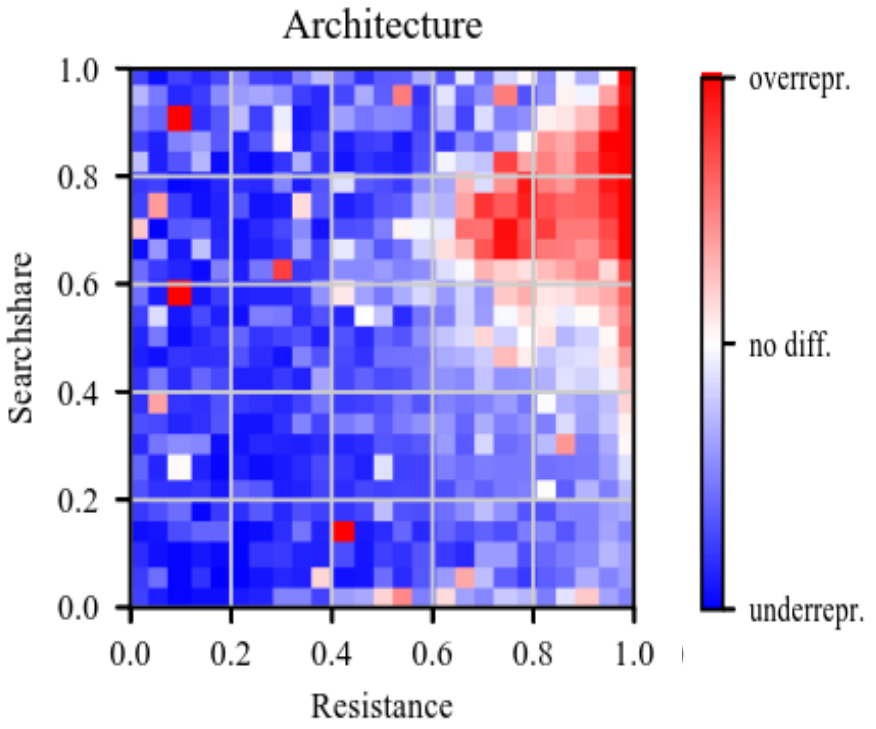


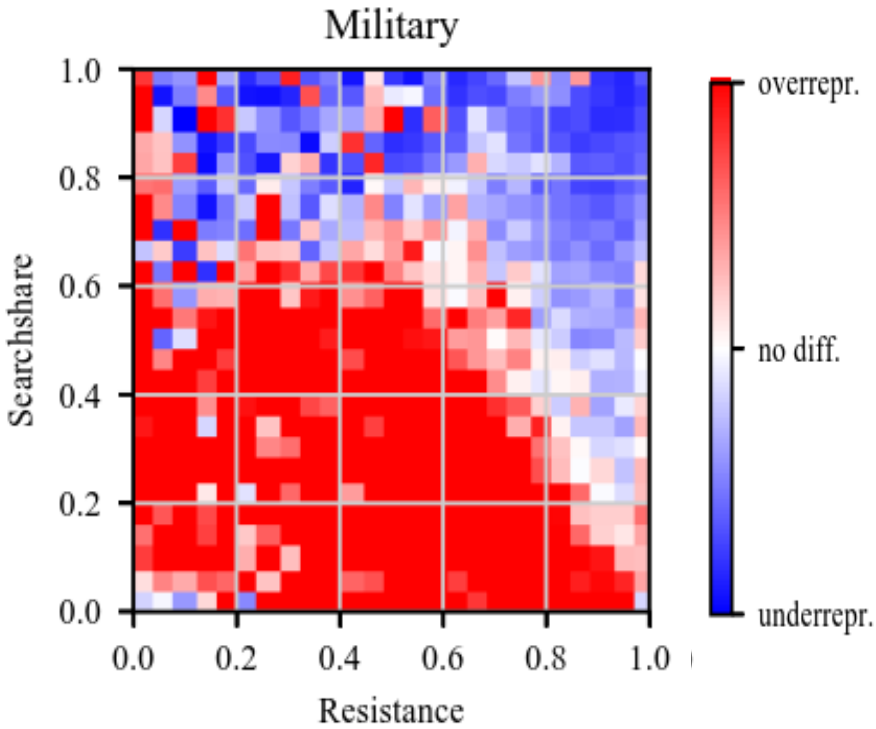


Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

What is the role of a topic for the traffic?





