



Progetto Advanced Data Science

Diego Cirillo 157176



Dataset - Darkweb

Crawling attraverso tor2web

Nodo -> dominio

Arco pesato -> numero di pagine su un dominio che puntano ad un altro dominio



Obiettivi

Esplorare la struttura del darkweb ed in particolare:

- Local Analysis
- Group Analysis
- Global Analysis



Numero di archi e di nodi

La rete analizzata è composta da 7178 nodi e 25104 archi



Archi con peso maggiore di 1

Sono stati rilevati 17091 archi con peso superiori a 1



Archi indiretti

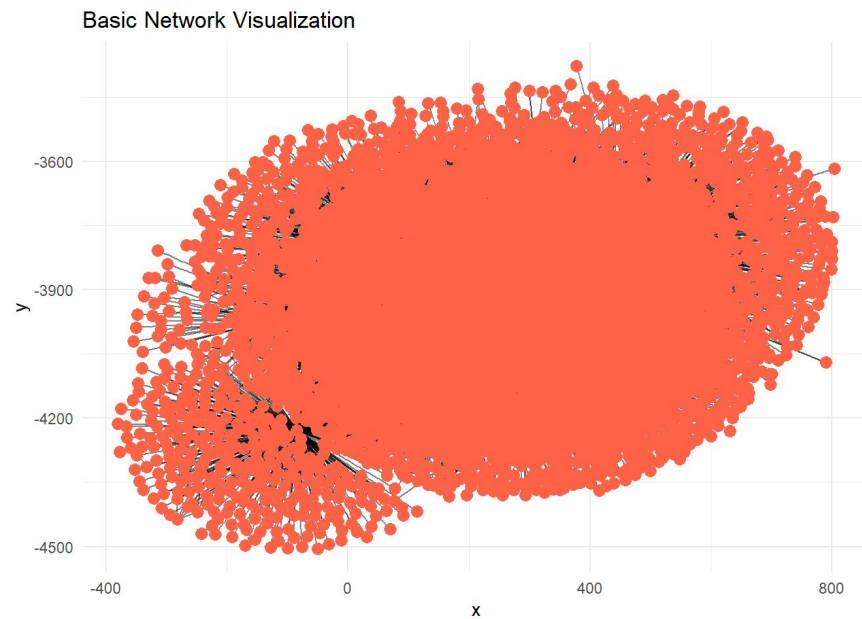
Sono stati rilevati 0 archi indiretti



Self loops

Sono stati rilevati 0 self loops

Visualizzazione del grafo





Local Analysis - 1

Ho calcolato le varie metriche di centralità ma ho deciso di concentrarmi sul grado dei nodi entranti ed uscenti

Sia la skewness dei nodi entranti (13.82) che quella dei nodi uscenti (45.84) è positiva, il che indica una asimmetria nei dati dove, graficamente, la maggior parte dei valori sono a sinistra ed è presente una lunga coda a destra



Local Analysis - 2

-Nodi entranti

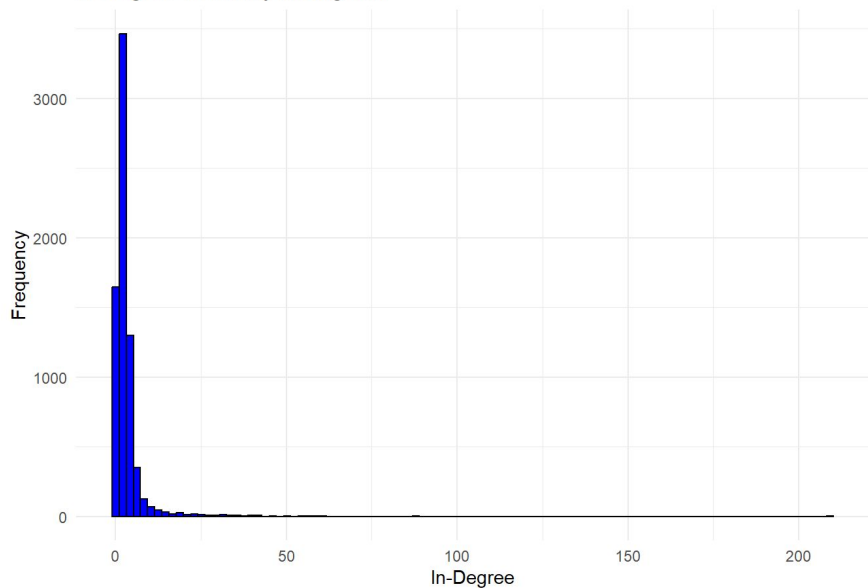
Min. 1st	Qu.	Median	Mean	3rd Qu.	Max.
1.000	2.000	3.000	3.497	4.000	209.000

