

**The
Alan Turing
Institute**

Wikipedia Game

A. Hickey | J. Knight | J. Bishop

2nd December 2024, Robots in Disguise



The Wikipedia Game

Get from a start page to a target page by clicking links in the pages. The fewer clicks the better.

≡ Papercuts (Illy song)

Article [Talk](#)

From Wikipedia, the free encyclopedia

"Papercuts" is a song by Australian [rapper Illy](#), which features Australian female [folk/indie-pop](#) vocalist [Vera Blue](#) and was produced by [M-Phazes](#). The song was released on 8 July 2016 and peaked within the top 10 on the Australian [ARIA Chart](#). The song is the lead single from Illy's fifth studio album, *Two Degrees*. "Papercuts" appears on the album as the sixth track.^[1]

≡ Baleine, Nova Scotia

Article [Talk](#)

From Wikipedia, the free encyclopedia

"Baleine" redirects here. For other uses, see [Bale](#)

Baleine (/bɜːrˈliːn/ *bay-LEEN*)^[1] (formerly Port aux Baleines) is a community in the [Canadian](#) province of [Nova Scotia](#), located in the [Cape Breton Regional Municipality](#) on [Cape Breton Island](#). The community is perhaps best known as the landing site for pilot [Beryl Markham](#)'s record flight across the Atlantic Ocean.

Motivation

The Wikipedia Game is a real-world example of a challenge that requires **linking concepts semantically**, which can also be assessed on different **languages and topics** allowing comparison between models **across domains** without requiring machine translation or generation.

Example Path 1

The **2015 World Wheelchair Curling Championship** was held from February 6 to 13 at the Kisakallio Sports Institute in Lohja, Finland.



Scotland is a country that is part of the United Kingdom. It contains nearly one-third of the United Kingdom's land area, consisting of the northern part of the island of Great Britain and more than 790 adjacent islands, principally in the archipelagos of the Hebrides and the Northern Isles. To the south-east, Scotland



The **FIFA World Cup**, often called the **World Cup**, is an international association football competition among the senior men's national teams of the members of the *Fédération Internationale de Football Association* (FIFA), the sport's global governing body. The tournament has been held every four years.

Weerathep Pomphan is a Thai professional footballer who plays as a defensive midfielder for Thai League 1 club Bangkok United and the Thailand national team.



The **Thailand national football team** represents Thailand in senior international football and is controlled by the Football Association of Thailand.

Example Path 2

"Morning Mood" is part of Edvard Grieg's *Peer Gynt*, Op. 23, written in 1875 as incidental music to Henrik Ibsen's play of the same name, and was also included as the first of four movements in *Peer Gynt Suite No. 1*, Op. 46.

Acacia, commonly known as **wattles** or **acacias**, is a genus of about 1,084 species of shrubs and trees in the subfamily Mimosoideae of the pea family Fabaceae. Initially, it comprised a group of plant species native to Africa, South America, and Australasia, but is now reserved



Paraserianthes lophantha, the Cape Leeuwin wattle, Bicol wattle, Cape wattle, Crested wattle, Brush wattle or plume albizia, is a fast-growing tree with creamy-yellow, bottlebrush like flowers. It is the sole species in genus **Paraserianthes**.



Falcataria falcata, commonly known as the **Moluccan albizia**, is a species of fast-growing tree in the family Fabaceae. It is native to the Maluku Islands, New Guinea, the Bismarck Archipelago, and the Solomon Islands. It is cultivated in South Asian countries.



The following is a list of **Streptomyces** species, organized alphabetically by species name. Names that have not been validated according to the Bacteriological Code are enclosed in "quotes".

Streptomyces longispororuber is a bacterium species from the genus of **Streptomyces**. *Streptomyces longispororuber* produces corallomycin.



Streptomyces is the largest genus of Actinomycetota, and the type genus of the family Streptomycetaceae. Over 700 species of *Streptomyces* bacteria have been described. As with the other Actinomycetota, streptomycetes are gram-positive, and have very large genomes with high GC content.

English Wikipedia Dataset

– 6 855 207 Pages

– 668 251 260 Links



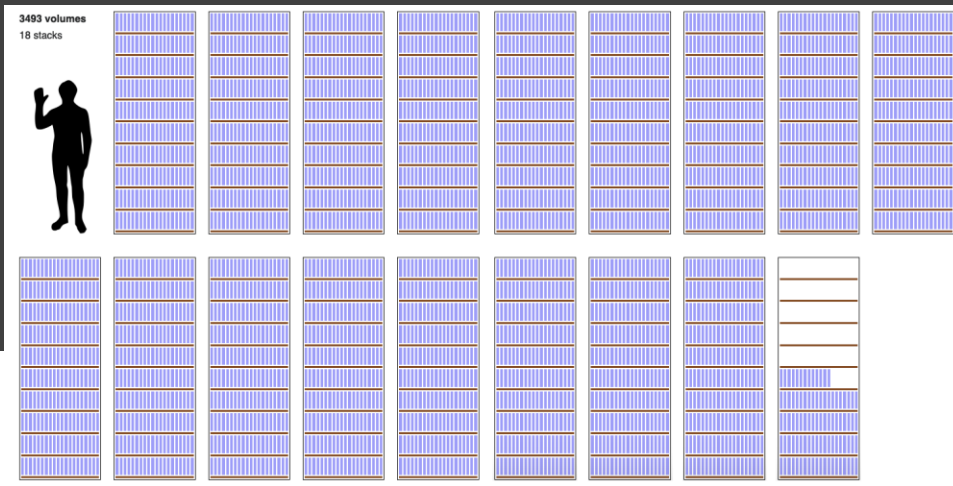
WIKIPEDIA
The Free Encyclopedia

≡ VW Jakarta Elektrik PLN

Article [Talk](#)

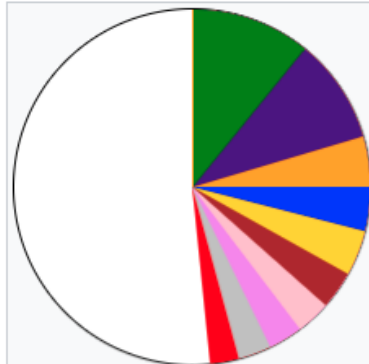
From Wikipedia, the free encyclopedia

Jakarta Elektrik PLN is an Indonesian professional women's volleyball club based in [Jakarta](#) and owned and managed by the [PT. Perusahaan Listrik Negara](#). The club was founded in 2004 and is currently a participant in the [women's Proliga](#). This club has won the title six times in the Proliga, being the most among other women's volleyball clubs in the Proliga.

