

# 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.

# Information Access on Wikipedia

- Search



london|

london weather  
london time  
london city airport  
london stansted  
london eye  
london tube map  
london heathrow  
london bridge  
london underground

WIKIPEDIA  
The Free Encyclopedia

london

From Wikipedia, the free encyclopedia

This article is about the capital city. For the region of England, see Greater London. For the historic city and financial district within London, see City of London. For other uses, see London (disambiguation).

London is London is the capital and most populous city of England and the United Kingdom.<sup>[1]</sup> Standing on the River Thames in the south east of the island of Great Britain, London has been a major settlement for two millennia. It was founded by the Romans, who named it Londinium.<sup>[2]</sup> London's ancient core, the City of London, largely retains its 1.18 square-mile (3.0 km<sup>2</sup>) medieval boundaries. Since at least the 18th century, "London" has also referred to the metropolis around the core, historically split between Middlesex, Essex, Surrey, Kent, and West-Middlesex,<sup>[3][4]</sup> which today largely makes up Greater London,<sup>[5][6][7]</sup> governed by the Mayor of London and the London Assembly.<sup>[8][9]</sup>

London is a leading global city<sup>[10]</sup> in the arts, commerce, education, entertainment, fashion, finance, healthcare, media, professional services, research and development, tourism, and transportation.<sup>[11][12]</sup> It is crowned as the world's largest financial centre<sup>[13][14]</sup> and has the fifth or sixth-largest metropolitan area GDP in the world.<sup>[15][16]</sup> London is a world cultural capital<sup>[17][18]</sup> and is the world's most visited city as measured by international arrivals<sup>[19]</sup> and has the world's largest city airport system measured by passenger traffic.<sup>[20]</sup> London is the world's leading economic destination<sup>[21][22]</sup> leading free international arrivals<sup>[23]</sup> and has the highest number of international<sup>[24]</sup> FDI flows in any other city. London's universities form the largest concentration of higher education institutions in Europe.<sup>[25]</sup> In 2012, London became the first city to have hosted the modern Summer Olympic Games three times.<sup>[26]</sup>

London has a diverse range of people and cultures, and more than 300 languages are spoken in the region.<sup>[27]</sup> Its estimated mid-2010s municipal population (comprising its Greater London area) was 8,873,737,<sup>[28]</sup> the largest of any city in the European Union,<sup>[29]</sup> and accounting for 12.1% of the UK population.<sup>[30]</sup> London's urban area is the second-most populous in the EU, after Paris, with 8,397,426 inhabitants at the 2011 census.<sup>[31]</sup> The city's metropolitan area is one of the most populous in Europe with 13,875,757 inhabitants,<sup>[32][33]</sup> while the Greater London Authority defines the population of the city-region (covering a large part of the south-east) as 21.7 million.<sup>[34]</sup> The city-region therefore has a large land area and population to that of the New York metropolitan area. London was the world's most populous city from around 1850 to 1925.<sup>[35]</sup>

London contains four World Heritage Sites: the Tower of London; New Garden; the site comprising the Palace of Westminster, Westminster Abbey, and St Margaret's Church; and the historic settlement of Greenwich in which the Royal Observatory, Greenwich marks the Prime Meridian, 0° longitude, and GMT.<sup>[36]</sup> Other famous landmarks include Buckingham Palace, the London Eye, Piccadilly Circus, St Paul's Cathedral, and the Shard.

Greater London Authority  
City of London  
Greater London

# Information Access on Wikipedia

- Search

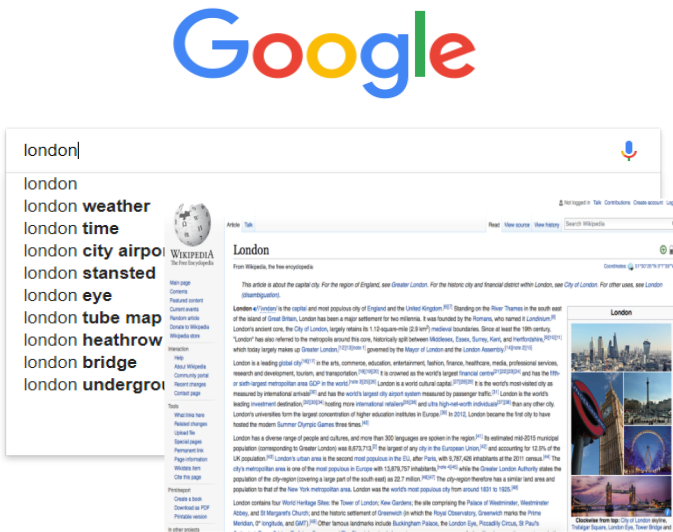
The image shows a Google search interface with the query 'london|'. Below the search bar, a list of suggestions includes 'london weather', 'london time', 'london city airport', 'london stansted', 'london eye', 'london tube map', 'london heathrow', 'london bridge', and 'london underground'. To the right, a Wikipedia article snippet for 'London' is visible, starting with 'From Wikipedia, the free encyclopedia'. The snippet describes London as the capital and most populous city of England and the United Kingdom, located on the River Thames in the south-east of Great Britain.

- Navigation



# Information Access on Wikipedia

- Search



~70% of all visits

- Navigation



~30% of all visits





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



WIKIPEDIA  
The Free Encyclopedia

- 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

Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Article [Talk](#)

[Read](#) [Edit](#) [View history](#)

## Amsterdam

From Wikipedia, the free encyclopedia

Coordinates: 52°22′N 4°54′E﻿ / ﻿

**Amsterdam** (/ˈæmstərdæm/<sup>[</sup>ɡɪˈlɔːlɪnɪ]<sup>[</sup>10]]<sup>[</sup>11<sup>]</sup>) Dutch: [ɑmˈstərˈdɑm] ⓘ listen) is the capital and most populous municipality of the Netherlands. Its status as the capital is mandated by the [Constitution of the Netherlands](#),<sup>[12]</sup> although it is not the seat of the government, which is The Hague.<sup>[13]</sup> Amsterdam has a population of 851,373 within the [city proper](#), 1,351,587 in the [urban area](#),<sup>[14]</sup> and 2,410,960 in the [Amsterdam metropolitan area](#).<sup>[8]</sup> 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.<sup>[15]</sup>

Amsterdam's name derives from *Amstelredamme*,<sup>[16]</sup> 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.<sup>[17]</sup> 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.<sup>[18]</sup> 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.<sup>[19]</sup> Also, many leading technology companies have their European headquarters in Amsterdam, such as [Uber](#), [Netflix](#) and [Tesla](#).<sup>[20]</sup> In 2012,

### Amsterdam

#### City and municipality





# 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?

# Traffic features

$$searchshare(a) = \frac{in_{se}(a)}{\underbrace{in_{se}(a) + in_{nav}(a)}_{\text{Total pageviews}}}$$

← Incoming search traffic  
← Incoming navigation traffic

**Searchshare** – the amount of views an article received by search



# Traffic features

$$searchshare(a) = \frac{in_{se}(a)}{\underbrace{in_{se}(a) + in_{nav}(a)}_{\text{Total pageviews}}}$$

← Incoming search traffic  
← Incoming navigation traffic

**Searchshare** – the amount of views an article received by search

$$resistance(a) = 1 - \frac{out_{nav}(a)}{\underbrace{in_{se}(a) + in_{nav}(a)}_{\text{Total pageviews}}}$$

← Outgoing navigation traffic

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





# Traffic features

$$searchshare(a) = \frac{in_{se}(a)}{\underbrace{in_{se}(a) + in_{nav}(a)}_{\text{Total pageviews}}}$$

