

Comunità Linguistiche

Un'analisi delle Lingue Simili nel Mondo
e dei Loro Contesti Geografici



Domanda di ricerca

1. Quali lingue sono simili ?

2 lingue sono simili se hanno più di 100 parole in comune

Otengo un grafo non orientato, dove i link tra due nodi (lingue) sono basati sulla quantità di parole condivise, rappresentate come pesi dei link

Data Gathering

«This repository contains all the words from every language that exists in the universe. Each word is separated by commas, allowing effortless integration into your project as a string array using basic comma-based splitting.»

abaca, abachi, abachista, abachiste, abaca, abariche, abarici, abarico, abasia, abaziale, abaziali, abazie, abbaca, abbacabbarono, abbacasse, abbacassero, abt acava, abbacavamo, abbacavano, abbacave menti, abbacchiamiento, abbacchiammo, at , abbacchiara, abbacchiare, abbacchiari iassero, abbacchiassi, abbacchiassimo, i, abbacchiato, abbacchiatura, abbacchi



1	abaca
2	abachi
3	abachista
4	abachiste
5	abachisti
6	abaco
7	abadessa
8	abadesse
9	abadia
10	abadie
11	abalietta
12	abarica
13	abariche
14	abarici



NETWORK ANALYSIS

Erdos-Renyi

--Number of nodes: 88

--Number of links: 1077

Statistiche della rete:

- Nodi: 88 Archi: 1104
- Densità: 0.288

Circa il 28% di tutte le possibili connessioni tra i nodi sono presenti nella rete

- Average Clustering Coefficient: 0.835

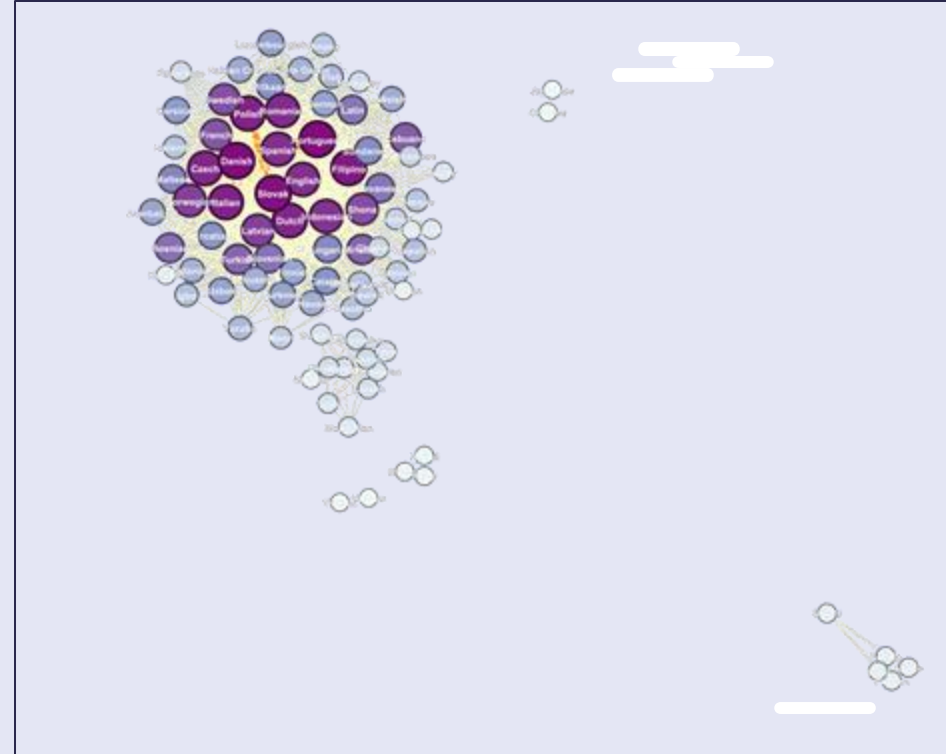
La rete è altamente clusterizzata.

