

# Structural Anatomy of the Dark Web: An Analysis of Hidden Service Connectivity

Victor A. Capa<sup>1</sup>

<sup>1</sup>*Institute for Cross-Disciplinary Physics and Complex Systems*

February, 2026

## Abstract

The Dark Web is frequently conceptualized as a distributed and resilient system. However, the topological structure of its underlying link network is a subject of debate. In this report, we analyze a network of Dark Web hidden services with 7,178 nodes and 24,879 edges. We employ canonical measures of network sciences, centrality analysis and community detection, comparing our findings against Erdős-Renyi and Barabási-Albert null models. Our results reveal a structure of ultra-small world combined with high local clustering. Crucially, we identify an extreme centralization hierarchy that exceeds theoretical scale-free predictions, alongside a strong disassortative mixing pattern. We analyze the roles of specific high-centrality nodes, identifying directory services that act as structural backbones. These findings suggest that the Dark Web's hyperlink structure is functionally oligarchic, relying strongly on central hubs, which renders the network potentially vulnerable to targeted attacks.

# 1 Introduction

The study of complex networks has changed our understanding of biological pathways, social interactions, or digital information ecosystems such as the World Wide Web. While the surface web has been extensively mapped, revealing scale-free properties and small-world architectures [Albert et al., 1999], the "Dark Web"—specifically hidden services accessible via anonymity networks like Tor—presents a unique case study. These networks are designed for anonymity and resistance to censorship, yet their navigational structure relies on hyperlinks created by users and administrators, which may not follow the decentralized ethos of the underlying protocol.

The primary objective of this report is to characterize the topological structure of a Dark Web hyperlink network. Unlike standard social networks where transitivity is high due to triadic closure, or biological networks shaped by evolutionary constraints, the Dark Web is shaped by a mix of secrecy and the need for *discoverability*. Specifically, we aim to answer the following research questions:

1. **Scale-Free Nature:** Does the Dark Web follow a power-law degree distribution similar to the surface web, or does it exhibit a different structural organization?
2. **Structural Cohesion:** How does its clustering and path length compare to random null models? Is it a "Small World"?
3. **Hierarchy and Control:** Is the network "democratic" or "oligarchic" in terms of centrality distribution? Do specific nodes dominate the information flow?

To address these questions, we perform a comprehensive structural analysis. We employ rigorous statistical fitting for heavy-tailed distributions [Clauset et al., 2009], quantify inequality/heterogeneity using the Gini coefficient, and validate the statistical significance of our findings by generating ensemble null models (Erdős-Rényi and Barabási-Albert).

# 2 Dataset and Methodology

## 2.1 Dataset Description

The dataset analyzed in this study represents a snapshot of the Dark Web structure extracted via Tor2web created between 2016 and 2017 [Griffith et al., 2017].

- **Nodes ( $V = 7,178$ ):** Represent distinct hidden services (websites ending in .onion).
- **Edges ( $E = 24,879$ ):** Represent directed hyperlinks from one site to another  $(i, j)$ , weighted by the number of pages on domain  $i$  linking to pages on domain  $j$ .

The network provided is fully connected, meaning there are no isolated components that require removal. Consequently, all analyses reported here correspond to the **full network**. Certain metrics require an undirected graph, so the network was symmetrized by treating all edges as bidirectional, while edge weights were disregarded as this study focuses purely on topological structure.

## 2.2 Computational Methods and Software

The analysis was conducted using **Python** (v3.10), leveraging the **NetworkX** library for graph algorithms. Rigorous statistical fitting of degree distributions was performed using the **powerlaw** package [Alstott et al., 2014]. Visualization was produced using **Gephi** (v0.10.1).

### 2.2.1 Degree Distribution Fitting

We tested the hypothesis that the degree distribution follows a Power Law  $P(k) \sim k^{-\alpha}$ . We employed the **Maximum Likelihood Estimation** (MLE) method. The lower bound cutoff ( $x_{min}$ ) was determined automatically by minimizing the Kolmogorov-Smirnov distance between the empirical data and the theoretical model. We compared the Power Law fit against Exponential and Log-normal alternatives using the Likelihood Ratio Test ( $R$ ), where a positive  $R$  with  $p < 0.05$  indicates a preference for the Power Law.