← Incoming search traffic  
← Incoming navigation traffic

**Searchshare** – the amount of views an article received by search

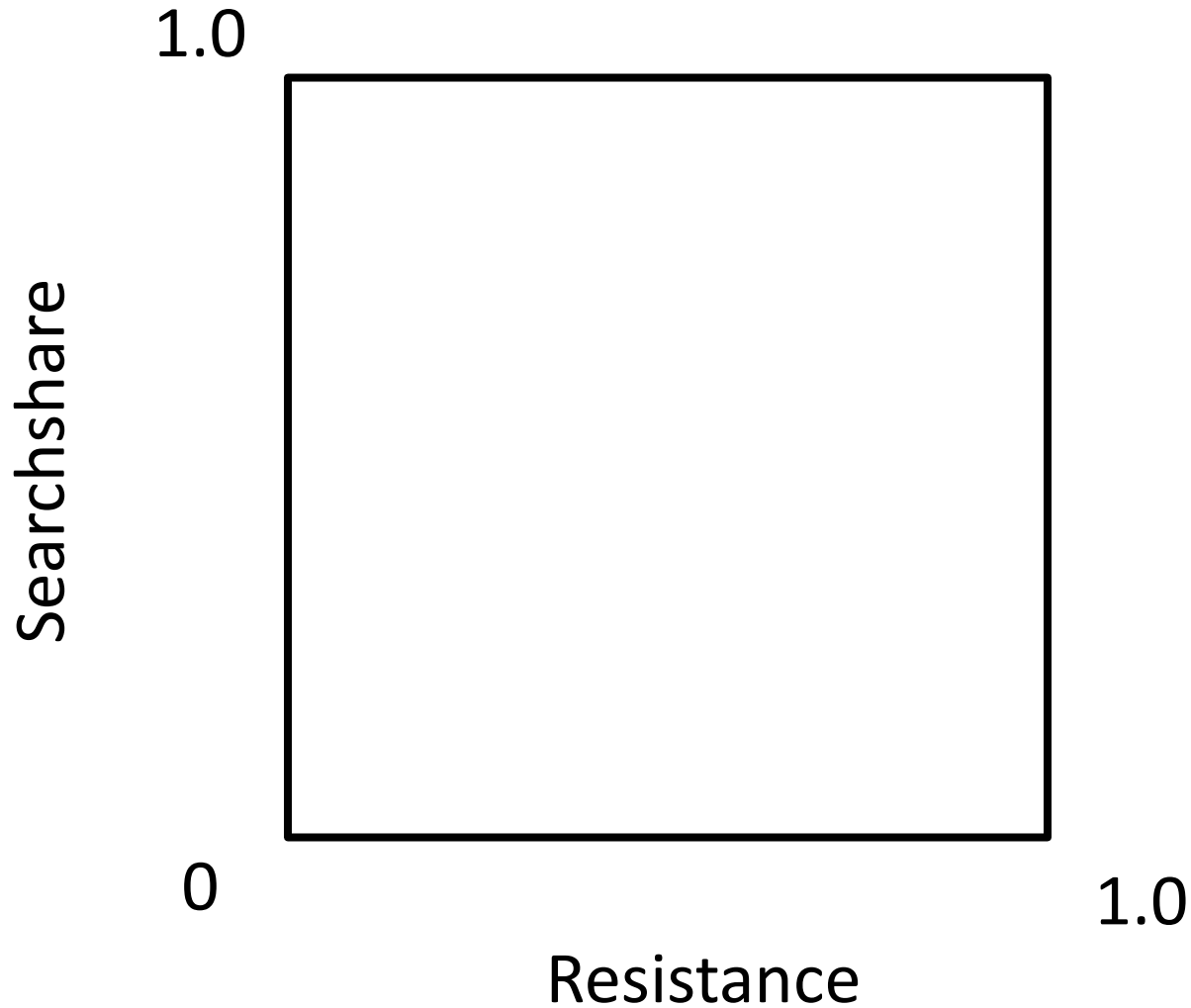
$$resistance(a) = 1 - \frac{out_{nav}(a)}{\underbrace{in_{se}(a) + in_{nav}(a)}_{\text{Total pageviews}}}$$

← Outgoing navigation traffic

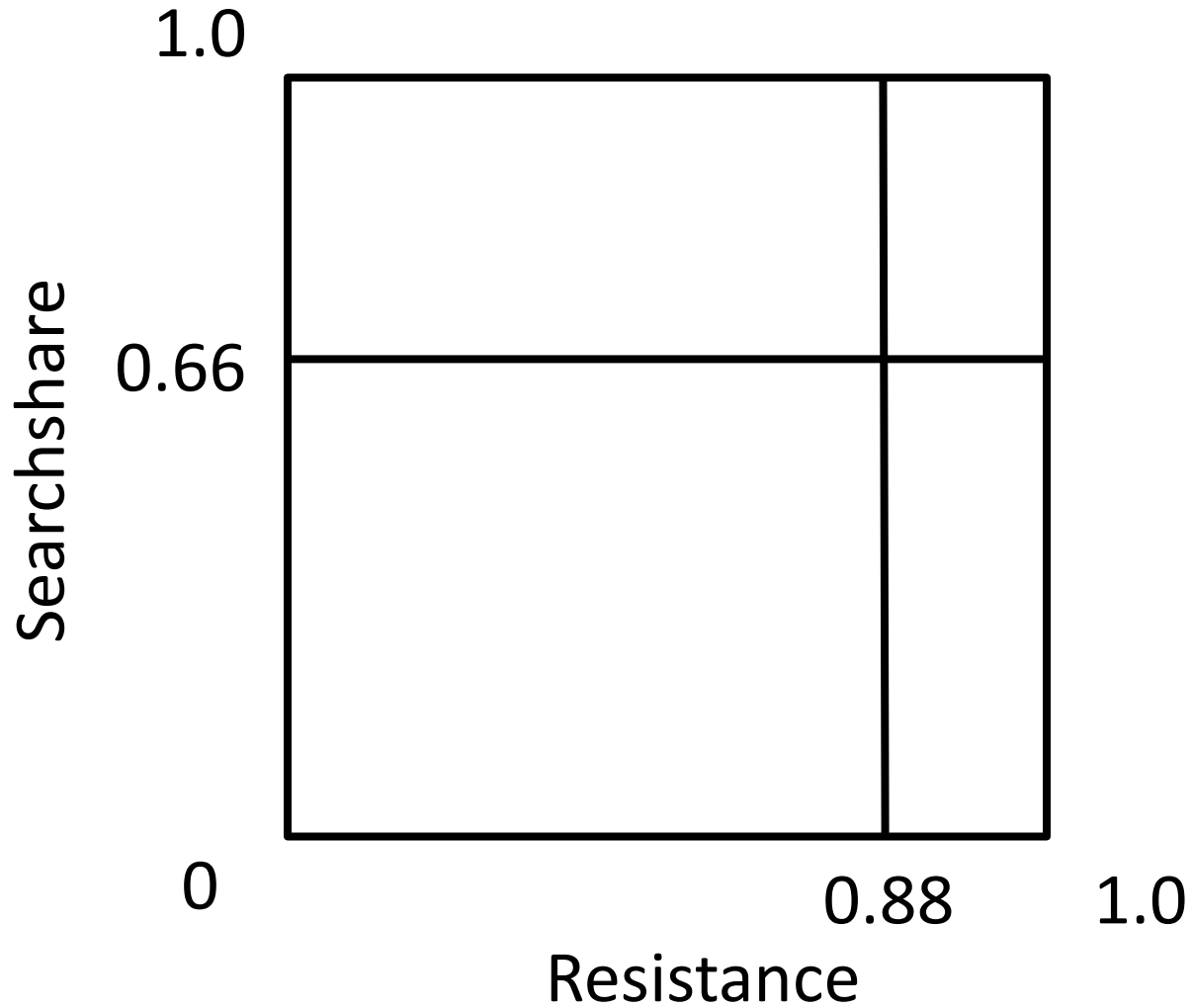
**Resistance** – the ability of an article to relay traffic to other Wikipedia articles