- Reciprocità: /
- Assortatività: -0.172

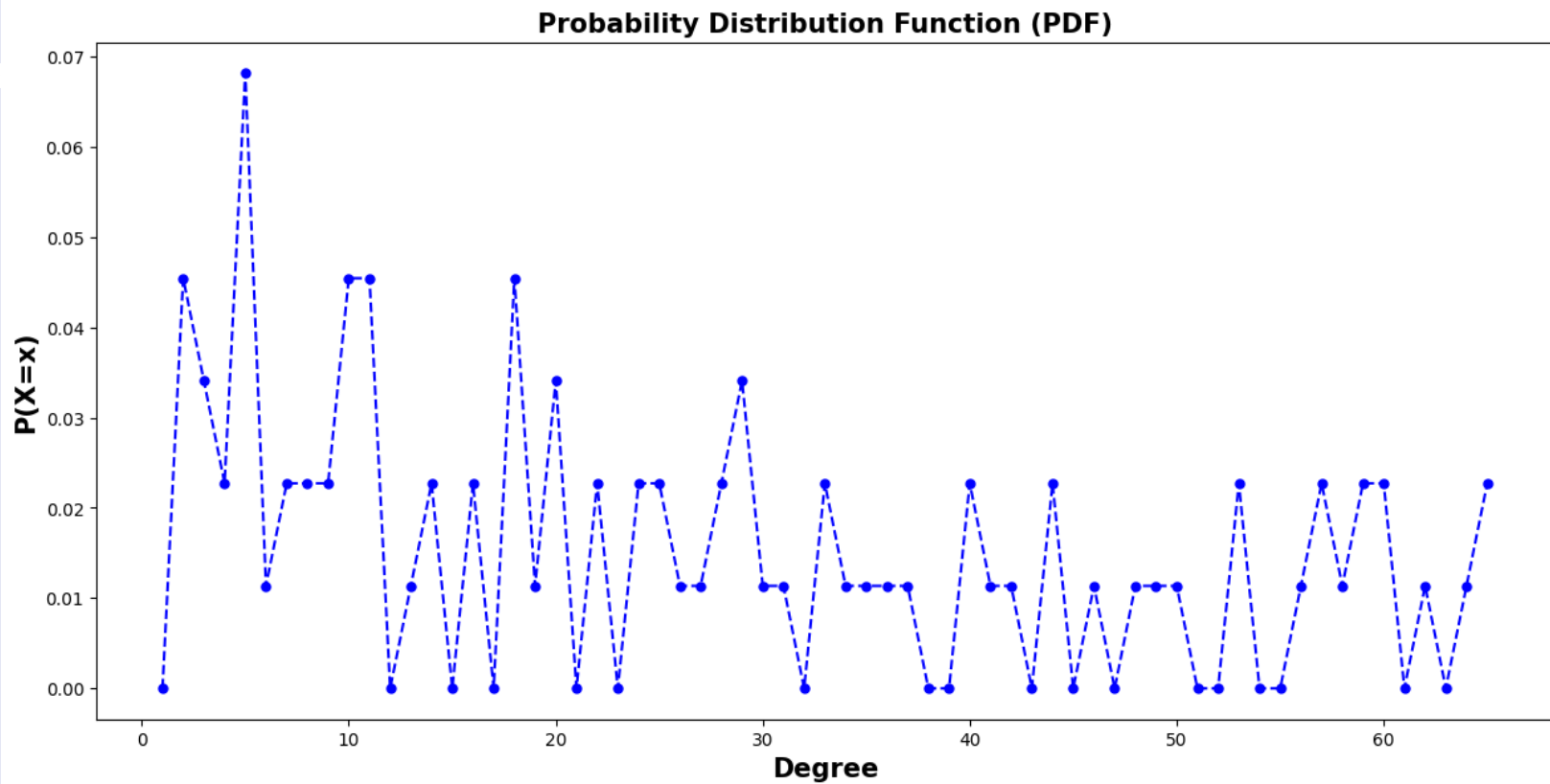
Valore negativo: i nodi con alto grado tendono ad associarsi con quelli con basso grado e viceversa.