### 2.2.2 Centrality and Inequality Quantification

We computed centrality metrics to capture different aspects of node importance:

- **Degree Centrality:** Measuring local popularity.
- **PageRank:** Measuring influence based on the recursive authority of neighbors (damping factor  $\alpha = 0.85$ ).
- **Eigenvector & Katz Centrality:** Measuring global influence based on the quality of connections. While Eigenvector captures influence at a global equilibrium, Katz introduces a distance attenuation factor that penalizes long paths, allowing to weigh immediate neighbors more heavily than distant ones.
- **Betweenness Centrality:** Measuring the control over information flow (bridges).
- **Subgraph Centrality:** Measuring participation in closed loops of all lengths.

To quantify the global heterogeneity of the centrality distributions, we adapted the **Gini Coefficient** from econometrics to network science. For a given centrality measure  $X = \{x_1, \dots, x_V\}$ , the coefficient is defined as the relative mean absolute difference [Kunegis and Preusse, 2012]:

$$G = \frac{\sum_{i=1}^V \sum_{j=1}^V |x_i - x_j|}{2N^2\bar{x}} \quad (1)$$

where  $\bar{x}$  represents the mean centrality. A value of  $G \approx 0$  indicates a homogeneous topology (e.g., a random graph), while  $G \approx 1$  reveals a highly centralized, "winner-takes-all" hierarchy, allowing us to compare the empirical network against the theoretical baselines.

### 2.2.3 Community Detection and Visualization

Communities were detected using the **Greedy Modularity Optimization** algorithm [Clauset et al., 2004], chosen for its efficiency on networks of this size.

For visualization in Gephi, we employed the **ForceAtlas2** layout algorithm. This force-directed algorithm is specifically designed for scale-free networks. We used a Scaling factor of 200.0 to prevent node overlap and enabled the "Dissuade Hubs" option to push high-degree nodes apart, revealing the structural core. Nodes were colored by their modularity class and sized according to their PageRank score.

### 2.2.4 Null Models and Statistical Significance

To ensure that our structural findings are not artifacts of the network's size or density, we generated an ensemble of 10 realizations for two types of null models, constrained to have the exact same number of nodes ( $V$ ) and edges ( $E$ ) as the real network:

1. **Erdős-Rényi (ER):** A random graph model  $G(n, m)$  where edges are placed uniformly at random. This serves as a baseline for a structureless system.
2. **Barabási-Albert (BA):** A preferential attachment model. This serves as a baseline for a standard scale-free network without community structure or assortative constraints. The BA parameter  $m$  (number of new edges per node) was set to  $m = \text{round}(E/V) = 3$ . This choice aligns the null model's density with the empirical network, resulting in an expected average degree of  $\langle k \rangle_{\text{BA}} = 2m = 6$ , which closely approximates the network's average degree ( $\langle k \rangle_{\text{real}} \approx 6.9$ ).

For each metric, we report the  $Z$ -Score, defined as  $Z = \frac{X_{\text{real}} - \mu_{\text{null}}}{\sigma_{\text{null}}}$ . A magnitude  $|Z| > 2$  indicates the result is statistically significant ( $p < 0.05$ ).

## 3 Results

### 3.1 Topology

The network exhibits a low density, characteristic of sparse real-world graphs. However, the connectivity patterns reveal an ultra-small world phenomenon.

The average shortest path length of the network is  $a \approx 2.37$ . This is remarkably short, even when compared to the random null models ( $a_{\text{ER}} \approx 4.80$ ,  $a_{\text{BA}} \approx 4.18$ ). The fact that the real network allows for faster navigation than a random graph suggests an efficient core structure, likely driven by central directories. The Diameter of the network is  $d = 5$ , further confirming its compactness.