	<i>Searchshare</i>	<i>Resistance</i>
Mean	0.66	0.88

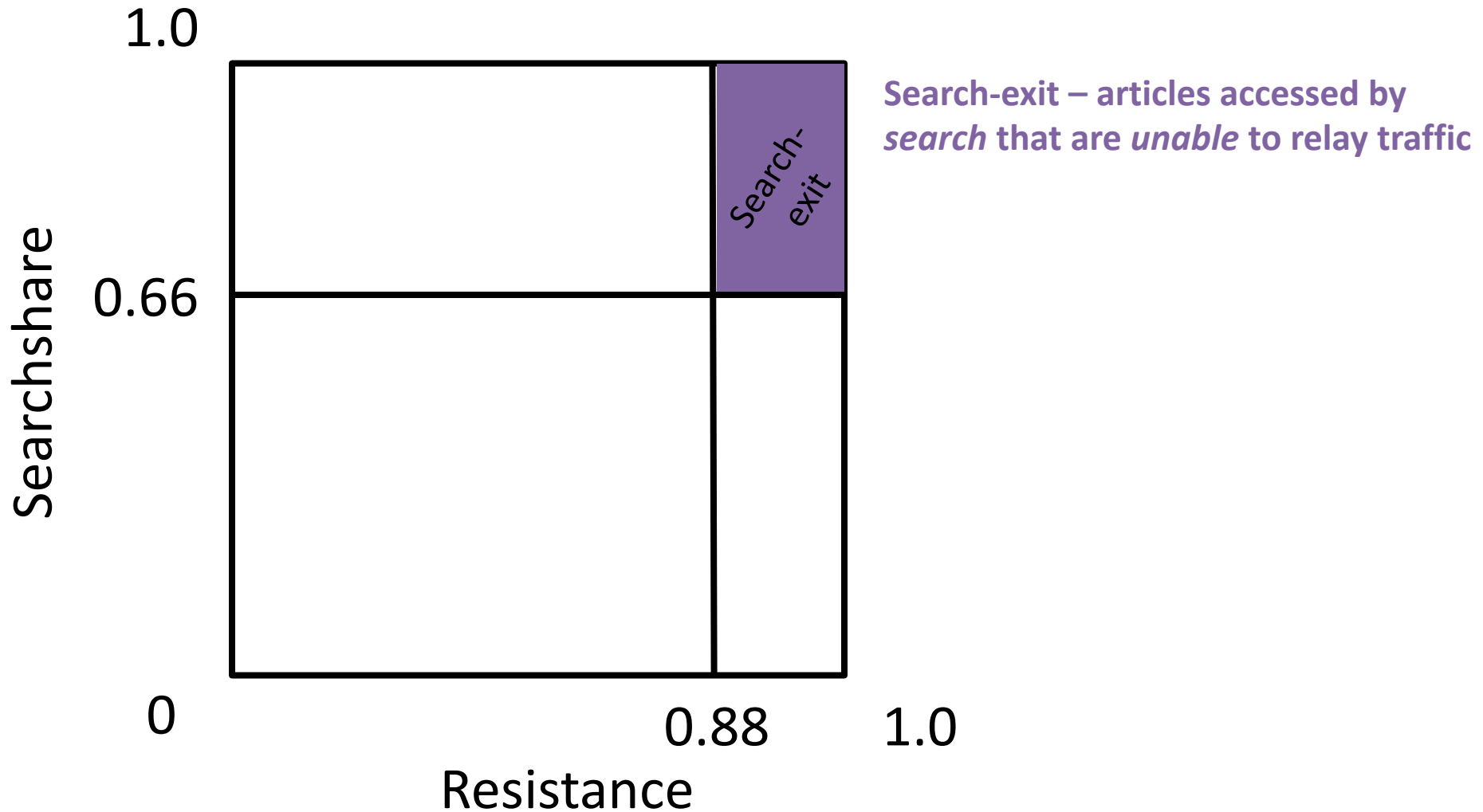
# Article roles



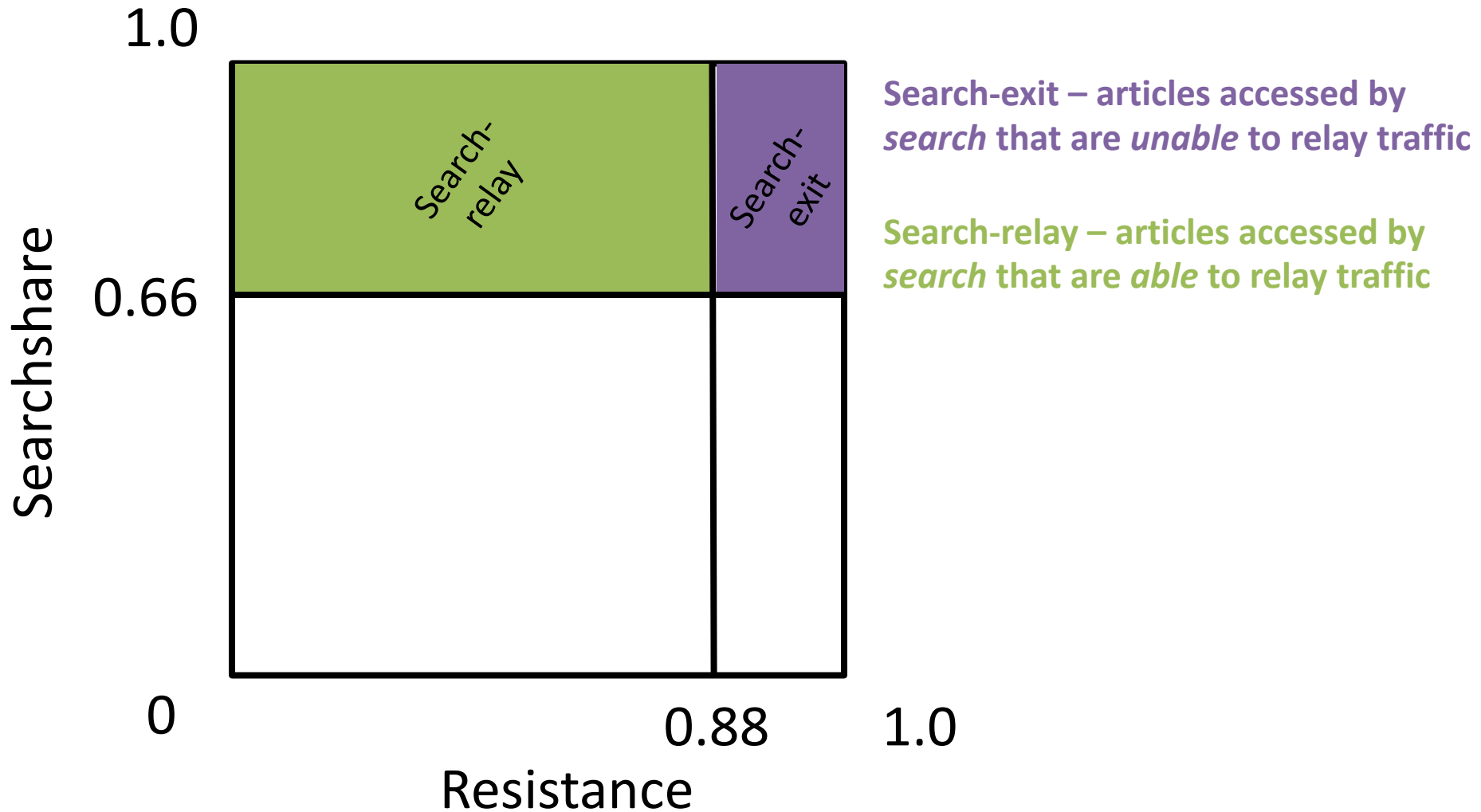
# Article roles



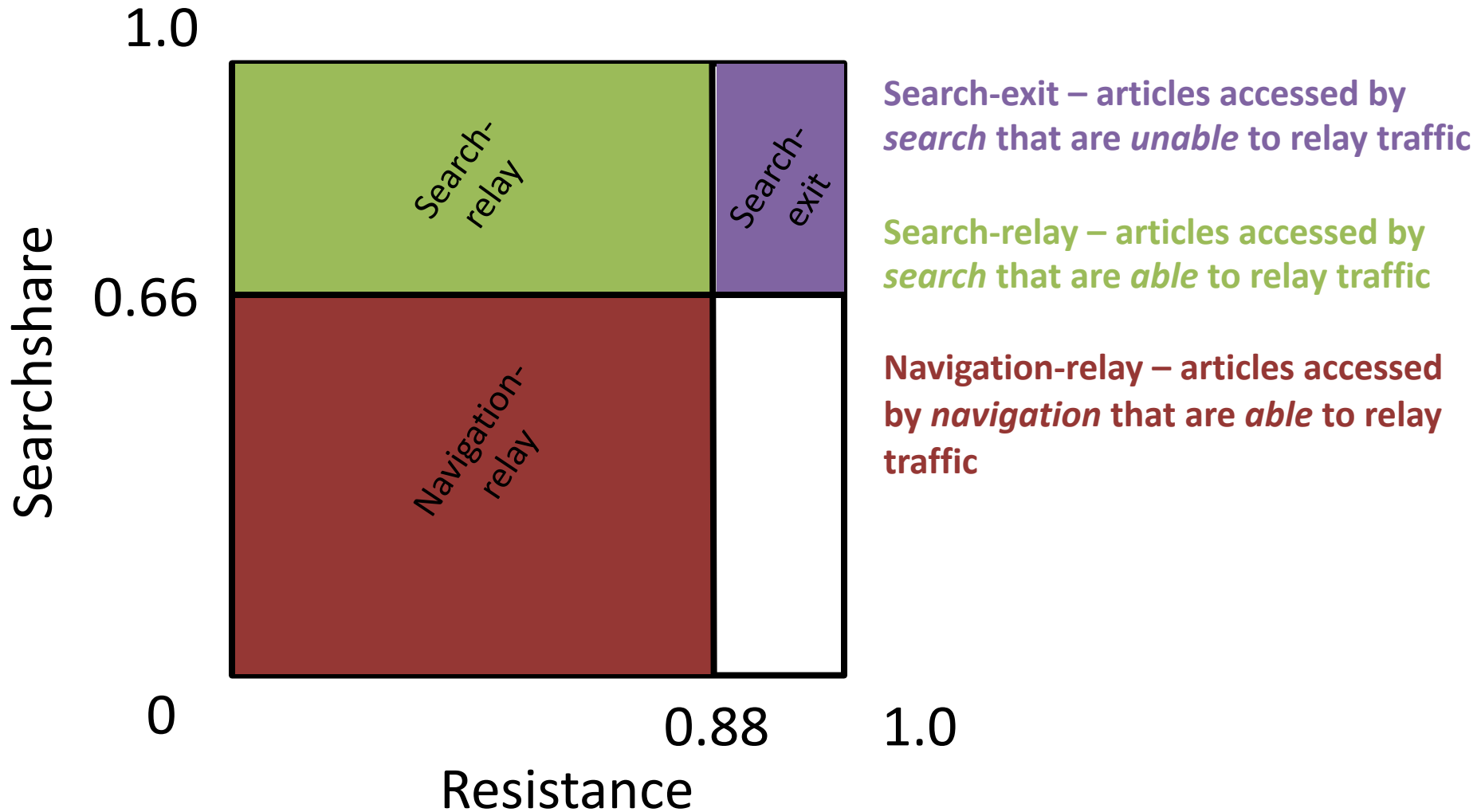
# Article roles



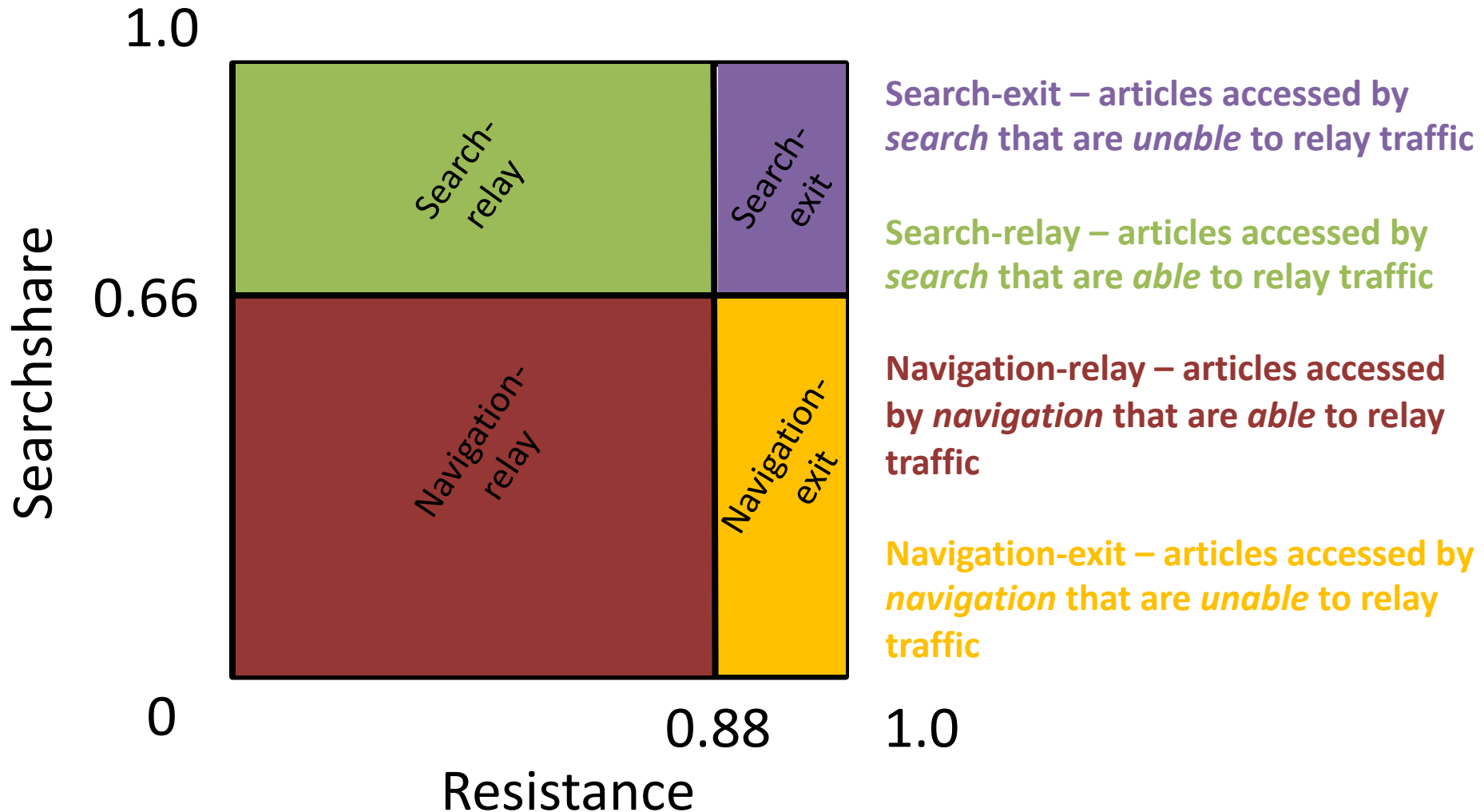
# Article roles



# Article roles

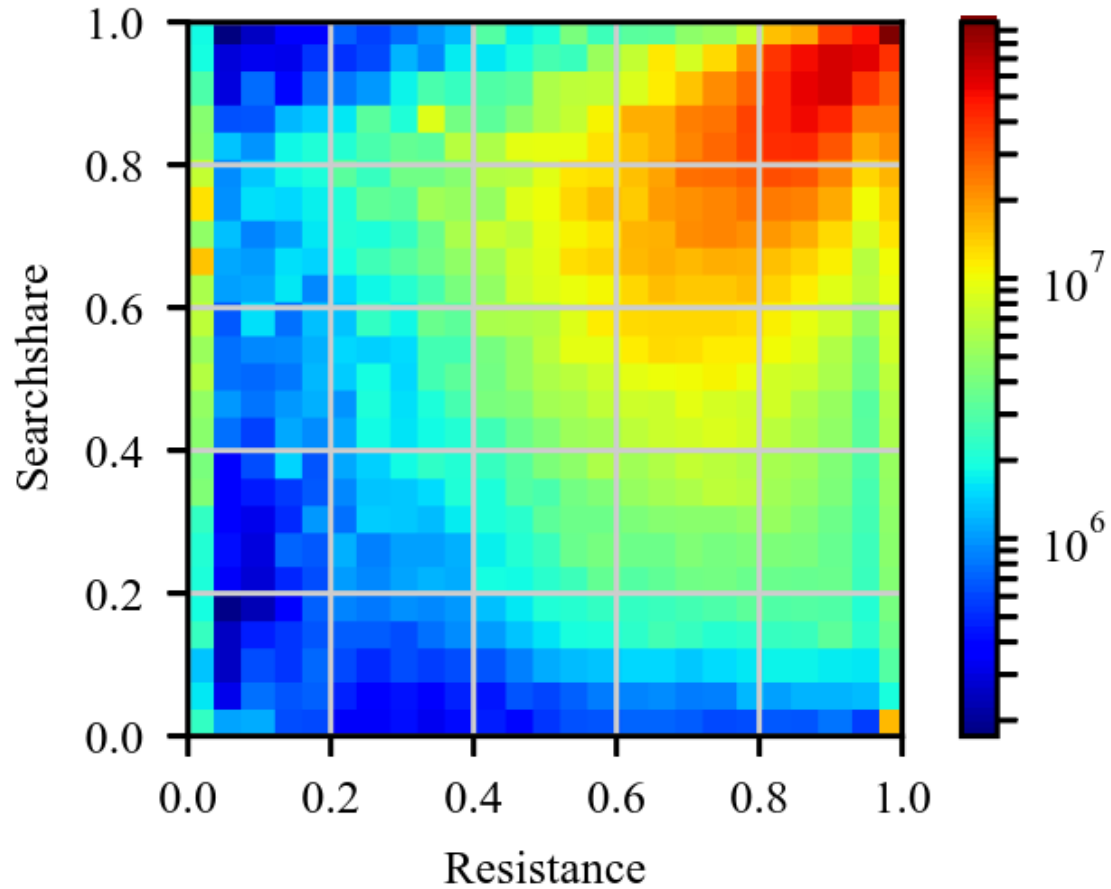


# Article roles



# Number of views

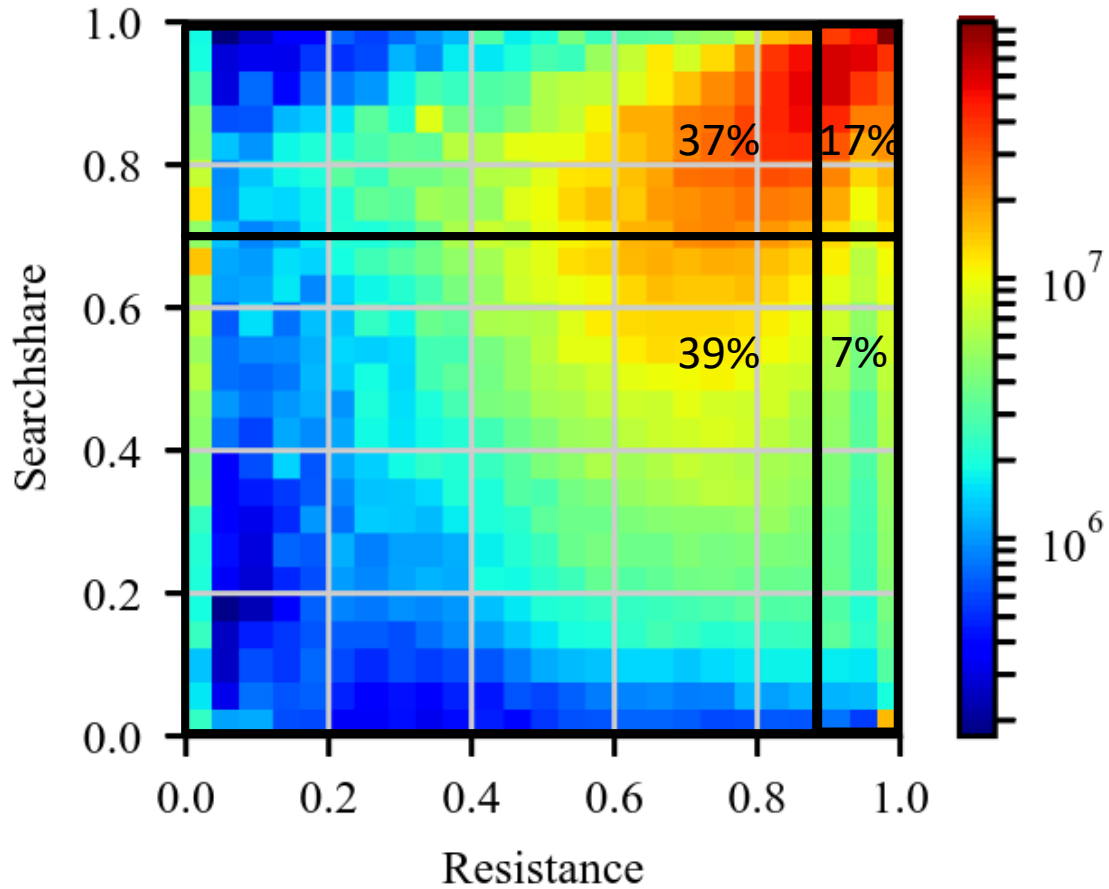
