Graph Theoretic Properties of the Darkweb

Virgil Griffith¹, Yang Xu¹, and Carlo Ratti²

¹Urban Mobility Laboratory, Singapore-MIT Alliance for Research and Technology, Singapore, 138602
²SENSEable City Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

April 27, 2017

Abstract

We collect and analyze the darkweb (a.k.a. the “onionweb”) hyperlink graph. We find properties highly dissimilar to the well-studied world wide web hyperlink graph; for example, our analysis finds that >87% of darkweb sites never link to another site. We compare our results to prior work on world-wide-web and speculate about reasons for their differences. We conclude that in the term “darkweb”, the word “web” is a connectivity misnomer. Instead, it is more accurate to view the darkweb as a set of largely isolated dark silos.

1 Introduction

Graph theory has long been a favored tool for analyzing social relationships [1] as well as quantifying engineering properties such as searchability [2]. For both reasons, there have been numerous graph-theoretic analyses of the World Wide Web (www) from the seminal [3 –7] to the modern [8]. Motivated by curiosity, we repeat these analyses for the notorious “darkweb”. The darkweb is sometimes loosely defined as “anything seedy on the Internet”, but we define the darkweb strictly, as simply all domains underneath the “.onion” pseudo-top-level-domain [9], i.e., we define the darkweb to be synonymous with the onionweb.

The darkweb is infamously mysterious, and any insight into it for both harnessing it or informing social policy is welcome. As far as we know we are among [10,11] the first to analyse the darkweb through graph theory. We analyze the darkweb foremost because it’s an interesting unexplored dataset, and second because, on the face of it, the world-wide-web and the darkweb are immensely similar—both are browsed through a standard web-browser. Therefore any differences between the structure of the darkweb versus the www likely indicate something about the societies inhabiting each. For comparable work on the world-wide-web, see especially [12] and [13].

Newcomers are often confused about the properties of the onionweb. Put simply, the onionweb is subset of the web where websites are identified not by a human-readable hostname (e.g., yahoo.com) or by a IP number (e.g., 206.190.36.45), but by a randomly generated 16-character address (specifically, a hash fingerprint). Each website can be accessed via its hash, but it is very difficult to learn the IP number of a website from its hash. The “dark” part means that, even though the website is accessible (via its hash), its IP number remains hidden. Without an IP number, it is exceedingly difficult to trace the geographical origin of a communication.

2 Data Collection

Crawling the darkweb is not much harder than crawling the regular web. In our case, we crawled the darkweb through the popular tor2web proxy onion.link. Onion.link does all of the interfacing with Tor, and one can crawl all darkweb pages not requiring a login simply by setting a standard crawler to specifically crawl the domain *.onion.link. Darkweb pages are written in the same HTML language as the regular web, and we
crawled onion.link using the commercial service scrapinghub.com. Starting from two popular lists of darkweb sites\footnote{\url{http://directoryvi6plzm.onion} and \url{https://ahmia.fi/onions/}} we accessed each page and crawled all linked pages using breadth-first search.

Most graph-theoretic analyses of the www consider the page-level, i.e., each node is an individual URL. The page-level description is important for engineering issues like crawlability, and one could certainly do a page-level analysis of the darkweb, however, it’s unclear how informative this level of description is on social behavior—the page-level graph is influenced more by the various choices of content management system than anything social.

Because of this, we follow \cite{13} to aggregate by second-level domain (for the onionweb the second-level domain is equivalent to \cite{13}’s “pay-level domain”). This means that links within a second-level domain are ignored as socially irrelevant self-connections. In our darkweb graph, each vertex is a domain and every directed edge from \( u \rightarrow v \) means there exists a page within domain \( u \) linking to a page within domain \( v \). The weight of the edge from \( u \rightarrow v \) is the number of pages on domain \( u \) linking to pages on domain \( v \).

The Tor Project Inc., the creators and custodians of the darkweb, maintain basic darkweb statistics \cite{14}. According to them, there are \(~60,000\) distinct, active .onion addresses. However, in our analysis we found merely 7,178 active .onion domains. We attribute this high-discrepancy to various messaging services—particularly TorChat \cite{15}, Tor Messenger \cite{16}, and Ricochet \cite{17}. In all of these services, each user is identified by a unique .onion domain.