Simultaneously, the average Clustering Coefficient is  $C \approx 0.706$ . This is orders of magnitude higher than the random expectation ( $C_{\text{ER}} \approx 0.0009$ ,  $C_{\text{BA}}$ ), yielding a massive Z-Score. This indicates high local cohesion: neighbors of a node are highly likely to be connected to each other, forming dense cliques.

The topological features of the network compared with BA and ER null models are presented in Table 1.

Metric	Real Value	ER Mean	ER Z-Score	BA Mean	BA Z-Score
Avg. Clustering ( $C$ )	0.7061	0.0009	5614.13	0.0066	1198.08
Transitivity ( $T$ )	0.0044	0.0009	35.98	0.0035	4.49
Avg. Path Length ( $a$ )	2.3704	4.8053	-1050.92	4.1774	-101.71
Diameter ( $d$ )	5.0000	8.9000	-12.33	7.0000	-
Assortativity ( $r$ )	-0.4430	-0.0012	-95.48	-0.0429	-62.43
Modularity ( $Q$ )	0.3100	0.3419	-23.56	0.3933	-47.41

Table 1: Comparison of the Real Network against ER and BA Null Models ( $N = 10$ ).

### 3.2 Degree Distribution

Figure 1 illustrates the Complementary cumulative Distribution Function (CCDF) and the Probability Density Function (PDF) of the node degrees. The distribution exhibits a clear heavy tail, characteristic of scale-free networks.

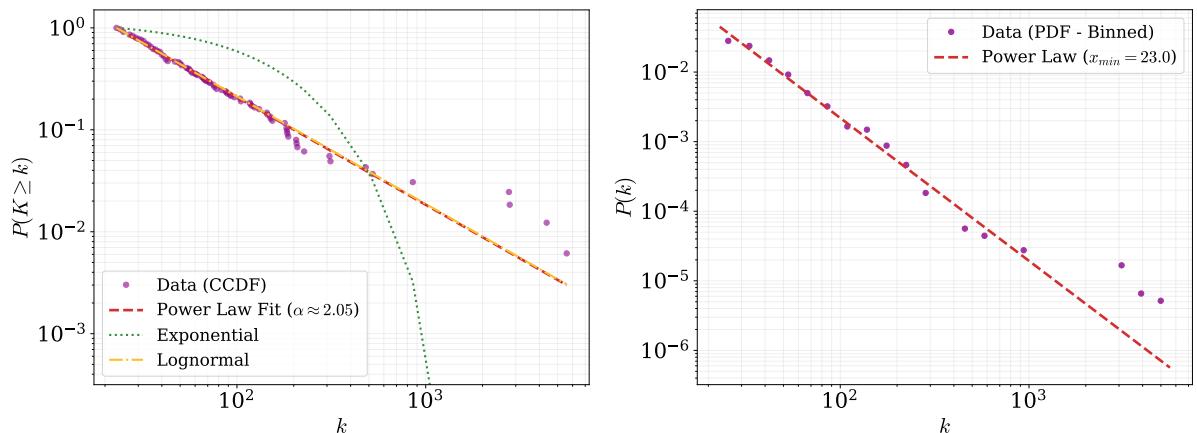


Figure 1: Degree distribution of the Dark Web network (log-log scale). Left: CCDF, the empirical data (purple) is compared against a power law (red), a log-normal (yellow) and an exponential fit (green). Right: PDF, empirical data (purple) compared with the power law fit (red).

The MLE method yielded a power-law exponent  $\alpha \approx 2.05$  with a lower bound cutoff at  $x_{min} = 23.0$ . We compared this fit against alternative distributions using the Likelihood Ratio Test:

- **Power Law vs. Exponential:**  $R = 157.54$ ,  $p = 7.05 \times 10^{-5}$ . The positive  $R$  and low  $p$ -value strongly reject the Exponential hypothesis in favor of the Power Law.
- **Power Law vs. Log-normal:**  $R = 0.0035$ ,  $p = 0.69$ . The test is inconclusive, suggesting that while the tail is heavy, a Log-normal distribution could also explain the data.

### 3.3 Centrality and Hierarchy

One of the most significant findings of this study is the extreme level of centralization. We quantified this using the Gini coefficient applied to centrality scores.