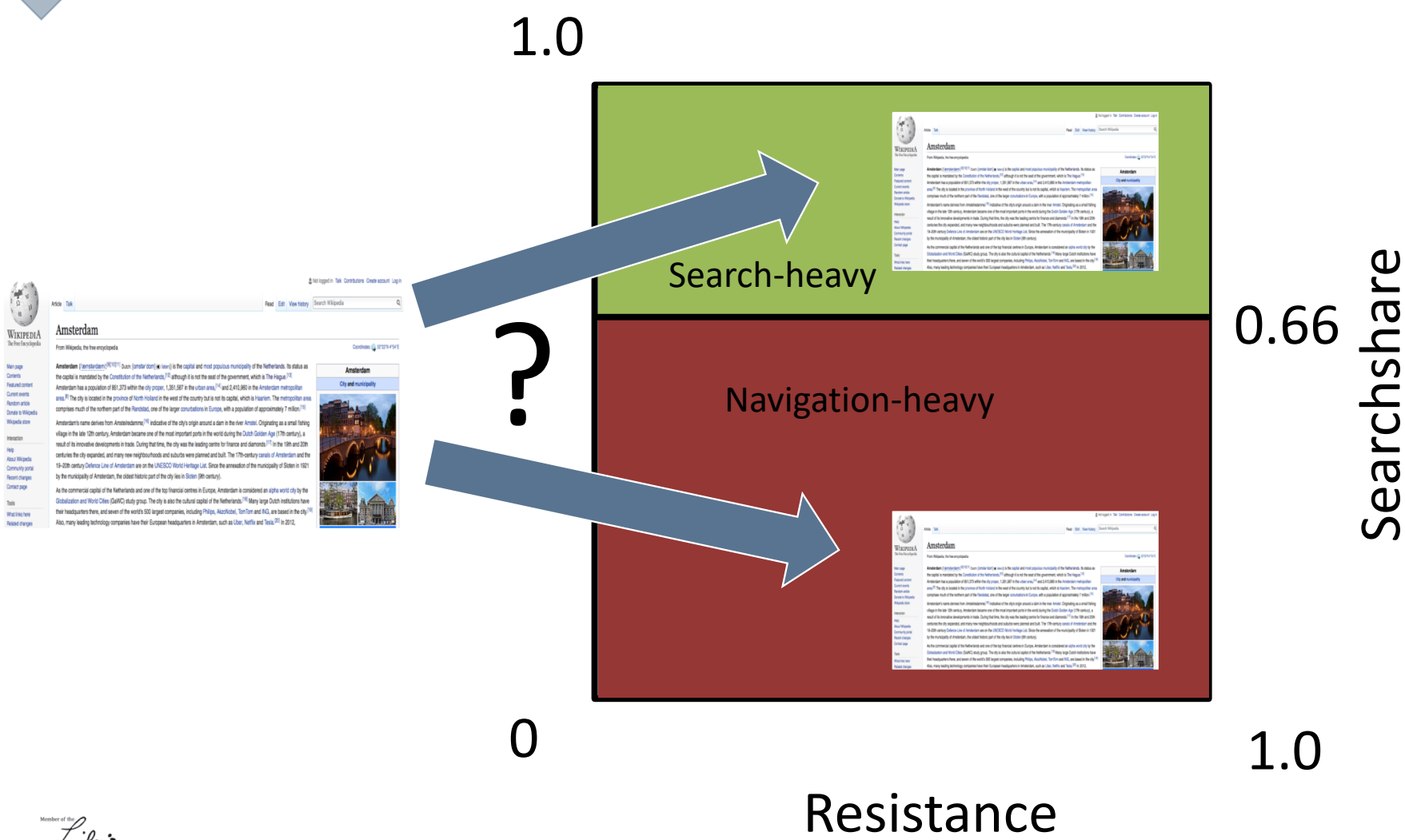
Query for „Architecture“, Click through „Military“

Which article properties are indicative of specific information access behavior?