The darkweb has a notoriously high attrition rate—sites regularly appear and disappear. This creates a substantial confound as there will be links to .onion domains that no longer exist. To account for this, we only include sites which responded. I.e., if we discover a link to a page on domain \( v \), but domain \( v \) could not be reached after > 10 attempts across November 2016–February 2017, we delete node \( v \) and all edges to node \( v \). In our analysis, before pruning nonresponding domains we found a graph of 13,117 nodes and 39,283 edges. After pruning, we have a graph of 7,178 nodes and 25,104 edges (55\% and 64\% respectively). In all results, we refer to this graph pruned of nonresponding domains as simply “the darkweb graph”.

## 3 Graph-theoretic Results

The darkweb graph consists of exactly one weakly connected component (WCC\footnote{A Weakly Connected Component is defined as a subset of a graph where every node in the subset, ignoring edge direction, can reach every other node in the subset. Finding exactly one weakly connected component is entailed by our original list of seed domains itself being part of the darkweb.} of 7,178 nodes and 25,104 edges.

The first step is to quantitatively analyze the degree distributions, which we do in Figure 1. Figures 1a and 1b vaguely resemble a powerlaw, but the \texttt{plfit} tool using the methods from \cite{18} report there’s not enough orders of magnitude in the data to affirm or deny a powerlaw. In Figure 1a we see that \(~30\%\) of domains have exactly one incoming link—of which 62\% come from one of the five largest out-degree hubs. Intrigued by the impact of these five large out-degree hubs, we found that almost all nodes (78\%) received a connection from at least one of them. For those curious about these specifics of these high out-degree hubs, the top fifteen hubs are listed in the Appendix (Table 2). In Figure 1b we see our most striking feature—that fully 87\% of sites do not link to any other site. This sole feature has immense impact on all graph-theoretic measures. In Figure 1c we see that \( > 98\% \) of domains are tied for the lowest pagerank \cite{19} of 2.095\texttimes{}4. In Figure 1d we see that when a site does link another, 32\% of the time it’s only a single page linking out. All together, the onionweb is a sparse hub-and-spoke place.

Our second result is a bow-tie decomposition from \cite{21}, shown in Figure 2. A bowtie decomposition divides the nodes of a directed graph into six disjoint parts, they are:

1. CORE — Also called the “Largest Strongly Connected Component”. It is defined as the largest subset of nodes such that there exists a directed path (both directions—from \( u \rightarrow \cdots \rightarrow v \) as well as \( v \rightarrow \cdots \rightarrow u \)) between every pair of nodes in the set.
2. IN — The set of nodes, excluding those in the CORE, that are ancestors of a CORE node.
3. OUT — The set of nodes, excluding those in the CORE, that are descendants of a CORE node.
4. TUBES — The set of nodes, excluding those in the CORE, IN, and OUT, who have an ancestor in IN as well as a descendant in OUT.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig1}
\caption{Illustration of a bow-tie decomposition.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig2}
\caption{Visualization of the darkweb graph.}
\end{figure}
Figure 1: The distribution of the in-degree, out-degree, pagerank, and edgeweights. In Figure 1c we exclude the three domains with the highest pagerank (they are listed in Table 3 in the Appendix) because they are such extreme outliers. For all plots with a log-scale axis, we follow following [8, 13] to use the Fibonacci binning from [20].

5. TENDRILS — Nodes that have an ancestor in IN but do not have a descendant in OUT. Also, nodes that have a descendant in OUT but do not have an ancestor in IN.

6. DISCONNECTED — Everything else.

We compare our results to the www results from [8, 12, 13]. We chose these reference points due to the size of their crawls and the rigor of their analyses. The most obvious difference between the world-web-web and the darkweb is that the darkweb only contains a CORE and an OUT component. We attribute this to the extraordinarily low out-degrees from Figure 1b. For the curious, the top 50 nodes of the CORE are provided in the Appendix (Table 4).