Metric	Real Value	ER Mean	ER Z-Score	BA Mean	BA Z-Score
HG (Degree)	0.6954	0.2120	386.55	0.3758	221.09
HG (PageRank)	0.5940	0.1769	416.82	0.3290	176.30
HG (Eigenvector)	0.3440	0.2604	34.43	0.5977	-18.84

Table 2: Heterogeneity (Gini Coefficient) comparison with the average coefficient for ten BA and 10 ER.

As shown in Table 2, the real network is almost twice as unequal as the Barabási-Albert model. While the BA model generates hubs through preferential attachment, the Dark Web network exhibits a "winner-takes-all" dynamic that is far more aggressive.

#### 3.3.1 Analysis of Top Nodes

Table 3 highlights the top 5 nodes ranked by different centrality measures. The node labeled `directoryvi6plzm` is the undisputed dominant hub, ranking no.1 in Degree, PageRank, Eigenvector, Betweenness, Closeness, Katz, and Subgraph centrality.

Rank	Degree	PageRank	Eigenvector	Betweenness
1	directoryvi6plzm (0.78)	directoryvi6plzm (0.11)	directoryvi6plzm (0.45)	directoryvi6plzm (0.48)
2	visitorfi5kl7q7i (0.61)	visitorfi5kl7q7i (0.08)	visitorfi5kl7q7i (0.39)	visitorfi5kl7q7i (0.23)
3	skunksworkedp2cg (0.39)	cratedvnn5z57xhl (0.05)	skunksworkedp2cg (0.26)	skunksworkedp2cg (0.15)
4	cratedvnn5z57xhl (0.38)	skunksworkedp2cg (0.05)	cratedvnn5z57xhl (0.25)	cratedvnn5z57xhl (0.14)
5	gxamjbnu7uknahng (0.12)	gxamjbnu7uknahng (0.02)	gxamjbnu7uknahng (0.07)	gxamjbnu7uknahng (0.10)

Table 3: Top 5 Nodes by Centrality Measures.

This node consistency indicates a robust "backbone". However, we observe interesting deviations. For instance, node `cratedvnn5z57xhl` ranks 4th in Degree (0.38) but jumps to 3rd in PageRank (0.05), suggesting that although it has fewer connections than

`skunksworkedp2cg`, the connections it does have come from highly important neighbors. Conversely, `skunksworkedp2cg` has a high degree but lower PageRank influence.

Furthermore, the Subgraph Centrality values for these top nodes are astronomical (order of  $10^{48}$ ), confirming their participation in an immense number of closed loops, which reinforces the high clustering coefficient found globally.

A complete list of the 25 top nodes by different centrality measures are presented in Appendix A.

### 3.4 Mixing Patterns and Community Structure

The network is strongly **disassortative**, with a degree assortativity coefficient of  $r \approx -0.44$ . This contrasts sharply with the neutral mixing found in the null models ( $r \approx 0$ ). This negative correlation implies that high-degree nodes tend to connect with low-degree nodes, a signature typical of technological networks (e.g., servers connecting to clients) or directory-based structures, rather than social networks where "rich nodes" tend to associate with other "rich nodes".

The Modularity ( $Q \approx 0.31$ ) is significantly higher than in random models ( $Q \approx 0$ ), confirming the existence of distinct functional communities.

## 4 Discussion

### 4.1 The Star-Like Directory Structure

The results present a topological paradox: a very high Clustering Coefficient (0.70) combined with a very low Transitivity (0.004). In many social networks, these two metrics correlate. However, in this Dark Web dataset, this discrepancy points to a structural motif: **local stars**.

Small groups of nodes form dense clusters (high clustering), but these clusters are linked via central hubs that do not close triangles with each other (low global transitivity). Supported by the strong disassortativity ( $r = -0.44$ ), this confirms that the network is organized around central authorities (directories) that point to many sites, but those sites do not link back to the directory or to each other.

### 4.2 Beyond Scale-Free: The Oligarchy

The comparison with the Barabási-Albert (BA) model is particularly important in this case. The BA model assumes that popularity is proportional to current degree (rich gets richer). However, the Gini coefficients for the Dark Web network (0.69) are significantly higher than the BA baseline (0.37).

