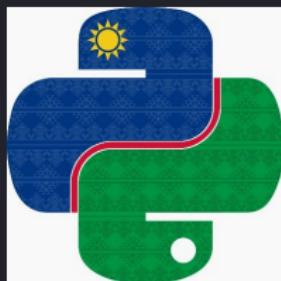


Web scraping, text analysis, and networks can help us
identify the most important Avenger

@NikoletaGlyn



Marvel Comics

1. An American media and entertainment company that was widely regarded as one of the “big two” publishers in the comic industry.

Marvel Cinematic Universe

1. Media franchise centered on a series of superhero films. The films are based on characters that appear in the comic books published by Marvel Comics.





23 movies

23 movies

8000+ characters

Who is the most important character in the MCU?





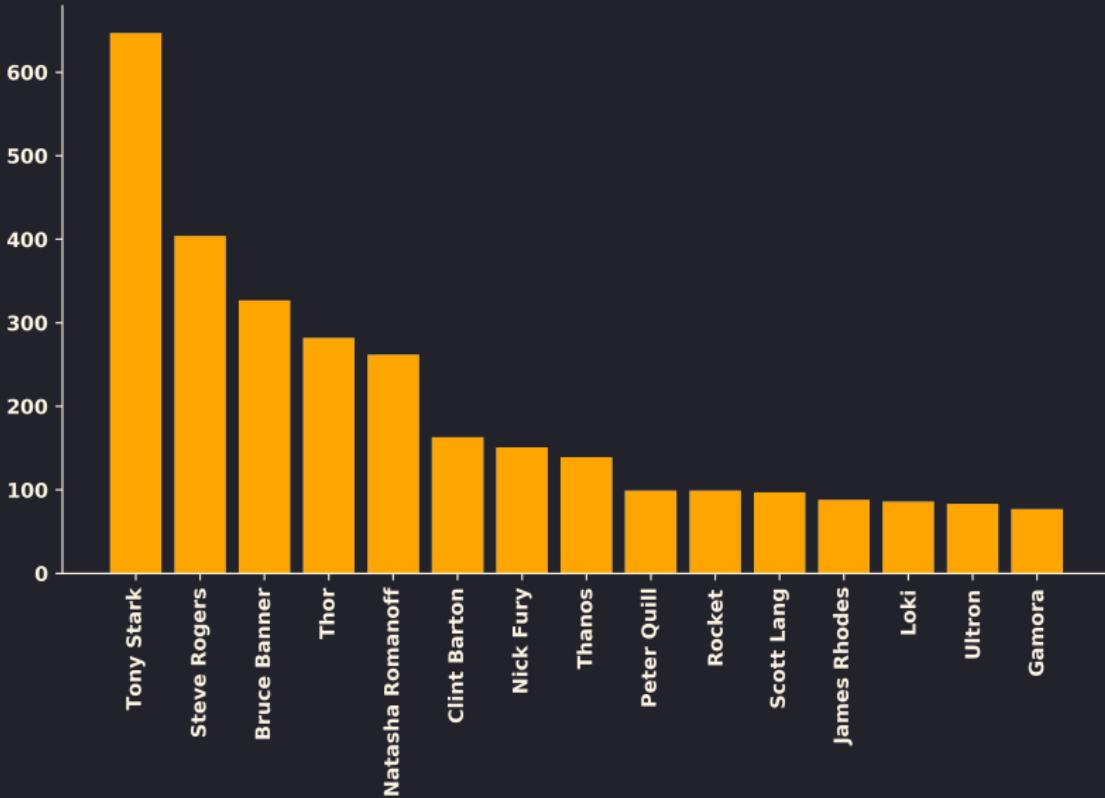
Which character has the most lines?