For our third result, we examine the darkweb’s internal connectivity via shortest-path-length (SPL), shown in Figure 3. First, for all pairs of nodes \( \{u, v\} \) in the darkweb, only 8.11% are connected by a directed path from \( u \rightarrow \cdots \rightarrow v \) or \( v \rightarrow \cdots \rightarrow u \). This is drastically lower than the \( \sim 43.42\% \) found in the www [13]. We again attribute this to the low out-degree per Figure 1b. Of the connected pairs, the darkweb’s average shortest path length is 4.35 compared to the 4.27 in the world-wide-web [13]. It’s surprising to see a graph as small as the darkweb have a higher mean SPL than the entire world-wide-web, and is a testament to how sparse the darkweb graph really is. In Figure 3c we plot the distribution of SPLs for the 297 nodes of the CORE, to our surprise, the mean SPL within the CORE is 3.97, only 9% less than the entire darkweb. From this we conclude the CORE is not any kind of densely interconnected core.
Figure 2: Bow-tie decomposition comparing the www versus the darkweb.

| World-Wide-Web | Darkweb |
|----------------|---------|
| CORE           | 22.3M   | 51.94% | 297 | 4.14% |
| IN             | 3.3M    | 7.65%  | 0   | 0.0%  |
| OUT            | 13.3M   | 30.98% | 6881| 95.86%|
| TUBES          | 17k     | .04%   | 0   | 0.0%  |
| TENDRILS       | 514k    | 1.2%   | 0   | 0.0%  |
| DISCONNECTED   | 3.5M    | 8.2%   | 0   | 0.0%  |

Figure 3: Comparing shortest path lengths between the world-wide-web and the darkweb considering directed edges. Whereas in the www 56.65% of node pairs have have \( \infty \) path-length (no path connecting them), in the darkweb 91.89% of node-pairs have no path connecting them. Moreover, even within that 8.11% of pairs with a directed path between them, the darkweb’s average SPL (\( \mu = 4.35 \)) is higher than that of the www (\( \mu = 4.27 \)).

3.1 Robustness and Fragility

In Figure 4 we show our fourth result—how quickly the entire network (WCC) as well as the CORE disintegrates under node removal. In Figures 4a and 4b we see the familiar resistance to random failure yoked with fragility to targeted attacks in the spirit of [22]. In Figure 4b we see that, unlike the www [13], the WCC is more susceptible to high in-degree deletions than the CORE. This elaborates the view from Figure 3c that the CORE is, in addition to not being strongly interconnected, is also not any kind of high in-degree nexus.

Figures 4c and 4d show the breakdown when removing central nodes. In Figure 4c the CORE is largely unaffected by low centrality deletions. In Figure 4d we see that although the CORE isn’t disproportionately held together by high in-degree nodes, it is dominated by very central nodes.
Comparing Figures 4b and 4f, we see the CORE relative to the entire network consists of more high-pagerank nodes than high in-degree nodes. This implies CORE nodes are not created by their high-indegree (Figure 4b), but by their high centrality, amply corroborated by Figures 4c and 4d. Likewise, Figure 4e recapitulates Figure 4a that nodes with especially low in-degree or centrality are, unsurprisingly, not in the CORE.

Figure 4: Deleting nodes from the darkweb graph and seeing how quickly the WCC and CORE disintegrate. In all plots, we shuffled the order of nodes with the same value until reaching stable statistics, e.g., in Figure 1c, 98% of nodes are tied for the lowest pagerank; so when removing only 10% of the nodes (e.g., Figure 4e), it’s ambiguous which nodes should be deleted first. So in our analysis we shuffled the ordering of the nodes with the same value and recomputed sizes of the WCC/CORE until the median was stable.

Given the hub-and-spoke nature of the graph, if the goal was destroy the connectivity of the hyperlink graph, the most effective method would be to attack the central hubs—but which hubs? In the Appendix (Figure 7)
we see the WCC breaks down at roughly the same rate when removing either the high in-degree or high out-degree nodes.

### 3.2 Reciprocal Connections

In [12] they stress the importance of reciprocal connections in maintaining the www’s graph properties. We compute two of their measures. First, we compute [12]’s measure $\langle k_{in}k_{out} \rangle = \frac{2\langle k_{in}\rangle\langle k_{out} \rangle}{\langle k_{in} \rangle \langle k_{out} \rangle}$, to quantify in-degree and out-degree’s deviation from independence. For the darkweb, we arrive at $\langle k_{in}k_{out} \rangle = 3.70$. This is in the middle of the road of prior estimates of the www, and means that the out-degree and in-degree are positively correlated. To be a better view, we plot the average out-degree as a function of the in-degree, given as,

$$\langle k_{out}(k_{in}) \rangle = \frac{1}{N_{k_{in}}} \sum_{i \in \Omega(k_{in})} k_{out,i} ,$$

which is simply “For all nodes of a given in-degree, what is the mean out-degree?” The results are given in Figure 5. In short, in the darkweb there’s no obvious pattern to the relationship between in-degree and out-degree, and is in fact mostly flat.