This implies that standard preferential attachment is insufficient to explain the concentration of power in the Dark Web. There are likely external mechanisms that artificially inflate the degree of specific nodes, creating an artificial oligarchy that dominates navigation.

We also note the anomaly in Eigenvector Centrality heterogeneity (HG Eigenvector). The Dark web network (0.34) is *less* unequal than the BA model (0.60). In BA models, hubs connect to other hubs (rich club effect), amplifying eigenvector scores. In our disassortative Dark Web, hubs avoid connecting to other hubs, which distributes the Eigenvector influence more evenly across the network, despite the high degree inequality.

### 4.3 Network Visualization

The visualization in Figure 2 corroborates these statistical findings.

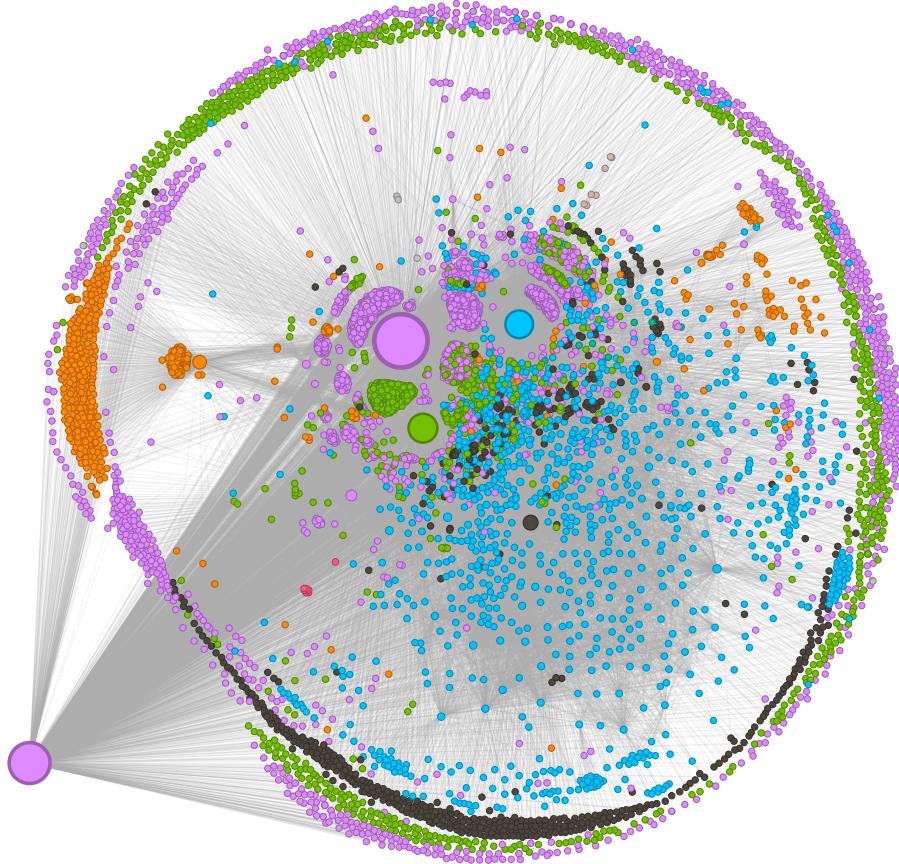


Figure 2: ForceAtlas2 visualization of the network generated in Gephi. Nodes are colored by modularity class (partition) and sized according to their PageRank score. The structure highlights the "star-like" hubs connecting the periphery. Note the presence of bridge nodes (high-ranking nodes located between distinct communities) which facilitate the low average path length.

The visual layout reveals a core-periphery structure where large hubs (high PageRank) act as bridges between distinct colored communities, validating the high modularity and

centralization scores.