Tony: Let's go, Steve, no pressure



Tony: Let's go, Steve, no pressure

Spoken lines





Tony: Let's go, Steve, no pressure

Tony: Let's go, Steve, no pressure

```
import spacy
nlp = spacy.load("en_core_web_sm")

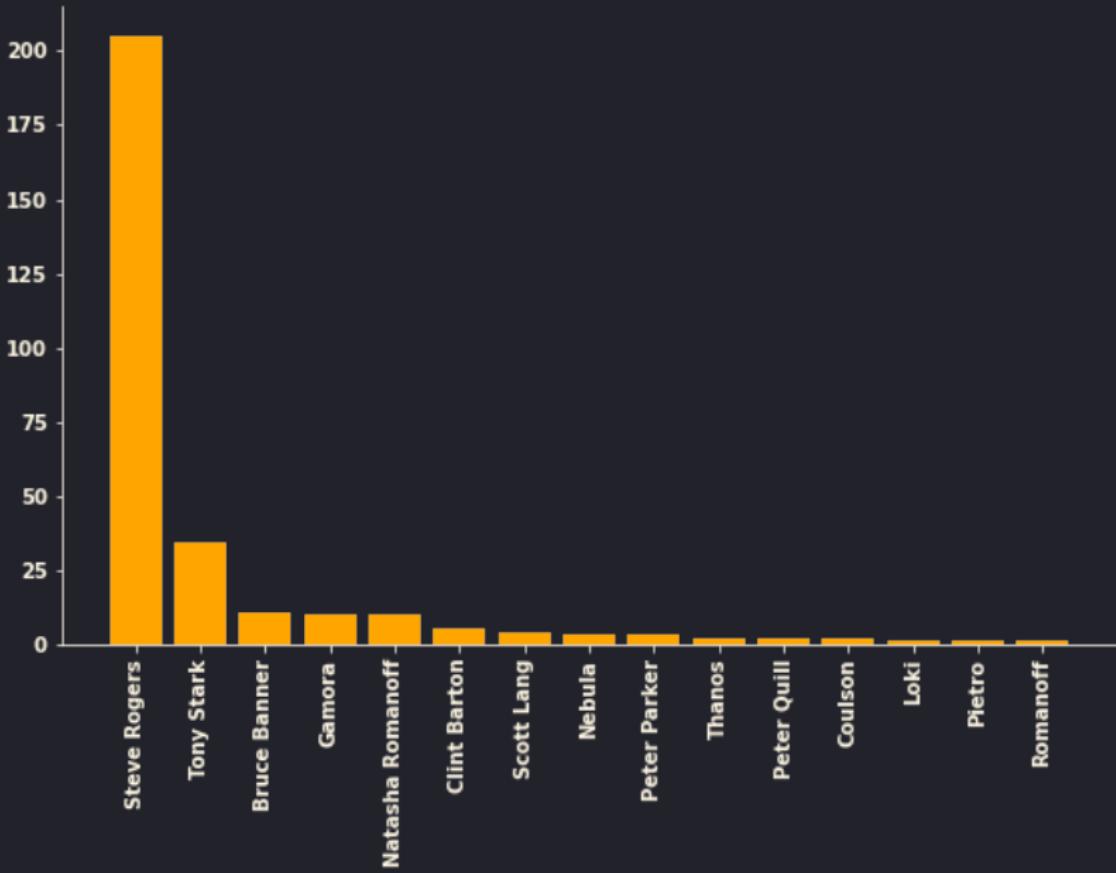
def identify_names(text):

... doc = nlp(text)
... names = [ent.text for ent in doc.ents if ent.label_ == 'PERSON']

... return names
```

Which character is mentioned the most?