Wikipedia





# Number of views

Wikipedia



# Which article properties are characteristic for specific article roles?

# Article features

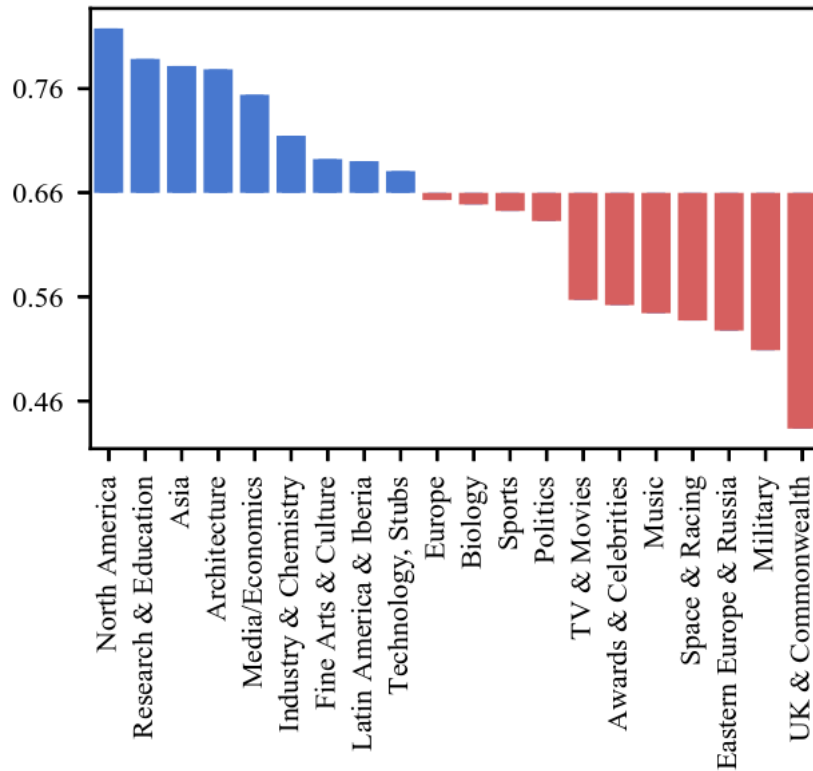
- Network features
  - ▶ In, out, and total degree
  - ▶ K-Core
- Content and edit features
  - ▶ Editors, revisions, age
  - ▶ Size(kB), sections, tables
- Topics
  - ▶ 20 topics discovered using LDA