Statistiche sul grado:

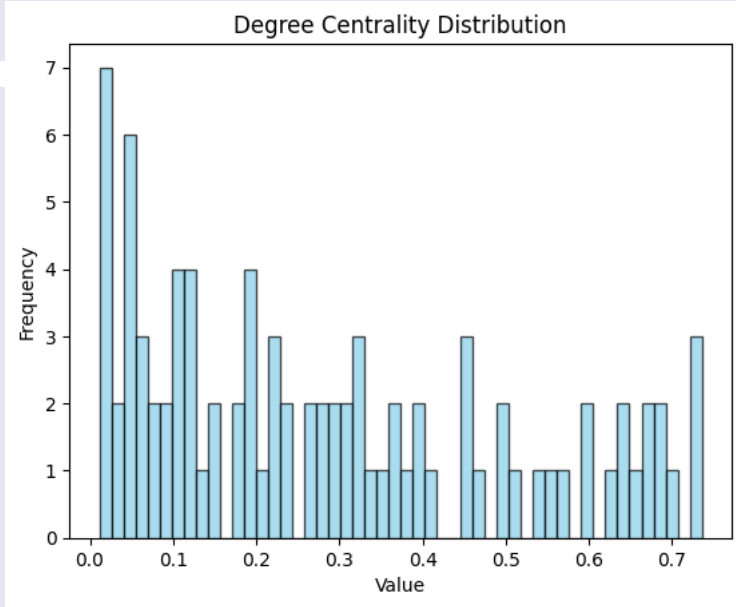
- Grado medio: 25.09
- Deviazione standard: 19.222
- Mediana: 21.0
- Grado Minimo: 1
- Grado Massimo: 64



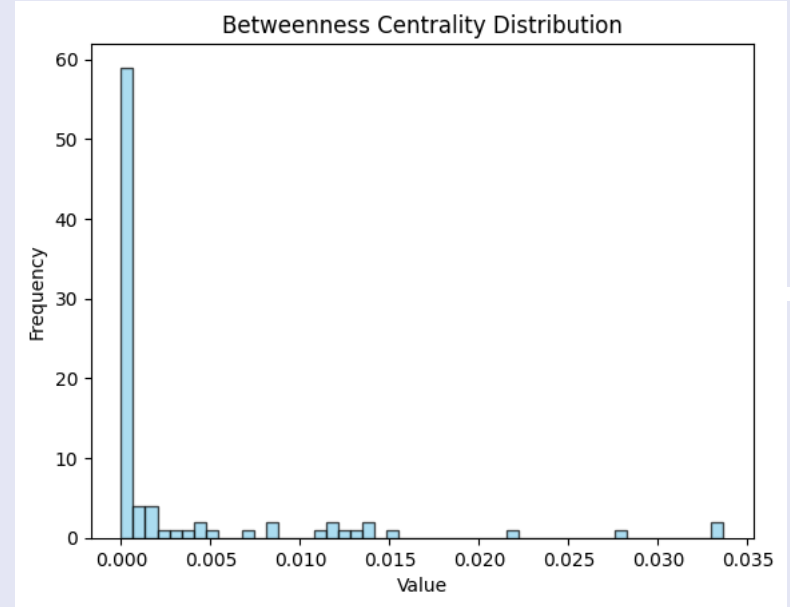
PDF



CENTRALITA'