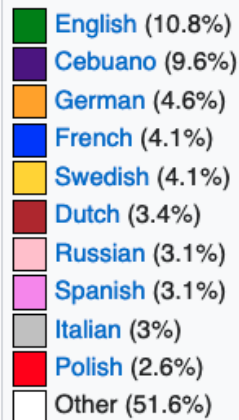


Multilingual Wikipedia

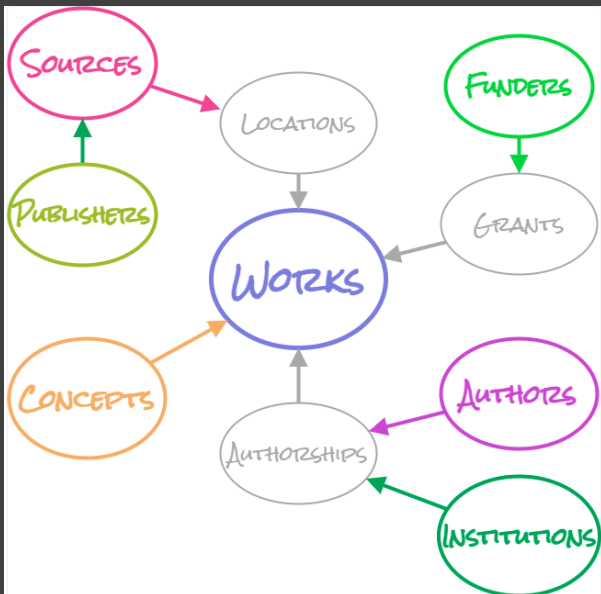
Language ↕	Language (local) ↕	Wiki ↕	Articles ↕	Users ↕	Active users ↕	Files ↕	Depth ↕
English	English	en	6,915,406	48,312,528	122,731	930,187	1,281
Cebuano	Cebuano	ceb	6,116,888	121,930	166	1	2
German	Deutsch	de	2,961,976	4,473,557	17,213	129,898	93
French	français	fr	2,649,109	5,055,882	17,621	72,559	267
Swedish	svenska	sv	2,597,953	929,689	1,913	0	17
Dutch	Nederlands	nl	2,172,490	1,375,028	3,497	20	19
Russian	русский	ru	2,011,600	3,665,235	9,282	256,455	159
Spanish	español	es	1,991,453	7,284,171	14,099	0	197
Italian	italiano	it	1,892,618	2,575,458	7,416	122,307	190
Polish	polski	pl	1,636,119	1,341,572	4,404	262	34
Egyptian Arabic	مصرى	arz	1,625,592	250,708	184	1,319	0
Chinese	中文	zh	1,452,085	3,622,321	7,116	68,295	210
Japanese	日本語	ja	1,438,122	2,259,110	12,480	4,734	91
Ukrainian	українська	uk	1,354,371	775,385	3,194	115,295	59
Vietnamese	Tiếng Việt	vi	1,294,086	973,098	1,465	26,414	729
Waray	Winaray	war	1,266,606	61,309	70	42	4
Arabic	العربية	ar	1,246,352	2,654,804	4,265	53,492	276
Portuguese	português	pt	1,138,512	3,153,820	8,783	64,061	199
Persian	فارسی	fa	1,019,952	1,352,325	4,944	94,019	152
Catalan	català	ca	763,723	498,880	1,570	11,443	40



Distribution of the 64,002,387 articles in different language editions (as of 25 November 2024)^[12]