![Graph with red outliers](image)

(a) With red outliers

![Graph without red outliers](image)

(b) Without red outliers

Figure 5: Average out-degree as a function of the in-degree. Figure 5b is the same as 5a but pruned of the three red outliers in 5a.

### 4 Discussion

Casual inspection of Figure 1 shows that the darkweb is indeed a very different graph than what in the worldwide web [3–8, 12]—particularly, the glaring fact of so little linking to other websites (Figure 1b). The fundamental question is why there’s so little linking. For this, the two major explanations are:

- **The technological explanation.** In the darkweb, sites go up and go down all the time. Why bother linking if there’s little chance that the destination will still exist?
- **The social explanation.** As-is, people who create sites on the darkweb are cut from a different social cloth than those who create sites on the www.

To disambiguate these two we performed a second crawl collecting instances of darkweb sites linking to the www and compared the relative rates of outbound linking, resulting in Figure 6. From the essentially equal rates of outbound linking to the www as well as the darkweb, we conclude:

1. The low outbound linking is not due to the notorious impermanence of onion sites.
2. If onionsites got drastically more stable, we would still see very low rates of linking.
3. By elimination of the technological explanation, we suggest that people creating darkweb sites are, on average, simply less social than those creating sites on the www.
The term “dark web” is commonplace, but based on our analysis, the “web” is a misnomer. It is more accurate to view it as a set of dark silos. Unlike the www, the darkweb is a place of isolation. In Table 1 we summarize our comparison statistics between the www and the darkweb.

![Graph showing the percentage of darkweb sites linking to the www versus linking to other darkweb sites.](image)

Figure 6: Comparing the rates of the darkweb sites linking to the www versus linking to other darkweb sites, we see they are essentially the same.

| Measure                          | www  | darkweb  |
|----------------------------------|------|----------|
| Num nodes                        | 43M  | 7,178    |
| Num edges                        | 623M | 25,104   |
| Prop. connected pairs            | ~42% | 8.11%    |
| Average SPL                      | 4.27 | 4.35     |
| Edges per node                   | 14.5 | 3.50     |
| Network diameter*                | 48   | 5        |
| Harmonic diameter                | ~9.55| 232.49   |

Table 1: Summarized network level properties between the www and the darkweb. Asterisk for the entries requiring conversion to an undirected graph.

References

[1] Borgatti SP, Jones C, Everett MG (1998) Network measures of social capital. Connections 21: 27–36.
[2] Wu P, Wen JR, Liu H, Ma WY (2006) Query selection techniques for efficient crawling of structured web sources. In: Data Engineering, 2006. ICDE’06. Proceedings of the 22nd International Conference on. IEEE, pp. 47–47.
[3] Barabási AL, Albert R (1999) Emergence of scaling in random networks. science 286: 509–512.
[4] Barabási AL, Albert R, Jeong H (2000) Scale-free characteristics of random networks: the topology of the world-wide web. Physica A: Statistical Mechanics and its Applications 281: 69–77.
[5] Adamic LA, Huberman BA (2000) Power-law distribution of the world wide web. Science 287: 2115–2115.
[6] Kumar R, Raghavan P, Rajagopalan S, Sivakumar D, Tompkins A, et al. (2000) The web as a graph. In: Proceedings of the nineteenth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems. ACM, pp. 1–10.
[7] Albert-László Barabási J (2001) The physics of the web. Physics Web http://www.physicsweb.org/article/world/14/7/09.

[8] Meusel R, Vigna S, Lehmburg O, Bizer C (2015) The graph structure in the web—analyzed on different aggregation levels. The Journal of Web Science 1.

[9] Appelbaum J, Muffet A (2015) The “.onion” special-use domain name. Internet Engineering Task Force RFC 7686.

[10] Everton SF (2012) Disrupting dark networks, volume 34. Cambridge University Press.

[11] De Domenico M, Arenas A (2017) Modeling structure and resilience of the dark network. Phys Rev E 95: 022313.

[12] Serrano MA, Maguitman A, Boguñá M, Fortunato S, Vespignani A (2007) Decoding the structure of the www: A comparative analysis of web crawls. ACM Trans Web 1.