gesis Leibniz Institute for the Social Sciences

Step 3

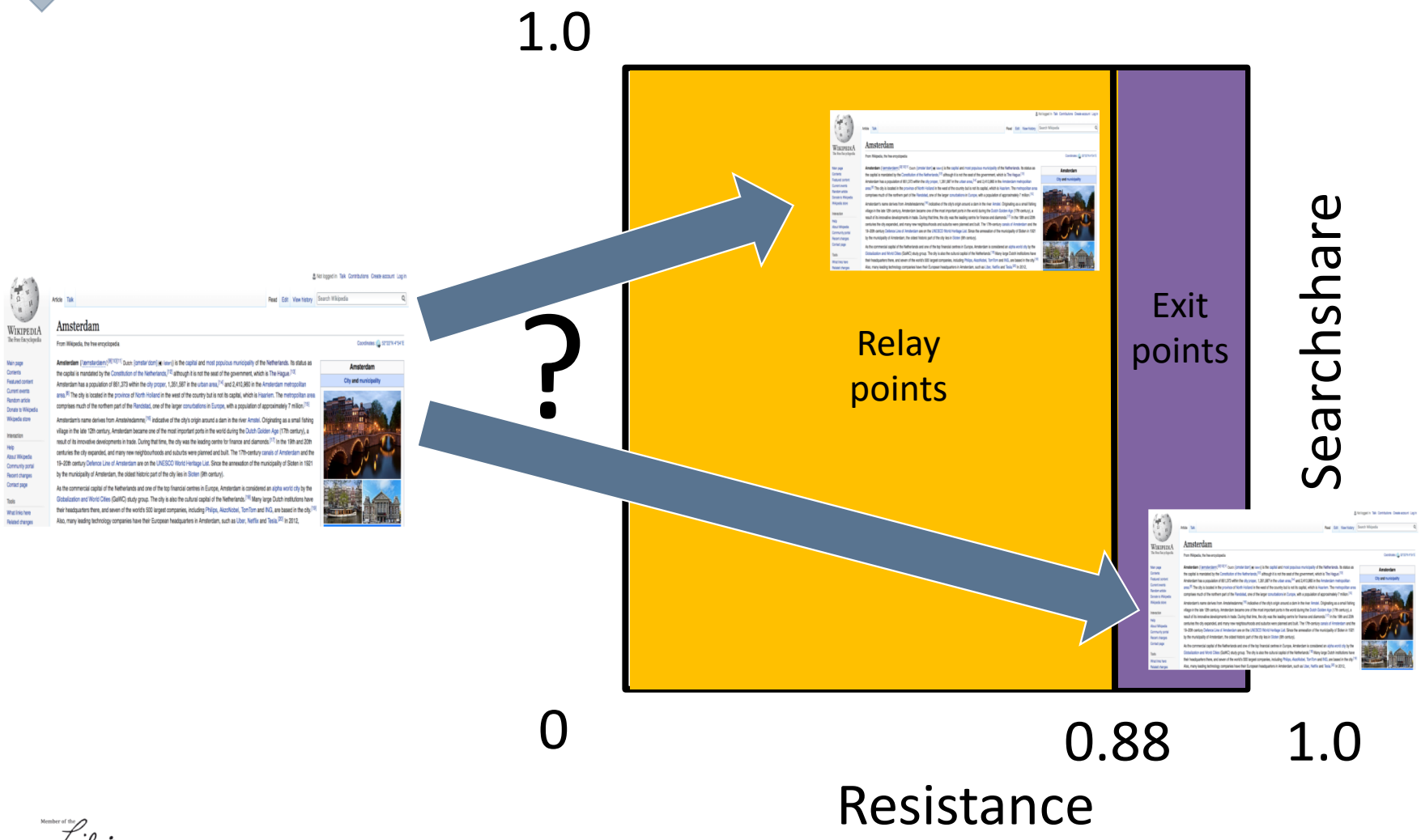
Modeling access behavior



gesis Leibniz Institute for the Social Sciences

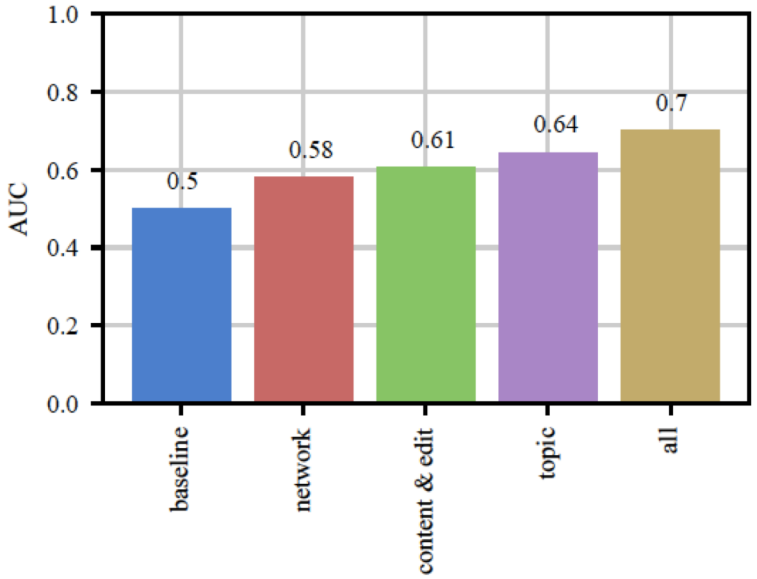
Step 3

Modeling access behavior



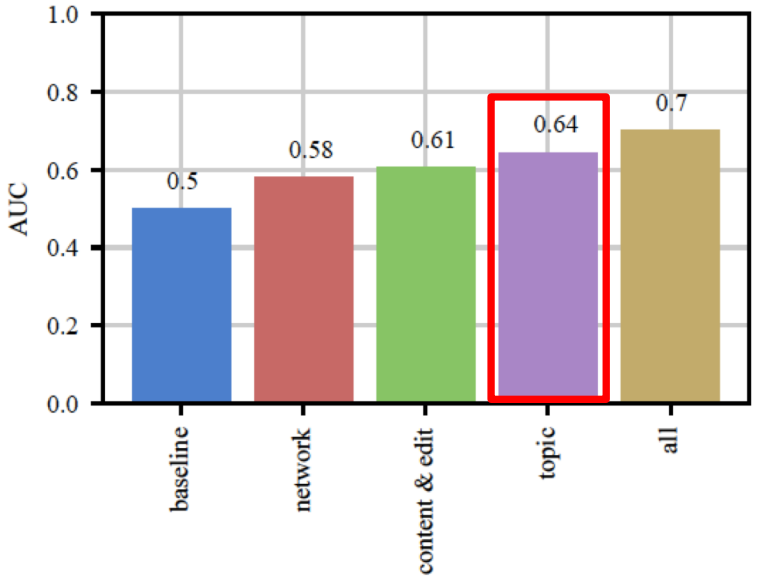
Member of the Leibniz Association

Results



Searchshare

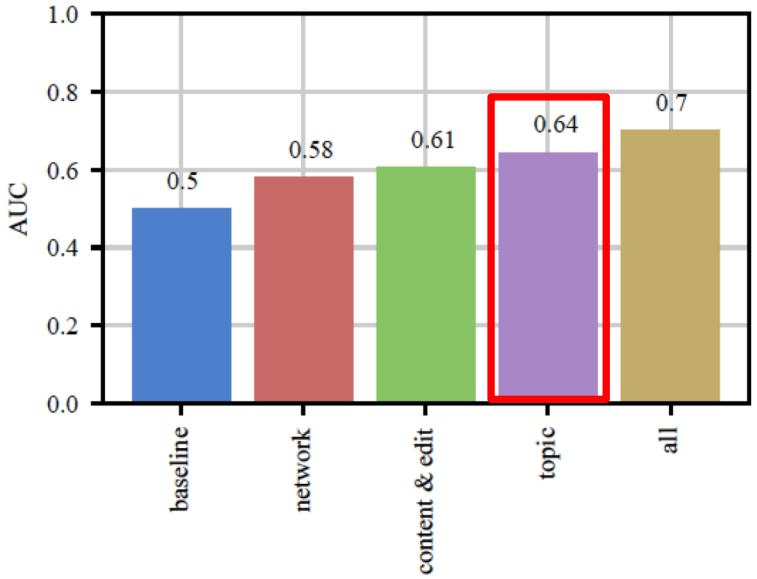
Results



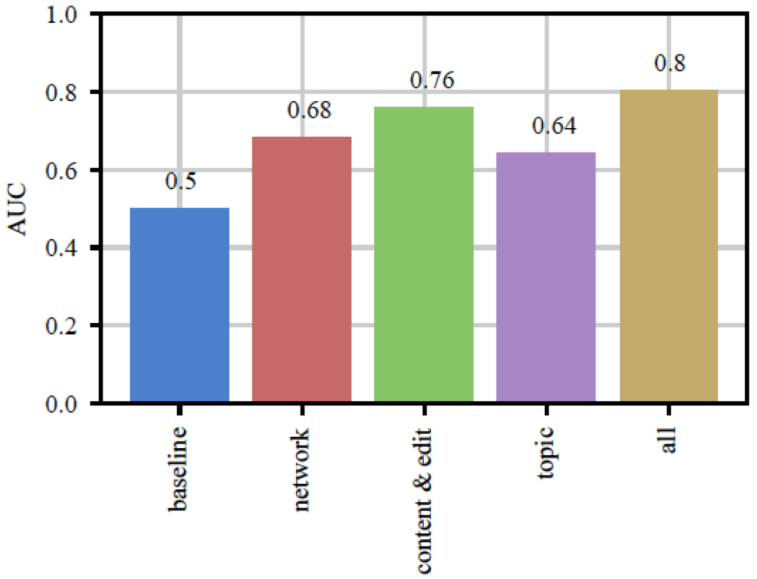
Searchshare