  - ▶ Topics are labeled by five researchers using the top words and articles for each topic

# Article features

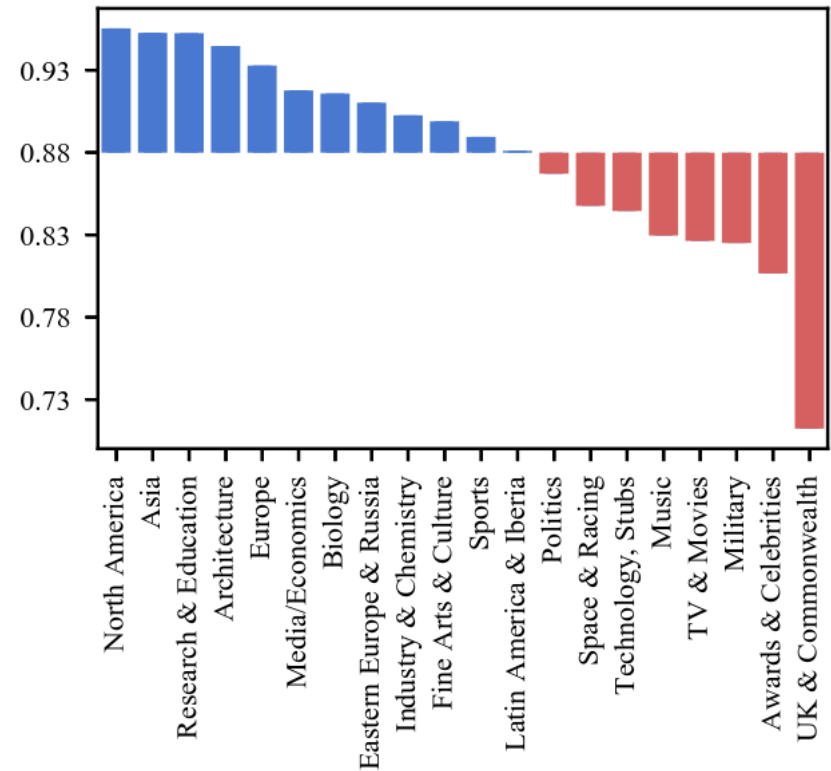
- Network features
  - ▶ In, out, and total degree
  - ▶ K-Core
- Content and edit features
  - ▶ Editors, revisions, age
  - ▶ Size(kB), sections, tables
- Topics
  - ▶ 20 topics discovered using LDA
  - ▶ Topics are labeled by five researchers using the top words and articles for each topic