Academic Literature

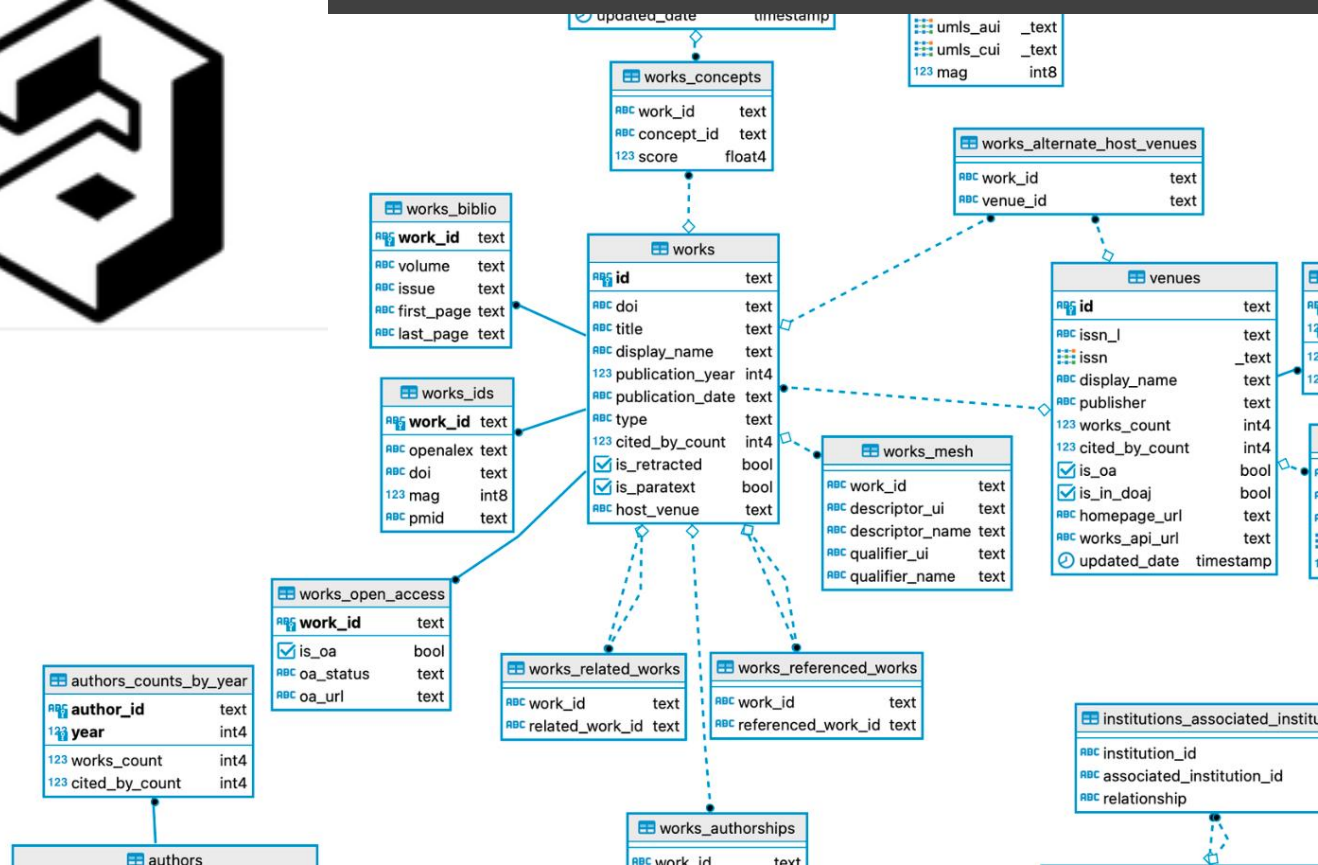


OpenAlex

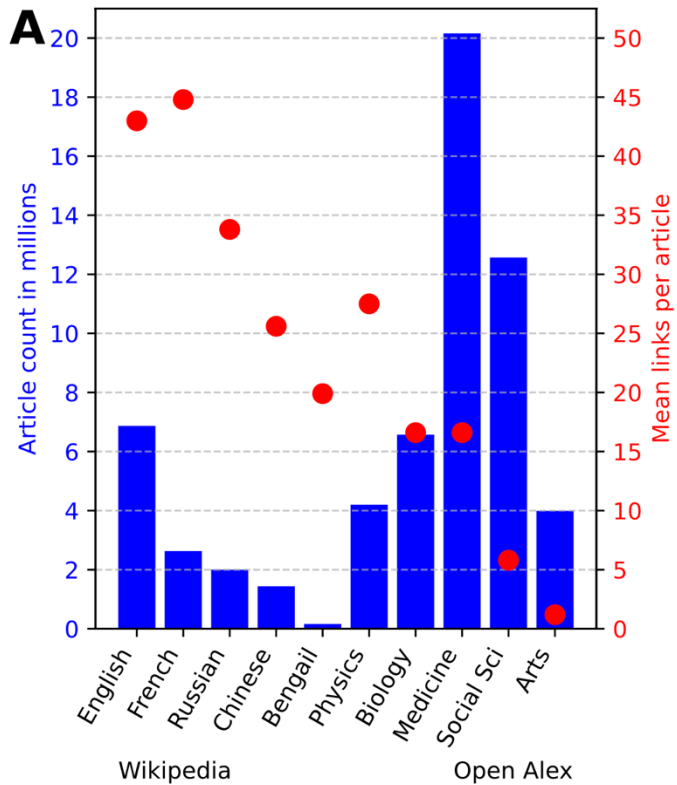
OpenAlex is a fully open catalog of the global research system. It's named after the [ancient Library of Alexandria](#) and made by the nonprofit [OurResearch](#).

Academic Literature – OA Works

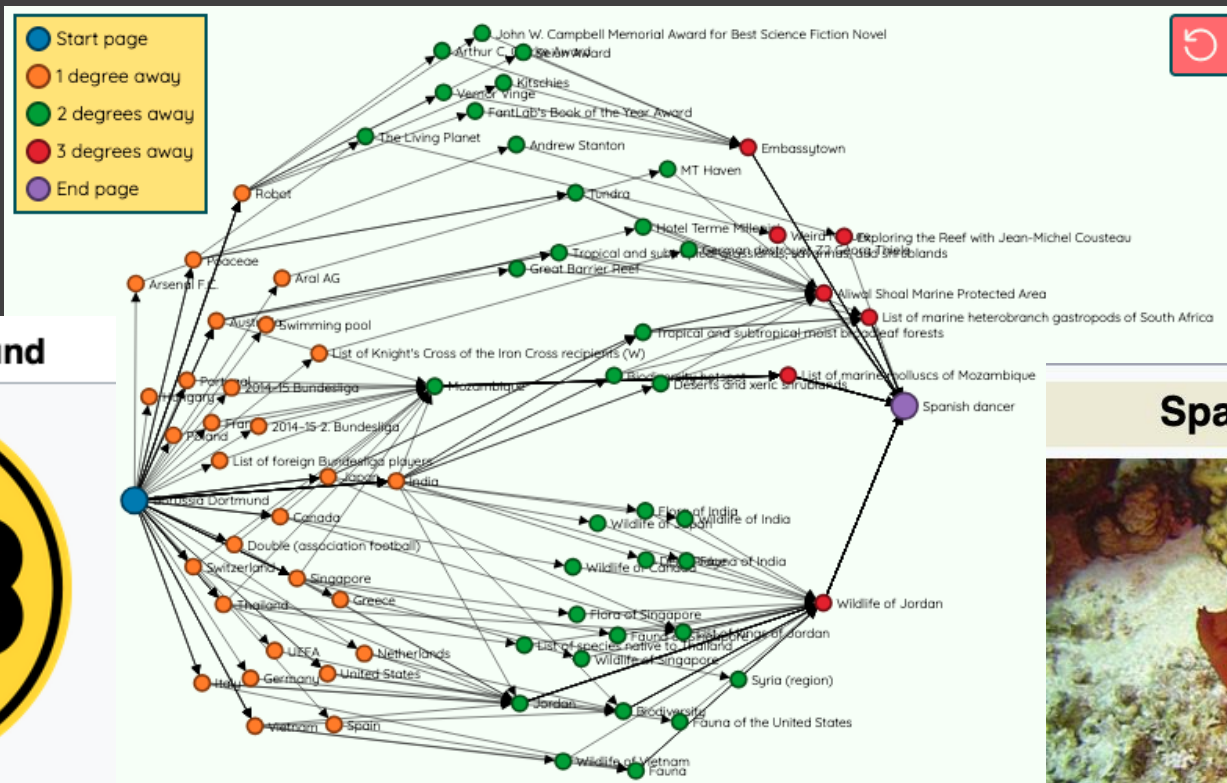
- Title
- Topic
- Language
- Papers Cited
- Work Type
- Abstract
- Year
- ...