## 5 Conclusion

This report analyzed the structural properties of a Dark Web hyperlink network consisting of 7,178 nodes. Our analysis leads to three main conclusions:

1. **Efficiency:** The network is an ultra-small world ( $a = 2.37$ ), allowing for information propagation speeds that exceed theoretical random models and standard scale-free networks.
2. **Organization:** It is not a random mesh but a highly structured, disassortative system likely organized around central directories. The Power Law fit ( $\alpha \approx 2.05$ ) confirms its heavy-tailed nature.
3. **Vulnerability:** The extreme centralization suggests that while the network is efficient, **it is likely fragile to targeted attacks on its main hubs** (such as `directoryvi6plzm`), contrary to the typical assumption of Dark Web resilience.

Future work could focus on the temporal evolution of these hubs to determine if this oligarchy is stable or shifting over time, and also regarding on the direction and weights of the edges.

## References

- Réka Albert, Hawoong Jeong, and Albert-László Barabási. Internet: Diameter of the world-wide web. *Nature*, 401(6749):130–131, 1999.
- Jeff Alstott, Ed Bullmore, and Dietmar Plenz. powerlaw: a python package for analysis of heavy-tailed distributions. *PloS one*, 9(1):e85777, 2014.
- Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.
- Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.
- Virgil Griffith, Yang Xu, and Carlo Ratti. Graph theoretic properties of the darkweb, 2017. URL <https://arxiv.org/abs/1704.07525>.
- Jérôme Kunegis and Julia Preusse. Fairness on the web: Alternatives to the power law. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 175–184. ACM, 2012.

## A Top 25 nodes by centrality

Table 4: Degree, PageRank, and Betweenness.

Rank	Degree	PageRank	Betweenness
1	directoryvi6plzm	0.7778	directoryvi6plzm
2	visitorfi5kl7q7i	0.6092	visitorfi5kl7q7i
3	skunksworkedp2cg	0.3882	cratedvnn5z57xhl
4	cratedvnn5z57xhl	0.3847	skunksworkedp2cg
5	gxamjbnu7uknahng	0.1190	gxamjbnu7uknahng
6	torvps7kzis5ujfz	0.0730	zlal32teyptf4tvi
7	zlal32teyptf4tvi	0.0670	torvps7kzis5ujfz
8	hwikis25cffertqe	0.0436	hwikis25cffertqe
9	w363zoq3ylux5rf5	0.0431	w363zoq3ylux5rf5
10	wikitorcwogtsifs	0.0316	fhostingesps6bly
11	fhostingesps6bly	0.0291	auutwvpt2zktxwng
12	y4yhci7273s2yeqk	0.0288	wikitorcwogtsifs
13	auutwvpt2zktxwng	0.0287	hdwikicorlcisiy
14	hdwikicorlcisiy	0.0261	y4yhci7273s2yeqk
15	mijpsrtgf54l7um6	0.0258	wikiwarixvouhwyn
16	wikiwarixvouhwyn	0.0256	nqigfqrxnkwcmiq
17	zqktlw4fecvo6ri	0.0255	mijpsrtgf54l7um6
18	32b5oz2bbtn6gqj3	0.0249	32b5oz2bbtn6gqj3
19	nqigfqrxnkwcmiq	0.0249	zqktlw4fecvo6ri
20	ntcixulmms4275vi	0.0215	soupkso3la22ltl3
21	kpvz7ki2lzvnwve7	0.0210	kpvz7ki2lzvnwve7
22	soupkso3la22ltl3	0.0209	ntcixulmms4275vi
23	wikihiddkz5w3hfg	0.0202	n6pbizsbykwxmydz
24	torwikignoueupfm	0.0201	godnotaba36dsabv
25	n6pbizsbykwxmydz	0.0183	hiddeninasdyooih

Table 5: Eigenvector and Closeness Centrality.