Mentions



Tony: Let's go, Steve, no pressure,
blah blah Infinity stones



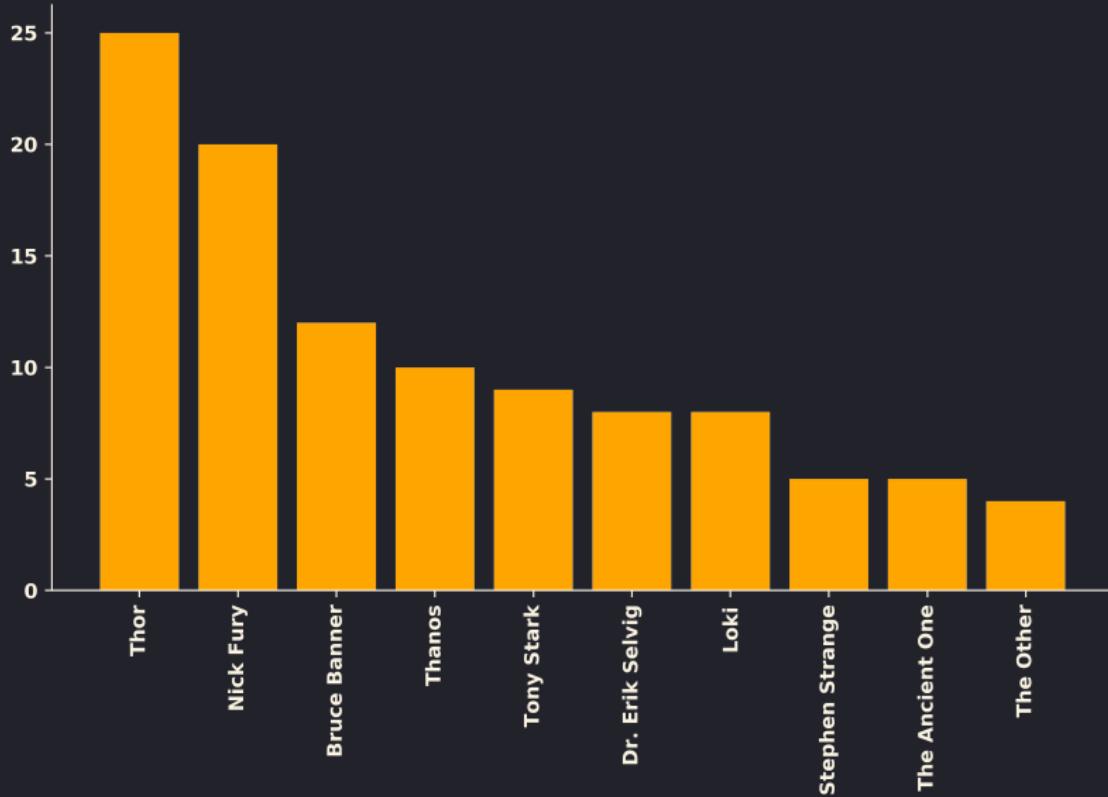
```
import spacy
nlp = spacy.load("en_core_web_sm")

def identify_names_and_stones(text):

    ... doc = nlp(text)
    ... names = [ent.text for ent in doc.ents if ent.label_ == 'PERSON']
    ... stones = [word.text for word in doc
    ...             if word.text in names_lookup['Infinity stones']]
    ...
    ... return names + stones
```

Which character interacts with the stones the most?