[13] Lehmburg O, Meusel R, Bizer C (2014) Graph structure in the web: aggregated by pay-level domain. In: Proceedings of the 2014 ACM conference on Web science. ACM, pp. 119–128.

[14] Project TT (2017). Tormetrics — unique .onion addresses. URL https://metrics.torproject.org/hidserv-dir-onions-seen.html

[15] Wikipedia (2016). Torchat — wikipedia, the free encyclopedia. URL https://en.wikipedia.org/w/index.php?title=TorChat&oldid=728783759 [Online; accessed 28-February-2017].

[16] Wikipedia (2017). Tor (anonymity network) — wikipedia, the free encyclopedia. URL https://en.wikipedia.org/w/index.php?title=Tor_(anonymity_network)#Tor_Messenger&oldid=767770868 [Online; accessed 28-February-2017].

[17] Wikipedia (2016). Ricochet (software) — wikipedia, the free encyclopedia. URL https://en.wikipedia.org/w/index.php?title=Ricochet_(software)&oldid=755044411 [Online; accessed 28-February-2017].

[18] Clauset A, Shalizi CR, Newman ME (2009) Power-law distributions in empirical data. SIAM review 51: 661–703.

[19] Page L, Brin S, Motwani R, Winograd T (1999) The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab. URL http://ilpubs.stanford.edu:8090/422/ Previous number = SIDL-WP-1999-0120.

[20] Vigna S (2013) Fibonacci binning. CoRR abs/1312.3749.

[21] Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, et al. (2000) Graph structure in the web. Computer networks 33: 309–320.

[22] Bollobás B, Riordan O (2004) Robustness and vulnerability of scale-free random graphs. Internet Mathematics 1: 1–35.

[23] Biryukov A, Pustogarov I, Weinmann RP (2013) Trawling for tor hidden services: Detection, measurement, deanonymization. In: Security and Privacy (SP), 2013 IEEE Symposium on. IEEE, pp. 80–94.

Appendix

A Miscellaneous figures
| Domain               | Title                                              | Out-degree | In-degree |
|---------------------|----------------------------------------------------|------------|-----------|
| directoryvi6plzm    | Tor Directory: A list of onion                    | 5582       | 1         |
| visitorfi5kl7q7i    | VisiTOR Search Engine - Tor Hi                    | 4367       | 18        |
| skunksworkedp2cg    | A portal containing lists of l                     | 2709       | 41        |
| createdvnn5z57xhl    | The onion crate REBOOT - Tor h                     | 2758       | 6         |
| gxamjbnu7uknahng     | The Uncensored Hidden Wiki New                   | 848        | 13        |
| torvps7kzis5uujfz    | TorVPS                                             | 498        | 33        |
| zla1a32teyptf4tvi    | Fresh Onions                                       | 478        | 4         |
| hwikis25cffertqe     | Hidden Wiki                                       | 309        | 15        |
| w363zoq3ylux5rf5     | Galaxy2 Social network                             | 294        | 32        |
| y4yhc17273s2yeqk     | 조선위키                                            | 203        | 9         |
| auutvvtpt2zktxwng    | Onion Dir                                          | 186        | 25        |
| wikiwarixvouhyn      | Light version of original Hidd                     | 180        | 7         |
| hdwikicorlcsiy       | HD Wiki                                            | 168        | 22        |
| ntciuxlmm4275vi      | Hidden Wiki - Outdated and fil                     | 150        | 6         |
| soupkso3la221t13     | This site contains an onion bl                     | 146        | 6         |
| torwikignouempf      | TorWiki                                            | 128        | 18        |
| vxzzqddt54racf3i     | Tor2Dir: A onion sites list                        | 119        | 3         |
| godnotaba36dsabv     | Годнотаба — мониторинг годноты                      | 115        | 5         |
| pduvohnvbtzph6sg     | pduvohnvbtzph6sg.onion                            | 103        | 1         |
| rbaco5flcou46wpd     | Welcome To Dark Web Links & M                      | 102        | 5         |

Table 2: The top 15 out-degree hubs on the darkweb.