# Are topics accessed differently?

## Searchshare

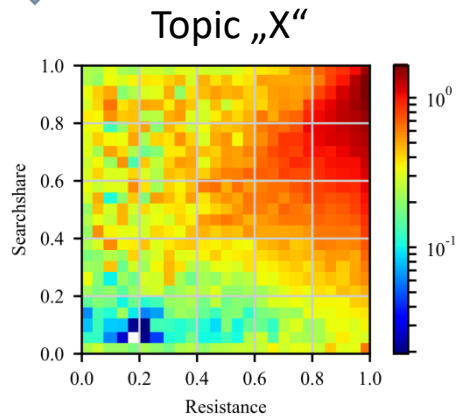


## Resistance





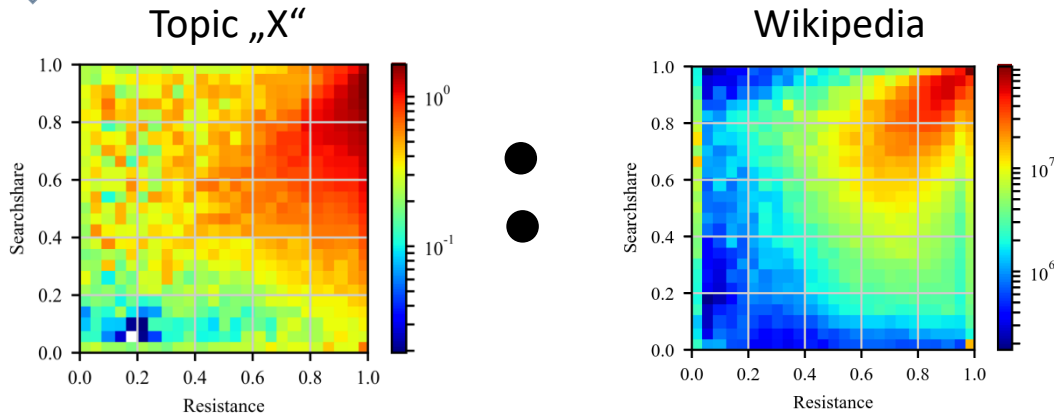
# Topic-specific heatmaps



- Create a topic-specific heatmap



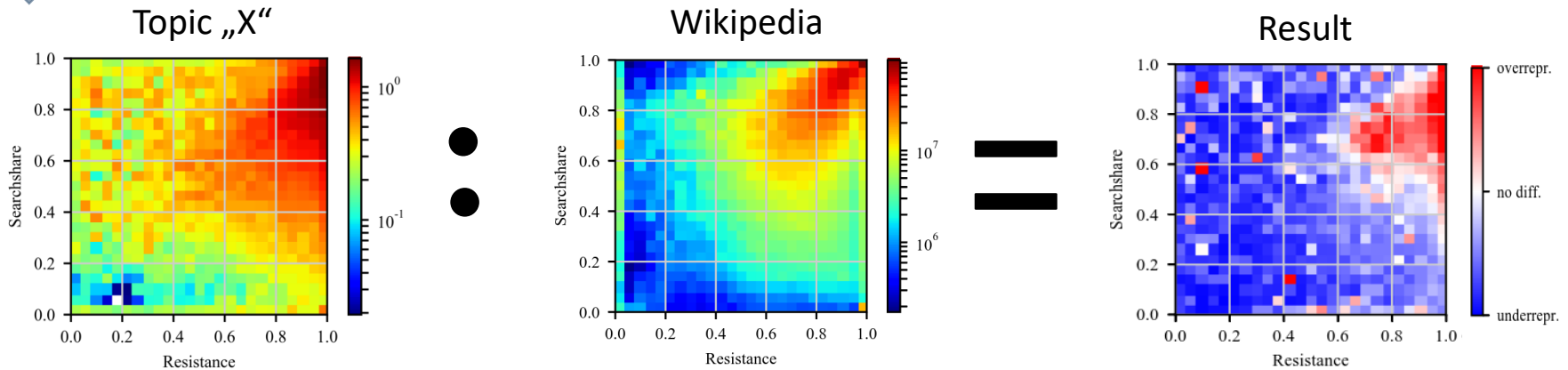
# Topic-specific heatmaps



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



# Topic-specific heatmaps

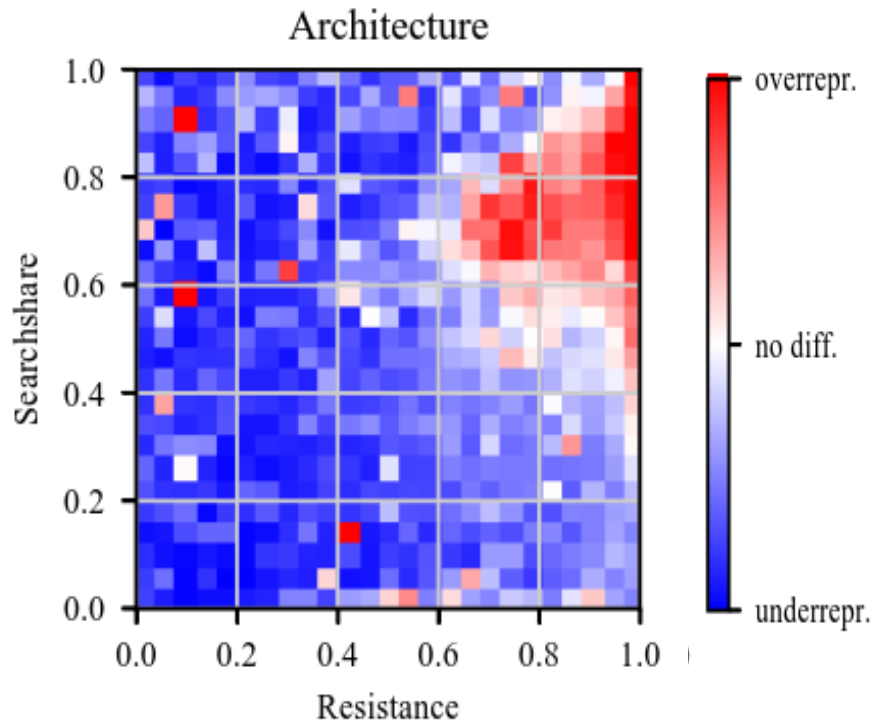


- Create a topic-specific heatmap
- Bin-wise division of topic-specific and Wikipedia heatmaps
- Result: heatmap showing the deviation of the topic-specific from the general access behavior

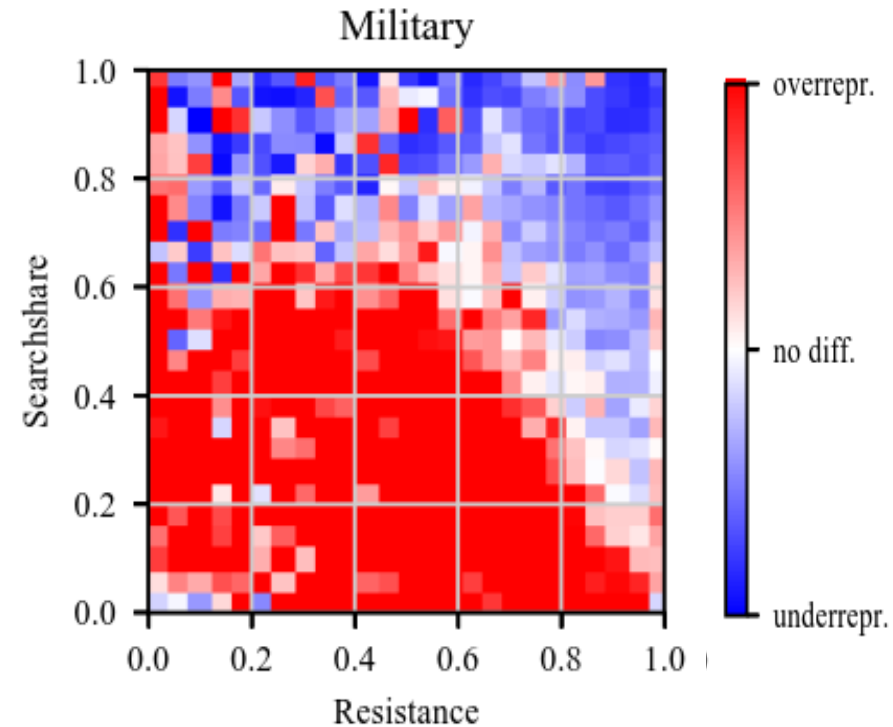
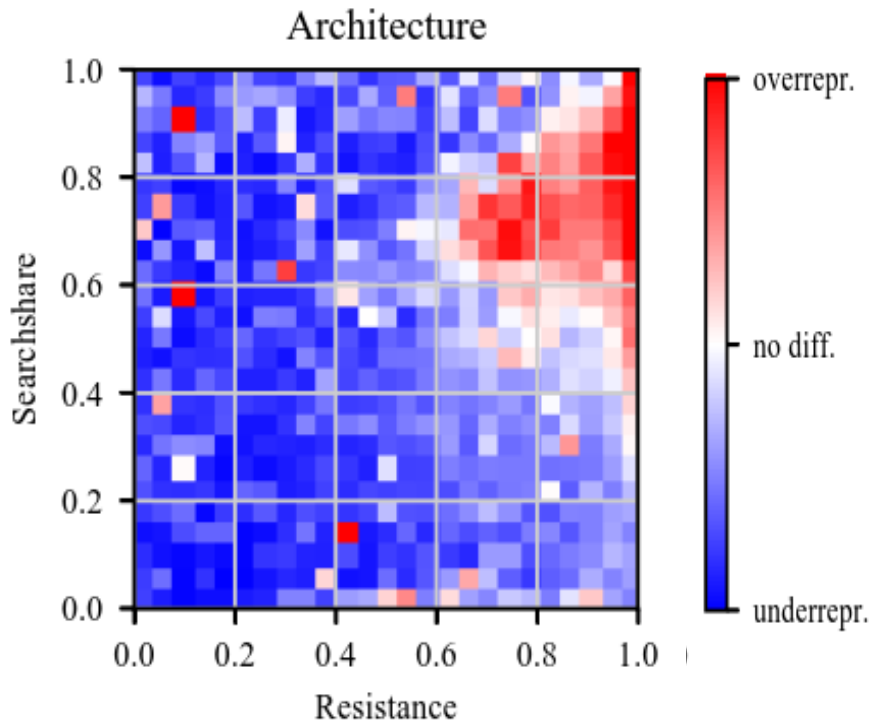


# What is the role of a topic for the traffic?

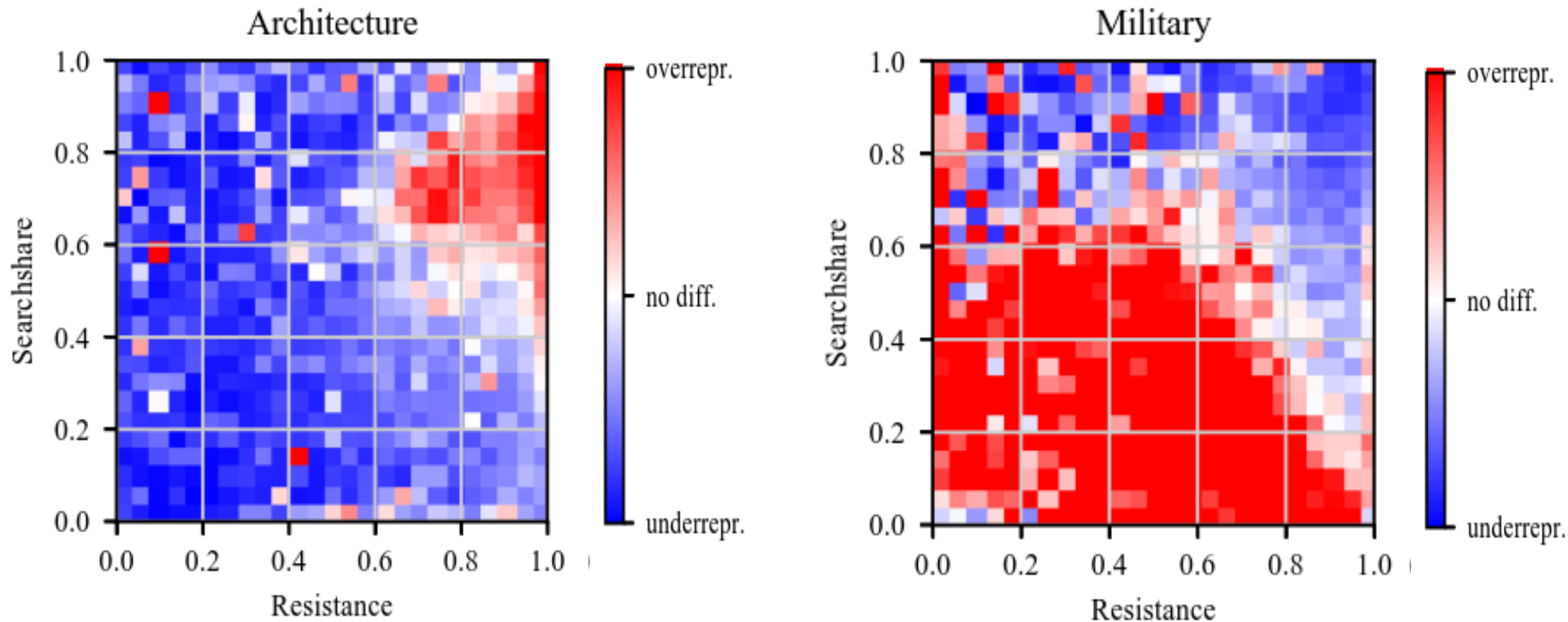
# What is the role of a topic for the traffic?