The Graphs to Navigate



Path Finding



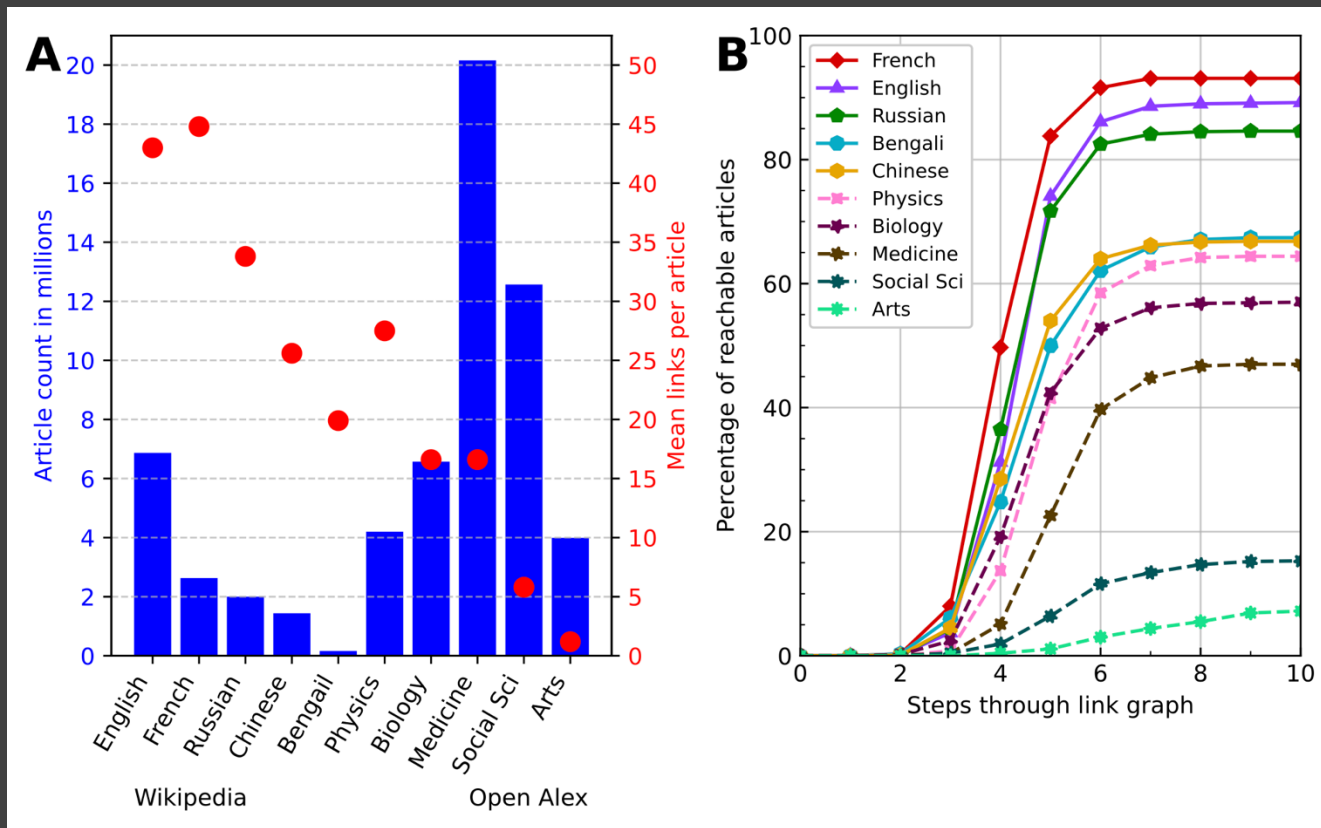
Borussia Dortmund



Spanish dancer



The Graphs' Connectivity



The Language Models

Embedding Models and Cosine Distance

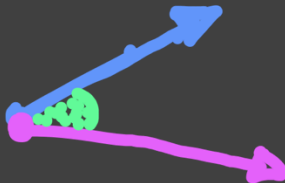
"The dog saw the swan and barked at it. The swan disapproved and started to..."

SBERT

[0.34, 0.45, -0.82, 0.22, -0.01,
0.87, ..., -0.54]

[0.2, -0.3, ..., 0.5]

[0.1, -0.9, ..., -0.2]



Sentence Transformers



Sentence Embedding Model	Parameters (Millions)	Multilingual
multilingual-e5-large	560	Y
all-roberta-large-v1	355	N
stsb-xlm-r-multilingual	278	Y
paraphrase-multilingual-mpnet-base-v2	278	Y
distiluse-base-multilingual-cased-v2	135	Y
distiluse-base-multilingual-cased-v1	135	N
average_word_embeddings_glove.6B.300d	120	N
multilingual-e5-small	118	Y
paraphrase-multilingual-MiniLM-L12-v2	118	Y
all-mpnet-base-v2	109	N
all-distilroberta-v1	82.1	N
all-MiniLM-L6-v2	22.7	N
paraphrase-albert-small-v2	11.7	N

Prompting - LLMs



```
self._system_content = (  
    "You are a helpful assistant that helps me to play the Wikipedia Game! "  
)  
self._query1_content = (  
    "If I give you the title and a summary of the target page that I am "  
    "trying to navigate to, can you help me navigate to the target page? "  
)  
self._response1_content = "Of course! What is your target page?"  
self._response2_content = (  
    "Sure! I can help you navigate to that page, what titles are there on "  
    "your current page?"  
)  
self._play_round_content = (  
    "Here is a list of titles, which one should I pick? Please provide your "  
    'answer in JSON format with the selected title given with key "title" and '  
    'a justification given with key "reason". The selected title must be from '  
    "the following list: "  
)
```

LLM – Forcing JSON

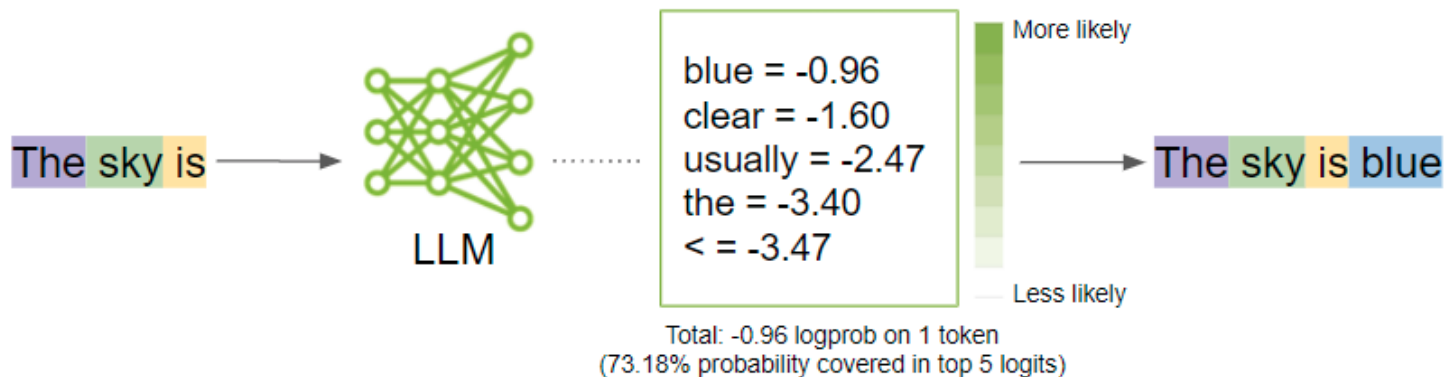


Figure 1. General working flow of an LLM predicting the next word

GPT-4o, Llama, Gemma



GPT-4o







Gemma






Model	Parameters (Billions)	Size (GB)
Llama3.2	1	2
	3	3
	8	5
Llama3.1	70	40
	405	230
Gemma2	27	16
Mistral NeMo	12	3
GPT-4o-mini	Unknown	Unknown
GPT-4o	Unknown	Unknown


Ollama on Baskerville

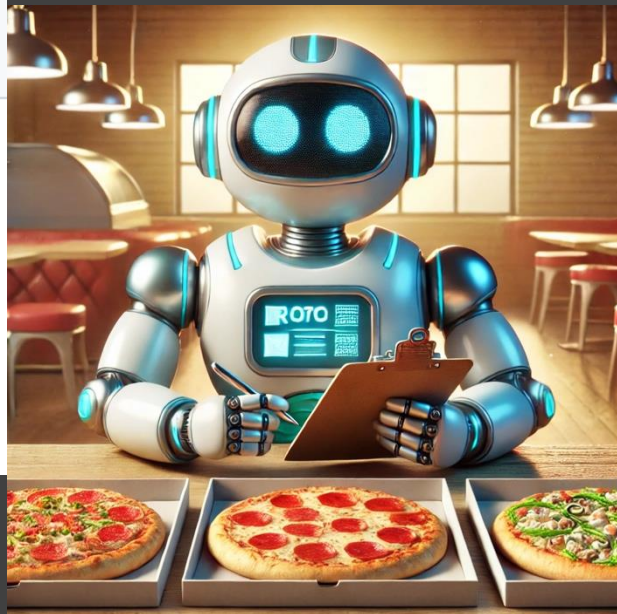
 alan-turing-institute / ollama-hpc-example 

[Code](#) [Issues](#) [Pull requests](#)



A minimal example package running ollama on baskerville.