1. Article topic is indicative of searchshare

Results



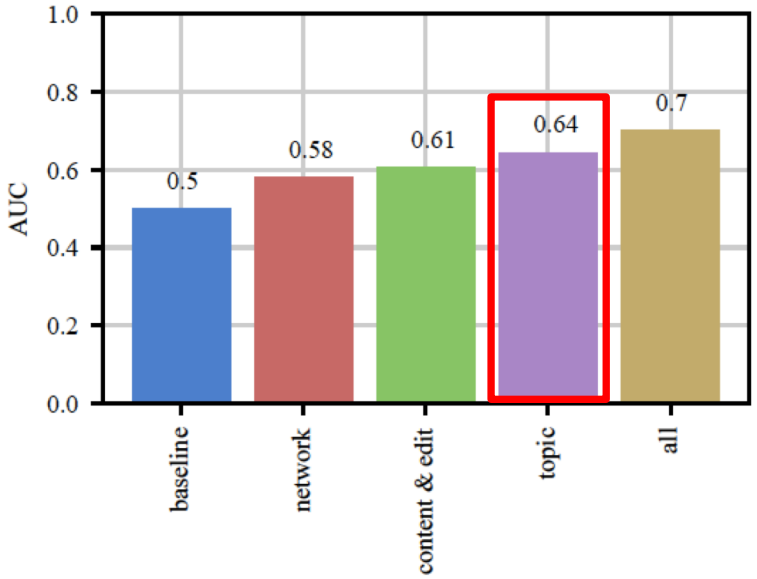
Searchshare



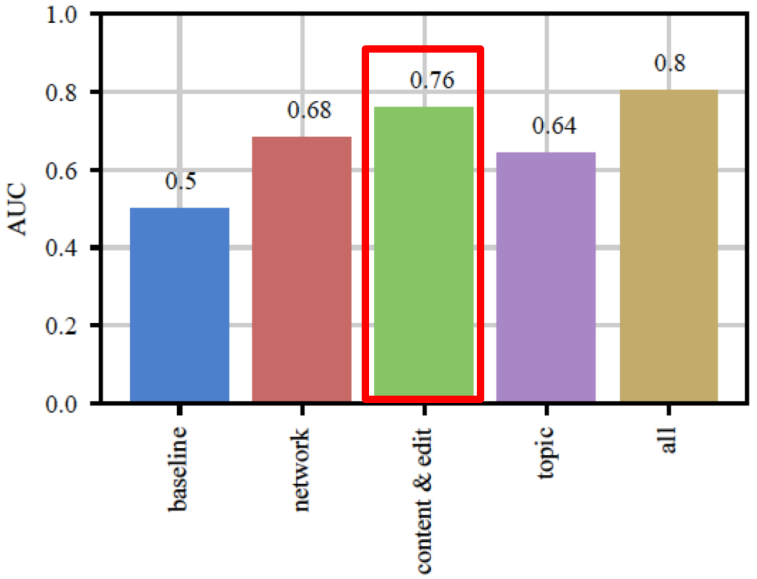
Resistance

1. Article topic is indicative of searchshare

Results



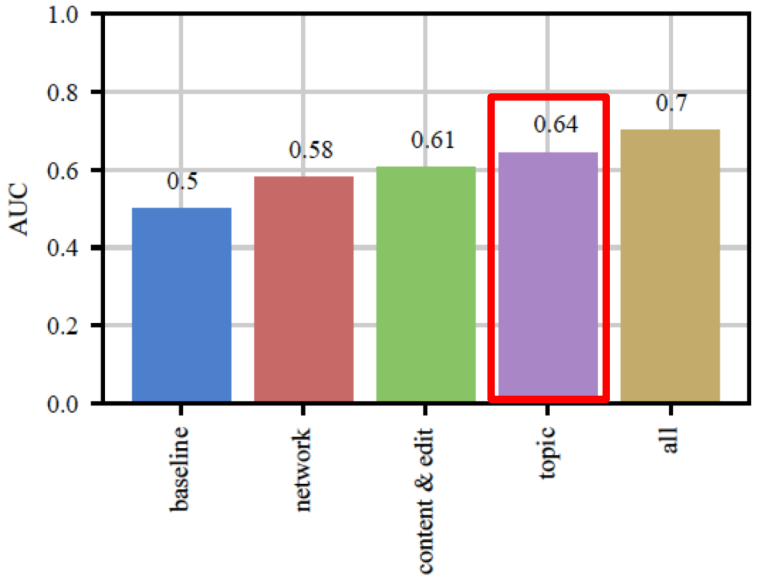
Searchshare



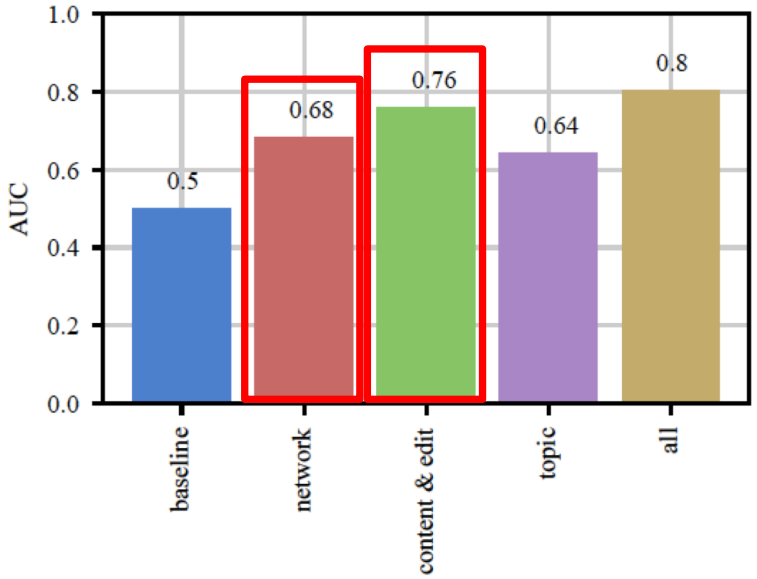
Resistance

1. Article topic is indicative of searchshare
2. Content and edit features are indicative of resistance

Results



Searchshare



Resistance

1. Article topic is indicative of searchshare
2. Content and edit features are indicative of resistance
3. Article position in the network is indicative of resistance

Implications

- Identification of search- and navigation-exit points
- Design of interfaces and new features
  - Topic-specific layouts to support search- and navigation-heavy reading behavior
  - Introduction of links to the page preview on search-heavy articles
- Detection of online misbehavior and vandalism
  - Unexpected changes in the access patters on articles

Summary and Conclusion

- Studied the interplay between search and navigation on Wikipedia
- Wikipedia is good at converting search to navigation traffic
- Search and navigation are used to access different articles
- Both information access forms are crucial as they complement each other

Questions?

Thank you!

Contact:

E-Mail: dimitar.dimitrov@gesis.org
Twitter: @trovdimi