# What is the role of a topic for the traffic?



# What is the role of a topic for the traffic?



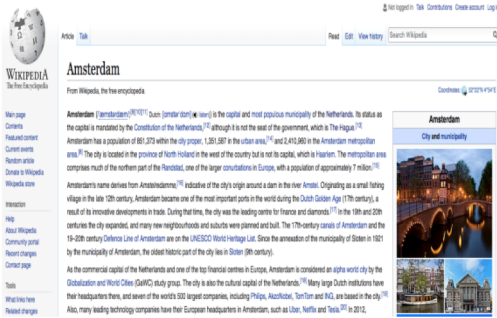
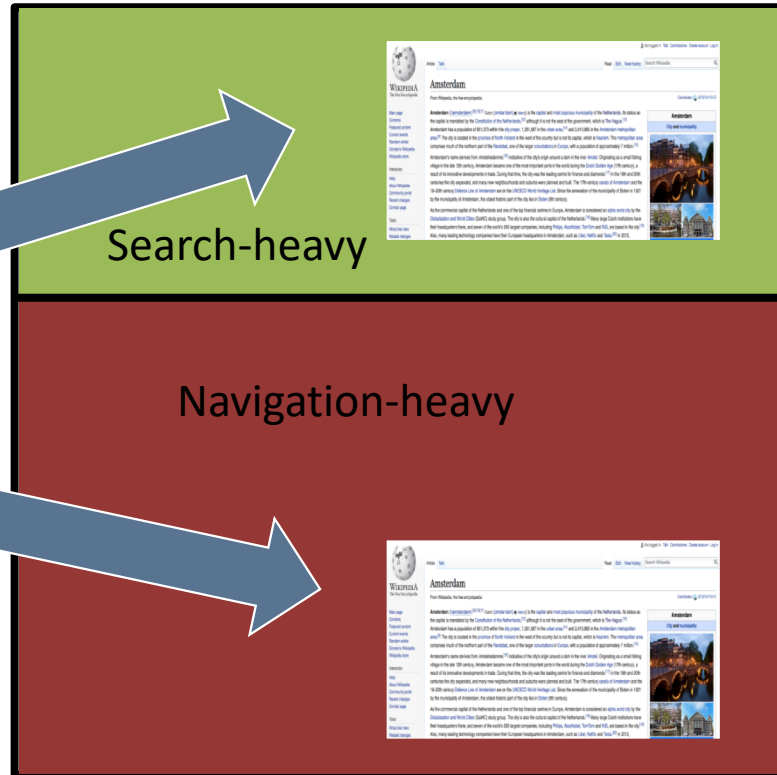
Query for „Architecture“, Click through „Military“

# Which article properties are indicative of specific information access behavior?

Step 3

# Modeling access behavior

1.0



?

Search-heavy

Navigation-heavy

0.66

Searchshare

0

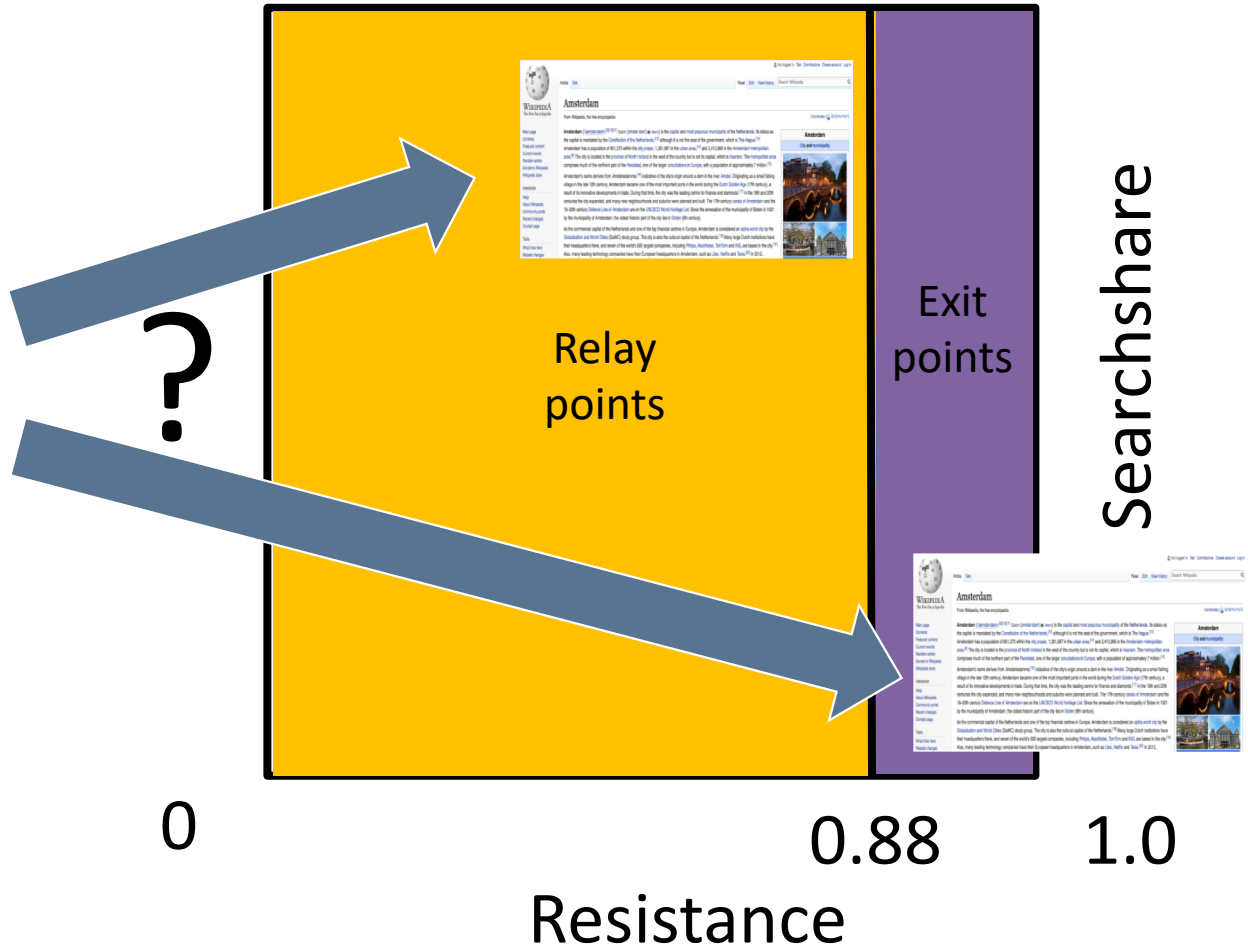
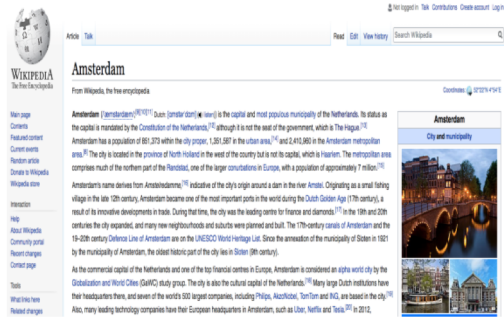
1.0