Rank	Eigenvector	Closeness
1	directoryvi6plzm	0.4463
2	visitorfi5kl7q7i	0.3883
3	skunksworkedp2cg	0.2562
4	cratedvnn5z57xhl	0.2493
5	gxamjbnu7uknahng	0.0662
6	torvps7kzis5ujfz	0.0541
7	w363zoq3ylux5rf5	0.0480
8	hwikis25cffertqe	0.0468
9	zqktlwi4fecvo6ri	0.0382
10	wikitorcwogtsifs	0.0381
11	y4yhci7273s2yeqk	0.0381
12	mijpsrtgf54l7um6	0.0347
13	32b5oz2bbtn6gqj3	0.0338
14	wikiwarixvouhwyn	0.0330
15	auutwvpt2zktxwng	0.0329
16	nqigfqrxnkwcmiq	0.0329
17	kpynyvym6xqi7wz2	0.0321
18	hdwikicorldcisiy	0.0311
19	kpvz7ki2lzvnwve7	0.0309
20	soupkso3la22ltl3	0.0305
21	ntcixulmms4275vi	0.0305
22	wikihiddkz5w3hfg	0.0296
23	fhostingesps6bly	0.0291
24	torwikignoueupfm	0.0280
25	dirnxxdraygbifgc	0.0257

Table 6: Top 25 Nodes: Subgraph Centrality and Katz Centrality.

<b>Rank</b>	<b>Subgraph</b>		<b>Katz</b>	
1	directoryvi6plzm	$1.17 \times 10^{49}$	directoryvi6plzm	0.4268
2	visitorfi5kl7q7i	$8.82 \times 10^{48}$	visitorfi5kl7q7i	0.3644
3	skunksworkedp2cg	$3.84 \times 10^{48}$	skunksworkedp2cg	0.2400
4	cratedvnn5z57xhl	$3.64 \times 10^{48}$	cratedvnn5z57xhl	0.2343
5	gxamjbnu7uknahng	$2.57 \times 10^{47}$	gxamjbnu7uknahng	0.0663
6	torvps7kzis5ujfz	$1.71 \times 10^{47}$	torvps7kzis5ujfz	0.0522
7	w363zoq3ylux5rf5	$1.35 \times 10^{47}$	w363zoq3ylux5rf5	0.0443
8	hwikis25cffertqe	$1.28 \times 10^{47}$	hwikis25cffertqe	0.0434
9	zqktlwi4fecvo6ri	$8.53 \times 10^{46}$	wikitorcwogtsifs	0.0354
10	wikitorcwogtsifs	$8.51 \times 10^{46}$	y4yhci7273s2yeqk	0.0350
11	y4yhci7273s2yeqk	$8.49 \times 10^{46}$	zqktlwi4fecvo6ri	0.0349
12	mijpsrtgf54l7um6	$7.04 \times 10^{46}$	mijpsrtgf54l7um6	0.0320
13	32b5oz2bbtn6gqj3	$6.69 \times 10^{46}$	32b5oz2bbtn6gqj3	0.0313
14	wikiwarixvouhwyn	$6.38 \times 10^{46}$	auutwvpt2zktxwng	0.0312
15	auutwvpt2zktxwng	$6.33 \times 10^{46}$	wikiwarixvouhwyn	0.0309
16	nqigfqrxnkwcmiq	$6.32 \times 10^{46}$	nqigfqrxnkwcmiq	0.0307
17	kpynyvym6xqi7wz2	$6.02 \times 10^{46}$	kpynyvym6xqi7wz2	0.0294
18	hdwikicorldcisiy	$5.65 \times 10^{46}$	hdwikicorldcisiy	0.0294
19	kpvz7ki2lzvnwve7	$5.59 \times 10^{46}$	kpvz7ki2lzvnwve7	0.0288
20	soupkso3la22ltl3	$5.44 \times 10^{46}$	soupkso3la22ltl3	0.0284
21	ntcixulmms4275vi	$5.43 \times 10^{46}$	ntcixulmms4275vi	0.0283
22	wikihiddkz5w3hfg	$5.14 \times 10^{46}$	fhostingesps6bly	0.0282
23	fhostingesps6bly	$4.97 \times 10^{46}$	wikihiddkz5w3hfg	0.0276
24	torwikignoueupfm	$4.59 \times 10^{46}$	torwikignoueupfm	0.0261
25	dirnxxdraygbifgc	$3.87 \times 10^{46}$	n6pbizsbykwxmydz	0.0244