 MIT license



Term frequency - Inverse Document Frequency



GENSIM
topic modelling for humans

ENGLISH

A salt seller used to carry the salt bag on his donkey to the market every day. On the way they had to cross a stream. One day the donkey suddenly tumbled down the stream and the salt bag also fell into the water. The salt dissolved in the water and hence the bag became very light to carry. The donkey was happy. Then the donkey started to play the same trick every day. The salt seller came to understand the trick and decided to teach a lesson to it. The next day he loaded a cotton bag on the donkey. Again it played the same trick hoping that the cotton bag would be still become lighter. But the dampened cotton became very heavy to carry and the donkey suffered. It learnt a lesson.



A=5
salt=5
seller=4
used=1
to=10
carry=1
the=21
that=1
bag=1
on=3
his=1
market=1
every=2
day=4
way=1
they=1
had=1
cross=1
stream=2
One=1
Donkey=7
Suddenly=1
Tumbled=1

Down=1
And=4
bag=4
also=1
fell=1
into=1
water=2
dissolved=1
in=1
hence=1
became=2
very=1
light=1
carry=1
was=1
happy=1
Then=1
Started=1
Play=1
Same=1
Trick=3
came=1
understand=1
decided=1
teach=1

lesson=2
next=1
he=1
loaded=1
cotton=1
Again=1
It=2
Played=1
Same=1
Hoping=1
Cotton=2
Would=1
Be=1
Still=1
Lighter=1
But=1
Dampened=1
Became=1
Very=1
Heavy=1
Carry=1
Suffered=1
It=1
Learnt=1

Humans

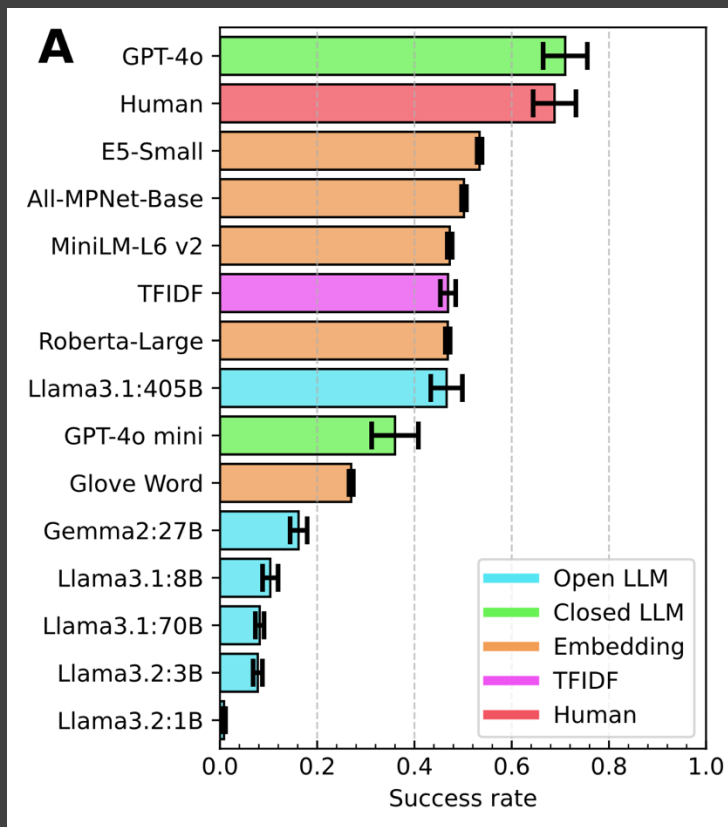
	A	B	C	D	E	F	G
1	Nur	Success	Steps	Time (m:	SpeedWikiU	TargetURL	StartPage
2	0	Yes	8	04:51	https://wikispeed	https://en.wikipedia	Cambusavie Platform railway station
3	1	No	18		https://wikispeed	https://en.wikipedia	Hannan (surname)
4	2	Yes	16	05:15	https://wikispeed	https://en.wikipedia	USRC Rush
5	3	Yes	17	05:55	https://wikispeed	https://en.wikipedia	Eva Trevisan
6	4	Yes	7	04:01	https://wikispeed	https://en.wikipedia	Operation Épaulard I
7	5				https://wikispeed	https://en.wikipedia	Galway Senior Football Championship
8	6	Yes	14	08:09	https://wikispeed	https://en.wikipedia	Christopher M. Tuckett
9	7	No	16		https://wikispeed	https://en.wikipedia	Jeffrey Gaines
10	8	No	18		https://wikispeed	https://en.wikipedia	Live on Two Legs
11	9				https://wikispeed	https://en.wikipedia	Afloqualone
12	10	Yes	7	05:02	https://wikispeed	https://en.wikipedia	Vrigny
13	11	Yes	9	04:52	https://wikispeed	https://en.wikipedia	Mtoto
14	12	No	16		https://wikispeed	https://en.wikipedia	Stenolechia bathrodyas
15	13				https://wikispeed	https://en.wikipedia	Jean-Christophe Attias
16	14	No			https://wikispeed	https://en.wikipedia	Bryn Celyn
17	15				https://wikispeed	https://en.wikipedia	Bulbophyllum aundense
			11	04:03	https://wikispeed	https://en.wikipedia	Carex spilocarpa
			10		https://wikispeed	https://en.wikipedia	Cape longclaw
			6	02:03	https://wikispeed	https://en.wikipedia	Shutdown (Skeptra song)
					https://wikispeed	https://en.wikipedia	Eirene
			5	03:12	https://wikispeed	https://en.wikipedia	Glaucocharis melli
			7		https://wikispeed	https://en.wikipedia	Grob G 109
			15	04:44	https://wikispeed	https://en.wikipedia	Seljačka buna
					https://wikispeed	https://en.wikipedia	George Gerken
26	24	Yes	8	03:43	https://wikispeed	https://en.wikipedia	Ross Berry