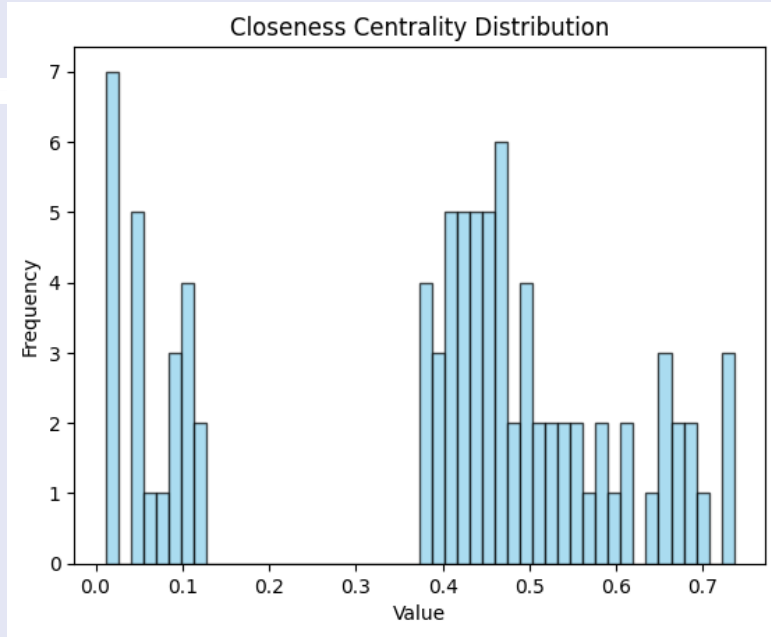


Nodes with the highest degree centrality (hubs):
['Danish', 'Portuguese'] 0.735632183908046

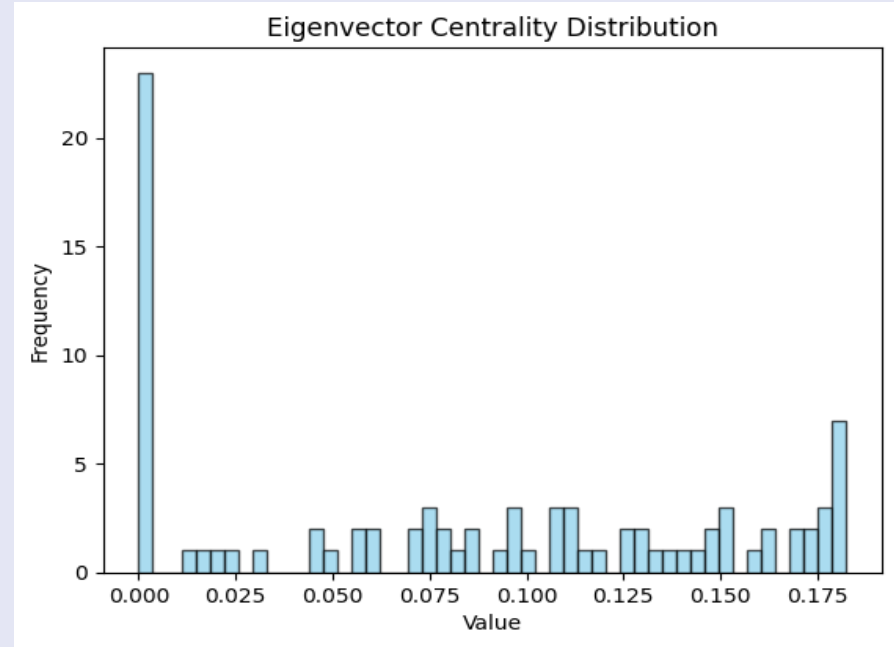


Nodes with the highest degree centrality (hubs):
['Danish', 'Portuguese'] 0.03365622032112665

CENTRALITA'