| PageRank | In-degree | Harmonic Centrality |
|----------|-----------|---------------------|
| Freedom Hosting II | Freedom Hosting II | Freedom Hosting II |
| Blockchain - Bitcoin Bloc | Blockchain - Bitcoin Bloc | Blockchain - Bitcoin Bloc |
| Outlaw Market | DuckDuckGo - Search engin | DuckDuckGo - Search engin |
| Tor Links | SIGAINT | SIGAINT |
| Dream Market - Dark web m | TORCH: Tor Search Engine | TORCH: Tor Search Engine |
| Tor Ads | Tor Ads | Tor Ads |
| Valhalla Market | Tor Shops | Tor Shops |
| Silkkitie Market | main.paraZite # Anarchy | main.paraZite # Anarchy |
| DreamMarket Forum | Dream Market - Dark web m | Dream Market - Dark web m |
| TOR FREE SPEECH:ask anyth | On my website you can upl | On my website you can upl |
| On my website you can upl | Alphabay Market | Alphabay Market |
| Alphabay Market | TorBox - Mailservice | TorBox - Mailservice |
| Скрытые Ответы | The Hidden Wiki - Outdate | The Hidden Wiki - Outdate |
| Hidden Answers pt | On my website you can upl | On my website you can upl |
| HD Wiki | OnionWallet | OnionWallet |
| Grams | HTTP File Server | HTTP File Server |
| DeepDotWeb - Surfacing th | The Pirate Bay - Torrent | The Pirate Bay - Torrent |
| OnionDir - Deep Web Link | A portal containing lists | A portal containing lists |
| The Darknet Company | Example rendezvous points | Example rendezvous points |
| OUTLAW Market | USA Citizenship | USA Citizenship |
| Carding Community | not Evil - Search Tor | not Evil - Search Tor |
| USA Citizenship | Example rendezvous points | Example rendezvous points |
| **Deep Web Radio** | **Deep Web Radio** | **Deep Web Radio** |