Resistance

Step 3

# Modeling access behavior

1.0



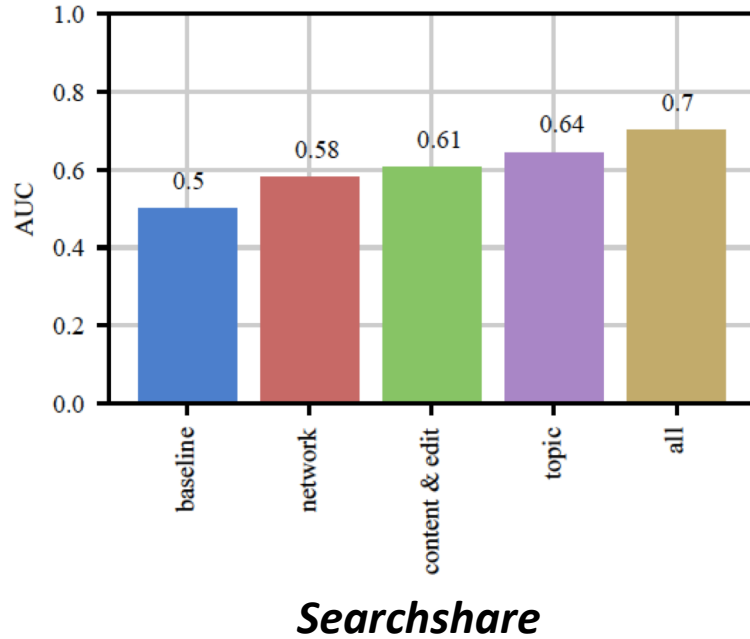
0

0.88

1.0

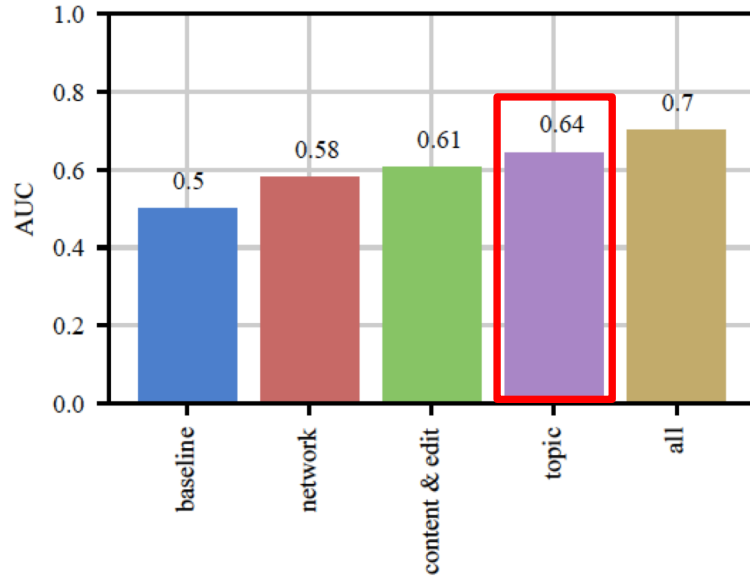
Resistance

# Results





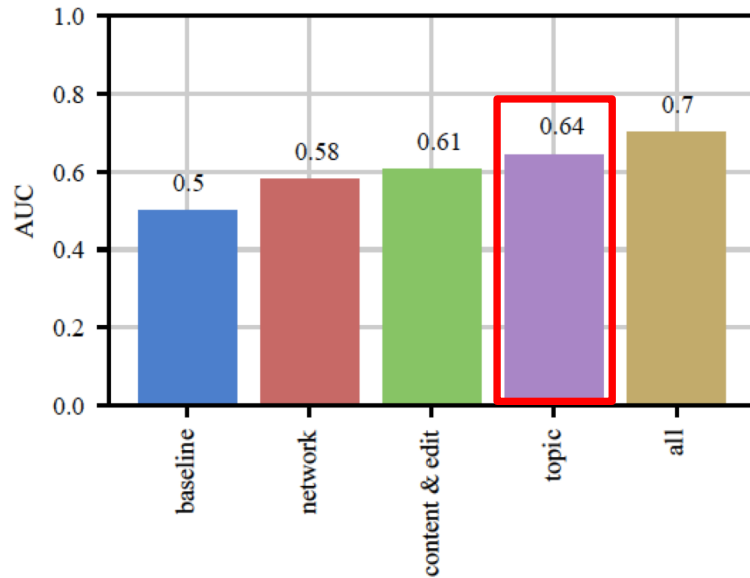
# Results



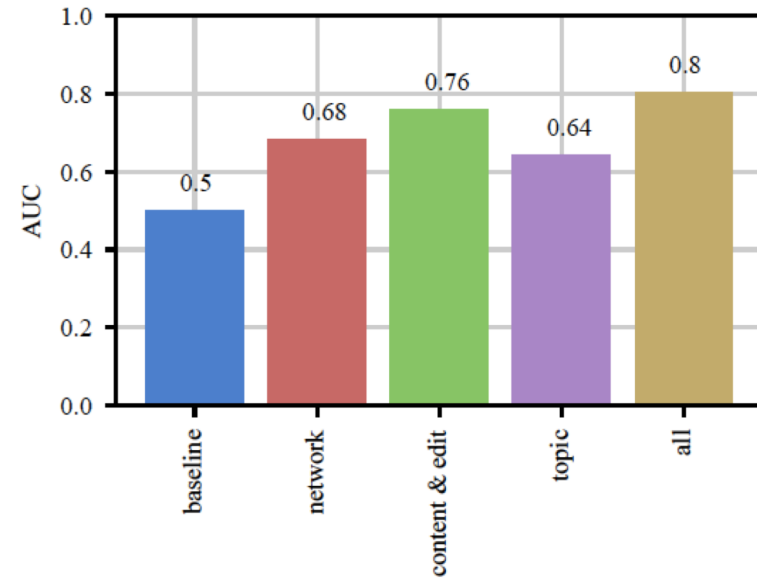
## *Searchshare*

1. Article topic is indicative of searchshare

# Results



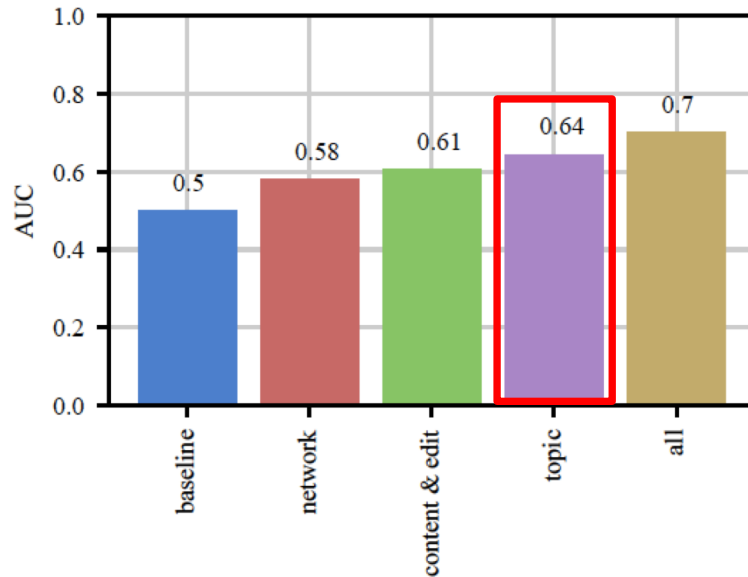
***Searchshare***



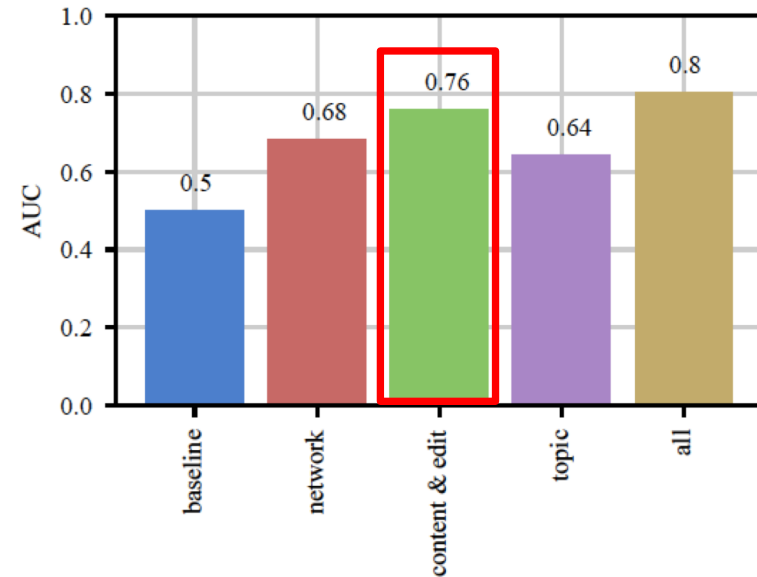
***Resistance***

1. Article topic is indicative of searchshare

# Results



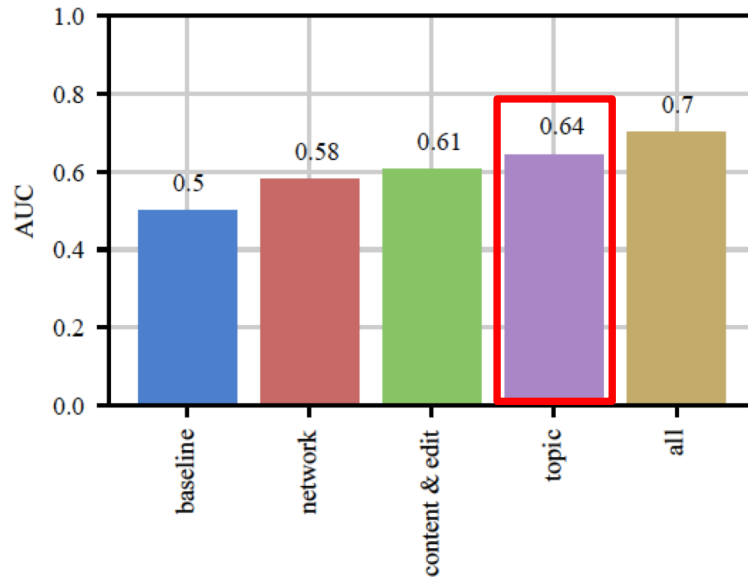
***Searchshare***



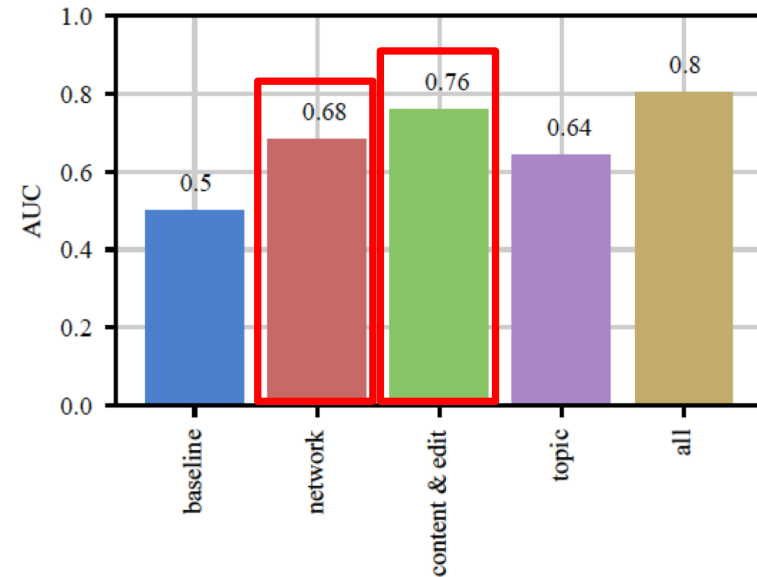
***Resistance***

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 patterns 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](mailto:dimitar.dimitrov@gesis.org)

Twitter: [@trovdimi](https://twitter.com/trovdimi)