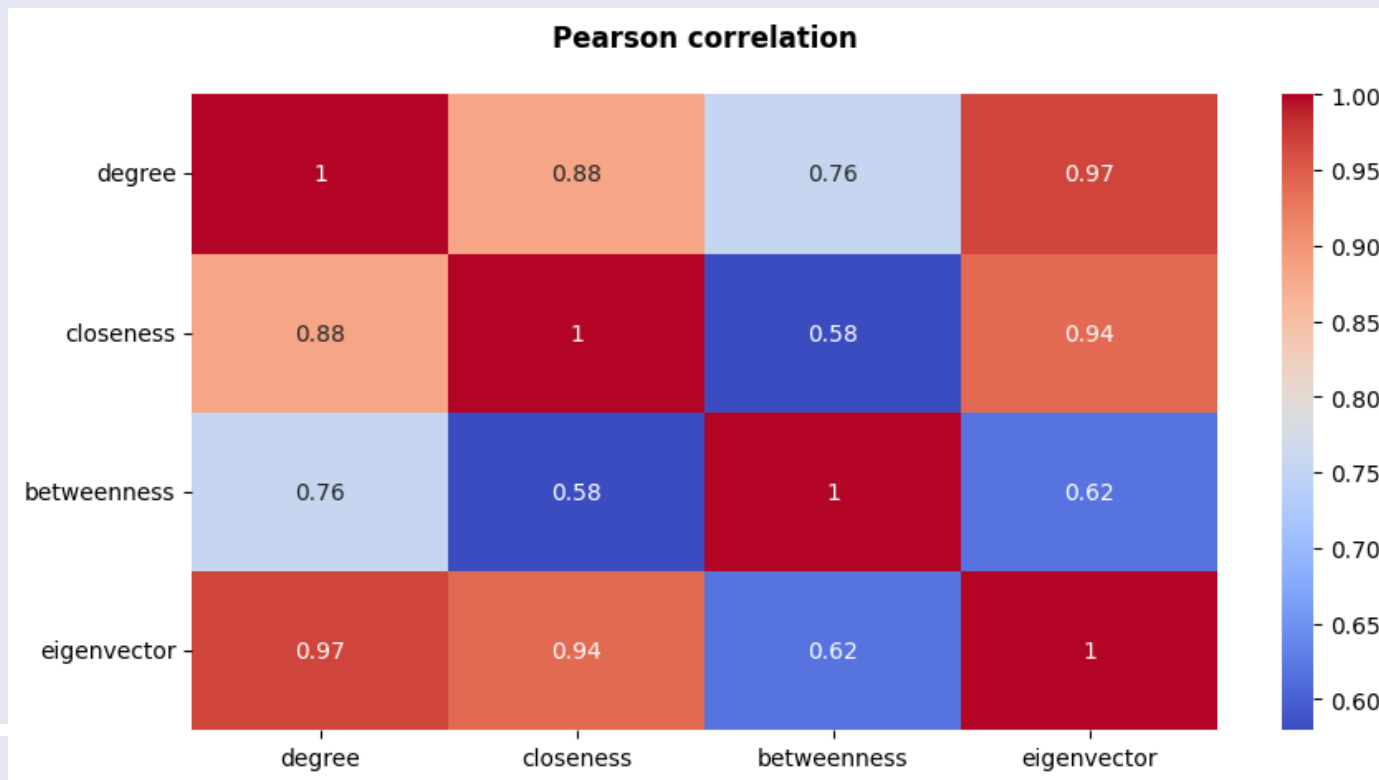


Nodes with the highest closeness centrality (hubs):
['Danish', 'Portuguese'] 0.735632183908046



Nodes with the highest eigenvector centrality (hubs):
['Danish', 'Portuguese'] 0.18247655755433131

CONFRONTO CENTRLITA'



LINGUE PIU' SIMILI TOP 15

('Czech', 'Slovak', {'weight': 101745})
('Portuguese', 'Spanish', {'weight': 85169})
('Polish', 'Slovak', {'weight': 52018})
('English', 'Slovak', {'weight': 45213})
('Danish', 'Slovak', {'weight': 44877})
('Danish', 'Norwegian', {'weight': 44430})
('Danish', 'Swedish', {'weight': 41988})
('Danish', 'English', {'weight': 41236})
('Filipino', 'Slovak', {'weight': 25774})
('English', 'Filipino', {'weight': 23767})
('Danish', 'Dutch', {'weight': 23283})
('Italian', 'Spanish', {'weight': 22998})
('Danish', 'Portuguese', {'weight': 22887})
('Persian', 'Urdu', {'weight': 22314})
('Italian', 'Portuguese', {'weight': 19880})

COMMUNITA'