-Nodi uscenti

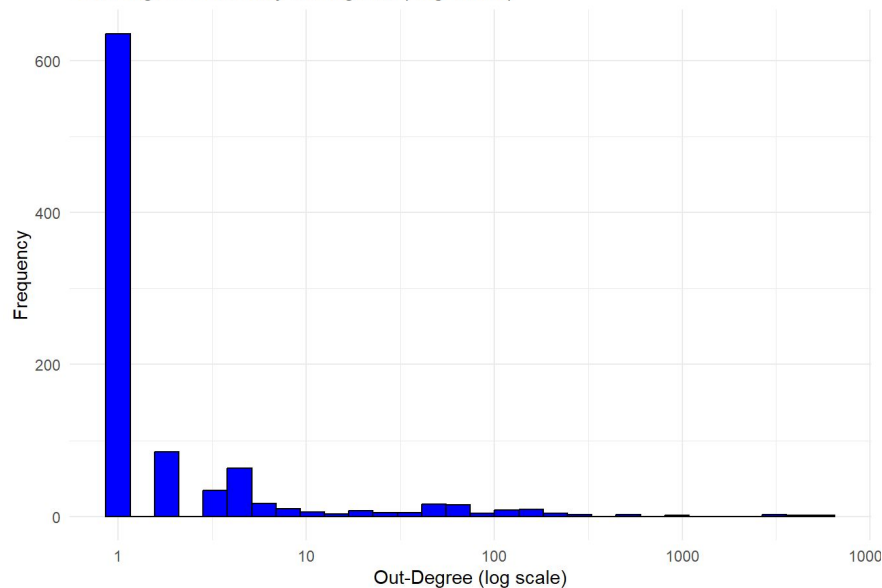
Min. 1st	Qu.	Median	Mean	3rd Qu.	Max.
0.000	0.000	0.000	3.497	0.000	5582.000

Local Analysis - 3

In-Degree Centrality Histogram



Out-Degree Centrality Histogram (Log Scale)





Local Analysis - 4

6242 nodi hanno 0 archi uscenti su 7178



Local Analysis - 5

Nodi con più archi uscenti

Id	Label	in_degree	out_degree	betweenness	closeness	eigen	pagerank
1 directoryvi6p...	NA	1	5582	138905.	0.000109	1	0.000114
2 visitorfi5kl7...	NA	18	4367	958759.	0.0000959	0.886	0.000141
3 skunksworke...dp...	NA	41	2769	762903.	0.0000828	0.591	0.000311
4 cratedvnn5z57...	NA	6	2758	169236.	0.0000803	0.570	0.000118
5 gxamjbnu7ukna...	NA	13	848	134093.	0.0000490	0.154	0.000196



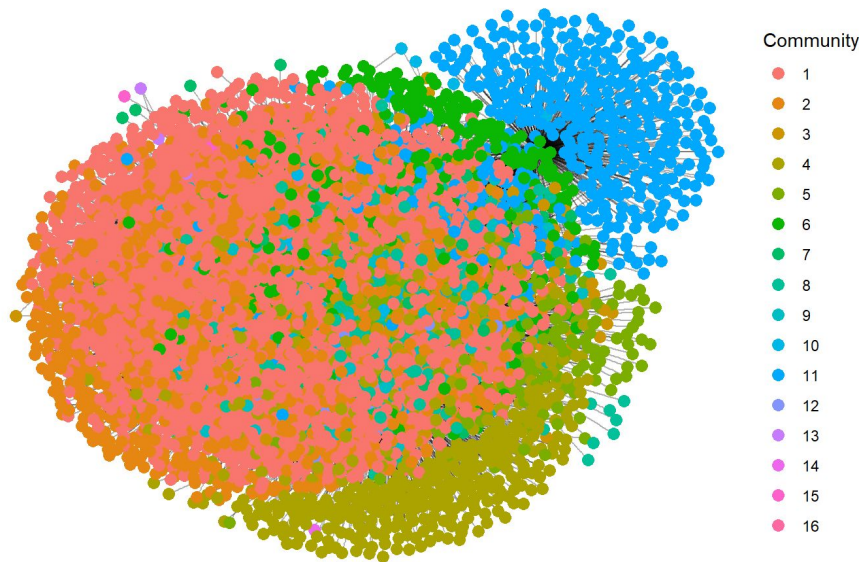
Group Analysis - 1

Ho calcolato le modularità di diversi algoritmi di community detection

	Method	Modularity
1.	Louvain	0.3629604
2.	Infomap	0.3397331
3.	Walktrap	0.3315492
4.	Fast Greedy	0.3093426
5.	Leading Eigenvector	0.2098401