Table 3: The top 20 sites of the darkweb using PageRank, In-degree, and Harmonic Centrality. Bolded entries are present across all three lists. † entries are in the CORE.
| Address                  | Name                                                                 | Pagerank  | In-degree | Out-degree |
|-------------------------|----------------------------------------------------------------------|-----------|-----------|------------|
| shopsat2dotfotbs        | TorShops                                                              | 6.30e-03  | 56        | 1          |
| torlinkbgp8sabnbs       | TorLinks — .onion Link List                                          | 6.00e-03  | 54        | 43         |
| lchudifyemq4ldjj        | Dream Market - Dark web market featuring                             | 5.59e-03  | 54        | 5          |
| valhallaxynm3fydu        | Valhalla Market                                                       | 3.75e-03  | 23        | 3          |
| slikkit1ehdg5mug        | Silkkit1 Market                                                       | 3.58e-03  | 28        | 7          |
| tmskhzavkcdufbr          | DreamMarket Forum                                                    | 3.00e-03  | 26        | 6          |
| jd6yhucivehvd4t          | Dream Market - Dark web market featuring                             | 2.39e-03  | 23        | 5          |
| t3e6ly3uoi4f2zw2         | Dream Market - Dark web market featuring                             | 2.39e-03  | 23        | 5          |
| 7ep7acrkunu3cw31         | Dream Market - Dark web market featuring                             | 2.28e-03  | 17        | 4          |
| tt3j2xk45ycaaa5zt        | On my website you can upload/download fi                            | 1.85e-03  | 42        | 53         |
| hdwikicor1dcs1y          | HD Wiki                                                               | 1.60e-03  | 22        | 168        |
| grams7enuf17mjd1         | Grams                                                                 | 1.60e-03  | 60        | 6          |
| deepdot35wvmedy5         | DeepDotWeb - Surfacing the News From The                             | 1.57e-03  | 28        | 13         |
| dirnxrxdraygb1fge        | OnionDir - Deep Web Link Directory                                    | 1.53e-03  | 28        | 65         |
| n2ha26qplph45ae6         | Welcome To A New Site                                                | 1.28e-03  | 7         | 1          |
| rbaco5f1cou46wpd         | Welcome To Dark Web Links & More!                                    | 1.23e-03  | 5         | 102        |
| radioicsb12q27toh        | Rádio CBS — Comunicações Brasileira de S                             | 9.85e-04  | 8         | 12         |
| 1wp1xqzvmgu3uff           | Runion - Russian Forum                                               | 8.58e-04  | 13        | 5          |
| hansamkt2rr6nfg3         | HANSA Market                                                          | 8.36e-04  | 34        | 2          |
| wallstiyijhkrvmj         | Wall Street Market                                                    | 7.99e-04  | 11        | 2          |
| x363zoq3yux5rf5          | Galaxy2 Social network                                               | 7.47e-04  | 32        | 294        |
| kpynyyym6xq17wzz          | main.piaraZite # Anarchy files and Underg                            | 7.13e-04  | 56        | 71         |
| bankshopioe013mv         | Store                                                                 | 7.00e-04  | 5         | 2          |
| zqktl1w14fev06ri          | The Hidden Wiki - Outdated, full of scam                             | 6.95e-04  | 45        | 145        |
| chattorucl7bcgygp        | This is ChatTor, the only 100% anonymous                              | 6.75e-04  | 16        | 3          |
| xmbh57jrrwm6insl         | TORCH: Tor Search Engine                                             | 6.35e-04  | 57        | 9          |
| slpw1klryjyjhc3t         | SleepWalker                                                           | 6.12e-04  | 7         | 2          |
| cocaineo5z66aluy         | Concerned Cocaine Citizens                                           | 6.04e-04  | 4         | 5          |
| darkmarabrstwfhf         | Darkmarket Market                                                     | 5.78e-04  | 7         | 2          |
| x7bwusmcore5fmx56         | Darknet Hacking Services                                             | 4.64e-04  | 2         | 1          |
| torvps7kzis5ujfzf         | TorVPS                                                                | 4.49e-04  | 33        | 498        |
| wikitorcogstafsfs        | The Hidden Wiki                                                       | 4.44e-04  | 11        | 222        |
| ntcixulmnm4275vi          | Hidden Wiki - Outdated and filled with s                              | 4.40e-04  | 6         | 150        |
| mystorew25hgytln         | Store                                                                 | 4.35e-04  | 7         | 6          |
| fantomw4f1uxar7u          | Fantom urls - Forum for paranoids                                     | 3.85e-04  | 4         | 1          |
| xfwnyig7olydpq5r          | USA Citizenship                                                       | 3.67e-04  | 40        | 1          |
| vfqnd6miecqyq1t          | UK Passports                                                          | 3.55e-04  | 36        | 1          |
| deeplinkdeatbm17          | DeepLink                                                              | 3.35e-04  | 4         | 11         |
| abbujjh5vt9q7th6g        | Onion Identity                                                        | 3.35e-04  | 31        | 1          |
| 54ogum7gwxhtgiyaa        | Krang Hidden Base in Tor. Technodrome. B                              | 3.28e-04  | 27        | 18         |
| costeiraizb2xecgs        | costeira.i2p.onion - Servidor de download                            | 3.21e-04  | 7         | 3          |
| allyour4nert7pkh          | A portal containing lists of links autom                              | 3.11e-04  | 41        | 2769       |
| lin4kskg1ymtyszdb         | LINKS Onion Web Link Directory - Your co                              | 3.15e-04  | 10        | 94         |
| skunksworkepd29c          | A portal containing lists of links autom                              | 3.11e-04  | 41        | 2769       |
| tfwd13izigxl1ure          | Apples 4 Bitcoin                                                      | 3.08e-04  | 39        | 1          |
| fdwocbsnt6vzwd           | French Deep Web                                                       | 3.06e-04  | 26        | 8          |
| roothiletpeyjiirta       | PT-BR: Site oficial do roothit EN-US: We                              | 3.02e-04  | 7         | 1          |
| tuu66yyvrnm3sotlf4       | UK Guns and Ammo Store                                                | 3.01e-04  | 31        | 1          |
| 3dbr5t4pgahedms          | ccpal - css - cvv2s - paypal                                         | 3.00e-04  | 31        | 1          |
| y3fpieiezy2sind4          | HQER - High Quality Euro Counterfeits                                 | 3.00e-04  | 31        | 1          |

Table 4: Top 50 domains by Pagerank in the CORE.
Figure 7: We see that removing the high in-degree and out-degree nodes are about equally effective in breaking the WCC.

B Peering beneath the Darkweb

Every darkweb domain can be accessed by a site-specified number of “Introduction Points”. These introduction points and hidden service directories have been explored before [23], but as far as we know, never statistically. In our analysis, we found that 96.8% of darkweb sites have exactly ≤ 3 introduction points (with 95% having exactly 3). The most salient fact is that if someone wanted to take down a darkweb site (presumably via DDOS), 97% of the time the attack would require taking out three publicly known relays. From 8b, the number of domains served from each Introduction Point decays roughly exponentially.

Figure 8: Number of domains found on each HSDir