G = GraphL

partition =

```
community_louvain.best_partition(G)
```

Number of communities: 9

Community 0: Number of nodes: 43

Density: 0.4008859357696567

Community 1: Number of nodes: 2

Density: 1.0

Community 2: Number of nodes: 10

Density: 0.7333333333333333

Community 3: Number of nodes: 2

Density: 1.0

Community 4: Number of nodes: 10

Density: 0.9111111111111111

Community 5: Number of nodes: 3

Density: 1.0

Community 6: Number of nodes: 5

Density: 1.0

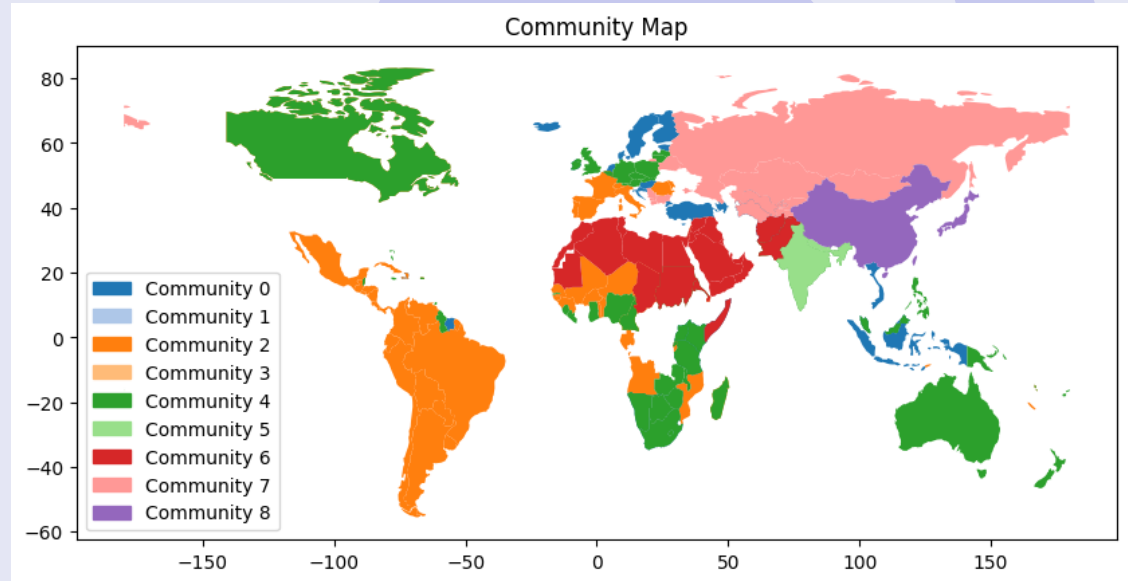
Community 7: Number of nodes: 11

Density: 0.8

Community 8: Number of nodes: 2

Density: 1.0

```
Community 7:  
Number of nodes: 11  
Density: 0.8  
Nodes in the community:  
- Belarus : Belarusian (Community 7)  
- Bulgaria : Bulgarian (Community 7)  
- Kazakhstan : Kazakh (Community 7)  
- Kyrgyzstan : Kyrgyz (Community 7)  
- North Macedonia : Macedonian (Community 7)  
- Tajikistan : Russian (Community 7)  
- Belarus : Russian (Community 7)  
- Kazakhstan : Russian (Community 7)  
- Kyrgyzstan : Russian (Community 7)  
- Uzbekistan : Russian (Community 7)  
- Turkmenistan : Russian (Community 7)  
- Russia : Russian (Community 7)  
- Tajikistan : Tajik (Community 7)  
- Mongolia : Mongolian (Community 7)  
- Bosnia and Herzegovina : Serbian (Community 7)  
- Kosovo : Serbian (Community 7)  
- Serbia : Serbian (Community 7)
```



Api usato: <https://restcountries.com/v3.1/lang/{language}>

SVILUPPI FUTURI

1. Controllare se i paesi che usano le lingue più simili sono vicini geograficamente
2. Perché 2 nodi che sono «lontani» sono collegati