Group Analysis - 2

Community Detection in Darkweb Network





Global Analysis - 1

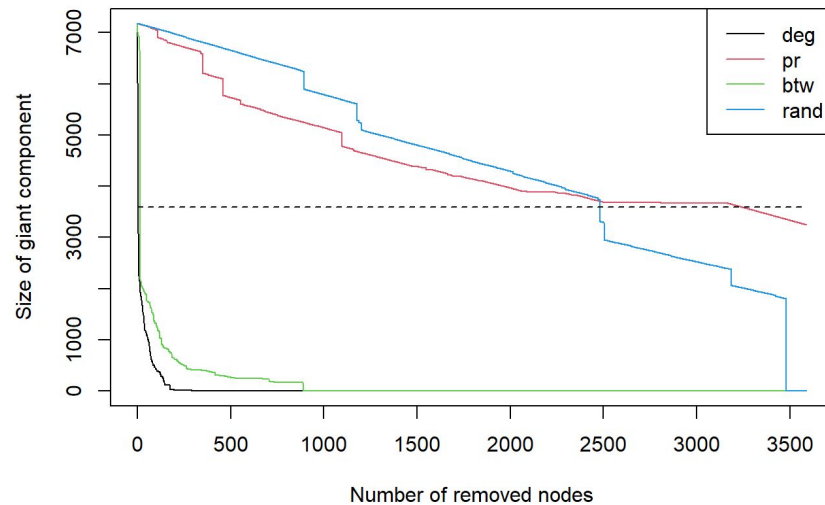
Numero di componenti: 1

Diametro: 10

Average path length 4.350535

Assortativity (degree): -0.3428776

Global Analysis 2

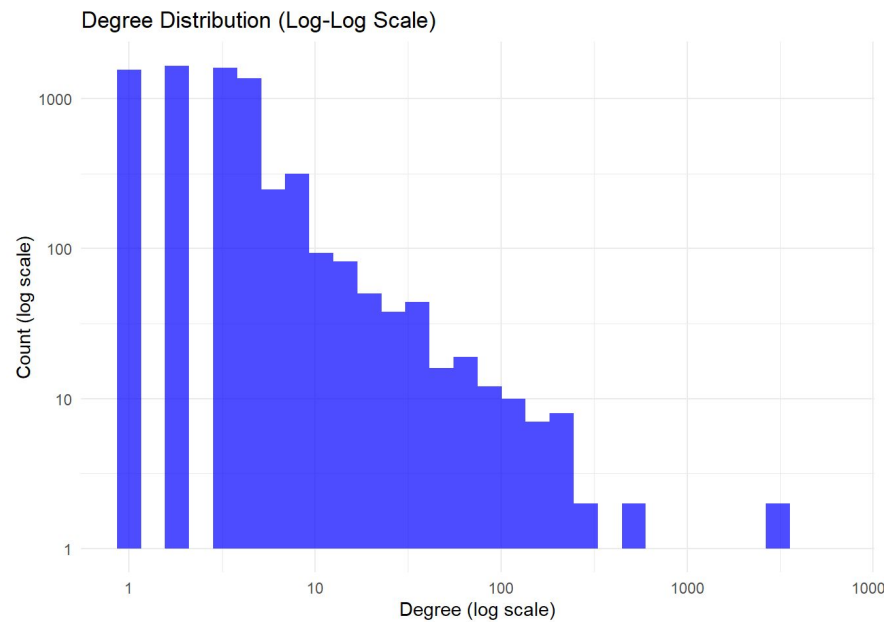




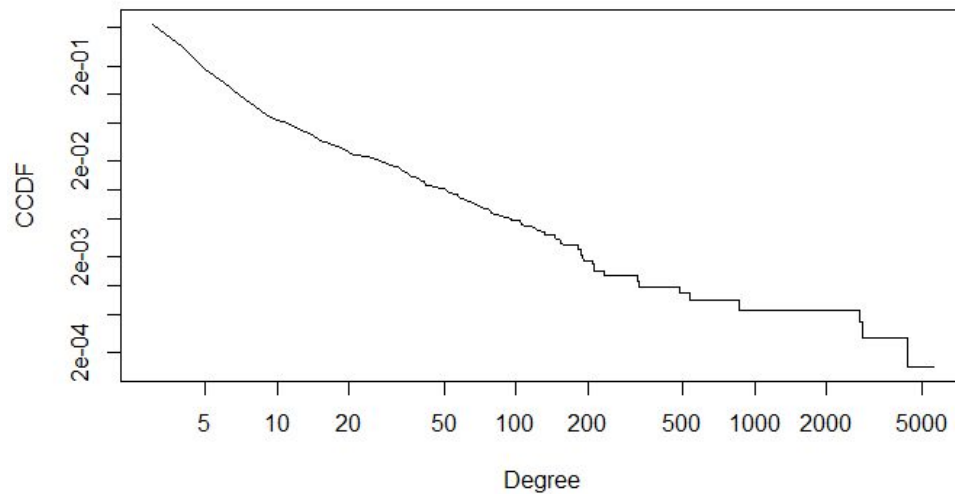
Verifica power law - 1

- 1) Grafo distribuzione LOGLOG
- 2) Grafo CCDF
- 3) Esponente power-law compreso fra 2 e 3 compresi

Verifica power law - 2



Verifica power law - 3





Verifica power law - 4

continuous = FALSE

alpha = 2.059684

xmin = 9

logLik = -1812.691

KS.stat = 0.03272644

Il range tipico per alpha per le power law è fra 2 e 3

Il comportamento power-law si applica ai nodi con degree ≥ 9 che è comune per reti reali

logLike basso significa che è significativamente un miglior fit di altre alternative

KS.stat basso implica un buon fit



Conclusioni

L'analisi locale ha evidenziato una distribuzione del grado fortemente sbilanciata

L'analisi di gruppo ha permesso di identificare comunità ben definite all'interno della rete

L'analisi globale ha confermato la natura scale-free della rete con distribuzione power-law