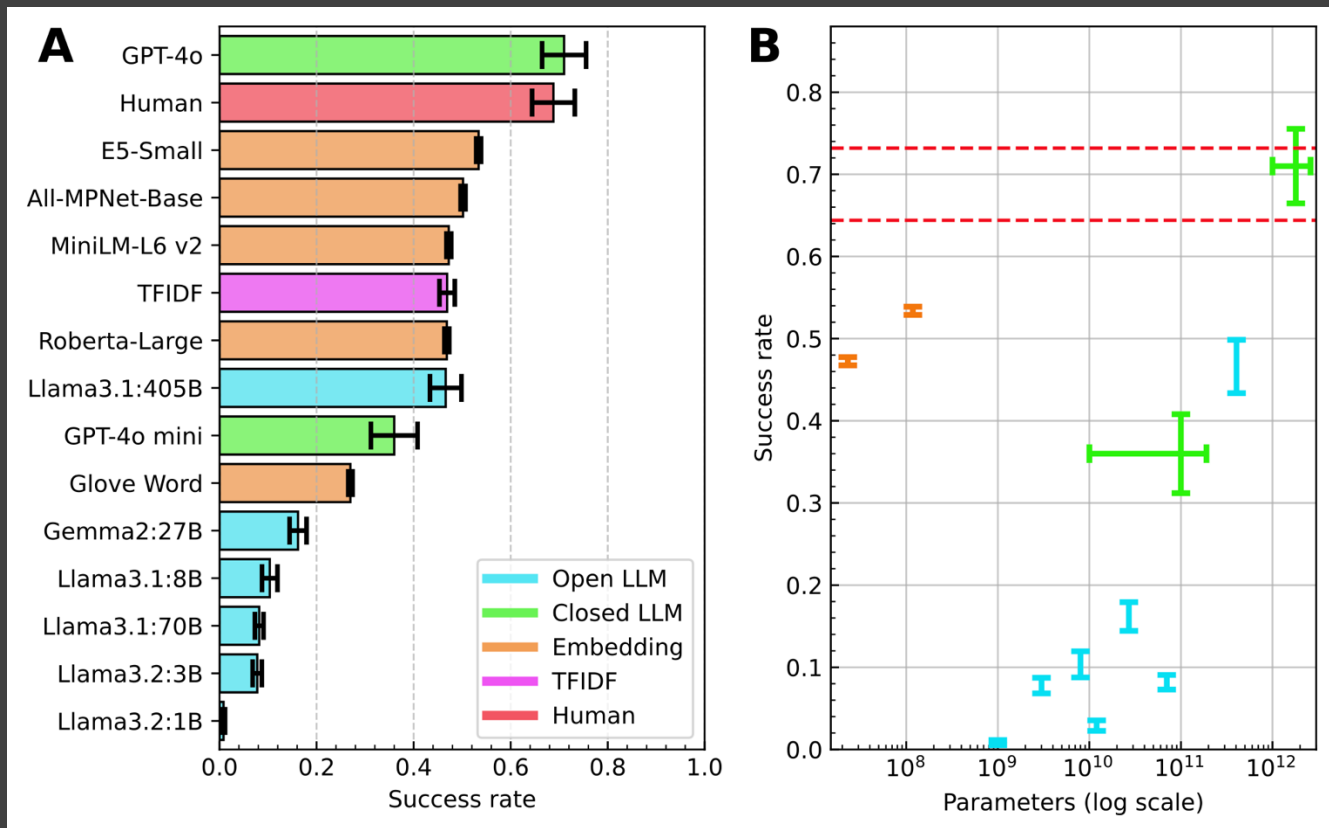
WIKIPEDIA
Speedrun

Performance on English Wikipedia

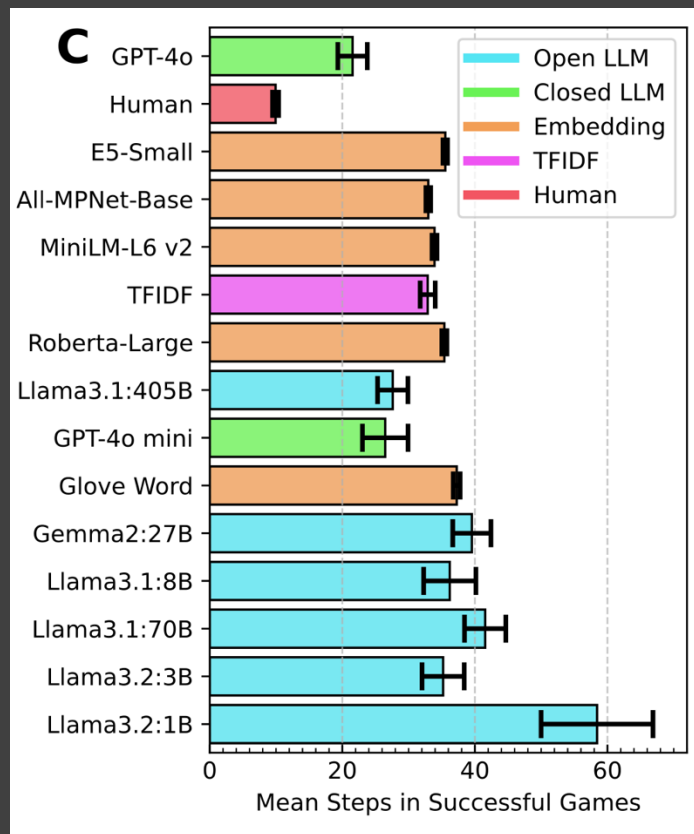
Performance on English Wikipedia



Performance with Model Size



Step Count on English Wikipedia



A Successful Game

Game to travel from `Tamagoyaki` to `Atlantic salmon` - Successful

0 : Tamagoyaki - Start page

1 : Eggs as food - The target page 'Atlantic salmon' is related to the topic of 'fish', ...

2 : Fish - The word 'fish' is a broad category that includes the Atlantic salmon, ...

3 : Osteichthyes - This is a broader category that includes fish, which is closely ...

4 : Sarcopterygii - The target page is about a group of fish that includes the ...

5 : Actinopterygii - The target page 'Atlantic salmon' is a member of the class ...

6 : Actinopteri - The target page 'Atlantic salmon' is a member of the class ...

7 : Atlantic salmon - The title matches the target page's summary, indicating that it is a relevant and likely correct choice.

Less Successful Examples

- The title 'Human' is a broad category that encompasses individuals, including Napoleon Bonaparte ...
 - The most relevant country in this context is Spain, as it shares a border with Scotland ...
- Pole vault is closely related to soil formation, as pole vaulting involves jumping over a bar set up on a foundation of soil or other material.
 - The UK is a country where wildebeests can be found...

Failing Successfully

Game to travel from `Great Barrier Reef` to `Limited-slip differential` - Successful

0 : Great Barrier Reef - Start page

1 : Great Barrier Reef Marine Park - The summary mentions that limited-slip differentials ...

2 : Exercise Talisman Sabre - The title is related to military exercises, which ...

3 : Fog Bay and Finniss River Floodplains - The target page 'Limited-slip differential' ...

4 : Intertidal zone - The mention of ZF and the specific reference to the Motor Vehicle ...

5 : **Littoral zone - The target page 'Limited-slip differential' contains the phrase 'littoral zone', which is mentioned in the context of marine habitats.**

This connection suggests that the Littoral Zone title might be relevant to the current page.

6 : Neritic zone - The Neritic zone is a coastal region of the ocean, which is ...

7 : Physical oceanography - Physical oceanography is a relevant field that deals with the ...

8 : Fluid dynamics - The title 'Limited-slip differential' is related to the topic of fluid ...