Interactions with stones



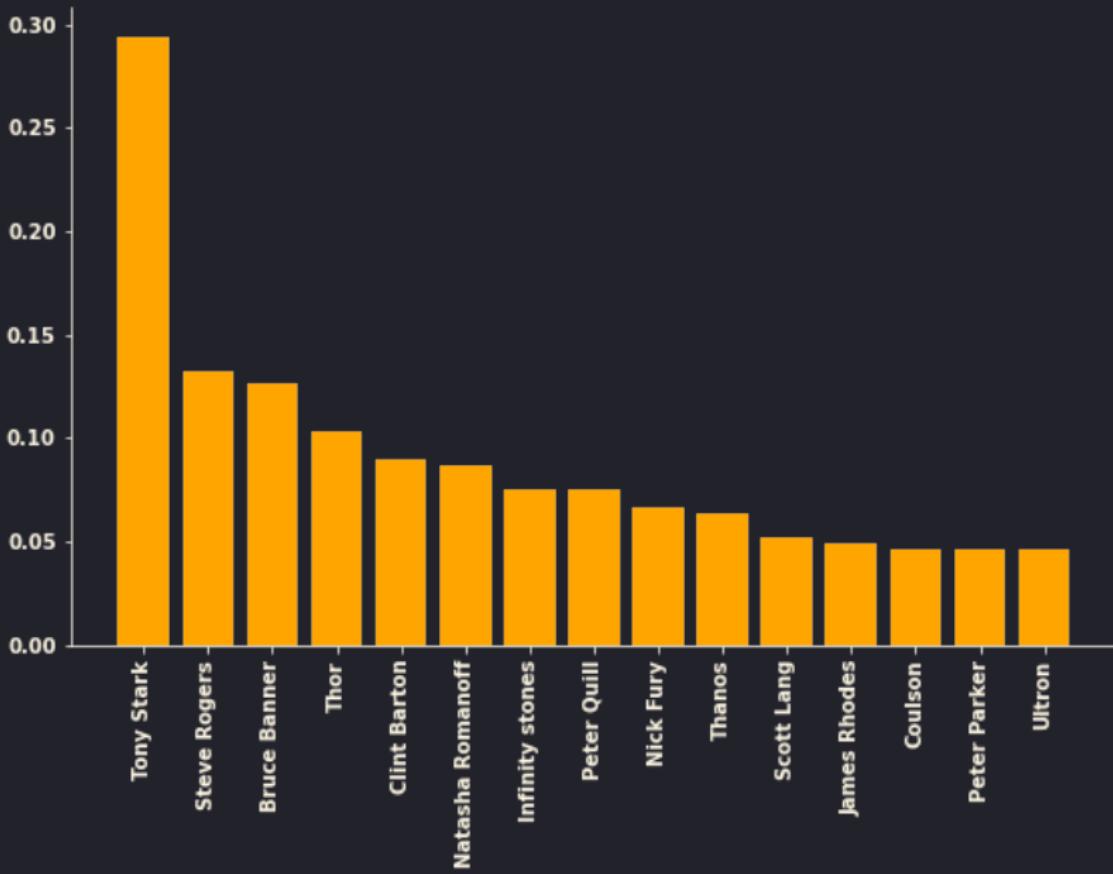
Social Network

1. A network of social interactions and personal relationships

Tony: Let's go, Steve, no pressure,
blah blah Infinity stones

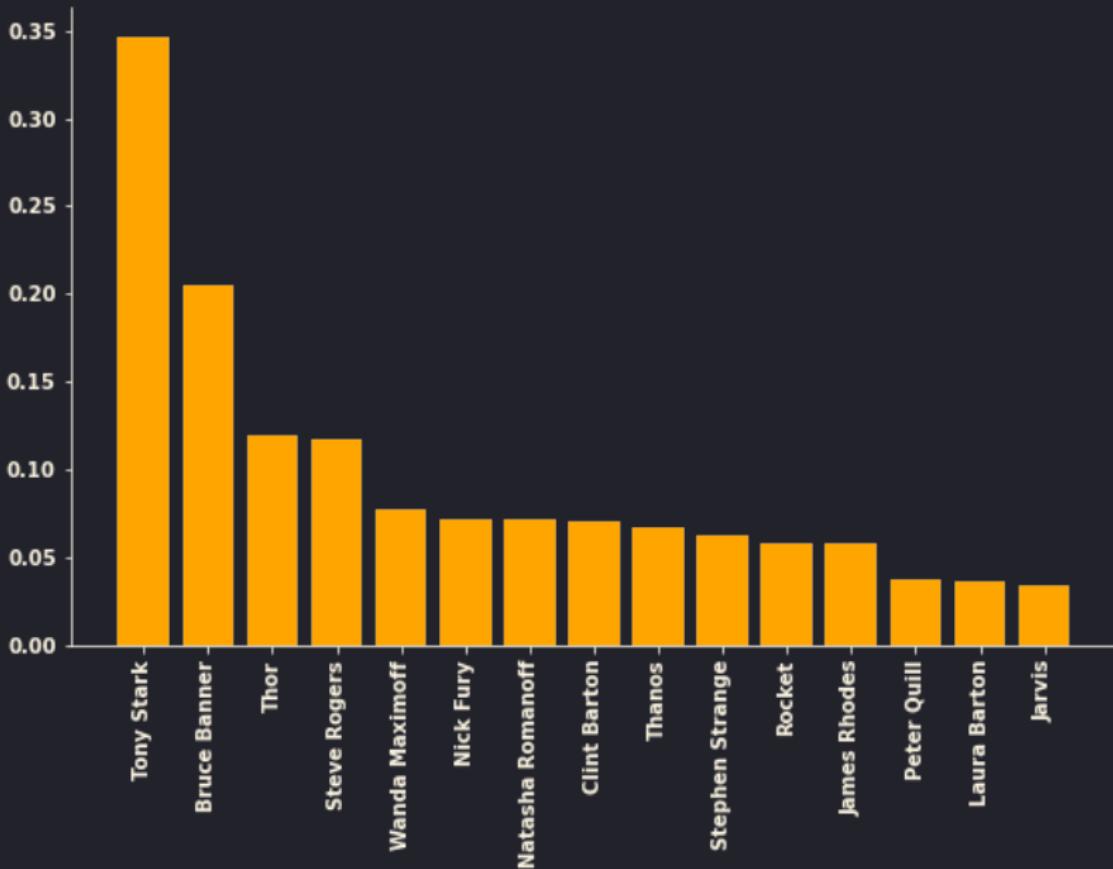
Who is the most connected character?