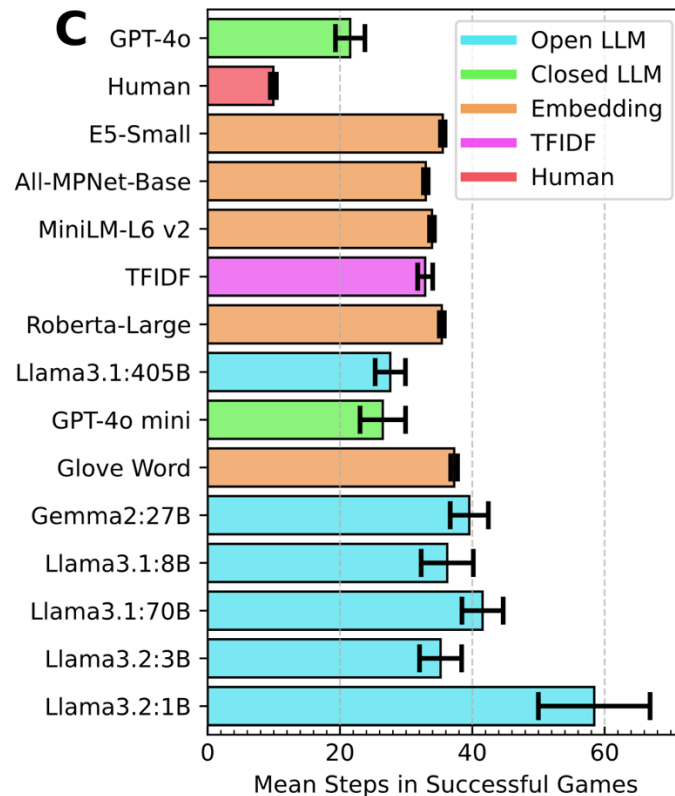
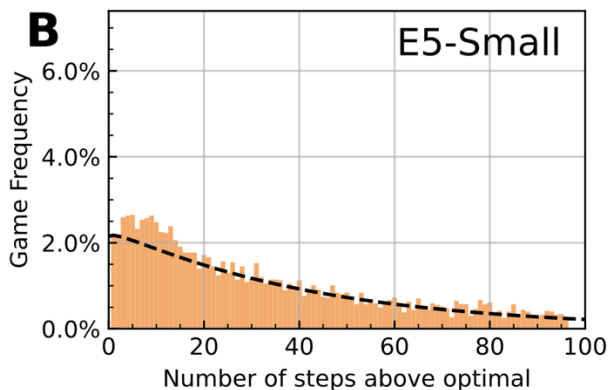
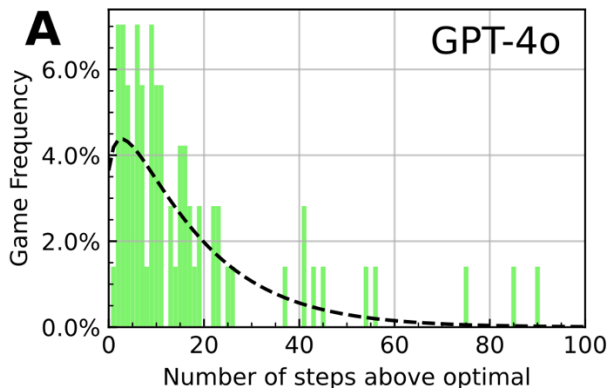
9 : Fluid mechanics - The target page 'Limited-slip differential' is closely related to the field ...

10 : Mechanical engineering - The target page 'Limited-slip differential' is closely ...

11 : Differential (mechanical device) - The target page 'Limited-slip differential' is ...

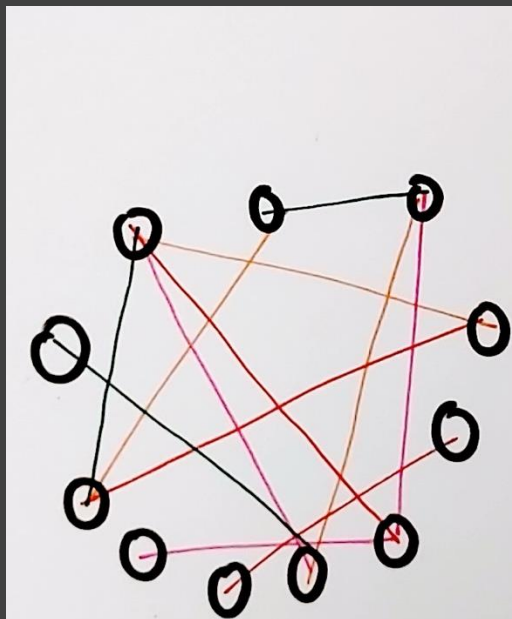
12 : Limited-slip differential - The title matches the target page's summary ...