Center



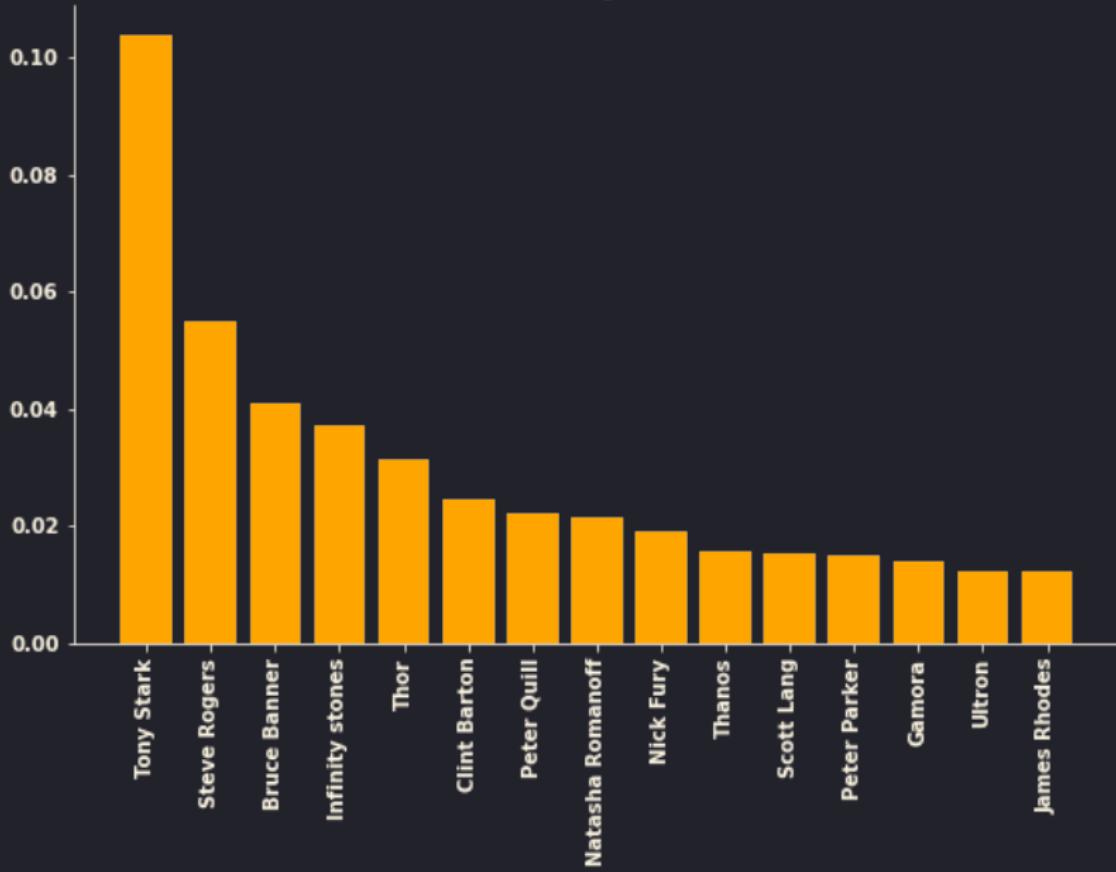
Which character stands between others the most?

In between



Which character is the most important based on Google?

Google



Who is the most important character in the MCU?

~~Who is the most important character in the MCU?~~

What do we mean by the most important?





@NikoletaGlyn

⌚ <https://github.com/Nikoleta-v3>

⌚ <https://github.com/Nikoleta-v3/avengers-analysis>

<https://medium.com/swlh/>

avengers-web-scraping-entity-extraction-and-network-graphs-in-python-ea6dc323eb7d