Step Count Histogram

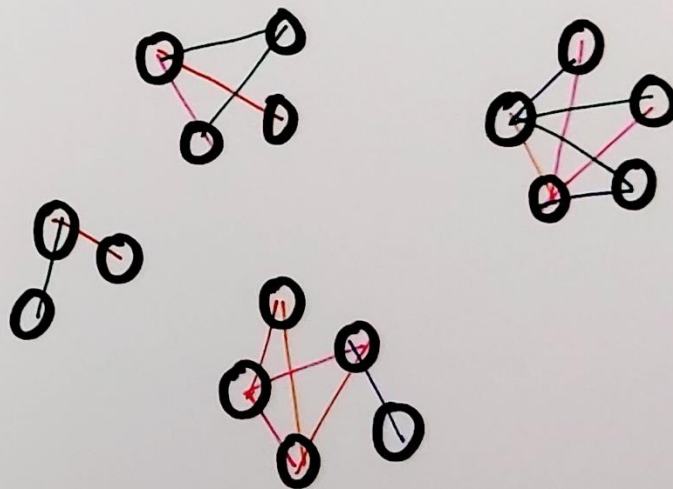


Random and Clustered Networks

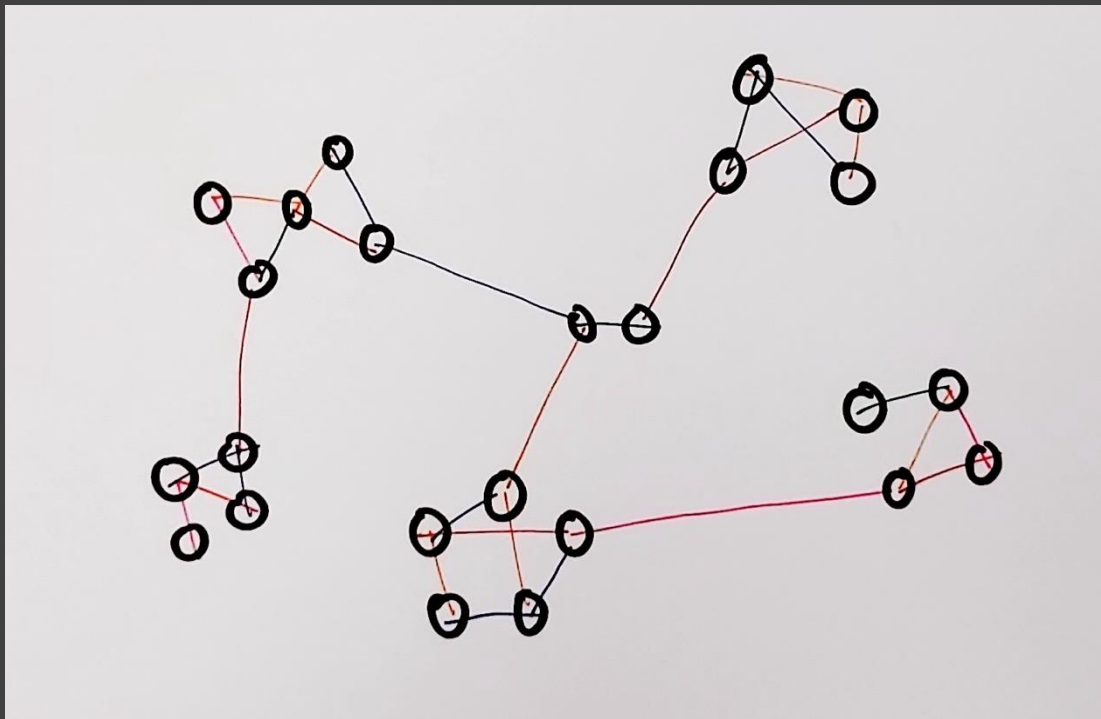
Random Network



Clustered Network



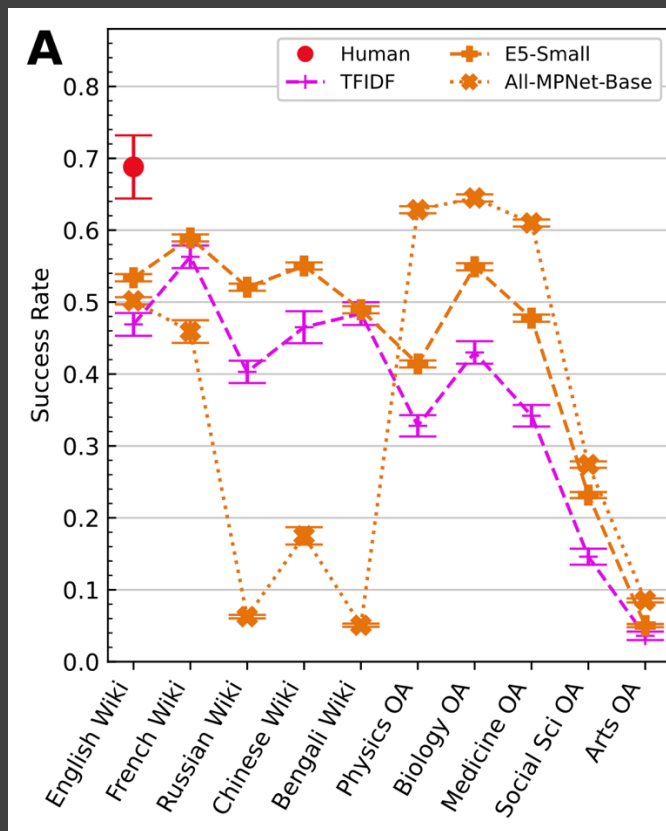
Semi-Clustered Networks



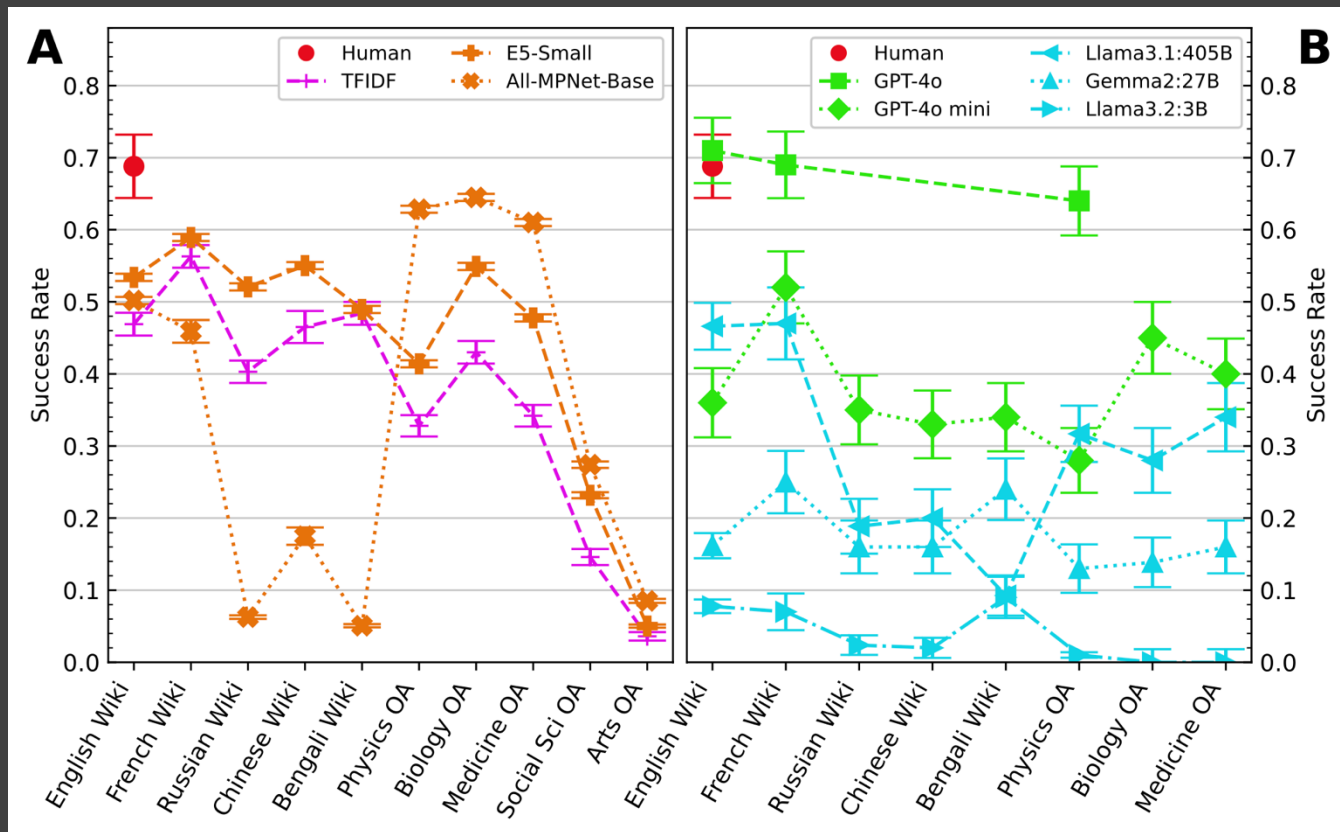
Idea: The most important factor is how quickly a hop is made to the cluster with the target node.

Language and Topic

Embedding Models



Topics and Language



Conclusion

Main Results

- Human performance similar to GPT-4o
- Embedding model performance better than open-source LLMs
- TFIDF is effective when corpus matches dataset
- The topic matters a lot and performance on English Wikipedia doesn't necessarily translate

Any Questions?

